Niranjan Sapkota

Essays on the New Blockchain-Based Digital Financial Market

Risks and Opportunities

ACTA WASAENSIA 491

Vaasan yliopisto
UNIVERSITY OF VAASA

ISBN 978-952-395-032-0 (print)
978-952-395-033-7 (online)

ISSN 0355-2667 (Acta Wasaensia 491, print)
2323-9123 (Acta Wasaensia 491, online)

URN http://urn.fi/URN:ISBN:978-952-395-033-7

Hansaprint Oy, Turenki, 2022.

ACADEMIC DISSERTATION

To be presented, with the permission of the Board of the School of Accounting and Finance of the University of Vaasa, for public examination in 14th of October, 2022, at noon.

Article based dissertation, School of Accounting and Finance, Finance

Author Niranjan Sapkota https://orcid.org/0000-0002-6549-8685

Supervisor(s) Professor Timo Rothovius
University of Vaasa. School of Accounting and Finance, Finance.

Professor Jussi Nikkinen
University of Vaasa. School of Accounting and Finance, Finance.

Associate Professor Klaus Grobys
University of Vaasa. School of Accounting and Finance, Finance.

Custos Professor Timo Rothovius
University of Vaasa. School of Accounting and Finance, Finance.

Reviewers Professor John Paul Broussard
University of Oklahoma. Price College of Business, Finance.

Professor John Goodell
University of Akron. College of Business, Finance.

Opponent Professor John Paul Broussard
University of Oklahoma. Price College of Business, Finance.

Tiivistelmä

Tämä väitöskirja koostuu viidestä esseestä, jotka käsittelevät uuden lohkoketjupohjaisen digitaalisen rahoitusmarkkinan riskejä ja mahdollisuuksia. Väitöskirjan tarkoituksena on analysoida, tunnistaa ja mahdollisuuksien mukaan ennustaa joitakin lohkoketjupohjaisten digitaalisten varojen markkinoiden suurimpia riskejä. Siinä analysoidaan, miten kryptovaluuttakohtaiset ominaisuudet liittyvät vakavaraisuusriskiin, kestävyysriskiin, eristäytymisriskiin ja sentimenttiriskiin. Tämän lisäksi se valottaa myös tämän rahoitusinnovaation mahdollisuuksia.

Tämän väitöskirjan ensimmäisessä esseessä keskitytään erityisesti kryptovaluuttaan maksukyvyttömyysriskinä. Tässä tutkimuksessa keskitytään muuttujiin, jotka ovat sijoittajan saatavilla korkeintaan 1 kuukausi sen jälkeen, kun kaupankäynti kryptovaluutalla on alkanut. Tämän tutkimuksen tulokset osoittavat, että kryptovaluuttojen konkurssit ovat ennustettavissa. Toisessa esseessä tutkitaan energiariskiä markkinoita ohjaavana voimana kryptovaluutan hinnoittelussa. Kryptovaluutat, jotka käyttävät paljon energiaa kuluttavaa konsensusprotokollaa, ovat muita riskialttiimpia, koska niiden louhintakustannukset ovat alttiimpia energian hinnan muutoksille. Yllättäen tutkimuksessa todetaan, että energiankulutuksella ei näytä olevan merkitystä kryptovaluuttojen hinnoittelussa. Kolmannessa esseessä hypoteesina on, että yksityisyyskolikot muodostavat erillisen alamarkkinan kryptovaluuttamarkkinoilla, ja tutkimus tarkastelee näiden eristäytymisriskiä. Siinä osoitetaan, että yksityisyyskolikot ja ei-yksityisyyskolikot ovat kaksi erillistä omaisuuserämarkkinaa kryptovaluuttamarkkinoilla. Neljäs essee käsittelee uutismedian sentimenttiriskiä. Siinä tutkitaan, vaikuttaako uutismedian sentimentti Bitcoinin volatiliteettiin. Siinä myös erotetaan toisistaan taloudellinen sentimentti ja psykologinen sentimentti ja todetaan, että taloudellisesti optimistiset sijoittajat ohjaavat Bitcoin-markkinoita.

Väitöskirjan viidennessä esseessä analysoidaan mahdollisuuksia, erityisesti rahoitusmahdollisuuksi, liittyen laajalti tunnettuihin digitaalisiin tokeneihin. Siinä havaitaan, että näihin omaisuuseriin sijoittavat sijoittajat toimivat pitkälti tunteidensa ohjaamina sijoituspäätöksiä tehdessään. Yllättävää kyllä, sääntelykehyksestä ei ole vielä tullut poliittisten päättäjien prioriteettia. Siksi tämä väitöskirja ei ainoastaan tue tulevaa tutkimusta, vaan auttaa myös viranomaisia lohkoketjupohjaisten rahoitusteknologioiden tulevaisuuden määrittelyssä.

Asiasanat: Lohkoketju, Digitaaliset Rahoitusmarkkinat, Kryptovaluutat, Riski, Kolikkoanti , Mahdollisuudet, Finanssiteknologia

Abstract

This doctoral thesis consists of five original essays on the risks and opportunities of the new blockchain-based digital financial market. The purpose of this dissertation is to analyze, identify, and, if possible, predict some of the major risks in the market for blockchain-based digital assets. It analyzes how crypto-specific characteristics are associated with solvency risk, sustainability risk, seclusion risk, and sentiment risk. On top of that, it also sheds light on the opportunity side of this financial innovation.

The first essay of this dissertation specifically focuses on cryptocurrency for solvency risks. To forecast potential cryptocurrency default at an early stage, this study focuses on variables that are part of the information set of the investor 1 month at most after the start of trading for a cryptocurrency. The results of this research show that bankruptcies among cryptocurrencies are predictable. The second essay explores energy risk as a fundamental market-driving force for the pricing of cryptocurrency. Cryptocurrencies using a high-energy-consumption consensus protocol are riskier than others because their mining costs are more exposed to changes in energy price. Surprisingly, the study finds that energy consumption does not seem to play a role in pricing cryptocurrency. The third essay hypothesizes that privacy coins form a distinct submarket in the cryptocurrency market, shedding light on seclusion risk. It shows that privacy coins and non-privacy coins are two distinct asset markets within the cryptocurrency market. The fourth essay is about news media sentiment risk. It explores whether news media sentiments have an impact on Bitcoin volatility. It also differentiates financial sentiment and psychological sentiment and finds that financially optimistic investors are driving the Bitcoin market.

On the other hand, the fifth essay in this dissertation analyzes opportunities, especially the funding opportunity in the widely known category of new digital assets defined as crypto tokens. It analyzes the determinants of the success of initial coin offerings and finds that initial-coin-offering investors are largely guided by their emotions when making investment decisions. Surprisingly, regulatory framework has not yet become a priority among policymakers. Therefore, this doctoral dissertation not only facilitates future research, but also helps regulators in shaping the future of blockchain-based financial technologies.

Keywords: Blockchain, Digital Financial Market, Cryptocurrencies, Risk, Initial Coin Offering (ICO), Opportunities, Financial Technology (FinTech)

ACKNOWLEDGEMENT

Ever since I knew about different levels of academic degree, I had a dream to achieve the highest in a field by obtaining a doctoral degree. I used to believe that one with a PhD degree knows everything. Now I am wise enough to know that learning is a never-ending process. I realized that a PhD is more than just an academic degree. It has taught me many more things besides how to deliver qualitative research. I learned that failure is the ingredient for greater success, that rejection leads to better acceptance, and most importantly, how to constructively handle criticism. I am extremely grateful to many people both inside and outside of academia for directly and indirectly helping me in achieving this huge milestone.

First of all, I would like to express my deepest gratitude to Professor Timo Rothovius and Professor Jussi Nikkinen for accepting me into the doctoral program and for supervising my doctoral study. I cannot begin to express my thanks to Associate Professor Klaus Grobys for his unwavering supervision, guidance, and collaboration on several projects during my PhD study. I would like to extend my deepest appreciation to the pre-examiners of my doctoral dissertation, Dr. John Paul Broussard, Professor of Finance, University of Oklahoma, and Dr. John Goodell, Professor of Finance, University of Akron, for their excellent feedback. I gratefully acknowledge Dean Marko Järvenpää and former Dean Helinä Saarela for their constant support.

Special thanks to the Director, Graduate School of Finance (GSF) and Finnish Doctoral Program in Economics (FDPE) Professor Mikko Leppämäki for all the courses and workshops and Professor Markku Kaustia for the PhD clinics. I appreciate Professor Mika Vaihekoski, Professor Gonul Colak, Associate Professor Petri Jylhä, University Lecturer Joni Kokkonen, Database Specialist Nikolas Beriktof for their comments and criticisms and Ali-Alha Reetta for her assistance in GSF workshops. I would like to thank the former Director of Innolab, Dr Mari K. Niemi, for her support for the "Business Renewal" project. I gratefully acknowledge the Director of the Digital Economy, Professor Heidi Kuusiniemi, for trusting me for the "powered-by-blockchain" project. I would also like to extend my gratitude to Professor Sami Vähämaa for his support during my study and the pre-examination process. Thanks should also go to Professor Janne Äijö for assigning me the suitable teaching responsibilities. I would like to thank Dr. Anupam Dutta, Dr. Tatiana Garanina, Dr. Denis Davydov, Dr. Vanja Piljak, Dr. Tuukka Järvinen, Dr. Nebojsa Dimic, Dr. Elina Haapamäki, Dr. Sara Yasar, Kaisa Penttinen, Shaker Ahmad, Shahbaz Qureshi, Veda Fatmy, and Amir Mobeshari for the friendly working environment. I would also like to thank Dr. Mikko Ranta for the Finnish translation of the abstract.

My sincere thanks go to the Foundation for Economic Education (Liikesivistysrahasto) and Graduate School for funding my PhD study. I would also like to thank the Erkki Paasikivi Foundation for the travel grant. I would like to thank the discussants and the participants, as well as the organizers and the host institutions, for the conferences. I am thankful to the participants of the 28th European Financial Management Association Conference (University of Azores, 2019), Finance, Property, Technology and the Economy Conference (University of South Australia- 2019), 17th Finance, Risk and Accounting Perspectives Conference (Hanken School of Economics, 2019), 3rd Cryptocurrency Research Conference (University of Reading, 2020), World Finance Conference (University of Agder, 2021), International Risk Management Conference (University of Cagliari, 2021), and 4th Cryptocurrency Research Conference (University of Southampton, 2021). I would like to extend my appreciation to Marjaana Tuppi-Hanka for arranging travel and accommodation during these conference trips.

My special thanks go to Simon Moore, Senior Contributor at Forbes, for media coverage of the first essays of this dissertation and Dr. Cristina Polizu, Senior Executive, Methodologies, Standard and Poor's (S&P) Global, for inviting me to present the different sides of cryptocurrency risk and for making my work visible to a larger audience and more practitioners. I gratefully acknowledge Specialist Juuli Honko for her guidance in the pre-examination process of my dissertation.

I am deeply indebted to my parents for their unconditional love and support. I would not have been in this position without the hard work and sacrifices of my father, Devendra Prasad Sapkota, and mother, Saraswati Sapkota. I would like to express my deepest gratitude to my wife, soul mate, best friend, and love of my life, Bimala Chapagain, for everything. She has always been there to boost my confidence and to give me unprecedented love and support. Thank you for making my life so beautiful. A lot of love goes to my son, Orion, and daughter, Azalea (soon to be born) for making this journey memorable. I would like to thank my sisters and their families for their love and care. I would like to remember my great grandfather, Kedarnath Sapkota for his contribution as a founder of the school where I completed my secondary education. I would also like to mention my Finnish Kummi family for integrating me into Finnish culture and making my stay in Finland enjoyable. Finally, I would like to remember everyone who has directly or indirectly helped me to become the person I am today. Thank you all!

Niranjan Sapkota
Vaasa, Finland

Dedication

This work is dedicated to my family.

Contents

Figures

Tables

Essays

Grobys, K., & Sapkota, N. (2020). Predicting cryptocurrency defaults. *Applied Economics*, *52*(46), 5060-5076. https://doi.org/10.1080/00036846.2020.1752903 [1]

Sapkota, N., & Grobys, K. (2019). Blockchain consensus protocols, energy consumption and cryptocurrency prices. *Journal of Energy Markets*, *13(4), 117-139*. https://doi.org/10.21314/JEM.2020.221 [2]

Sapkota, N., & Grobys, K. (2021). Asset market equilibria in cryptocurrency markets: evidence from a study of privacy and non-privacy coins. *Journal of International Financial Markets, Institutions and Money*, *74*, 101402. https://doi.org/10.1016/j.intfin.2021.101402 [3]

Sapkota, N. (2022). News-based sentiment and bitcoin volatility. *International Review of Financial Analysis*, *82*, 102183. https://doi.org/10.1016/j.irfa.2022.102183 [4]

Sapkota, N., & Grobys, K. (2021). Fear sells: On the sentiment deceptions and fund-raising success of initial coin offerings. *Proceedings of the World Finance Conference, WFC 2021; Cryptocurrency Research Conference CRC 2021; and 14th International Risk Management Conference IRMC 2021*

[1] Printed with kind permission of Taylor and Francis
[2] Printed with kind permission of Risk Journals
[3] Printed with kind permission of Elsevier
[4] Printed with kind permission of Elsevier

1 INTRODUCTION

After the United States (US) investment bank giant, Lehman Brothers Holdings Incorporation filed for bankruptcy in 2008, the failure of a systemically important financial institution with more than $700 billion of liabilities shook the entire global financial system and froze global money markets.[5] People started losing trust in the traditional banking system, eventually giving birth to a new class of digital assets without the requirement of any trusted third parties or central banks. Bitcoin, the most popular blockchain-based digital asset known as cryptocurrency, was introduced in November 2008 by its anonymous creator, Satoshi Nakamoto, about 2 months after the Lehman Brothers crisis. This new electronic payment mechanism is a fully peer-to-peer, decentralized system without any financial intermediaries or controlling bodies. Being designed as completely decentralized means that users of the currency do not need to trust in a central authority, such as traditional central banks. Rather than being based on trust, as in a fiat currency system, the governance mechanism of this new blockchain system is based on cryptographic proof that follows certain consensus protocols to manage and store transactions. The concept of digital currency began in the 1980s long before Bitcoin, but all attempts failed.[6] The financial technology (FinTech) landscape shifted with the publication of a Bitcoin white paper in 2008 that presented a peer-to-peer, decentralized system that uses the concept of blockchain. According to Harvey and Ramachandran (2021), the concept of blockchain was not, however, pioneered by Satoshi Nakamoto, but was created by Haber and Stornetta (1991) as a time-stamping system to keep track of different versions of a document. Satoshi Nakamoto combined two key innovations to create Bitcoin: (a) recording and storing transactions (i.e., time stamping or "blockchain") and (b) a consensus mechanism called "proof of work" (PoW) introduced by Back (2002).

According to Nakamoto (2008), commerce on the internet has come to rely almost exclusively on financial institutions serving as trusted third parties to process electronic payments. He stated that the system works well enough for most transactions, but it still suffers from inherent weaknesses in the trust-based model. The central bank must be trusted not to degrade the currency, but history is full of breaches of that trust. Even though this blockchain-based digital financial market is creating significant opportunities, it is not free from numerous weaknesses and

---

[5] https://www.independent.co.uk/news/business/analysis-and-features/financial-crisis-2008-why-lehman-brothers-what-happened-10-years-anniversary-a8531581.html

[6] In the early 1980s, Digicash was founded by American computer scientist and cryptographer David Chaum. Digicash went bankrupt in the 1990s as it failed to persuade banks to embrace its technology.

threats, which this dissertation explores. The paper proposes several insights for understanding blockchain-based FinTech innovations in both theory and practice. Theoretically, it highlights various aspects of risk and opportunity for the new blockchain-based digital financial market. Even though it is evident that FinTech, especially blockchain, have caused a paradigm shift in the financial sector, it comes at significant financial and nonfinancial costs. Therefore, this study attempts to offer theoretical and empirical analyses to aid in understanding the challenges, threats, and benefits of blockchain-based digital innovations like cryptocurrencies and initial coin offerings (ICOs). From a theoretical stance, the cost/benefit and risk/reward analyses of the digital financial innovations examined in this doctoral thesis offer several aspects of consideration, including various perspectives related to their issuance, implementation, and governance.

The practical implications for this study are specifically related to helping and supporting decision-makers and policymakers, such as central banks and other financial organizations, to improve the governance and regulation of these financial instruments. In particular, this study can offer much-needed guidance to help establish various governance practices and policies to facilitate the successful adoption of cryptocurrencies and tokens within the finance sector. Moreover, this study also benefits practitioners, especially cryptocurrency traders, in avoiding investment in potential default cryptocurrencies by understanding in detail what features within cryptocurrencies make them vulnerable. This study also presents opportunities to businesses, including entrepreneurs, to explore blockchain-based digital innovations as a new tool to finance startup projects. Unfortunately, given the existence of several risks, even after more than a decade in the financial world, the adoption of FinTech innovations like cryptocurrency still represents less than 4% of the global population.[7] This dissertation aims to widen the understanding of the risks and opportunities of this new market and to help increase the adoption of cryptocurrencies and ICOs. Furthermore, the key findings of this study can guide others toward adopting empirically-based work and research within this identified FinTech context. Moreover, this emerging field within finance is still in its crude phase, so the analytical literature perspective and empirical analysis are carried out by identifying key factors related to cost, benefit, risk, and opportunity in the context of digital finance following the literature on cryptocurrencies like Bitcoin.

The recent blockchain revolution started when Satoshi Nakamoto (presumed pseudonym) developed the Bitcoin blockchain in 2008. Thereafter and ever since, many startups, individuals, and developers have issued versions of cryptocurrency.

[7] https://blockchain.news/analysis/crypto-users-stand-at-300m-representing-nearly-3.7-percent-the-global-population

In this unregulated and decentralized financial market, even individuals and institutions are issuing their own cryptocurrencies, causing the number of cryptocurrencies to increase exponentially—with approximately 15,000 traded on more than 400 exchanges around the world.[8] Unfortunately, almost half of these have defaulted over a short period of time. Because of the enormous amount of capital involved, one cannot ignore the fact that cryptocurrency default is a major risk of the new digital financial market, as stated by Fry and Cheah (2016). In addition to the default risk, in contrast to traditional investments, cryptocurrencies carry different risks. For instance, Hileman and Rauchs (2017) reported that the chance of cryptocurrency exchanges being hacked is 74 to 79%. Taking the legal perspective, Kethineni and Cao (2019) argued that cryptocurrencies have become the currency of choice for many drug dealers and extortionists because of the ability to hide behind their presumed privacy and anonymity. Maume (2020) highlighted that the potential lack of regulation and enforcement is particularly tempting for scammers and other miscreants.

In the course of adopting new technologies such as blockchain, one must be prepared to tackle numerous associated risks. The risk may vary based on the sector in which blockchain is being implemented, but in most cases, technical, operational, regulatory, legal, and reputational risks are the common types of risk that may arise. Operational risks include hardware and software failure, service disruption, and compromised system and database protection mainly arising from external hackers. In addition, clients may (un)intentionally misuse the platform, leading to increased operational risk. Overall, inadequate technical and operational controls, policies, and procedures can create operational and security risks that can lead to theft and fraud in a digital environment. All of these risks are critical, as relationships in the sector are mostly based on trust in the technology, as posited by Osmani et al. (2020).

Besides the operational risk, there is a major debate on the sustainability and the environmental risk of cryptocurrencies like Bitcoin due to heavy energy consumption by their governance. The PoW-mining process of cryptocurrencies like Bitcoin consumes an enormous amount of energy to maintain their distributed ledger system, the blockchain. The main cost for cryptocurrency is the operational cost of running hardware, which mainly includes energy costs. There has been much debate on the total energy consumption of Bitcoin mining, especially because not all cryptocurrency-mining facilities use renewable energy. De Vries (2019) found that Bitcoin consumed as much electrical energy as all of Hungary in 2018. Also in 2018, Bitcoin generated as much electronic waste as a small nation, such as Luxembourg. Therefore, fundamental economic risk factors like

[8] See more at https://coinmarketcap.com/

environmental risk, energy price shock, cybercrime, governance failure, and financial crisis are persistent in the market. Undoubtedly, even advanced and sophisticated financial innovations like cryptocurrencies cannot overcome these risks. Based on the economic fundamentals of risk and return, one could easily argue that cryptocurrencies should compensate the risk holders associated with the energy price shock with higher returns, which can be observed if past returns reflect the compensation for energy risk or not (e.g., Hayes, 2017; Li and Wang, 2017; Symitsi and Chalvatzis, 2018; Dimitri, 2017; Bendiksen et al., 2018).

In addition to energy risk, there is growing concern over the governance, or regulatory, risk of cryptocurrency. The majority of risks threatening global economies have existed for centuries, but newer threats, like cybercrime, continue to emerge and develop along with the emergence of FinTech innovation. The risk of cyberattack and other criminal activity previously carried out using nonprivacy coins like Bitcoin via the dark web has increased significantly in the normal web environment, too, after the emergence of privacy coins like Monero and DASH (e.g., Brenig et al., 2015; Kethineni and Cao, 2019; Maume, 2020). One reason for the growing usage of privacy coins could be the complete transparency of nonprivacy coins like Bitcoin and Ethereum. A certain degree of financial privacy regarding transactions and balances is important for both institutional and individual users. However, complete anonymity offered by some privacy coins might be an imposing threat to the global economy due to the borderless, or global, nature of cryptocurrency.

Next, because of the growing use of social-networking sites and mobile applications like Facebook and Twitter, the power of news media has increased significantly; it can easily reach a global audience in a short period of time. Similarly, the audience of online news media is also growing because of increasing access to the internet around the globe. The worldwide coverage of news offers both benefits and drawbacks because news can be either positive or negative. Media sentiment risk is emerging as an essential factor to consider in the digital financial market. For example, in the past decade, cryptocurrencies like Bitcoin have made several news headlines in mainstream media. While many newspapers have covered Bitcoin as a possible scam or a bubble, some newspapers have highlighted the opportunities it has created. News about hackings, crypto-exchange collapses, government bans, regulations, taxes, scams, etc. have made many headlines in global news media outlets. According to 99bitcoins.com, Bitcoin has "died" 432 times in the news.[9] Nonetheless, there has also been positive news about Bitcoin such as a legal tender, means of payment, futures, exchange-traded funds, etc. The Bitcoin market has reacted to the good and bad news with upward

[9]See the details at https://99bitcoins.com/bitcoin-obituaries/

and downward swings. Capturing the true sentiment generated from the news has become an important factor to consider in digital finance. Thankfully, we can now easily quantify news and articles such that they can provide more accurate, more efficient proxies for investor sentiment (e.g., Ho et al., 2018; Shen et al., 2018).

While one cannot ignore different types of risk associated with the new digital financial market, one also needs to highlight the immense opportunities brought to financial markets by this technology. Recently, ICOs have received considerable attention as a new form of crowdfunding based on blockchain technology, in which startups sell their tokens in exchange for cryptocurrencies. Recent research has documented that more than $30 billion has been raised via the ICO market (Sapkota, Grobys, and Dufitinema, 2020; Howell, Niessner, and Yermack, 2020). Given their nature as unregulated offerings of digital tokens on the internet aiming to collect funding for a project, ICOs disintermediate any external platform, payment agent, or professional investor and, thus, disrupt the current financial system [i.e., the market for initial public offerings (IPOs)]. In corporate finance, an IPO has several requirements, such as a good track record of earnings above a minimum earnings threshold, whereas other financial criteria are set by the exchange on which the firm plan is listed. Generally, anyone who has an innovative idea or is willing to create a company is eligible to issue an ICO. One could even argue that companies that are financially qualified for an IPO are actually overqualified for an ICO.

Looking at the big picture, the purpose of this doctoral dissertation is to analyze, identify, and predict some of the major risks in the market for cryptocurrencies. It analyzes how crypto-specific characteristics are associated with default risk, regulatory risk, media risk, and risk related to energy consumption and sustainability. Furthermore, financial institutions are one of the world's fastest-growing competitive sectors. Unsurprisingly, their regulatory framework has recently become a priority among policymakers. This doctoral dissertation not only facilitates future research, but also helps regulators in shaping the future of FinTechs.

This dissertation contributes to the ongoing recent discussion on different types of risk and opportunity for the new blockchain-based digital financial market, either by identifying or predicting them. In doing so, essays (1), (2), (3), and (4) study default/solvency risk, sustainability/energy/environment risk, seclusion and systematic/governance/regulatory risk, and sentiment/media risk, respectively, whereas essay (5) studies the opportunities in this market, especially focusing on the ICO as a digital crowdfunding tool.

The first essay exclusively focuses on the first category of new digital assets defined as cryptocurrency. It explores cryptocurrency-specific variables that are accessible to the naïve investor. To forecast potential cryptocurrency defaults at an early stage, this study focuses on variables that are a part of the information set of the investor 1 month at most after a cryptocurrency starts trading. The results of this research show that bankruptcies among cryptocurrencies are predictable. The results from this study agree with the literature on predicting firm bankruptcy (Altman, 1968, 2000, 2002; Altman, Haldeman, and Narayanan, 1977; Altman, Hartzell and Peck, 1995; Lugovskaya, 2010).

The second essay, extending single-asset and single-consensus, studies (Hayas, 2017; Li and Wang, 2017; Bendiksen et al., 2018; Biais et al., 2018; Symitsi and Chalvatzis, 2018; Li, et al., 2019) follows a portfolio approach to explore energy as a fundamental market-driving force for the pricing of cryptocurrency. Because cryptocurrencies incorporating the PoW consensus protocol are riskier than other groups because their mining costs are more exposed to changes in energy price, finance theory suggests that they should compensate investors with higher returns. Hence, the study tests whether the differences between those average portfolio returns are statistically different from each other. Surprisingly, it finds that energy consumption does not seem to play a role in pricing cryptocurrency.

The third essay hypothesizes that privacy coins form a distinct submarket in the cryptocurrency market and highlights the risk of seclusion between users demanding privacy and users demanding transparency. Following Urquhart (2016), Dyhrberg (2016), and Bouri, Molnár, Azzi, Roubaud, and Hagfors (2017), who adopted the perspective that cryptocurrency is an asset market, the privacy and nonprivacy coin markets are considered to be two different asset markets. The study explores whether market equilibria exist in the cryptocurrency market, which, in turn, would imply the existence of submarkets. In this regard, it explicitly tests whether privacy coins form such a distinct submarket in the cryptocurrency market. In doing so, it also investigates the efficiency of the overall cryptocurrency market and shows that the asset market equilibrium of privacy coins appears to originate from privacy coins with the highest market capitalizations. The reason for this finding could be that privacy coins may be the first choice for criminals who might prefer cryptocurrencies exhibiting high levels of anonymity and liquidity.

The fourth essay is about modeling Bitcoin volatility by incorporating sentiment risk. As an accurate estimation of volatility is vital for investors to develop an adequate strategy to hedge potential risks associated with an investment, the study explores whether news media sentiments have an impact on Bitcoin volatility. It extends Corsi's (2009) heterogeneous autoregressive realized volatility (HAR-RV) model with news-based sentiments as additional explanatory variables with the past RVs of Bitcoin to predict its future RVs. It includes different range-based volatility estimation methods for a better understanding of the range and significance in forecasting future volatilities. Furthermore, it also differentiates financial sentiment and psychological sentiment cached in the news and their impact on Bitcoin volatility in different time horizons to capture the heterogeneity of news arrival time and sentiment memory lengths among investors. Moreover, to explore the sentiments of different human emotions, it also extends the HAR-RV model to the emotional level.

Finally, the fifth essay in this dissertation analyzes the funding opportunities in the widely known category of new digital assets defined as cryptotokens in the emerging FinTech market. FinTech is helping businesses, entrepreneurs, and customers to have better control of their financial activities, procedures, and lives by reducing frictions and other costs for intermediaries. As many digital crowdfunding projects are unable to generate enough funding for startup, this dissertation highlights the factors affecting successful fundraising. It explores opportunities in the new digital financial market, and in doing so, it analyzes cross-sectionally the determinants of ICO success as measured by the amount of raised funds. It finds that ICO investors are largely guided by their emotions when making investment decisions. Furthermore, the study finds that a higher assessed riskiness of an ICO project turning out to be a scam does not lower the predicted amount of raised funding. Unlike documented in earlier literature (e.g., Meyer and Ante, 2020; Amsden and Schweizer, 2018; Fish, 2019; Howell et al., 2020), this study shows the weak association between quality signals and readability of ICO white papers and ICO success.

The new blockchain-based digital finance, or FinTech, might become a prime example of how digital innovation can be revolutionary for the whole finance industry, one that drives nearly every other field around the globe. By utilizing the current state of the art in this specific research domain, this study proposes a framework for benefits, opportunities, costs, and risks. It outlines several challenges that affect the adoption of blockchain-based FinTechs; future research should look into providing solutions for these issues.

This doctoral dissertation consists of the introduction chapter and five essays. In the introduction, Section 2 elaborates on the contributions of each essay and the dissertation as a whole. Section 3 provides a brief discussion of the theoretical and fundamentals, and then Section 4 briefly summarizes the five essays of this dissertation.

2 CONTRIBUTION OF THE DISSERTATION

This dissertation contributes to understanding the risks and opportunities of new blockchain-based digital financial instruments like cryptocurrencies and tokens from both theoretical and practical stances. Even though the focus areas of this dissertation are cryptocurrencies and cryptotokens, the risks and opportunities of these markets are representative of the entire blockchain-based digital financial market.

The first paper contributes to the new strand of digital finance literature exploring cryptocurrency and credit risk. Much current literature has investigated the volatility of cryptocurrency (Katisiampa, 2017; Balcilar, Bouri, Gupta, and Roubaud, 2017; Osterrieder and Lorenz, 2017; Ardia, Bluteau, and Rüede, 2018; Baur and Dimpfl, 2018; Borri, 2019), price spillovers between cryptocurrencies (Fry and Cheah, 2016), predictability of cryptocurrency time series (Catania, Grassi and Ravazzolo, 2019; Lahmiri and Bekiros, 2019; Omane-Adjepong, Alagidede and Akosah, 2019; Shen, Urquhart, and Wang, 2019), cryptocurrencies as investment assets (Urquhart 2016; Dyhrberg, 2016; Dwyer, 2015), and speculative bubbles in the cryptocurrency market (Cheah and Fry, 2015; Chaim and Laurini, 2019; Li, Tao, Su, and Lobont, 2019). Even though evidence has shown that the majority of cryptocurrencies go into default, there is no paper available on the predictability of such cryptocurrency bankruptcy. Being able to forecast potential cryptocurrency default is important because the sums of money involved are substantial (Fry and Cheah, 2016). This paper fills this important gap in the literature while also complementing the large body of literature exploring the predictability of commercial bankruptcy. From a more practical point of view, it also supports the finance industry by proposing a model that could be applied for investment decisions. For instance, new digital asset management could use our model to determine which cryptocurrencies should be treated with caution owing to a high probability of default.

The second paper contributes to the literature on cryptocurrency market efficiency and energy risk. While previous studies have mostly considered a single digital currency or a single consensus protocol as a case study exploring either energy efficiency or energy as a fundamental market-driving force for the pricing of cryptocurrency (Hayas, 2017; Li and Wang, 2017; Bendiksen et al., 2018; Biais et al., 2018; Symitsi and Chalvatzis, 2018; Li, et al., 2019), there is no paper available that takes a portfolio perspective across cryptocurrencies. This paper fills this gap in the literature for each of the three groups of consensus protocol—PoW, proof of stake (PoS), and hybrid (PoW and PoS)—and forms equally weighted portfolios of high-, medium-, and low-energy-consuming cryptocurrencies that correspond to the groups of PoW, hybrid, and PoS consensus protocols, respectively. Because cryptocurrencies incorporating the PoW consensus protocol are riskier than the other groups because their mining costs are more exposed to changes in energy price, finance theory has suggested that they should compensate investors with higher returns. Hence, the study tests whether the differences between those average portfolio returns are statistically different from each other. According to finance theory, one would expect the PoW portfolio to generate significantly higher portfolio returns on average than the hybrid and PoS portfolios if energy is a fundamental risk factor affecting the cryptocurrency market. In the same manner, one would also expect a hybrid portfolio to generate significantly higher portfolio returns on average than the PoS portfolio. This study fills some important gaps in the literature. First, it extends the current literature investigating the role of economic fundamentals in pricing cryptocurrencies. For instance, it extends the works of Biais et al. (2018) and Li et al. (2019) by taking a market-wide perspective. While Biais et al. (2018) and Li et al. (2019) focused on Bitcoin and Montero as single cryptocurrencies only, this study employs Fama and French's (2008, 2015, 2017) portfolio analysis, enabling us to make a much more generalized, i.e., market-wide, conclusion. Moreover, this study empirically tests Hayes's (2017) and Li and Wang's (2017) argument that fundamental factors are drivers of cryptocurrency price dynamics. So far, there has been no consensus achieved on energy efficiency as a market fundamental for cryptocurrency. While one group of scholars has argued that mining cost plays an important role in determining the market price (e.g., Hayes, 2017; Li and Wang, 2017; Symitsi and Chalvatzis, 2018), another group of scholars has argued that mining costs do not matter in the long run (e.g., Dimitri, 2017; Bendiksen et al., 2018).

The third paper contributes to the cryptocurrency literature in several important ways and also focuses on the regulatory risk of cryptocurrency. On one hand, Urquhart (2016) and Al-Yahyaee, Mensi, and Yoon (2018) studied the market efficiency of Bitcoin and found the cryptocurrency to be inefficient; however, Nadarajah and Chu (2017) revisited Urquhart's (2016) work and found that Bitcoin returns do satisfy the efficient market hypothesis. Moreover, Vidal-Tomás and Ibañez (2018) and Sensoy (2019) argued that Bitcoin has become more efficient over time, whereas Bariviera's (2017) findings suggested that Bitcoin has met the standards of informational efficiency since 2014. The different views existent in the literature indicate that there is no consensus on the market efficiency of cryptocurrency. While the papers cited above consider a single asset (e.g., Bitcoin), this paper adds to this strand of literature by taking a market-wide perspective. In doing so, it considers a whole set of cryptocurrencies that exhibit the largest market capitalization and employs Johansen's (1991, 1992, 1994, 1995) multivariate cointegration methodology to explore whether or not asset market equilibria exist that align with Engle and Granger’s (1987) cointegration theory. There is also a new strand of literature emerging that has discussed the features of privacy and nonprivacy coins. This literature has mostly adopted a technological perspective and has explored the privacy implications of Bitcoin (Androulaki et al., 2013), identification of a particular user’s blockchain transactions (Goldfeder et al., 2017, Khalilov and Levi 2018), technological interventions that could address the privacy issues of cryptocurrency (Ouaddah, Elkalam, and Ouahman, 2017; Kopp, Mödinger, Hauck, Kargl, and Bösch, 2017), and potential failures to guarantee privacy in terms of unlinkability and nontraceability (Kumar, Fischer, Tople, and Saxena, 2017; Möser, Soska, Heilman, Lee, Heffan, Srivastava, Hogan, Hennessey, Miller, Narayanan, and Christin, 2018). This paper adopts the financial perspective and considers the cryptocurrency market as an asset market comprising two submarkets, the privacy coin market and the nonprivacy coin market. This is the first paper that explicitly explores the existence of cointegration relationships among asset prices in the cryptocurrency market.

The fourth paper contributes to the recent stream of literature on news media sentiment risk in the cryptocurrency market in numerous ways. While earlier research studies have focused on news sentiments around events mostly related to macroeconomic announcements (e.g., Andersen, Bollerslev, and Diebold, 2007; Corsi, Pirino, and Reno, 2010; Corbet et al., 2020; Entrop, Frijns, and Seruset, 2020), this study covers in its analysis all the Bitcoin-related news sentiments published in major English language–based newspapers from around the globe. Previous studies on sentiment and Bitcoin price movements have mostly relied on news blogs and search websites rather than mainstream newspapers (e.g., Kristoufek, 2013; Garcia et al., 2014; Karalevicius et al., 2018), but this study chooses to analyze major newspapers. The main concern with creating a corpus using news blogs is the possible repetition or inclusion of advertisement texts along with the main news corpus. If the screening is not done properly, sentiments will tilt more in one direction as these advertisements mostly trigger either positive or negative emotions. Next, by further classifying sentiments into psychological and finance-specific and extending them into three different horizons to capture the heterogeneity in news arrival time among readers, this paper contributes to a better understanding of time-varying news sentiments, as well as their memory lengths and their effect on Bitcoin volatility. On top of that, this paper also studies the roles played by different human emotions by applying emotion lexicon–based sentiments and their implications on digital financial innovations like Bitcoin. From the practitioner's point of view, this paper also sheds light on capturing different sentiments from the news because accurate estimation of volatility is vital for investors to develop an adequate strategy in hedging potential risks associated with their investments.

The fifth and final essay contributes to the recent literature on ICOs in various fundamentally important aspects. First and foremost, taking the broader finance perspective, our paper adds to the literature on entrepreneurial finance and especially to the emerging literature on ICOs. Existing studies on ICOs have dealt with modeling the choice between ICO and venture capital funding (Catalini and Gans 2018; Chod and Lyandres 2020), exploring the need for ICO market regulations (Kaal 2018), the role of country-specific availability of investment-based crowdfunding platforms for ICOs (Huang, Meoli, and Vismara 2019), the pricing of exchange-traded ICOs (Howell, Niessner, and Yermack 2020; Benedetti and Kostovetsky 2021; Lyandres, Palazzo, and Rabetti 2019), and mechanisms through which tokens and ICO structures create value for both entrepreneurs and platform users (Li and Mann 2018). In this regard, this study adds to the literature exploring the determinants of success for ICOs. Specifically, the studies by Adhami, Giudici, and Martinazzi (2018), Fisch (2019), Amsden and Schweizer (2019) and Howell et al. (2020) investigated samples ranging from 253 to 1,500 ICOs. This work extends those studies by accessing the entire population of ICOs; i.e., it retrieves all 5,033 ICOs launched from August 2014 to December 2019. As a consequence, this study is not exposed to potential small-sample bias as it accounts for the whole population of available data. On top of that, another important novel aspect of this study is that it explores the question of whether financial sentiment—as opposed to psychological sentiment—cached in white papers has an impact on the success of ICOs. The white paper of an ICO is of major importance as it reveals the intended production outcome of the proposed business project; consequently, potential investors may or may not invest in an ICO merely based on its content. Hence, several natural and important questions arise. First, does the content of an ICO white paper matter? Second, should a white paper be written in simple terms for easy readability so that even a naïve investor can grasp the project idea? Third, should a white paper's choice of words trigger the positive or the negative side of the sentiment to successfully attract investors? This study is the first that seeks to answer these important questions. Apart from these first-order questions specifically related to an ICO white paper, there are also second-order questions that arise. For instance, what about other characteristics of ICOs not usually found in white papers, such as social media followers (e.g., as measured in terms of hype score), projects backed up by people disclosing their identities [e.g., as measured by know-your-customers (KYC) score], or potential risk for fraud (e.g., as measured by risk scores)? Do these factors also have an impact on attracting potential investors? Because there is no study available addressing the breadth of these issues, this study seeks to answer all these novel questions by accumulating information from all ICO white papers published during the 2014–2019 period.

FinTech is not only limited to cryptocurrency and blockchain implementation in finance but also relates to technological innovations in finance and financial services like payments, exchange, online banking, investments, financing, data analysis, machine learning, etc. **Figure 1** highlights the focus areas of this dissertation within the broader areas of FinTech.

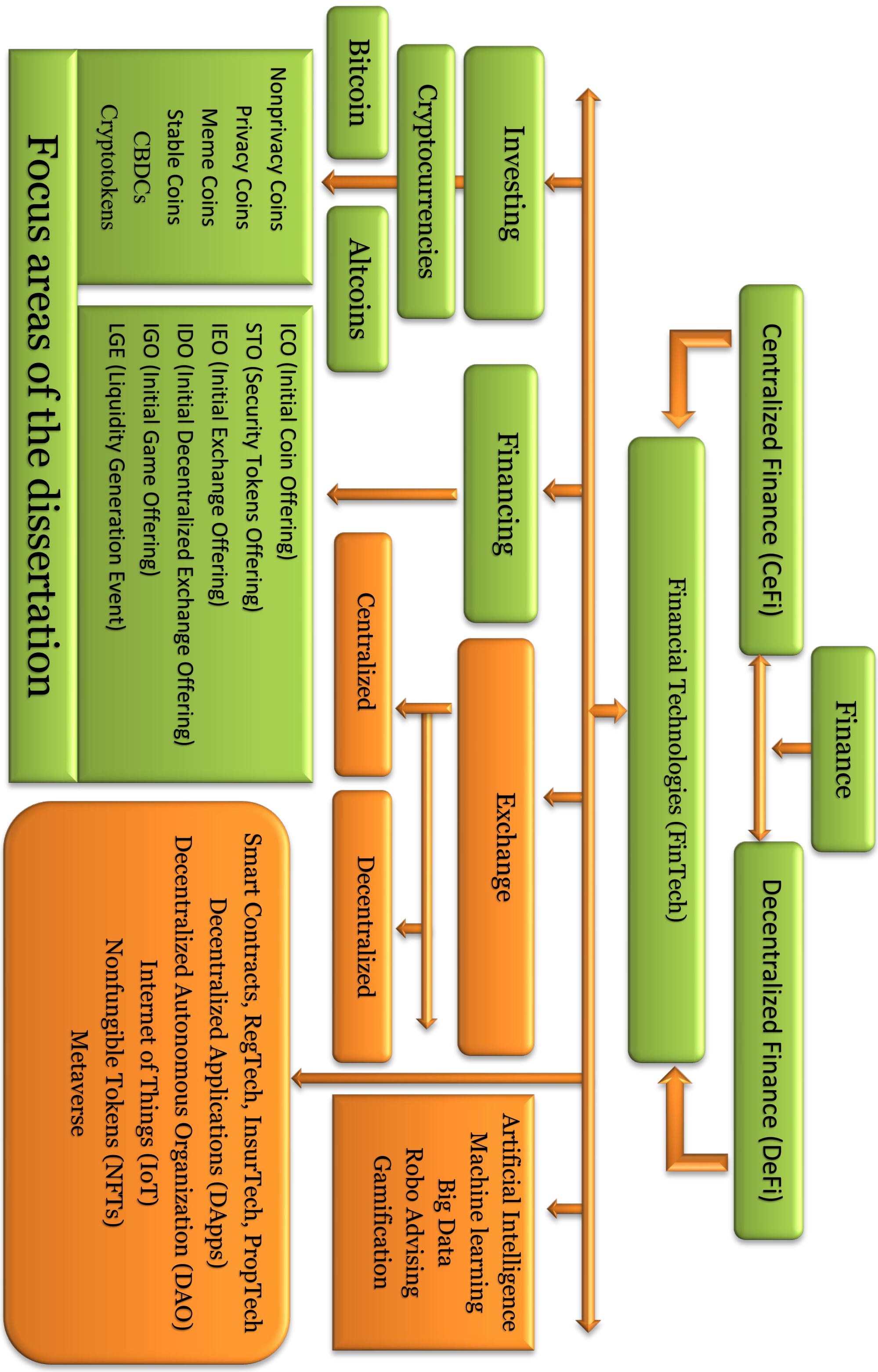


Figure 1. Focus areas of the dissertation within fintech (highlighted in green)

FinTech is driving innovation in financial services globally and is changing the nature of financial services. While the term FinTech carries broad meaning, it generally covers the application of blockchain, artificial intelligence (AI), cloud computing, machine learning, and big data in areas such as payments, clearing and settlement, deposits, lending, financing, insurance, investment and wealth management, regulation and market support, and various other aspects. At its core, FinTech helps companies, business owners, and consumers better manage their financial operations, processes, and lives by utilizing specialized software and algorithms that are used on computers and smartphones. Some of the most active areas of FinTech innovation include or revolve mostly around cryptocurrency and digital cash.

There is a huge misconception about FinTech mostly when it comes to the comparison between conventional and modern finance. It is not wise to disentangle FinTech from conventional finance. Generally, conventional finance is mostly centralized finance (CeFi) as there is a central regulating body. Blockchain technology, including Ethereum, is a distributed ledger technology that maintains records on a network of computers but has no central ledger. However, blockchain can be both centralized and decentralized. Conventional finance systems like traditional banks have used FinTechs more than any other sector—payment cards, online banking, digital wallets, and mobile pay have already been in use for several years.

This dissertation covers the investment and financing aspects of FinTech. As mentioned previously, FinTech is a broader concept that includes vast dimensions of financial instruments like cryptocurrency, tokens, open banking, smart contracts, regulation technology (RegTech), insurance technology (InsurTech), property technology (PropTech), decentralized applications (DApps), decentralized autonomous organization (DAO), internet of things (IoT), nonfungible tokens (NFTs), metaverse, AI, machine learning, big data, robo-advising, gamification, and much more. Even though the risks and opportunities discussed in this dissertation are mainly focused on cryptocurrencies and cryptotokens, they can be considered the representative risks and opportunities of the entire blockchain-based digital financial market.

In light of rapid technological advancements and innovations, it is very important to analyze the benefits and risks brought by FinTechs and to support adaptation. For example, according to Sapkota, Grobys, and Dufitinema (2020), $30 billion has been raised via the ICO market; however, they identified 576 ICOs launched before December 2019 as scams. The cumulative loss due to scams corresponds to $10.12 billion, suggesting an enormous societal impact of this criminal activity.

Therefore, it is also crucial to maintain a high level of cybersecurity and data security to maintain the trust of the public in the new blockchain-based digital financial market. Covering the investing and financing aspect of cryptocurrency as mentioned in Figure 1, this dissertation is composed of the following articles, as shown in Figure 2, which capture risks and opportunities in this market.

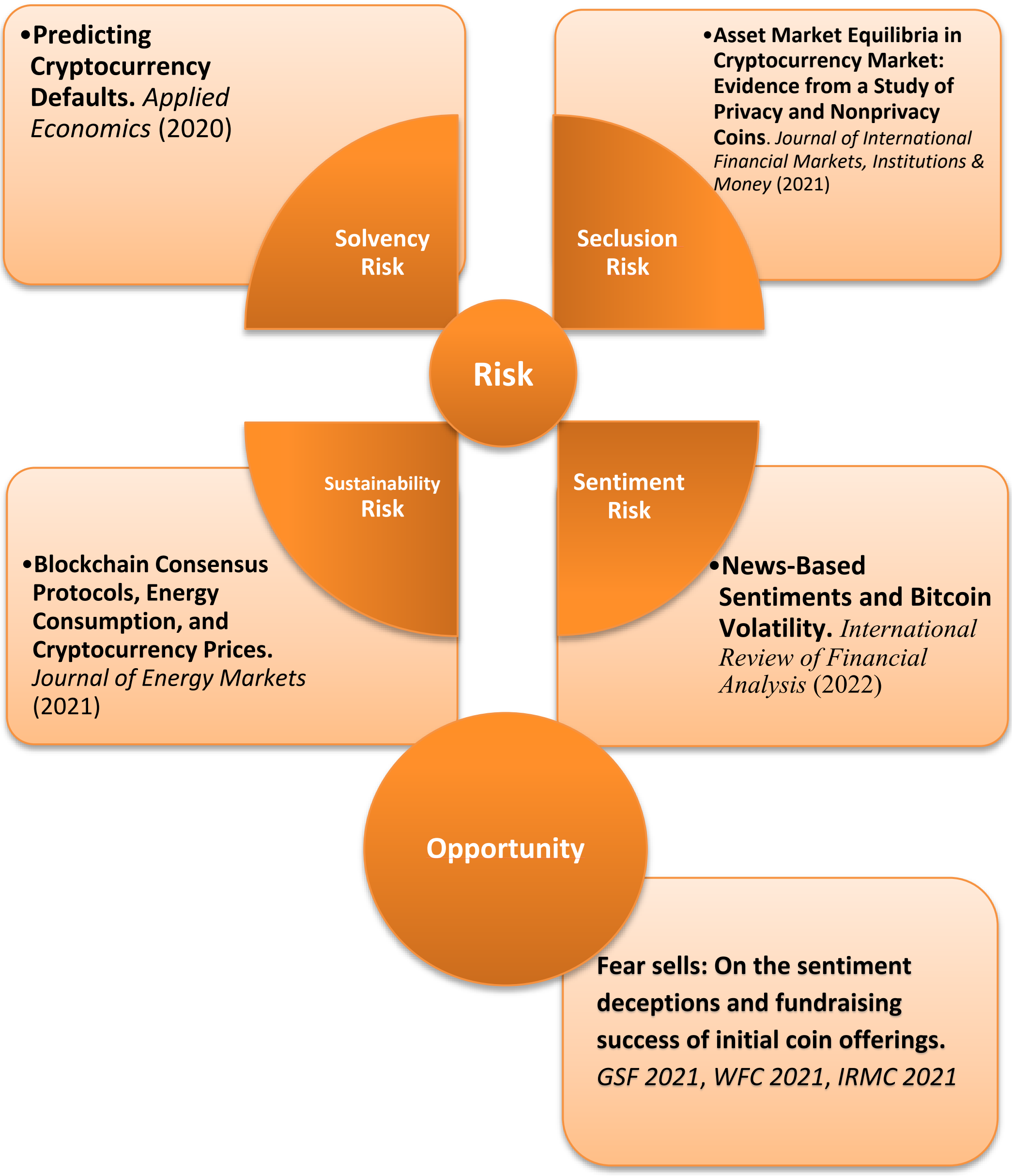


Figure 2. Composition of the articles for the dissertation.

3 THEORETICAL BACKGROUND

3.1 Blockchain

Blockchains are distributed digital ledgers of cryptographically signed transactions that are grouped and time-stamped into blocks. Each block in the chain is cryptographically linked to the previous block, making it hard for anyone to tamper with. Blockchains are validated using one or more consensus protocols. The older the blocks get, the more difficult they become to modify. For example, the Bitcoin blockchain is considered confirmed after six blocks as it becomes tamper-resistant. Each new block is distributed across copies of the ledger within the network, also known as nods, and is solved automatically using pre-established rules (i.e., consensus algorithms) by the miners. The users in the blockchain place their transactions on the network via different interfaces such as desktop applications, smartphone applications, digital wallets, web services, etc. The interface sends these transactions to all the distributed nodes within the blockchain. The selected nodes may be publishing full nodes as well as non-publishing nodes. A pending transaction is typically queued once it has been distributed to the nodes and later added to the blockchain by publishing nodes.

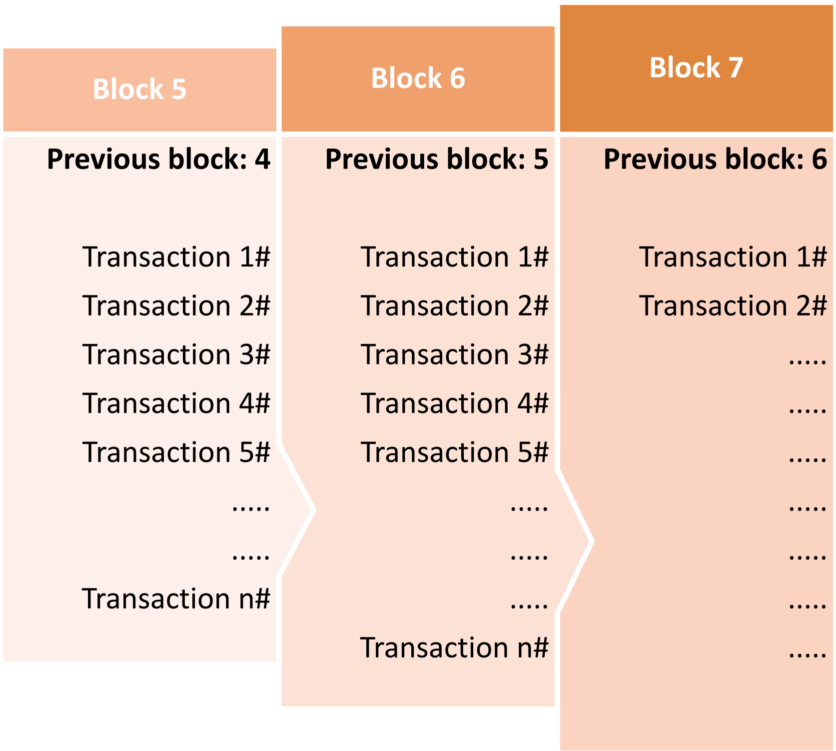


Figure 3. Generation of blocks in a blockchain.

Table 1 and **Table 2** show the Bitcoin blockchain taken from the btc.com block explorer.

Table 1. Latest blocks (1.12.2021 10:50 AM, source: btc.com)

Height	Mined by	Time	Reward
712,057	AntPool	15 secs ago	6.34841289 BTC
712,056	unknown	3 mins 24 secs ago	6.34861849 BTC
712,055	unknown	5 mins 22 secs ago	6.39187682 BTC
712,054	BTC.com	16 mins 43 secs ago	6.42923210 BTC
712,053	ViaBTC	36 mins 9 secs ago	6.39655711 BTC
712,052	AntPool	50 mins 16 secs ago	6.38541175 BTC

Table 2. Inside the block height: 712057

Tx Hash	Time	Amount
f945bab726a5...06465e7fb21e	3 secs ago	0.01439432 BTC
ccc8c2a51b50...4eb08ff96ece	3 secs ago	0.00707393 BTC
8932d7b99202...5b5a7274f29d	3 secs ago	6.44661761 BTC
39624c703a67...14bc5ef45d4f	3 secs ago	0.01536438 BTC
1fc3a0b38f17...76d67d26dc99	3 secs ago	0.02000896 BTC
1b24e03f0daf...a80092090409	3 secs ago	0.00979418 BTC

There are three types of blockchain network. The first type is called the permission-less network, or the public blockchain, which allows anyone to anonymously create an account and participate. It delivers a level of trust among participants with no prior knowledge about each other. Therefore, it enables individuals and institutions to engage directly, resulting in faster delivery at a lower cost. Similarly, the second type is called the permissioned blockchain; it can be controlled by one authority or by a consortium where the blockchain access is controlled to strengthen trust, e.g., in the corporate database. The third type of blockchain is called hybrid; it is a mix of permission-less (public) and permissioned (private) blockchain. A social media site database is one example of a hybrid blockchain where the network is controlled by one authority with some permission-less process.

3.2 Blockchain-based digital financial market vs. conventional finance

A plain explanation of FinTech does not convey its true definition. Financial firms like banks are early adaptors of any technological innovation that benefits them and their clients. Therefore, it is not wise to distinguish FinTech and the banking system as two different entities. In conventional finance, an entity like a bank or another financial firm acts as an intermediary; it is the dealer. However, in FinTech, the technology itself is an intermediary and provides brokerage between lenders and borrowers. Unfortunately, despite the ubiquitous presence of technology in banks, the basic business of financial intermediation has been largely unchanged over decades of technological development (Allen et al., 2020). Not only the financial services sector, but also the nonfinancial services sector can take advantage of blockchain's security, immutability, transparency, and ability to cut out the middleman. Like the internet, blockchain is a global infrastructure. It allows institutions and individuals making transactions to remove the third party, or the middleman, reducing cost and time. The blockchain ledger is not owned or controlled by one central authority or institution as in conventional finance and can be viewed by all users on the network.

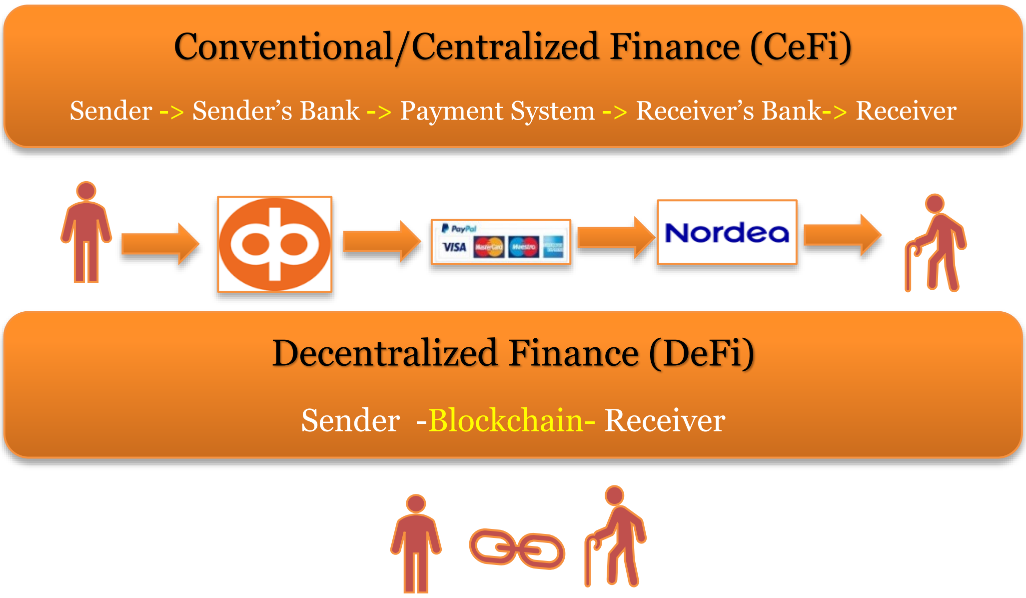


Figure 4. A simple difference between CeFi and DeFi

3.3 Risks in the blockchain-based digital financial market

As mentioned previously, even though the focus areas of this dissertation are cryptocurrencies and crypto-tokens, the risks and opportunities of these markets are representative of the entire blockchain-based digital financial market. This thesis uncovers 13 different types of cryptocurrency risk.

3.4 The 13Ss: The 13 sides of cryptocurrency risk

In the course of adopting new technologies such as blockchain-based digital currencies, one must be prepared to tackle several risks associated with it. However, the risks may vary based on the sector in which blockchain is being implemented, but in most cases, technical, operational, regulatory, legal, and security risks are the common types of risks that may arise. The risks from regulation, government, hardware and software failure, service disruption, and compromised system and database protection are considered the systematic risk, mainly arising from external hackers. Figure 5 shows the tridecagon (13 sides) of crypto risk and this dissertation covers the risks highlighted in green arrows.

Figure 5. The 13 sides of cryptocurrency risk

3.4.1 Solvency risk

Solvency risk is one of the most common risks in the new blockchain-based market. The first paper of this dissertation is about predicting cryptocurrency default; it finds that 59% of cryptocurrencies go into default within 4 years. If we look at the current situation of cryptocurrency, there are already more than 15,000 cryptocurrencies traded on more than 400 exchanges around the globe.[10] The list of actively trading cryptocurrencies published by coinmarketcap.com is 8,234 coins, of which around 800 coins are on the verge of default with zero trading activities. Out of 15,000 cryptocurrencies, only 7,400 are actively trading, which shows that more than 50% of cryptocurrencies are still going into default as suggested by the first paper in this dissertation. There are various reasons for the insolvency of cryptocurrency. As suggested by our paper, one of the main reason is

[10] https://coinmarketcap.com/?page=83 (as of 15.12.2021)

the heavy premine of the coin. Fake teams and scammers are another reason, in which the developer of the coin conducts different scams like exit, pump-and-dump, etc.

3.4.2 Sustainability risk

According to Digiconomist, Bitcoin mining generates 30,000 tons of electronic waste, as well as 96 million tons of carbon dioxide emissions, each year, which is equivalent to the amount generated by some smaller countries.[11] PoW consensus protocols used by cryptocurrencies like Bitcoin and Ethereum require a massive amount of electricity for their computational power. On one hand, this feature is a strength: computing power provides unprecedented security for networks. However, it is also a weakness given that most of the energy used is generated by fossil fuels, according to Harvey and Ramachandran (2021). In addition to energy consumption and emissions, cryptocurrency mining also generates a significant amount of electronic waste; specialized machines and hardware for mining the most popular cryptocurrencies become obsolete quickly, especially for application-specific integrated circuit (ASIC) miners. These circuits are unlike other computer hardware, and they quickly become outdated as they cannot be repurposed.

3.4.3 Seclusion risk

One of the main motives for the existence of Bitcoin is to bring financial inclusion to everyone. However, because of the full-transparency nature of cryptocurrencies like Bitcoin, users in the crypto space are being divided into two subgroups. One group favors full transparency, while the other group favors some level of privacy. Therefore, to understand this risk, one should first know the difference between privacy and non-privacy coins and also further understand concepts like clean vs. dirty coins and fungible vs. nonfungible coins. An anti-money laundry (AML) checker can be used to check if any Bitcoin has previously been used in illegal activities or not. If the checker finds that the Bitcoin has been previously involved in illegal activities, it can mark that particular Bitcoin as “dirty,” and it will be nonfungible to regular, or “clean,” Bitcoin.

---

[11] https://ticotimes.net/2021/12/27/what-is-the-environmental-impact-of-cryptocurrency

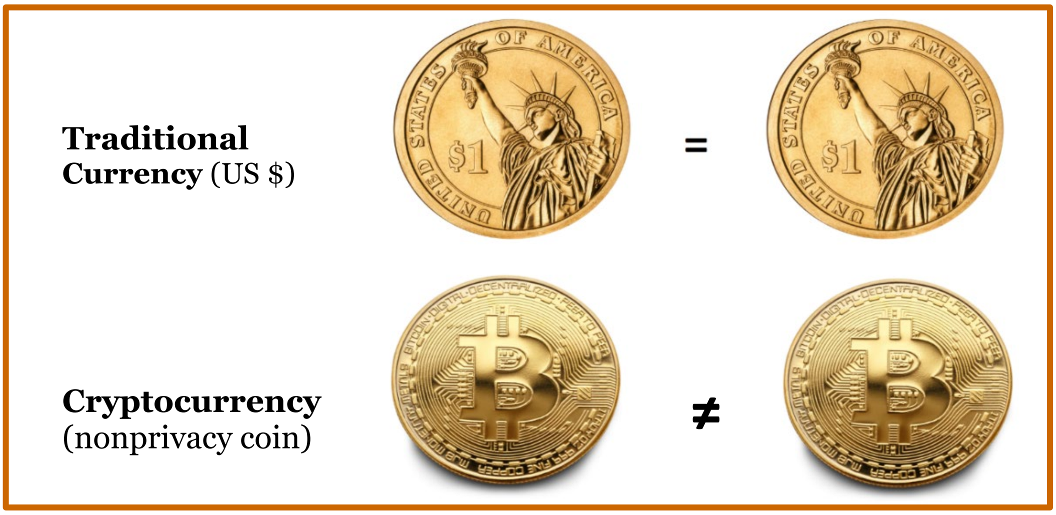


Figure 6. The fungibility problem (clean vs. dirty nonprivacy coins)

Another main factor for seclusion risk is the fungibility problem in the crypto world. With traditional currencies like the dollar or Euro, there is no fungibility problem because no matter how illegally one earns a fiat currency, it remains fungible, meaning a dollar earned from crime is worth the same in value as a dollar earned from hard labor. But this is not so for non-privacy cryptocurrencies like Bitcoin. As mentioned earlier, Bitcoin is nonfungible, so Bitcoin earned through illegal activities is marked as dirty through an AML checker, and dirty coins can be of far less (or zero) value than clean coins. This situation is possible because there is no user-level, transaction-level, or account balance–level privacy in the Bitcoin blockchain.

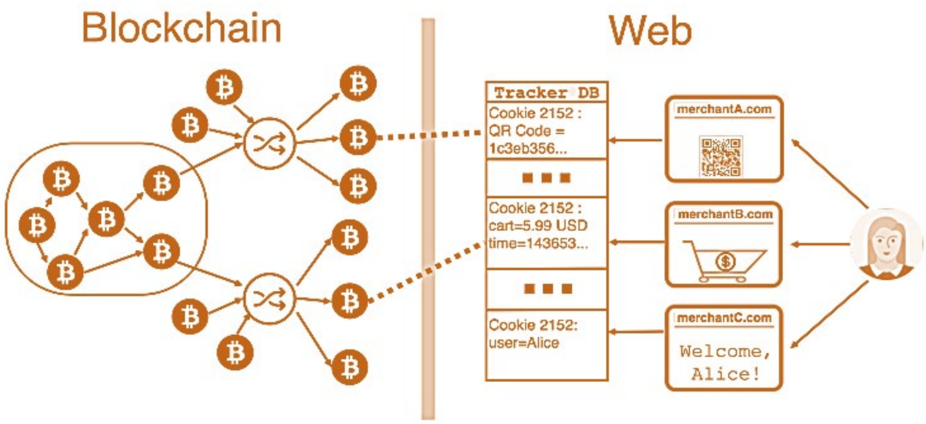


Figure 7. Bitcoin and anonymity (Source: Goldfeder et al., 2017, MIT Technology Review)

One might argue that the wallet address of a big corporate house is public and can be traced by anyone, but is the case the same for small-scale individual investors? Unsurprisingly, it is the same because we usually accept all cookies while surfing the internet and give away our identification easily while shopping online. In this way, a website can easily track the owner of a wallet address.

Figure 8. The fungibility problem (clean vs. dirty privacy coins)

The same may not be true for privacy coins like DASH and Monero. For example, even though Monero is a public and decentralized ledger, all transaction details are hidden. One can see transaction ID and payment ID, but no wallet information can be found. Therefore, privacy coins are fungible. Now, criminals are using nonprivacy coins like Bitcoin on the dark web for privacy, and those who want privacy on the surface web can still use privacy coins for anonymity. This is the seclusion risk.

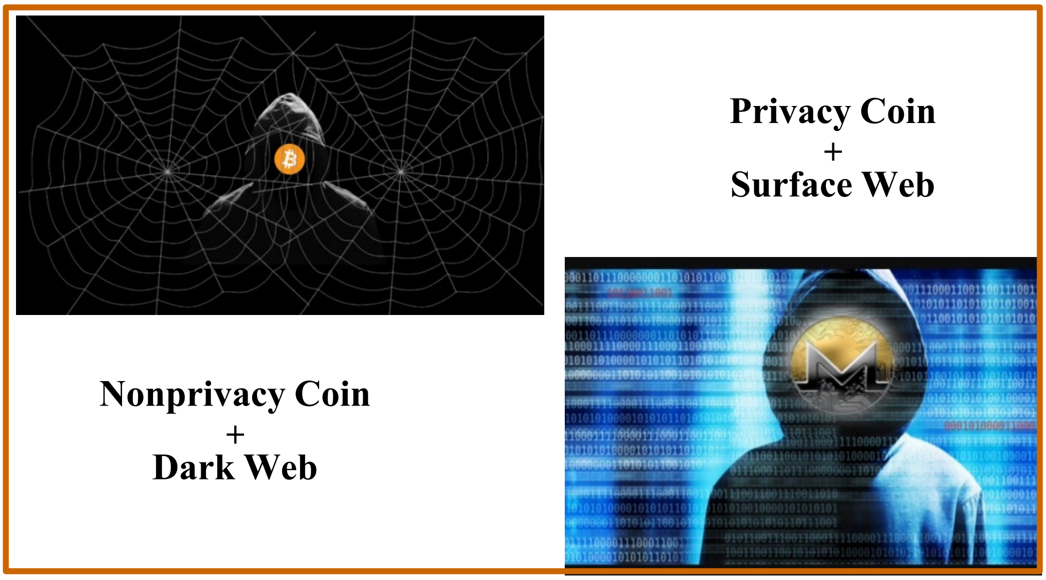


Figure 9. The privacy choice

Crypto users might divide into different subgroups instead of following the financial inclusion motto of Bitcoin. We know that the original purpose of Bitcoin was to provide financial inclusion and transparency, but the fully transparent nature of Bitcoin is also problematic; the developers need to come up with a privacy coin for those who want some privacy in their financial matters.

3.4.4 Sentiment risk

In the past decade, Bitcoin and other cryptocurrencies have made many news headlines in mainstream media. While many newspapers have covered cryptocurrencies as a possible scam or bubble, some newspapers have highlighted the opportunities created. Hackings, exchange shutdowns, government bans, regulations, taxes, scams, etc. have made many headlines in global news media outlets. The sentiment conveyed or imposed by the media is easily noticeable by looking at the Bitcoin price swing after a big news item either in favor of or against it. Volatility could be triggered by sentiment from a small news item as well.

3.4.5 Security risk

According to Foley, Karlsen, and Putniņš (2019), about one-quarter of all users (26%) and close to one-half of Bitcoin transactions (46%) are associated with illegal activity. Hileman and Rausch (2017, p.39) documented that 73% of exchanges control customers’ private keys, making them a potentially attractive “honeypot” for hackers as these exchanges have possession of user funds denominated in cryptocurrency. These studies showed that, unlike traditional asset markets, the new digital financial market carries different types of risks such as fraud, money laundering, credit, or hacking. This is a crucial issue, especially because institutional investors have recently entered cryptocurrency exchanges that store considerably larger funds in their digital wallets than retail investors do. It is relatively difficult to hack the Bitcoin blockchain, but hackers can easily steal Bitcoin by gaining access to the digital wallets of naïve users through various scamming techniques. Hackers also manipulate the market by gaining access to the trading system of crypto exchanges.

3.4.6 Scalability risk

Ethereum and other cryptocurrency blockchains with PoW consensus protocols have a fixed block size. According to Harvey and Ramachandran (2021), for a block to become part of the chain, each miner must execute all the included transactions on their machine, and expecting each miner to process all the financial transactions for a global financial market is unrealistic because consensus protocols are limited to a certain number of transactions per second. For example, the current version of the Ethereum blockchain can handle a maximum of 30 transactions per second. Given that almost all of decentralized finance (DeFi) resides on the Ethereum blockchain and comparing its transactions with the number of transactions of Visa, which can execute more than 65,000 transactions per second, Ethereum's lack of scalability places DeFi at risk of being unable to meet requisite demand. Therefore, one should focus on increasing Ethereum's scalability or replacing Ethereum with an alternative blockchain that can handle a higher number of transactions. Emerging blockchains, such as Solana, can achieve more than 50,000 transactions per second. Many approaches are addressing the scalability risks facing DeFi today. As long as the growth of any FinTech innovation is limited by blockchain scaling, the true power of applications will be limited.

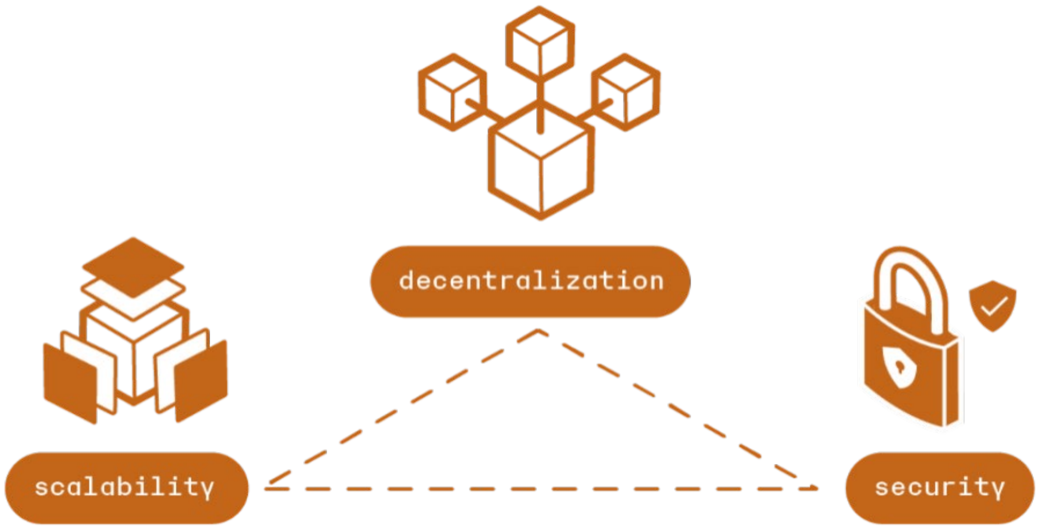


Figure 10. The scalability trilemma (source: ledger.com)

Moreover, scalability risk is often referred to as the scalability trilemma or the trade-offs of blockchain. Blockchain has to be built in such a way that a large number of transactions can be handled per second without compromising the effectiveness or security of the network. However, no blockchain can have all three—scalability, decentralization, and security—in one place. One has to compromise either on decentralization or security to achieve blockchain scalability. This is called the trilemma, or trade-off because if a project wants to improve one feature, it has to compromise on two other features; e.g., if it wants to improve scalability, it might have to compromise on security and decentralization, or if it wants to focus on security, it might have to compromise on scalability or decentralization or both. For example, we know the Bitcoin blockchain is secured and decentralized, but when it comes to scalability, it is measurable. Currently, each Bitcoin block is added every 10 minutes on average. Bitcoin transactions are confirmed after six blocks, so it takes almost an hour for a single transaction to be confirmed. This is a serious scalability issue with Bitcoin; nevertheless, off-chain payments like Bitcoin Lightning Network for handling micropayments could be one solution to the scalability issue.

3.4.7 Societal risk

According to *MIT Technology Review*, criminal entities laundered approximately $2.8 billion through crypto-asset exchanges in 2019.[12] Even though peer-to-peer

[12] https://www.technologyreview.com/2020/01/16/130843/cryptocurrency-money-laundering-exchanges/

exchange has its benefits, criminals have found ways to take advantage of its shortcomings for their benefit and fraudulent use. Thankfully, regulators have started paying attention to identifying these shortcomings and addressing them. For example, the Financial Conduct Authority (FCA) requires any company carrying out activity related to crypto assets in the United Kingdom (UK) to register and comply with AML and counterterrorist financing requirements.[13]

Each blockchain system, or exchange, has the potential to display risks that can be exploited. For example, many acts of money laundering are made possible by the relative anonymity of cryptocurrency transactions or by the security vulnerabilities present in some of the systems used to facilitate those transactions. Essentially, while banks have a distinct and heavily regulated global system of legal protections and obligations, the crypto-asset market is not as universally protected or regulated. Many jurisdictions and businesses boast their territory is safe and secure, but not all of them can ensure it.

3.4.8 Scam risk

There are several different ways by which investors can be fooled by scammers. Sapkota, Grobys, and Dufitinema (2020) focused on the market for ICOs and categorized scam incidents into different types based on their nature. These types of scams are common across all crypto-asset exchanges. For example, spam emails sent to users are called phishing scams in all sectors. There are many fake crypto projects with fake teams and scammers. Another scamming tool is a so-called bounty scam, where financial rewards are promised, mostly in tokens, for activities such as promoting projects on different forums, but the projects do not pay a bounty to those promotors. Another common type of scam among crypto projects is the exit scam, where developers and promotors suddenly disappear, dumping their holdings; the same group of scammers is likely involved in many previous scams. An airdrop is a fraud where scammers steal wallets' private keys from their users. Specifically, scammers create a booby trap, and users wanting to acquire free tokens click on the links and give away their private information. It is important to note that there are more than 400 crypto exchanges around the world, which makes it difficult for users to identify scam exchanges. Scammers who take advantage of this situation preferably launch the crypto project on a fraudulent exchange. Furthermore, many scam projects have fully or partially plagiarized white papers describing successful or promising projects. Pump-and-dump is another scam technique. Investors and traders rush to buy tokens at an

[13] https://www.coindesk.com/markets/2020/12/16/uk-fca-grants-crypto-firms-temporary-registration-as-it-deals-with-applications-backlog//

early phase when the price is low. Once the scammers complete the sale, the price suddenly crashes because of scammer sell-off. Other scamming techniques like Ponzi schemes, pyramid schemes, premine, and porn are gaining popularity.

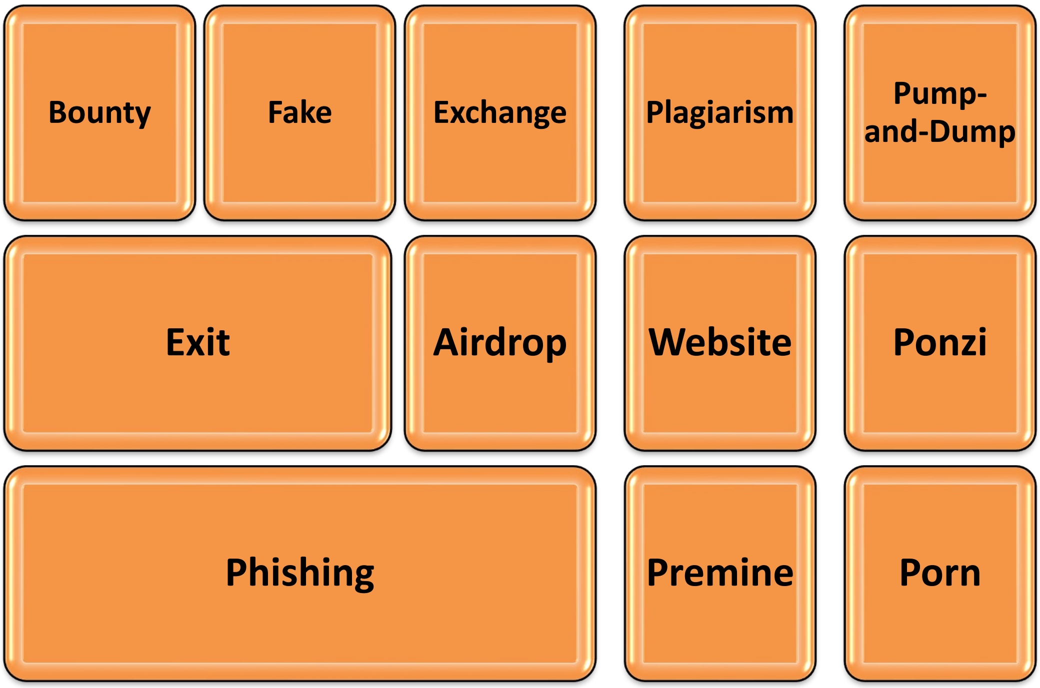


Figure 11. The 12 most common scamming techniques of a serial scammer (Source: Sapkota, Grobys, and Dufitinema, 2020)

3.4.9 Schooling risk

"In a few years, there will be more users in #crypto than there are users for internet today." –Changpeng Zhao (Chef Executive Officer, Binance tweet on 30.4.2019)

Schooling risk, or crypto-literacy risk, is another important factor to worry about. The current world population is 7.9 billion (as of May 2022).[14] As of today, 68% of the total world population uses the internet, but only around 4% are involved with cryptocurrency.[15] Cryptocurrencies like Bitcoin are already more than a decade old, but the adoption rate is still very low. The question arises, then—who is going

[14] See more at: https://www.worldometers.info/world-population

[15] See more at https://datareportal.com/global-digital-overview

to educate all these people and how long will it take for cryptocurrencies to become mainstream?

3.4.10 Steward risk

Technically, custodianship involves securing private keys, the cryptographic alphanumeric combinations needed to transfer and store digital assets. Private keys represent the single point of failure within the space and cannot be recovered once lost. A wide range of storage options is available, each with unique benefits for holders. One can manage assets using hardware and software wallets and third-party offline storage. Some users require access to their assets on a more frequent basis than others. Some investors may need third-party cold storage for legal or insurance reasons. Institutional investors in the US, for example, are required to use a qualified custodian for safety reasons.

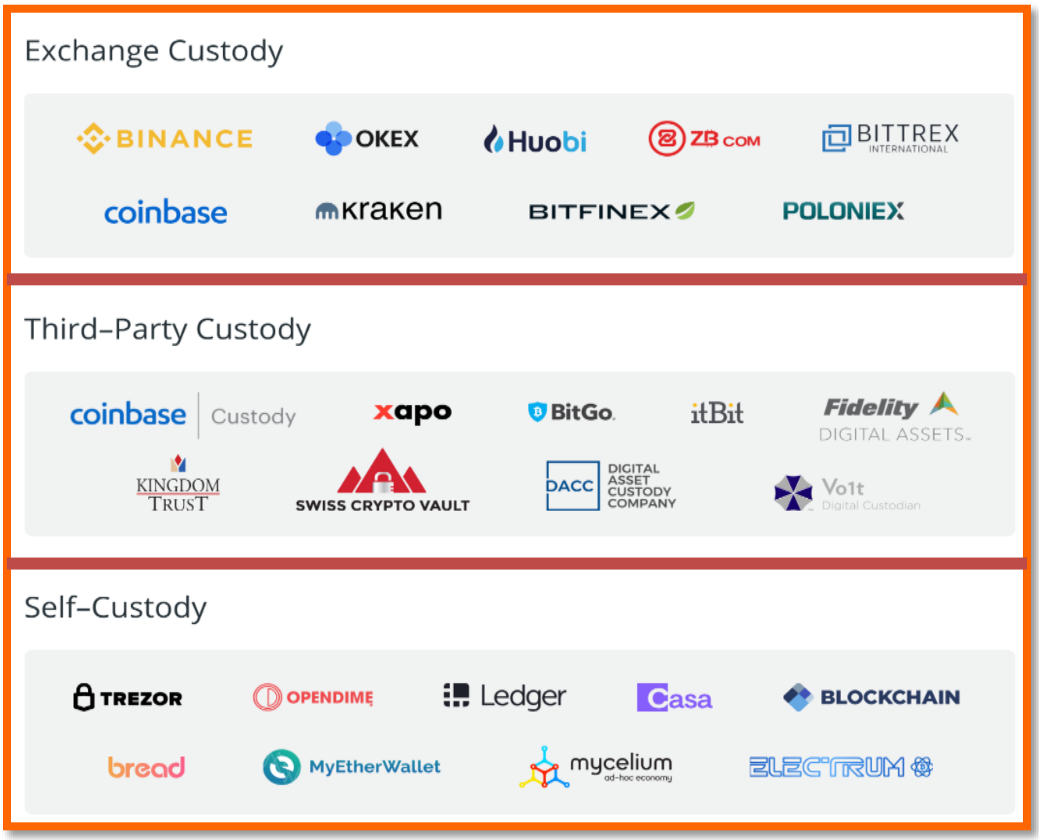


Figure 12. Different crypto-custody solutions (source: cointelegraph.com)

Steward risk, or custodian risk, is present in all types of storage. Among the exchange custodies, third-party custody and self-custody, also known as noncustody, are the safest. Most recently, though, after news of Coinbase's bankruptcy, third-party custody also seems risky as the third party might seize all the digital assets in the hosted wallet.[16] In this particular case, one might think that self-custody is safer, but unfortunately, we have seen some past events where losing a device results in losing everything without any way to retrieve it other than to find the private key or the device itself. One person from the UK lost almost half a billion dollars worth of cold storage devices in the dump that is gone forever.[17]

3.4.11 Stability risk

Cryptocurrencies are highly volatile; to overcome this issue, developers have come up with stablecoins. To make the coins stable, they are pegged with dollars or crypto portfolios or algorithms. Unfortunately, an incident with Terra stablecoin (USDT) showed us that even stablecoins are not in fact stable, as shown by Grobys, Junttila, Kolari, and Sapkota (2021). Therefore, USDT's collapse will probably be the end of most algorithmic stablecoins.[18]

3.4.12 Systematic risk

The trustless nature of cryptocurrency is its main feature; users can participate in a peer-to-peer network and exchange digital assets without the involvement of a central authority or trusted intermediary like a traditional bank. The systematic, or regulatory, risk of cryptocurrencies like Bitcoin increases as the concentration of power shifts to a handful of miners and cryptocurrency holders. It is evident that if any country or state has stricter regulations, innovation will move offshore or to a different state. However, if regulations are not as strict, many individuals and institutions can easily exploit their clients. Therefore, regulators should impose optimal or balanced regulations that are neither too strict nor too flexible.

FinTech is technically challenging, and regulators need to invest significant time getting up to date. Even after training, regulators find that their knowledge quickly depreciates given the speed of change. The innovation direction of FinTech is shifting so fast that regulators may find it challenging to catch up with innovation

---

[16] See more at https://nypost.com/2022/05/11/coinbase-warns-customers-they-may-lose-crypto-if-company-goes-bankrupt/
[17] See more at https://www.newyorker.com/magazine/2021/12/13/half-a-billion-in-bitcoin-lost-in-the-dump
[18] https://www.cnbc.com/2022/06/02/ust-debacle-will-probably-be-the-end-of-algorithmic-stablecoins.html

speed. For example, during 2017 and 2018, many startups were focused on ICOs, 2019 on Stablecoins, 2020 on central bank digital currencies (CBDCs), and in 2021 on DAO and NFTs.

However, this is not the only challenge given the shift in innovation; it becomes even more challenging to maintain a balance between the potential benefits of cryptocurrency and its costs and potential financial and nonfinancial risks. Therefore, it is important to regularly update the global regulatory framework for governing these digital assets. Regulators should review all the activities within the Bitcoin network and its interfaces and their interferences with the conventional financial system—e.g., conversion of Bitcoin into fiat currency, as well as trading in Bitcoin-based assets or tokens as securities or derivatives.

It is critical given the increasing popularity of Bitcoin to consider that the concentration of miners and holders creates an opportunity for market manipulation that could hurt other stakeholders. On the other hand, the operational risks of hardware and software failure, service disruption, and compromised system and database protection mainly arise from external hackers. In addition, clients may (un)intentionally misuse the platform, leading to increased operational risk.

Overall, inadequate technical and operational controls, policies, and procedures can create operational and security risks that can lead to theft and fraud in a digital environment. All of these risks are critical, as relationships in the sector are mostly based on trust, as stated by Osmani et al. (2020). Blockchain consensus protocols create rules about how participant nodes can interact, and some cryptocurrencies like Ethereum are based on standards, such as ERC-721. However, most crypto governance within or across cryptocurrencies is opaque. Moving forward, it will be imperative to understand how the risks will be governed when more crypto exchanges are, by design, decentralized. Questions arise about the monitoring of global platforms, especially when it comes to making important decisions and the fact that blockchain nodes exist across the globe and include competitive economic giants like the US, Russia, China, Europe, and others.

3.4.13 Speculative risk

Speculation carries a substantial risk of losing all of its value, but it also carries the potential for a sizable gain. The level of risk involved is the primary distinction between an investment and a speculative transaction. The majority of financial investments, including buying stock, include speculative risk. The value of the shares may increase, resulting in a gain, or decrease, resulting in a loss. A form of risk known as speculative risk is one that, when taken, has an undetermined degree of gain or loss. Accordingly, cryptocurrencies have grown in popularity since the establishment of Bitcoin, which had an initial market capitalization of zero, and a current market capitalization of almost US$400 billion.[19] At its peak, Bitcoin's market capitalization hit almost US$1.3 trillion. [20] Other widely adopted cryptocurrencies, such as Ethereum, have seen similar dramatic increases in market capitalization. The adoption of cryptocurrencies along with other forms of digital innovation as a form of a speculative asset has substantially increased the risk in this market.

3.5 Opportunities in the blockchain-based digital financial market

Even though there are different types of risk associated with the new blockchain-based digital FinTech, there is no doubt that this technology has brought a paradigm shift in finance. Global business today is being affected by FinTech, one of the fastest-growing segments in the digital and online industry. While defining FinTech can be difficult, it relates to companies and products that employ digital and online technologies in the banking and financial services industries. As a matter of fact, in a relatively short period of time, the emergence of a new generation of FinTech has drastically changed the way customers transact and how businesses complete transactions. Furthermore, it allows business services such as banking, advisory, and technology providers to appear almost identical, bridging the gaps between the various financial services providers.

[19] As of 8.7.2022 (source: coinmarketcap.com)
[20] As of 9.11.2021 (source: ycharts.com)

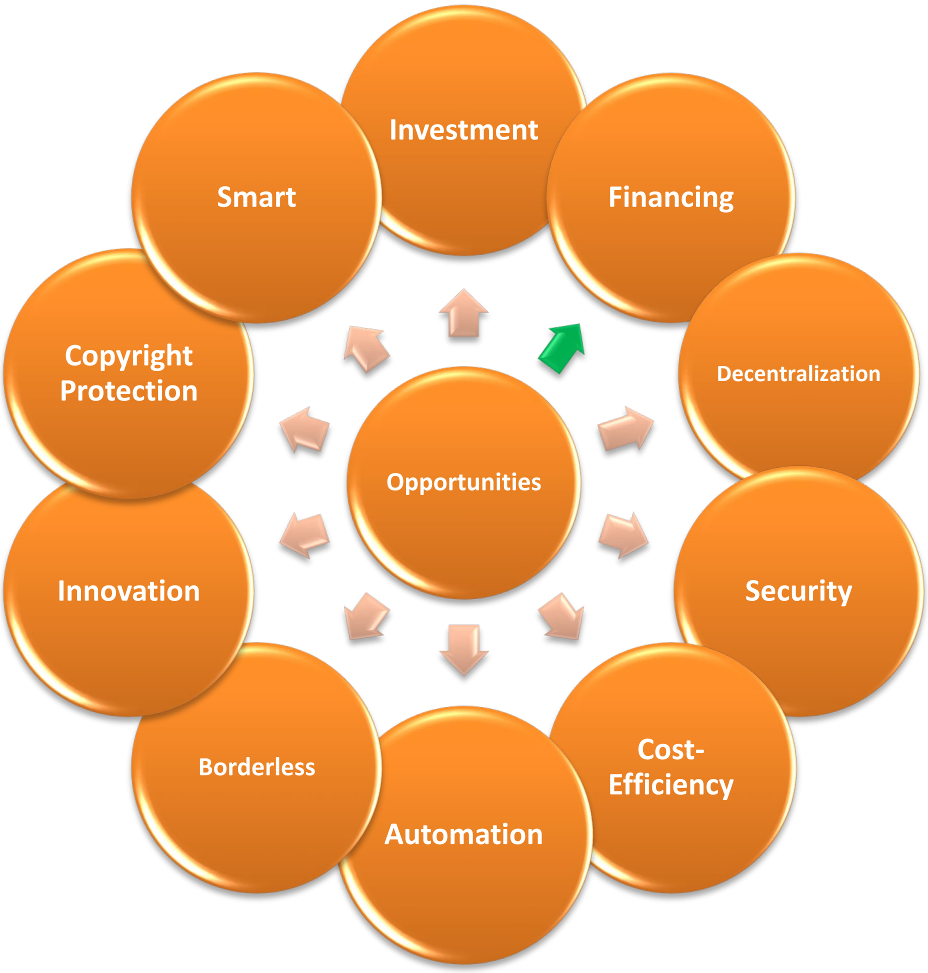


Figure 13. Opportunities in blockchain-based digital financial market (Note: This thesis covers the financing opportunity)

3.5.1 Investment opportunity

Financial platform providers are growing rapidly, and investment activities are also growing in emerging markets as FinTech adoption speeds up. The market is expanding rapidly, and governments and regulators around the world are working to build appropriate FinTech ecosystems so that they can keep pace with the market. Platforms help in building an impactful FinTech investment portfolio. One can see that the potential of inclusive FinTech solutions and financial inclusion can help investors with higher returns. Investing in FinTech now also opens up an array of possibilities for investors from developing and emerging markets. Automation in asset rebalancing has increased significantly because of data-processing and analytics tools. Additionally, cloud-based, robo-advisory-enabled platforms offer users investing and asset management advice using algorithms.

3.5.2 Financing opportunity

With crowdfunding and peer-to-peer lending, the traditional way of borrowing money from a bank has become much more transparent and less centralized. Through nontraditional methods like venture capital and others, new methods of sharing money have allowed investors to prosper while allowing those who may not qualify for traditional loans to access the money they need. With ICOs, startups can sell their tokens in exchange for cryptocurrency as a form of crowdfunding based on blockchain technology. Recent research has documented more than $30 billion raised via ICOs (Sapkota, Grobys, and Dufitinema, 2020; Howell, Niessner, and Yermack, 2020). ICOs are unregulated and unsupervised offerings of digital tokens over the internet that disintermediate platforms, payment agents, and professional investors and, thus, disrupt the current financial system (i.e., the market for IPOs). To be eligible for a corporate IPO, a firm must show a successful track record of earnings above a certain threshold, but other financial criteria are determined by the exchange where the firm plans to list. Generally, an ICO can be initiated by anyone with a creative idea or who wishes to start a company. Such a low bar often means that companies with high financial qualifications for an IPO may actually be overqualified for an ICO.

3.5.3 Decentralization opportunity

Markets without any central authorities facilitate transactions in DeFi. Free trade occurs without the influence of any centralized authority, allowing for a user-centric economy. Additionally, smart contracts on the blockchain can be used to exchange collectables and games as NFTs. Prompt settlement of transactions and easy synchronization of physical and digital assets are the key advantages of decentralization. Furthermore, transacting with a wide range of users in a decentralized marketplace democratizes the process by keeping borrowers' and lenders' decision-making options open.

3.5.4 Security opportunity

RegTech is becoming more prominent in the FinTech security space as governments across the globe seek to enforce broader cybersecurity regulations. In addition to deidentification processes and data encryption, RegTech generally utilizes new technologies to monitor and assess big data usage while ensuring that FinTech is complying with regulatory requirements. FinTech enterprises, their users, and their devices are going to handle increasingly sensitive information such that they need smart security solutions. Increasing in number are threats seeking to exploit their vulnerabilities, and implementing technologies such as IoT can be extremely useful for identifying security vulnerabilities and exchanging valuable information between devices. In addition to making it difficult for hackers to attack and enter sensitive business infrastructure, AI has also made it harder to exploit vulnerabilities and perform cyberattacks.

3.5.5 Cost-efficiency opportunity

Thanks to FinTech, payment services have become more transparent and efficient, and FinTech has increased its market share. It also connects banks with key customer bases to implement easy payments through wallets, payment applications, and online banking. FinTech’s solutions are not only cheap, they are also easy, transparent, and unified. Its benefits are flexibility and fast and cheap transactions. It is a technology-driven engine without the burden of regulation. Virtual operations and flexibility, along with lack of counting on funds from venture capital, make it possible for FinTech companies to attract customers at a lower price.

3.5.6 Automation opportunity

FinTech's robotic process automation (RPA) offers efficiency gains by automating routine, repetitive tasks, reducing labor costs, and speeding up internal processes. It simplifies operations, reduces labor costs, and reduces the time needed for repetitive tasks associated with managing a financial institution, such as maintaining accounts, connecting new customers, and processing loans. An enterprise automation platform is an engine that runs and controls a company's business events and provides real-time outcomes. FinTech automation can be defined as the use of automation tools to streamline end-to-end financial operations. Automating processes can help FinTech companies deliver fast, cost-effective services as opposed to traditional financial institutions. FinTech companies are known for their quick and lower-cost service compared to traditional financial institutions, and automation may be able to assist. Integration and automation present different challenges for FinTech organizations. Growing technology ecosystems, automated processes, and integration help FinTech companies streamline their services and accelerate their growth. FinTech automation offers several benefits, including a rapid time-to-deployment platform, real-time event-driven data operations, seamless adaptation to dynamic environments, an integrated platform for automated and integrated processes, and a cloud-based enterprise. Therefore, the enterprise must choose automation technology that enables them to deliver faster and more seamless services across various digital channels.

3.5.7 Borderless opportunity

FinTech can also be a key component in the future to bridging the gap to the unbanked. By operating primarily in the cryptocurrency world, FinTech services can enable people to send and receive assets such as Bitcoin and Ethereum via a digital wallet, which generally requires no KYC verification. Essentially, it means that users from all over the world can send and receive the same cryptocurrency, which requires no international transaction fees, without providing identifying documents that they may not possess. Moreover, FinTech is gaining popularity rapidly and is providing more intricate services, such as borrowing and lending, that can help increase access to financial services for more individuals. Nevertheless, there are still challenges to overcome. Although FinTech appears to be a global force for disruption, it will be some time before widespread access to the technology is sufficient to realize significant improvements. According to the Federal Deposit Insurance Corporation (FDIC), 3 out of 10 unbanked Americans lack a mobile device, and 7 out of 10 do not have home internet access (Harvey and Ramachandran, 2021). Technology must be accessible to unbanked communities, and FinTech has been able to accomplish this in the last decade with far more comprehensive solutions than their traditional banking counterparts could manage.

3.5.8 Innovation opportunity

Financial institutions are experimenting with several technological innovations in FinTech to better serve their customers, including automation, predictive analytics, digital-only banking, blockchain technology, and more. Companies in the financial sector have become more aligned with consumer expectations as a result of these changes, providing higher-quality services at lower prices. Consumer experiences are positively impacted by digital FinTech enhancements that create better digital processes that can be personalized. The ability to tailor a financial solution to a specific consumer in real time was previously unthinkable, yet today, banks have the capability. By using AI-driven predictive analytics, banks can customize financial packages for each customer. The use of AI in FinTech allows businesses to retain customers, accelerate loan approvals, and deter fraud.

3.5.9 Copyright protection opportunity

Tokens can be fungible or nonfungible, both of which are based on blockchain technology. The drawback of this is that the blockchain cannot store the actual underlying digital asset, particularly in the case of large files like pieces of artwork. If someone purchases an NFT, they don't purchase the digital artwork itself; they purchase a link to the digital artwork. Thus, the token in an NFT is a digital asset that is used to identify the asset by its token ID and ascertain who owns it. From the time the NFT was initially created and recorded on the blockchain, it has been publicly accessible. It is the blockchain equivalent of the county assessor's deeds that prove the ownership of real property, or if one works in intellectual property, the trademark and patent office or copyright office that proves the assignment of a right.

3.5.10 Smart finance opportunity

The financial sector has undergone rapid changes since the COVID-19 pandemic and lockdown took place. According to Statista, in 2020, companies worldwide preferred FinTech solutions for business-to-business payments over the payment options provided by banks.[21] In addition to peer-to-peer transfers, cryptocurrency sales, and contactless payments with no fees, digital banking has become more popular since the pandemic. There will continue to be significant advances in this field, enabling people to meet all of their financial needs anywhere, anytime. As Forbes wrote, banks have a tremendous amount of customer information, which is leading to the move to the cloud. This popularity is driven by open banking, which improves transparency. Many people understand that a hybrid cloud can provide a smooth path forward, combining public and private clouds for a seamless experience. However, their legacy systems may have prevented them from quickly gathering information from that data before it became outdated. Managing personal finances can be time-consuming and tedious. With FinTech, customers can automate financial decision-making and save valuable time. AI and machine-learning technologies offer analysis services that can help clients save money. Customer analytics platforms can be used by FinTech companies to collect and analyze user data, including baseline insights, brand interactions, and survey data. This will allow them to tailor their services to better match customer needs through personalization. Because of rapidly developing financial innovation trends, both personal finance and banking are now benefiting from FinTech trends. Rather than waiting in line at the bank, there are new products and mobile applications

[21] https://www.statista.com/statistics/1084937/future-usage-fintech-vs-bank-payment-solutions-companies-global/

available to accomplish the same task. Many companies provide software development services that can create the most convenient and secure ways to deal with finances. As a result, financial companies can obtain a personalized solution that corresponds to current trends.

4 SUMMARY OF THE ESSAYS

This dissertation includes five essays. The authors' contributions to each essay are detailed in **Table 3**.

Table 3. Authors' contributions across five different essays of this dissertation

	Essay and Authors				
Area of Contribution	***1***	***2***	***3***	***4***	***5***
Defining the research design	A1, A2	A1, A2	A1, A2	A1	A1
Collecting and managing research data	A1	A1	A1	A1	A1
Data analysis	A1, A2	A1, A2	A1, A2	A1	A1
Acquisition of research funding					A1, A2
Methodology and research methods	A1, A2	A1, A2	A1, A2	A1	A1
Project management					A2
Acquisition of research resources	A1, A2	A1, A2	A1, A2	A1	A1
Guidance in the research process	A2	A2	A2	A1	A2
Verification and analysis of findings	A1, A2	A1, A2	A1, A2	A1	A1
Visualization of results	A1, A2	A1, A2	A1, A2	A1	A1
Writing the body text of the original article	A1, A2	A1, A2	A1, A2	A1	A1
Editing the article at different stages	A1, A2	A1, A2	A1, A2	A1	A1, A2

Note

***Author 1** (A1):* Niranjan Sapkota, PhD Student

***Author 2** (A2):* Klaus Grobys, Associate Professor {PhD Supervisor (2nd)}

Essay 1: Predicting cryptocurrency defaults.

Essay 2: Blockchain consensus protocols, energy consumption and cryptocurrency prices.

Essay 3: Asset market equilibria in cryptocurrency markets: evidence from a study of privacy and non-privacy coins.

Essay 4: News-based sentiment and bitcoin volatility.

Essay 5: Fear sells: On the sentiment deceptions and fundraising success of initial coin offerings.

4.1 Predicting cryptocurrency defaults

The first essay of this dissertation studies the predictability cryptocurrency default. This study considers cryptocurrencies that were first traded before December 2014 and finds that 59% had gone bankrupt by the end of 2018. It presents a model for predicting cryptocurrency default by using data gathered between January 2009 and December 2014 and then by establishing whether these cryptocurrencies had gone bankrupt by December 2018. It identifies nearly two dozen crypto-specific variables that may provide predictor variables and uses those variables only for the model setups available to naïve investors 1 month after the launch of a new cryptocurrency. This study is similar to that of Altman (1968), who proposed a model to predict bankruptcies for firms in a specific industry in the US (e.g., manufacturing industry). This paper takes the first step in exploring predictable patterns in defaults in the newly emerging digital financial market and proposes a model capable of predicting defaults across different categories of blockchain-based digital financial assets like cryptocurrencies and cryptotokens. As a methodology, this study applies multiple linear discriminant analysis (MLDA), a type of cluster analysis used to model credit risks. For instance, Altman (1968) explored bankruptcy among companies in the manufacturing industry and proposed the z-score to predict the probability that a firm will go bankrupt within 2 years. That research led to many modifications being applied to predict various types of financial failure (Altman, 2002; Altman, Hartzell, and Peck, 1995; Altman, Haldeman, and Narayanan, 1977; Altman, Danovi, and Falini, 2013; Altman, and Rijken, 2010). This is the first paper to apply MLDA to model defaults in the cryptocurrency market. Again, all input variables used in this model were available to the naïve investor within 1 month after the start of trading for a cryptocurrency.

To check robustness, using bootstrapping techniques, the study establishes that the estimates are statistically significant for each of the selected variables for both the default sample and the functioning sample. The model correctly predicts 75 of 86 bankruptcy cases (87%). Moreover, the new market for digital assets is subject to the effects of technology: for example, many cryptocurrencies launched after 2015 adopted the PoS (mining) consensus mechanism, which uses less energy than the PoW (mining) consensus mechanism. Another significant change identified by this paper is that new cryptocurrencies are adopting more efficient and profitable cryptographic algorithms like X11, X12, and X17 over established algorithms like SHA and Scrypt. The proposed model may be utilized in the asset management industry, for example, as a screening tool for investment decisions. A rational investor would avoid investing in digital assets exhibiting overly high default risks. Portfolios that filter out cryptocurrencies predicted to be at risk of bankruptcy and that precondition the set of cryptocurrencies with those that are not could create a more favorable risk-return profile for investors.

4.2 Blockchain consensus protocols, energy consumption, and cryptocurrency prices

The second essay of this dissertation studies whether energy is a fundamental risk factor for pricing cryptocurrency. Cryptocurrency mining uses substantial energy. Consensus has not been reached about whether energy is a market fundamental for pricing. This study divides cryptocurrencies into three groups, depending on the consensus protocols' energy consumptions: PoW, PoS, and hybrid. Cryptocurrencies with the PoW consensus protocol consume the most energy, whereas cryptocurrencies with the PoS consensus protocol consume the least energy. For the empirical analysis, this study retrieves weekly data on 20 cryptocurrencies with the highest market capitalization as of January 3, 2016 for each of the three groups of cryptocurrency. One would expect a well-diversified portfolio of cryptocurrencies with the high-energy-consumption PoW consensus algorithm to result in higher fees and, therefore, higher returns on average than a well-diversified portfolio of cryptocurrencies with the low-energy-consumption PoS consensus protocol. Surprisingly, this study finds a well-diversified portfolio of cryptocurrencies with the medium-energy-consumption hybrid consensus protocol to generate considerably higher returns than either the PoW or PoS portfolio. One important implication of these findings is that the energy consumption of cryptocurrency does not appear to have anything to do with its price. Moreover, cryptocurrency price is largely determined by demand, as the supply side is controlled by algorithms specific to each coin. Because cryptocurrencies that incorporate hybrid consensus protocols generate significantly higher returns on average than cryptocurrencies incorporating PoW or PoS protocols, these results imply that investors' relative demand for cryptocurrencies incorporating hybrid protocols is greater than investors' relative demand for cryptocurrencies employing PoW and PoS protocols. A possible explanation for this phenomenon is that investors perceive cryptocurrencies that employ hybrid protocols as "more trustworthy" than the others. Even if it is unlikely that the blockchain of cryptocurrencies employing PoW and PoS protocols is manipulated by a miner holding either 51% of the entire network's computing power or 51% of a cryptocurrency-specific network's stake, market participants may still overestimate the risk for market manipulation.

The prospect theory states that individuals are generally risk-averse, and investors tend to overestimate events with low chances. As a result, investors' demand for cryptocurrencies that incorporate hybrid consensus protocols is relatively higher because they involve a lower risk of manipulation. However, other explanations are possible as well, and future research is needed to elaborate on this concept. Another important question that has recently been intensively investigated in the literature is whether the cryptocurrency market is efficient. The findings of this study provide evidence that the cryptocurrency market exhibits a strong pattern of higher-order autocorrelation on a portfolio level, although consensus has not yet been reached on whether the crypto market is efficient. Additionally, this study finds the time-series momentum to have a strong impact on the cryptocurrency market. Both results suggest that the cryptocurrency market is far from efficient.

4.3 Asset market equilibria in cryptocurrency market: Evidence from a study of privacy and non-privacy coins

The third essay of this dissertation explores whether asset market equilibria in the cryptocurrency market exist. In doing so, it considers privacy and nonprivacy coins as two different submarkets within the cryptocurrency market. Recently, users have been attracted to a new segment of the crypto market where one can hide his or her identity behind the privacy function offered by some cryptocurrencies. Goldfeder et al. (2017) reported that third-party trackers can deanonymize users of Bitcoin and other nonprivacy cryptocurrencies. Furthermore, with financial transparency, institutions are hesitant to use nonprivacy cryptocurrencies as a medium of exchange. As privacy coins have emerged, this problem has been resolved by utilizing features such as master node technology, ring signatures, and a stealth wallet address, which prohibit third parties from tracking the actual parties involved in a transaction. Cryptocurrency prices are typically predetermined by their total supply, suggesting that traders who value privacy over complete transparency are emerging as a new group within the digital financial market. One would expect privacy coins to form a submarket of cryptocurrencies detached from the market of cryptocurrencies, given that user demand for privacy coins is different from that of nonprivacy coins. This study considers a whole set of cryptocurrencies that exhibit the largest market capitalization and employs Johansen's (1991, 1992, 1994, 1995) multivariate cointegration methodology to explore whether or not asset market equilibria exist that align with Engle and Granger's (1987) cointegration theory. This methodology yields at least three advantages: (1) one can determine whether the privacy coin market generates a cointegration equilibrium that is distinct from the market for nonprivacy coins; (2) it does not prescribe a particular formulation of an equilibrium price mechanism given that there exists a cointegration equilibrium; and (3) it demonstrates four cointegration equilibrium points using all 20 cryptocurrencies as the basis of estimation.

This study accounts for liquidity and selects four variables to represent the two largest and the two smallest market capitalizations of privacy and nonprivacy coins. To compute the equation for modeling the privacy coin with the lowest market capitalization, Prime-XI (PXI), both privacy and nonprivacy coins are entered; only two of the nonprivacy coins that enter the equation exhibit statistical significance in the equation modeling DASH, which is the privacy coin with the highest market capitalization. Performing a likelihood-ratio test finds that one cannot reject the null hypothesis that the whole set of nonprivacy coins is jointly insignificant. First, an immediate implication of cointegration is the existence of Granger-causal orderings among cointegrated series, implying that asset prices determined in a weakly efficient market cannot be cointegrated. Hence, the findings provide evidence for market inefficiency. Second, the cointegration equilibrium associated with DASH appears to be disconnected from the market for nonprivacy coins. A novel aspect of this study is that it provides evidence that the underlying forces that cause large-cap privacy coin equilibrium are unrelated to those at work in the nonprivacy coin market.

Results of this study show strong evidence for the existence of four cointegration relationships in the market for cryptocurrencies, suggesting that the privacy coin market forms a distinct asset market equilibrium. Employing 10 large-cap cryptocurrencies with privacy functions and 10 nonprivacy cryptocurrencies in a joint model, the study uses a fully specified vector-error-correction model (VECM) to estimate the four distinct cointegration equilibria. To do so, it employs from both groups of cryptocurrency (e.g., privacy and nonprivacy coins) those cryptocurrencies that exhibit the respective highest and lowest market capitalizations as left-side variables. Using DASH as the privacy coin with the highest market capitalization, it finds that only two nonprivacy coins, Peercoin (PPC) and MaidSafeCoin (MAID), exhibit *t*-statistics indicating statistical significance on a common 5% level. One explanation could be behavioral type: it could be that the market actors in the privacy coin market are different from those who trade in the nonprivacy coin market. For instance, criminals involved in money laundering could favor privacy coins exhibiting a high level of liquidity as the sums involved could be substantial, making small-cap privacy coins an inappropriate choice for money laundering. Future studies might explore the heterogeneity existent in the cryptocurrency market in more detail. Potential factors causing the cointegration relationships should also be the subject of future research. If the stable cointegration relationship between asset prices is known to market participants, they would be able to exploit it and position themselves to profit. There is a broad stream of literature dealing with pairs trading, for instance, which requires the presence of cointegrated assets. Future research is needed to investigate this issue in the context of the new digital currency market. Because cryptocurrencies are by definition privy to intrinsic value—unlike fiat money issued by governments—the extent to which shocks may propagate across cryptocurrencies raises questions as to whether any observed spillover stems from investors' rebalancing activity and the accompanying price pressures rather than from fundamental information transmission. Future research is needed to elaborate on these mechanisms.

4.4 News-based sentiment and Bitcoin volatility

The fourth essay of this dissertation studies the effect of news-based sentiment on the volatility of Bitcoin. Bitcoin has received significant attention from mainstream media in the past decade. Some publications have considered it a positive innovation, while others have doubted its true value and authenticity. Many academic journals have also published on the cryptocurrency market in recent years, especially those studying the volatility of this new blockchain-based digital asset. An accurate estimation of volatility is vital for investors who wish to develop a strategy to hedge the potential risks associated with an investment. Volatility models can be beneficial in helping to estimate the risk associated with an asset or portfolio of assets. By extending Corsi's (2009) work to include news-based sentiments as an additional explanatory variable, this work examines whether news media sentiments have an impact on Bitcoin volatility. Corsi (2009) proposed the HAR-RV, a type of autoregressive model commonly used to assess volatility transmission. The model consists of a cascade of partial volatilities coupled with information flow and differences in agent risk allocations. The multiple components of the volatility structure are related to institutional structures, information flow, and differences in investors' risk portfolios. This study utilizes past RVs of Bitcoin and news sentiments to predict its future RVs with the aim to understand the nature and significance of ranges in predicting future volatility by considering three different range-based volatility estimates as proposed by Parkinson (1980), Garman and Klass (1980), and Rogers and Satchell (1991). Furthermore, the study differentiates financial and psychological sentiments cached in the news from different time periods to see the effect of heterogeneity of news arrival times and investor sentiment memory lengths and their impact on Bitcoin volatility by considering four various sentiment dictionaries, Harvard-IV, Loughran and McDonald (2011), Henry (2008), and Quantitative Discourse Analysis Package (QDAP). The work also extends the HAR model to the emotional level to explore the effects of different human emotions on

Bitcoin volatility by considering the National Research Council (NRC) Emotion Lexicon. As a result, it finds trust and fear to be the two most common human emotions affected by Bitcoin news. For all the range-based estimators, the average psychological sentiment is not significant either at the daily or monthly level, with the most likely explanation being the arrival time of news to potential investors or readers. Most people don't read newspapers on the day they're published. Furthermore, over a long period of time, they tend to forget the news, resulting in a deterioration of the sentiment originally generated by the news articles and coverages. In contrast to psychological sentiments and discourse sentiments, it is surprising that finance-specific sentiments are significant over the long run. This could be because Bitcoin is more closely associated with finance than psychology. It might also be the case that investors remember news with more financial sentiment for a longer time period than news with more psychological sentiment.

This study also uses the decomposition of overall sentiment into purely positive and purely negative sentiments to capture the heterogeneity between optimistic and pessimistic investors, assuming that optimistic traders are mainly driven by positive news sentiments and pessimistic traders are generally guided by negative news sentiments. The results indicate the volatility of Bitcoin to be mostly driven by positive financial sentiment. In other words, financially optimistic investors seem to be the main drivers of this market. Furthermore, applying the NRC Emotion Lexicon as a robustness check shows trust and anticipation to be significant throughout all the volatility measures. Based on the fact that NRC considers trust and anticipation to be positive sentiments and fear and anger to be negative sentiments, results confirm that it is not the negative, but the positive sentiment that is largely responsible for the volatility of the Bitcoin market. The out-of-sample forecasting accuracy of the model displays HAR-RV with sentiment extension to have good forecasting accuracy regardless of the choice of volatility measure. Furthermore, results reveal that information about time-varying sentiments could be important for analyzing the media risk associated with Bitcoin. These findings are thus critical for volatility modeling and trading strategies. This paper has significant implications for investors holding assets in the cryptocurrency market, more specifically, Bitcoin, as true sentiment in the news is critical for risk management and portfolio optimization. One limitation of this study is the use of non-FinTech dictionaries to analyze news sentiments. Previous studies have shown that dictionaries borrowed from other disciplines tend to misrepresent real sentiment; therefore, the author believes that a FinTech-specific sentiment dictionary would contribute to a better understanding of the true sentiments of the new digital financial market.

4.5 Fear sells: On the sentiment deceptions and fundraising success of initial coin offerings

The fifth essay of this dissertation studies the success factors of a blockchain-based digital crowdfunding tool commonly known as ICO. Recent developments have enabled entrepreneurs to receive funding in the form of digital currency based on blockchain technology. An ICO, unlike an IPO, is not subject to regulatory oversight and typically involves only two steps, namely a white paper and issuance of tokens that give the investor access to the ICO project's platform or service. White papers are important because they reveal the intended outcome of a business project and because they are the only document upon which potential investors can rely. A white paper's ability to communicate a product's benefits may be a key factor in attracting investors. In this regard, this study addresses several questions. The first-order questions include the white paper’s length, number of words, readability, and sentiment, and the second-order questions refer to external factors, such as social media followers, signature campaigns, attracted attention, and external assessments of risk associated with ICOs. A total of 5,033 ICOs were retrieved from the 2014–2019 time period. Of those, 1,507 ICOs contained data on the amount raised in funding. This study extends previous studies by retrieving the entire population of ICOs launched during the specified time period. Data were collected on ICO characteristics from different websites, along with a textual analysis of those white papers, resulting in 37 potential characteristics that might contribute to the success of ICOs. Although this study explores several novel factors, it controls for other factors related to ICO success as established in earlier studies, including disclosure, credible commitment to the project, and quality signals (e.g., token listings). This paper focuses on three major features associated with ICOs. The first is the level of information disclosure in terms of availability of necessary information on the white paper itself such as Roadmap/Milestone, softcap, hardcap, and disclosure of token numbers. Second, it quantifies the qualitative aspects of white papers such as sentiment and readability following four different sentiment dictionaries and seven different readability scores. Third, it accounts for the characteristics of the ICO project found outside of a white paper, such as social media followers, possible scams, and KYC score. Finally, for the methodology, it applies the multiple linear regression model based on pooled ordinary least squares for parameter estimation

The findings of this study, based on the entire available population of ICOs between 2014 and 2019, suggest that quality signals such as token number and softcap/hardcap do not seem to predict ICO success. Also, this study hypothesizes that an ICO white paper is intensively examined by rational investors who assess a project's quality based on quality and risk assessment. The readability of a white paper is not associated with ICO success. Specifically, results suggest that negative sentiment in ICO white papers is positively correlated with the amount of funding raised by ICO investors. This finding provides strong evidence that ICO investors are mainly guided by their emotional experiences when investing in the ICO market. Moreover, negative emotions are a significant factor determining whether companies acquire funding through ICOs. Also influencing an ICO's success are the number of followers on Twitter and the amount of attention it receives. ICOs with the most Twitter followers raise the most funding—indicating herding behavior. Considering that many of these behaviors are also associated with a desire to remain connected with others, this study suggests that indicators such as number of Twitter followers, signature campaigns, and attention scores point to the importance of this behavior. Further research is strongly encouraged on this important issue.

This study finds fear generated from a white paper to be the main emotional factor behind negative sentiment. Using the NRC Emotion Lexicon, this study factors the words associated with fear into their constituent elements. It indicates that investor behavior in the ICO market is mainly driven by fears associated with risk, problem, change, and regulation, among others. Concerning fear associated with risk, for instance, people nowadays face (a) risk of inflation due to extremely low-interest rates associated with quantitative easing, (b) risk of global warming due to pollution, and (c) risk of cyberattack due to lack of technological advancement, among others. The work finds projects that communicate in their white papers how they address these risks to be more successful in acquiring funding. Additionally, Loughran and McDonald (2011) reported that conventional word lists developed for other disciplines misclassify the words found in financial text, though research has shown the overall sentiment of Loughran and McDonald's dictionary to be around 60% in financial contexts. In the analysis, one can observe neither finance-specific dictionary to provide significant results, indicating that a dictionary borrowed from another discipline is likely to misjudge sentiment exponentially. For an additional robustness check, this study applies the artificial neural network, a machine-learning approach, and finds that it also favors the Harvard GI psychological sentiment dictionary. Two reasons may explain this phenomenon: either the finance-specific sentiment is not important to investors, or these dictionaries fail to capture the true sentiment when applied to FinTech-related contexts, such as an ICO. To analyze sentiment within the context of the new digital financial market, the authors argue that a FinTech-specific sentiment dictionary is essential.

References

Adhami, S., Giudici, G. and Martinazzi, S., 2018. Why do businesses go crypto? An empirical analysis of initial coin offerings. *Journal of Economics and Business, 100*, pp.64-75. https://doi.org/10.1016/j.jeconbus.2018.04.001

Allen, F., Gu, X. and Jagtiani, J., 2020. A survey of fintech research and policy discussion. *Review of Corporate Finance, 1*, pp.259-339. http://dx.doi.org/10.1561/114.00000007

Altman E., 2002. Revisiting Credit Scoring Models in a Basel 2 Environment. *Salomon Center for the Study of Financial Institutions* 2, 2-37. Available at SSRN: https://ssrn.com/abstract=1295815

Altman, E. 2000, Predicting Financial Distress of Companies: Revisiting the Z-score and Zeta Models, *New York University Working Paper*. https://doi.org/10.4337/9780857936080.00027

Altman, E. I., Haldeman, R.G. and Narayanan, P. 1977. ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking and Finance* 129-54. https://doi.org/10.1016/0378-4266(77)90017-6

Altman, E., 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* 23, 189–209. https://doi.org/10.2307/2978933

Altman, E., Hartzell, J. and Peck, M., 1995. Emerging Markets Corporate Bonds: A Scoring System. *New York: Wiley and Sons*. DOI: 10.1007/978-1-4615-6197-2_25

Al-Yahyaee, K.H., Mensi, W. and Yoon, S.M., 2018. Efficiency, multifractality, and the long-memory property of the Bitcoin market: A comparative analysis with stock, currency, and gold markets. *Finance Research Letters*, *27*, pp.228-234. https://doi.org/10.1016/j.frl.2018.03.017

Amsden, R. and Schweizer, D., 2018. Are Blockchain Crowdsales the New'Gold Rush'? Success Determinants of Initial Coin Offerings. Available at SSRN: https://ssrn.com/abstract=3163849

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics 89*, 701–720. https://doi.org/10.1162/rest.89.4.701

Androulaki, E., Karame, G.O., Roeschlin, M., Scherer, T. and Capkun, S., 2013, April. Evaluating user privacy in bitcoin. In *International Conference on Financial Cryptography and Data Security* (pp. 34-51). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39884-1_4

Ardia, D., K. Bluteau, and M. Rüede, 2018. Regime changes in Bitcoin GARCH volatility dynamics, *Finance Research Letters* 29, 266-271. https://doi.org/10.1016/j.frl.2018.08.009

Back, A., 2002. Hashcash-a denial of service counter-measure. Available at: http://www.hashcash.org/papers/hashcash.pdf

Balcilar, M., Bouri, E., Gupta, R. and Roubaud, D., 2017. Can volume predict Bitcoin returns and volatility? A quantiles-based approach. *Economic Modelling*, *64*, pp.74-81. https://doi.org/10.1016/j.econmod.2017.03.019

Bariviera, A. F. (2017). The inefficiency of Bitcoin revisited: A dynamic approach. *Economics Letters*, *161*, 1-4.
https://doi.org/10.1016/j.econlet.2017.09.013

Baur, D.G., and T. Dimpfl, 2018. Asymmetric volatility in cryptocurrencies. *Economics Letters* 173, 148-151. https://doi.org/10.1016/j.econlet.2018.10.008

Bendiksen, C., Gibbons, S., & Lim, E. (2018). The Bitcoin Mining Network-Trends, Marginal Creation Cost, Electricity Consumption & Sources. *CoinShares Research*, *21* pp.3-19.

Benedetti, H. and Kostovetsky, L., 2021. Digital tulips? Returns to investors in initial coin offerings. *Journal of Corporate Finance*, *66*, p.101786.
https://doi.org/10.1016/j.jcorpfin.2020.101786

Biais, B., Bisiere, C., Bouvard, M., Casamatta, C., & Menkveld, A. J., 2018. Equilibrium bitcoin pricing. Available at SSRN:
https://ssrn.com/abstract=3261063

Borri, N., 2019. Conditional tail-risk in cryptocurrency markets. *Journal of Empirical Finance* 50, 1-19. https://doi.org/10.1016/j.jempfin.2018.11.002

Bouri, E., Molnár, P., Azzi, G., Roubaud, D. and Hagfors, L.I., 2017. On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?. *Finance Research Letters*, *20*, pp.192-198. https://doi.org/10.1016/j.frl.2016.09.025

Brenig, C., Accorsi, R. and Müller, G., 2015, May. Economic Analysis of Cryptocurrency Backed Money Laundering. In *ECIS*.
https://doi.org/10.18151/7217279

Catalini, C. and Gans, J.S., 2018. Initial coin offerings and the value of crypto tokens. *National Bureau of Economic Research* (No. w24418). Available at: https://www.nber.org/papers/w24418

Catania, L., Grassi, S. and F. Ravazzolo, 2019. Forecasting cryptocurrencies under model and parameter instability. *International Journal of Forecasting* 35, 485-501. https://doi.org/10.1016/j.ijforecast.2018.09.005

Chaim, P., and M.P. Laurini, 2019. Is Bitcoin a bubble? *Physica A: Statistical Mechanics and its Applications* 517, 222-232.
https://doi.org/10.1016/j.physa.2018.11.031

Chod, J. and Lyandres, E., 2021. A theory of icos: Diversification, agency, and information asymmetry. *Management Science*.
https://doi.org/10.1287/mnsc.2020.3754

Corbet, S., Larkin, C., Lucey, B.M., Meegan, A. and Yarovaya, L., 2020. The impact of macroeconomic news on Bitcoin returns. *The European Journal of Finance*, *26*(14), pp.1396-1416. https://doi.org/10.1080/1351847X.2020.1737168

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), pp.174-196. https://doi.org/10.1093/jjfinec/nbp001

Corsi, F., Pirino, D. and Reno, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, *159*(2), pp.276-288. https://doi.org/10.1016/j.jeconom.2010.07.008

de Vries, A. (2019). Renewable Energy Will Not Solve Bitcoin's Sustainability Problem. *Joule*, *3*(4), 893-898. https://doi.org/10.1016/j.joule.2019.02.007

Dimitri, N., 2017. Bitcoin mining as a contest. *Ledger*, *2*, pp.31-37. https://doi.org/10.5195/ledger.2017.96

Dwyer, G.P., 2015. The economics of Bitcoin and similar private digital currencies. *Journal of Financial Stability* 17, 81-91. https://doi.org/10.1016/j.jfs.2014.11.006

Dyhrberg, A.H., 2016. Hedging capabilities of bitcoin. Is it the virtual gold?. *Finance Research Letters*, *16*, pp.139-144. https://doi.org/10.1016/j.frl.2015.10.025

Engle, R. F., and Granger, C. W., 1987, Cointegration and error correction: representation, estimation, and testing. *Econometrica* 55, pp. 251 – 276. https://doi.org/10.2307/1913236

Entrop, O., Frijns, B. and Seruset, M., 2020. The determinants of price discovery on bitcoin markets. *Journal of Futures Markets*, *40*(5), pp.816-837. https://doi.org/10.1002/fut.22101

Fisch, C., 2019. Initial coin offerings (ICOs) to finance new ventures. *Journal of Business Venturing*, *34*(1), pp.1-22. https://doi.org/10.1016/j.jbusvent.2018.09.007

Foley, S., Karlsen, J. R., and Putniņš, T. J., 2019. Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?. *The Review of Financial Studies* 32(5), pp.1798-1853. https://doi.org/10.1093/rfs/hhz015

Fry, J. and Cheah, E.T., 2016. Negative bubbles and shocks in Cryptocurrency markets. International *Review of Financial Analysis* 47, 343-352. https://doi.org/10.1016/j.irfa.2016.02.008

Garcia, D., Tessone, C.J., Mavrodiev, P. and Perony, N., 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. https://doi.org/10.1098/rsif.2014.0623

Garman, M.B. and Klass, M.J., 1980. On the estimation of security price volatilities from historical data. *Journal of business*, pp.67-78. Available at: https://www.jstor.org/stable/2352358

Goldfeder, S., Kalodner, H., Reisman, D. and Narayanan, A., 2017. When the cookie meets the blockchain: Privacy risks of web payments via cryptocurrencies. *arXiv preprint arXiv:* 1708.04748. https://doi.org/10.48550/arXiv.1708.04748

Grobys, K., Junttila, J., Kolari, J.W. and Sapkota, N., 2021. On the stability of stablecoins. *Journal of Empirical Finance*, *64*, pp.207-223. https://doi.org/10.1016/j.jempfin.2021.09.002

Haber, S. and Stornetta, W.S., 1990, August. How to time-stamp a digital document. In *Conference on the Theory and Application of Cryptography* (pp. 437-455). Springer, Berlin, Heidelberg. Available at: https://link.springer.com/chapter/10.1007/3-540-38424-3_32

Harvey, C.R., Ramachandran, A. and Santoro, J., 2021. DeFi and the Future of Finance. *John Wiley & Sons*. Available at SSRN: https://ssrn.com/abstract=3711777

Hayes, A. S. (2017). Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. *Telematics and Informatics*, *34*(7), 1308-1321. https://doi.org/10.1016/j.tele.2016.05.005

Henry, E., 2008. Are investors influenced by how earnings press releases are written?. *The Journal of Business Communication (1973)*, *45*(4), pp.363-407. https://doi.org/10.1177/0021943608319388

Hileman, G. and Rausch, M., *Global Cryptocurrency Benchmarking Study*, 2017 (Cambridge University, Cambridge Center for Alternative Finance: Cambridge). Available at SSRN: https://ssrn.com/abstract=2965436

Ho, K.Y., Shi, Y. and Zhang, Z., 2020. News and return volatility of Chinese bank stocks. *International Review of Economics & Finance*, *69*, pp.1095-1105. https://doi.org/10.1016/j.iref.2018.12.003

Howell, S.T., Niessner, M. and Yermack, D., 2020. Initial coin offerings: Financing growth with cryptocurrency token sales. *The Review of Financial Studies*, *33*(9), pp.3925-3974. https://doi.org/10.1093/rfs/hhz131

Huang, W., Meoli, M. and Vismara, S., 2020. The geography of initial coin offerings. *Small Business Economics*, *55*(1), pp.77-102. https://doi.org/10.1007/s11187-019-00135-y

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica: journal of the Econometric Society*, pp.1551-1580. https://doi.org/10.2307/2938278

Johansen, S., 1992a. Determination of cointegration rank in the presence of a linear trend. *Oxford Bulletin of Economics and Statistics*, *54*(3), pp.383-397. https://doi.org/10.1111/j.1468-0084.1992.tb00008.x

Johansen, S., 1992b. Testing weak exogeneity and the order of cointegration in UK money demand data. *Journal of Policy modeling*, *14*(3), pp.313-334. https://doi.org/10.1016/0161-8938(92)90003-U

Johansen, S.R., 1994. The role of the constant and linear terms in cointegration analysis of nonstationary variables. *Econometric reviews*, *13*(2), pp.205-229. https://doi.org/10.1080/07474939408800284

Kaal, W.A., 2018. Initial Coin Offerings: The top 25 jurisdictions and their comparative regulatory responses (as of May 2018). *Stan. J. Blockchain L. & Pol'y*, *1*, p.41. Available at SSRN: https://ssrn.com/abstract=3117224

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

Katsiampa, P., 2017. Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters* 158, 3-6.
https://doi.org/10.1016/j.econlet.2017.06.023

Kethineni, S. and Cao, Y., 2020. The rise in popularity of cryptocurrency and associated criminal activity. *International Criminal Justice Review*, *30*(3), pp.325-344. https://doi.org/10.1177/1057567719827051

Khalilov, M.C.K. and Levi, A., 2018. A survey on anonymity and privacy in bitcoin-like digital cash systems. *IEEE Communications Surveys & Tutorials*, *20*(3), pp.2543-2585. https://10.1109/COMST.2018.2818623

Kristoufek, L., 2013. Can Google Trends search queries contribute to risk diversification? *Scientific reports*, *3*(1), pp.1-5.
https://doi.org/10.1038/srep02713

Kumar, A., Fischer, C., Tople, S. and Saxena, P., 2017, September. A traceability analysis of monero’s blockchain. In *European Symposium on Research in Computer Security* (pp. 153-173). Springer, Cham. https://doi.org/10.1007/978-3-319-66399-9_9

Lahmiri, S., and S. Bekiros, 2019. Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solutions and Fractals* 118, 35-40. https://doi.org/10.1016/j.chaos.2018.11.014

Li, J., Li, N., Peng, J., Cui, H., & Wu, Z. (2019). Energy consumption of cryptocurrency mining: A study of electricity consumption in mining cryptocurrencies. *Energy*, *168*, 160-168.
https://doi.org/10.1016/j.energy.2018.11.046

Li, X., & Wang, C. A. (2017). The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin. *Decision Support Systems*, *95*, 49-60. https://doi.org/10.1016/j.dss.2016.12.001

Li, Z.-Z., Tao, R., Su, C.-W., and O.-R. Lobont, 2019. Does Bitcoin bubble burst? *Quality and Quantity* 53, 91-105. https://doi.org/10.1007/s11135-018-0728-3

Loughran, T. and McDonald, B., 2013. IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, *109*(2), pp.307-326. https://doi.org/10.1016/j.jfineco.2013.02.017

Lugovskaya, L., 2010. Predicting default of Russian SMEs on the basis of financial and non-financial variables. *Journal of financial services marketing*, *14*(4), pp.301-313. https://doi.org/10.1057/fsm.2009.28

Lyandres, E., Palazzo, B. and Rabetti, D., 2019. Do tokens behave like securities? An anatomy of initial coin offerings. *SSRN Electronic Journal*. Available at: https://www.runi.ac.il/media/10tedfav/do-tokens-behave-like-securities-lyandres.pdf

Maume, P., 2020. Initial Coin Offerings and EU Prospectus Disclosure. *European Business Law Review*, *31(2)* pp.185-208. https://doi.org/10.54648/eulr2020008

Meyer, A. and Ante, L., 2020. Effects of initial coin offering characteristics on cross-listing returns. *Digital Finance*, *2*(3), pp.259-283. https://doi.org/10.1007/s42521-020-00025-z

Mohammad, S.M. and Turney, P.D., 2013. Nrc emotion lexicon. *National Research Council, Canada*, *2,* p.234. Available at: http://www.saifmohammad.com/WebDocs/NRCemotionlexicon.pdf

Möser, M., Soska, K., Heilman, E., Lee, K., Heffan, H., Srivastava, S., Hogan, K., Hennessey, J., Miller, A., Narayanan, A. and Christin, N., 2017. An empirical analysis of traceability in the monero blockchain. *arXiv preprint arXiv:1704.04299*. https://doi.org/10.48550/arXiv.1704.04299

Nadarajah, S. and Chu, J., 2017. On the inefficiency of Bitcoin. *Economics Letters*, *150*, pp.6-9. https://doi.org/10.1016/j.econlet.2016.10.033

Nakamoto, S., 2008. Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review*, p.21260. Available at: https://www.debr.io/article/21260.pdf

Omane-Adjepong, M., Alagidede, P., and N.K. Akosah, 2019. Wavelet time-scale persistence analysis of cryptocurrency market returns and volatility. *Physica A: Statistical Mechanics and its Applications* 514, pp.105-120. https://doi.org/10.1016/j.physa.2018.09.013

Osmani, M., El-Haddadeh, R., Hindi, N., Janssen, M. and Weerakkody, V., 2020. Blockchain for next generation services in banking and finance: cost, benefit, risk and opportunity analysis. *Journal of Enterprise Information Management, 34(3)* pp.884-899. https://doi.org/10.1108/JEIM-02-2020-0044

Osterrieder, J. and Lorenz, J., 2017. A statistical risk assessment of Bitcoin and its extreme tail behavior. *Annals of Financial Economics 12(1)*, p.1750003. https://doi.org/10.1142/S2010495217500038

Ouaddah, A., Elkalam, A.A. and Ouahman, A.A., 2017. Towards a novel privacy-preserving access control model based on blockchain technology in IoT. In *Europe and MENA Cooperation Advances in Information and Communication Technologies* (pp. 523-533). Springer, Cham. https://doi.org/10.1007/978-3-319-46568-5_53

Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return. *Journal of business*, pp.61-65. Available at: https://www.jstor.org/stable/2352357

Rogers, L.C.G. and Satchell, S.E., 1991. Estimating variance from high, low and closing prices. *The Annals of Applied Probability*, pp.504-512. Available at: https://www.jstor.org/stable/2959703

Sapkota, N., Grobys, K. and Dufitinema, J., 2020. How much are we willing to lose in cyberspace? On the tail risk of scam in the market for Initial Coin Offerings. *On the Tail Risk of Scam in the Market for Initial Coin Offerings (November 18, 2020)*. Available at SSRN: https://ssrn.com/abstract=3732747

Sensoy, A., 2019. The inefficiency of Bitcoin revisited: A high-frequency analysis with alternative currencies. *Finance Research Letters*, 28, pp.68-73 https://doi.org/10.1016/j.frl.2018.04.002

Shen, D., Urquhart, A. and Wang, P., 2019. Does twitter predict Bitcoin? *Economics Letters*, *174*, pp.118-122. https://doi.org/10.1016/j.econlet.2018.11.007

Symitsi, E., & Chalvatzis, K. J. (2018). Return, volatility and shock spillovers of Bitcoin with energy and technology companies. *Economics Letters*, *170*, 127-130. https://doi.org/10.1016/j.econlet.2018.06.012

Urquhart, A., 2016. The inefficiency of Bitcoin. *Economics Letters*, *148*, pp.80-82. https://doi.org/10.1016/j.econlet.2018.02.017

Vidal-Tomás, D. and Ibañez, A., 2018. Semi-strong efficiency of Bitcoin. *Finance Research Letters*, *27*, pp.259-265. https://doi.org/10.1016/j.frl.2018.03.013

APPLIED ECONOMICS
2020, VOL. 52, NO. 46, 5060–5076
https://doi.org/10.1080/00036846.2020.1752903

Predicting cryptocurrency defaults

Klaus Grobys and Niranjan Sapkota

School of Accounting and Finance, University of Vaasa, Vaasa, Finland

ABSTRACT
We examine all available 146 Proof-of-Work-based cryptocurrencies that started trading prior to the end of 2014 and track their performance until December 2018. We find that about 60% of those cryptocurrencies were eventually in default. The substantial sums of money involved mean those bankruptcies will have an enormous societal impact. Employing cryptocurrency-specific data, we estimate a model based on linear discriminant analysis to predict such defaults. Our model is capable of explaining 87% of cryptocurrency bankruptcies after only one month of trading and could serve as a screening tool for investors keen to boost overall portfolio performance and avoid investing in unreliable cryptocurrencies.

KEYWORDS
Cryptocurrency; bitcoin; bankruptcy; default; credit risk

JEL CLASSIFICATION
G10; G12; G14; G17

I. Introduction

Facing the zeitgeist of digitalization, Bill Gates stated that 'the future of money is digital currency'.[1] Since the advent of Bitcoin – the first cryptocurrency traded – the number of cryptocurrencies has increased exponentially and there are now over 2,000 cryptocurrencies traded on over 16,000 markets around the world. The main advantages of cryptocurrencies are transparency and 24-h accessibility. Transactions of cryptocurrencies are all recorded on the open public ledger called the blockchain. This decentralized mechanism gives cryptocurrencies an unparalleled transparency. The technology behind the blockchain is revolutionary, but understanding it is challenging, especially for people without a technical background.

In contrast to traditional investments, cryptocurrencies carry different risks. For instance, Rauchs and Hileman (2017) report that the chance of cryptocurrency exchanges being hacked is 74–79%. Taking the legal perspective, Kethineni and Cao (2019) argue that cryptocurrencies became the currency of choice for many drug dealers and extortionists because of the opportunities to hide behind the presumed privacy and anonymity. Maume (2019), who explores Initial Coin Offerings (ICO), highlights that the potential lack of regulation and enforcement is particularly tempting for scammers and other miscreants. In contrast to traditional currency markets, cryptocurrency markets also involve credit risk: As a stylized fact, among all the cryptocurrencies launched prior to 31 December 2014, 59% went in default by the end of 2018, and the reasons for defaults are manifold.[2]

As of February 2019, the overall market capitalization in the digital asset market is more than 120 billion USD with Bitcoin dominating slightly with more than 50%.[3] In this regard, Fry and Cheah (2016, 350) highlight that 'from an economic perspective, the sums of money involved are substantial', and accordingly, the societal impact of losses due to defaults in the digital asset market may be enormous. Howell, Niessner, and Yermack (2019, 1) define three types of digital assets which are often referred to as coins. Specifically, the first type of digital asset is defined as a general-purpose medium of exchange and store of value cryptocurrency, such as Bitcoin. The second type of digital asset is a security token, which represents a conventional security that is recorded and exchanged on a blockchain

CONTACT Klaus Grobys klaus.grobys@uva.fi Department of Accounting and Finance, University of Vaasa, Wolffintie 34, Vaasa 65200, Finland

[1]The Bloomberg interview took place on 2 October 2014.

[2]The dead coin tracking website coinopsy.com lists the following as the main reasons for default: abandoned, abandoned/website, abandoned/volume, abandoned/buyback, abandoned/scam, scam, scam project/virus, joke, no exchanges/struggling, failed fork, failed/pre-mine no/low trade volume, pump and dump, and crashed (see https://www.coinopsy.com/dead-coins/).

[3]See https://coinmarketcap.com (accessed on 15 February 2019, 11:00 EST).

to reduce transaction costs and create a record of ownership, whereas the third type of digital assets is a utility token, which gives its holder consumptive rights to access a product or service. In Tables 1 and 2, we provide a demographic overview of new and bankrupt cryptocurrencies in different years.

Figure 1 shows that the numbers of default cryptocurrencies are increasing in comparison to the new cryptocurrencies added to the digital finance world after 2018. Specifically, we find that as of April 2019 there are altogether 1778 defaulted coins; however, 2147 coins are still in the digital asset market.[4]

It should not be surprising that in a zero-interest regime even the asset management industry pays ever more attention to digital assets as an investment alternative. Given the likelihood of digital assets ending up in default, it is surprising that there is no paper available exploring the extent to which a default of a digital asset is forecastable. This current paper fills this important gap in the new age of digital finance literature.

In our paper, we exclusively focus on the first category of digital asset defined as cryptocurrencies. As this type of digital asset is considered general-purpose medium of exchange, it is an alternative to traditional currency. We start our analysis by exploring which cryptocurrency-specific variables are accessible to the naïve investor. As we are interested in forecasting potential cryptocurrency defaults at an early stage, we focus on variables that are a part of the information set of the investor at most one month after a cryptocurrency started trading. Accordingly, we downloaded data for all cryptocurrencies launched before 2015 and followed up those cryptocurrencies until the end of 2018.[5] Specifically, our data set consists of 146 cryptocurrencies, of which 86 went bankrupt before the end of 2018. We divided our dataset into two subsamples: The first subsample contains data on those cryptocurrencies that went into default and the second subsample contains the data of those cryptocurrencies that functioned until the end of our sample period. To analyse which of our variables have discriminative power, we then test which of the mean differences of

Table 1. Population of cryptocurrencies including tokens.

Year	Before 28 April 2013	Mar–Dec 2013	2014	2015	2016	2017	2018	2019
New cryptocurrencies	7	60	452	207	298	800	1187	168
Default cryptocurrencies	0	0	**2**	**150**	**363**	**213**	**743**	**307**
Total cryptocurrencies	7	67	517	577	663	1353	2073	2147

Note: This table reports the numbers of new, bankrupt, and total cryptocurrencies during each year from April 2013 until April 2019. It is generated using the historical snapshot available at coinmarketcap.com.

Table 2. Life span of default cryptocurrencies including tokens.

	Default year	2014	2015	2016	2017	2018	Jan–Apr 2019	Total
		Number of default cryptocurrencies						
Year of issuance	**Before 28 April 2013**	0	0	0	0	1	0	**1**
	Mar–Dec 2013	2	1	17	3	11	4	**38**
	2014		149	133	34	58	13	**387**a
	2015			213	56	98	20	**387**
	2016				120	174	26	**320**
	2017					401	72	**473**
	2018						172	**172**
	Total number of default cryptocurrencies	**2**	**150**	**363**	**213**	**743**	**307**	**1778**

Note: This table reports the numbers of bankrupt cryptocurrencies with their specific year of issuance and year of bankruptcy. It is generated using the historical snapshot available at coinmarketcap.com.

aIt includes all cryptocurrencies and tokens using different consensus mechanisms. Out of which 86 cryptocurrencies that we used for our analysis are based on PoW consensus protocols.

[4]Note that Howell, Niessner, and Yermack (2019, 1) define three types of coins, Figure 1 accounts for the whole universe of digital assets. For instance, as of 2014, 146 out of 517 coins were cryptocurrencies that have the Proof-of-Work consensus protocol which are subject of examination in this study.

[5]It is also noteworthy that cryptocurrencies exhibit different types of consensus protocols to verify transactions such as Proof-of-Work, Proof-of-Stake or a mixture of both which is often referred to as Hybrid. Before 2015, however, there were only few cryptocurrencies issued that were implemented using the Proof of Stake (PoS) mechanism. PoS was first introduced by Sunny King and Scott Nadal in 2012 and later in 2013 Sunny King created the first cryptocurrency Peercoin (PPC) implementing the PoS protocol. PoS is created to solve the high energy consumption problem of Bitcoin which uses the Proof-of-Work mechanism. In order to keep our sample homogenous, we exclude those cryptocurrencies using a PoS mechanism from our sample.

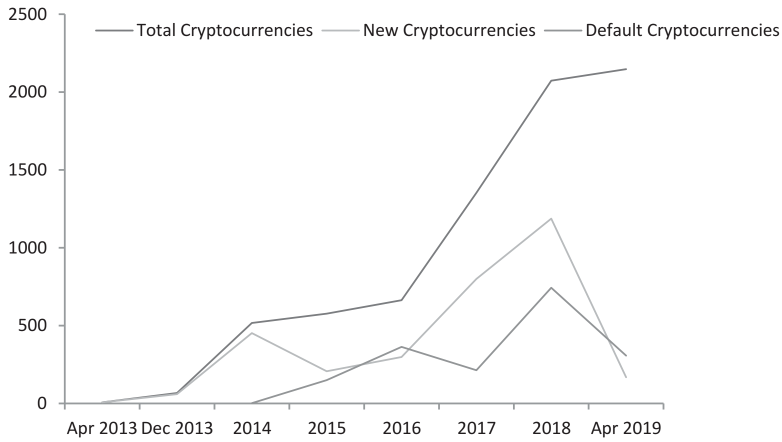


Figure 1. Demography of cryptocurrencies (Apr 2013 – Apr 2019).

Note: This figure shows the evolutions of new, default and total cryptocurrencies. Cryptocurrencies correspond to all three types of digital assets as defined in Howell, Niessner, and Yermack (2019, 1) where the first type is defined as a general-purpose medium of exchange and store of value cryptocurrency, such as Bitcoin. The second type of cryptocurrencies is a security token, which represents a conventional security that is recorded and exchanged on a blockchain to reduce transaction costs and create a record of ownership, whereas the third type of cryptocurrencies is a utility token, which gives its holder consumptive rights to access a product or service.

our cryptocurrency-specific variables for our two subsamples were statistically significant. We made use of those variables that exhibited significant differences in sample means in a multiple linear discriminant model. We compared the estimated bankruptcies with the actual numbers. Moreover, we applied bootstrapping techniques to investigate the robustness of our model involving Type-I and Type-II errors.

Our paper contributes to the new strand of digital finance literature exploring cryptocurrencies. Recent literature investigates the volatility of cryptocurrencies (Katsiampa 2017; Balcilar, Bouri, Gupta, and Roubaud, 2017; Osterrieder and Lorenz 2017; Ardia, Bluteau, and Rüede 2018; Baur and Dimpfl 2018; Borri 2019), price spillovers between cryptocurrencies (Fry and Cheah 2016), predictability of cryptocurrency time series (Catania, Grassi, and Ravazzolo 2019; Lahmiri and Bekiros 2019; Omane-Adjepong, Alagidede, and Akosah 2019; Shen, Urquhart, and Wang 2019), cryptocurrencies as investment assets (Urquhart 2016; Dyhrberg 2016; Dwyer 2015), and speculative bubbles in the cryptocurrency market (Cheah and Fry 2015; Chaim and Laurini 2019; Li et al. 2019). Even though empirical evidence shows that the majority of cryptocurrencies goes into default, there is no paper available on the predictability of such cryptocurrency bankruptcy. Being able to forecast potential cryptocurrency defaults is important because the sums of money involved are substantial (Fry and Cheah 2016). This paper fills this important gap in the literature while also complementing the large body of literature exploring the predictability of commercial bankruptcy. The publication of Altman's (1968) z-score model for predicting bankruptcy among manufacturing firms in the U.S.A, led to a wealth of research (Satish and Janakiram 2011; Wang and Campbell, 2010; Lugovskaya, 2010), and Altman (2018) has recently provided an excellent overview of the relevant literature.

Moreover, Cheah and Fry (2015) and Osterrieder and Lorenz (2017) express concern that academic research on cryptocurrency is often focused on the legality of cryptocurrencies (Kethineni and Cao 2019; Maume 2019) rather than offering a comprehensive analysis of their statistical or financial aspects. Therefore, our paper contributes to the finance literature by adding a new perspective, credit risk. Finally, and from a more practical point of view, our paper also supports the finance industry by proposing a model that could be used for investment decisions. For instance, new digital asset management could use our model to determine which cryptocurrencies should be treated with caution owing to a high probability of default.

The results of this research show that bankruptcies among cryptocurrencies are predictable. Specifically, our model shows that we can predict 75 out of 86 cryptocurrency defaults. Employing 5000 bootstrap replications shows that the confidence interval for the point estimate indicating *default* does not overlap with the point estimate for the Type-I error. This shows that the discriminative power of our model is significant. Our results are in line with the literature on predicting firm bankruptcy (Altman 1968, 1983, 2000, 2002; Altman, Haldeman, and Narayanan 1977; Altman, Hartzell, and Peck 1995; Lugovskaya, 2010). Surprisingly, our model is not suitable for predicting the fate of functioning cryptocurrencies unlike Altman's (1968) z-score model or Altman, Haldeman, and Narayanan (1977) ZETA model. We strongly encourage future research to elaborate on this issue.

The paper is organized as follows: The next section presents the empirical framework, including the model setup and robustness checks and the last section concludes.

II. Empirical framework

Multiple linear discriminant analysis

Our analysis is supported by data from various sources.[6] Each cryptocurrency has certain characteristics related to its history, specification, trading activities, reward, privacy, and scaling among others. Table A1 in the appendix shows the categorized specification details of cryptocurrencies. We downloaded all cryptocurrencies that incorporated the Proof-of-Work (PoW)[7] mechanism and started trading between 2010 and the end of 2014 and considered a data period of four years ahead.[8] In total, we retrieved 146 cryptocurrencies, of which 86 went into default in the sample period and 60 continued functioning. We define a cryptocurrency as being in a 'default state' when the cryptocurrency stopped trading, that is, there is no more evidence of any trading.[9] Altman and Rijken (2010, 4–5) emphasize the importance of ratio analysis as an empirical tool in assessing the performance of business enterprises. Identifying variables that exhibit discriminative power is ultimately an empirical question. Therefore, the first step in our analysis was to explore variables that potentially discriminate between cryptocurrencies that ended up in default and those that remained functioning. Moreover, we wanted to account for variables that only the investor has access to at most one month after starting trading that support a decision on whether to invest in the relevant cryptocurrency at an early stage. Table A2 in the appendix records 20 cryptocurrency-specific variables that exhibit information that could be utilized. Unfortunately, some information was not available for some now defunct cryptocurrencies.

There are many cryptocurrencies that are pre-mined before being offered to the public. Pre-mining has some advantages like rewarding the developers or creating a balanced distribution of coins (e.g., units of a cryptocurrency) between developers and traders. However, a larger number of pre-mined coins could be a negative indication, as when the developer has a large percentage of available coins and could therefore opt to leverage the price before selling quickly. Cryptocurrencies exhibiting higher levels of pre-mining are under constant attack and carry a high manipulation risk.[10] Therefore, investors are generally concerned about whether a particular cryptocurrency is pre-mined or not (which we account for by using a simple binary dummy variable), and also the fractions of pre-mined coins (it measures the extent to which the developers retain control over that particular cryptocurrency if the total coins are

[6] We used the following sources: mapofcoins.com (name of the respective cryptocurrency, categorization of 'running' and 'defunct'), coinmarketcap.com (historical price data), deadcoins.com (confirmation of the categorization 'defunct'), coinopsy.com/dead-coins (life span and founder information of dead coins), bitcointalk.org (announcement date and other technical specifications), and personal websites of coins for gathering any missing data.

[7] PoW is the very first consensus algorithm in decentralized public blockchains where miners solve complex cryptographic puzzles to add a block to a specific blockchain in exchange for coins as rewards.

[8] We downloaded price history from the coinmarketcap.com. The earliest data provided by this website start on 28 April 2013. Though Bitcoin (BTC), Litecoin (LTC), Namecoin (NMC), Terracoin (TRC), Devcoin (DVC) and Novacoin (NVC) started trading before this date. To have uniformity and consistency across our data set; however, we set 28 April 2013 as the first day of trade for the above-mentioned coins.

[9] There are a few cryptocurrencies in the list of functioning cryptocurrencies in coinmarketcap.com even though these cryptocurrencies do not exhibit any trading activities. In our data set, we adjusted for these errors.

[10] See https://cryptodaily.co.uk/2018/08/premined-coins-like-xrp-trx-xlm-and-neo-are-causing-problems-for-index-funds (published on 29 August 2018).

mined as the *Pre-mined-to-Total-Coins-Ratio* (PMTTCR)). Moreover, we accounted for block time, Day-1 return, Week-1 return, and Month-1 return after the respective cryptocurrency started trading. For instance, a positive return in the initial trading period could indicate popularity of a particular cryptocurrency. We also compounded the corresponding time-congruent volatilities (Day-1, Week-1, and Month-1) simply as the corresponding squared return. Instead of interpreting each variable in isolation, our variables should be considered in the respective context. For instance, a slightly negative first-day trading return with a low volatility in association with a high monthly volatility could indicate that the cryptocurrency did not attract attention following the announcement owing to a lack of social promotion, but the cryptocurrency could be subject to excessive speculation within the first month after trading. Generally, assets that are subject to excessive speculation may end up in trouble – or in default – at a later stage. Furthermore, *reward per block* shows the level of coin supply during that particular block interval. We include *Minimum-Reward-To-Total-Coin-Ratio* (MTTCR) as a common comparative tool to measure the minimum level of controlled supply among the cryptocurrencies in our sample. Our model includes both an individual and a comparative level of minimum controlled supply. Finally, we also coded dummy variables for identifying both the cryptocurrency-specific algorithm and whether the cryptocurrency has a known founder.[11]

We report the descriptive statistics of our selected variables in Table A3 in the appendix. Moreover, Table 3 reflects the variable means and the results of testing the difference in means for significance. We used a simple two-sample *t*-test to test the difference in sample means (Snedecor and Cochran 1989). The sample differences of minimum reward, Day-1 and Month-1 returns, Day-1 volatility, and PMTTCR are statistically significant on at least a 5% level (see Table 3). Moreover, the Month-1 volatility is at least marginally significant on a 10% level. Interestingly, we also find that among functioning cryptocurrencies, 58% of the founders remain anonymous, whereas among bankrupt cryptocurrencies that figure rises to 79%. For a 95% confidence interval, the critical values for the binary-distributed variable *known founder* in the sample of functioning cryptocurrencies is between 0.50 and 0.66, implying that the sample of bankrupt cryptocurrencies exhibits a significantly higher probability of the founder being anonymous, given a 5% significance level. Moreover, for a 95% confidence interval, the critical value for the binary-distributed variable *scrypt algorithm* in the sample of functioning cryptocurrencies is between 0.52 and 0.68. As the sample average in the default sample is 0.80, we can reject the null hypothesis that the sample means are equal, implying that those cryptocurrencies that ended up in default exhibit this specific algorithm more frequently.

More precisely, the definitions of our variables are as following:

$$Ret\_D1_t = \frac{(Day_1Close)_t - (Day_1Open)_t}{(Day_1Open)_t},$$

where Ret_D1_t denotes the first day's return of cryptocurrency t, $(Day_1Close)_t$ denotes the first day's closing price of cryptocurrency t, and $(Day_1Open)_t$ denotes the first day's opening price of cryptocurrency t.

$$Ret\_W1_t = \frac{(Day_7Close)_t - (Day_1Open)_t}{(Day_1Open)_t},$$

where Ret_W1_t denotes the first week's return of cryptocurrency t, $(Day_7Close)_t$ denotes the closing price after the seventh day of cryptocurrency t, and $(Day_1Open)_t$ denotes the first day's opening price of the cryptocurrency t.

$$Ret\_M1_t = \frac{(Day_{30}Close)_t - (Day_1Open)_t}{(Day_1Open)_t},$$

where Ret_M1_t denotes the first month's return of the cryptocurrency t, $(Day_{30}Close)_t$ denotes the closing price of cryptocurrency t after 30 trading days, and $(Day_1Open)_t$ denotes the first day's opening price of cryptocurrency t.

$$Vol\_D1_t = (Ret\_D1_t)^2,$$

[11]We categorized algorithms into three types: 'SHA' (Secure Hash Algorithm), 'Scrypt', and 'others'. 'Others' contains all other algorithms besides SHA and Scrypt family algorithms.

Table 3. Testing the differences-in-means between functioning and default cryptocurrencies.

	Default (D)	Functioning (F)	Difference (F–D)
Minimum reward	65880.65	3377.064	−62503.6b (−1.97)
Block time	160.79	152.92	7.87 (0.30)
Ret_D1	0.0403	0.7124	0.6721c (3.17)
Ret_W1	0.2541	0.2849	0.0309 (0.19)
Ret_M1	0.2454	0.1197	−0.1257b (−2.31)
Vol_D1	3.1776	10.2559	7.0783b (2.43)
Vol_W1	2.7361	4.8749	2.1388 (0.93)
Vol_M1	0.6147	0.3131	−0.3016a (−1.78)
MTTCR	3.2E-05	8.0E-06	−2.4E-05 (−1.57)
PMTTCR	0.0152	0.0041	−0.0111b (−2.47)
Pre-mined	4.89E+07	6.14E+08	−5.65E+08b (−2.03)
Known founder	0.79	0.58	−0.21c (−8.62)
Scrypt algorithm	0.80	0.60	−0.20c (−8.33)

Note: This table reports the differences of the means of our predictor variable candidates between our sample of functioning cryptocurrencies and those that went into default. As potential predictor variable candidates we consider the minimum reward, block time, first-day return (Ret_D1), first-week return (Ret_W1), first-month return (Ret_M1), first-day volatility (Vol_D1), first-week volatility (Vol_W1), first-month volatility (Vol_M1), Minimum-Reward-to-Total-Coins-Ratio (MTTCR), Pre-Mined-To-Total-Coins-Ratio (PMTTCR), and pre-mined coins (pre-mined). Our data set consists of all cryptocurrencies that incorporated the Proof-of-Work mechanism and started trading prior to 31 December 2014. We followed up those cryptocurrencies until the end of 2018. We retrieved 146 cryptocurrencies, of which 86 went into default (D) in the sample period and 60 remained functioning (F). (F-D) measures the mean-difference between the functioning and default sample. The corresponding *t*-statistics are given in parentheses.
aStatistically significant on a 10% level.
bStatistically significant on a 5% level.
cStatistically significant on a 1% level.

where Vol_D1_t denotes the first day's volatility of cryptocurrency t, and Ret_D1_t denotes the first day's return of cryptocurrency t.

$$Vol\_W1_t = (Ret\_W1_t)^2,$$

where Vol_W1_t denotes the first week's volatility of cryptocurrency t, and Ret_W1_t denotes the first week's return of cryptocurrency t.

$$Vol\_M1_t = (Ret\_M1_t)^2,$$

where Vol_M1_t denotes the first month's volatility of cryptocurrency t, and Ret_M1_t denotes the first month's return of cryptocurrency t.

Moreover, PMTTCR (*Pre-Mined-To-Total-Coins -Ratio*) indicates the fraction of coins that are allocated to the developers in relation to the total coins in circulation, given that a cryptocurrency is fully mined. (Note that developers with a large portion of coins in stake can manipulate the market with a so-called pump-and-dump strategy. Note also that if a large proportion of a cryptocurrency is pre-mined, this cryptocurrency could be subject to potential scam.) Further,

$$PMTTCR_t = \frac{(NUMBER\ OF\ PRE-MINED\ COINS)_t}{(TOTAL\ COINS\ WHEN\ FULLY\ MINED)_t},$$

where $PMTTCR_t$ denotes the *Pre-Mined-To-Total-Coins-Ratio* of cryptocurrency t, $(NUMBER\ OF\ PRE-MINED\ COINS)_t$ denotes the number of pre-mined coins of cryptocurrency t, and $(TOTAL\ COINS\ WHEN\ FULLY\ MINED)_t$ denotes the number of total coins of cryptocurrency t when being fully mined.

The number of coins received by miners as a reward per block for any cryptocurrency shows how new coins are generated after every *block time* interval (which, in turn, varies among cryptocurrencies). Specifically, *block time* is the time it takes to verify one block. This also indicates how frequently the new coins are generated to reward the miners for verifying the block. Moreover, the coins rewarded for the miners are the new coins supplied to the market. Due to the limited supply of coins (at least for the majority of cryptocurrencies), the reward decreases over time. *Minimum reward* measures the lowest number of coins as a reward given to the miners. The mining of a cryptocurrency continues only if the rewards cover the mining cost. If the *minimum reward* is meagre such that the mining cost cannot be covered, miners will stop mining and eventually that cryptocurrency is likely to end up in default. Therefore, *minimum reward* may be an important factor to consider in our current research's context. On the other hand, MTTCR (*Minimum-Reward-to-Total-Coins-Ratio)* measures the minimum level of controlled supply until the cryptocurrency is fully mined. Both, too much or too little supply of coins are not beneficial for the crypto economy. Further,

$$MTTCR_t = \frac{(MINIMUM\ REWARDS\ PER\ BLOCK\ EXCLUDING\ BONUS\ REWARDS)_t}{(TOTAL\ COINS\ WHEN\ FULLY\ MINED)_t},$$

where $MTTCR_t$ denotes the *Minimum-Reward-to-Total-Coins-Ratio* of cryptocurrency t, $(MINIMUM\ REWARDS\ PER\ BLOCK\ EXCLUDING\ BONUS\ REWARDS)_t$ denotes the minimum number of coins rewarded for the miners of cryptocurrency t, and $(TOTAL\ COINS\ WHEN\ FULLY\ MINED)_t$ denotes the number of total coins of cryptocurrency t, given the cryptocurrency is fully mined.

Next, we employed Multiple Linear Discriminant Analysis (MLDA) to address our research question. MDLA, which is a type of cluster analysis, that has been used to model credit risks. For instance, in his seminal paper, Altman (1968) explored bankruptcy among companies in the manufacturing industry and proposed the z-score to predict the probability that a firm will go bankrupt within two years. That research led to many modifications being applied to predict various types of financial failure (Altman 1983, 2002; Altman, Hartzell, and Peck 1995; Altman, Haldeman, and Narayanan 1977; Altman, Danovi, and Falini 2013; Altman and Rijken 2010). This is the first paper to make use of MLDA to model defaults in the cryptocurrency market. Again, all input variables used in our model were available to the naïve investor within one month after a cryptocurrency started trading. Since there are different methodologies to perform cluster analysis, below we explain how we set up our model.

We divided the data into two groups, the default group, and the group that consists of functioning cryptocurrencies. We stacked the data of those two groups into two matrices defined as X_1 and X_2, where X_1 denotes the default group and X_2 denotes the functioning group. Moreover, the matrix X defines the whole data set, that is

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} [x_{1,1} & \dots & x_{1,K}] \\ [x_{2,1} & \dots & x_{2,K}] \end{bmatrix}. \quad (1)$$

Let us assume that we consider K variables of the cryptocurrency-specific data and let us also assume that we deal with T_1 cryptocurrencies that went into default and T_2 cryptocurrencies that were functioning during our sample period. For instance in Equation's (1) notation, $x_{1,1}$ defines a T_1x1 column vector that contains the values for variable 1 for the default sample (e.g., group 1), whereas $x_{1,K}$ defines a T_1x1 column vector that contains the values for variable K in the default sample, and so forth. More concretely,

$$x_{1,1} = \begin{bmatrix} x_{1,1} \\ x_{2,1} \\ \vdots \\ x_{T_1,1} \end{bmatrix}, \text{ or } x_{1,K} = \begin{bmatrix} x_{1,K} \\ x_{2,K} \\ \vdots \\ x_{T_1,K} \end{bmatrix}, \text{ and analogously } x_{2,1} = \begin{bmatrix} x_{T_1,2} \\ x_{T_1+1,2} \\ \vdots \\ x_{T,2} \end{bmatrix}, \text{ or } x_{1,K} = \begin{bmatrix} x_{T_1,K} \\ x_{T_1+1,K} \\ \vdots \\ x_{T,K} \end{bmatrix}.$$

Then, for the matrices X_1 and X_2, the sample average of each column can be stacked into the *1*x*K* vectors $\mu_1$and μ_2, given by

$\mu_1 = [\bar{x}_{1,1} \quad \dots \quad \bar{x}_{1,K}]$ and $\mu_2 = [\bar{x}_{2,1} \quad \dots \quad \bar{x}_{2,K}]$.(2)

For instance, the element $\bar{x}_{1,1} = \frac{1}{T_1}\sum_{t=1}^{T_1} x_{t,1}$ defines the sample average of the first cryptocurrency-specific variable of the default group and $\bar{x}_{2,1} = \frac{1}{(T-T_1)}\sum_{t=T_1+1}^{T} x_{t,2}$ defines the corresponding sample average of the first cryptocurrency-specific variable of the functioning group. Moreover, the global mean vector μ stacks the overall sample averages for each column of the matrix X into a *1*x*K* row vector. Note that μ can be simply calculated as

$$\mu = \frac{1}{T}(T_1\mu_1 + (1 - T_1)\mu_2) = \frac{1}{T}(T_1\mu_1 + T_2\mu_2) \equiv [\mu_1 \quad \mu_2 \quad \dots \quad \mu_K] \quad (3)$$

Then, we calculated the mean-corrected matrices X_1^0 and X_2^0 defined as

$$X_1^0 = \begin{bmatrix} x_{1,1} - \mu \\ x_{2,1} - \mu \\ \vdots \\ x_{T_1,1} - \mu \end{bmatrix} \text{ and } X_2^0 = \begin{bmatrix} x_{T_1+1,2} - \mu \\ x_{T_1+2,2} - \mu \\ \vdots \\ x_{T,2} - \mu \end{bmatrix}, (4)$$

where obviously $T - T_1 = T_2$ and given Equation's (4) notation, $x_{t,i} - \mu$ defines a $1xK$ row vector i in each respective matrices, X_1 and X_2, subtracted by the global mean vector μ. For instance,

$x_{1,1} - \mu = [(x_{1,1} - \mu_1) \quad (x_{1,2} - \mu_2) \quad \ldots \quad (x_{1,K} - \mu_K)]$, or $x_{2,1} - \mu = [(x_{2,1} - \mu_1) \quad (x_{2,2} - \mu_2) \ldots (x_{2,K} - \mu_K)]$, for the default group and analogously,

$x_{T_1+1,2} - \mu = [(x_{T_1+1,1} - \mu_1) \quad (x_{T_1+1,2} - \mu_2) \ldots (x_{T_1+1,K} - \mu_K)]$, or $x_{T_1+2,2} - \mu = [(x_{T_1+2,1} - \mu_1) \quad (x_{T_1+2,2} - \mu_2) \quad \ldots \quad (x_{T_1} + 2, K - \mu_K)]$, for the functioning group, respectively.

We compounded the corresponding empirical sample covariance matrices as

$$C_1 = \frac{X_1^{0^T} X_1^0}{T_1} \text{ and } C_2 = \frac{X_2^{0^T} X_2^0}{T_2}, \tag{5}$$

where the dimension of C_1 and C_2 must be the same, that is, KxK as we want to investigate the characteristic-specific differences in cryptocurrencies. Then, we employed the estimated sample covariance matrices C_1 and C_2 to calculate the pooled within-group covariance matrix, simply defined as $C(r, s)$ and given by

$$C(r, s) = \frac{1}{(T_1 + T_2)} \sum_{i \in (1,2)} T_i \cdot C_i(r, s) \tag{6}$$

where $r = 1, \ldots, K$ and $s = 1, \ldots, K$. As $rank(C) = K$ was satisfied, we then compounded the inverse of C, defined as C^{-1}. Moreover, the prior-probability vector, based on the empirical data, can simply be calculated as

$$P = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \left(\frac{T_1}{(T_1+T_2)}\right) \\ \left(\frac{T_2}{(T_1+T_2)}\right) \end{bmatrix} \tag{7}$$

Finally, for our T cryptocurrencies, we can estimate the discriminant function depending on the default and non-default cluster as

$$f_{1,t} = \mu_1 C^{-1} x_{t,K}^T - 0.5 \cdot \mu_1 C^{-1} \mu_1^T + ln(p_1) \tag{8.a}$$

$$f_{2,t} = \mu_2 C^{-1} x_{t,K}^T - 0.5 \cdot \mu_2 C^{-1} \mu_2^T + ln(p_2) \tag{8.b}$$

where $x_{t,K}^T$ is the corresponding transposed $1xK$ vector of characteristics of cryptocurrency t. If $f_{1,t} > f_{2,t}$, cryptocurrency t is predicted to be in the default group, otherwise it is predicted to be in the functioning group. Given the subsamples, we defined $f_{1,t} > f_{2,t}$ as *success* for the default group and $f_{1,t} < f_{2,t}$ as *success* for the functioning group meaning that the discriminant model correctly assigned the respective cryptocurrency to its corresponding group. Furthermore, for each group, we coded two vectors of dummy variables denoted as d_1 and d_2 that have a value of one in case of *success* and a value of zero otherwise. The prediction accuracy of predicting a cryptocurrency's default within four years is then simply given by $\sum_{t=1}^{T_1} d_{1,t}/T_1$, whereas the Type-I error of this model is $1 - \sum_{t=1}^{T_1} d_{1,t}/T_1$. In the same manner, the prediction accuracy for predicting a cryptocurrency's continued functioning can be calculated as $\sum_{t=1}^{T_2} d_{2,t}/T_2$, while the Type-II error is then given by $1 - \sum_{t=1}^{T_2} d_{2,t}/T_2$.

Setting up the empirical model is ultimately an empirical question. After our initial analysis of differences in sample means, we decided to employ the following cardinal variables in our discriminant analysis: Day-1 return, Month-1 return, and the corresponding volatilities, PMTTCR, and minimum reward. We also account for a set of dummy variables for measuring the qualitative variables *algorithm, anonymous founder*, and *pre-mined*. Specifically, we employ K=9 predictor variables in our analysis. Since we have $T_1 = 86$ cryptocurrencies that ended up in default and $T_2 = 60$ that remained functioning, our matrices X_1 and X_2 have the dimension 86x12 and 60x12, respectively. Our model operates with 12 instead of nine columns because we employ three dummy variables for indicating the algorithm (scrypt, SHA, or others), one binary dummy variable for indicating whether a cryptocurrency is pre-mined, one dummy variable indicating whether the founder is anonymous, and continuous variables for Day-1 return, Day-1 volatility, Month-1 return, Month-1 volatility, actual amount of pre-mined coins, actual amount of minimum reward, and PMTTCR. The estimated discriminant functions are reported in

Table 4. Predicting cryptocurrency default.

	Predicted group by the Multiple Linear Discriminant Function	
Actual Group	Default group	Functioning group
Default group	87.21%	12.79%
Functioning group	43.33%	56.67%

Note: This table reports the results of our multiple linear discriminant analysis. Our dataset consists of all cryptocurrencies that incorporated the Proof-of-Work mechanism and started trading between 2009 and the end of 2014. We followed up those cryptocurrencies until the end of 2018. We retrieved 146 cryptocurrencies, of which 86 went into default in the sample period and 60 remained functioning. Our model incorporates the following predictor variables: minimum reward, pre-mined, Day-1 return, Month-1 return, Day-1 volatility, Month-1 volatility, and PMTTCR. Moreover, we include a set of dummy variables for indicating 'algorithm' and 'founder anonymity'.

Tables A4 and A5 in the appendix. Finally, these estimates are used to calculate the results reported in Table 4. For example, from Table A4, we learn that the discriminant function correctly predicts in 75 out of 86 cryptocurrencies that they are in group 1 because the value of the discriminant function is larger for group1 than for group 2. Consequently, 87.21% of cryptocurrency defaults are predicted correctly.

Given the data of bankrupt and functioning cryptocurrencies, as reported in Table A6 that are in either the default group or the functioning group our model is able to correctly predict 87% of the defaults corresponding to a Type-I error of 13%. Our estimates are close to models that predict bankruptcy of enterprises. For instance, the popular multiple discriminant model from Altman (1968) predicted 94% of bankruptcies of U.S. firms in the manufacturing industry. It is important to note, however, that first Altman's (1968) benchmark model uses recent information on those companies investigated because he employed data from balance sheets that were released about one year before the bankruptcy occurred. Second, he matched that sample of companies that went bankrupt with a sample of matched companies having the same number of firms and the same firm characteristics, whereas our analysis accounts for the whole sample of available cryptocurrencies. Furthermore, we use only information available at an early stage, that is, after one month of trading. Our model predicts bankruptcy within the next four years, which is very different from Altman's findings. Even though Altman's (1968) model performed remarkably well for a one- and two-year period prior to bankruptcy, a robustness check shows that its success rate is only 29% for a four-year period.[12] Even though our results suggest that our cryptocurrency default prediction model is an accurate forecaster of failure, Table 4 shows that the Type-II error is 43%. This result implies that our model struggles to predict functioning cryptocurrencies.

Robustness checks

Since we only have one sample available, our estimates reported in Table 4 are only point estimates. To investigate how sensitive our model is with respect to resampling and to compound confidence intervals for our estimates, we employed bootstrapping. It seems reasonable to assume that characteristic k of cryptocurrency t is uncorrelated with the characteristic k of cryptocurrency s, that is, $cov(x_{t,k}, x_{s,k}) = 0$.[13] However, characteristic k of cryptocurrency t is not necessarily uncorrelated with characteristic l, meaning $cov(x_{t,k}, x_{t,l}) \neq 0$. We have ensured this is ex-ante by simply the way we chose our research set-up because all cryptocurrencies have the same consensus protocol and are therefore homogenous. However, characteristics of a cryptocurrency could be – at least potentially – correlated with other characteristics of the same cryptocurrency. To account for this issue, we employed a pairs bootstrap as detailed by Godfrey (2009, pp.183–185). In doing so, we constructed new data matrices defined as X_1^b and X_2^b where each row vector in X_1 and X_2 is randomly resampled with replacement where each row in X_1 and X_2 is drawn with probability $1/T_1$ and $1/T_2$ respectively. We employ $B = 5000$ bootstrap samples and re-estimate the corresponding discriminant functions to estimate the empirical confidence interval.

[12]The average success rate of Altman's (1968) model between year one and four prior to bankruptcy is 61%.
[13]Note that Altman (1968, 592) highlights that 'there is reason to believe that some of the measurements will have a high degree of correlation or collinearity with each other'.

More concretely, for each bootstrap run b, we employ the original data matrix X_1 that has the dimension 86x12, as described in section 2.1. Then, we randomly draw with replacement and with probability 1/86 = 0.0116 a row from matrix X_1 and add that row into matrix X_1^b to construct a new data matrix. For each run b, this procedure is stopped when all 86 rows in the new data matrix X_1^b are filled. In the same manner, for each bootstrap run b, we employ the original data matrix X_2 that has the dimension 60x12, as described in section 2.1. Then, we randomly draw with replacement and with probability 1/60 = 0.0167 a row from matrix X_2 and add that row into matrix X_2^b to construct a new data matrix. For each run b, this procedure is stopped when all 60 rows in the new data matrix X_2^b are filled. These new data matrices are used to run the linear discriminant analysis described in section 2.1 for each bootstrap iteration $b = 1, \ldots, 5000$. The corresponding point estimates are stored in vector. These vectors are sorted in an increasing order. Then, the 125th observation gives us the value of the lower bound and the 4875th observation gives us the upper bound of our confidence interval covering 95% probability.

The results of our analysis can be found in Table 5. Using bootstrapping, the 95% confidence interval for our point estimate concerning successfully predicting cryptocurrency default is between 83.72% and 89.53% again suggesting a high level of accuracy. However, the 95% confidence interval for the Type-II error ranges between 41.67% and 53.33%. Assuming that the point estimate for the Type-II error is distributed as $N(\mu_1, \sigma)$ with $\mu_1 = 47.50$ and $\sigma = 2.97$, and that the corresponding point estimate for the correctly predicted functioning cryptocurrencies is distributed as $N(\mu_2, \sigma)$ with $\mu_2 = 52.50$, 60.99% of the confidence intervals are overlapping.[14] This result implies that our model overpredicts defaults in the sample of functioning cryptocurrencies.

Table 5. Estimated confidence intervals using bootstrapping.

	Predicted group by the Multiple Linear Discriminant Function	
Actual group	Default group	Functioning group
Default group	[83.72%; 89.53%]	[10.47%; 16.28%]
Functioning group	[41.67%; 53.33%]	[46.67%; 58.33%]

Note: This table reports the results of $B = 5000$ bootstrap replications using a pairs bootstrap. We constructed new data matrices by random resampling with replacement using a probability of $1/T_1$ for the subsample of default cryptocurrencies and a probability of $1/T_2$ for the subsample of functioning cryptocurrencies. Then, we re-estimated our model B times and sorted the estimated probabilities in an increasing order. The 125th observation gives us the value of the lower bound and the 4875th observation gives us the upper bound of our confidence interval covering 95% probability.

It is important to note that the new digital financial markets evolve fast over time. For instance, during 2016 few cryptocurrencies were launched implementing SHA algorithms.[15] Moreover, we would like to stress that new cryptocurrencies are applying more advanced algorithms and security mechanisms; there are only few new cryptocurrencies implementing the SHA hashing algorithms and PoW mechanism because these methods have slower speeds and higher energy consumption. Moreover, as technology advances so does the blockchain too. The most common algorithms for the cryptocurrencies issued before 2015 were SHA and Scrypt, but new cryptocurrency algorithms like X11–X17 were created specifically for Graphical Processing Unit (GPU) mining and provide good profit levels to the Portable Instant Mining Platform (PiMP) community since the rise of large Application-Specific Integrated Circuits (ASICs) for Scrypt. Table A7 in the appendix provides a brief overview of those new algorithms X11–X17. New research needs to account for technological changes associated with cryptocurrencies.

In the same way like Altman (1968) proposed in his seminal paper a model that had the ability to predict bankruptcies for firms in a specific industry in the U.S. (e.g., manufacturing industry), our paper takes the first step in exploring predictable patterns in defaults in new emerging digital financial markets. Expecting that we could propose a universal model being capable of predicting defaults across different categories of digital asset would be an illusion. As pointed out in Howell, Niessner, and Yermack (2019) there are three types of digital assets.

While cryptocurrencies that are defined as 'general-purpose medium of exchange and store of value cryptocurrency' can be considered as alternative to traditional fiat currency, utility tokens or

[14]Note, $\mu_1 = \frac{(41.67+53.33)}{2} = 47.50$, $\sigma = \frac{(53.33-47.50)}{1.96} = \frac{(58.33-52.50)}{1.96} = 2.9745$, $\mu_2 = \frac{(46.67+58.33)}{2} = 52.50$.

[15]See footnote 8.

security tokens have a very different purpose, which is financing business projects. Due to their nature, those digital assets have very different characteristics compared to cryptocurrencies. While our paper specifically governs cryptocurrencies that incorporate the PoW consensus protocol – which obviously was the dominant consensus protocol in our sample of investigation – future research is needed to explore the predictability of defaults for either cryptocurrencies that follow other consensus protocols (e.g., Proof-of-Stake or Hybrid), or other types of digital currencies, such as tokens issued in Initial Coin Offerings.

Moreover, the research methodology of our paper is related to the literature applying MLDA to predict various types of financial failure (e.g., Altman 1968, 1983, 2002; Altman, Hartzell, and Peck 1995; Altman, Haldeman, and Narayanan 1977; Altman, Danovi, and Falini 2013; Altman and Rijken 2010). However, there are other strands of literature dealing with analysing credit risks and employ methodologies such as Probit/Logit models. Future research is encouraged to investigate the predictability of defaults using other methodologies than MLDA also.

III. Conclusion

In this age of digital finance, investors can now choose from more than 2,000 cryptocurrencies to invest in. Among cryptocurrencies that started trading prior to December 2014, we found 59% went bankrupt by the end of 2018. This paper proposes a model to predict cryptocurrency default. We downloaded data for all cryptocurrencies launched between January 2009 and December 2014 and established if they went bankrupt by December 2018. We explored almost two dozen cryptocurrency-specific variable candidates that might serve as predictor variables. From those variables, we only used data for the model setup available to the naïve investor one month after a new cryptocurrency started trading. For each of the selected variables, we estimated the sample means for both the default sample and the functioning sample. Our model correctly predicts 75 of 86 bankruptcies (87%). Employing bootstrapping established that the estimates are statistically significant. Notably, the new digital asset markets are subject to technological changes: For instance, many cryptocurrencies issued after 2015 adopted the PoS (minting) consensus mechanism due to greater energy consumption of PoW (mining). Other major changes that we identified are among others that new cryptocurrencies adopt more efficient and profitable algorithms like X11, X12 and X17 over previously popular algorithms like SHA and Scrypt. Therefore, future research on such technological changes is warranted. Nevertheless, our proposed model could be employed in the asset management industry, for instance, as a screening tool for investment decision-making. A rational investor would avoid investing in digital assets exhibiting overly high default risks. Portfolios that pre-condition the set of digital assets on those cryptocurrencies that are not predicted as at risk of bankruptcy might generate a better risk-return profile for investors.

Acknowledgments

We would like to thank Simon Moore for a lively, interesting, and detailed discussion of our paper in the U.S. Business Magazine *Forbes* entitled “How To Tell If Your Cryptocurrency Will Go Bust” on 28 May 2019. Moreover, valuable comments were received from participants of the *2019 Graduate School Seminar* at the University of Vaasa. We also received valuable comments from the *2019 Aalto University Graduate School of Finance Workshop* at the University of Jyväskyla. In particular, we would like to thank Mika Veihekoski for an encouraging discussion and helpful comments. Furthermore, we are grateful for having received valuable comments from the *2019 Economics & Finance Seminar* at Hanken School of Economics. Especially we would like to thank Timo Korkeamäki, Kenneth Högholm and Emilia Vähämaa for useful comments. Moreover, we received interesting comments from the participants from the *2019 Blockchain Seminar* at the University of Vaasa. We would like to thank Robert Faff and the participants of the FINANCE, PROPERTY, TECHNOLOGY, AND THE ECONOMY CONFERENCE 2019, University of South Australia. Finally, we would like to thank an anonymous reviewer for helpful comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Altman, E. 1968. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *Journal of Finance* 23: 189–209.

Altman, E. 1983. *Corporate Financial Distress*. New York: Wiley Interscience.

Altman, E. 2000, "Predicting Financial Distress of Companies: Revisiting the Z-score and Zeta Models." *New York University Working Paper*.

Altman, E. 2002. "Revisiting Credit Scoring Models in a Basel 2 Environment." *Salomon Center for the Study of Financial Institutions* 2: 2–37.

Altman, E., J. Hartzell, and M. Peck. 1995. *Emerging Markets Corporate Bonds: A Scoring System*. New York: Wiley and Sons.

Altman, E. I. 2018. "Applications of Distress Prediction Models: What Have We Learnt after 50 Years from the Z-Score Models?" *International Journal of Financial Studies* 6: 1–15. doi:10.3390/ijfs6030070.

Altman, E. I., A. Danovi, and A. Falini. 2013. "Z-Score Models' Application to Italian Companies Subject to Extraordinary Administration." *Journal of Applied Finance* 23: 128–137.

Altman, E. I., and H. Rijken. 2010. "Assessing Sovereign Debt Default Risk: A Bottom-up Approach." *Journal of Applied Corporate Finance* 41: 1–30.

Altman, E. I., R. G. Haldeman, and P. Narayanan. 1977. "ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations." *Journal of Banking and Finance* 1 (1): 129–154.

Ardia, D., K. Bluteau, and M. Rüede. 2018. "Regime Changes in Bitcoin GARCH Volatility Dynamics." *Finance Research Letters* 29: 266–271. doi:10.1016/j.frl.2018.08.009.

Balcilar, M., E. Bouri, R. Gupta, and D. Roubaud. 2017. "Can Volume Predict Bitcoin Returns and Volatility? A Quantiles-based Approach." *Economic Modelling* 64: 74–81.

Baur, D. G., and T. Dimpfl. 2018. "Asymmetric Volatility in Cryptocurrencies." *Economics Letters* 173: 148–151. doi:10.1016/j.econlet.2018.10.008.

Borri, N. 2019. "Conditional Tail-risk in Cryptocurrency Markets." *Journal of Empirical Finance* 50: 1–19. doi:10.1016/j.jempfin.2018.11.002.

Catania, L., S. Grassi, and F. Ravazzolo. 2019. "Forecasting Cryptocurrencies under Model and Parameter Instability." *International Journal of Forecasting* 35: 485–501. doi:10.1016/j.ijforecast.2018.09.005.

Chaim, P., and M. P. Laurini. 2019. "Is Bitcoin a Bubble?" *Physica A: Statistical Mechanics and Its Applications* 517: 222–232. doi:10.1016/j.physa.2018.11.031.

Cheah, E. T., and J. Fry. 2015. "Speculative Bubbles in Bitcoin Markets? An Empirical Investigation into the Fundamental Value of Bitcoin." *Economics Letters* 130: 32–36. doi:10.1016/j.econlet.2015.02.029.

Dwyer, G. P. 2015. "The Economics of Bitcoin and Similar Private Digital Currencies." *Journal of Financial Stability* 17: 81–91. doi:10.1016/j.jfs.2014.11.006.

Dyhrberg, A. H. 2016. "Hedging Capabilities of Bitcoin. Is It the Virtual Gold?" *Finance Research Letters* 16: 139–144. doi:10.1016/j.frl.2015.10.025.

Fry, J., and E. T. Cheah. 2016. "Negative Bubbles and Shocks in Cryptocurrency Markets." *International Review of Financial Analysis* 47: 343–352. doi:10.1016/j.irfa.2016.02.008.

Godfrey, L. 2009. *Bootstrap Test for Regression Models*. New York: Palgrave MacMillan.

Howell, S. T., M. Niessner, and D. Yermack 2019. "Initial Coin Offerings: Financing Growth with Cryptocurrency Token Sales." *Working paper, Leonard N. Stern School of Business.*

Katsiampa, P. 2017. "Volatility Estimation for Bitcoin: A Comparison of GARCH Models." *Economics Letters* 158: 3–6. doi:10.1016/j.econlet.2017.06.023.

Kethineni, S., and Y. Cao. 2019. "The Rise in Popularity of and Associated Criminal Activity." *International Criminal Justice Review* 105756771982705. forthcomingCryptocurrency. doi:10.1177/1057567719827051.

Lahmiri, S., and S. Bekiros. 2019. "Cryptocurrency Forecasting with Deep Learning Chaotic Neural Networks." *Chaos, Solutions and Fractals* 118: 35–40. doi:10.1016/j.chaos.2018.11.014.

Li, -Z.-Z., R. Tao, C.-W. Su, and O.-R. Lobont. 2019. "Does Bitcoin Bubble Burst?" *Quality and Quantity* 53: 91–105. doi:10.1007/s11135-018-0728-3.

Lugovskaya, L. 2010. "Predicting Default of Russian Smes on the Basis of Financial and Non-financial Variables." *Journal of Financial Services Marketing* 14 (4): 301–313.

Maume, P. 2019. "Initial Coin Offerings and EU Prospectus Disclosure." *European Business Law Review* 31 (2): 185–208.

Omane-Adjepong, M., P. Alagidede, and N. K. Akosah. 2019. "Wavelet Time-scale Persistence Analysis of Cryptocurrency Market Returns and Volatility." *Physica A: Statistical Mechanics and Its Applications* 514: 105–120. doi:10.1016/j.physa.2018.09.013.

Osterrieder, J., and J. Lorenz. 2017. "A Statistical Risk Assessment of Bitcoin and Its Extreme Tail Behavior." *Annals of Financial Economics* 12: 1750003. doi:10.1142/S2010495217500038.

Rauchs, M., and G. Hileman, 2017. *Global Cryptocurrency Benchmarking Study (No. 201704-gcbs)*. Cambridge Centre for Alternative Finance, Cambridge Judge Business School, University of Cambridge.

Satish, Y. M., and B. Janakiram. 2011. "Turnaround Strategy Using Altman Model as a Tool in Solar Water Heater Industry in Karnataka." *International Journal of Business and Management* 6: 199–206.

Shen, D., A. Urquhart, and P. Wang. 2019. "Does Twitter Predict Bitcoin?" *Economics Letters* 174: 118–122. doi:10.1016/j.econlet.2018.11.007.

Snedecor, G. W., and W. G. Cochran. 1989. *Statistical Methods.* Eighth ed. Ames (Iowa), USA: Iowa State University Press.

Urquhart, A. 2016. "The Inefficiency of Bitcoin." *Economics Letters* 148: 80–82. doi:10.1016/j.econlet.2016.09.019.

Wang, Y., and M. Campbell. 2010. "Do Bankruptcy Models Really Have Predictive Ability? Evidence Using China Publicly Listed Companies." *International Management Review* 6 (2).

Appendix

Table A1. Cryptocurrency characteristics.

Category	Details
Resource and history	Website, announcement, whitepaper, block explorer, github, etc.
Coin specifications	Coin name, type, founder(s), contributor(s), block time, security mechanism, algorithm, staking maturity, block size, launch type, etc.
Daily trading, supply and distribution	rank, market cap, price ($), price (BTC), volume(24 h), market dominance (Volume, Value), supply (max, total, circulating), etc.
Economics	Block reward, inflation, fees recipient, funding model, etc.
Privacy	Cryptographic privacy, sender privacy, recipient privacy, hidden transaction amount, transaction link broken, balances visible, anonymous holdings, network trust required, quantum-proof privacy, trusted setup, auditable supply, mobile privacy, etc.
Features and scaling	Instant send, protocol level, governance, voters, multi-signature support, scaling model, transparent transaction size (bytes), private transaction size (bytes), most throughput in a block, prunable blockchain, etc.
Wallets	Ledger, trezor, coinomi, jaxx, native mobile wallet binaries for all major OS, webwallet, etc.
Network and masternodes	Largest miner or pool, entities controlling, staking supply, public nodes, masternodes, masternode cost (coin), masternode cost ($), etc.
Community	Percentage of active users, number of subscribers, facebook likes, Twitter followers, Alexa rank, Google/Bing searches, etc.

Note: This table provides an overview of different features and characteristics of a cryptocurrency.
(Source: https://news.bitcoin.com).

Table A2. Potential cryptocurrency-specific variables for the model.

Potential categorical variable candidates	
1. Security mechanism	PoW/PoS/Hybrid/Others
2. Launch type*	Standard/ICO/Fork/Coinswap/others
3. Algorithms	SHA/Scrypt/others
4. Funding model*	ICO/donations/founders/others
Potential Binary Variable Candidates	
5. Pre-mined, **6.** Privacy choice*, **7.** Sender privacy*, **8.**Recipient Privacy*, **9.** Network trust required*, **10**. Multi-signature Support*, **11.** Founder anonymity	YES/NO
Potential Continuous Variable Candidates	
12. Block time, **13.** Block reward, **14.** Block size*, **15.** Pre-mined ratio, **16.** Total coins, **17**. Volume*, **18.** Return, **19.** Volatility, **20**. Reward percentage	

Note: This table provides an overview of different quantifiable (continuous/categorical/binary) cryptocurrency-specific variables that could potentially be used to develop a model. Our model incorporates 9 variables from the 20 candidates. The remaining variables were excluded owing to the non-accessibility of websites for many dead coins highlighted with an asterisk (*).

Table A3. Descriptive statistics of functioning and bankrupt cryptocurrencies.

	N	Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	Jarque-Bera
Panel A. Descriptive statistics of the functioning sample									
PMTTCR	60	0.0041	0.0000	0.0842	0.0000	0.0153	4.5660	23.2754	1236.2090
MTTCR	60	0.0000	0.0000	0.0004	0.0000	0.0000	7.3888	56.3109	7651.0780
Ret_D1	60	0.7124	−0.3779	18.3576	−0.9655	3.1486	3.7353	18.8502	767.6010
Ret_W1	60	0.2849	−0.0753	16.2296	−0.8501	2.2079	6.4717	46.9459	5246.9390
Ret_M1	60	0.1197	0.0000	2.6973	−0.7500	0.5512	2.6563	11.9317	269.9996
Vol_D1	60	10.2559	0.4855	337.0018	0.0000	45.5486	6.4127	45.7924	4989.2090
Vol_W1	60	4.8749	0.0694	263.4009	0.0000	33.9713	7.5306	57.8118	8077.9110
Vol_M1	60	0.3131	0.0116	7.2756	0.0000	1.1050	5.0058	29.2786	1976.9990
Block time	60	152.9167	60.0000	600.0000	5.0000	185.9005	1.7929	4.6947	39.3247
Minimum reward	60	3377.0640	50.0000	100000.0000	0.0000	14473.7900	5.6653	36.1092	3061.5140
Pre-mined	60	6.1400E+08	0.0000E+00	3.6800E+10	0.0000E+00	4.7500E+09	7.5509E+00	5.8017E+01	8.1373E+03
Total coins	60	2.9500E+10	8.4000E+07	5.0000E+11	4.2000E+01	1.0000E+11	3.8711E+00	1.6909E+01	6.3351E+02
Panel B. Descriptive statistics of the default sample									
PMTTCR	86	0.0152	0.0000	0.5000	0.0000	0.0754	6.1213	39.2322	5241.1920
MTTCR	86	0.0000	0.0000	0.0024	0.0000	0.0003	9.0733	83.5495	24429.4600
Ret_D1	86	0.0403	−0.4245	13.4928	−0.9815	1.7926	5.4962	39.2686	5146.5500
Ret_W1	86	0.2541	−0.0911	13.4928	−0.9827	1.6441	6.3551	50.4816	8657.5260
Ret_M1	86	0.2454	0.0405	4.6154	−0.4512	0.7490	3.7423	18.5781	1070.3270

(*Continued*)

Table A3. (Continued).

	Panel A. Descriptive statistics of the functioning sample								
	N	Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	Jarque-Bera
Vol_D1	86	3.1776	0.3235	182.0544	0.0000	19.8910	8.6488	77.9195	21185.1800
Vol_W1	86	2.7361	0.1414	182.0544	0.0000	19.6609	8.9798	82.3686	23728.5600
Vol_M1	86	0.6147	0.0152	21.3018	0.0000	2.6716	6.1461	44.6106	6745.7830
Block time	86	160.7907	60.0000	3600.0000	10.0000	407.9158	7.2223	60.3602	12537.5000
Minimum reward	86	65880.6500	50.0000	5000000.0000	0.2500	541325.8000	8.9668	82.1671	23610.7400
Pre-mined	86	4.8800E+07	0.0000E+00	1.8200E+09	0.0000E+00	2.3400E+08	6.1705E+00	4.3285E+01	6.3612E+03
Total coins	86	1.4000E+10	1.0000E+08	5.5000E+11	3.2000E+04	6.3300E+10	7.3333E+00	6.1404E+01	1.2994E+04

Note: This table reports the descriptive statistics for the following 12 variables: Pre-Mined-To-Total-Coin-Ratio (PMTTCR), Minimum-Reward-To-Total-Coin-Ratio (MTTCR), first-day return (Ret_D1), first-week return (Ret_W1), first-month return (Ret_M1), first-day volatility (Vol_D1), first-week volatility (Vol_W1), first-month volatility (Vol_M1), block time (in seconds), minimum reward for the miners per block (minimum reward), number of coins mined before issued to the public (pre-mined), and the total number of coins of the cryptocurrency (total coins). The figures are reported for both, the running sample (Panel A) and the default sample (Panel B).

Table A4. Discriminant function for the default group.

t	Predicted group 1	Predicted group 2	t	Predicted group 1	Predicted group 2
1	27.3805	27.7258	44	24.5862	23.9807
2	25.7955	26.5482	45	25.1247	24.3461
3	29.1404	28.5218	46	23.6011	24.0290
4	28.5050	28.3511	47	23.7185	24.0802
5	28.4949	28.3009	48	26.6271	27.7649
6	26.7095	25.0682	49	23.6858	24.0478
7	29.0677	29.6769	50	22.7344	22.5435
8	27.3892	27.7573	51	24.1190	23.7822
9	28.7611	28.5356	52	24.6971	24.2462
10	23.3757	22.8601	53	25.1351	24.5761
11	26.2717	23.9860	54	24.8930	24.4421
12	−24.1924	−24.2417	55	23.4671	23.1464
13	24.5396	23.8052	56	25.5403	24.3222
14	25.0503	24.3044	57	25.0424	23.9827
15	25.2780	24.4149	58	27.0499	25.4708
16	28.3193	28.2433	59	29.1556	28.5972
17	28.7954	28.5569	60	25.4615	24.2150
18	27.1257	25.3986	61	25.7609	24.4991
19	27.4460	25.7223	62	25.6308	24.3650
20	24.9086	23.8277	63	25.5393	24.4207
21	28.8249	28.2669	64	24.8019	23.8088
22	24.3948	23.4981	65	26.0950	24.7936
23	25.5785	24.3492	66	24.2334	23.3495
24	24.5392	23.7033	67	25.4390	24.2220
25	25.5527	24.3702	68	25.3581	24.1425
26	25.2765	24.4344	69	25.9102	24.6305
27	25.1694	24.2734	70	26.1701	24.7816
28	27.1309	24.7269	71	23.4963	23.2327
29	25.4649	24.5927	72	25.0475	24.2988
30	25.4501	24.6327	73	24.6484	23.9325
31	24.4508	24.0105	74	25.4414	24.5399
32	26.4005	25.3303	75	25.3188	24.4492
33	25.0495	24.2478	76	26.1223	25.0992
34	25.7881	24.8164	77	24.1029	25.0230
35	24.8941	24.1895	78	24.9693	24.2222
36	28.7836	28.5683	79	25.9494	24.9470
37	27.7893	27.7997	80	24.8711	24.0899
38	25.4256	24.4933	81	25.8052	24.8806
39	24.8828	24.1603	82	23.0776	22.8346
40	23.6977	23.4344	83	24.9735	23.4067
41	28.5788	28.3744	84	24.4830	23.9345
42	28.5522	28.3965	85	24.4737	23.8478
43	28.6334	28.1262	86	24.1142	24.5685

Note: This table reports the values for the discriminant function (Equation 8.a) for the default group (e.g., group 1).

Table A5. Discriminant function for the functioning group.

t	Predicted group 1	Predicted group 2	t	Predicted group 1	Predicted group 2
1	27.82065	27.73309	31	23.81871	23.66742
2	28.21115	28.06119	32	22.69601	22.69963
3	27.56446	27.8617	33	25.47647	24.56926
4	24.88217	26.79222	34	26.46251	27.41852
5	24.18907	23.8235	35	24.05865	24.03279
6	27.64701	27.89729	36	23.55639	23.60259
7	28.47438	28.74628	37	26.2968	27.56041
8	28.23995	28.78224	38	26.63752	27.78143
9	28.43512	28.02157	39	23.94774	24.32457
10	27.85306	27.72618	40	27.08734	28.17422
11	24.93163	23.89347	41	26.38169	27.58451
12	25.24417	24.02586	42	26.65863	27.79309
13	24.07879	23.36516	43	26.46865	27.67717
14	25.39226	24.0771	44	21.97226	22.86871
15	25.19025	24.36664	45	28.81073	29.50451
16	31.00211	30.156	46	24.47789	24.03402
17	24.91486	24.19066	47	24.29782	25.81669
18	25.57118	24.62532	48	25.75796	24.53365
19	26.27598	25.23614	49	26.17337	25.1366
20	25.00121	24.35303	50	28.75249	28.56327
21	28.64485	28.4349	51	24.91487	24.19067
22	26.96299	27.15382	52	25.01501	24.15877
23	24.91487	24.19067	53	25.19886	24.33645
24	24.4346	23.66766	54	24.09113	24.42255
25	26.47026	26.85933	55	23.3386	23.74375
26	26.01675	25.00639	56	−23.2155	−22.1288
27	22.73395	23.34834	57	28.3670	28.27678
28	24.55047	23.99539	58	24.27292	23.87617
29	25.37536	24.50556	59	25.17099	24.26115
30	22.85169	23.04122	60	−22.6525	−21.7619

Note: This table reports the values for the discriminant function (Equation 8.b) for the functioning group (e.g., group 2).

Table A6. Name and symbol of cryptocurrencies used for the study.

S.No.	Cryptocurrency	Symbol	S.No.	Cryptocurrency	Symbol	S.No.	Cryptocurrency	Symbol	S.No.	Cryptocurrency	Symbol	S. No.Cryptocurrency	Symbol
Panel A. Name and symbol of running cryptocurrencies													
1	Blakecoin	BLC	13	Fedoracoin	TIPS	25	Reddcoin	RDD	37	Bitcoin	BTC	49 SmartCoin	SMC
2	Maxcoin	MAX	14	Novacoin	NVC	26	NobleCoin	NOBL	38	Peercoin	PPC	50 Gridcoin	GRC
3	Zurcoin	ZUR	15	Spots	SPT	27	Mooncoin	MOON	39	Zetacoin	ZET	51 Lucky7Coin	LK7
4	DimeCoin	DIME	16	Diamond	DMD	28	Phoenixcoin	PXC	40	Unobtanium	UNO	52 42coin	42C
5	Quark	QRK	17	Royalcoin	RYC	29	Fastcoin	FST	41	Bytecoin	BCN	53 Goldcoin	GLD
6	Animecoin	ANI	18	Worldcoin	WDC	30	Argentum	ARG	42	Terracoin	TRC	54 Huntercoin	HUC
7	Primecoin	XPM	19	Mincoin	MNC	31	Florincoin	FLO	43	Namecoin	NMC	55 Curecoin	CURE
8	Vertcoin	VTC	20	Megacoin	MEC	32	Annoncoin	ANC	44	Tekcoin	TEK	56 Stellar	XLM
9	Litecoin	LTC	21	Feathercoin	FTC	33	Grandcoin	GDC	45	Skeincoin	SKC	57 Trollcoin	TROLL
10	Bullion	CBX	22	Dogecoin	DOGE	34	Supercoin	SUPER	46	CDNcoin	CDN	58 SecureCoin	SRC
11	Communitycoin	COMM	23	Galaxycoin	GLX	35	Betacoin	BET	47	Bela Coin	BELA	59 Marscoin	MARS
12	Emerald	EMD	24	BitBar	BTB	36	iXcoin	IXC	48	Redcoin	RED	60 Pandacoin	PND
Panel B. Name and symbol of default cryptocurrencies													
1	Datacoin	DTC	18	Sexcoin	SXC	35	Doubloons	DBL	52	Metiscoin	MTS	69 Cagecoin	CAGE
2	Tagcoin	TAG	19	Xivra	XIV	36	CHNCoin	CNC	53	Unioncoin	UNC	70 Electric	VOLT
3	Nyancoin	NYAN	20	Extremecoin	EXC	37	Globalcoin	GLC	54	Frozencoin	FZ	71 Bottlecaps	CAP
4	Paycoin	XPY	21	Americancoin	AMC	38	Krugercoin	KGC	55	KingdomCoin	KING	72 Neocoin	NEC
5	Infinitecoin	IFC	22	Lottocoin	LOTTO	39	Franko	FRK	56	Memecoin	MEM	73 Bitgem	BTG
6	Qubitcoin	Q2C	23	Graincoin	GRA	40	Netcoin	NET	57	Solcoin	SOL	74 Lebowskis	LBW
7	Freicoin	FRC	24	Xencoin	XNC	41	BBQcoin	BQC	58	Hypercoin	HYC	75 Growthcoin	GRW
8	AllAgesCoin	AAC	25	Batcoin	BAT	42	Catcoin	CAT	59	Craftcoin	CRC	76 Prospercoin	PRC
9	Joincoin	J	26	Junkcoin	JKC	43	Memorycoin	MMC	60	Nanotoken	NAN	77 Hobbitcoin	HBC
10	Cthulhu Offerings	OFF	27	StarCoin	STR	44	Luckycoin	LKY	61	GIL	GIL	78 Hotcoin	HTC
11	Colossuscoin	COL	28	Zenithcoin	ZTC	45	Yacoin	YAC	62	Usecoin	USE	79 Lovecoin	LOVE
12	Particle	PRT	29	HoboNickels	HBN	46	AsicCoin	ASC	63	AlphaCoin	ALF	80 Bells	BEL
13	Pennies	CENT	30	Philosopherstone	PHS	47	Tigercoin	TGC	64	Richcoin	RCH	81 Zeuscoin	ZEU
14	Qqcoin	QQC	31	CACHeCoin	CACH	48	Devcoin	DVC	65	Zedcoin	ZED	82 ELACoin	ELC
15	Applecoin	APC	32	Microcoin	MRC	49	Teacoin	TEA	66	Stablecoin	SBC	83 EzCoin	EZC
16	ZcCoin	ZCC	33	Kittehcoin	MIAOW	50	Copperlark	CLR	67	Socialcoin	SOC	84 Noirbits	NRB
17	Onecoin	ONE	34	Gamecoin	GME	51	Chaincoin	CHC	68	ElephantCoin	ELP	85 Nibble	NBL
												86 Globe	GLB

Note: This table reports the name and symbol of cryptocurrencies with PoW consensus protocol issued prior to the end of year 2014. These cryptocurrencies were tracked until the end of year 2018 and categorized them into running (Panel A) and default (Panel B) samples.

Table A7. Algorithms X11, X13, X14, X15 and X17.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Blake	BMW	Groestl	JH	Keccak	Skein	Luffa	Cubehash	Shavite	SIMD	Echo						
					X11						Hamsi	Fugue				
						X13							Shabal			
							X14							Whirlpool		
							X15								Loselose	Djb2
								X17								

Note: This table shows the chain of different hashing algorithms X11, X13, X14, X15 and X17 with their sub-algorithms (Source: getpimp.org).

Journal of Energy Markets **13**(4), 117–139
DOI: 10.21314/JEM.2020.221

Research Paper

Blockchain consensus protocols, energy consumption and cryptocurrency prices

Niranjan Sapkota and Klaus Grobys

Department of Accounting and Finance, University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland; emails: niranjan.sapkota@uva.fi, klaus.grobys@uva.fi

(Received January 24, 2020; accepted October 14, 2020)

ABSTRACT

Cryptocurrencies employ different consensus protocols to verify transactions. While the "proof-of-work" consensus protocol is the most energy-consuming protocol, "proof-of-stake" and the hybrid of these two consensus protocols, which consume considerably less energy, have also been introduced. We employ portfolio analysis to explore whether energy is a fundamental economic factor affecting cryptocurrency prices. Surprisingly, our results suggest that, on average, cryptocurrencies employing proof-of-work consensus protocols do not generate returns that are significantly different from those that incorporate proof-of-stake consensus protocols. Even more surprising is that our results show that cryptocurrencies that incorporate the hybrid version of these consensus protocols generate significantly higher average returns than the other groups. A possible explanation for this phenomenon may be that the cryptocurrency market is still driven by the trust factor rather than the energy factor.

Keywords: cryptocurrency; Bitcoin; blockchain; consensus protocol; energy efficiency; momentum.

Corresponding author: N. Sapkota

Print ISSN 1756-3607 | Online ISSN 1756-3615

1 INTRODUCTION

In a recent study, de Vries (2019) found that Bitcoin consumed as much electrical energy as all of Hungary in 2018. He also highlighted that even if Bitcoin mining devices could run on renewable energy alone, they would still be discarded as electronic waste at the end of their lifespans. Specifically, the most popular machine on the market, an application-specific integrated circuit (ASIC) miner, cannot be repurposed, because it is hardwired solely for mining Bitcoin. This implies that it is likely to wind up with other cast-off electronics in a landfill or incinerator, causing damage to the environment. Summing up, de Vries's calculations demonstrate that Bitcoin currently generates as much electronic waste as a small nation, such as Luxembourg. Unfortunately, most cryptocurrencies do not have an accurate estimate of their electricity consumption. However, what is known is how many cryptocurrencies make use of certain types of consensus protocol.[1] Not all cryptocurrencies that employ the proof-of-work (PoW) consensus protocol consume as much electricity as Bitcoin. For example, Ethereum (ETH) and Monero (XMR) also follow the PoW consensus protocol, but they consume considerably less electricity than Bitcoin.[2] In general, cryptocurrencies that do not employ the PoW consensus protocol consume less energy per transaction, as pointed out by Nguyen and Kim (2018). Due to the energy inefficiency of PoW mining, proof-of-stake (PoS), a new consensus protocol, emerged. PoS was first introduced by Sunny King and Scott Nadal in 2012. Later, in 2013, Sunny King created the first cryptocurrency to implement this consensus protocol: Peercoin (PPC).[3]

According to Bach *et al* (2018), PoW will eventually be replaced by newer and more efficient algorithms. In this regard, de Vries (2019) mentions the Ethereum blockchain has been planning a transition from PoW to PoS. In terms of energy efficiency, PoW and the hybrid consensus protocol (which is a mixture of PoW and PoS) are inefficient, whereas PoS is highly efficient (Nguyen and Kim 2018). The hybrid consensus protocol (hereafter hybrid) follows a two-stage verification process. In the first stage, a coin is mined using PoW consensus. In the second stage, the authenticity of that coin is validated by stakeholders using PoS consensus. Even though this algorithm applies both consensuses, electricity consumption is significantly lower for hybrid than for PoW. In hybrid, hashing operations are done over

---

[1] Table 1 provides an overview of the numbers and market capitalization of cryptocurrencies using different consensus mechanisms.
[2] See https://digiconomist.net/ethereum-energy-consumption.
[3] Although PPC was the first cryptocurrency to use the PoS protocol, it was implemented along with PoW, which makes PPC a hybrid coin rather than a pure PoS coin.

TABLE 1 Comparison of PoW, PoS and hybrid cryptocurrencies based on their numbers and market capitalization.

(a) Including Bitcoin, Ethereum and Litecoin

	PoW	PoS	Others	Total
Number of cryptocurrencies	517	402	1222	2141
Percentage (%)	24.14	18.78	57.08	100
Market capitalization (US$ billions)	141.38	10.86	36.40	188.64
Dominance (%)	75.14	5.77	19.09	100

(b) Excluding Bitcoin, Ethereum and Litecoin

	PoW	PoS	Others	Total
Number of cryptocurrencies	514	402	1222	2138
Percentage (%)	24.04	18.80	57.16	100
Market capitalization (US$ billions)	12.48	10.86	36.57	59.91
Dominance (%)	20.83	18.12	61.05	100

This table was generated using the information available at cryptoslate.com as of May 7, 2019. Panel (b) excludes the three largest cryptocurrencies under the PoW protocol. “Others” includes hybrids, delegated proof of state (dPoS) and other consensus protocol cryptocurrencies.

a limited search space (more specifically, one hash per unspent wallet output per second) instead of over an unlimited search space as in PoW. Therefore, no significant consumption of energy is involved (King and Nadal 2012). Hence, we can categorize these three consensus mechanisms, PoW, hybrid and PoS, as high-, medium- and low-energy consuming cryptocurrencies, respectively. Further, we would also like to stress that the optimal level of energy consumption also depends on mining rewards and fees. Dimitri (2017) argues that as long as the reward is positive (after deducing the mining costs), the optimal amount of energy consumption differs between miners. Summing up, the energy consumption for PoW is considerably higher than that for cryptocurrencies incorporating PoS and hybrid consensus protocols.[4]

Hayes (2017) argues that, in an economy, the cost of production plays an important role in determining the market price. Likewise, in the case of Bitcoin, anything that reduces the cost of production will have a negative influence on its price. He further argues that mining efficiency lowers the marginal cost of production. After

[4] This issue will be discussed in detail in Section 2.

considering the technological progress as energy efficiency and the size of the mining network as mining difficulty, he concludes that both of these factors have implications for the cost of production and, ultimately, the market value. Similarly, Li and Wang (2017) confirm the relevance of both technological and economic factors in determining Bitcoin market exchange rates by studying a market earlier and later. They find that speculators drive the early market exchange rates, whereas economic factors drive the long-run price dynamics. Further, they confirm that the market price is anchored by mining costs, but the impact of mining difficulties diminishes over the long-run as mining technology advances. Interestingly, Bendiksen *et al* (2018) find that the average marginal costs of Bitcoin creation are significantly lower than the market price of Bitcoin. Contrary to previous studies, Biais *et al* (2018) explore the reflection of fundamental events on the Bitcoin price and show that a large part of the variation in prices does not relate to those fundamental events. They further find that the large fluctuation in Bitcoin prices is driven by the noisy changes in the trust system but not by the fundamentals.

There are few research papers available that study the relationship between cryptocurrency and energy. In a recent case study, Li *et al* (2019) estimate the electricity consumption of Monero mining. Monero, like Bitcoin and Ethereum, uses the PoW consensus protocol. By studying different hashing algorithms, they conclude that mining efficiency largely depends on the hashing algorithms rather than on the consensus mechanism. They also suggest that more studies should be done that investigate the relationship between energy consumption and cryptocurrency mining. Taking into account the financial aspects of cryptocurrency- and energy-related research, Symitsi and Chalvatzis (2018) analyze the spillover effects between Bitcoin, energy and technology companies. They find evidence of short-run volatility spillover from technology stocks to Bitcoin, and of long-run volatility spillover from energy stocks to Bitcoin.

As previous research has mostly considered single digital currencies and single consensus protocols for its case studies, exploring either energy efficiency or energy as a fundamental market driving force for the pricing of cryptocurrencies (Hayes 2017; Li and Wang 2017; Bendiksen *et al* 2018; Biais *et al* 2018; Symitsi and Chalvatzis 2018; Li *et al* 2019), there is no paper available that adopts a portfolio perspective across cryptocurrencies. Our paper fills this gap in the literature. Specifically, for each of the three groups of consensus protocols – PoW, hybrid and PoS – we retrieve weekly cryptocurrency price data for 20 cryptocurrencies that exhibited the highest market capitalization as of January 3, 2016. By doing so, we ensure that our cryptocurrencies exhibit a high level of liquidity. We form equal-weighted portfolios of high-, medium- and low-energy consuming cryptocurrencies that correspond to the PoW, hybrid or PoS consensus protocols, respectively. Since cryptocurrencies incorporating the PoW consensus protocol are more risky than the other groups, because their mining costs are more exposed to changes in energy prices, finance theory suggests that they should compensate investors with higher returns. Hence, we test whether the differences between those average portfolio returns are statistically different from each other. According to finance theory, we expect our PoW portfolio to generate significantly higher portfolio average returns than our hybrid or PoS portfolios if energy is a fundamental risk factor affecting the cryptocurrency market. In the same manner, we also expect our hybrid portfolio to generate significantly higher portfolio average returns than our PoS portfolio. Moreover, we investigate the statistical properties of our three portfolios in more detail and test different autoregressive models for determining the data-generating processes. Since finance theory suggests that, in an efficient market, past price information should not embed information in future returns, we do not expect to find any autocorrelations if the cryptocurrency market exhibits weak-form market efficiency.

Our paper fills some important gaps in the literature. First, our study extends the current literature that investigates the role of economic fundamentals in pricing cryptocurrencies. For instance, our paper expands on those of Biais *et al* (2018) and Li *et al* (2019) in taking a market-wide perspective. While Biais *et al* (2018) and Li *et al* (2019) focus on Bitcoin and Montero as single cryptocurrencies, we employ the portfolio analysis of Fama and French (2008, 2015, 2017), which enables us to make much more generalized, that is, market-wide, conclusions. As Sensoy (2019) finds that liquidity is positively related to market efficiency in cryptocurrency markets, we control for market liquidity by accounting only for those 20 cryptocurrencies that exhibit the highest market capitalization in each group. Moreover, our study empirically tests the argument of Hayes (2017) and Li and Wang (2017) that fundamental factors are drivers of cryptocurrencies' price dynamics. So far, no consensus has been reached on energy efficiency as a market fundamental for cryptocurrencies. While one group of scholars argues that mining costs play an important role in determining the market price (see, for example, Hayes 2017; Li and Wang 2017; Symitsi and Chalvatzis 2018), another group argues that mining costs do not matter in the long run (see, for example, Dimitri 2017; Bendiksen *et al* 2018). Our paper explores this issue from a new angle.

In doing so, our paper is the first to employ portfolio analysis, which has been used to investigate, for instance, the carry trade or momentum effect in traditional currency markets (Lustig *et al* 2011; Menkhoff *et al* 2012a,b) and, more recently, the momentum effect in cryptocurrencies (Grobys and Sapkota 2019). Finally, our paper contributes to the literature that tests the market efficiency of cryptocurrencies. While some scholars argue that Bitcoin is an efficient market (see, for example, Tiwari *et al* 2018; Bariviera 2017; Nadarajah and Chu 2017; Wei 2018), other scholars take a contrary view, that is, they argue that Bitcoin is an inefficient market (see, for example, Zhang *et al* 2018; AI-Yahyaee *et al* 2018; Urquhart 2016). Our paper adds to this literature by taking a market-wide perspective similar to that of Grobys and Sapkota (2019), who employ the whole cross-section of cryptocurrencies. In contrast to Grobys and Sapkota (2019), who use monthly return data, we employ higher-frequency weekly data and analyze our well-diversified energy risk portfolios by running autoregressive models to determine return predictability.

Our results show that cryptocurrencies that employ the PoW consensus protocol do not, on average, generate higher average returns than cryptocurrencies that employ either hybrid or PoS consensus protocols. This finding supports that of Dimitri (2017) and Bendiksen *et al* (2018), who argue that mining costs are not relevant for pricing cryptocurrencies in the long term. Surprisingly, our findings indicate that cryptocurrencies that employ the hybrid consensus protocol generate significantly higher average returns during our sample period than cryptocurrencies that employ either PoW or PoS consensus protocols. Further subsample analysis provides evidence that this effect is not sample specific. One possible explanation for this finding could be that market participants' demand for cryptocurrencies incorporating hybrid technology is relatively higher than that for cryptocurrencies incorporating either PoS or PoW consensus protocols. The hybrid consensus protocol creates a balance between miners and stakeholders that improves the governance of the underlying cryptocurrency. Therefore, cryptocurrencies incorporating hybrid consensus protocols are possibly considered as "more trustworthy". Next, we find that our portfolios sorted by energy consumption exhibit strong patterns of higher-order autocorrelations. Since autocorrelation implies return predictability, our study is in line with the literature arguing that cryptocurrencies are inefficient (Zhang *et al* 2018; AI-Yahyaee *et al* 2018; Urquhart 2016). Finally, we find strong evidence for time series momentum; this is in contrast to Grobys and Sapkota (2019), who do not find any evidence for either cross-sectional or time series momentum. The main difference between our study and that of Grobys and Sapkota (2019) is that we employ weekly data while those authors employ monthly data. As our findings indicate return autocorrelations up to three weeks, our results are not entirely comparable with those of Grobys and Sapkota (2019).

Our paper is organized as follows. The next section takes a closer look at consensus protocols of cryptocurrencies. Section 3 presents our empirical analysis and Section 4 concludes.

2 BACKGROUND

There are now more than 2100 cryptocurrencies being traded at over 18 000 cryptocurrency exchanges around the world, with a Bitcoin market capitalization dominance of 56%.[5] Bitcoin, the first decentralized digital currency, follows the PoW consensus mechanism.[6] There are currently 500 cryptocurrencies in the market that are running on PoW protocols. PoW cryptocurrency's sector dominance based on market capitalization is slightly over 75%.[7] Employing the PoW protocol, miners on a network compete against each other to keep transactions by using a huge amount of electricity. Whoever solves a complicated cryptographic puzzle gets the reward. Since there is a unique solution for each transaction, only one miner is rewarded for every puzzle solved. Unfortunately, PoW in not an energy-efficient consensus protocol.

According to the Bitcoin Energy Consumption Index 2019, Bitcoin's level of energy consumption is equivalent to that of a small country: it places 44th amongst the world's biggest energy-consuming nations. Moreover, one Bitcoin transaction consumes as much electricity as 100 000 Visa (another payment system, which is, however, centralized) transactions. It was estimated that the Bitcoin network would be consuming as much electricity as Denmark by 2020. Further, there are different types of costs associated with mining, such as energy, warehouse maintenance, supercomputers and mining software.[8]

Due to the high energy consumption of PoW, PoS was introduced. There are now over 400 cryptocurrencies operating with the PoS consensus protocol, giving it a market capitalization dominance of 6%.[9] PoS compares the fraction of an asset (eg, cryptocurrency) owned by miners and rewards them based on their percentage of the stake. Instead of relying on a massive amount of power to incentivize the network, PoS relies on the proportion of wealth that miners stake relative to everyone else. Therefore, in the PoS consensus protocol, all of the active miners get their reward based on their stake. As a result, this protocol is far more energy efficient than the PoW consensus protocol.

In addition to being energy efficient, PoS has many potential benefits over PoW: its increased overall security, superior decentralization and transparent pooling consensus are just some examples. Moreover, PoW is vulnerable to a so-called 51% attack, meaning that whoever holds 51% of the entire network's computing power can manipulate the blockchain for their own ends, preventing new transactions from being confirmed. However, with 51% of a cryptocurrency-specific network's stake, one can manipulate the blockchain, too. As a consequence, a pure PoS system

[5] This information according to coinmarketcap.com, as of May 7, 2019.
[6] In PoW, the unique cryptographic puzzle is the "proof" and the energy consumption and other resources used to solve this puzzle are the "work". For more information, see Nakamoto (2008).
[7] This information according to https://cryptoslate.com/cryptos/proof-of-work/, as of May 7, 2019.
[8] See https://digiconomist.net/bitcoin-energy-consumption.
[9] This information according to https://cryptoslate.com/cryptos/proof-of-stake/, as of May 7, 2019.

TABLE 2 Comparison between PoW, PoS and hybrid consensus protocols.

Criteria	PoW	PoS	Hybrid (PoW + PoS)
Energy efficiency	No	Yes	No
Modern hardware	Very important	No need	Important
Forking	When two nodes find the suitable nonce at the same time	Very difficult	Probably
Double spending attack	Yes	Difficult	Yes, but less serious than in PoW
Block creating speed	Low, depends on variant	Fast	Low, depends on variant
Pool mining	Yes, but it can be prevented	Yes, and it is difficult to prevent	Yes
Example	Bitcoin (BTC), Litecoin (LTC)	BitShares (BTS), Nextcoin (NXT)	Dash (DASH), Peercoin (PPC)

Source: Nguyen and Kim (2018).

can be unstable, and stakeholders with 51% could easily generate fake timestamp histories for fake blocks. As a result, a hybrid PoS + PoW consensus protocol has been implemented by many cryptocurrencies to create a balance between miners and stakeholders in order to provide improved governance for the respective cryptocurrency.[10]

More specifically, according to the Bitcoin Energy Consumption Index 2019, the estimated annual electricity consumption of Bitcoin is 61.74 TWh; the estimated consumption for Ethereum is almost 10 times lower, even though both cryptocurrencies use the same consensus protocol. However, there is no exact estimation of how much energy the PoS consensus protocol uses. Comparatively, PoW uses thousands of heavily powered supercomputers, investing millions in infrastructure and energy. It should be noted that when operating with the PoS consensus protocol, one is able to stake something as simple as Raspberry Pi, which uses around 950 mA (5.0 W) of energy with an estimated 43.95 kWh annually.[11] For example, if one cryptocurrency using the PoS consensus protocol uses one million of these mini computers for minting, it will still consume 1000 times less energy than Bitcoin and almost 100 times less than Ethereum.

[10] Table 2 provides an overview of the basic differences between PoW, PoS and hybrid consensus protocols.
[11] For more information, visit www.raspberrypi.org/forums/viewtopic.php?t=18043.

3 METHODOLOGY

Given our sample, we can divide cryptocurrencies into three groups, (1) PoW, (2) hybrid and (3) PoS, depending on the consensus protocol's energy consumption. The cryptocurrencies with the PoW consensus protocol are the highest energy-consuming cryptocurrencies, whereas the cryptocurrencies with the PoS consensus protocol are the lowest energy-consuming ones.

We start our analysis by retrieving weekly data for 20 cryptocurrencies that have the highest market capitalization as of January 3, 2016 for each of these three groups.

The three data sets of cryptocurrencies are reported in Table 3. Due to their enormous market capitalizations, we also consider a data set that excludes Bitcoin, Litecoin and Ethereum from the sample. This data set, which we refer to as the PoW* group, contains 17 cryptocurrencies and exhibits a total market capitalization of US$45 388 634. The total market capitalizations of the hybrid and PoS groups are US$36 089 983 and US$36 229 758, respectively. For these three groups of cryptocurrencies (eg, PoW*, hybrid and PoS), which comprise a total of 57 cryptocurrencies, their relative shares in terms of market capitalization correspond to 38%, 31% and 30%, respectively. As a consequence, the market capitalizations of our three groups are roughly similar.

For each employed cryptocurrency, we retrieve weekly data from January 1, 2016 until December 31, 2018.[12]

3.1 Cryptocurrency portfolios based on energy consumption and hypothesis testing

We start our analysis by forming equal-weighted portfolios for each group. The cumulative return evolutions are plotted in Figure 1.

From Figure 1, we observe that the cumulative return evolutions of our PoW and PoW* groups are very similar due to the chosen equal-weighting scheme, which is common practice in studies that analyze traditional currencies. Interestingly, cryptocurrencies that employ the hybrid consensus protocol generate considerably higher returns than the other groups.

In Table 4, we report the descriptive statistics of our cryptocurrency portfolios. For our hypothesis tests, we employ simple t-statistics of the zero-cost portfolios (PoW-PoS), (PoW*-PoS), (PoW-hybrid) and (hybrid-PoS).

The corresponding point estimates of the zero-cost portfolios are −0.54%, −0.33%, −9.12% and 8.58% per week with corresponding t-statistics of −0.49,

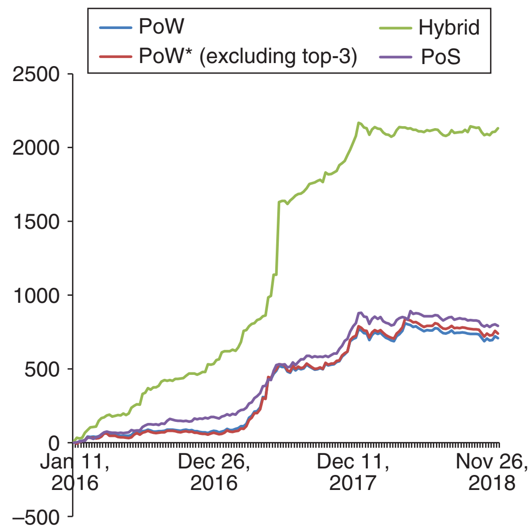
[12] In choosing weekly data, we follow Gutierrez and Kelley (2008). In weekly terms, we can retrieve 156 observations, which makes our statistical inferences more accurate than they would have been if we had employed monthly data.

TABLE 3 Top 20 coins using PoW, PoS and hybrid consensus protocols.

No	PoW coins	Symbol	Market cap	PoS coins	Symbol	Market cap	Hybrid coins	Symbol	Market cap
1	Bitcoin	BTC	6 467 437 080	BitShares	BTS	8 591 688	Dash	DASH	19 794 713
2	Litecoin	LTC	152 873 521	Nxt	NXT	6 863 998	Peercoin	PPC	9 756 959
3	Ethereum	ETH	73 843 278	Factom	FCT	5 646 935	Emercoin	EMC	2 729 184
4	Dogecoin	DOGE	14 940 681	NuShares	NSR	3 176 456	Novacoin	NVC	1 162 587
5	Namecoin	NMC	6 073 338	Rubycoin	RBY	2 763 547	I0Coin	I0C	697 337
6	Bytecoin	BCN	5 582 979	Clams	CLAM	2 199 047	ReddCoin	RDD	577 185
7	Monero	XMR	5 295 952	BlackCoin	BLK	2 000 105	Diamond	DMD	484 841
8	GridCoin	GRC	3 206 756	GlobalCurrency	GCR	1 387 263	I/O Coin	IOC	310 983
9	MonaCoin	MONA	1 627 740	StorjcoinX	SJCX	687 999	CloakCoin	CLOAK	201 995
10	Startcoin	START	1 561 435	MintCoin	MINT	661 925	WhiteCoin	XWC	70 092
11	Tether	USDT	951 600	NuBits	USNBT	849 743	Bitstar	BITS	62 056
12	Primecoin	XPM	900 600	SolarCoin	SLR	522 662	BeanCash	BITB	53 056
13	VeriCoin	VRC	796 319	Rimbit	RBT	330 286	LiteDoge	LDOGE	42 925
14	DigiByte	DGB	740 007	Blocknet	BLOCK	312 057	UltraCoin	UTC	31 851
15	DNotes	NOTE	711 592	FairCoin	FAIR	247 002	Capricoin	CPC	27 313
16	GameCredits	GAME	666 554	HyperStake	HYP	173 539	Kobocoin	KOBO	27 149
17	Quark	QRK	647 684	BitBay	BAY	141 822	Sprouts	SPRTS	25 115
18	WorldCoin	WDC	581 449	AudioCoin	ADC	129 387	Amsterdam	AMS	13 420
19	PayCoin	XPY	581 404	Orbitcoin	ORB	122 492	SuperCoin	SUPER	12 954
20	BoostCoin	BOST	522 544	NavCoin	NAV	121 805	BitSend	BSD	8 268
	Total market cap		45 388 634*			36 929 758			36 089 983
	Market cap share		38.33%			31.19%			30.48%

The top 20 cryptocurrencies using PoW, PoS and hybrid (PoW/Pos) consensus protocols as of January 3, 2016. Data generated using the historical snapshot available at coinmarketcap.com. *Excluding Bitcoin, Litecoin and Ethereum.

FIGURE 1 Cumulative returns of cryptocurrency portfolios sorted by energy consumption.



The evolution of the cumulative returns for the PoW, PoW*, hybrid and PoS portfolio groups from January 1, 2016 until December 31, 2018.

TABLE 4 Descriptive statistics of cryptocurrency portfolios sorted by energy consumption.

	PoW	PoW*	Hybrid	PoS
Mean	4.5354	4.7480	13.6552	5.0736
Median	0.2240	−0.2533	6.4679	2.8859
Maximum	130.9860	147.4168	492.9698	66.7729
Minimum	−44.0888	−45.3746	−45.8466	−45.6972
SD	19.5048	21.2037	45.4339	16.4352
Skewness	2.4676	2.7013	7.8663	0.9413
Kurtosis	14.8132	16.5463	81.2525	5.5172
Jarque–Bera	1 065.4040	1 382.4810	41 411.3300	64.2257
Probability	0.0000	0.0000	0.0000	0.0000

The descriptive statistics for four cryptocurrency portfolios ranked by energy consumption from January 1, 2016 until December 31, 2018. *Excluding Bitcoin, Litecoin and Ethereum. SD denotes standard deviation.

−0.27, −2.76 and 2.50. The t-statistics of the return differences of the (PoW-PoS) and (PoW*-PoS) groups imply that high-energy-consuming cryptocurrencies do not generate higher average returns than low-energy-consuming cryptocurrencies, irrespective of whether or not we include the three cryptocurrencies that exhibit the highest market capitalizations (Bitcoin, Litecoin and Ethereum). On the one hand, this is a surprising result, because finance theory suggests that assets that carry a higher risk should compensate investors with a higher return. The PoW cryptocurrency portfolio is considerably more highly exposed to changes in the energy price. For instance, if the energy price increases, the mining costs will increase as well. In this regard, this result is contrary to that of Hayes (2017) and Li and Wang (2017), who argue that economic factors drive the long-run price dynamics of cryptocurrencies. On the other hand, this result is in line with that of Biais *et al* (2018), who document that a large part of the variation in Bitcoin prices does not relate to economic fundamental events.

Further, the t-statistics of the return differences of the (PoW-hybrid) and (hybrid-PoS), that is, the cryptocurrencies with the hybrid consensus protocol, on average, generate higher average returns than the PoW, PoW* or PoS groups. One could argue that the results are driven by outliers because, as we observe from Table 4, the hybrid group in particular is highly positively skewed. Therefore, we trim the data set and cut the four most extreme observations off the right and left tails, corresponding to 5% of the sample. The observations that were cut off from the sample are reported in Table 10. Then, we employ a simple t-test of the zero-cost portfolios (PoW-PoS), (PoW*-PoS), (PoW-hybrid) and (hybrid-PoS). The point estimates of the zero-cost portfolios are -0.99%, -0.92%, -5.97% and 5.22% per week with corresponding t-statistics of -1.41, -1.24, -4.94 and 4.49. A robustness check shows that even though the economic magnitude of the point estimates of the zero-cost portfolios is lower, the t-statistics are considerable higher for the (PoW-hybrid) and (hybrid-PoS) portfolios. As our previous conclusions remain unchanged, it would appear that our findings are not driven by outliers. Finally, Biais *et al* (2018) argue that the large fluctuation in Bitcoin prices is driven by the noisy changes in the trust system but not by the fundamentals. If trust is a driver of cryptocurrencies, a possible explanation for this phenomenon could be that investors' relative demand for cryptocurrencies incorporating the hybrid consensus protocol is higher than that for cryptocurrencies incorporating PoW or PoS consensus protocols because the hybrid consensus protocol is less likely to be subject to manipulation.

3.2 Statistical analysis of the data-generating cryptocurrency portfolio processes

The next step in our portfolio analysis is to explore the data-generating processes of our PoW, PoW*, hybrid and PoS portfolios in more detail. To investigate any potential autocorrelation, we employ the Akaike information criterion (AIC) to assess the optimal lag order. To estimate the optimal lag order, we choose a maximum lag length

TABLE 5 Estimated autoregressive models.

Variable	Constant	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	R-squared
PoW	2.45**	−0.03	0.29**	0.21***	—	—	0.13
	(2.30)	(−0.56)	(2.22)	(2.75)			
PoW*	2.66**	−0.04	0.29**	0.21**	—	—	0.13
	(2.27)	(−0.71)	(2.35)	(2.60)			
Hybrid	10.06**	−0.00	0.26***	—	—	—	0.07
	(2.33)	(−0.12)	(2.76)				
PoS	2.83***	0.09	0.21***	0.15***	−0.15	0.14**	0.12
	(2.90)	(1.08)	(2.81)	(3.15)	(−1.16)	(2.41)	

The parameter estimates for different autoregressive models. The optimal lag order is determined using the AIC. The Newey–West (1986) t-statistics accounting for five lags are given in parenthesis. The sample period is from January 1, 2016 until December 31, 2018. *Excluding Bitcoin, Litecoin and Ethereum. **Statistically significant at the 5% level. ***Statistically significant at the 1% level.

of 12 weeks. In Table 5, we report the estimated autoregressive model accounting for the optimal lag order with respect to the AIC. Surprisingly, we find that having a second lag that is highly statistically significant while the first lag is not is a trait common to all portfolios. Moreover, each cryptocurrency portfolio has a different autoregressive profile. The results reported in Table 5 also reveal that the stability conditions are fulfilled for all cryptocurrency portfolios (see Lütkepohl *et al* 2004, p. 23).

To investigate whether the regression residuals exhibit any remaining autocorrelation, we employ the Lagrange multiplier (LM) test for autocorrelation. Specifically, for each model, we perform the LM test for autocorrelation successively for one to five lags. The corresponding test statistics are, under the null hypothesis, distributed as chi-squares with one to five degrees of freedom. The results are reported in Table 6.

The LM tests reveal that there is no evidence of any remaining autocorrelation in the regression residuals, which indicates that the chosen lag order (see Table 5) is indeed appropriate for modeling the underlying data-generating process. This result has some important implications. First of all, some studies document that cryptocurrency markets move toward efficiency. Specifically, Vidal-Tomás and Ibañez (2018) and Sensoy (2019) argue that Bitcoin has become more efficient over time. Moreover, Bariviera (2017) documents Bitcoin's informational efficiency since 2014, whereas Sensoy (2019) finds that Bitcoin markets have become more informatively efficient at the intraday level since the beginning of 2016. In another study, Nadarajah and Chu (2017) report that Bitcoin returns do satisfy the efficient market hypothesis. Further, the study of Khuntia and Pattanayak (2018) supports Vidal-Tomás and Ibañez (2018) and Sensoy (2019) in finding that Bitcoin exhibits market efficiency over time, validating the adaptive market hypothesis. In contrast to those studies, we use

TABLE 6 LM test statistics for remaining autocorrelation.

Model	Lags 1	2	3	4	5
PoW	1.87	1.88	2.75	3.80	3.91
	(0.1716)	(0.3915)	(0.4316)	(0.4343)	(0.5611)
PoW*	2.00	2.00	2.91	4.22	4.35
	(0.1577)	(0.3684)	(0.4065)	(0.3767)	(0.4995)
Hybrid	0.45	2.45	2.56	2.99	3.01
	(0.5046)	(0.2942)	(0.4644)	(0.5594)	(0.6991)
PoS	1.79	2.65	2.87	3.61	4.63
	(0.1807)	(0.2665)	(0.4115)	(0.4604)	(0.4619)

Different LM test statistics for the regression residuals of Table 5. Under the null hypothesis, the test statistics are asymptotically distributed as chi-squares with one to five degrees of freedom. The corresponding critical values for the 5% significance levels are 3.84, 5.99, 7.82, 9.49 and 11.07. The corresponding p-values are given in parentheses.

a more recent sample from 2016 to 2018. Since our findings indicate higher-order autocorrelation, our results provide strong evidence that cryptocurrency markets are inefficient. As we employ only the cryptocurrencies that exhibit the highest market capitalizations at the time of portfolio formation, one cannot argue that our results are driven by microcryptocurrencies.

3.3 Implementing time series momentum strategies

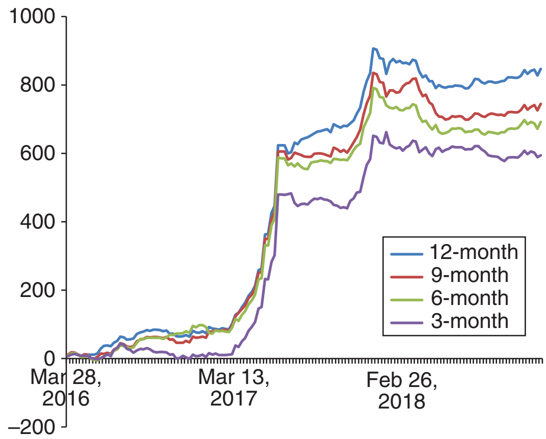
Another implication of autocorrelation among cryptocurrency portfolios is potential time series momentum. Moskowitz *et al* (2012) have proposed a time series momentum strategy that performs well even in different market scenarios. Therefore, we estimate time series momentum (TSMOM) as

$$r^{\mathrm{TSMOM},s}_{t,t+1} = \mathrm{sign}(r^s_{t-K})r^s_{t,t+1},$$

where r^s_{t-K} is the return of security s over the past K months and $r^s_{t,t+1}$ is next month's return, which indicates taking a long position when the sign of the cumulative past K-month return is positive and a short position otherwise. First, we implement strategies for $K = \{12, 9, 6, 3\}$ and employ the whole cross-section of 60 cryptocurrencies. The cumulative return series of our strategies over the sample period are plotted in Figure 2.

On the left-hand side of Table 7, we observe that all of the strategies generate statistically significant average payoffs, ranging from 4.09% per week for $K = 3$ to 5.84% per week for $K = 12$. The payoffs are statistically significant at, at least, the

FIGURE 2 Cumulative returns of time series momentum payoffs.



The evolution of the cumulative returns for the different time series momentum portfolios. The sample period is from March 28, 2016 until December 31, 2018.

5% level. Another interesting finding is that the payoffs of the TSMOM portfolios are linear increasing as we move from $K = 3$ to $K = 12$. One explanation for this phenomenon could be that some individual cryptocurrencies have a higher-order autocorrelation that exceeds an order of five. Therefore, employing a cumulative return window of 12 weeks instead of three might offer more precise information. One may be concerned that the results are driven by outliers. To address this concern, we trim the data and cut off 5% of the observations, that is, the most extreme observations from the right and left tails of the distribution. The results are reported on the right-hand side of Table 7. Even though the portfolios' average payoffs are lower after trimming, the t-statistics increase and show that all average payoffs are statistically significant at, at least, the 1% level.

One may also wonder to what extent the results are driven by Bitcoin, Litecoin and Ethereum, because those three cryptocurrencies dominate the overall market in terms of market capitalization (see Table 3). To address this issue, we exclude those three cryptocurrencies from the sample and repeat the previous analysis. The results are reported on the left-hand side of Table 8. The results are virtually the same. Next, we again trim the data and cut off the most extreme 5% of observations. The observations that were cut off are reported in Table 11. The results for the trimmed data excluding Bitcoin, Litecoin and Ethereum are reported on the right-hand side of Table 8. The conclusions remain unchanged.

TABLE 7 Descriptive statistics of different time series momentum strategies.

Statistic	Strategy: Untrimmed data 12-week	9-week	6-week	3-week	Trimmed data 12-week*	9-week*	6-week*	3-week*
Mean	5.8406*** (3.25)	5.1375*** (2.89)	4.7690*** (2.72)	4.0946** (2.40)	4.0765*** (4.48)	3.3978*** (3.73)	2.7740*** (3.22)	2.4013*** (2.81)
Median	2.1003	1.2360	0.6012	0.3523	2.1003	1.2360	0.6012	0.3523
Maximum	177.5289	176.9290	178.4604	177.2945	47.0611	45.7703	48.5731	43.1194
Minimum	−43.6178	−40.6550	−32.3301	−36.0245	−19.1971	−22.9566	−19.8712	−17.0311
SD	21.6429	21.3942	21.1121	20.5486	10.6520	10.6530	10.0834	9.9986
Skewness	4.3059	4.3729	4.8064	4.8256	1.2205	1.0119	1.2830	1.4688
Kurtosis	31.7670	32.8757	36.0594	38.0875	5.8958	5.7910	7.2381	6.2550
Jarque–Bera	5447.7940	5854.6400	7161.3580	8000.8420	81.8794	67.8476	140.1137	109.7399
Probability	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The descriptive statistics for different cryptocurrency momentum portfolios using untrimmed and trimmed data. The trimmed data cuts off the most extreme observations on the right and left tails of the distribution, corresponding to 5% of the empirical probability distribution. The sample period is from March 28, 2016 until December 31, 2018. *This data is trimmed and excludes the four most extreme time series observations on the left and right tails of the corresponding return distributions. **Statistically significant at the 5% level. ***Statistically significant at the 1% level.

TABLE 8 Descriptive statistics of different time series momentum strategies excluding Bitcoin, Litecoin and Ethereum.

	Strategy							
	Untrimmed data				Trimmed data			
Statistic	**12-week**	**9-week**	**6-week**	**3-week**	**12-week***	**9-week***	**6-week***	**3-week***
Mean	5.9707*** (3.21)	5.2727*** (2.87)	4.8860*** (2.69)	4.1547** (2.34)	4.1088*** (4.42)	3.4477*** (3.72)	2.8014*** (3.19)	2.3748*** (2.71)
Median	2.1498	1.2782	0.8633	0.1245	2.1498	1.2782	0.8633	0.1245
Maximum	185.8179	185.1864	186.7984	185.5712	47.1799	45.8212	48.7715	43.0308
Minimum	−43.9765	−40.8578	−32.9985	−36.4867	−20.4995	−23.3220	−19.9503	−18.3225
SD	22.4212	22.1504	21.8863	21.3493	10.8845	10.8471	10.2639	10.2624
Skewness	4.4225	4.4952	4.9134	4.9222	1.1995	1.0025	1.2498	1.4606
Kurtosis	32.9569	34.1708	37.3207	39.1820	5.7590	5.6745	7.0168	6.1719
Jarque–Bera	5894.5460	6358.5230	7699.9440	8494.8950	76.3038	63.7816	127.7652	106.1410
Probability	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The descriptive statistics for different cryptocurrency momentum portfolios using untrimmed and trimmed data. The data set excludes Bitcoin, Litecoin and Ethereum when implementing the momentum strategies. The trimmed data cuts off the most extreme observations on the right and left tails of the distribution, corresponding to 5% of the empirical probability distribution. The sample period is from March 28, 2016 until December 31, 2018. *This data is trimmed and excludes the four most extreme time series observations on the left and right tails of the corresponding return distributions. **Statistically significant at the 5% level. ***Statistically significant at the 1% level.

3.4 Controlling for the hybrid cryptocurrency effect

In Section 3.1 we have shown that, over the 2016–2018 sample period, cryptocurrencies with the hybrid consensus protocol generated on average significantly higher returns than cryptocurrencies implementing either PoS or PoW technology. This raises the question of whether momentum payoffs are exposed to this cross-sectional phenomenon and, if so, to what extent. To investigate this issue in detail, we regress the payoffs of our time series momentum strategies on a zero-cost portfolio that is long on the cryptocurrency portfolio with the hybrid consensus protocol and short on the cryptocurrency portfolio with the PoS consensus protocol.[13] The regression equation is then simply given by

$$r_{t,t+1}^{\mathrm{TSMOM},s} = c + \gamma(R_{\mathrm{hybrid},t+1} - R_{\mathrm{PoS},t+1}) + e_{t+1},$$

where $r_{t,t+1}^{\mathrm{TSMOM},s}$ denotes the momentum payoff of strategy s at time $t+1$, $R_{\mathrm{hybrid},t+1}$ is the return of the portfolio comprising cryptocurrencies with the hybrid consensus protocol, $R_{\mathrm{PoS},t+1}$ is the return of the portfolio comprising cryptocurrencies with the PoS consensus protocol and e_{t+1} denotes the error term.

We estimate the time series regressions using a Newey–West (1986) covariance estimator with five lags. We report the result for the whole data set and for the data set excluding Bitcoin, Litecoin and Ethereum in Table 9. From Table 9, we observe that – depending on the time series momentum strategy – the exposures against the hybrid factor, defined as $(R_{\mathrm{hybrid},t+1} - R_{\mathrm{PoS},t+1})$, vary between 0.33 and 0.37.

The corresponding t-statistics indicate that those exposures are statistically significant at any level. This result implies that our time series momentum strategies are investment strategies that are invested in cryptocurrencies with the hybrid consensus protocol. As the intercepts become insignificant, after controlling for our hybrid factor, time series momentum strategies implemented in cryptocurrencies' time series momentum strategies do not generate returns in excess of this factor.

On the one hand, our results are in line with Asness *et al* (2013), who explore the pervasiveness of the momentum phenomenon, arguing that momentum payoffs are positively co-moving across otherwise unrelated asset markets. Our results especially extend the study of Menkhoff *et al* (2012a), who explore different momentum strategies implemented among traditional currencies. Our results are, however, contrary to the recent findings of Grobys and Sapkota (2019) because they do not find any momentum effects in cryptocurrencies. It is noteworthy that our study is different from that of Grobys and Sapkota (2019), as they use monthly data and we

[13] One could also implement the short leg as a strategy as $0.5 \cdot \mathrm{PoW} + 0.5 \cdot \mathrm{PoS}$ or alternatively as $0.5 \cdot \mathrm{PoW}^b + 0.5 \cdot \mathrm{PoS}$.

TABLE 9 Controlling for the hybrid cryptocurrency effect.

Strategy	Constant	(Hybrid-PoS)	R-squared
12-month	2.82 (1.42)	0.36*** (26.04)	0.53
9-month	2.10 (1.06)	0.36*** (29.72)	0.55
6-month	1.81 (1.13)	0.35*** (17.14)	0.53
3-month	1.27 (0.75)	0.33*** (10.73)	0.51
12-month*	2.81 (1.38)	0.37*** (26.33)	0.54
9-month*	2.10 (1.04)	0.37*** (30.19)	0.56
6-month*	1.79 (1.12)	0.37*** (17.13)	0.54
3-month*	1.19 (0.70)	0.35*** (10.83)	0.52

The parameter estimates for regression equations that regress different momentum strategies on the hybrid factor, which is a portfolio that is long on cryptocurrencies that incorporate the hybrid consensus protocol and short on cryptocurrencies that incorporate the PoS consensus protocol. Newey–West (1986) t-statistics accounting for five lags are given in parentheses. The sample period is from March 28, 2016 until December 31, 2018. *Excluding Bitcoin, Litecoin and Ethereum. ***Statistically significant at the 1% level.

TABLE 10 Trimmed observations for energy-sorted cryptocurrency portfolios.

(PoW-PoS)	(PoW*-PoS)	(Hybrid-PoW)	(Hybrid-PoS)
−62.6284	−62.5129	−459.372	−56.8231
−26.2100	−29.8355	−107.789	−38.5675
−25.8078	−28.2217	−74.8582	−31.6448
−25.7638	−25.6716	−63.8474	−27.2312
27.88344	30.27623	24.7859	72.3571
30.63994	33.7027	35.82428	78.75633
57.07233	64.70682	52.2831	82.02508
87.16298	103.5939	53.16028	487.2559

employ weekly data. Gutierrez and Kelley (2008), who also use weekly data to investigate short-term momentum in equities, find strong patterns of momentum effects. The findings of Gutierrez and Kelley (2008) are, however, different from those of Jegadeesh and Titman (1993) and Jegadeesh (1990), who employ monthly data, which implies different data frequencies incorporate different types of information.

TABLE 11 Trimmed observations for cryptocurrency momentum portfolios.

12-week	9-week	6-week	3-week
177.5289	176.929	178.4604	177.2945
100.4631	97.91368	96.34356	81.66023
72.01109	66.20229	73.25367	68.439
61.89667	59.80724	60.48544	51.3516
−24.11458	−23.78528	−20.45539	−18.07823
−24.28796	−24.64637	−21.93967	−28.09708
−31.46949	−32.33011	−22.35219	−31.80449
−43.61776	−40.65498	−32.33011	−36.02447

4 CONCLUSION

The mining process of cryptocurrencies consumes an enormous amount of energy. So far, there has been no consensus as to whether energy is a market fundamental for pricing cryptocurrencies. If energy, as an economic factor, has an impact on the market values of cryptocurrencies and plays a role in determining cryptocurrency prices, we would expect our well-diversified portfolio of cryptocurrencies incorporating the high-energy-consuming PoW consensus protocol to generate, on average, higher returns than our well-diversified portfolio of cryptocurrencies incorporating the low-energy-consuming PoS consensus protocol. However, we do not find such evidence. Surprisingly, our results demonstrate that our well-diversified portfolio of cryptocurrencies incorporating the medium-energy-consuming hybrid consensus protocol generates considerably higher average returns than our PoW or PoS portfolios.

One important implication of our findings is that energy consumption does not seem to play a role in pricing cryptocurrencies. Moreover, the price of cryptocurrencies is mostly determined by the demand side because the supply side is controlled by a cryptocurrency-specific algorithm for the majority of coins. Since cryptocurrencies incorporating the hybrid consensus protocol generate significantly higher average returns than cryptocurrencies incorporating PoW or PoS protocols, our results imply that investors' relative demand for the former is greater than that for the latter.

A possible explanation for this phenomenon is that investors perceive cryptocurrencies that employ hybrid protocols as "more trustworthy" than the others. Even if it is unlikely that a blockchain of cryptocurrencies employing PoW or PoS protocols is manipulated either by a miner that holds 51% of the entire network's computing power or by a miner that holds 51% of a cryptocurrency-specific network's stake, market participators may still overestimate the risk of market manipulation.

Psychologically, investors tend to overestimate events that have small probabilities, and individuals are, according to prospect theory, generally risk averse. As a consequence, investors' demand for cryptocurrencies incorporating the hybrid consensus protocol is relatively larger because these cryptocurrencies have a lower risk of manipulation. However, other explanations are possible as well, and future research is needed to elaborate on this issue.

Another important question that has been discussed recently and intensively in the literature is whether cryptocurrency markets are efficient. Until now, no consensus had been reached as to whether cryptocurrency markets are efficient. Our findings provide evidence that cryptocurrencies exhibit, on a portfolio level, strong patterns of higher-order autocorrelation. We also found strong effects of time series momentum. Both results imply that cryptocurrency markets are not efficient.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

ACKNOWLEDGEMENTS

We thank the participants of the 3rd Cryptocurrency Research Conference 2020, organized by the Centre for Digital Finance, the University of Southampton and the ICMA Centre, for their valuable comments. We also thank the participants of the 17th ACRN Oxford FRAP Conference 2019 at Hanken School of Economics for their valuable comments. We are grateful to Professor Joni Kokkonen for his insightful comments as a discussant at the Graduate School of Finance (GSF) and the Finnish Doctoral Programme in Economics (FDPE) Winter Workshop in Finance, Aalto University, 2019. We also thank the participants of the 2019 Research Seminar of the School of Accounting & Finance at the University of Vaasa for their encouraging comments, and we thank Helinä Saarela and Jussi Nikkinen specifically for their helpful comments.

REFERENCES

Al-Yahyaee, K. H., Mensi, W., and Yoon, S. M. (2018). Efficiency, multifractality, and the long-memory property of the Bitcoin market: a comparative analysis with stock, currency, and gold markets. *Finance Research Letters* **27**, 228–234 (https://doi.org/10.1016/j.frl.2018.03.017).

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and momentum everywhere. *Journal of Finance* **68**(3), 929–985 (https://doi.org/10.1111/jofi.12021).

Bach, L. M., Mihaljevic, B., and Zagar, M. (2018). Comparative analysis of blockchain consensus algorithms. In *41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1545–1550. IEEE (https://doi.org/10.23919/MIPRO.2018.8400278).

Bariviera, A. F. (2017). The inefficiency of Bitcoin revisited: a dynamic approach. *Economics Letters* **161**, 1–4 (https://doi.org/10.1016/j.econlet.2017.09.013).

Bendiksen, C., Gibbons, S., and Lim, E. (2018). The Bitcoin mining network-trends, marginal creation cost, electricity consumption and sources. *CoinShares Research* **21**, 3–19.

Biais, B., Bisiere, C., Bouvard, M., Casamatta, C., and Menkveld, A. J. (2018). Equilibrium bitcoin pricing. Working Paper, Social Science Research Network (https://doi.org/10.2139/ssrn.3261063).

de Vries, A. (2019). Renewable energy will not solve Bitcoin's sustainability problem. *Joule* **3**(4), 893–898 (https://doi.org/10.1016/j.joule.2019.02.007).

Dimitri, N. (2017). Bitcoin mining as a contest. *Ledger* **2**, 31–37 (https://doi.org/10.5195/ledger.2017.96).

Fama, E. F., and French, K. R. (2008). Dissecting anomalies. *Journal of Finance* **63**(4), 1653–1678 (https://doi.org/10.1111/j.1540-6261.2008.01371.x).

Fama, E. F., and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics* **116**(1), 1–22 (https://doi.org/10.1016/j.jfineco.2014.10.010).

Fama, E. F., and French, K. R. (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics* **123**(3), 441–463 (https://doi.org/10.1016/j.jfineco.2016.11.004).

Grobys, K., and Sapkota, N. (2019). Cryptocurrencies and momentum. *Economics Letters* **180**, 6–10 (https://doi.org/10.1016/j.econlet.2019.03.028).

Gutierrez, R. C., Jr., and Kelley, E. K. (2008). The long-lasting momentum in weekly returns. *Journal of Finance* **63**(1), 415–447 (https://doi.org/10.1111/j.1540-6261.2008.01320.x).

Hayes, A. S. (2017). Cryptocurrency value formation: an empirical study leading to a cost of production model for valuing bitcoin. *Telematics and Informatics* **34**(7), 1308–1321 (https://doi.org/10.1016/j.tele.2016.05.005).

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance* **45**(3), 881–898 (https://doi.org/10.1111/j.1540-6261.1990.tb05110.x).

Jegadeesh, N., and Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance* **48**(1), 65–91 (https://doi.org/10.1111/j.1540-6261.1993.tb04702.x).

Khuntia, S., and Pattanayak, J. K. (2018). Adaptive market hypothesis and evolving predictability of Bitcoin. *Economics Letters* **167**, 26–28 (https://doi.org/10.1016/j.econlet.2018.03.005).

King, S., and Nadal, S. (2012). PPcoin: peer-to-peer cryptocurrency with proof-of-stake. Working Paper, August 19, Decred. URL: https://decred.org/research/king2012.pdf.

Li, J., Li, N., Peng, J., Cui, H., and Wu, Z. (2019). Energy consumption of cryptocurrency mining: a study of electricity consumption in mining cryptocurrencies. *Energy* **168**, 160–168 (https://doi.org/10.1016/j.energy.2018.11.046).

Li, X., and Wang, C. A. (2017). The technology and economic determinants of cryptocurrency exchange rates: the case of Bitcoin. *Decision Support Systems* **95**, 49–60 (https://doi.org/10.1016/j.dss.2016.12.001).

Lustig, H., Roussanov, N., and Verdelhan, A. (2011). Common risk factors in currency markets. *Review of Financial Studies* **24**(11), 3731–3777 (https://doi.org/10.1093/rfs/hhr068).

Lütkepohl, H., Krätzig, M., and Phillips, P. C. (eds) (2004). *Applied Time Series Econometrics*. Cambridge University Press (https://doi.org/10.1017/CBO9780511606885).

Menkhoff, L., Sarno, L., Schmeling, M., and Schrimpf, A. (2012a). Currency momentum strategies. *Journal of Financial Economics* **106**(3), 660–684 (https://doi.org/10.1016/j.jfineco.2012.06.009).

Menkhoff, L., Sarno, L., Schmeling, M., and Schrimpf, A. (2012b). Carry trades and global foreign exchange volatility. *Journal of Finance* **67**(2), 681–718 (https://doi.org/10.1111/j.1540-6261.2012.01728.x).

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics* **104**(2), 228–250 (https://doi.org/10.1016/j.jfineco.2011.11.003).

Nadarajah, S., and Chu, J. (2017). On the inefficiency of Bitcoin. *Economics Letters* **150**, 6–9 (https://doi.org/10.1016/j.econlet.2016.10.033).

Nakamoto, S. (2008). Bitcoin: a peer-to-peer electronic cash system. Working Paper. URL: http://bitcoin.org/bitcoin.pdf.

Newey, W. K., and West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Technical Working Paper 55, National Bureau of Economic Research (https://doi.org/10.3386/t0055).

Nguyen, G. T., and Kim, K. (2018). A survey about consensus algorithms used in blockchain. *Journal of Information Processing Systems* **14**(1), 101–128.

Sensoy, A. (2019). The inefficiency of Bitcoin revisited: a high-frequency analysis with alternative currencies. *Finance Research Letters* **28**, 68–73 (https://doi.org/10.1016/j.frl.2018.04.002).

Symitsi, E., and Chalvatzis, K. J. (2018). Return, volatility and shock spillovers of Bitcoin with energy and technology companies. *Economics Letters* **170**, 127–130 (https://doi.org/10.1016/j.econlet.2018.06.012).

Tiwari, A. K., Jana, R. K., Das, D., and Roubaud, D. (2018). Informational efficiency of Bitcoin: an extension. *Economics Letters* **163**, 106–109 (https://doi.org/10.1016/j.econlet.2017.12.006).

Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters* **148**, 80–82 (https://doi.org/10.1016/j.econlet.2016.09.019).

Vidal-Tomás, D., and Ibañez, A. (2018). Semi-strong efficiency of Bitcoin. *Finance Research Letters* **27**, 259–265 (https://doi.org/10.1016/j.frl.2018.03.013).

Wei, W. C. (2018). Liquidity and market efficiency in cryptocurrencies. *Economics Letters*, **168**, 21–24 (https://doi.org/10.1016/j.econlet.2018.04.003).

Zhang, W., Wang, P., Li, X., and Shen, D. (2018). The inefficiency of cryptocurrency and its cross-correlation with Dow Jones Industrial Average. *Physica A: Statistical Mechanics and its Applications* **510**, 658–670 (https://doi.org/10.1016/j.physa.2018.07.032).

J. Int. Financ. Markets Inst. Money 74 (2021) 101402

Contents lists available at ScienceDirect

Journal of International Financial Markets, Institutions & Money

journal homepage: www.elsevier.com/locate/intfin

Asset market equilibria in cryptocurrency markets: Evidence from a study of privacy and non-privacy coins

Niranjan Sapkota [*,1], Klaus Grobys [1]

University of Vaasa, Vaasa, Finland

A R T I C L E I N F O

JEL Classification:
G12
G14

Keywords:
Cryptocurrency
Bitcoin
Privacy coins
Non-privacy coins
Market equilibria

A B S T R A C T

This paper explores whether asset market equilibria in cryptocurrency markets do exist. In doing so, it distinguishes between privacy and non-privacy coins. Most recently, privacy coins have attracted increasing attention in the public debate as non-privacy cryptocurrencies, such as Bitcoin, do not satisfy some users' demands for anonymity. Analyzing ten cryptocurrencies with the highest market capitalization in each submarket in the 2016–2018 periods, we find that privacy coins exhibit a distinct market equilibrium. Contributing to the current debate on the market efficiency of cryptocurrency markets, our findings provide evidence of market inefficiency. Moreover, the asset market equilibrium of privacy coins appears to originate from non-privacy coins with highest market capitalizations. We argue that the reason for this finding could be that non-privacy coins may be the first choice for criminals who might prefer cryptocurrencies exhibiting both a high level of anonymity and liquidity.

1. Introduction

Recently, empirical investigations of cryptocurrency markets have attracted considerable attention in the academic literature. This is not surprising given that from an economic perspective the sums of money involved are substantial (Fry and Cheah, 2016, p.350). Bitcoin (BTC), Ethereum (ETH), and Ripple (XRP) are among the top three cryptocurrencies with the highest market capitalization[2]. Interestingly, in recent years, another type of cryptocurrency, exemplified by Monero (XMR), Dash (DASH), and Verge (XVG) came into existence claiming to provide transaction privacy and account balance privacy to their users that the public blockchains like BTC, ETH or XRP do not provide. Androulaki, Karame, Roeschlin, Scherer, and Capkun (2013) argue that almost 40% of Bitcoin users could be identified despite the privacy measures in place; today, however, virtually all Bitcoin users could be identified. Goldfeder, Kalodner,

* Corresponding author at: Department of Accounting and Finance, University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland.
E-mail addresses: niranjan.sapkota@uva.fi (N. Sapkota), klaus.grobys@uva.fi (K. Grobys).

[1] We would like to thank the editor Jonathan Batten and an anonymous reviewer for giving us very helpful and very insightful comments and advice. We thank the participants of the 3rd Cryptocurrency Research Conference 2020, organized by the Centre for Digital Finance, the University of Southampton and the ICMA Centre, for their valuable comments. We are also thankful to the participants of the FINANCE PROPERTY, TECHNOLOGY AND THE ECONOMY CONFERENCE, 2019, University of South Australia, Adelaide, Australia. We also received valuable comments from the participants of the Graduate School Seminar 2018-2019 and Blockchain Seminar 2019, University of Vaasa, Finland.

[2] https://coinmarketcap.com/, accessed on 18.3.2019

https://doi.org/10.1016/j.intfin.2021.101402
Received 24 May 2019; Accepted 23 July 2021
Available online 27 July 2021

Reisman, and Narayanan (2018) show how third-party web trackers can de-anonymize users of cryptocurrencies.[3] By implementing sophisticated cryptographic protocols, privacy coins can hide senders and receivers addresses as well as the transaction amount while executing each transaction. The complete financial transparency of Bitcoin and other non-privacy coins is deterring many institutional and private investors from adopting these decentralized cryptocurrencies (Liu, Li, Karame and Asokan, 2018). The website Bitinfocharts provides user level, account balance level and transaction level information for the top 100 richest wallet addresses of fourteen different cryptocurrencies[4]. Non-privacy coins like Bitcoin are fully transparent; therefore, institutions are hesitant to use it as a medium of exchange. Baur, Hong, and Lee (2018) show that Bitcoins are mainly used as a speculative asset rather than as a medium of exchange.

An alternative is presented by Brenig, Accorsi, and Müller (2015) who outline the structure of a money laundering process and anti-money-laundering controls to examine whether cryptocurrencies constitute a driver for money laundering or not. The study argues that privacy coins have potential benefits for criminals; an important yet neglected factor in the circulation of cryptocurrencies as a money laundering instrument. In this regard, Kethineni and Cao (2020) support Brenig et al. (2015) and argue that cryptocurrencies became the currency of choice for many drug dealers and extortionists because of the opportunities to hide behind the assumed privacy and anonymity.

Privacy coins also use the decentralized public blockchain, but by using many features like masternode technology, ring signature, and a stealth wallet address, privacy coins make it impossible for third parties to trace transactions to the real parties involved. Privacy coins are different from non-privacy coins not only in the cryptographic level, but probably also at the user level. One reason why a separate user base for privacy coins might be growing could be due to the complete transparency of non-privacy coins like BTC and ETH. A certain degree of financial privacy about transaction and balance may be important for both the institutional and individual users. Therefore, one can hypothesize that the traders who favor privacy over complete transparency are emerging as a different subgroup in digital financial markets. Since the total supply of cryptocurrencies is (in most cases) predetermined, the price processes of cryptocurrencies depend solely on the demand side (e.g., the users). As a consequence, if the user base for privacy coins is different from that for non-privacy coins, we would expect that each submarket of cryptocurrencies potentially forms an individual asset market equilibrium.

Hence, we hypothesize that privacy coins form a distinct submarket in the cryptocurrency market. Following Urquhart (2016), Dyhrberg (2016), Bouri, Molnár, Azzi, Roubaud, and Hagfors (2017), who adopt the perspective that cryptocurrencies are an asset market, we consider the privacy and non-privacy coin markets to be two different asset markets. We explore whether market equilibria in cryptocurrency markets exist which, in turn, would imply the existence of submarkets. In this regard, we explicitly test whether privacy coins form such a distinct submarket in the cryptocurrency market. In doing so, we also investigate the efficiency of the overall cryptocurrency market. A test for market efficiency that does not require the specific formulation of an equilibrium price mechanism goes back to an argument by Granger (1986): If two or more asset prices show a stable common relationship in the long-run, it is possible that the movement of one asset price is linked to the movement of other asset prices. In turn, the price of one asset does not only depend on its own past prices but also on the history of a different asset's prices. As a consequence, the weak-form of market efficiency is violated (Richards, 1995, p.632). In line with Engle and Granger's (1987) cointegration theory, we employ Johansen's (1991, 1992a, 1992b, 1994, 1995) multivariate methodology to model and test the long-term equilibrium process. Specifically, if the stable long-run relationship between asset prices is known to the market participants they are able to exploit them to make excess profits (Copeland, 1991, p. 187).

This paper contributes to the literature in several important ways. On the one hand, Urquhart (2016) and Al-Yahyaee, Mensi, and Yoon (2018) studied the market efficiency of Bitcoin and found the cryptocurrency to be inefficient; however, Nadarajah and Chu (2017) revisited Urquhart's (2016) paper and found that Bitcoin returns do satisfy the efficient market hypothesis. Moreover, Vidal-Tomás and Ibañez (2018) and Sensoy (2019) argue that Bitcoin has become more efficient over time, whereas Bariviera's (2017) findings suggest Bitcoin has met the standards of informational efficiency since 2014. The different views in the literature indicate there is no consensus on the market efficiency of cryptocurrencies. While the past papers cited above consider a single asset (e.g., Bitcoin) our paper adds to this strand of literature by taking a market-wide perspective. In doing so, we consider a whole set of cryptocurrencies that exhibit the largest market capitalization and employ Johansen's (1991, 1992a, 1992b, 1994, 1995) multivariate cointegration methodology to explore whether or not asset market equilibria in line with Engle and Granger's (1987) cointegration theory exist.

There is also a new strand of literature emerging that discusses the features of privacy and non-privacy coins. This literature however mostly adopts a technological perspective and explores the privacy implications of Bitcoin (Androulaki et al., 2013), identification of a particular user's blockchain transactions (Goldfeder et al., 2018, Khalilov and Levi 2018), technological interventions that could address the privacy issues of cryptocurrencies (Ouaddah, Elkalam, and Ouahman, 2017; Kopp, Mödinger, Hauck, Kargl, and Bösch, 2017), and potential failures to guarantee privacy in terms of unlinkability and untraceability (Kumar, Fischer, Tople, and Saxena, 2017; Möser et al., 2017). Inspired by Fry and Cheah (2016) and Osterrieder and Lorenz (2017) who discuss the need for academic research on cryptocurrency from a financial point of view, our paper adopts the financial perspective and considers the cryptocurrency market as an asset market comprising two submarkets, the privacy coin market and the non-privacy coin market. This is the first paper that explicitly explores the existence of cointegration relationships among asset prices in the cryptocurrency market.

Finally, there is a wide strand of literature performing market efficiency tests using cointegration analysis of traditional currencies.

[3] Third parties typically receive user information for advertising purposes even if users pay via cryptocurrencies, which is enough to identify that particular user's blockchain transactions. By linking the user's cookie to that particular blockchain transaction, one can identify the real user behind it.

[4] https://bitinfocharts.com, accessed on 18.3.2019

Studies that also employ multivariate cointegration analysis as a methodological framework for exploring the European Monetary System (EMS) are Norrbin (1996), Woo (1999), Haug, MacKinnon, and Michels (2000), Rangvid and Sørensen (2002), and Aroskar, Sarkar, and Swanson (2004). Interestingly, those studies are mostly able to reject the null hypothesis of no cointegration for the EMS currencies. No paper has explicitly examined cryptocurrency markets for cointegration.[5] Hence, our paper extends the literature to reveal potential cointegration equilibria in new digital currency markets.

Our results show strong evidence for the existence of four cointegration relationships in the market for cryptocurrencies. Our evidence suggests that the privacy coin market forms a distinct asset market equilibrium. Employing ten large cap cryptocurrencies with privacy function and ten non-privacy cryptocurrencies in a joint model, we use a fully-specified Vector-Error-Correction Model (VECM) to estimate the four distinct cointegration equilibria. To do so, we employ from both groups of cryptocurrencies (e.g., privacy and non-privacy coins) those cryptocurrencies that exhibit the respective highest and lowest market capitalizations as left-hand side variables. Using Dash (DASH) as the privacy coin with highest market capitalization, we find that only two non-privacy coins, Peercoin (PPC) and MaidSafeCoin (MAID), exhibit *t*-statistics indicating statistical significance on a common 5% level.

On the other hand, six out of eight privacy coins exhibit *t*-statistics indicating statistical significance on at least a 5% level, which strongly suggests that this distinct market equilibrium is primarily driven by privacy coins. A joint test for exploring whether or not non-privacy coins are a part of that cointegration relationship shows that we cannot reject the null hypothesis, implying that DASH is a part of a submarket equilibrium consisting of privacy coins that is statistically unrelated to non-privacy coins. Our results are in line with the literature offering evidence of cointegration equilibria in traditional currency markets such as the European Monetary System (Norrbin, 1996; Woo, 1999; Haug et al., 2000; Rangvid and Sørensen, 2002; Aroskar, Sarkar, and Swanson, 2004). The presence of cointegration equilibria is also in line with Urquhart (2016) and Al-Yahyaee, Mensi, and Yoon (2018), who argue that Bitcoin is inefficient because a cointegration relationship implies weak-form market inefficiency. Finally, our results provide some new evidence on market heterogeneity: Privacy coins with high market capitalizations appear to build a submarket of cryptocurrencies as they form their own market equilibrium that is unrelated to the remaining cointegration equilibria in the cryptocurrency market. This market heterogeneity phenomenon in the cryptocurrency market may be subject to future investigations.

2. Background

Alexander and Dimitriu (2005) investigate the performance of a cointegration-based index tracking strategy and, by this, test the Efficient Market Hypothesis (EMH), as according to Jensen's (1978) definition of efficient markets, a trading strategy producing significant risk-adjusted payoffs is evidence against the EMH. In this regard, Alexander and Dimitriu (2005, p.215) argue that "the rationale for constructing portfolios based on a cointegration relationship with the benchmark, rather than correlation, rests on the following features of cointegration: the price difference between the benchmark and the replica portfolio is, by construction, stationary; the stock weights, being based on a large amount of history, have an enhanced stability; finally, there is a full use of the information contained in level variables such as stock prices". As pointed out in Alexander and Dimitriu (2005), assets in price levels contain a higher level of information as opposed to returns. Note that a cointegration relationship is based on an 'historical equilibrium', that is, a price equilibrium estimated from a long historical sample. In this cointegration equilibrium, exogenous cryptocurrencies and endogenous cryptocurrencies share a common stochastic trend that is stationary and mean-reverting. Whereas exogenous cryptocurrencies follow their own stochastic evolutions, endogenous cryptocurrencies adjust to deviations from the long-term equilibrium condition as defined by the cointegration vector. A cointegration relationship is evidence against the EMH because traders could gain profits by exploiting the mean-reverting characteristic defined by a cointegration relationship. In this regard, we follow Alexander and Dimitriu (2005) and test the EMH in the market for cryptocurrencies. In doing so, our paper contributes to the literature on testing the market efficiency of cryptocurrency markets using cointegration theory (Engle and Granger's, 1987).

Furthermore, Borri's (2019) study indicates that cryptocurrency returns are highly correlated one with the other. However, she finds that using cryptocurrency portfolios may substantially reduce idiosyncratic risk and may offer better risk-adjusted and conditional returns than individual cryptocurrencies. In this regard, it is interesting to note that Borri (2019) does not study cryptocurrency portfolios that form submarkets within the overall cryptocurrency ecosystem. Using popular moving average strategies, a recent study from Ahmed, Grobys, Sapkota (2020) explores the profitability of technical trading rules implemented among cryptocurrencies with privacy function. Averaging the average returns across the entire set of cryptocurrencies with privacy function, they did not find any positive average portfolio returns in excess of the equally-weighted average buy-and-hold portfolio. This is a surprising finding because earlier literature suggests that technical trading rules are profitable for cryptocurrency markets (Grobys et al., 2020; Gerritsen et al., 2020; Corbet et al., 2019; Miller et al., 2019). Ahmed et al.'s (2020) study does not include any fully elaborated dynamic general equilibrium asset-pricing models to assess whether the reported payoffs are merely the equilibrium rents that accrue to investors willing to carry the risks associated with such strategies (Lo, Mamaysky, and Wang, 2000) and, hence, encourages studies to discern the economic sources of return differentials among cryptocurrency submarkets. Our study contributes to this strand of literature by taking an important next step by exploring whether statistical equilibriums in cryptocurrency submarkets do exist.

Moreover, we note that Ahmed et al.'s (2020) study provides some evidence for that the market forces driving the market for cryptocurrencies with privacy function are somewhat different from those that drive the overall cryptocurrency market. It is interesting to note that recent research suggests that about half of Bitcoin transactions is associated with criminal activities such as drugs,

[5] Other relevant papers that document mixed results are Jeon and Seo (2003) and Phengpis (2006), who examine structural instability in particular and investigate currency crises for market efficiency.

Table 1
Top 10 Privacy and Non-privacy Coins.

Panel A: Top 10 Non-Privacy Coins

S.No	Non-Privacy Coin	Symbol	January 3, 2016 Coin Rank /572 Coins	Capitalization ($)	December 30, 2018 Coin Rank/2073 Coins	Capitalization ($)	3 Years' Market Capitalization Growth%
	Bitcoin	BTC	1	6,467,437,080	1	67,475,512,827	943.31
1	Ripple	XRP	2	201,799,631	2	15,076,740,856	7371.14
2	Litecoin	LTC	3	152,873,521	7	1,912,263,648	1150.90
3	Ethereum	ETH	4	73,843,278	3	14,560,066,114	19617.52
4	Dogecoin	DOGE	6	14,940,681	23	278,470,304	1763.84
5	Peercoin	PPC	7	9,756,959	181	14,254,314	46.10
6	BitShares	BTS	8	8,591,688	44	106,520,276	1139.81
7	Stellar	XLM	9	8,436,465	6	2,250,048,215	26570.51
8	Nxt	NXT	10	6,863,998	113	29,426,627	328.71
9	MaidSafeCoin	MAID	11	6,789,470	65	61,335,019	803.38
10	NameCoin	NMC	12	6,073,338	220	10,480,336	72.56
						Average (Excluding Bitcoin)	**5886.45**

Panel B: Top 10 Privacy Coins

S.No	Privacy Coin Name						
1	Dash	DASH	5	19,794,713	15	698,091,183	3426.65
2	Bytecoin	BCN	14	5,582,979	38	134,517,567	2309.42
3	Monero	XMR	15	5,295,952	13	806,939,516	15136.91
4	DigitalNote	XDN	51	447,057	252	8,565,009	1815.87
5	CloakCoin	CLOAK	125	201,995	318	6,249,733	2994.00
6	Aeon	AEON	137	137,088	393	4,475,790	3164.90
7	NavCoin	NAV	142	121,805	213	11,127,542	9035.54
8	Verge	XVG	149	109,968	43	110,307,808	100209.00
9	Stealth	XST	161	8352	515	2,539,074	30300.79
10	Prime-XI	PXI	322	8889	1701	4236	−52.35
						Average	**16834.07**

Note: This table reports the top 11 non-privacy coins (including Bitcoin) and top ten privacy coins based on their market capitalization as of January 3, 2016. There were 572 cryptocurrencies available (including both privacy and non-privacy coins) as of January 3, 2016, and 2073 coins as of December 30, 2018. Coin Rank shows the position based on a coin's market capitalization. Three Years' Market Capitalization Growth shows the percentage growth in market capitalization from January 3, 2016 until December 30, 2018. Panel A shows the top 11 non-privacy coins and Panel B shows the top ten privacy coins in terms of market capitalization (Source: coinmarketcap.com/historical/).

Table 2
ADF tests for privacy and non-privacy coins.

Privacy coins					Non-privacy coins				
	Model 1		Model 2			Model 1		Model 2	
Coin	Intercept[a]	Lags[d]	Intercept and trend[b]	Lags[d]	Coin	Intercept[a]	Lags[d]	Intercept and trend[b]	Lags[d]
DASH	−1.79	0	0.25	0	XRP	−0.99	2	−1.19	2
BCN	−1.34	1	−0.90	1	LTC	−1.13	0	−0.23	0
XDN	−1.39	0	−0.91	0	ETH	−2.93**	0	−0.63	0
XMR	−2.16	0	−0.15	0	DOGE	−1.30	0	−1.41	0
CLOAK	−1.51	2	−0.72	2	PPC	−1.43	0	−1.18	0
AEON	−1.46	1	−0.57	1	BTS	−1.24	0	−0.41	0
XST	−1.49	1	−1.45	1	XLM	−0.93	1	−1.45	1
PXI	−1.44	2	−1.16	2	NXT	−1.37	0	−0.62	0
NAV	−1.86	4	−0.60	4	MAID	−2.86**	0	−1.69	1
XVG	−1.23	4	−1.27	4	NMC	−1.65	1	−1.59	1
					BTC	−1.31	0	−0.06	0

Note: This table reports the results for Augmented Dickey Fuller tests of the daily price series in logs for privacy and non-privacy coins. Model 1 accounts for an intercept in the test regression, whereas model 2 accounts for both an intercept and trend term. The sample period is from January 1, 2016 until December 31, 2018 corresponding to 1096 observations.
**Statistically significant on a 5% level.
[a]Critical values for 10%, 5% and 1% significance levels are −2.57, −2.86 and −3.44.
[b]Critical values for 10%, 5% and 1% significance levels are −3.13, −3.41 and −3.97.
[d]Lag-order is chosen by using the Schwarz info criterion. The maximum lag length is chosen by default is 21.

for instance (Foley, Karlsen, Putniņš, 2019). Because Bitcoin is a non-privacy coin, Ahmed et al. (2020) argue that the only option how market participants might achieve (full) anonymity is via the dark web. On the other hand, the usage of the dark web is per se a criminal offence; as a consequence, traders might prefer choosing privacy coins for their transactions instead of non-privacy coins. "This enables users making transactions in cryptocurrency (e.g., privacy coins) in the legal world-wide-web domain while still meeting their demands for legal transfers of digital currency, security, and confidentiality through anonymous transactions. Moreover, such security features may be of considerable importance for traders from countries where economic and political freedom is limited." (Ahmed et al., 2020). Since the price of cryptocurrencies is in most cases (due to the limited supply) driven by the demand side, we hypothesize in line with Ahmed et al. (2020) that market participants engaging in criminal activities shift their demand from non-privacy to privacy cryptocurrencies. If those market participants constituted a distinct user base it would imply that a potential market equilibrium in the submarket for cryptocurrencies exhibiting the privacy function should consequently be detached from the market equilibrium formed by non-privacy cryptocurrencies. Motivated by this literature, we take a novel perspective using cointegration theory to explore whether distinct market equilibria in those two different cryptocurrency submarkets do exist.

3. Methodology

For each group—privacy and non-privacy coins—we retrieved daily closing prices[6] for ten cryptocurrencies that exhibit the highest market capitalization as of January 3, 2016. We also downloaded data for Bitcoin, which dominates the cryptocurrency market. Our sample is from January 1, 2016 until December 31, 2018 accounting for 1096 daily observations. For all asset prices we compound the log-price series that we used in the following analyses. The trend in data levels of non-privacy and privacy coins are reported in Fig. A1 and Fig. A2 in the appendix.

Table 1 illustrates that at the beginning of our sample period, the average market capitalization of non-privacy coins (excluding Bitcoin) is about 19 times larger than that of privacy coins. Interestingly, the three-year growth in market capitalization of these top ten cryptocurrencies in each category is about three times higher for privacy coins than for non-privacy coins. This suggests that the relative popularity of privacy coins has increased over time. Even though both privacy and non-privacy categories of cryptocurrencies use the decentralized public blockchain technology, they differ on other technological levels, particularly in relation to either the public node, wallet form, or the signature. Privacy coins also differ from non-privacy coins in terms of their usage. People who use non-privacy coins face complete financial transparency which might be less appealing in a competitive environment. Traders might want to maintain a certain degree of financial privacy at both the transaction and account balance level. Thus, we hypothesize that, as independent players emerge in digital currency markets, privacy coins form a distinct submarket.

To investigate whether or not distinct market equilibria, forming submarkets in the overall cryptocurrency market, exist and to explore how they relate to each other, the first step of our empirical analysis is to identify whether or not our set of cryptocurrencies exhibits stationarity. This is an important step because cointegration equilibria require that the input variables are integrated of order one, that is, $I(1)$ stochastic processes. To test the order of integration, we follow the common literature and employ the well-known Augmented Dickey Fuller (ADF) unit root test (Dickey-Fuller, 1979). The model for estimating the test statistics is given by:

[6] Because cryptocurrencies are traded 24/7, coinmarketcap.com takes the latest data in the range (UTC time) to determine the closing price.

Table 3
Trace test excluding Bitcoin.

Rank	Eigenvalue	Trace test	p-value	Lmax test	p-value
0	0.1390	1094.7000	[0.0000]	163.7500	[0.0000]
1	0.1227	931.0000	[0.0000]	143.2300	[0.0001]
2	0.0926	787.7700	[0.0043]	106.3600	[0.2285]
3	0.0830	681.4100	[0.0402]	94.7980	[0.4534]
4	0.0752	586.6200	[0.1597]	85.4960	[0.6187]
5	0.0676	501.1200	[0.3743]	76.5640	[0.7612]
6	0.0526	424.5500	[0.6062]	59.0730	[0.9899]
7	0.0519	365.4800	[0.6468]	58.2950	[0.9560]
8	0.0491	307.1900	[0.7431]	55.1160	[0.9217]
9	0.0407	252.0700	[0.8473]	45.4780	[0.9817]
10	0.0386	206.5900	[0.8753]	43.0240	[0.9529]
11	0.0303	163.5700	[0.9204]	33.6620	[0.9921]
12	0.0265	129.9100	[0.9093]	29.4100	[0.9895]
13	0.0243	100.5000	[0.8892]	26.9620	[0.9671]
14	0.0233	73.5350	[0.8887]	25.8330	[0.8615]
15	0.0140	47.7020	[0.9404]	15.4110	[0.9891]
16	0.0130	32.2900	[0.8650]	14.2800	[0.9204]
17	0.0104	18.0110	[0.8192]	11.4910	[0.8060]
18	0.0052	6.5194	[0.8243]	5.7163	[0.8390]
19	0.0007	0.8031	[0.3702]	0.8031	[0.3702]

Note: This table reports the results for the trace test for cointegration applied to a set of twenty cryptocurrencies consisting of ten privacy coins and ten non-privacy coins (excluding Bitcoin) exhibiting the highest market capitalization as of Jan 3, 2016. Our model uses daily data of log prices. The test statistic allows for linear deterministic trend in data. The model is estimated with the econometric software Gretl. The sample period is from January 1, 2016 until December 31, 2018 corresponding to 1096 observations.
Johansen test:
Number of equations = 20
Lag order = 2
Estimation period: 2016-01-03-2018-12-31 (T = 1094)
Case 5: Unrestricted trend and constant
Log-likelihood = 45917.6 (including constant term: 42813)

$$\Delta p_{i,t} = \delta_{0,i} + \delta_{1,i}t + \delta_{2,i}p_{i,t-1} + \sum_{j=1}^{T} \delta_{2+j,i}\Delta p_{i,t-j} + \varepsilon_{i,t} \tag{1}$$

where $p_{i,t}$ denotes the log-price of cryptocurrency i at time t, δ_0 is the corresponding estimate for the intercept of the regression, δ_1 is the corresponding estimate for the time trend t, δ_{2+j} is the estimate for the lagged differences in log prices of lag $2+j$ where $j = 1, \cdots, T$, ε_t denotes a white-noise error term, and the parameter δ_2 is assumed to be zero under the null hypothesis. Under the null hypothesis, $p_{i,t}$ is an $I(1)$ stochastic process. Given our research context, we assume stationarity under the alternative hypothesis. Moreover, we follow a common practice by choosing the lag-order j with respect to the Schwarz criterion.

Specifically, under some circumstances, a linear combination of some $I(1)$ processes becomes a stationary process, that is, an $I(0)$ process. According to Granger (1986), and Engle and Granger (1987), those $k = 1, \cdots, K$ stochastic processes are said to be co-integrated of order (1, …, 1), denoted as CI(1, …, 1). To test the order of cointegration, we employ the trace test using the overall set of $K = 20$ cryptocurrencies, consisting of ten privacy coins and ten non-privacy coins, given by:

$$LR(r_0) = -T \sum_{j=r_0+1}^{K} log\left(1 - \lambda_j\right) \tag{2}$$

where λ_j are the eigenvalues obtained by applying Reduced Rank (RR) regression techniques to the fully unrestricted Vector-Error Correction model (VECM) (Johansen 1991, 1992a, 1992b, 1994, 1995) given by:

$$\Delta \boldsymbol{y}_t - \mu_1 = \Pi(\boldsymbol{y}_{t-1} - \mu_0 - \mu_1(t-1)) + \sum_{j=1}^{p-1} \Gamma_j(\Delta \boldsymbol{y}_{t-1} - \mu_1) + \boldsymbol{u}_t \tag{3}$$

where the 20x1 vector $\Delta \boldsymbol{y}_t$ contains the log-returns of cryptocurrencies $i = \{1, 2, \cdots, 19, 20\}$.[7] Moreover, μ_0 and μ_1 denote constant and trend, Π and $\Gamma_1, \Gamma_2, \ldots, \Gamma_{p-1}$ are KxK parameter matrices. The trace test tests the sequence of hypotheses given by:

[7] The model is more detailed in Lütkepohl, Krätzig, and Phillips (2004, p.114).

Table 4
Trace test including Bitcoin.

Rank	Eigenvalue	Trace test	p-value	Lmax test	p-value
0	0.1408	1196.4000	[0.0000]	166.0300	[0.0000]
1	0.1268	1030.4000	[0.0000]	148.3400	[0.0002]
2	0.1057	882.0600	[0.0015]	122.1900	[0.0407]
3	0.0873	759.8800	[0.0344]	99.9500	[0.4744]
4	0.0821	659.9300	[0.1408]	93.7200	[0.5010]
5	0.0693	566.2100	[0.3807]	78.5400	[0.8712]
6	0.0632	487.6700	[0.5692]	71.4400	[0.9119]
7	0.0520	416.2300	[0.7268]	58.4500	[0.9921]
8	0.0498	357.7800	[0.7611]	55.9000	[0.9797]
9	0.0422	301.8800	[0.8170]	47.1600	[0.9955]
10	0.0393	254.7300	[0.8125]	43.8500	[0.9906]
11	0.0384	210.8800	[0.8174]	42.8300	[0.9558]
12	0.0332	168.0500	[0.8682]	36.9300	[0.9671]
13	0.0256	131.1200	[0.8939]	28.3400	[0.9940]
14	0.0245	102.7900	[0.8492]	27.0900	[0.9653]
15	0.0239	75.7000	[0.8440]	26.5700	[0.8288]
16	0.0170	49.1400	[0.9169]	18.7900	[0.9322]
17	0.0126	30.3400	[0.9183]	13.9100	[0.9331]
18	0.0092	16.4300	[0.8896]	10.0300	[0.8961]
19	0.0050	6.4000	[0.8346]	5.5400	[0.8541]
20	0.0008	0.8600	[0.3548]	0.8600	[0.3548]

Note: This table reports the results for the trace test for cointegration applied to a set of twenty-one cryptocurrencies consisting ten privacy coins and eleven non-privacy coins (including Bitcoin) exhibiting the highest market capitalization as of Jan 3, 2016. Our model uses daily data of log prices. The test statistic allows for linear deterministic trend in data. The model is estimated with the econometric software Gretl. The sample period is from January 1, 2016 until December 31, 2018 corresponding to 1096 observations.
Johansen test:
Number of equations = 21
Lag order = 2
Estimation period: 2016-01-03-2018-12-31 (T = 1094)
Case 5: Unrestricted trend and constant
Log-likelihood = 49307.7 (including constant term: 46203.1)

$$\begin{array}{lll} H_0(0): rank(\Pi) = 0 & \text{versus} & H_1(0): rank(\Pi) > 0, \\ H_0(1): rank(\Pi) = 1 & \text{versus} & H_1(0): rank(\Pi) > 1, \\ & \vdots & \\ H_0(K-1): rank(\Pi) = K-1 & \text{versus} & H_1(K-1): rank(\Pi) = K. \end{array}$$

The corresponding cointegration rank is selected when the null hypothesis cannot be rejected for the first time.[8] If a cointegration relationship of order r exists, the matrix Π has a reduced rank form and can be decomposed into $\boldsymbol{\alpha\beta}$', where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are Kxr matrices. In this regard, the term,

$$\boldsymbol{\beta}'(\boldsymbol{y}_{t-1} - \mu_0 - \mu_1(t-1))$$

contains the cointegration equilibrium relationships, whereas $\boldsymbol{\alpha}$ is the loading matrix. Each cryptocurrency in submarket r that has an insignificant weight attached to the respective cointegration relationship is said to be exogenous as it does not respond to disturbances of the long-term equilibrium. If cointegration holds, the linear combination conditioned by the matrix

$\boldsymbol{\beta}$ results in r stationary stochastic processes that are mean-reverting. While exogenous cryptocurrencies entering the respective cointegration equilibrium do not mean-revert, those cryptocurrencies that are endogenous adjust to deviations from the long-term equilibrium defined by $\boldsymbol{\beta}$ (via mean-reversion). Corresponding significant loadings in the matrix $\boldsymbol{\alpha}$ suggest endogeneity. Traders could exploit this pattern by betting on this mean-reversion property. As pointed out in Alexander and Dimitriu (2005), the presence of a cointegration relationship is evidence against the EMH because traders could gain profits by exploiting the mean-reverting characteristic defined by a cointegration relationship.

4. Results

The results of testing the order of integration are presented in Table 2. Initially imposing the restriction $\delta_{1,i} = 0$, we find that none of the privacy coins appears to be stationary, whereas the null hypothesis is rejected for only two non-privacy coins (e.g., Ethereum and MaidSafeCoin). However, when estimating the fully unrestricted ADF test—involving testing a random walk with drift under the null hypothesis against a trend-stationary process under the alternative—we find that all cryptocurrencies appear to be $I(1)$. This result is in line with the earlier literature indicating that Bitcoin is an $I(1)$ process (Urquhart, 2016).

[8] Using the general-to-specific rule, we made use of the fully unrestricted VECM as given in Eq. (3).

Table 5
Vector-Error Correction model estimates using all 20 cryptocurrencies excluding Bitcoin.

i	Coins	$\hat{\beta}_{1,i}$	$\hat{\beta}_{2,i}$	$\hat{\beta}_{3,i}$	$\hat{\beta}_{4,i}$	$\hat{\alpha}_{1,i}$	$\hat{\alpha}_{2,i}$	$\hat{\alpha}_{3,i}$	$\hat{\alpha}_{4,i}$
1	**XRP(−1)**	1.0000	0.0000	0.0000	0.0000	−0.0038	−0.0186***	0.0141*	0.0020
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(−0.4440)	(−2.5970)	(1.6730)	(0.3506)
2	**DASH(−1)**	0.0000	1.0000	0.0000	0.0000	0.0111	−0.0005	−0.0030	0.0156***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(1.6430)	(−0.0812)	(−0.4512)	(3.352)
3	**NMC(−1)**	0.0000	0.0000	1.0000	0.0000	0.0180*	0.0076	−0.0187*	0.0218***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(1.6900)	(0.8479)	(−1.769)	(2.975)
4	**PXI(−1)**	0.0000	0.0000	0.0000	1.0000	−0.0024	0.0217	0.0657***	–
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(−0.1158)	(1.2260)	(3.1520)	(−4.5020)
5	**LTC(−1)**	−0.22378	0.25449	–	−1.2630***	0.0094	–	–	0.0193***
		(−0.6291)	(0.5822)	(−1.6127)	(−3.8480)	(1.4610)	(−1.1890)	(−0.7491)	(4.3810)
6	**BCN(−1)**	0.6816***	–	–	−1.2040***	0.0498***		0.0660***	0.0147*
		(2.6323)	(−1.5670)	(−6.7180)	(−5.0368)	(3.9960)	(2.2860)	(5.3260)	(1.7130)
7	**ETH(−1)**	0.1403	–	0.1487	0.7983***	0.0176**	0.0120**	0.0165**	0.0005
		(0.5667)	(−1.6679)	(0.9668)	(3.4939)	(2.4650)	(2.0000)	(2.3310)	(0.0930)
8	**XMR(−1)**	2.7010***	–	–	−2.4440***	0.0253***		0.0022	0.0132**
		(7.7472)	(−5.7110)	(−4.8846)	(−7.5971)	(3.1810)	(1.9850)	(0.2730)	(2.4090)
9	**DOGE(−1)**	–	–	–	0.7019**	0.0365***	0.0127*	0.0236***	0.0064
		(−0.7941)	(−0.6409)	(−2.3270)	(2.4723)	(4.7090)	(1.9390)	(3.0690)	(1.2100)
10	**XDN(−1)**	–	1.9534***	–	−0.7147***	0.0660***		0.0019	0.0466***
		(−5.429)	(6.5663)	(−0.1281)	(−3.1993)	(5.7470)	(0.3348)	(0.1648)	(5.8950)
11	**PPC(−1)**	0.4894	–	–	−1.8360***	0.0420***		0.0127*	0.0183***
		(1.2068)	(−6.0611)	(−2.8568)	(−4.9070)	(5.4370)	(4.2350)	(1.6530)	(3.4450)
12	**CLOAK(−1)**	−0.12281	0.3844*	0.1176	0.3329**	−0.0163	−0.0138	−0.0140	–
		(−0.7380)	(1.8797)	(1.1370)	(2.1681)	(−1.0640)	(−1.0690)	(−0.9187)	(−0.6890)
13	**BTS(−1)**	–	0.46371	0.6770***	1.3303***	0.0327***	0.0180**	0.0274***	–
		(−2.1728)	(1.2042)	(3.4764)	(4.6004)	(3.6690)	(2.3950)	(3.0940)	(−1.6760)
14	**AEON(−1)**	–	0.7939***	0.1772	1.0691***	0.0487***	0.0315***	0.0123	–
		(−4.1714)	(2.7546)	(1.2160)	(4.9397)	(3.8170)	(2.9310)	(0.9696)	(−0.7280)
15	**XLM(−1)**	0.0170	0.4911*	0.5639***	0.5981**	−0.0031	−0.0268***	−0.0055	0.0145**
		(0.0783)	(1.8431)	(4.1814)	(2.9892)	(−0.3247)	(−3.3150)	(−0.3150)	(2.1960)
16	**NAV(−1)**	–	0.1297	1.1108***	1.2463***	0.0613***	0.0273**	–	0.0048
		(−5.0319)	(0.4953)	(8.3859)	(6.3364)	(4.2940)	(2.2660)	(−2.0240)	(0.4846)
17	**NXT(−1)**	0.4986**	–	–	−0.2720	0.0087	–	0.0006	0.0076
		(2.4449)	(−0.3914)	(−0.1526)	(−1.4451)	(0.9944)	(−0.5809)	(0.0663)	(1.2640)
18	**XVG(−1)**	–	–	0.0723	0.0919	0.1073***	0.0538***	0.0588***	–
		(−0.2168)	(−2.3863)	(0.7427)	(0.6356)	(6.4200)	(3.8140)	(3.5350)	(−0.7020)
19	**MAID(−1)**	−0.5737*	1.4236***	–	−0.2948	0.0030	–	–	0.0121**
		(−1.7481)	(3.5305)	(−0.8360)	(−0.9736)	(0.3909)	(−0.9620)	(−1.9150)	(2.3350)
20	**XST(−1)**	0.4933***	–	–	−0.2360	0.0183	0.0071	0.0278*	0.0007
		(2.6938)	(−2.2542)	(−3.1445)	(−1.3968)	(1.2540)	(0.5770)	(1.9190)	(0.0690)

Note: This table reports the estimates for a fully specified Vector-Error-Correction Model using all twenty cryptocurrencies in our sample excluding Bitcoin. Our model uses daily data of log prices. The model has a lag-order of $p = 2$. We report the estimates for the matrix $\boldsymbol{\beta}$ and the estimates for the adjustment parameter matrix $\boldsymbol{\alpha}$. The model allows for linear deterministic trend in data. The corresponding *t*-statistics are given in parentheses. The model is estimated with the econometric software Gretl. The sample period is from January 1, 2016 until December 31, 2018 corresponding to 1096 observations. Log-likelihood (lu) = 42519.662.
*Statistically significant on a 10% level.
** Statistically significant on a 5% level.
*** Statistically significant on a 1% level.

We employ the Akaike criterion (AIC), Schwarz Bayesian criterion (BIC) and the Hannan-Quinn criterion (HQC), to assess the optimal lag-length od the Vector-Autoregression (VAR) model and find that the optimal lag-length is one for the HQC and BIC of two for the AIC (see Table A.1 in the appendix).

Next, using the optimal lag-length of two lags and running the trace test, we find that on the 5% level, there are four significant cointegration equilibrium relationships. The results are reported in Table 3.[9]

Since Bitcoin dominates the non-privacy cryptocurrency market, it would be useful to understand how the results are affected when adding Bitcoin to our sample. Adding Bitcoin to our set of cryptocurrencies and re-running the trace test, we find that including Bitcoin does not alter our main result (see Table 4); still there is evidence of four cointegration equilibria in the sample. Hence, our conclusions remain unchanged. While earlier studies found cointegration relationships in traditional currency markets (Norrbin, 1996; Woo, 1999; Haug et al., 2000; Rangvid and Sørensen, 2002; Aroskar et al., 2004) our novel findings suggest the presence of cointegration equilibria even in new digital currency markets.

[9] We decided to use two lags for the VECM specifications because testing for the optimal lag length in the VAR model representation suggests according to the AIC a lag-order of 2. The corresponding results are reported in Table A.1 in the appendix. Furthermore, the trend in data levels of non-privacy and privacy coins are reported in Fig. A.1 and Fig. A.2 in the appendix, which suggests to estimate the fully unrestricted VECM.

Table 6
Restricted Vector-Error Correction model estimates using all 20 cryptocurrencies (excluding Bitcoin).

i	Coins	$\widehat{\beta}_{1,i}$	$\widehat{\beta}_{2,i}$	$\widehat{\beta}_{3,i}$	$\widehat{\beta}_{4,i}$	$\widehat{\alpha}_{1,i}$	$\widehat{\alpha}_{2,i}$	$\widehat{\alpha}_{3,i}$	$\widehat{\alpha}_{4,i}$
1	**XRP(−1)**	4.9650	23.8620	−4.4180	1.0840	0.0020*	0.0003	0.0037***	–
						(1.9370)	(1.2750)	(3.6500)	(−4.0660)
2	**DASH(−1)**	1.4050	20.3200	−5.4600	2.2000	0.0030***	–	0.0004	0.0007
						(3.6390)	(−1.0270)	(0.4767)	(0.9418)
3	**NMC(−1)**	1.0470	−37.6080	−6.7530	−12.7680	0.0026	–	−0.0011	0.0033
						(1.9660)	(−0.6535)	(−0.8434)	(2.8370)
4	**PXI(−1)**	3.9350	8.8340	2.9110	5.0080	–	0.0013**	−0.0027	–
						(−2.512)	(2.5550)	(−1.0790)	(−3.5770)
5	**LTC(−1)**	−6.1150	0.0000	−1.8420	−1.8950	0.0036***	–	0.0014*	0.0004
						(4.5920)	(−1.1500)	(1.8660)	(0.5406)
6	**BCN(−1)**	−3.3150	31.2420	3.1460	6.5820	0.0102***	0.00002	−0.0018	–
						(6.6300)	(0.0573)	(−1.1860)	(−2.7810)
7	**ETH(−1)**	3.3910	0.0000	3.8950	2.1490	0.0021**	0.0001	−0.0014	–
						(2.3520)	(0.3820)	(−1.6150)	(−1.2890)
8	**XMR(−1)**	−0.0490	66.3660	4.0290	4.8860	0.0032***	–	−0.0015	0.0009
						(3.2910)	(−0.0275)	(−1.6250)	(0.9700)
9	**DOGE(−1)**	0.4290	0.0000	6.5460	6.3010	0.0043***	0.0007***	−0.0004	–
						(4.4830)	(3.4680)	(−0.4156)	(−3.5040)
10	**XDN(−1)**	−7.3700	−33.2890	−9.5790	−7.0180	0.0107***	0.0008***	0.0023*	–
						(7.5970)	(2.7990)	(1.6880)	(−1.6080)
11	**PPC(−1)**	−8.9510	0.0000	16.8170	0.8960	0.0054***	–	−0.0035***	0.0016*
						(5.6730)	(−0.6052)	(−3.8050)	(1.8640)
12	**CLOAK(−1)**	1.0890	−9.6940	−2.3870	−1.5050	–	0.0004	0.0021	–
						(−1.7080)	(1.1300)	(1.1330)	(−0.2703)
13	**BTS(−1)**	3.6980	0.0000	1.3240	1.9320	0.0014	0.0009***	−0.0014	–
						(1.2850)	(3.9260)	(−1.3380)	(−4.2180)
14	**AEON(−1)**	0.2140	−24.6440	0.3550	0.1280	0.0007	0.0012***	−0.0034**	–
						(0.4381)	(3.7950)	(−2.2330)	(−1.4560)
15	**XLM(−1)**	3.8690	0.0000	−4.4170	−2.1640	0.0027**	0.0002	0.0049***	–
						(2.3080)	(0.8360)	(4.3380)	(−2.6010)
16	**NAV(−1)**	1.1000	−44.8930	0.8390	−7.1840	–	0.0018***	−0.0024	–
						(−0.3513)	(4.9360)	(−1.4310)	(−0.6023)
17	**NXT(−1)**	1.0430	0.0000	−2.9420	−2.2270	0.0018*	0.0002	0.0013	–
						(1.6780)	(1.0550)	(1.2040)	(−1.0840)
18	**XVG(−1)**	−0.2140	−3.2080	3.1300	0.1480	0.0063***	0.0026***	−0.0042**	–
						(3.0350)	(6.0900)	(−2.1000)	(−4.8370)
19	**MAID(−1)**	−2.5510	0.0000	−6.4290	−0.3290	0.0012	−0.0001	0.0010	0.0009
						(1.2400)	(−0.4403)	(1.1520)	(1.0200)
20	**XST(−1)**	0.5260	16.4420	2.5880	3.4220	0.0032*	0.0003	−0.0001	–
						(1.7580)	(0.7125)	(−0.0739)	(−1.5830)

Note: This table reports the estimates for a restricted Vector-Error-Correction Model using the set of all cryptocurrencies that are endogenous. Our model uses daily data of log prices. The model has a lag-order of $p = 2$. We report the estimates for the matrix $\boldsymbol{\beta}$ and the estimates for the adjustment parameter matrix $\boldsymbol{\alpha}$. The model allows for linear deterministic trend in data. The corresponding t-statistics for the adjustment parameter matrix $\boldsymbol{\alpha}$ are given in parentheses. The model is estimated with the econometric software Gretl. The sample period is from January 1, 2016 until December 31, 2018 corresponding to 1096 observations. Unrestricted loglikelihood (lu) = 42519.662, Restricted loglikelihood (lr) = 42518.064, 2 * (lu - lr) = 3.19601, P(Chi-square(5) > 3.19601) = 0.669797.
* Statistically significant on a 10% level.
** Statistically significant on a 5% level.
*** Statistically significant on a 1% level.

Establishing the existence of four cointegration relationships makes it possible to estimate the reduced form of model (3) using $\boldsymbol{\alpha\beta}$' where the dimension of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is 20x4. To test our hypothesis of interest, we order the vector $\boldsymbol{y}_{t-1}$ as,

$$\boldsymbol{y}_{t-1} = (XRP_{t-1}, DASH_{t-1}, NMC_{t-1}, PXI_{t-1}, LTC_{t-1}, BCN_{t-1}, ETH_{t-1}, XMR_{t-1}, DODGE_{t-1}, XDN_{t-1}, \cdots$$

$$PPC_{t-1}, CLOAK_{t-1}, BTS_{t-1}, AEON_{t-1}, XML_{t-1}, NAV_{t-1}, NXT_{t-1}, XVG_{t-1}, MAID_{t-1}, XST_{t-1})$$

The chosen ordering implies for the normalized matrix $\boldsymbol{\beta}$ that,

$$\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4] = \begin{bmatrix} \beta_{XRP,1} & 0 & 0 & 0 \\ 0 & \beta_{DASH,2} & 0 & 0 \\ 0 & 0 & \beta_{NMC,3} & 0 \\ 0 & 0 & 0 & \beta_{PXI,4} \\ \beta_{LTC,1} & \beta_{LTC,2} & \beta_{LTC,3} & \beta_{LTC,4} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{XST,1} & \beta_{XST,2} & \beta_{XST,3} & \beta_{XST,4} \end{bmatrix}$$

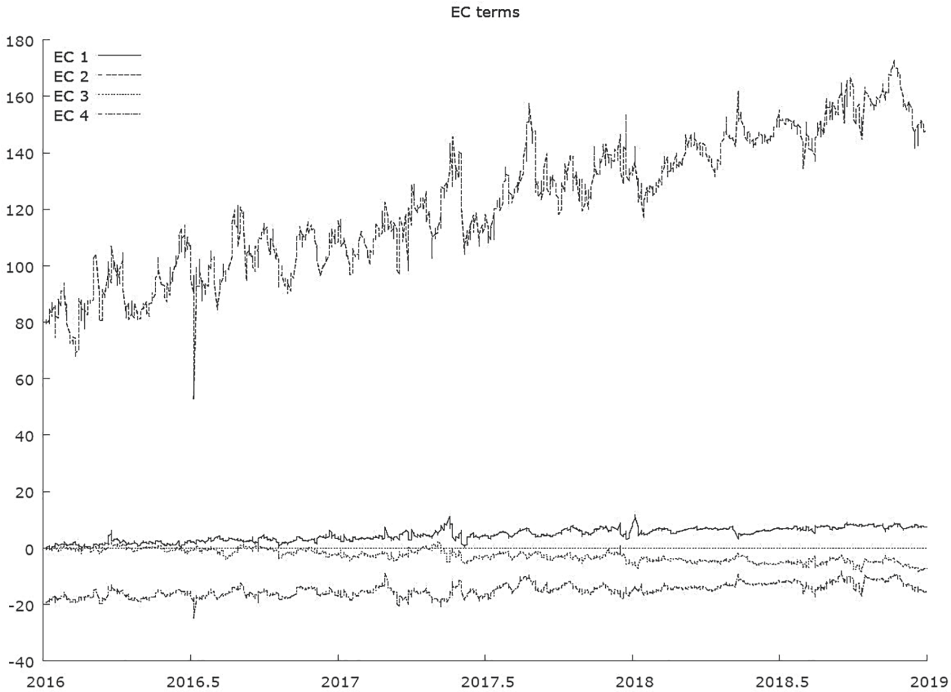


Fig. 1. This figure shows the cointegration equilibrium for privacy coins. The sample period is from January 1, 2016 until December 31, 2018 corresponding to 1096 observations.

where $\boldsymbol{\beta}_k$ with $k = \{1, \cdots, 4\}$ are parameter vectors that have the dimension 20x1. Specifically, in our ordering we selected those cryptocurrencies that have the highest and lowest market capitalization in both categories, privacy coins (i.e., DASH and PXI) and non-privacy coins (i.e., XRP and NMC) as the ones to be explained by the remaining cryptocurrencies. This approach allows us to identify whether the liquidity of cryptocurrencies matters for the formation of submarkets. Next, using our normalized matrix $\boldsymbol{\beta}$, we estimated a fully specified VECM using again all twenty cryptocurrencies. Since we are interested in the cointegration equilibria, we report only the point estimates for the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. The results are reported in Table 5.

Table 5 reveals some interesting results. First, the privacy coin market defined by PXI, which is in terms of market capitalization the smallest cryptocurrency investigated here, and defined by,

$$\boldsymbol{\beta}_4'\left(\boldsymbol{y}_{t-1} - \mu_{0,4} - \mu_{1,4}(t-1)\right)$$

seven out of eight non-privacy coins entering the cointegration equilibrium exhibit t-statistics indicating statistical significance on at least a 5% level. This result implies that the submarket defined here is not a submarket detached from the market for non-privacy coins. Second, the privacy coin market defined by DASH, which is in terms of its market capitalization the largest cryptocurrency investigated here, and defined by,

$$\boldsymbol{\beta}_2'\left(\boldsymbol{y}_{t-1} - \mu_{0,2} - \mu_{1,2}(t-1)\right)$$

only two out of eight non-privacy coins entering the cointegration equilibrium exhibit t-statistics indicating statistical significance on at least a 5% level. This result could indicate that this submarket defined is potentially detached from the market for non-privacy coins.

To test whether the submarket of privacy coins, defined by $\boldsymbol{\beta}_2'\left(\boldsymbol{y}_{t-1} - \mu_{0,2} - \mu_{1,2}(t-1)\right)$ is detached from the market for non-privacy coins, we set the following restrictions: $\beta_{2,5} = 0$, $\beta_{2,7} = 0$, $\beta_{2,9} = 0$, $\beta_{2,11} = 0$, $\beta_{2,13} = 0$, $\beta_{2,15} = 0$, $\beta_{2,17} = 0$, $\beta_{2,19} = 0$ and then re-estimate the VECM. Note that the imposed restrictions imply that all eight non-privacy coins are set equal to zero and the equilibrium is formed by privacy coins only. Note also that our normalization accounts for $\beta_{2,1} = 0$ and $\beta_{2,3} = 0$ (i.e., $\beta_{2,XRP} = 0$ and $\beta_{2,NMC} = 0$) in the original model specification. The point estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for the restricted model are reported in Table 6.

The log-likelihood of the unrestricted model (Table 5) is 42519.66, whereas the log-likelihood of the unrestricted model is 42518.06. Employing the Likelihood-Ratio (LR) test gives us an estimated value of,

$$\widehat{\lambda} = 2(42519.66 - 42518.06) = 3.20 < 11.07 = \chi^2_{0.95} \tag{5}$$

Since the estimated test statistic is clearly below the critical value of the corresponding reference distribution under null hypothesis

Table 7
Vector-Error Correction model estimates using all 21 cryptocurrencies (including Bitcoin).

i	Coins	$\widehat{\beta}_{1,i}$	$\widehat{\beta}_{2,i}$	$\widehat{\beta}_{3,i}$	$\widehat{\beta}_{4,i}$	$\widehat{\alpha}_{1,i}$	$\widehat{\alpha}_{2,i}$	$\widehat{\alpha}_{3,i}$	$\widehat{\alpha}_{4,i}$
1	**XRP(−1)**	1.0000	0.0000	0.0000	0.0000	−0.0076	−0.0172**	0.0004	0.0076
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(−0.870)	(−2.1270)	(0.0613)	(1.4020)
2	**DASH(−1)**	0.0000	1.0000	0.0000	0.0000	0.0025	−0.0036	−0.0008	0.0153***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.3580)	(−0.5500)	(−0.1675)	(3.5430)
3	**NMC(−1)**	0.0000	0.0000	1.0000	0.0000	−0.0015	0.0091	−0.0024	0.0146
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(−0.1372)	(0.8872)	(−0.3147)	(2.1390) **
4	**PXI(−1)**	0.0000	0.0000	0.0000	1.0000	0.0292	0.0255	0.0324**	−0.0524***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(1.3410)	(1.2660)	(2.1740)	(−3.8890)
5	**LTC(−1)**	−3.2199***	0.8418***	3.1825***	0.2241	0.0090	−0.0097	0.0011	0.0170***
		(−5.4338)	(2.7834)	(3.1956)	(0.6288)	(1.3500)	(−1.5790)	(0.2422)	(4.1440)
6	**BCN(−1)**	−0.1283	−0.0653	−0.3487	−0.6140***	0.0715***	0.0272**	0.0429***	0.0184**
		(−0.3555)	(−0.3546)	(−0.5747)	(−2.8271)	(5.5220)	(2.2730)	(4.8310)	(2.3010)
7	**ETH(−1)**	0.4228	−0.5460***	−0.1801	0.7983***	0.0119*	0.0150**	0.0108**	0.0021
		(1.2135)	(−3.0696)	(−0.3076)	(3.1123)	(1.6120)	(2.1920)	(2.1300)	(0.4621)
8	**XMR(−1)**	−0.9931**	−1.1750***	3.0670***	−0.6145**	0.0053	0.0143*	0.0064	0.0096*
		(−2.0396)	(−4.7280)	(3.7483)	(−2.0980)	(0.6446)	(1.8750)	(1.1320)	(1.8870)
9	**DOGE(−1)**	−1.4630***	−0.02607	1.0270	1.2550***	0.0385***	0.0114	0.0256***	0.0069
		(−3.4138)	(−0.1192)	(1.4259)	(4.8685)	(4.7920)	(1.5390)	(4.6610)	(1.3840)
10	**XDN(−1)**	1.3271***	0.9191***	−2.8790***	−2.0910***	0.0533***	−0.0002	0.0301***	0.0383***
		(3.9482)	(5.3576)	(−5.0964)	(−10.3413)	(4.4480)	(−0.0224)	(3.6660)	(5.1720)
11	**PPC(−1)**	0.0423	−2.5830***	−0.2824	−1.7190***	0.0303***	0.0309***	0.0176***	0.0165***
		(0.0738)	(−8.8348)	(−0.2934)	(−4.9881)	(3.7830)	(4.1750)	(3.2110)	(3.3290)
12	**CLOAK(−1)**	−1.2639***	0.4614***	1.5613***	0.8283***	0.0061	−0.0192	−0.0019	−0.0086
		(−4.8640)	(3.4794)	(3.5751)	(5.2993)	(0.3853)	(−1.3120)	(−0.1792)	(−0.8799)
13	**BTS(−1)**	2.0520***	−0.1234	−2.5420***	0.0164	0.0281***	0.0179**	0.0255***	−0.0083
		(4.2823)	(−0.5041)	(−3.1560)	(0.0568)	(3.0490)	(2.1030)	(4.0430)	(−1.4530)
14	**AEON(−1)**	−0.6005*	0.4365***	−0.10107	0.8054***	0.0378***	0.0318***	0.0317***	−0.0139*
		(−1.7821)	(2.5380)	(−0.1785)	(3.9733)	(2.8540)	(2.5940)	(3.4890)	(−1.690)
15	**XLM(−1)**	−1.0560***	0.7228***	1.8220***	1.1150***	−0.0086	−0.0339***	−0.0047	0.0142**
		(−3.3453)	(4.4880)	(3.4360)	(5.8756)	(−0.8655)	(−3.6870)	(−0.6872)	(2.3100)
16	**NAV(−1)**	−0.7261***	0.0948**	1.5061***	0.2218	−0.0013	0.0247*	0.0212**	−0.0204**
		(−5.3674)	(2.0063)	(4.2492)	(0.0271)	(−0.0883)	(1.7890)	(2.0720)	(−2.2070)
17	**NXT(−1)**	−0.7261**	0.0948	1.5061***	0.2218	0.0029	−0.0086	0.0024	0.0073
		(−2.3679)	(0.6059)	(2.9224)	(1.2024)	(0.3190)	(−1.0270)	(0.3853)	(1.2960)
18	**XVG(−1)**	1.0755***	−0.4860***	−1.3560***	−0.3850***	0.0789***	0.0526***	0.0706***	−0.0123
		(4.7743)	(−4.2275)	(−3.5814)	(−2.8405)	(4.5510)	(3.2800)	(5.9410)	(−1.1450)
19	**MAID(−1)**	−1.1073**	1.2521***	0.57559	−0.04574	−0.0048	−0.0093	−0.0051	0.0099**
		(−2.3552)	(5.2182)	(0.7284)	(−0.1617)	(−0.6112)	(−1.2850)	(−0.9578)	(2.0560)
20	**XST(−1)**	−2.0344***	0.0583	2.6091***	0.9837***	0.0508***	0.0029	0.0246**	0.0041
		(−7.4054)	(0.4159)	(5.6501)	(5.9526)	(3.3790)	(0.2049)	(2.3910)	(0.4403)
21	**BTC(−1)**	4.3059***	−1.1047***	−5.0060***	−2.0850***	0.0056	0.0050	0.0030	0.0127
		(6.3630)	(−3.1986)	(−4.4017)	(−5.1220)	(1.2340)	(1.1780)	(0.9537)	(4.4910)

Note: This table reports the estimates for a fully specified Vector-Error-Correction Model using all twenty-one cryptocurrencies in our sample including Bitcoin. Our model uses daily data of log prices. The model has a lag-order of $p = 2$. We report the estimates for the matrix $\boldsymbol{\beta}$ and the estimates for the adjustment parameter matrix $\boldsymbol{\alpha}$. The model allows for linear deterministic trend in data. The corresponding t-statistics are given in parentheses. The model is estimated with the econometric software Gretl. The sample period is from January 1, 2016 until December 31, 2018 corresponding to 1096 observations. Log-likelihood(lu) = 45873.105
*Statistically significant on a 10% level.
** Statistically significant on a 5% level.
*** Statistically significant on a 1% level.

which is chi-square distributed with five degrees of freedom, we cannot reject the null hypothesis (p-value 0.6698). This result implies that the privacy coin market, as defined by $\boldsymbol{\beta}_2{}'(\boldsymbol{y}_{t-1} - \mu_{0,2} - \mu_{1,2}(t-1))$ fulfills the conditions of a statistical equilibrium which is indeed detached from the market for non-privacy coins. In Fig. 1 we plot the cointegration equilibria $\boldsymbol{\beta}_k{}'(\boldsymbol{y}_{t-1} - \mu_0 - \mu_1(t-1))$ for $= \{1, \cdots, 4\}$, where $k = 2$ defines our cointegration equilibrium for privacy coins. We observe from Fig. 1 that the cointegration equilibrium for privacy coins shows a clear linear stationary trend which underpins the necessity to employ a fully specified VECM accounting for both constant and trend term (see Equation (3)).

Next, we have shown that adding Bitcoin to our set of coins and re-running the trace test provided evidence that the system of equations still generated four cointegration equilibria. To test the robustness of our results, we again add Bitcoin to our sample so that $\boldsymbol{y}_{t-1}$ is now a 21x1 vector given by

$$\boldsymbol{y}_{t-1} = (XRP_{t-1}, DASH_{t-1}, NMC_{t-1}, PXI_{t-1}, LTC_{t-1}, BCN_{t-1}, ETH_{t-1}, XMR_{t-1}, DODGE_{t-1}, XDN_{t-1}, \cdots$$

$$PPC_{t-1}, CLOAK_{t-1}, BTS_{t-1}, AEON_{t-1}, XML_{t-1}, NAV_{t-1}, NXT_{t-1}, XVG_{t-1}, MAID_{t-1}, XST_{t-1}, BTC_{t-1})$$

Since we are interested in the cointegration equilibria, we again report only the point estimates for the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$,

Table 8
Restricted Vector-Error Correction model estimates using all 21 cryptocurrencies (including Bitcoin).

i	Coins	$\widehat{\beta}_{1,i}$	$\widehat{\beta}_{2,i}$	$\widehat{\beta}_{3,i}$	$\widehat{\beta}_{4,i}$	$\widehat{\alpha}_{1,i}$	$\widehat{\alpha}_{2,i}$	$\widehat{\alpha}_{3,i}$	$\widehat{\alpha}_{4,i}$
1	**XRP(−1)**	2.7920	−23.7910	0.2950	3.6640	0.0031***	−0.0001	0.0010	0.0012
						(2.9240)	(−0.3261)	(0.9609)	(1.2960)
2	**DASH(−1)**	0.6610	−19.1560	−5.4660	0.2790	0.0033***	0.0001	0.0003	−0.0005
						(4.0070)	(0.8537)	(0.3134)	(−0.6296)
3	**NMC(−1)**	0.1800	31.0960	−3.9900	12.3180	0.0029**	0.00003	−0.0018	−0.0023**
						(2.186)	(0.1268)	(−1.3880)	(−2.025)
4	**PXI(−1)**	4.3980	−7.8020	2.3050	−3.9460	−0.0076***	−0.0011**	−0.0012	0.0071***
						(−3.0060)	(−2.1690)	(−0.4614)	(3.1430)
5	**LTC(−1)**	−3.6900	0.0000	−8.8260	−4.0540	0.0034***	0.0001	0.0018**	0.0002
						(4.2730)	(0.4629)	(2.3360)	(0.3161)
6	**BCN(−1)**	−2.7300	−26.8060	1.7350	−7.4520	0.0079***	−0.0005	−0.0001	0.0070***
						(5.1460)	(−1.5720)	(−0.0748)	(5.2810)
7	**ETH(−1)**	3.4570	0.0000	4.7460	−1.3560	0.0022**	−0.0001	−0.0020**	0.0012
						(2.4650)	(−0.4883)	(−2.3180)	(1.6270)
8	**XMR(−1)**	−1.4310	−64.3710	4.3720	−4.8660	0.0034***	−0.00008	−0.0026***	−0.0003
						(3.5030)	(−0.42620)	(−2.6420)	(−0.3547)
9	**DOGE(−1)**	2.6610	0.0000	1.7870	−9.1560	0.0035***	−0.0008***	0.0006	0.0036***
						(3.6700)	(−4.0000)	(0.5781)	(4.3010)
10	**XDN(−1)**	−8.5870	33.4290	−6.7550	8.2280	0.0099***	−0.0011***	0.0028*	0.0029**
						(7.0250)	(−3.8160)	(1.9610)	(2.3340)
11	**PPC(−1)**	−8.0090	0.0000	14.2560	−7.3220	0.0051***	−0.0001	−0.0032***	0.0006
						(5.3590)	(−0.6383)	(−3.3300)	(0.7335)
12	**CLOAK(−1)**	1.6940	10.7940	−4.2270	1.0230	−0.0042**	−0.0005	0.0047**	0.0004
						(−2.2340)	(−1.2590)	(2.5280)	(0.2280)
13	**BTS(−1)**	2.8940	0.0000	4.3040	1.1000	0.0011	−0.0008***	−0.0012	0.0037***
						(0.9834)	(−3.7300)	(−1.1370)	(3.8850)
14	**AEON(−1)**	1.5060	24.8650	−1.8870	−1.1170	−0.0006	−0.0013***	−0.0021	0.0030**
						(−0.3696)	(−4.2610)	(−1.3250)	(2.2220)
15	**XLM(−1)**	3.9810	0.0000	−5.6250	2.8620	0.0032***	−0.0001	0.0038***	0.0007
						(2.6700)	(−0.3685)	(3.2000)	(0.6931)
16	**NAV(−1)**	1.1950	39.9050	3.0150	6.6620	−0.0010	−0.0018***	−0.0047***	0.00003
						(−0.5785)	(−5.1130)	(−2.6730)	(0.0217)
17	**NXT(−1)**	0.4050	0.0000	−2.7740	2.9580	0.0019*	−0.0002	0.0012	0.0006
						(1.7750)	(−0.9222)	(1.0740)	(0.5848)
18	**XVG(−1)**	−0.4000	1.6220	4.2720	−0.2050	0.0049**	−0.0025***	−0.0039*	0.0087***
						(2.3800)	(−6.2330)	(−1.9210)	(4.8620)
19	**MAID(−1)**	−1.9200	0.0000	−8.4200	−0.4170	0.0013	−0.0005	0.0037	0.0047
						(1.6440)	(0.6050)	(1.0090)	(−1.2690)
20	**XST(−1)**	1.8850	−14.8190	−1.4020	−5.6850	0.0013	−0.0005	0.0037**	0.0047***
						(0.7027)	(−1.4040)	(2.1020)	(3.0230)
21	**BTC(−1)**	−3.1250	0.0000	9.9740	5.4360	0.0032***	0.00001	−0.0009	−0.0004
						(5.9210)	(0.0585)	(−1.5900)	(−0.8382)

Note: This table reports the estimates for a restricted Vector-Error-Correction Model using the set of all cryptocurrencies that are endogenous including Bitcoin. Our model uses daily data of log prices. The model has a lag-order of $p = 2$. We report the estimates for the matrix $\boldsymbol{\beta}$ and the estimates for the adjustment parameter matrix $\boldsymbol{\alpha}$. The model allows for linear deterministic trend in data. The corresponding t-statistics for the adjustment parameter matrix $\boldsymbol{\alpha}$ are given in parentheses. The model is estimated with the econometric software Gretl. The sample period is from January 1, 2016 until December 31, 2018 corresponding to 1096 observations. Unrestricted loglikelihood (lu) = 45873.105, Restricted loglikelihood (lr) = 45869.056, 2 * (lu - lr) = 8.09739, P(Chi-square (6) > 8.09739) = 0.231055
* Statistically significant on a 10% level.
** Statistically significant on a 5% level.
*** Statistically significant on a 1% level.

respectively. The results are reported in Table 7.

Again, we normalized the matrix $\boldsymbol{\beta}$ which is now of dimension 21x4, and the choose the same ordering as before, that is,

$$\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4] = \begin{bmatrix} \beta_{XRP,1} & 0 & 0 & 0 \\ 0 & \beta_{DASH,2} & 0 & 0 \\ 0 & 0 & \beta_{NMC,3} & 0 \\ 0 & 0 & 0 & \beta_{PXI,4} \\ \beta_{LTC,1} & \beta_{LTC,2} & \beta_{LTC,3} & \beta_{LTC,4} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{XST,1} & \beta_{XST,2} & \beta_{XST,3} & \beta_{XST,4} \\ \beta_{BTC,1} & \beta_{BTC,2} & \beta_{BTC,3} & \beta_{BTC,4} \end{bmatrix}$$

where $\boldsymbol{\beta}_k$ with $k = \{1, \cdots, 4\}$ are parameter vectors that have the dimension 21x1. Again, we test whether the submarket of privacy coins, defined by $\boldsymbol{\beta}_2{}'\left(\boldsymbol{y}_{t-1} - \mu_{0,2} - \mu_{1,2}(t-1)\right)$ is detached from the market for non-privacy coins by re-estimating the VECM using the following restrictions: $\beta_{2,5} = 0, \beta_{2,7} = 0, \beta_{2,9} = 0, \beta_{2,11} = 0, \beta_{2,13} = 0, \beta_{2,15} = 0, \beta_{2,17} = 0, \beta_{2,19} = 0$, and additionally $\beta_{2,21} = 0$. The imposed restrictions imply that all nine non-privacy coins including Bitcoin are set equal to zero, which means that the equilibrium is formed by privacy coins only. Note again that our chosen normalization accounts for $\beta_{2,1} = 0$ and $\beta_{2,3} = 0$ (e.g., $\beta_{2,XRP} = 0$ and $\beta_{2,NMC} = 0$) in the original model specification. The point estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for the restricted model are reported in Table 8.

The log-likelihood of the unrestricted model (Table 7) is 45873.11, whereas the log-likelihood of the unrestricted restricted model is 45869.06. Employing the Likelihood-Ratio (LR) test gives us an estimated value of

$$\widehat{\lambda} = 2(45873.11 - 45869.06) = 8.10 < 12.59 = \chi^2_{0.95} \tag{6}$$

Since the estimated test statistic is below the critical value of the corresponding reference distribution under null hypothesis which is chi-square distributed with six degrees of freedom, we cannot reject the null hypothesis (p-value 0.2311). This result implies that despite of accounting for Bitcoin as the cryptocurrency that exhibits the highest market capitalization the privacy coin market, as defined by $\boldsymbol{\beta}_2{'}\left(\boldsymbol{y}_{t-1} - \mu_{0,2} - \mu_{1,2}(t-1)\right)$ fulfills the conditions of a statistical equilibrium. This statistical equilibrium is detached from the market for non-privacy coins. Hence, the results of our robustness check strongly confirms our previous findings.

5. Conclusion

Goldfeder et al. (2018) show how third-party web trackers can de-anonymize users of Bitcoin and other non-privacy coins. Due to financial transparency, institutions are hesitant to use non-privacy cryptocurrencies as a medium of exchange. The emerge of privacy coins remedies this issue by using features like masternode technology, ring signature, and a stealth wallet address, to make it impossible for third parties to trace transactions to the real parties involved. We hypothesized that the traders, who favor privacy over complete transparency, are emerging as a different subgroup in digital financial markets. A common feature of cryptocurrencies is that the total supply of cryptocurrencies is often predetermined. As a consequence, the price processes depend solely on the demand side, that is, the users. Given that the user base for privacy coins is different from that for non-privacy coins, we would expect that privacy coins form a submarket of cryptocurrencies that is detached from the market for non-privacy coins.

To explore this issue, we make use of cointegration analysis which has at least three benefits; first, we can test whether the market for privacy coins generates a cointegration equilibrium that is detached from the market for non-privacy coins. Second, given that a cointegration equilibrium exists, the model allows us to test at the same for market efficiency. Third, and finally, making use of this model has the advantage that is does not require the specific formulation of an equilibrium price mechanism. Using the whole set of cryptocurrencies consisting of twenty cryptocurrencies to estimate the model, we find evidence for four cointegration equilibria. Accounting for liquidity, we estimate a fully-specified VECM by selecting the privacy and non-privacy coins with highest and lowest market capitalization as the four left-hand side variables in the model. Whereas in the equation modeling the privacy coin with lowest market capitalization (PXI) enter both categories of cryptocurrencies, that is, privacy and non-privacy coins, in the equation modeling DASH, which is the privacy coin with highest market capitalization, only two of non-privacy coins entering the equation exhibit statistical significance. Performing a Likelihood-Ratio test, we find that we cannot reject the null hypothesis that the whole set of non-privacy coins is jointly insignificant.

First, an immediate implication of cointegration is the existence of Granger-causal orderings among cointegrated series, which implies that asset prices determined in a weakly efficient market cannot be cointegrated. Hence, our findings provide evidence for market inefficiency. Second, the cointegration equilibrium associated with DASH appears to be disconnected from the market for non-privacy coins. A novel aspect of our study is it providing evidence that the underlying forces that cause large cap privacy coins equilibrium are unrelated to those at work in the non-privacy coins market. One explanation could be behavioral type: It could be that the market actors in the privacy coin market are different from those that trade in the non-privacy coin market. For instance, criminals involved in money laundry could favor privacy coins exhibiting a high level of liquidity as the sums involved could be substantial making small cap privacy coins an inappropriate choice for money laundry.

However, future studies might explore the market heterogeneity in the cryptocurrency market in more detail. Also, potential factors that might have caused the cointegration relationships should be the subject of future research. Third, if the stable cointegration relationship between asset prices is known to the market participants they would be able to exploit it and be in a position to profit. There is a broad stream of literature dealing with pairs trading, for instance, that requires the presence of cointegrated assets. Future research investigating this issue in the context of new digital currency markets would be welcome. Finally, since cryptocurrencies are by definition privy of intrinsic value - unlike fiat money issued by governments - the extent to which shocks may propagate across cryptocurrencies raises questions as to whether any observed spillover stems from investors' rebalancing activity and the accompanying price pressures rather than from fundamental information transmission. Future research is needed to elaborate on these mechanisms.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

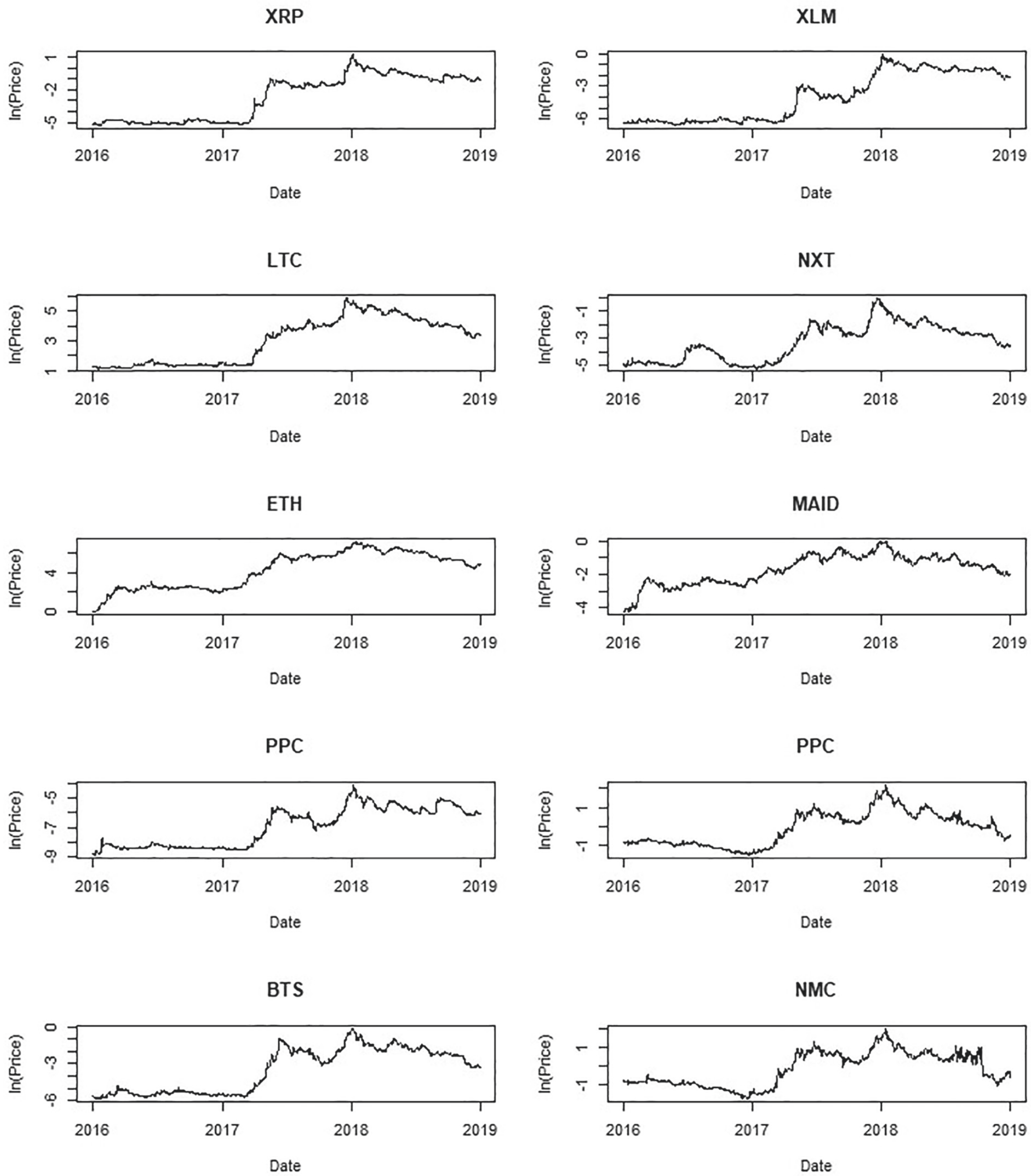


Fig. A1. Trend in the data levels of top 10 non-privacy coins 2016–2018(excluding Bitcoin).

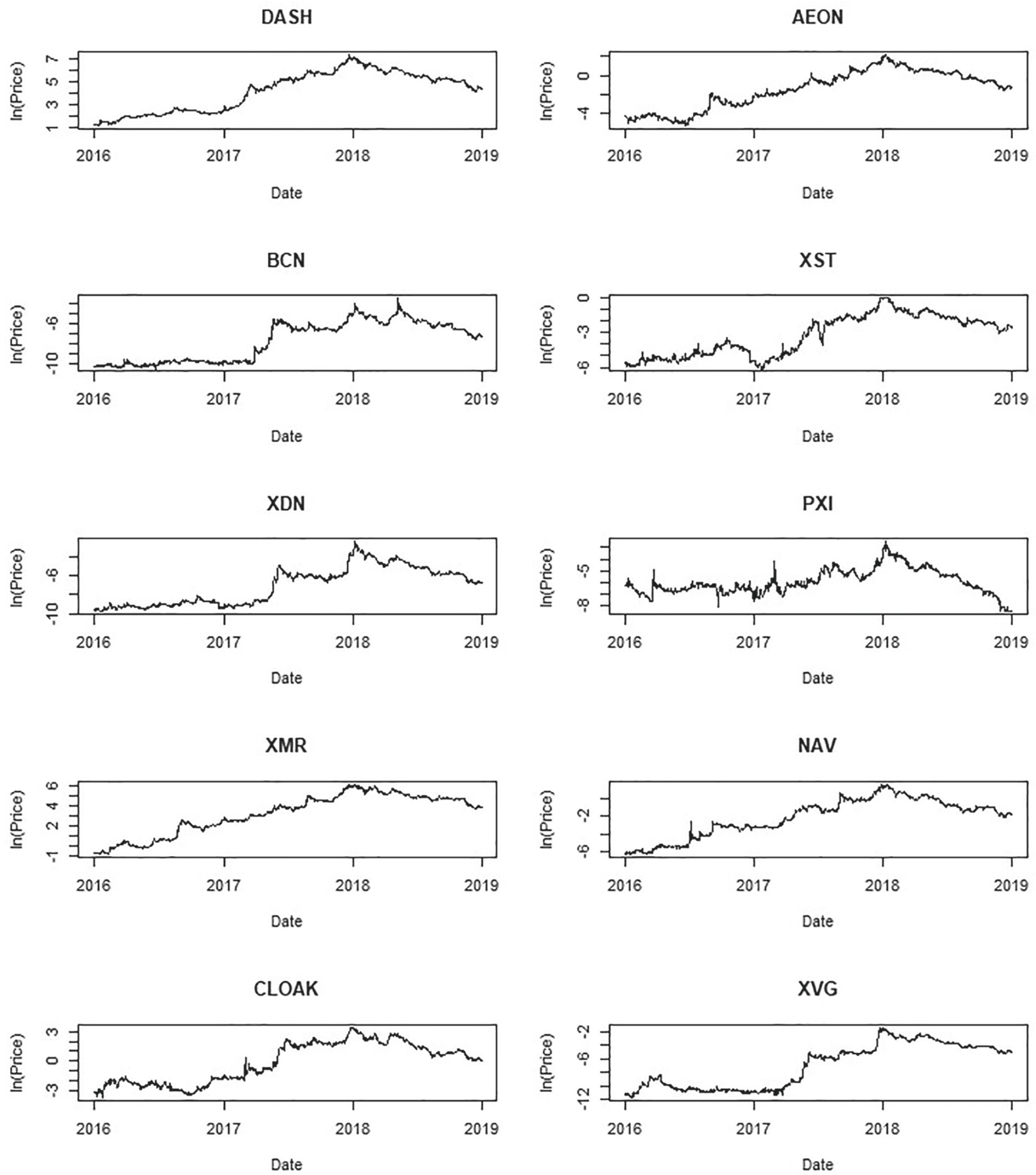


Fig. A2. Trend in the data levels of top 10 privacy coins 2016–2018.

Table A1
VAR optimal lag order selection.

lags	loglik	p(LR)	AIC	BIC	HQC
1	40392.0140		−74.3365	−72.4894*	−73.6369*
2	40898.6327	0.0000	−74.6071*	−71.0889	−73.2747
3	41194.6165	0.0000	−74.4863	−69.2970	−72.5210
4	41495.7699	0.0000	−74.3750	−67.5146	−71.7769
5	41753.5155	0.0000	−74.1831	−65.6515	−70.9521
6	42007.6063	0.0000	−73.9844	−63.7817	−70.1205
7	42302.2864	0.0000	−73.8611	−61.9873	−69.3643
8	42549.7210	0.0000	−73.6500	−60.1050	−68.5203
9	42845.2508	0.0000	−73.5283	−58.3122	−67.7658
10	43120.1391	0.0000	−73.3683	−56.4810	−66.9728
11	43442.7987	0.0000	−73.2970	−54.7386	−66.2687
12	43730.9191	0.0000	−73.1616	−52.9320	−65.5003
13	44012.0285	0.0000	−73.0131	−51.1124	−64.7189
14	44335.3340	0.0000	−72.9430	−49.3712	−64.0160
15	44645.5180	0.0000	−72.8485	−47.6056	−63.2886
16	44942.0211	0.0000	−72.7287	−45.8146	−62.5359
17	45257.0731	0.0000	−72.6433	−44.0580	−61.8176
18	45574.6329	0.0000	−72.5625	−42.3061	−61.1039
19	45894.2404	0.0000	−72.4856	−40.5581	−60.3941
20	46206.6556	0.0000	−72.3953	−38.7966	−59.6709

Note: VAR system, maximum lag order 20
The asterisks indicate the best (that is, minimized) values
of the respective information criteria, AIC = Akaike criterion,
BIC = Schwarz Bayesian criterion and HQC = Hannan-Quinn criterion.

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intfin.2021.101402.

References

Ahmed, S., Grobys, K., Sapkota, N., 2020. Profitability of Technical Trading Rules among Cryptocurrencies with Privacy Function. Finance Res. Lett. 35, 101495. https://doi.org/10.1016/j.frl.2020.101495.
Alexander, C., Dimitriu, A., 2005. Indexing Cointegration and Equity Market Regimes. *J. Int. Finance Econ.* 10 (3), 213–231.
Al-Yahyaee, K.H., Mensi, W., Yoon, S.-M., 2018. Efficiency, multifractality, and the long-memory property of the Bitcoin market: A comparative analysis with stock, currency, and gold markets. Finance Res. Lett. 27, 228–234.
Androulaki, E., Karame, G.O., Roeschlin, M., Scherer, T. and Capkun, S., 2013, April. Evaluating user privacy in bitcoin. In International Conference on Financial Cryptography and Data Security (pp. 34-51). Springer, Berlin, Heidelberg.
Aroskar, R., Sarkar, S.K., Swanson, P.E., 2004. European foreign exchange market efficiency: Evidence based on crisis and noncrisis periods. Int. Rev. Financial Analy. 13 (3), 333–347.
Bariviera, A.F., 2017. The inefficiency of Bitcoin revisited: A dynamic approach. Economics Letters 161, 1–4.
Baur, D.G., Hong, K., Lee, A.D., 2018. Bitcoin: Medium of exchange or speculative assets? J. Int. Financial Markets, Inst. Money 54, 177–189.
Borri, N., 2019. Conditional tail-risk in cryptocurrency markets. J. Empirical Finance 50, 1–19.
Bouri, E., Molnár, P., Azzi, G., Roubaud, D., Hagfors, L.I., 2017. On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? Finance Res. Lett. 20, 192–198.
Brenig, C., Accorsi, R., Müller, G., 2015, May. Economic Analysis of Cryptocurrency Backed Money Laundering. In ECIS.
Copeland, L.S., 1991. Cointegration tests with daily exchange rate data. Oxford Bull. Econ. Stat. 53 (2), 185–198.
Corbet, S., Eraslan, V., Lucey, B., Sensoy, A., 2019. The effectiveness of technical trading rules in cryptocurrency markets. Finance Res. Lett. 31, 32–37.
Dyhrberg, A.H., 2016. Hedging capabilities of bitcoin. Is it the virtual gold? Finance Res. Lett. 16, 139–144.
Engle, R.F., Granger, C.W.J., 1987. Cointegration and error correction: representation, estimation, and testing. Econometrica 55 (2), 251. https://doi.org/10.2307/1913236.
Foley, S., Karlsen, J.R., Putniņš, T.J., 2019. Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? Rev. Financial Stud. 32 (5), 1798–1853.
Fry, J., Cheah, E.-T., 2016. Negative bubbles and shocks in cryptocurrency markets. Int. Rev. Financial Analy. 47, 343–352.
Gerritsen, D.F., Bouri, E., Ramezanifar, E., Roubaud, D., 2020. The profitability of technical trading rules in the Bitcoin market. Finance Res. Lett. 34, 101263. https://doi.org/10.1016/j.frl.2019.08.011.
Goldfeder, S., Kalodner, H., Reisman, D., Narayanan, A., 2018. When the cookie meets the blockchain: Privacy risks of web payments via cryptocurrencies. Proc. Privacy Enhancing Technol. 2018 (4), 179–199.
Granger, C.J., 1986. Developments in the study of cointegrated economic variables. Oxford Bull. Econ. Stat. 48 (3), 213–228.
Grobys, K., Ahmed, S., Sapkota, N., 2020. Technical Trading Rules in the Cryptocurrency Market. Finance Res. Lett. 32, 101396. https://doi.org/10.1016/j.frl.2019.101396.
Haug, A.A., MacKinnon, J.G., Michelis, L., 2000. European Monetary Union: a cointegration analysis. J. Int. Money Finance 19 (3), 419–432.
Jensen, M.C., 1978. Some anomalous evidence regarding market efficiency. J. Financ. Econ. 6 (2-3), 95–101.
Jeon, B.N., Seo, B., 2003. The impact of the Asian financial crisis on foreign exchange market efficiency: The case of East Asian countries. Pacific-Basin Finance J. 11 (4), 509–525.
Johansen, S., 1994. The role of the constant and linear terms in cointegration analysis of nonstationary variables. Econometric Rev. 13 (2), 205–229.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica: J. Econometric Soc., pp.1551-1580.
Johansen, S., 1992a. Determination of cointegration rank in the presence of a linear trend. Oxford Bull. Econ. Stat. 54 (3), 383–397.
Johansen, S., 1992b. Testing weak exogeneity and the order of cointegration in UK money demand data. J. Policy Model. 14 (3), 313–334.
Johansen, S., 1995. Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press on Demand.
Kethineni, S., Cao, Y., 2020. The Rise in Popularity of Cryptocurrency and Associate Criminal Activity. International Criminal Justice Review 30 (3), 325–344.
Kus Khalilov, M.C., Levi, A., 2018. A survey on anonymity and privacy in bitcoin-like digital cash systems. IEEE Commun. Surv. Tutorials 20 (3), 2543–2585.
Kopp, H., Mödinger, D., Hauck, F., Kargl, F., Bösch, C., 2017, April. Design of a privacy-preserving decentralized file storage with financial incentives. In 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW) (pp. 14-22). IEEE.
Kumar, A., Fischer, C., Tople, S., Saxena, P., 2017, September. A traceability analysis of monero's blockchain. In European Symposium on Research in Computer Security (pp. 153-173). Springer, Cham.
Lo, A.W., Mamaysky, H., Wang, J., 2000. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. J. Finance 55 (4), 1705–1765.
Liu, J., Li, W., Karame, G.O., Asokan, N., 2018. Toward Fairness of Cryptocurrency Payments. IEEE Secur. Priv. 16 (3), 81–89.
Lütkepohl, H., Krätzig, M., Phillips, P.C. (Eds.), 2004. Applied time series econometrics. Cambridge university press.
Möser, M., Soska, K., Heilman, E., Lee, K., Heffan, H., Srivastava, S., Hogan, K., Hennessey, J., Miller, A., Narayanan, A. and Christin, N., 2017. An empirical analysis of traceability in the monero blockchain. *arXiv preprint arXiv:1704.04299*.
Miller, N., Yang, Y., Sun, B., Zhang, G., 2019. Identification of technical analysis patterns with smoothing splines for bitcoin prices. J. Appl. Statist. 46 (12), 2289–2297.
Nadarajah, S., Chu, J., 2017. On the inefficiency of Bitcoin. Econ. Lett. 150, 6–9.
Norrbin, S.C., 1996. Bivariate cointegration among European monetary system exchange rates. Appl. Econ. 28 (12), 1505–1513.
Osterrieder, J., Lorenz, J., 2017. A statistical risk assessment of Bitcoin and its extreme tail behavior. Annals of Financial Economics 12 (01), 1750003.
Phengpis, C., 2006. Market efficiency and cointegration of spot exchange rates during periods of economic turmoil: Another look at European and Asian currency crises. J. Econ. Bus. 58 (4), 323–342.
Rangvid, J., Sørensen, C., 2002. Convergence in the ERM and declining numbers of common stochastic trends. Journal of Emerging Market Finance 1 (2), 183–213.
Richards, A.J., 1995. Comovements in national stock market returns: Evidence of predictability, but not cointegration. J. Monetary Econ. 36 (3), 631–654.
Ouaddah, A., Elkalam, A.A., Ouahman, A.A., 2017. Towards a novel privacy-preserving access control model based on blockchain technology in IoT. In Europe and MENA Cooperation Advances in Information and Communication Technologies(pp. 523-533). Springer, Cham.
Sensoy, A., 2019. The inefficiency of Bitcoin revisited: A high-frequency analysis with alternative currencies. Finance Res. Lett. 28, 68–73.
Urquhart, A., 2016. The inefficiency of Bitcoin. Econ. Lett. 148, 80–82.
Vidal-Tomás, D., Ibañez, A., 2018. Semi-strong efficiency of Bitcoin. Finance Res. Lett. 27, 259–265.
Woo, K.Y., 1999. Cointegration analysis of the intensity of the ERM currencies under the European Monetary System. Journal of International Financial Markets, Institutions and Money 9 (4), 393–405.

Niranjan Sapkota completed his Master's degree in Finance from the University of Oulu, Finland with an excellent grade in 2015. He is now working as a Ph.D. student in the Accounting and Finance department with Finance as a major at the University of Vaasa, Finland. His primary focus area is Digital Finance, Financial Uncertainty and Risk. He is also interested in the area of Market Microstructure and High Frequency Trading.

Klaus Grobys studied Economics and Econometrics at the University of Kiel and the University of Stockholm. He holds a Ph.D. degree in Finance passed with distinction from the University of Vasa where he graduated in December 2014. Since January 2015 he works as Assistant Professor of Finance at the University of Vaasa. He has published about 20 articles related to Asset Pricing in journals such as Quantitative Finance, International Review of Financial Analysis, and Journal of Futures Markets among others. His most recent research interests are related to cryptocurrencies.

International Review of Financial Analysis 82 (2022) 102183

Contents lists available at ScienceDirect

International Review of Financial Analysis

journal homepage: www.elsevier.com/locate/irfa

News-based sentiment and bitcoin volatility☆

Niranjan Sapkota *

University of Vaasa, Vaasa, Finland

ARTICLE INFO

JEL classification:
G11
G14
G17
G41

Keywords:
Bitcoin
News sentiments
Natural language processing
Range-based volatility
HAR-RV (heterogeneous autoregressive realized volatility)

ABSTRACT

In this work, I studied whether news media sentiments have an impact on Bitcoin volatility. In doing so, I applied three different range-based volatility estimates along with two different sentiments, namely psychological sentiments and financial sentiments, incorporating four various sentiment dictionaries. By analyzing 17,490 news coverages by 91 major English-language newspapers listed in the LexisNexis database from around the globe from January 2012 until August 2021, I found news media sentiments to play a significant role in Bitcoin volatility. Following the heterogeneous autoregressive model for realized volatility (HAR-RV)—which uses the heterogeneous market idea to create a simple additive volatility model at different scales to learn which factor is influencing the time series—along with news sentiments as explanatory variables, showed a better fit and higher forecasting accuracy. Furthermore, I also found that psychological sentiments have medium-term and financial sentiments have long-term effects on Bitcoin volatility. Moreover, the National Research Council Emotion Lexicon showed the main emotional drivers of Bitcoin volatility to be anticipation and trust.

1. Introduction

Around 2.5 billion people in the world read newspapers regularly in hard copy, whereas more than 600 million read newspapers in digital form.[1] The growing audience of news media, social media, and blogging websites has widened the scope of textual analysis beyond linguistic studies. From high-frequency traders using real-time news sentiment for trading activities to predicting future volatility, news sentiment has also become an essential market factor in finance. Therefore, an accurate estimation of positive or negative sentiment from the news is crucial for investment decision-making and portfolio management (Mishev, Gjorgjevikj, Vodenska, Chitkushev, & Trajanov, 2020). News also provides the opportunity to see a context analytically from a wider perspective. As a consequence, the reader can assess the quality of a project, product, or service through understanding the general sentiment of the crowd. However, many readers often misjudge the true sentiment behind the news. Fürsich (2009) argued that media texts present a distinctive discursive moment between encoding and decoding that requires special scholarly engagement. News generally provides either negative or positive sentiment to its readers. People are less interested in reading news articles of neutral sentiment (Dos Rieis et al., 2015). In this regard, psychological literature has frequently confirmed the priority of processing words with negative or positive emotion against words with neutral emotion (for example, Chen, Lin, Chen, Lu, & Guo, 2015; Kissler, Herbert, Peyk, & Junghofer, 2007; Yap & Seow, 2014; Zhang et al., 2014).

In the past decade, Bitcoin (BTC) has made a lot of news in mainstream media. According to 99bitcoins.com, BTC has "died" 432 times in the news.[2] While many newspapers have covered BTC as a possible scam or a bubble, some newspapers have highlighted the opportunities it has created. BTC hackings, crypto exchange collapses, government bans, regulations, taxes, scams, etc. have made many headlines in global news media outlets. Nonetheless, there has also been positive news of BTC such as a legal tender, means of payment, futures, exchange-traded funds, etc. Furthermore, news like Tesla acquiring $1.5 billion worth of BTC has given this digital innovation a significant positive-sentiment

☆ I am thankful to Prof. Mikko Leppämäki and Prof. Markku Kaustia, for providing useful comments at Aalto University PhD clinic, Finland, 2021. I am extremely thankful to Database Specialist Dr. Nikolas Breitkopf from Aalto University for his insightful comments as a discussant at the Graduate School of Finance (GSF) and the Finnish Doctoral Programme in Economics (FDPE) WINTER WORKSHOP IN FINANCE, at Aalto University, Finland, 2021. I would also like to thank Prof. Timo Rothovius, A.Prof. Klaus Grobys, and Researcher Dr. Anupam Dutta from the University of Vaasa, Finland for their encouraging comments.

* Corresponding author at: Department of Accounting and Finance, University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland.
E-mail address: niranjan.sapkota@uva.fi.

[1] See more at: http://www.ifabc.org/news/More-People-Read-Newspapers-Worldwide-Than-Use-Web

[2] See the details at: https://99bitcoins.com/bitcoin-obituaries/

https://doi.org/10.1016/j.irfa.2022.102183
Received 10 December 2021; Received in revised form 20 January 2022; Accepted 19 April 2022
Available online 29 April 2022

hype. Unfortunately, there was a reversal of sentiment when Tesla Chief Executive Officer (CEO) Elon Musk removed it as a payment option on the Tesla website, stating the energy and environmental risk imposed by this financial technology innovation. Both the Tesla acceptance and rejection were sensational in the news and crypto world. The sentiment conveyed or imposed by the media has been easily noticeable by looking at the BTC price swing after a big news item either in favor of it or against it. However, a swing could be triggered by sentiment from a small news item as well. Even if we could see the direction of the sentiment after a big news item, we cannot exactly measure the degree or magnitude of this qualitative phenomenon. In this regard, Bonato, Gkillas, Gupta, and Pierdzioch (2020) stated that investor sentiment cannot be directly measured or observed.

Thankfully, machine-learning tools like natural language processing are becoming handy in quantifying qualitative transcripts. We can now easily quantify news and articles, and they can provide more accurate, more efficient proxies for investor sentiments (for example, Ho, Shi, & Zhang, 2020; Shen et al., 2018) in comparison to traditional approaches like that of Baker and Wurgler (2006, 2007), who used several market-based measures as proxies. Besides the market-based measure, the other most common approach applied in earlier research has been survey-based indices. More recently, building an investor-sentiment index employing daily news, internet search, and social media data has gained popularity because traditional approaches like market-based and survey-based methods seem to be less transparent. Furthermore, the advantage of internet-based sentiment is that it can be extracted in real time and at a lower frequency, like every second, minute, hour, or day, compared to traditional approaches that extract every month, quarter, or year (Bonato et al., 2020). One can extract sentiment scores following various sentiment dictionaries. Studies comparing various sentiment dictionaries have focused on those of Henry (2008) and Loughran and McDonald (2011). Using the Harvard-IV general dictionary, Loughran and McDonald (2011) found its word list to be largely inapplicable to financial contexts and created a finance-specific list. Henry (2008) captured the tone of earnings press releases in order to create a word list for financial texts. Both studies found finance-specific word lists to be more powerful than general word lists.

In recent years, cryptocurrency markets have attracted considerable attention in the academic literature. This is not surprising given that from an economic perspective, the sums of money involved are substantial (Fry & Cheah, 2016). The current market cap of cryptocurrency is 2.6 trillion dollars, whereas the dominance of BTC is 45%—slightly less than half of the total market cap.[3] However, the dominance of BTC tends to fluctuate heavily following good or bad news, making this digital currency highly volatile.[4]

Modeling volatility is an important step to precisely measure the risk associated with an asset or portfolio of assets. An accurate estimation of volatility is vital to investors to develop an adequate strategy to hedge potential risks associated with an investment. In this paper, I used the past realized volatilities (RVs) of BTC to predict its future RVs by following the popular volatility-forecasting model proposed by Corsi (2009). The heterogeneous autoregressive model for realized volatility (HAR-RV) utilizes three AR(1) volatility processes at daily, weekly, and monthly windows. A natural economic interpretation of this model according to the author is that each volatility component in the model corresponds to a market component that forms expectations for the next period's volatility based on the observation of the current RV and the expectation for the longer-term volatility. By decomposing volatility into short-term (daily), medium-term (weekly), and long-term (monthly) frequencies, this model captures the heterogeneity among short-term, medium-term, and long-term investors. One of the flexibilities of the HAR-RV is that one can easily include additional explanatory variables in the ordinary-least-squares (OLS) regression equation.

Furthermore, in this study, I sought to explore whether news media sentiments have an impact on BTC volatility by including different sentiments as additional explanatory variables. On top of that, I differentiated between financial sentiment and psychological sentiment cached in the news and analyzed their impact on BTC volatility in different time windows, similar to the RV estimation, as daily, weekly, and monthly. I addressed the issue of heterogeneity in news arrival time and investors' sentiment memory length by incorporating short-term, medium-term, and long-term sentiment windows. While I incorporated sentiments as additional explanatory variables, one could also include RVs over other time windows besides daily, weekly, and monthly. An example can be found in the work of Busch, Christensen, and Nielsen (2011), who used implied volatility as an additional explanatory variable on top of RVs.

In this study, a comparison between the baseline HAR-RV and the HAR-RV extended with news sentiment index (HAR-RV-SI) showed the HAR-RV-SI to be a better fit. The out-of-sample forecast also showed that sentiments as explanatory variables in the HAR-RV have a higher forecasting accuracy. Furthermore, I also found psychological sentiments to have short-term and financial sentiments to have long-term effects on BTC volatility. Moreover, the results showed that either a mixture of positive and negative sentiments or purely positive sentiment is more responsible for BTC volatility, as compared with purely negative sentiment. Implementing the National Research Council (NRC) Emotion Lexicon showed the main emotional drivers of BTC volatility to be anticipation and trust.

This article contributes towards the multiple aspects in financial research such as literature, data, methodologies, sentiment approach, and practical implications. Firstly (*i*), this article contributes to the recent stream of financial literature in numerous ways. While earlier research studies have mostly focused on news sentiments around events related to macroeconomic announcements (for example, Andersen, Bollerslev, & Diebold, 2007; Corbet, Larkin, Lucey, Meegan, & Yarovaya, 2020; Corsi, Pirino, & Reno, 2010; Entrop, Frijns, & Seruset, 2020), this analysis covered all the BTC-related news sentiments published in major English-language newspapers from around the globe. Secondly (*ii*), this article contributes towards unique data as previous studies on sentiment and BTC price movements have mostly relied on news blogs and search websites rather than on mainstream newspapers (for example, Garcia, Tessone, Mavrodiev, & Perony, 2014; Karalevicius, Degrande, & De Weerdt, 2018; Kristoufek, 2013). I chose to go with the major newspapers. The main concern with creating a corpus with news blogs is the possible repetition or inclusion of advertisement texts along with the main news story. If screening is not done properly, sentiments will tilt more towards one direction as advertisements mostly trigger either positive or negative emotions. In this regard, Mcduff and Berger (2020) stated that when it comes to engaging the audience, what matters is not just provoking positive emotions, but provoking activating emotions, and those can be positive or negative. Another major issue with news blogs is that they are mostly clickbait, so sentiments in the headline and the main body do not necessarily always match. Even though I used a different data source in this study, I addressed the possibility of this issue by extracting sentiments from the whole body of the news story, not just the headline. While including as many blogs as possible might sound good, the majority of small news blogs generally copy or share their content from well-known cryptocurrency websites like cointelegraph.com, coindesk.com, etc., creating redundancies in the data sample and resulting in inaccurate estimation of investor sentiment. Furthermore, to overcome the issue of redundancies in the newspaper articles, I also used the filter of "maximum similarity" while searching the LexisNexis news database.

Next (*iii*), this article contributes methodologically by extending HAR-RV towards a new direction. To the best of my knowledge, no

[3] www.coinmarketcap.com (as of 26.10.2021)

[4] See more at: https://www.goldmansachs.com/insights/pages/crypto-a-new-asset-class-f/report.pdf

article has yet explored newspaper-based sentiment as an additional explanatory variable in the HAR-RV environment in forecasting future volatilities of BTC or any other digital financial asset. Furthermore (*iv*), by further classifying sentiments into psychological- and finance-specific and extending them into three different horizons to capture heterogeneity in news arrival time among readers, this article contributes towards a better understanding of time-varying news sentiments, their memory length, and their effect on BTC volatility. On top of that, this work studied the role of different human emotions by applying Emotion Lexicon–based sentiments and their implications on digital financial innovations like BTC. Finally (*v*), from the practitioner point of view, this paper also sheds light on capturing different sentiments in the news because accurate estimation of volatility is vital to investors for developing an adequate strategy in hedging potential risks associated with their investments.

2. Literature review

Sentiment analysis in general investigates opinions expressed in texts and their polarity as positive, negative, or neutral (Muhammad, Wiratunga, & Lothian, 2016). The Emotion Lexicon can further categorize sentiments into different human emotions like fear, trust, etc. Sometimes, the tone of news is perhaps more influential than its substantive content in the body. There have been plenty of studies exploring the sentiment of news content, political speeches, blogs, advertisements, financial statements, earnings announcements, etc. Earlier research has shown how news sentiment affects individual decision-making, especially political judgment. In this regard, Young and Soroka (2012) highlighted that negative sentiment has more impact on human psychology and political interactions. Furthermore, Tetlock (2007) highlighted the advantage of applying sentiment analysis to predict or forecast the return of financial assets as being able to measure the impact of a wide range of events without the need to specify them.

Previous literature has covered sentiments in a wide range of asset classes. News and social media sentiments impact foreign exchange, stocks, bonds (for example, Busch et al., 2011), and commodities (for example, Qadan & Nama, 2018; Zhang & Li, 2019; Dutta, Bouri, & Saeed, 2021). Investigating the sentiment in the oil market, Bonato et al. (2020) used the HAR-RV to analyze whether a measure of investor happiness predicts the daily RV of oil-price returns. They used high-frequency intraday data to measure RV and found it to be significantly negatively linked to investor happiness in the short term. Furthermore, they also found that investor happiness significantly improves the accuracy of RV forecasts in the short term. Besides the conventional asset class, a new strand of literature is exploring sentiment in new blockchain-based digital financial markets (for example, Entrop et al., 2020; Hu, Kuo, & Härdle, 2019; Sapkota & Grobys, 2021).

Karalevicius et al. (2018) highlighted that only a small number of studies have considered the sentiment of publicly available textual information as an indicator for BTC price movements. However, a growing number of studies are stepping into the exploration of this relationship. Aalborg, Molnár, and de Vries (2019) studied how the return and trading volume of BTC depends on other variables such as trading volume, number of unique BTC addresses, and Google search trends on BTC. They found that the past RV of BTC predicts its future RV on the HAR-RV setup. In addition to that, they found that trading volume improves volatility prediction. They further identified a causal relationship between Google search trends to trading volume and trading volume to BTC volatility. Another BTC sentiment paper that applied HAR-RV is that of Bouri, Gkillas, Gupta, and Pierdzioch (2021), who analyzed the role of the United States–China trade war in predicting the daily RV of BTC returns. They extended the HAR-RV to include a metric of United States–China trade tensions. Their findings revealed that United States–China trade uncertainty improves forecast accuracy.

Baillie, Calonaci, Cho, and Rho (2019) stated that long memory in RV is a widespread stylized fact. Long memory in RV has been synonymized with jumps, structural breaks, and nonlinearities. They highlighted the forecasting power of the HAR model and its extensions. They assessed the separate roles of fractionally integrated, long memory models, extended HAR models, and time-varying-parameter HAR models and found the presence of the long memory parameter to be often important in addition to the HAR model. Andersen et al. (2007), from analyses of exchange rates, equity index returns, and bond yields, found that the volatility jump component is highly important and that separating the rough-jump moves from the smooth-jump moves results in significant improvement in volatility forecast. Furthermore, they also found many of the significant jumps to be associated with specific macroeconomic news announcements. In this regard, Corsi et al. (2010) showed that fragmenting volatility into jumps and continuous variation substantially improves volatility forecasting because of the significant positive impact of past jumps on future volatility. Corbet et al. (2020) also examined the link between macroeconomic news announcements and BTC returns. They constructed a sentiment index based on news stories following the announcements of four macroeconomic indicators. They found that an increase in positive news surrounding unemployment rates and durable goods results in a corresponding increase in equity returns and a decrease in BTC returns. Furthermore, they also observed that an increase in the percentage of negative news surrounding the announcement is linked with an increase in BTC returns. They concluded that the cryptocurrency market is further maturing through interactions with macroeconomic news. On the contrary, Entrop et al. (2020) found that attention and macroeconomic news have no impact on the price discovery of BTC. In addition to that, they also showed higher news-based BTC sentiment to increase the informational role of the futures market. Rognone, Hyde, and Zhang (2020) also contributed to the current debate on the nature of BTC implementing news sentiment. They explored whether the digital currency should be considered a financial asset or a medium of exchange. They investigated the intraday relationship between BTC and the major fiat currencies to assess whether there exists a similar reaction to high-frequency unscheduled news sentiment applying a vector autoregression (VAR) model.

The rest of the paper is organized as follows. Section 3 briefly discusses data, data sources, and sentiment data generation from the news corpus. Section 4 provides detailed descriptions of the methodologies. Section 5 presents the results, and Section 6 concludes.

3. Data

I retrieved daily open, high, low, and close (OHLC) prices for BTC from the website investing.com. The BTC OHLC data sample was from January 1, 2012, until August 31, 2021, accounting for 3530 daily observations. I also downloaded BTC-related news covered by major English-language newspapers from around the globe from the LexisNexis news database. LexisNexis has a list of 91 English news media in its “Major Newspapers” category, which is reported in Appendix A.1. During the sample period, from January 1, 2012, until August 31, 2021, there were a total of 17,490 news pieces covered on BTC by these major newspapers, which were extracted using the search term “Bitcoin”. While searching the news database, one can limit LexisNexis search by Contents (for example, news, cases, etc.), Publication type (for example, newspapers, blogs, etc.), Language (for example, English, German, etc.), Industry (finance, media, technology) and many other criteria. However, there is no further filter within the “Newspaper” segment to see whether the news is a daily coverage or an article on that particular search topic. Therefore, the news items that I downloaded from LexisNexis includes not just the BTC-related news articles but also the news coverages on it. Furthermore, to reduce redundancies, I applied the filter of “maximum similarity” while searching this news database. The geographical representation of the newspapers included in the list of major newspapers can be considered global because it also covers big non-English–speaking countries around the world. The newspapers listed as major English language newspapers from around the globe

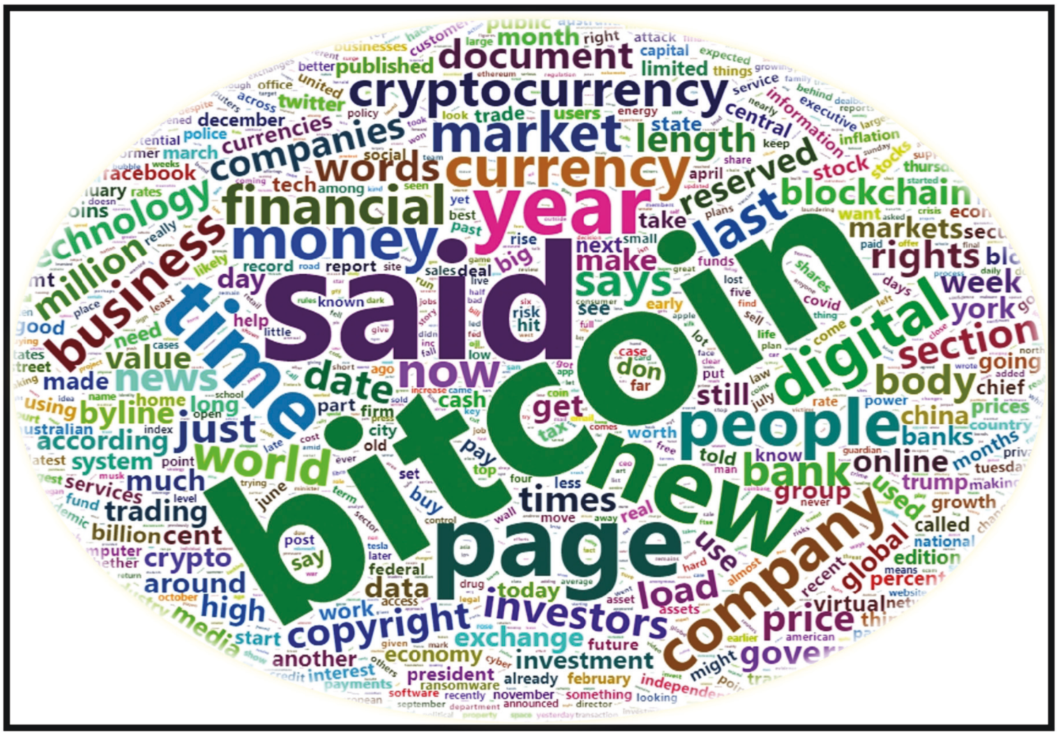


Fig. 1. Wordcloud of the most used words on BTC news by major English-language global newspapers (Jan 2012–Aug 2021). Note: This wordcloud was created using the wordcloud2 package in R.

includes newspapers from English speaking countries like the USA, UK, Australia, New Zealand, Canada, etc. as well as non-English speaking countries like China, Brazil, Singapore, Japan, Malaysia, Israel, etc.

Furthermore, earlier research has found that Google trends have a positive relationship with the future volatility of BTC. For example, Aalborg et al. (2019) found Google searches on BTC to predict trading volume and found trading volume to predict BTC volatility. As a control for the model, I also downloaded the daily Google search trends on BTC for the entire sample period of January 1, 2012, until August 31, 2021. Daily Google trends, by default, are not available through web interface or application programming interface (API). However, applying open-source code along with the gtrendsR package, I generated a daily time series of Google search trends on the term "Bitcoin" for the given sample period.[5]

3.1. *Extracting sentiment scores of news on Bitcoin*

From the LexisNexis database, one can either download all news files as a single file or by individual news article in PDF and other document formats.[6] I choose the latter, as the first option would make it complicated to convert the news into a time-series object. First of all, to create the corpus in a time-series format, I extracted the publication dates of all 17,490 news PDF files. Each news corpus starts with a header that includes the page number and the headline. The heading section is structured as follows:

1st line: Headline.

2nd line: Newspaper name (with/without location).

3rd line: Publication date.

By applying the "readLines" function, I extracted publication dates from most of the news files. To have the dates from the rest of the news articles, I used an alternative method for extracting the publication date by applying the "Key Word in Context" (KWIC) function. The PDF version of the downloaded news comes with the key word "load date" in the footer section of each file. The majority of the news files are uploaded in the LexisNexis database on the same day of publication; therefore, the "load date" can be used as the publication date.

Fig. 1 shows a "word cloud" of the most frequently used words by the 91 major English language–based newspapers during the sample period. Next, I extracted the BTC news sentiments by applying four different sentiment dictionaries. The Sentiment Analysis package in R statistical software supplies *positive*, *negative*, and *overall* sentiment scores for four different sentiment dictionaries applied to the corpus of news. I further grouped these four dictionaries into two subgroups, *psychological and discourse sentiment dictionaries* and *finance-specific sentiment dictionaries*. The "analyzeSentiment" function in this package gives sentiment scores for:

i. *Psychological and discourse sentiment dictionaries*
    a. Harvard-IV general-purpose psychological dictionary (GI)
    b. Quantitative Discourse Analysis Package (QDAP) dictionary
ii. *Finance-specific sentiment dictionaries*
    a. Henry's (2008) finance-specific dictionary (HE)
    b. Loughran and McDonald (2011) finance-specific dictionary (LM)

The Harvard-IV psychological dictionary is a general-purpose dictionary that maps a corpus with counts on positive, negative, and overall sentiments. Similarly, QDAP provides quantitative analysis of a

[5] R script for daily Google search trends can be found at http://alexdyachenko.com/all/how-to-get-daily-google-trends-data-for-any-period-with-r/

[6] A sample news file is attached as Appendix A.2.

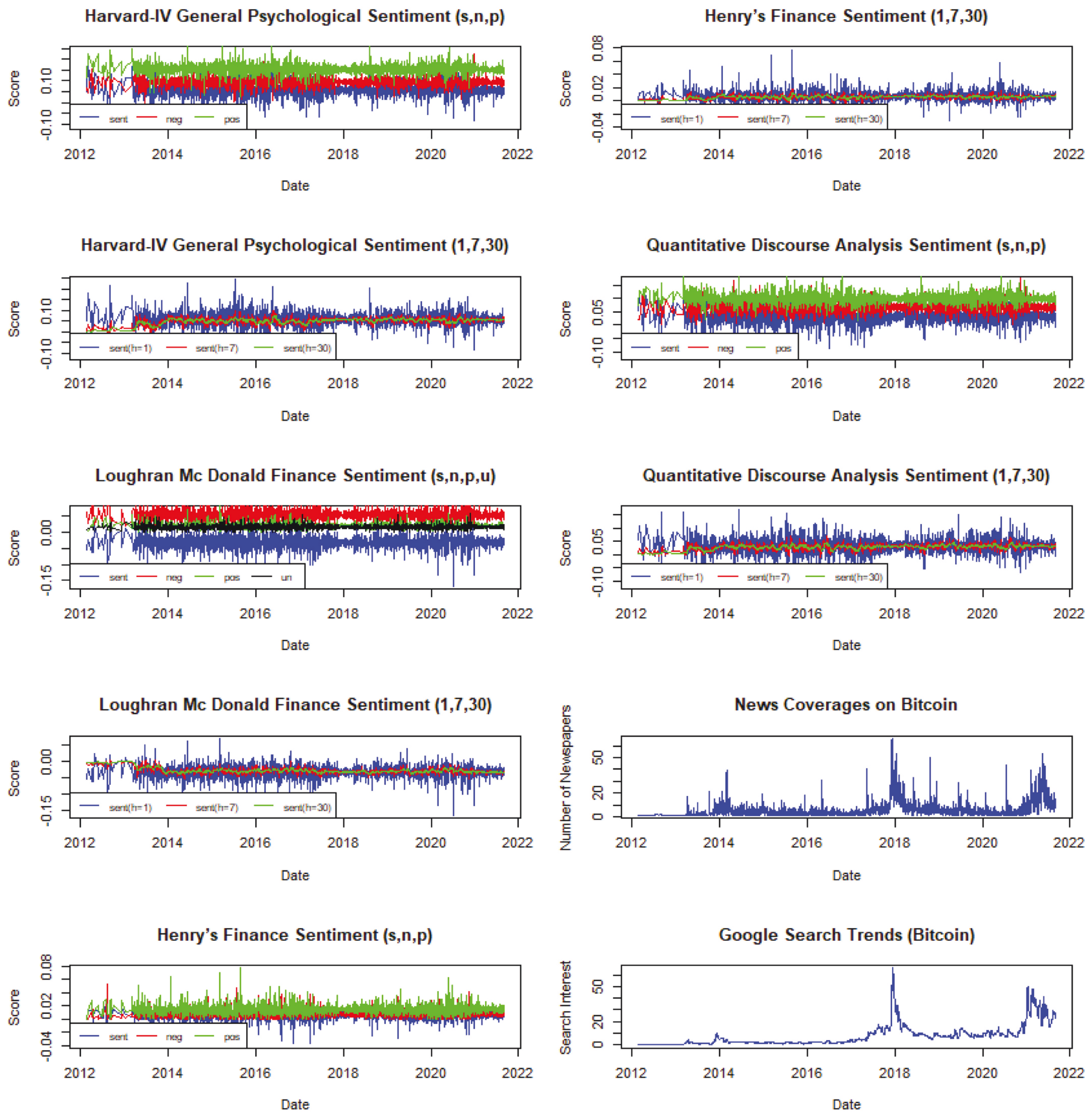


Fig. 2. BTC news coverage, sentiments, and daily worldwide Google search trends (Jan 2012–Aug 2021).
Note:
s = overall sentiment (p-n-u).
n = negative sentiment.
p = positive sentiment.
u = sentiment uncertainty.
1 = daily.
7 = weekly.
30 = monthly.

qualitative corpus. It also gives positive, negative, and overall sentiments of the news. For the finance-specific sentiments, Henry (2008) studied capital market data to assess the impact on investors of tone and other stylistic attributes. The dictionary categorizes sentiments as positive, negative, and overall. Another popular finance-specific dictionary is that of Loughran and McDonald (2011). This measure gives positive, negative, risk for sentiment uncertainty, and overall sentiment scores of the text. Fig. 2 shows the polarity and time-varying sentiments on BTC news coverage and daily worldwide Google search trends.

Table 1.a summarizes the statistics of the news coverage, Google

Table 1.a
Summary Statistics on Sentiments and BTC News Pieces (Jan 2012–Aug 2021).

Variable	Nobs	Min	Max	1st Qu	3rd Qu	Mean	Median	StDev	Skew	Kurt
Btc_G	3525	0.00	100.0000	1.7400	9.9000	7.6108	4.1300	9.4241	2.80	11.81
WordC	2802	31.00	13,780	375	606.00	574.35	476.50	638.62	12.33	208.62
NewsC	2802	1.00	66.0000	2.0000	7.0000	6.1934	4.0000	7.0613	3.08	12.67
S_GI_1	2802	−0.11	0.2427	0.0426	0.0739	0.0585	0.0583	0.0302	0.06	3.78
S_GI_7	2802	−0.02	0.0989	0.0464	0.0617	0.0536	0.0552	0.0136	−0.68	1.51
S_GI_30	2802	0.00	0.0742	0.0489	0.0585	0.0527	0.0551	0.0097	−1.94	5.98
N_GI_1	2802	0.00	0.2234	0.0837	0.1047	0.0942	0.0940	0.0200	0.24	3.44
N_GI_7	2802	0.00	0.1221	0.0809	0.0961	0.0865	0.0908	0.0154	−1.82	4.66
N_GI_30	2802	0.00	0.1052	0.0822	0.0936	0.0851	0.0887	0.0135	−2.86	11.07
P_GI_1	2802	0.03	0.2868	0.1412	0.1640	0.1528	0.1522	0.0221	0.11	4.00
P_GI_7	2802	0.02	0.1834	0.1322	0.1543	0.1401	0.1481	0.0236	−2.10	5.57
P_GI_30	2802	0.00	0.1591	0.1326	0.1510	0.1378	0.1445	0.0216	−2.96	11.62
S_HE_1	2802	−0.04	0.0769	0.0031	0.0101	0.0067	0.0065	0.0075	0.53	9.37
S_HE_7	2802	−0.01	0.0185	0.0045	0.0080	0.0062	0.0064	0.0030	−0.28	1.43
S_HE_30	2802	0.00	0.0116	0.0053	0.0071	0.0061	0.0063	0.0017	−0.67	1.36
N_HE_1	2802	0.00	0.0531	0.0048	0.0097	0.0079	0.0072	0.0051	2.10	9.21
N_HE_7	2802	0.00	0.0172	0.0060	0.0084	0.0072	0.0073	0.0021	0.00	1.26
N_HE_30	2802	0.00	0.0106	0.0064	0.0080	0.0071	0.0073	0.0014	−1.43	4.86
P_HE_1	2802	0.00	0.0769	0.0107	0.0176	0.0146	0.0140	0.0068	1.63	8.12
P_HE_7	2802	0.00	0.0257	0.0117	0.0154	0.0134	0.0138	0.0033	−0.47	1.36
P_HE_30	2802	0.00	0.0183	0.0123	0.0147	0.0132	0.0136	0.0024	−1.90	6.37
S_LM_1	2802	−0.17	0.0694	−0.0418	−0.0222	−0.0328	−0.0322	0.0183	−0.57	4.37
S_LM_7	2802	−0.05	0.0044	−0.0348	−0.0263	−0.0301	−0.0309	0.0076	0.73	1.50
S_LM_30	2802	−0.04	0.0002	−0.0329	−0.0274	−0.0296	−0.0308	0.0053	2.02	6.51
N_LM_1	2802	0.00	0.1696	0.0423	0.0589	0.0512	0.0504	0.0156	0.93	4.16
N_LM_7	2802	0.00	0.0717	0.0430	0.0524	0.0470	0.0488	0.0089	−1.44	3.61
N_LM_30	2802	0.00	0.0564	0.0439	0.0507	0.0462	0.0482	0.0074	−2.73	10.62
P_LM_1	2802	0.00	0.0833	0.0146	0.0217	0.0184	0.0178	0.0071	1.38	6.72
P_LM_7	2802	0.00	0.0284	0.0150	0.0190	0.0169	0.0174	0.0036	−0.74	1.82
P_LM_30	2802	0.00	0.0219	0.0155	0.0183	0.0166	0.0172	0.0028	−2.34	8.70
RU_LM_1	2802	0.00	0.0552	0.0108	0.0172	0.0144	0.0139	0.0061	1.14	4.51
RU_LM_7	2802	0.00	0.0220	0.0117	0.0150	0.0132	0.0136	0.0029	−0.93	1.93
RU_LM_30	2802	0.00	0.0165	0.0121	0.0143	0.0130	0.0134	0.0021	−2.49	10.20
S_QDAP_1	2802	−0.12	0.1705	0.0189	0.0475	0.0330	0.0327	0.0270	0.04	2.41
S_QDAP_7	2802	−0.01	0.0716	0.0243	0.0364	0.0301	0.0306	0.0106	−0.19	0.86
S_QDAP_30	2802	0.00	0.0475	0.0267	0.0336	0.0296	0.0308	0.0066	−1.09	2.38
N_QDAP_1	2802	0.00	0.1760	0.0559	0.0771	0.0668	0.0663	0.0187	0.38	1.96
N_QDAP_7	2802	0.00	0.0932	0.0562	0.0689	0.0613	0.0642	0.0117	−1.52	3.62
N_QDAP_30	2802	0.00	0.0737	0.0578	0.0667	0.0604	0.0625	0.0098	−2.68	10.15
P_QDAP_1	2802	0.02	0.2174	0.0899	0.1088	0.0998	0.0991	0.0181	0.52	3.31
P_QDAP_7	2802	0.01	0.1222	0.0862	0.1010	0.0915	0.0962	0.0158	−1.91	4.89
P_QDAP_30	2802	0.00	0.1061	0.0865	0.0986	0.0900	0.0944	0.0142	−2.89	11.17

Note: BTC_G (Google daily BTC search intensity), WordC (News word count), NewsC (Daily news count), GI (Harvard psychological sentiment), HE (Henry's finance sentiment), LM (Loughran's and McDonald's finance sentiment), S (overall sentiment), P (purely positive Sentiment), N (purely Negative Sentiment), 1 (daily), 7 (weekly), and 30 (monthly).

Table 1.b
Summary Statistics on Lexicon-Based Sentiments (2012−2021).

Variable	Nobs	Min	Max	1st Qu	3rd Qu	Mean	Median	StDev	Skew	Kurt
syuzhet_1	2802	−28.10	67.55	2.50	9.70	6.28	5.95	6.98	0.83	5.34
syuzhet_7	2802	−8.99	17.64	3.85	7.55	5.71	5.74	3.05	−0.07	1.01
syuzhet_30	2802	−2.67	12.43	4.51	6.91	5.62	5.82	2.03	−0.50	0.93
bing_1	2802	−71.00	42.00	−4.60	2.33	−1.15	−1.00	6.86	−0.56	8.14
bing_7	2802	−20.24	9.86	−2.67	0.65	−1.11	−1.00	2.96	−0.83	4.01
bing_30	2802	−11.63	4.04	−2.23	0.17	−1.08	−0.95	1.99	−0.82	2.90
afinn_1	2802	−95.00	96.00	−5.50	10.98	2.40	3.33	15.45	−0.26	3.64
afinn_7	2802	−30.29	22.86	−1.56	6.71	2.18	2.74	6.80	−0.72	1.66
afinn_30	2802	−17.07	14.39	−0.78	5.43	2.13	2.95	4.57	−0.84	1.21
nrc_1	2802	−15.00	86.00	6.50	15.60	11.72	10.83	8.71	1.55	6.92
nrc_7	2802	−1.14	27.36	8.16	12.94	10.69	10.56	3.82	0.33	0.65
nrc_30	2802	0.40	18.65	9.17	12.07	10.51	10.67	2.48	−0.55	1.44
anger	2802	1.00	102.00	5.00	12.00	9.66	8.00	7.44	3.19	20.07
anticipation	2802	0.00	146.00	8.00	18.00	14.48	12.00	10.63	3.17	19.78
disgust	2802	0.00	69.00	1.00	6.00	4.19	3.00	4.63	3.97	32.11
fear	2802	0.00	135.00	5.00	14.00	11.17	9.00	10.13	3.37	22.33
joy	2802	0.00	107.00	4.00	11.00	8.59	7.00	7.59	3.27	21.54
sadness	2802	0.00	102.00	3.00	11.00	8.19	6.00	7.88	3.91	27.99
surprise	2802	0.00	71.00	2.00	8.00	5.60	4.00	5.11	3.20	20.72
trust	2802	0.00	227.00	11.00	26.00	20.62	17.00	15.65	3.51	24.29
negative	2802	1.00	231.00	10.00	25.75	20.29	17.00	16.96	3.78	26.28
positive	2802	0.00	358.00	17.00	40.00	31.90	27.00	25.14	3.73	26.52

Note: 1 (daily), 7 (weekly), and 30 (monthly).

Table 1.c
Summary Statistics on BTC OHLC and Different RV Estimates (2012–2021).

Variable	Nobs	Min	Max	1st Qu	3rd Qu	Mean	Median	StDev	Skew	Kurt
Open	3525	4.70	63,544.2000	441.1250	8922.0250	7694.7779	3434.0000	12,472.3915	2.59	6.32
High	3525	4.80	64,778.0000	450.8250	9189.2500	7930.1425	3490.4000	12,867.4114	2.57	6.22
Low	3525	4.50	62,067.5000	428.8500	8723.6750	7433.9823	3380.9500	12,025.7293	2.60	6.42
Close	3525	4.60	63,540.9000	441.1250	8915.5000	7711.4243	3440.0000	12,495.0922	2.58	6.28
Volume	3525	0.00	13,328,655	90,898.5	1,469,082.50	961,943.36	487,111.00	1,285,228.79	2.35	8.45
PK_RV1	3525	0.00	0.4367	0.0002	0.0019	0.0034	0.0007	0.0174	15.75	296.66
PK_RV7	3525	0.00	0.1888	0.0005	0.0023	0.0033	0.0011	0.0121	10.38	125.08
PK_RV30	3525	0.00	0.0745	0.0007	0.0027	0.0032	0.0012	0.0080	6.37	45.41
GK_RV1	3525	0.00	0.8706	0.0004	0.0033	0.0061	0.0012	0.0348	17.54	366.45
GK_RV7	3525	0.00	0.4833	0.0008	0.0040	0.0059	0.0019	0.0239	11.95	170.21
GK_RV30	3525	0.00	0.1580	0.0013	0.0047	0.0059	0.0022	0.0159	7.16	57.65
RS_RV1	3525	0.00	0.6599	0.0003	0.0020	0.0038	0.0007	0.0234	18.71	430.50
RS_RV7	3525	0.00	0.2161	0.0005	0.0024	0.0037	0.0010	0.0140	9.42	101.26
RS_RV30	3525	0.00	0.0832	0.0007	0.0030	0.0037	0.0013	0.0094	6.05	39.93

Note: PK (Parkinson), GK (Garman-Klass), RS (Rogers-Satchel), RV (realized variance), 1 (daily), 7 (weekly), and 30 (monthly).

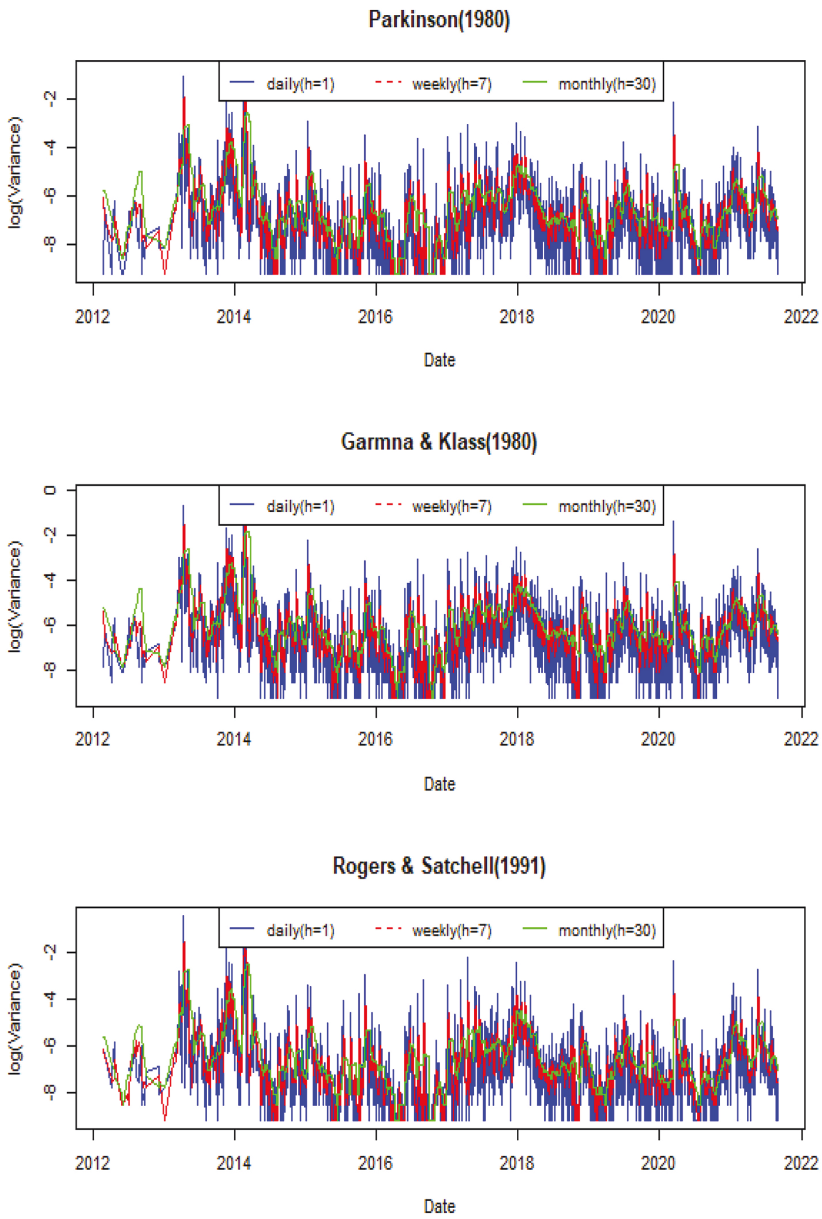


Fig. 3. Range-based volatility estimates, daily, weekly, and monthly (Jan 2012–Aug 2021).

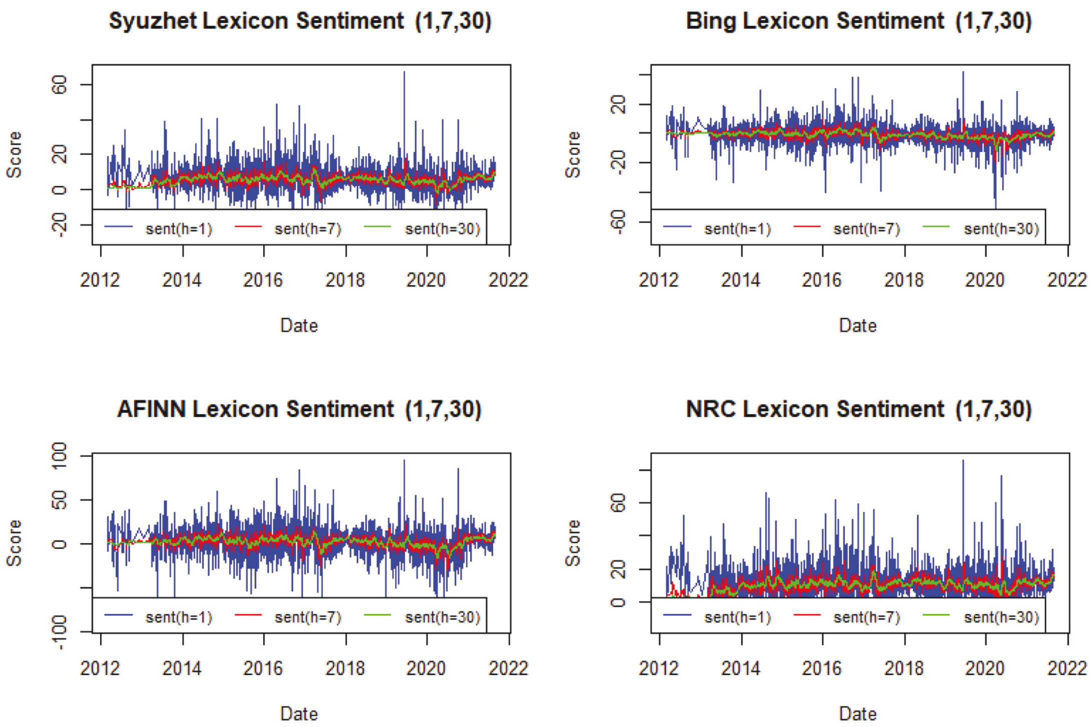


Fig. 4. Emotion Lexicon–based sentiments on BTC news, daily, weekly, and monthly (Jan 2012–Aug 2021).

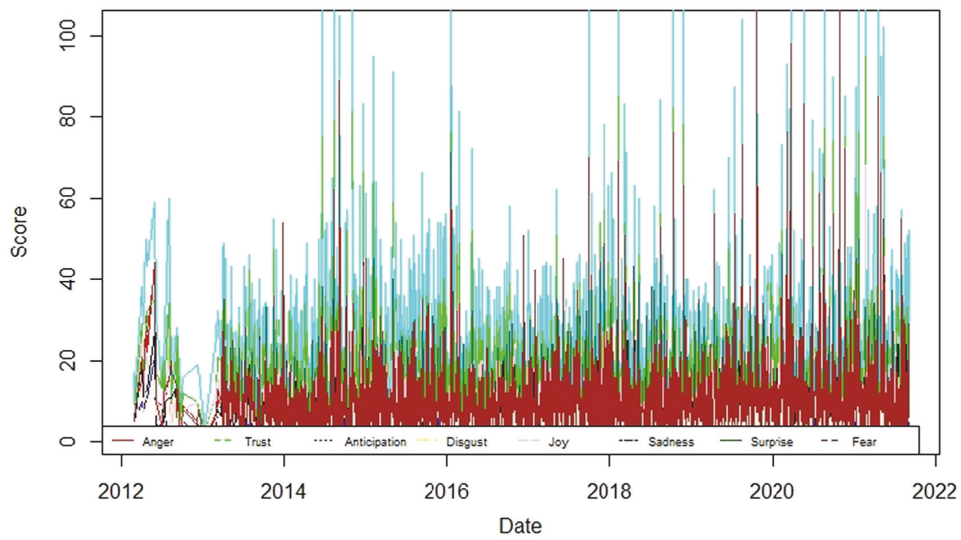


Fig. 5. NRC word-emotion association on global BTC news (Jan 2012–Aug 2021).

search trends, and various sentiments of different memory lengths, whereas Table 1.b includes the Emotion Lexicon–based summary statistics for 2802 days of news observations, totaling 17,490 news items. We can observe that there were on average 6 news pieces on BTC published daily within these 91 major newspapers, ranging from a minimum of 1 to a maximum of 66 daily news pieces. I aggregated the sentiment scores of a day if it had more than one news item.

4. Methodologies

4.1. *Estimating realized variance of Bitcoin*

Parkinson (1980) introduced the high/low range–based volatility estimation technique. Thereafter, new range-based volatility estimation methods emerged including opening and closing prices. These new range-based estimators use OHLC prices in an intraday setting. By including different methods, we can gain a better understanding of the nature of ranges and their significance in forecasting future volatilities. Volatility plays a central role in many areas of finance, and price range provides an intuitive and efficient estimator of volatility (Chou, Chou, & Liu, 2010). Including various range-based estimation methods along with different sentiment dictionaries contributes to a comparative analysis for deciding the most suitable volatility estimation method with the right sentiment dictionary.

Utilizing the BTC intraday OHLC data from investing.com, I created three different, daily, range-based, BTC RV series applying the methods of Parkinson (1980), Garman and Klass (1980), and Rogers and Satchell (1991). Parkinson (1980) introduced a volatility measure that uses the high and low prices of the day instead of only a closing price, considering that large price movements could have happened during the day itself. Thus, Parkinson's volatility is considered to be more precise than the regular close-by-close volatility estimation.

$$\sigma^2_{PK} = \frac{1}{4ln(2)n}\sum_{i=1}^{n} ln^2\left(\frac{H_i}{L_i}\right) \tag{1}$$

where σ^2_{PK} is the Parkinson (1980) variance estimator, and H_i is the highest and L_i the lowest intraday price of asset *i*.

However, it does not consider price movements after market close, systematically undervaluing volatility. The Garman and Klass (1980) volatility estimator overcomes this drawback by incorporating OHLC prices of a security. Considering market swings during the opening and closing hours makes volatility estimation more accurate.

$$\sigma^2_{GK} = \frac{1}{n}\left(\sum_{i=1}^{n}\frac{1}{2}ln^2\left(\frac{H_i}{L_i}\right) + (2ln(2)-1)ln^2\left(\frac{C_i}{O_i}\right)\right) \tag{2}$$

where σ^2_{GK} is the Garman and Klass (1980) variance estimator, H_i is the highest and L_i the lowest intraday price, O_i is the opening and C_i the closing price of asset *i*.

One criticism of the Garman-Klass method is that it is not robust for opening jumps in price and trend movements. Nevertheless, it is still more effective than the regular close-by-close volatility estimation because it considers not only the opening and closing prices but also intraday price extrema. Rogers and Satchell (1991) proposed a more efficient method for assessing historical volatility that takes into account price trends.

$$\sigma^2_{RS} = \frac{1}{n}\sum_{i=1}^{n}(u_i(u_i-c_i)+d_i(d_i-c_i)) \tag{3}$$

where σ^2_{RS} is the Rogers and Satchell (1991) variance estimator, u_i is the normalized high and d_i the normalized low, and c_i is the normalized closing price of asset *i*.

The Rogers-Satchell method incorporates the drift term; as a result, it provides a better volatility estimation when the underlying is trending. These three range-based estimation methods were applied to the BTC data sample from January 2012 until August 2021, revealing a close-by-close volatility of 3.37%, Parkinson (1980) volatility of 5.56%, Garman and Klass (1980) volatility of 7.44%, and Rogers and Satchell (1991) volatility of 5.59%. All three range-based estimation methods showed higher volatilities than the regular close-by-close technique. Table 1.c presents the summary statistics of the BTC OHLC data, as well as a summary of the variance series of the three different range-based variance estimators with three different memory lengths.

Table 2
Estimation of Baseline HAR-RV with Range-Based RVs.

	Dependent variable: $\log(RV_{t+1})$		
	Parkinson (PK)	Garman-Klass (GK)	Rogers-Satchel (RS)
log(PK_RV$_1$)	0.343***		
	(15.729)		
log(PK_RV$_7$)	0.384***		
	(11.795)		
log(PK_RV$_{30}$)	0.146***		
	(4.862)		
log(GK_RV$_1$)		0.299***	
		(13.632)	
log(GK_RV$_7$)		0.416***	
		(12.408)	
log(GK_RV$_{30}$)		0.158***	
		(5.043)	
log(RS_RV$_1$)			0.285***
			(18.626)
log(RS_RV$_7$)			0.306***
			(9.945)
log(RS_RV$_{30}$)			0.145***
			(4.997)
Constant	−0.512***	−0.506***	−0.611***
	(−9.159)	(−9.304)	(−10.998)
Observations	2802	2802	2802
Adjusted R^2	0.464	0.464	0.469
F Statistic (df = 3; 2798)	808.729***	809.825***	825.827***

Notes: This table reports the estimates for the daily HAR-RV models. The estimation period spans from 1 January 2012 to 31 August 2021. Three different RVs are considered in this empirical analysis. The basic HAR-RV is presented in Eq. (8). T-stats are reported in the parenthesis.
***Significant at the 1% level.
**Significant at the 5% level.
*Significant at the 10% level.

4.2. HAR-RV for forecasting Bitcoin volatility

The HAR-RV proposed by Corsi (2009) is one of the most popular models for forecasting volatility. Recently, HAR-type models have received considerable attention in academic research. The HAR approach separates RVs into short-term, medium-term, and long-term volatility components. Previous studies (for example, Andersen et al., 2007; Ma, Wei, Huang, & Chen, 2014) have found that the HAR-RV process outperforms other approaches when forecasting future RV. Considering its outperformance in forecasting future RV, the basic HAR-RV has been extended in several other dimensions. First, the definition of RV for day *t* is:

$$RV_t = \sum_{j=1}^{M} r^2_{t,j};\ t = 1, 2, \ldots, T \tag{4}$$

where $r_{t,\,j}$ is the logarithmic return for period *j* of day *t*, M indicates the number of intraday observations at time *t*, and *T* refers to the number of periods in the sample.

$$RV^{(d)}_t = RV^{(X)}_t \tag{5}$$

where $RV^{(X)}_t$ can refer to any measure of volatility.

The original HAR model was proposed to model daily RV with 5 days in a week and 22 days in a month. However, the BTC market is 24 h a day and 7 days a week. Therefore, the weekly and monthly RVs are aggregated as:

$$RV^{(w)}_t = \frac{1}{7}\sum_{h=0}^{6} RV^{(d)}_{t-h} \tag{6}$$

where $RV^{(w)}_t$ is weekly RV.

$$RV^{(m)}_t = \frac{1}{30}\sum_{h=0}^{29} RV^{(d)}_{t-h} \tag{7}$$

where $RV^{(m)}_t$ is monthly RV.

Therefore, the HAR-RV for BTC can be written as:

$$HAR\_RV_{t,t+1} = \beta_0 + \beta^d RV^{(d)}_t + \beta^w RV^{(w)}_t + \beta^m RV^{(m)}_t + \varepsilon_t \tag{8}$$

Table 3
Estimation of HAR-RV with Range-Based Volatilities and Overall Psychological & Discourse Sentiment.

	Dependent variable: $log(RV_{t+1})$					
	PK (1980)		GK (1980)		RS (1991)	
	(GI)	(QDAP)	(GI)	(QDAP)	(GI)	(QDAP)
logPK_RV1	0.326***	0.325***				
	(14.861)	(14.782)				
logPK_RV7	0.371***	0.373***				
	(11.404)	(11.488)				
logPK_RV30	0.148***	0.147***				
	(4.920)	(4.880)				
logGK_RV1			0.300***	0.299***		
			(14.427)	(14.355)		
logGK_RV7			0.385***	0.387***		
			(12.219)	(12.299)		
logGK_RV30			0.155***	0.155***		
			(5.256)	(5.221)		
logRS_RV1					0.278***	0.275***
					(13.161)	(13.009)
logRS_RV7					0.368***	0.371***
					(11.790)	(11.883)
logRS_RV30					0.154***	0.154***
					(5.201)	(5.171)
S_GI_1	−0.099		−0.090		0.011	
	(−0.312)		(−0.267)		(0.035)	
S_GI_7	2.371***		2.381**		2.307**	
	(2.653)		(2.510)		(2.519)	
S_GI_30	−0.943		−1.216		−1.534	
	(−0.785)		(−0.954)		(−1.245)	
S_QDAP_1		−0.064		−0.046		0.003
		(−0.176)		(−0.119)		(0.008)
S_QDAP_7		2.851***		2.854**		2.930***
		(2.590)		(2.442)		(2.597)
S_QDAP_30		−1.902		−2.251		−2.560
		(−1.151)		(−1.283)		(−1.508)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.644***	−0.619***	−0.845***	−0.823***	−0.894***	−0.844***
	(−8.258)	(−8.600)	(−11.329)	(−12.085)	(−11.414)	(−11.683)
Observations	2802	2802	2802	2802	2802	2802
Adjusted R^2	0.498	0.498	0.499	0.499	0.456	0.456
F Statistic (df = 8; 2793)	348.569***	348.740***	349.406***	349.675***	295.042***	294.731***

Notes: This table reports the estimates for the HAR-RV-SI models. The estimation period spans from 1 January 2012 to 31 August 2021. Three different RVs along with two overall psychological sentiments are considered in this empirical analysis. The extended HAR-RV-SI is presented in Eq. (9). T-stats are reported in the parenthesis.
***Significant at the 1% level.
**Significant at the 5% level.
*Significant at the 10% level.

where d is the daily, w is the weekly, and m is the monthly horizon.

4.3. HAR-RV-SI: Adding news sentiment in Bitcoin volatility modeling

In this study, I extended the baseline HAR-RV model by adding news sentiments and other control variables, called HAR-RV-SI. I used the logarithm of the RV for getting the time series of daily volatilities. Furthermore, weekly and monthly variances were also calculated in a rolling window fashion, and to get the respective volatilities, I also log-transformed the weekly and monthly series. The time-series plot of RVs of BTC separated into short-term, medium-term, and long-term volatility components following different estimation methods is presented in Fig. 3. Similarly, to capture the effect of sentiment variables at each frequency, in addition to the daily sentiment indices, I also derived the weekly and monthly sentiment indices of BTC-related news for all four dictionaries, which can be observed in Fig. 2.

Moreover, to test the order of integration, I followed the common literature and employed the well-known augmented Dickey-Fuller (ADF) unit root test (Dickey & Fuller, 1979). All the right-hand and left-hand variables including log-transformed variance and sentiments at different time scales, including control variables, showed stationarity.

The log-transformed realized variance (log-RV), log-HAR-RV, then, according to Corsi (2009), can be specified as:

$$\begin{aligned} logHAR\_RV\_SI_{t,t+1} = & \alpha + \beta^d logRV_t^{(d)} + \beta^w logRV_t^{(w)} + \beta^m logRV_t^{(m)} \\ & + \delta^d SI_t^{(d)} \\ & + \delta^w SI_t^{(w)} + \delta^m SI_t^{(m)} + \gamma^d X_t^{(d)} + \varepsilon_t \end{aligned} \tag{9}$$

where $SI_t^{(w)}$ is the weekly sentiment index for each sentiment measure.

$$SI_t^{(w)} = \frac{1}{7}\sum_{h=0}^{6} SI_{t-h}^{(d)} \tag{9.a}$$

where $SI_t^{(m)}$ is the monthly sentiment index for each sentiment measure.

$$SI_t^{(m)} = \frac{1}{30}\sum_{h=0}^{29} SI_{t-h}^{(d)} \tag{9.b}$$

where $\boldsymbol{X}_t^{(d)}$ are the control variables, daily Google search intensity, and daily news count.

4.4. HAR-RV-PS/NS: Decomposing sentiment into purely positive and purely negative sentiments

I also extended the benchmark HAR-RV in several other dimensions. Specifically, I extended the benchmark HAR-RV to feature a measure of

Table 4
Estimation of HAR-RV with Range-Based Volatilities and Overall Financial Sentiment.

	Dependent variable: $log(RV_{t+1})$					
	PK (1980)		GK (1980)		RS (1991)	
	(HE)	(LM)	(HE)	(LM)	(HE)	(LM)
logPK_RV1	0.328***	0.326***				
	(14.923)	(14.799)				
logPK_RV7	0.375***	0.378***				
	(11.527)	(11.602)				
logPK_RV30	0.142***	0.143***				
	(4.732)	(4.759)				
logGK_RV1			0.302***	0.299***		
			(14.498)	(14.373)		
logGK_RV7			0.389***	0.392***		
			(12.337)	(12.406)		
logGK_RV30			0.150***	0.151***		
			(5.076)	(5.111)		
logRS_RV1					0.278***	0.277***
					(13.163)	(13.053)
logRS_RV7					0.373***	0.377***
					(11.930)	(12.058)
logRS_RV30					0.147***	0.146***
					(4.951)	(4.943)
S_HE_1	1.112		1.135		0.812	
	(0.844)		(0.862)		(0.592)	
S_HE_7	1.672		1.881		1.633	
	(0.432)		(0.487)		(0.406)	
S_HE_30	−1.155		−2.105		5.580	
	(−0.180)		(−0.328)		(0.834)	
S_LM_1		0.242		0.154		0.299
		(0.457)		(0.292)		(0.544)
S_LM_7		2.894*		2.962*		2.987*
		(1.821)		(1.872)		(1.811)
S_LM_30		−4.315**		−1.151		−4.571*
		(−1.999)		(−0.516)		(−1.959)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.648***	−0.588***	−0.858***	−0.791***	−0.860***	−0.835***
	(−9.177)	(−7.336)	(−12.949)	(−10.328)	(−12.108)	(−10.301)
Observations	2802	2802	2802	2802	2802	2802
Adjusted R^2	0.497	0.498	0.498	0.498	0.455	0.455
F Statistic (df = 8; 2793)	347.028***	347.863***	347.823***	348.711***	293.061***	293.784***

Notes: This table reports the estimates for the HAR-RV-SI models. The estimation period spans from 1 January 2012 to 31 August 2021. Three different RVs along with two overall financial sentiments are considered in this empirical analysis. The extended HAR-RV-SI is presented in Eq. (9). T-stats are reported in the parenthesis.
***Significant at the 1% level.
**Significant at the 5% level.
*Significant at the 10% level.

positive and negative sentiments, called HAR-RV-PS/NS. To capture the heterogeneity between the optimistic and pessimistic investors, I decomposed sentiment into purely positive and purely negative. The general idea of this decomposition is that optimistic investors are mainly guided by positive sentiments in the news, whereas pessimistic investors are mainly guided by negative sentiments.

4.4.1. Positive sentiment

$$logHAR\_RV\_PS_{t,t+1} = \alpha + \beta^{d} logRV_t^{(d)} + \beta^{w} logRV_t^{(w)} + \beta^{m} logRV_t^{(m)} + \delta^{d} PS_t^{(d)} + \delta^{w} PS_t^{(w)} + \delta^{m} PS_t^{(m)} + \gamma^{d} X_t^{(d)} + \varepsilon_t \tag{10}$$

where $PS_t^{(d)}$ is the daily positive sentiment, and weekly (w) and monthly (m) positive sentiments are captured via the following equations:

$$PS_t^{(w)} = \frac{1}{7}\sum_{h=0}^{6} PS_{t-h}^{(d)} \tag{10.a}$$

$$PS_t^{(m)} = \frac{1}{30}\sum_{h=0}^{29} PS_{t-h}^{(d)} \tag{10.b}$$

4.4.2. Negative sentiment

$$logHAR\_RV\_NS_{t,t+1} = \alpha + \beta^{d} logRV_t^{(d)} + \beta^{w} logRV_t^{(w)} + \beta^{m} logRV_t^{(m)} + \delta^{d} NS_t^{(d)} + \delta^{w} NS_t^{(w)} + \delta^{m} NS_t^{(m)} + \gamma^{d} X_t^{(d)} + \varepsilon_t \tag{11}$$

where $NS_t^{(d)}$ is the daily negative sentiment, and weekly (w) and monthly (m) negative sentiments are captured via the following equations:

$$NS_t^{(w)} = \frac{1}{7}\sum_{h=0}^{6} NS_{t-h}^{(d)} \tag{11.a}$$

$$NS_t^{(m)} = \frac{1}{30}\sum_{h=0}^{29} NS_{t-h}^{(d)} \tag{11.b}$$

4.5. Emotion lexicon sentiment of the news and Bitcoin volatility: Robustness check

The NRC Emotion Lexicon is a list of 5636 English words and their associations, with 8 basic emotions—anger, fear, anticipation, trust, surprise, sadness, joy, and disgust—and 2 sentiments—negative and positive (Mohammad & Turney, 2013). To further explore the sentiments of different human emotions, I followed the NRC Emotion Lexicon. Fig. 4 shows the polarity and time-varying sentiments of BTC news items with different emotions. Fig. 5 shows a histogram of the

Table 5
Estimation of HAR-RV with Range-Based Volatilities and Positive Psychological Sentiments.

	Dependent variable: $log(RV_{t+1})$					
	PK (1980)		GK (1980)		RS (1991)	
	(GI)	(QDAP)	(GI)	(QDAP)	(GI)	(QDAP)
logPK_RV1	0.328***	0.327***				
	(14.911)	(14.892)				
logPK_RV7	0.371***	0.371***				
	(11.353)	(11.367)				
logPK_RV30	0.145***	0.145***				
	(4.811)	(4.789)				
logGK_RV1			0.285***	0.285***		
			(12.903)	(12.891)		
logGK_RV7			0.401***	0.402***		
			(11.928)	(11.941)		
logGK_RV30			0.157***	0.156***		
			(4.976)	(4.953)		
logRS_RV1					0.366***	0.365***
					(17.526)	(17.472)
logRS_RV7					0.292***	0.293***
					(9.458)	(9.480)
logRS_RV30					0.146***	0.146***
					(4.996)	(4.983)
P_GI_1	−0.050		−0.115		0.318	
	(−0.121)		(−0.260)		(0.746)	
P_GI_7	0.629		0.632		0.636	
	(1.065		(1.008		(1.052	
P_GI_30	−0.796		−0.938		−1.442**	
	(−1.206)		(−1.339)		(−2.125)	
P_QDAP_1		−0.064		−0.100		0.362
		(−0.125)		(−0.184)		(0.692)
P_QDAP_7		0.746		0.755		0.804
		(0.848)		(0.809)		(0.894)
P_QDAP_30		−1.176		−1.380		−2.221**
		(−1.178)		(−1.303)		(−2.166)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.614***	−0.600***	−0.576***	−0.570***	−0.697***	−0.670***
	(−5.952)	(−6.303)	(−5.379)	(−5.794)	(−6.748)	(−7.036)
Observations	2802	2802	2802	2802	2802	2802
Adjusted R^2	0.497	0.497	0.467	0.467	0.474	0.474
F Statistic (df = 8; 2793)	347.062***	347.026***	308.321***	308.305***	316.986***	317.082***

Notes: This table reports the estimates for the HAR-RV-PS models. The estimation period spans from 1 January 2012 to 31 August 2021. Three different RVs along with two positive psychological sentiments are considered. The extended HAR-RV-PS is presented in Eq. (10). T-stats are reported in the parenthesis.
***Significant at the 1% level.
**Significant at the 5% level.
*Significant at the 10% level.

corresponding eight basic human emotions in the news data sample. We can observe from the graph that trust and fear were the two emotions most triggered by BTC-related news in the time period.

Next, I extended the benchmark HAR-RV model to feature different human emotions from the lexicon index beyond positive and negative sentiments. This extended HAR-RV-LI with all eight daily normalized emotions, along with Google search trends and news counts as controls, is calculated as follows:

$$\begin{aligned} logHAR\_RV\_LI_{t,t+1} = & \alpha + \beta^{d} logRV_t^{(d)} + \beta^{w} logRV_t^{(w)} + \beta^{m} logRV_t^{(m)} \\ & + \delta^{d1} nAng_t^{(d)} + \delta^{d2} nAnt_t^{(d)} + \delta^{d3} nDis_t^{(d)} + \delta^{d4} nFear_t^{(d)} \\ & + \delta^{d5} nJoy_t^{(d)} + \delta^{d6} nSad_t^{(d)} + \delta^{d7} nSur_t^{(d)} + \delta^{d8} nTru_t^{(d)} \\ & + \gamma^{d} X_t^{(d)} + \varepsilon_t \end{aligned} \tag{12}$$

where δ^{d1} to δ^{d8} are the eight different normalized emotions extracted implementing the NRC Emotion Lexicon. The eight emotions are anger, anticipation, disgust, fear, joy, sadness, surprise, and trust, as mentioned above.

5. Results

5.1. Summary statistics

In the summary statistics in Table 1.a, the Google search trends show a minimum value of 0 and a maximum value of 100. The lowest search intensity days during the whole sample period of January 1, 2012, to August 31, 2021, score 0, and the highest search intensity days score 100. Because the result is normalized between 0 and 100, the search data series shows stationarity at the given level. The mean Google search score of 7.61 shows that BTC was not intensively searched on Google daily during the time period. During the sample period, there were a total of 17,490 news items which also includes the news articles on BTC published by the major English language newspapers from around the globe. The total days within the sample period was 3525; however, there was at least one news item on BTC by at least one of the major English-language newspapers for only 2802 days.

I downloaded the BTC Google search intensity and OHLC data for the full sample period and matched to the respective sentiment days. On average, there were 6 news pieces on BTC daily, ranging from a minimum of 1 news piece to a maximum of 66 news pieces in a day. One newspaper might have had more than one news item on BTC on a particular day. In other words, all 66 news items were not published by

Table 6
Estimation of HAR-RV with Range-Based Volatilities and Negative Psychological Sentiments.

	Dependent variable: log(RV_{t+1})					
	PK (1980)		GK (1980)		RS (1991)	
	(GI)	(QDAP)	(GI)	(QDAP)	(GI)	(QDAP)
logPK_RV1	0.326***	0.326***				
	(14.815)	(14.798)				
logPK_RV7	0.378***	0.380***				
	(11.567)	(11.630)				
logPK_RV30	0.141***	0.140***				
	(4.667)	(4.644)				
logGK_RV1			0.284***	0.284***		
			(12.819)	(12.814)		
logGK_RV7			0.408***	0.410***		
			(12.131)	(12.182)		
logGK_RV30			0.152***	0.152***		
			(4.841)	(4.825)		
logRS_RV1					0.364***	0.363***
					(17.368)	(17.356)
logRS_RV7					0.298***	0.299***
					(9.641)	(9.694)
logRS_RV30					0.143***	0.143***
					(4.904)	(4.915)
N_GI_1	−0.299		−0.381		−0.024	
	(−0.640)		(−0.768)		(−0.050)	
N_GI_7	−0.890		−0.885		−0.834	
	(−1.014)		(−0.950)		(−0.928)	
N_GI_30	0.122		−0.070		−1.099	
	(0.120)		(−0.065)		(−1.052)	
N_QDAP_1		−0.331		−0.378		0.005
		(−0.656)		(−0.706)		(0.009)
N_QDAP_7		−1.790		−1.771		−1.821
		(−1.581)		(−1.474)		(−1.570)
N_QDAP_30		0.859		0.617		−0.754
		(0.631)		(0.427)		(−0.540)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.547***	−0.558***	−0.514***	−0.532***	−0.593***	−0.597***
	(−5.934)	(−6.375)	(−5.420)	(−5.939)	(−6.432)	(−6.829)
Observations	2802	2802	2802	2802	2802	2802
Adjusted R^2	0.497	0.498	0.468	0.468	0.474	0.475
F Statistic (df = 8; 2793)	347.315***	347.723***	308.607***	308.877***	317.145***	317.354***

Notes: This table reports the estimates for the HAR-RV-NS models. The estimation period spans from 1 January 2012 to 31 August 2021. Three different RVs along with two negative psychological sentiments are considered. The extended HAR-RV-NS is presented in Eq. (11). T-stats are reported in the parenthesis.
***Significant at the 1% level.
**Significant at the 5% level.
*Significant at the 10% level.

66 unique newspapers in a single day. Another interesting summary statistic is that the Harvard-IV psychological dictionary, QDAP discourse dictionary, and Henry's finance-specific dictionary showed similar patterns, whereas Loughran's and McDonald's finance-specific dictionary was overall negative on daily, weekly, and monthly aggregates. If we remove Henry's finance-specific dictionary, one can observe that the psychological and discourse sentiments from the BTC-related news were overall positive and that the finance sentiments were overall negative during the full sample period.

Similarly, in the summary statistics shown in Table 1.b, we can observe the basic statistics of four different lexicon sentiments. The Syuzhet R package with the "get_sentiment" function gives scores for *Syuzhet, Bing, Afinn*, and *NRC* sentiments. In this study, the NRC Emotion Lexicon had a higher mean score for positive sentiments than negative sentiments. Dividing the sentiments into eight different human emotions showed that the news articles during the sample period, on average, triggered mostly trust, anticipation and fear in its readers. Furthermore, on average, news during the sample period equally conveyed the emotions of joy and sadness to the public.

Table 1.c includes statistics summarizing BTC OHLC and three range-based variances on daily, weekly, and monthly averages calculated by a rolling window method. BTC closing price ranged from a minimum of 4.6 dollars to a maximum of 63,540.90 dollars, which is 13,813 times higher than its minimum value. The daily, average, range-based variance following Parkinson (1980) was 0.0034, Garman and Klass (1980) was 0.0061, and Rogers and Satchell (1991) was 0.0038. The weekly and monthly averages following each range-based method showed no vast differences in the average variance.

5.2. *Basic fitting of the HAR-RV*

The first step in the analysis was to fit the basic HAR(3) to compare if sentiments, as additional explanatory variables, improve the model fitting or not. The HAR(3) model utilizes three AR(1) volatility processes at daily, weekly, and monthly windows. As a natural economic interpretation of this model according to Corsi (2009), each component in the model corresponds to a short-term, medium-term and long-term volatilities. The baseline model–fitting results presented in Table 2 show

Table 7
Estimation of HAR-RV with Range-Based Volatilities and Positive Financial Sentiments.

	Dependent variable: log(RV_{t+1})					
	PK (1980)		GK (1980)		RS (1991)	
	(HE)	(LM)	(HE)	(LM)	(HE)	(LM)
logPK_RV1	0.327***	0.327***				
	(14.877)	(14.876)				
logPK_RV7	0.373***	0.370***				
	(11.459)	(11.348)				
logPK_RV30	0.144***	0.148***				
	(4.764)	(4.888)				
logGK_RV1			0.285***	0.285***		
			(12.882)	(12.873)		
logGK_RV7			0.403***	0.400***		
			(12.031)	(11.929)		
logGK_RV30			0.155***	0.159***		
			(4.934)	(5.054)		
logRS_RV1					0.365***	0.366***
					(17.486)	(17.532)
logRS_RV7					0.293***	0.291***
					(9.510)	(9.425)
logRS_RV30					0.145***	0.148***
					(4.987)	(5.082)
P_HE_1	1.168		1.317		1.316	
	(0.821)		(0.872)		(0.904)	
P_HE_7	0.921		0.788		2.099	
	(0.246)		(0.199)		(0.548)	
P_HE_30	−5.361		−6.300		−11.589**	
	(−1.038)		(−1.150)		(−2.181)	
P_LM_1		−0.515		−0.710		0.619
		(−0.380)		(−0.493)		(0.446)
P_LM_7		6.389*		6.718*		6.280*
		(1.838)		(1.821)		(1.764)
P_LM_30		−8.030*		−9.343*		−12.020***
		(−1.776)		(−1.947)		(−2.584)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.604***	−0.607***	−0.582***	−0.577***	−0.656***	−0.673***
	(−7.642)	(−7.197)	(−7.258)	(−6.694)	(−8.366)	(−8.047)
Observations	2802	2802	2802	2802	2802	2802
Adjusted R^2	0.497	0.498	0.468	0.468	0.475	0.475
F Statistic (df = 8; 2793)	347.187***	347.686***	308.478***	308.926***	317.146***	317.398***

Notes: This table reports the estimates for the HAR-RV models. The estimation period spans from 1 January 2012 to 31 August 2021. Three different RVs along with two positive financial sentiments are considered. The extended HAR-RV-PS is presented in Eq. (10). T-stats are reported in the parenthesis.
***Significant at the 1% level.
**Significant at the 5% level.
*Significant at the 10% level.

similar results to those of Corsi (2009, page 187).

Table 2 reports the results of the estimation of the basic HAR-RV for three range-based volatility series. T-statistics confirmed all three RVs aggregated over the three different horizons to be highly significant. This result is in line with the results of Aalborg et al. (2019) where they found that the past RV of BTC predicts its future RV on the HAR-RV setup. One surprising finding in the current study is that RV aggregated weekly seemed to be less noisy and received more weight compared with RV aggregated daily and monthly, as in the cases of Parkinson (1980), Garman and Klass (1980), and Rogers and Satchell (1991). According to Corsi (2009), weekly and monthly RVs averaged over longer periods contain less noise and more information on the volatility process and, hence, receive higher weight from the model. However, in the table, the range-based volatilities seem to lose information or memory over a longer time period.

5.3. Extension of the HAR-RV with overall psychological and financial sentiments

While the daily, weekly, and monthly volatilities remained equally significant, as shown in the baseline HAR model presented in Table 2, we can observe an R-squared in Table 3 showing that after adding psychological sentiment, the quality of the extended HAR-RV was improved. A similar weight pattern on different scalings of sentiment can be observed. The weekly aggregates had more weight compared to daily and monthly averages. Another interesting finding is that only weekly aggregated psychological and discourse sentiments extracted from the news had a statistically significant impact on BTC volatility. For all the range-based estimators, neither daily nor monthly sentiments were significant. One possible reason behind this result might be the arrival of news to potential investors or readers. Not all audiences read newspapers on the same day they are published. Furthermore, general readers easily tend to forget the news over the long run, resulting in the decay of sentiment generated from the news within the month. Moreover, the extended model also accounted for Google search intensity and news counts as controls.

Similarly, in Table 4, results on the two separate financial sentiment dictionaries, along with three different range-based volatilities, are reported. Two dictionaries, those of Henry and Loughran and McDonald, are specifically targeted to the domain of finance. Loughran and McDonald (2011) used Harvard-IV and Henry (2008) used earnings press releases to capture tone. Both found finance-specific word lists to be more powerful than general word lists. However, in this study, Henry's finance-specific dictionary did not seem to show any significance in any time length for any of the range-based volatility estimators. As opposed to psychological and discourse sentiments, Loughran's and

Table 8
Estimation of HAR-RV with Range-Based Volatilities and Negative Financial Sentiments.

	Dependent variable: $log(RV_{t+1})$					
	PK (1980)		GK (1980)		RS (1991)	
	(HE)	(LM)	(HE)	(LM)	(HE)	(LM)
logPK_RV1	0.325***	0.326***				
	(14.752)	(14.833)				
logPK_RV7	0.373***	0.377***				
	(11.392)	(11.562)				
logPK_RV30	0.141***	0.142***				
	(4.644)	(4.719)				
logGK_RV1			0.283***	0.284***		
			(12.765)	(12.839)		
logGK_RV7			0.403***	0.407***		
			(11.945)	(12.123)		
logGK_RV30			0.152***	0.154***		
			(4.826)	(4.895)		
logRS_RV1					0.359***	0.364***
					(17.163)	(17.413)
logRS_RV7					0.293***	0.296***
					(9.475)	(9.610)
logRS_RV30					0.142***	0.146***
					(4.857)	(5.018)
N_HE_1	−0.206		0.079		−1.484	
	(−0.106)		(0.039)		(−0.750)	
N_HE_7	−1.574		−1.954		2.042	
	(−0.273)		(−0.320)		(0.347)	
N_HE_30	−14.557		−15.981*		−31.128***	
	(−1.643)		(−1.701)		(−3.413)	
N_LM_1		−0.507		−0.504		−0.615
		(−0.840)		(−0.787)		(−0.995)
N_LM_7		−1.453		−1.522		−1.243
		(−0.985)		(−0.973)		(−0.824)
N_LM_30		0.346		0.095		−1.992
		(0.193)		(0.050)		(−1.085)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.549***	−0.559***	−0.528***	−0.535***	−0.580***	−0.571***
	(−7.403)	(−6.340)	(−7.042)	(−5.921)	(−7.891)	(−6.492)
Observations	2802	2802	2802	2802	2802	2802
Adjusted R^2	0.498	0.497	0.468	0.468	0.476	0.475
F Statistic (df = 8; 2793)	347.900***	347.353***	309.130***	308.561***	319.438***	317.335***

Notes: This table reports the estimates for the HAR-RV-NS models. The estimation period spans from 1 January 2012 to 31 August 2021. Three different RVs along with two negative financial sentiments are considered. The extended HAR-RV-NS is presented in Eq. (11). T-stats are reported in the parenthesis.
***Significant at the 1% level.
**Significant at the 5% level.
*Significant at the 10% level.

McDonald's overall finance sentiments significantly impacted weekly and monthly volatilities. As rationale for the insignificance of daily financial sentiments, we can again use the argument of news arrival delay in the context of psychological and discourse sentiments. However, it is surprising that the effect of financial sentiments was significant over a longer time period in comparison to the other two psychological and discourse dictionaries.

5.4. Decomposing overall sentiments into positive and negative sentiments

Overall sentiment is a combination of positive and negative sentiments. However, some readers might be influenced by either of these emotions. Pessimistic and optimistic readers have different choices and perceptions of events. Pessimistic readers are mostly influenced by negative events, whereas optimistic are influenced by positive events. In this regard, McAfee, Doubleday, Geiger, and Connell (2019) stated that optimism and pessimism inform our expectation that events will turn out positively or negatively. Therefore, I further decomposed the overall sentiment into positive and negative sentiments and extended the basic HAR-RV to see which polarity of emotion is more responsible for BTC volatility.

Table 5 shows the results of HAR-RV with positive psychological and quantitative discourse sentiments. Table 6 presents the results of negative sentiments from the same dictionaries. Sentiments being fragmented into only negative or only positive showed BTC market volatility to be subject to a mixture of negative and positive sentiments rather than a purely negative or purely positive sentiment. Comparing this result with the findings of Corbet et al. (2020) who constructed a sentiment index based on news surrounding macroeconomic indicators found that negative news related to these indicators is positive for BTC and vice-versa. Nevertheless, the result is not fully comparable as the sentiment generated by their study is based on the news surrounding macroeconomic announcements whereas the sentiment index generated in the current study is fully based on the news specific to BTC. On the other hand, Entrop et al. (2020) used the news-based BTC sentiment data from Thomson Reuters MarketPsych (TRMI) to study the dynamic relation between bitcoin spot and futures prices and found that higher news-based BTC sentiment increases the informational role of the BTC futures market. Furthermore, they found news-based BTC sentiment to be a relevant measure of BTC price discovery, which is in line with this current research as we can observe in Table 3 and Table 4 that the extended HAR-RV model with news-based BTC sentiment has improved

Table 9
Estimation of HAR-RV with Range-based Volatilities and NRC Emotion Lexicon.

	Dependent variable: $log(RV_{t+1})$		
	(PK)	(GK)	(RS)
logPK_RV1	0.325***		
	(14.809)		
logPK_RV7	0.377***		
	(11.597)		
logPK_RV30	0.144***		
	(4.802)		
logGK_RV1		0.284***	
		(12.840)	
logGK_RV7		0.406***	
		(12.130)	
logGK_RV30		0.156***	
		(4.983)	
logRS_RV1			0.367***
			(17.713)
logRS_RV7			0.296***
			(9.649)
logRS_RV30			0.146***
			(5.010)
NormalizedAnger	0.473	0.411	0.789**
	(1.213)	(0.993)	(1.976)
NormalizedAnticipation	−0.880**	−0.883*	−0.790*
	(−2.058)	(−1.944)	(−1.807)
NormalizedDisgust	−0.371	−0.352	−0.147
	(−1.117)	(−0.998)	(−0.433)
NormalizedFear	−0.268	−0.177	−0.926**
	(−0.627)	(−0.389)	(−2.120)
NormalizedJoy	−0.011	−0.044	0.084
	(−0.031)	(−0.115)	(0.226)
NormalizedSadness	−0.303	−0.316	−0.232
	(−0.893)	(−0.877)	(−0.671)
NormalizedSurprise	0.419	0.437	0.266
	(1.321)	(1.299)	(0.820)
NormalizedTrust	1.078***	1.059**	1.223***
	(2.765)	(2.558)	(3.067)
Controls	Yes	Yes	Yes
Constant	−0.649***	−0.638***	−0.762***
	(−10.180)	(−10.182)	(−12.016)
Observations	2802	2802	2802
Adjusted R^2	0.498	0.468	0.476
F Statistic (df = 13; 2788)	215.003***	190.749***	196.677***

Notes: This table reports the estimates for the HAR-RV-LI models. The estimation period spans from 1 January 2012 to 31 August 2021. Three different RVs along with eight different daily emotions based on NRC are considered. The extended HAR-RV-LI is presented in Eq. (12). T-stats are reported in the parenthesis. ***Significant at the 1% level, **Significant at the 5% level, *Significant at the 10% level.

the model. Furthermore, in Table 5, we can observe that both monthly psychological sentiments and monthly discourse sentiments were significant in the Rogers and Satchell (1991) volatility estimation. We can argue that the effect of positivity, or positive sentiment, lasts longer than negativity, or negative sentiment. However, we can see in the results that the effect of positive sentiment was significantly negative over the long term. We can relate this result with that of PH and Rishad, 2020 who found the impact of sentiment on volatility to cause market uncertainty and lead to fewer returns. If investors fail to earn a risk premium for their expected volatility, they will move away from the market, which further causes volatility in the market.

In Table 7 and Table 8, we see the results of purely positive and purely negative finance-specific sentiments and their significance in predicting future volatilities of BTC. On the contrary, for psychological sentiments, both the purely positive and purely negative sentiments showed a significant effect on BTC volatilities in monthly aggregated sentiments. However, the purely negative finance-specific sentiments incorporating Loughran's and McDonald's dictionary was insignificant in all time scales in all volatility estimators. The result is again similar to negative psychological sentiment.

5.5. The HAR-RV and emotion lexicon sentiments: A robustness check

As an additional robustness check and to further explore the sentiments of different human emotions, I followed the NRC Emotion Lexicon. It is a list of English words and their associations with eight basic human emotions—anger, fear, anticipation, trust, surprise, sadness, joy, and disgust—and two sentiments—negative and positive. Fig. 5 shows a histogram of the corresponding eight basic human emotions in the news data sample. We can observe from the graph that trust and fear were the two emotions most triggered by BTC-related news. Because the weight of the emotions largely depends upon the number of words appearing in the news, I applied the min-max normalization process to scale these emotions. All the normalized emotions showed stationarity at their normalized levels. Next, I extended the HAR-RV with all eight daily normalized emotions, along with Google search intensity and news counts as controls. The results are presented in Table 9.

We can observe from the results that trust and anticipation were significant throughout all the volatility measures. In addition, fear and anger were significant at the 5% level in the volatility model incorporating the Rogers and Satchel (1991) method. Furthermore, Mohammad and Turney (2013) categorized trust and anticipation as positive sentiments and fear and anger as negative sentiments. In line with previous results presented in this paper, we can argue that it is not the negative

Table 10
Out-of-Sample Forecast Evaluation Statistics with Overall Sentiment.

			Sentiment			
	Measures	Basic-HAR	GI	QDAP	HE	LM
PK (1980)	ME	0.039	0.011	0.017	0.029	0.013
	RMSE	0.442	0.439	0.441	0.440	0.440
	MAE	0.315	0.313	0.314	0.314	0.316
	MAEP	0.115	0.113	0.114	0.114	0.114
	U2	0.816	0.816	0.817	0.814	0.818
GK (1980)	ME	0.095	0.010	0.019	0.034	0.014
	RMSE	0.467	0.463	0.464	0.464	0.464
	MAE	0.345	0.344	0.346	0.346	0.347
	MAEP	0.141	0.137	0.138	0.138	0.138
	U2	0.782	0.783	0.786	0.781	0.786
RS (1991)	ME	0.017	−0.014	−0.006	0.009	−0.014
	RMSE	0.472	0.469	0.470	0.469	0.468
	MAE	0.332	0.326	0.328	0.327	0.327
	MAEP	0.122	0.119	0.120	0.120	0.119
	U2	0.832	0.835	0.836	0.832	0.836

Notes: This table reports the values of various forecasting accuracy test results. The in-sample estimation period spans from 1 January 2012 to 31 December 2020, whereas the out-of-sample period ranges from 1 January 2021 to 31 August 2021. ME (mean error), RMSE (root mean square error), MAE (mean absolute error), MAEP (mean absolute error percentage), U2 (Thely's U2).

but positive sentiments that largely trigger volatility in the BTC market.

5.6. Out-of-sample forecast

To compare the out-of-sample accuracy of the different HAR-RV applications, first each alternative model was fitted to the in-sample RV data. Next, it was used to generate one-step-ahead out-of-sample forecasts. Because the data on volatility were generated with a daily range-based method, I focused on one-step-ahead forecasts in this study. However, multistep-ahead forecasts can be obtained similarly.

The in-sample data used for training purposes in this study were from January 1, 2012, until December 31, 2020. For testing the forecasting accuracy of the model, the out-of-sample data were from January 1, 2021, until August 31, 2021. The out-of-sample forecast accuracy measured by different methods is presented in Table 10.

$$ME = \frac{1}{n}\sum_{t=1}^{n} e_{it};\ RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_{it}^2};\ MAE = \frac{1}{n}\sum_{t=1}^{n}|e_{it}|;\ MAEP = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{e_{it}}{a_{it}}\right|$$

$$U2 = \sqrt{\frac{\sum_{t=1}^{n-1}\left(\frac{f_{t+1}-a_{t+1}}{a_t}\right)^2}{\sum_{t=1}^{n-1}\left(\frac{a_{t+1}-a_t}{a_t}\right)^2}} \tag{13}$$

There are different techniques in measuring the forecast accuracy of the statistical model. Let's define the forecast error as $e_{it} = a_{it} - f_{it.}$

where a_{it} is the actual and f_{it} is the forecasted value. Then, the five accuracy measures are defined by:

In Table 10, for a better comparison, we can observe the value of mean absolute error percentage (MAEP) for all three range-based volatility estimations of the sentiment dictionaries. MAEPs ranged between 11% and 13%, which according to Lewis (1982, p.40), is good forecasting accuracy. On the other hand, Theil's U2, which looks at the accuracy of one-step-ahead forecasts, showed the HAR-RV extended with sentiment to be better than the naive forecasting method. Furthermore, the error statistics of the extended HAR models with sentiments as additional explanatory variables gave lower errors, implying higher forecasting accuracy. Appendix A.3 and Appendix A.4 show two out-of-sample forecast accuracy plots.

6. Conclusion

In the past decade, BTC has made a lot of news in mainstream media. Some news media have portrayed it as a positive phenomenon, while many others have doubted its worth and authenticity. In recent years, cryptocurrency markets have also attracted considerable attention in academic literature, especially in finance and economics journals studying volatility of this new blockchain-based digital asset.

Modeling volatility is an important step to precisely measure the risk associated with an asset or portfolio of assets. An accurate estimation of volatility is vital for investors to develop an adequate strategy to hedge potential risks associated with an investment. In this study, I explored whether news media sentiments have an impact on BTC volatility by extending the work of Corsi (2009) with an HAR-RV with news-based sentiments as additional explanatory variables. I used past RVs of BTC and news sentiments to predict its future RVs. This study applied different range-based volatility estimation methods to obtain a better understanding of the nature of ranges and their significance in forecasting future volatilities. Furthermore, I differentiated financial sentiments and psychological sentiments cached in the news and their impact on BTC volatility in different time spans to capture the heterogeneity of news arrival times and sentiment memory lengths among investors. Moreover, to further explore the sentiments of different human emotions, I also extended the HAR model to the emotional level. As a result, I found trust and fear to be the two human emotions most triggered by BTC-related news and ultimately affecting its volatility.

Results for all the range-based estimators showed neither daily nor monthly psychological sentiments as being significant. The most likely reason behind this result might be the arrival of news to potential investors or readers. Not all audiences read newspapers on the same day they get published. Furthermore, general readers easily tend to forget the news over the long run, resulting in the decay of sentiment generated from the news within the month. However, it is surprising that the effect of finance-specific sentiments was significant over the long term in comparison to the other two psychological and discourse sentiments. One possible explanation of this result could be that BTC is more related to the field of finance than psychology. Another possible explanation could be that investors remember the news with more finance-specific sentiments for longer periods of time than news with more psychological sentiments. Moreover, I used the decomposition of overall sentiments into purely positive and purely negative sentiments to capture the heterogeneity between optimistic and pessimistic investors. The general idea is that optimistic investors are mainly guided by positive sentiments originating from the news, whereas pessimistic investors are mainly guided by negative sentiments. The results showed purely positive financial sentiment as being more responsible for BTC volatility. In other words, financially optimistic investors seem to be the main drivers of this market. Furthermore, the NRC Emotion Lexicon as a robustness check also showed trust and anticipation to be significant throughout all the volatility measures. Because NRC categorizes trust and anticipation as positive sentiments and fear and anger as negative sentiments, we can confirm the result that it is not the negative but the positive sentiment that largely triggers volatility in the BTC market. The out-of-sample forecasting accuracy of the model also showed the HAR-RV with sentiment extension to have a good forecasting accuracy irrespective of the choice of volatility measure.

Overall, the results reveal that information on time-varying sentiments could play a major role in analyzing the news media risk associated with BTC. Thus, the findings seem important for volatility modeling and developing a trading strategy. Given that capturing true sentiment in news plays a significant role in risk management and portfolio optimization, this paper has important implications for investors holding assets in the cryptocurrency market, more specifically, BTC. Moreover, one possible limitation of this study is the consideration of news sentiment generated from the news covered by the major English language newspapers only. Therefore, future research is encouraged on news media versus social media sentiment and volatility of digital assets like BTC. Furthermore, analyzing news sentiments with non-FinTech dictionaries might be another limitation of this study. Previous studies have shown that a borrowed dictionary from a different discipline is likely to misjudge true sentiment, I would also like to highlight the need for a FinTech-specific sentiment dictionary that helps to explore the true sentiments of the new digital financial market.

Declaration of Competing Interest

I declare that I have no significant competing financial, professional, or personal interests that might have influenced the performance or presentation of the work described in this manuscript.

Appendix A. Appendices

Appendix A.1. Major newspapers in English listed by lexisnexis.com from around the world

S.No.	Major Newspapers in English	S.No.	Major Newspapers in English
1	The Advertiser/Sunday Mail (Adelaide, South Australia)	47	The Independent (London)
2	The Age (Melbourne, Australia)	48	The Indianapolis Star (Indiana)
3	APN Australian Newspapers	49	The Irish Times
4	The Arizona Republic (Phoenix)	50	The Japan News
5	Arkansas Democrat-Gazette	51	The Jerusalem Post
6	The Atlanta Journal-Constitution	52	The Kansas City Star
7	The Australian	53	Los Angeles Times
8	Australian Financial Review	54	The Miami Herald
9	The Baltimore Sun	55	The Milwaukee Journal Sentinel
10	The Boston Globe	56	New Straits Times (Malaysia)
11	The Boston Herald	57	Newsday (New York)
12	The Buffalo News (New York)	58	The New York Post
13	Business Times (Malaysia)	59	The New York Times
14	The Business Times Singapore	60	The New Zealand Herald
15	The Canberra Times	61	Northern Territory News (Australia)
16	The Charlotte Observer	62	The Observer
17	Chicago Sun-Times	63	The Orange County Register
18	Chicago Tribune	64	The Oregonian
19	The Christian Science Monitor	65	Orlando Sentinel (Florida)
20	The Chronicle (Australia)	66	Ottawa Citizen
21	The Cincinnati Enquirer (Ohio)	67	The Philadelphia Daily News (PA)
22	The Columbus Dispatch	68	The Philadelphia Inquirer
23	The Courier Mail/The Sunday Mail (Australia)	69	Pittsburgh Post-Gazette
24	The Courier-Journal (Louisville, Kentucky)	70	The Plain Dealer
25	Daily News (New York)	71	The Press (Christchurch, New Zealand)
26	The Daily Oklahoman (Oklahoma City, OK)	72	Sacramento Bee
27	The Daily Telegraph (London)	73	San Antonio Express-News
28	Daily Telegraph and Sunday Telegraph (Sydney, Australia)	74	San Diego Union-Tribune
29	The Dallas Morning News	75	The San Francisco Chronicle
30	The Denver Post	76	The Seattle Times
31	Detroit Free Press	77	South China Morning Post
32	The Detroit News (Michigan)	78	St. Louis Post-Dispatch (Missouri)
33	The Dominion Post (Wellington, New Zealand)	79	The Star-Ledger (Newark, New Jersey)
34	Financial Times (London)	80	Star Tribune (Minneapolis MN)
35	Fort Worth Star-Telegram	81	The Straits Times (Singapore)
36	The Gazette (Montreal)	82	Sun-Sentinel (Fort Lauderdale)
37	Gazeta Mercantil Online	83	The Sunday Herald (Glasgow)
38	The Globe and Mail (Canada)	84	The Sydney Morning Herald (Australia)
39	Grand Rapids Press (Michigan)	85	Tampa Bay Times
40	The Guardian	86	The Tampa Tribune (Florida)
41	The Hartford Courant	87	Times - Picayune (New Orleans)
42	The Herald (Glasgow)	88	The Toronto Star
43	Herald Sun/Sunday Herald Sun (Melbourne, Australia)	89	USA Today
44	Het Financieele Dagblad	90	The Wall Street Journal
45	Hobart Mercury/Sunday Tasmanian (Australia)	91	The West Australian (Perth)
46	The Houston Chronicle		

Appendix A.2. Sample news file (LexisNexis.com)

LexisNexis

Virtual currency Bitcoin registers with European regulators

Guardian.com

December 7, 2012 Friday

Length: 478 words

Byline: Jemima Kiss, guardian.co.uk

Body

ABSTRACT

Site takes step towards legitimacy as euro accounts now subject to same protection as bank holdings

FULL TEXT

The virtual currency ***Bitcoin*** took a step towards legitimacy today as its eurozone wing joined the ranks of PayPal and Worldpay by becoming a registered payment services provider (PSP) under European law.

Under a deal made in France with the investment firm Aqoba and the Crédit Mutuel bank, ***Bitcoin***-Central now has an international bank ID number, meaning the network will be able to send and receive transfers to and from other banks and issue debit cards for users.

In a post on the *Bitcoin forum*, ***Bitcoin*** staffer davout announced: "At Paymium we spent lots of time and energy talking about ***Bitcoin*** to our regulating bodies, the Banque de France, the ACP (French equivalent of the American SEC), TRACFIN (AML French supervising body) etc. We engaged all these resources with one goal in mind: get these people to know ***Bitcoin***, advocate our beloved crypto-currency and listen to them, help them think until they finally reach the same conclusion as we did: there's nothing wrong with people being free.

"There's nothing wrong with people freely exchanging value, we don't hurt anybody, we're not forcing anyone to use ***Bitcoin***, we simply want to see our dream and the future of money become a reality."

The virtual currency has seen significant growth since it launched in 2009 with an estimated 10.5m bitcoins currently being traded. One ***bitcoin*** is currently worth £8.54, after peaking at nearly £18 in June 2011, meaning the ***Bitcoin*** empire represents £89.6m of *trading value*.

Bitcoin magazine's editor, Vitalik Buterin, told the BBC the deal would encourage more growth and make it more accessible to new users.

It will also mean balances held in euros by ***Bitcoin*** will be subject to the same protection and compensation laws as cash held in conventional banks.

"The more we see governments and banks being willing to deal with ***Bitcoin***, the more comfortable a lot of organisations are going to be making the step forward themselves," he said.

Despite a fiercely dedicated userbase in the tech community, ***Bitcoin***'s ubiquity and the anonymity of its users have also made it an attractive exchange platform for criminals, leading to a call by the US Senate in 2011 to investigate the site for tax evasion and money laundering.

The chairman of the non-profit ***Bitcoin*** Foundation, *Peter Vessenes, said in October that the site was battling against barriers to more widespread adoption*.

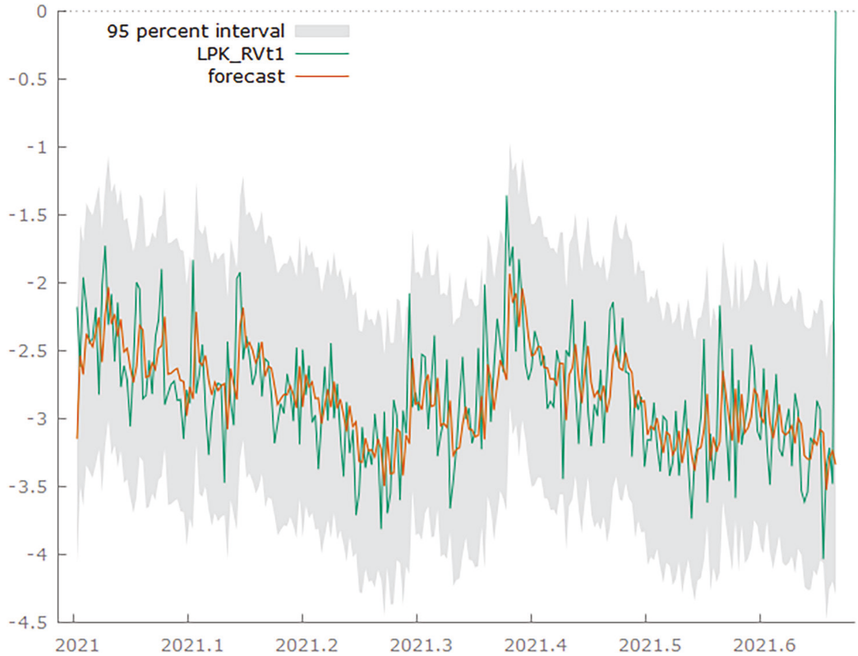
"There's a lot to love [but] ... there are botnet operators, hackers, and Ponzi-scheme runners floating around our space," he said.

"As the ***Bitcoin*** economy has evolved, we have all noticed barriers to its widespread adoption - [programs] that attempt to undermine the network, hackers that threaten wallets, and an undeserved reputation stirred by ignorance and inaccurate reporting."

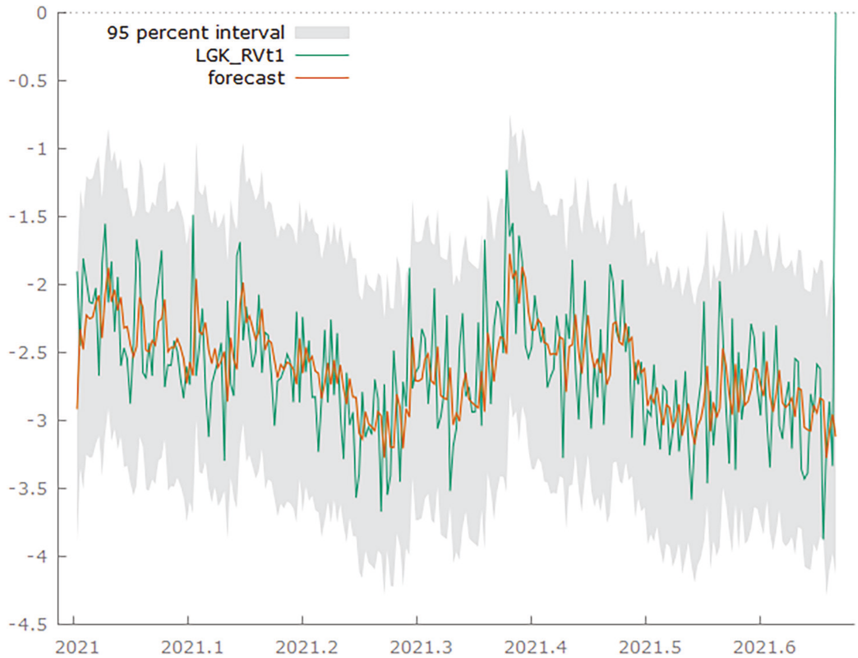
Load-Date: December 7, 2012

End of Document

Appendix A.3. Out-of-sample forecast accuracy plot (Parkinson volatility and Harvard psychological sentiment)



Appendix A.4. Out-of-sample forecast accuracy plot (Garman-Klass volatility and Harvard psychological sentiment)



References

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

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics, 89*, 701–720.

Baillie, R. T., Calonaci, F., Cho, D., & Rho, S. (2019). Long memory, realized volatility and heterogeneous autoregressive models. *Journal of Time Series Analysis, 40*(4), 609–628.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance, 61*(4), 1645–1680.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives, 21*(2), 129–152.

Bonato, M., Gkillas, K., Gupta, R., & Pierdzioch, C. (2020). Investor happiness and predictability of the realized volatility of oil price. *Sustainability, 12*(10), 4309.

Bouri, E., Gkillas, K., Gupta, R., & Pierdzioch, C. (2021). Forecasting realized volatility of Bitcoin: The role of the trade war. *Computational Economics, 57*(1), 29–53.

Busch, T., Christensen, B. J., & Nielsen, M.Ø. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics, 160*(1), 48–57.

Chen, P., Lin, J., Chen, B., Lu, C., & Guo, T. (2015). Processing emotional words in two languages with one brain: ERP and fMRI evidence from Chinese–English bilinguals. *Cortex, 71*, 34–48.

Chou, R. Y., Chou, H., & Liu, N. (2010). Range volatility models and their applications in finance. In *Handbook of quantitative finance and risk management* (pp. 1273–1281). Boston, MA: Springer.

Corbet, S., Larkin, C., Lucey, B. M., Meegan, A., & Yarovaya, L. (2020). The impact of macroeconomic news on Bitcoin returns. *The European Journal of Finance, 26*(14), 1396–1416.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics, 7*(2), 174–196.

Corsi, F., Pirino, D., & Reno, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics, 159*(2), 276–288.

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

Dos Rieis, J. C. S., de Souza, F. B., de Melo, P. O. S. V., Prates, R. O., Kwak, H., & An, J. (2015 April). Breaking the news: First impressions matter on online news. In *Ninth International AAAI conference on web and social media*.

Dutta, A., Bouri, E., & Saeed, T. (2021). News-based equity market uncertainty and crude oil volatility. *Energy, 222*, Article 119930.

Entrop, O., Frijns, B., & Seruset, M. (2020). The determinants of price discovery on bitcoin markets. *Journal of Futures Markets, 40*(5), 816–837.

Fry, J., & Cheah, E. T. (2016). Negative bubbles and shocks in cryptocurrency markets. *International Review of Financial Analysis, 47*, 343–352.

Fürsich, E. (2009). In defense of textual analysis: Restoring a challenged method for journalism and media studies. *Journalism Studies, 10*(2), 238–252.

Garcia, D., Tessone, C. J., Mavrodiev, P., & Perony, N. (2014). The digital traces of bubbles: Feedback cycles between socio-economic signals in the Bitcoin economy. *Journal of the Royal Society Interface, 11*(99), 20140623.

Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, 67–78.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *The Journal of Business Communication (1973), 45*(4), 363–407.

Ho, K. Y., Shi, Y., & Zhang, Z. (2020). News and return volatility of Chinese bank stocks. *International Review of Economics and Finance, 69*, 1095–1105.

Hu, J., Kuo, W., & Härdle, W. K. (2019). *Risk of Bitcoin market: Volatility, jumps, and forecasts. Availabe at SSRN 3658078 (august 25, 2021)*.

Karalevicius, V., Degrande, N., & De Weerdt, J. (2018). Using sentiment analysis to predict interday Bitcoin price movements. *The Journal of Risk Finance, 19*(1).

Kissler, J., Herbert, C., Peyk, P., & Junghofer, M. (2007). Buzzwords: Early cortical responses to emotional words during reading. *Psychological Science, 18*(6), 475–480.

Kristoufek, L. (2013). Can Google trends search queries contribute to risk diversification? *Scientific Reports, 3*(1), 1–5.

Lewis, C. D. (1982). *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth-Heinemann.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35–65.

Ma, F., Wei, Y., Huang, D., & Chen, Y. (2014). Which is the better forecasting model? A comparison between HAR-RV and multifractality volatility. *Physica A: Statistical Mechanics and its Applications, 405*, 171–180.

McAfee, D., Doubleday, Z. A., Geiger, N., & Connell, S. D. (2019). Everyone loves a success story: Optimism inspires conservation engagement. *BioScience, 69*(4), 274–281.

Mcduff, D., & Berger, J. (2020). Why do some advertisements get shared more than others?: Quantifying facial expressions to gain new insights. *Journal of Advertising Research, 60*(4), 370–380.

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of sentiment analysis in finance: From lexicons to transformers. *IEEE Access, 8*, 131662–131682.

Mohammad, S. M., & Turney, P. D. (2013). *Nrc emotion lexicon* (p. 2). Canada: National Research Council.

Muhammad, A., Wiratunga, N., & Lothian, R. (2016). Contextual sentiment analysis for social media genres. *Knowledge-Based Systems, 108*, 92–101.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 61–65.

Haritha, P. H., & Rishad, A. (2020). An empirical examination of investor sentiment and stock market volatility: Evidence from India. *Financial Innovation, 6*(1), 1–15.

Qadan, M., & Nama, H. (2018). Investor sentiment and the price of oil. *Energy Economics, 69*, 42–58.

Rogers, L. C. G., & Satchell, S. E. (1991). Estimating variance from high, low and closing prices. *The Annals of Applied Probability*, 504–512.

Rognone, L., Hyde, S., & Zhang, S. S. (2020). News sentiment in the cryptocurrency market: An empirical comparison with Forex. *International Review of Financial Analysis, 69*, Article 101462.

Sapkota, N., & Grobys, K. (2021). *Fear sells: Determinants of fund-raising success in the cross-section of initial coin offerings*. Available at SSRN 3843138.

Shen, C., Sun, C., Wang, J., Kang, Y., Li, S., Liu, X., … Zhou, G. (2018). Sentiment classification towards question-answering with hierarchical matching network. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 3654–3663).

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance, 62*(3), 1139–1168.

Yap, M. J., & Seow, C. S. (2014). The influence of emotion on lexical processing: Insights from RT distributional analysis. *Psychonomic Bulletin & Review, 21*(2), 526–533.

Young, L., & Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. *Political Communication, 29*(2), 205–231.

Zhang, D., He, W., Wang, T., Luo, W., Zhu, X., Gu, R., Li, H., & Luo, Y. J. (2014). Three stages of emotional word processing: An ERP study with rapid serial visual presentation. *Social Cognitive and Affective Neuroscience, 9*(12), 1897–1903.

Zhang, Y. J., & Li, S. H. (2019). The impact of investor sentiment on crude oil market risks: evidence from the wavelet approach. *Quantitative Finance, 19*(8), 1357–1371.

Fear sells: On the sentiment deceptions and fundraising success of initial coin offerings

Niranjan Sapkota [a, ǂ, *], Klaus Grobys [a, ǂ]

[a] University of Vaasa, Vaasa, Finland

ABSTRACT

Retrieving information from an intensive hand-collected whitepaper data library covering 5,033 ICOs launched before 2020, we analyze the determinants of ICO success as measured by the amount of raised funding. We assess the sentiment and readability in ICO whitepapers in addition to other information disclosures. Whereas we do not find any evidence for that the riskiness of ICO projects would lower the predicted amount of raised funding, our results strongly suggest that ICO investors are largely guided by emotions when making investment decisions. Contrary to earlier literature, we find a weak association between quality signals of whitepapers and its success.

JEL Classification: G11, G20, G41

Key Words: Initial Coin Offering (ICO), Natural Language Processing (NLP), Sentiment Dictionaries, Deep Learning, Artificial Neural Networks (ANN)

[ǂ] We gratefully acknowledge the Project Research Grant (Grant no. 190405) by the Foundation for Economic Education (LIIKESIVISTYSRAHASTO), Finland. We would like to thank the discussant, and other participants for providing insightful comments to our paper at the Aalto University, Graduate School of Finance (GSF), The Finnish Doctoral Programme in Economics (FDPE), Summer Workshop 2021, Helsinki, Finland. We are also thankful to the discussant and participants for providing us useful and encouraging comments at the World Finance Conference, University of Agder, Norway (Virtual), August 3-6, 2021. We are also grateful to the participants of the Cryptocurrency Research Conference 2021, September 16-17, 2021, organized by the University of Southampton, UK and ICMA Centre, UK. Moreover, we also received insightful and encouraging comments from the participants of the IRMC 2021, 14th Edition October 1-2, 2021, organized by the Risk Banking and Finance Society, in collaboration with the University of Florence, NYU Stern Salomon Center, USA along with the co-organizer Joint Research Centre (JRC)-European Commission and the host institution, the University of Cagliari, Italy.

[*] Correspondence to: Department of Accounting and Finance, University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland, Tel.: +358 468 418 558.

E-mail addresses: niranjan.sapkota@uva.fi (N. Sapkota), klaus.grobys@uva.fi (K. Grobys)

1. INTRODUCTION

Recently, initial coin offerings (ICOs) have received considerable attention as a new form of crowdfunding based on blockchain technology. Recent research documents that more than $30 billion has been raised via the ICO market (Howell, Niessner, and Yermack, 2020). Due to their nature as unregulated offerings of digital tokens on the Internet, aiming to collect funding for a project, ICOs disintermediate any external platform, payment agent or professional investor and thus disrupt the current financial system, i.e. the market for Initial Public Offerings (IPO).[1] Unsurprisingly, due to its easy-to-execute approach to attaining external funding, ICOs have recently attracted enormous attention.

Recent finance literature on ICOs explores the link between ICOs price responses and investor attention (Tsukioka, Yanagi, and Takada, 2018), potential factors affecting ICO market outcomes (Momtaz, 2020; Domingo, Piñeiro-Chousa and López-Cabarcos, 2020; Yu, 2019), the link between determinants of the characteristics of the advisory board and ICO fundraising success (Giudici, Moncayo, and Martinazzi, 2020), the usefulness of information availability as a market signal of quality (Meyer and Ante, 2020; Amsden and Schweizer, 2018), and the link between Twitter followers and activity and the success of ICOs (Benedetti and Kostovetsky, 2021). Another important contribution is the studies of Roosenboom, Kolk and Jong (2020) and Howell, Niessner, and Yermack (2020) who find that ICO success is associated with (*i*) disclosure, (*ii*) credible commitment to the project and (*iii*) quality signals such as token listings. Finally, the study which is perhaps the closest related to ours is the one of Fish (2019) uses data on 431 ICOs to investigate

[1] In corporate finance, IPO has several requirements, such as a good track record of earnings above a minimum earnings threshold, whereas other financial criteria are set by the exchange where the firm plans are listed. Whereas the NASDAQ requires a total of $11 million pre-tax earnings in the previous three years and more than $2.2 million in each of the two most recent years, none of these financial requirements apply for ICOs. Generally, anyone who has an innovative idea or is willing to create a company, is eligible to issue an ICO. One could even argue that companies that are financially qualified for an IPO are overqualified for an ICO. Furthermore, firms might have to wait for many years before fulfilling the criteria set by the stock exchanges to issue an IPO. In this regard, using CRSP data, Jovanovic and Rousseau (2001) show that it takes years and even decades for firms to be listed on the stock exchange.

the role of signalling ventures' technological capabilities in ICOs. Fish's (2019) results indicate that technical whitepapers and high-quality source codes increase the amount of raised funding.

Motivated by this recent stream of literature, the purpose of this study is to explore the factors that determine the success of ICOs in terms of raised funding. In doing so, a novel issue that we consider is whether sentiment embedded in ICO whitepapers serves as a predictor variable for ICO success. To do so, we retrieve a unique hand-collected data set using all 5,033 ICOs that were launched between August 2014 to December 2019 period. We combine and match the information from ICO whitepapers with various databases allowing us to identify plenty of ICO-specific information. In exploring the sentiment embedded in whitepapers, we applied four different sentiment dictionaries in association with seven different readability measures.

Our study contributes to the recent literature in various fundamentally important aspects. First and foremost, taking the broader finance perspective, our paper adds to the literature on entrepreneurial finance (e.g., Catalini and Gans, 2018; Chod and Lyandres, 2020; Kaal, 2018; Huang, Meoli, and Vismara, 2019; Lyandres, Palazzo, and Rabetti, 2019; Li and Mann, 2018). Specifically, our study adds to the literature exploring the determinants of success in ICOs (Adhami, Giudici, and Martinazzi, 2018; Fisch, 2019; Howell et al., 2020) by first (*i*) accessing the entire population of ICOs, that is, we retrieve all 5,033 ICOs launched in August 2014 to December 2019 period. As a consequence, our study is not exposed to potential small sample biases as it accounts for the whole population of available data. Second, (*ii*) we employ a total of 37 potential predictor variables that could have an impact on the success of ICOs. In doing so, we replicate Fisch's (2019) method as the raised amount represents our dependent variable. Given the current research context, there is no other study available covering our extensive data set and extracting such a large set of potential predictor variables. Moreover, Fisch (2019) uses ICO whitepaper for extracting word count only. Our study does not only extract the word count characteristics of the ICO whitepapers but also other important characteristics such as sentiments, emotions and readability. Besides the

ICO whitepaper, our study also incorporates, the possibility of an ICO project ending up as a scam, using 'Risk Score' as one of the predictor variables. On top of that our statistical model also accounts for social media hype, measured as the "Hype Score". Where Fisch (2019) accounts for qualitative disclosure in terms of high-quality source codes, patents and copyright, our model incorporates qualitative disclosures such as Country of Origin disclosure, Roadmap/Milestone disclosure, etc. and quantitative disclosures such as SoftCap, HardCap, Number of Tokens, Number of Categories, Team Size, etc.

Next, the psychological literature has repetitively confirmed the priority of processing emotion words against neutral words. (Anooshian and Hertel, 1994; Chen, Lin, Chen, Lu, and Guo, 2015; Kissler, Herbert, Peyk, Junghofer, Buzzwords, 2007; Yap and Seow, 2013; Zhang, He, Wang, Luo, Zhu, Gu, Li, Luo, 2018). In this regard, researchers in psycholinguistics and linguistics distinguished two kinds of emotion words, that is, "emotion words" and "emotion-laden words" (Altarriba, 2006; Pavlenko, 2008). Wu, Zhang and Yuan (2021) argue that behavioral and electrophysiological studies supported that also emotion-laden words (e.g., war, death, disaster, risk) affect human behavior. The psychological literature has not yet explored the effect of emotion-laden words in whitepapers. Due to the enormous amount of money involved in the market for ICOs, this is definitely not a trivial issue which needs to be investigated. In this regard, our study is the first that explores the psychological sentiment in ICO whitepapers and clarifies the following questions: Firstly (*i*), which emotional content dominates whitepapers, and secondly (*ii*), how psychological sentiment affects the success of ICOs.

Another important novel aspect of our study is that we explore the question of whether financial sentiment–as opposed to psychological sentiment–cached in whitepapers has an impact on the success of ICOs.[2] There are various sentiment dictionaries from which sentiment scores can be

[2] In general, sentiment is a genre of the appraisal theory. Specifically, sentiment analysis investigates opinions expressed in texts and comprises (*i*) the extraction of opinion polarity (positive or negative), (*ii*) the target (or specific aspects of the target) to which the opinion refers, (*iii*) the holder of the opinion, and (*iv*) the time at which the opinion was expressed

calculated. Studies comparing sentiment in the finance-specific domain have focused on Henry's (2008) and Loughran and McDonald's (2011) sentiment dictionaries. Using the Harvard-IV general dictionary, Loughran and McDonald (2011) found its word list to be largely inapplicable to financial contexts and created a finance-specific list. Henry (2008) captured the tone of earnings press releases to create a word list for financial texts. These studies found a finance-specific dictionary to be more powerful than the general psychological dictionary (Sapkota, 2022). There are already many studies (for example; Li et al., 2014; Alessia et al., 2015; Pröllochs, Feuerriegel, and Neumann, 2015; Yekrangi and Abdolvand, 2021) on finance domain-specific sentiments as opposed to psychological sentiment. This is because the previous studies have shown that the borrowed sentiment dictionary from a different discipline is likely to misjudge the true sentiment in that particular context.

A growing body of research suggests that affects a central component of individual decision-making, political judgment, and especially the processing of media contents. Notably, Young and Soroka (2012) highlight that negative affect seems to be extraordinarily important in the human psyche, and in political interactions. The whitepaper of an ICO is of major importance as it reveals the intended production outcome of the proposed business project: Consequently, potential investors may or may not invest in the ICO merely based on its content. Hence, several natural and important questions arise: First, (*i*) does it matter how an ICO whitepaper is conducted with respect to its content? Second, (*ii*) should the whitepaper be written in simple terms for easy readability so that even a naïve investor is able to grasp the project idea? Third, (*iii*) should the whitepaper's choice of words trigger the positive or the negative side of the sentiment in order to successfully attract investors? While earlier studies explored sentiment associated with IPOs (Loughran, McDonald,

(Muhammad, Wiratunga, and Lothian, 2016). Indeed, the saying *It is not what you say that matters but the manner in which you say it; there lies the secret of the ages* (William Carlos Williams) indicates that the tone of a text is perhaps more influential than its substantive content. In fact, plenty of studies have been devoted to exploring the sentiment of news content, political speeches, and advertisements.

2013; Bajo and Raimondo, 2017; Guldiken, Tupper, Nair, and Yu, 2017), our study is the first that seeks to answer these three important sentiment-related questions for ICO whitepapers.[3]

A final contribution of our study is a methodological one. Specifically, our study uses a different approach than previous studies. Earlier studies mainly focus on the performance of the ICO (e.g. token performance as measured in terms of returns), which is an ex-post ICO phenomenon (for example, Howell et al., 2020). Our paper extends earlier research by first (*i*) focusing on the sentiment side of the ICO and second, (*ii*) by utilizing cross-sectional data to identify potential key factors in determining the size of the raised amount. It is important to note that in doing so, our study controls for a battery of factors evidently associated with ICO success.

Identifying a total of 37 potential predictor variables, our result shows that investors in FinTech sectors are not immune to behavioral biases. The key results of our study can be summarized as follows. First, only the Harvard Psychological Sentiment Dictionary appears to provide useful information that can be linked to ICO success. Specifically, negative sentiment is associated with a higher amount of raised funding, whereas positive sentiment does not have any significant impact. Our study identifies that the prevalent emotion cached in whitepapers is 'fear'. Factorizing this emotion into its specific components shows that investors' behavior in the ICO market is mainly driven by fears associated with 'risk', 'problem', 'change', and 'regulation', among others.

Next, another important finding is that popularity in terms of media attention is a key determinant for the success of ICOs. We observe a linear relation as we move from a low to a high level of media attention. Whereas a low level of media attention does not correlate with ICO success,

[3] Apart from these first-order questions that are specifically related to the ICO whitepaper, there are also second-order questions that arise. For instance, what about other characteristics of ICOs that are usually not found in whitepapers, such as social media followers (e.g., as measured in terms of *Hype Score*), projects backed up by people disclosing their identity (e.g., as measured by Know Your Customers *KYC Score*), or potential risk for fraud (e.g., as measured by *Risk Scores*)? Do these factors also have an impact in attracting potential investors?

a higher level of media attention is associated with increased raised funding. Our results support the study of Benedetti and Kostovetsky (2021) in recognizing the impact of Twitter: Specifically, our findings indicate that a higher number of followers on Twitter correlate positively with ICO success. In this regard, a novel finding of our study is that signature campaigns are of significance. Signature campaigns, which are often referred to as bounty programs, may have different procedures[4].

An unforeseen finding is that readability does not have any impact on the success of ICOs, which is in stark contrast to what has been documented in the corresponding literature on IPOs (Loughran, McDonald, 2013; Bajo and Raimondo, 2017; Guldiken, Tupper, Nair, and Yu, 2017).[5] Further, team size only marginally influences ICO success, whereas risk assessments, country disclosure, category (e.g., industry), or the similarity of a whitepaper with another project's whitepaper do not. Finally, the evidence documented in the current research suggests that ICO investors are, generally, not acting as rational investors because they (*i*) are biased towards negative sentiment, (*ii*) do not take into account the risk assessments, and (*iii*) do not even consider whether a whitepaper is conducted in an understandable manner or (*iv*) if it violates copyrights.

This paper is organized as follows: The next section provides a literature review. The third section presents the data and methodology. Furthermore, the fourth section documents the results and the last section concludes.

[4] Essentially, a signature campaign is a subscription campaign where ICOs release signatures with an embedded code. The bounty stake associated with the campaign is based on the ranking of the participants. Generally, for most bounty campaigns, only people on Bitcointalk forum who are at least ranked as Junior Member can participate. We find that signature campaigns appear to be useful predictor variables for ICO success.

[5] This finding is also contrary to Fish's (2019) study, which finds that the way a whitepaper is conducted has an impact on ICO success.

2. LITERATURE REVIEW

In the wake of the increase of social media and blogging websites, textual sentiment analysis has increased significantly. Nowadays, companies are using Twitter and Facebook to analyze the sentiment of their clients. Since user-generated content such as posts, shares, likes, tweets and retweets are openly available, firms and companies have enormous opportunities to understand the customers (He et al., 2015). With the rise of digital crowdfunding, there are tremendous opportunities for the clients to understand their companies too. Openly available ICO whitepapers consist of essential information about the startups. As a consequence, potential investors have the opportunity to assess the quality of the project by understanding the hidden sentiments in the whitepaper. Investors can also perform the sentiment analysis of any social media or the blog posts shared by the startups. Previous studies have identified various factors associated with the success of ICOs and in this study, we group these papers into four different groups.

2.1.Information disclosure and ICO success

Howell et al. (2020) find that the success of ICO depends on the disclosure, credible commitment to the project along with other quality signals. Their study shows that ICO token exchange listing causes higher future employment and giving access to token liquidity has a positive outcome for the enterprise. Using the database of 1,000 ICO whitepapers, Zetzsche et al. (2017) show that many ICOs offer inadequate disclosure of information: the majority of the ICO whitepapers are either silent on the initiators or backers/promotors or do not provide contact details. Furthermore, more than half of the ICOs do not elaborate on the applicable law, segregation or pooling of client funds, and the existence of an external auditor[6]. Therefore, the decision to frequently invest in ICOs can perhaps not be the outcome of a rational thought process. Similarly, using hand-collected data on

[6] In November 2017, the European Supervisory Markets Authority issued statements notifying investors and firms of potential risks native to certain ICOs. The authority notified that certain feature ICOs may be governed by existing EU legislation.

472 public token sales over the period of 2013–2017, Boreiko and Risteski (2020) find that some contributors often invest in more than one campaign, and such serial investors contribute earlier. However, they are not more informed and fail to pick better quality ICOs. On the other hand, Hornuf, Kück and Schwienbacher (2021) show that issuers who disclose their source code are more likely to be targeted by hackers and scammers, highlighting the risk of disclosing the code. They find it extremely difficult to predict fraud with the information available (whitepapers and other sources like websites, social media accounts, etc.) at the time of ICO issuance. Zhang Aerts, Lu, and Pan (2019) study the data from the four largest tokens exchanges in Asia and their findings indicate that whitepapers with more readable disclosures are likely to result in a higher first-day return.

2.2. Investors' sentiments, whitepaper readability, and ICO performance

Baker and Wurgler (2007) use several market-based measures as proxies for investor sentiment. Besides the market-based measure, the other most common approach applied in earlier research has been survey-based indices. More recently, building an investor-sentiment index employing qualitative transcripts (for example 10K filings, whitepapers, earning announcements, etc), daily news, internet search, social media content, blogs, etc. have gained popularity because traditional approaches like market-based and survey-based methods seem to be less transparent.

Drobetz, Momtaz, and Schröder (2018) examine to what extent the market for ICOs is driven by investor sentiment. Their results, based on sentiment and coin price data, show that the ICO market is driven by digital financial market sentiment, whereas it is almost unrelated to general capital market sentiment. Their results show that social media channels overdrive traditional news channels as the main source of investor sentiment. Similarly, Domingo, Piñeiro-Chousa, and López-Cabarcos (2020) also find that sentiment extracted from social networks positively influences ICO returns. Specifically, the authors document that Bitcoin spot and Bitcoin futures returns are positively correlated with ICO returns, whereas the existence of a presale period has a negative influence. Zhang Aerts, Lu, and Pan (2019) study the linkage between the readability of whitepapers and the first-day return. Using data from the four largest tokens exchanges in Asia, their findings indicate that whitepapers with more readable disclosures are likely to result in a higher first-day return for ICO investors. Similarly, Qadan (2019) uses readily available 11 different sentiment indices as different proxies of risk appetite. These indices are Baker and Wurgler's (2006) index; Huang et al. (2015) HJTZ Index; Baker et al. (2016) Economic /Monetary Policy Uncertainty Index (EPU); American Association of Individual Investors' (AAII) Sentiment Survey, Consumer Sentiment Index (CSI); Consumer Confidence Index(CCI); Louis Fed Financial Stress Index (STLFSI). This is not the same in our case as we quantified the qualitative data (whitepapers) using natural language processing tools and extracted sentiment scores by applying four different sentiment dictionaries and further expand it to the emotional level. We also extracted the seven different readability scores from each of the ICO whitepapers.

2.3. Connectivity of CEO and the advisors and ICO performance

Giudici, Moncayo, and Martinazzi (2020) find a relationship between the number of advisors' connections and their capability. They conclude that advisors in connection with multiple ICOs bridge the gap between the network and result in ICO success. They also show that the well-connected advisors in other ICOs are directly related to a larger amount of raised funding. Similarly, Amsden and Schweizer (2018) study 1,009 ICOs from 2015 to March 2018 and highlight those better-connected CEOs are positively correlated with ICO success. Moreover, providing information on a hard cap in a pre-ICO can help investors measure success in the pre-sale. Momtaz (2020) explores the factors affecting the ICO market outcomes and finds that management quality and project quality are positively correlated with the funding amount and returns. This study also finds that highly visionary projects harm success. Furthermore, the study shows that highly visionary projects are more likely to fail, as 21% of all tokens got delisted from a major exchange platform during the sample period.

2.4. Technical factors among ICOs and their signaling capabilities

Analyzing the data of 1,392 projects, Yu (2019) shows that the volatility of the main cryptocurrencies has a significant impact on the success of ICOs. For example, the success of ICOs on smart contracts built upon the ERC-20 token primarily depends on the volatility of Ethereum and secondarily on all other factors such as team quality. Furthermore, Meyer and Ante (2020) analyze 250 cross-listings of 135 different tokens and calculate abnormal returns for specific samples using an event study. They find that returns are driven by success in terms of token performance and project funding as well as characteristics such as regulation and domestic market size of the ICO issuing party.

Other characteristics such as blockchain infrastructure, token distribution, team, campaign duration, and whitepaper characteristics also seem to influence the perceived project quality as well as the cross-listing returns. Fisch (2019) assesses the determinants of the amount raised in 423 ICOs. The study explores the role of technological capabilities among ICOs and their signaling capabilities. The results also show that technical whitepapers and high-quality source codes are positively related to the amount of raised funding. Surprisingly, patents and copy rights, which can be considered quality signals, are not associated with increased amounts of funding.

3. DATA AND METHODOLOGY

3.1. Preparing the data set

We applied *rvest and xml2* web scrapping packages in the standard statistical software R to download the data from icorating.com., icosbull.com, and tokendata.io. Icorating.com has the *risk score* (e.g., a score for measuring potential fraud) and the *hype score* (e.g., a score resembling the number of social media followers) for more than 5,000 ICOs with additional information on the amount of raised funding which we denote in our study as *raised* measured in terms of USD. Similarly, the website icosbull.com provides basic data (i.e., data such as the *Name, Symbol, Description, Country*), financial data (i.e., data such as *Softcap, Hardcap, Raised Amount*) and data on social signal views (i.e., data such as Telegram or Twitter followers covering around 3,000 ICOs.) Moreover, the website tokendata.io has information on the daily price and return data on the token sales of the ICOs including the raised amount. We downloaded the financial information of those listed ICOs using the same web scrapping packages. Unfortunately, the financial information for many ICOs is missing on these three websites. Furthermore, financial information, especially on the raised amount of funding, which is of major importance in our study, is missing on many websites (including major ICO database providers such as icobench, neironix, icoholders) Nevertheless, after combining the data retrieved from icorating.com, iscosbull.com and tokendata.io, we were able to collect the information about the raised amount for 1,507 ICOs issued in the August 2014 to December 2019 period. Furthermore, we also observe some non-uniformity in the reported raised amount for some completed ICOs on these websites. Most of the reported raised amounts on these websites are rounded in thousands and millions. If available, our sample takes the exact raised amount (not rounded) from the above-mentioned websites.

As a result, we manually collected all 5,033 ICO whitepapers from various sources. Neironix.io provides access to the majority of ICO whitepapers by providing a direct link to the website of the ICO providers. However, some failed, scam and unsuccessful ICOs, unfortunately, removed whitepaper access from their websites. Fortunately, some websites have copies of whitepapers in their own databases. Using intensive manual work, we are able to collect the whitepapers of all unique ICOs for this study. **Figure 1** shows the step-by-step process of the data retrieval highlighting various sources used to obtain the data set.

Fig. 1. ICO data accumulation process (2014-2019)
This figure shows our sample of 1508 ICO with raised amount info data generation process.

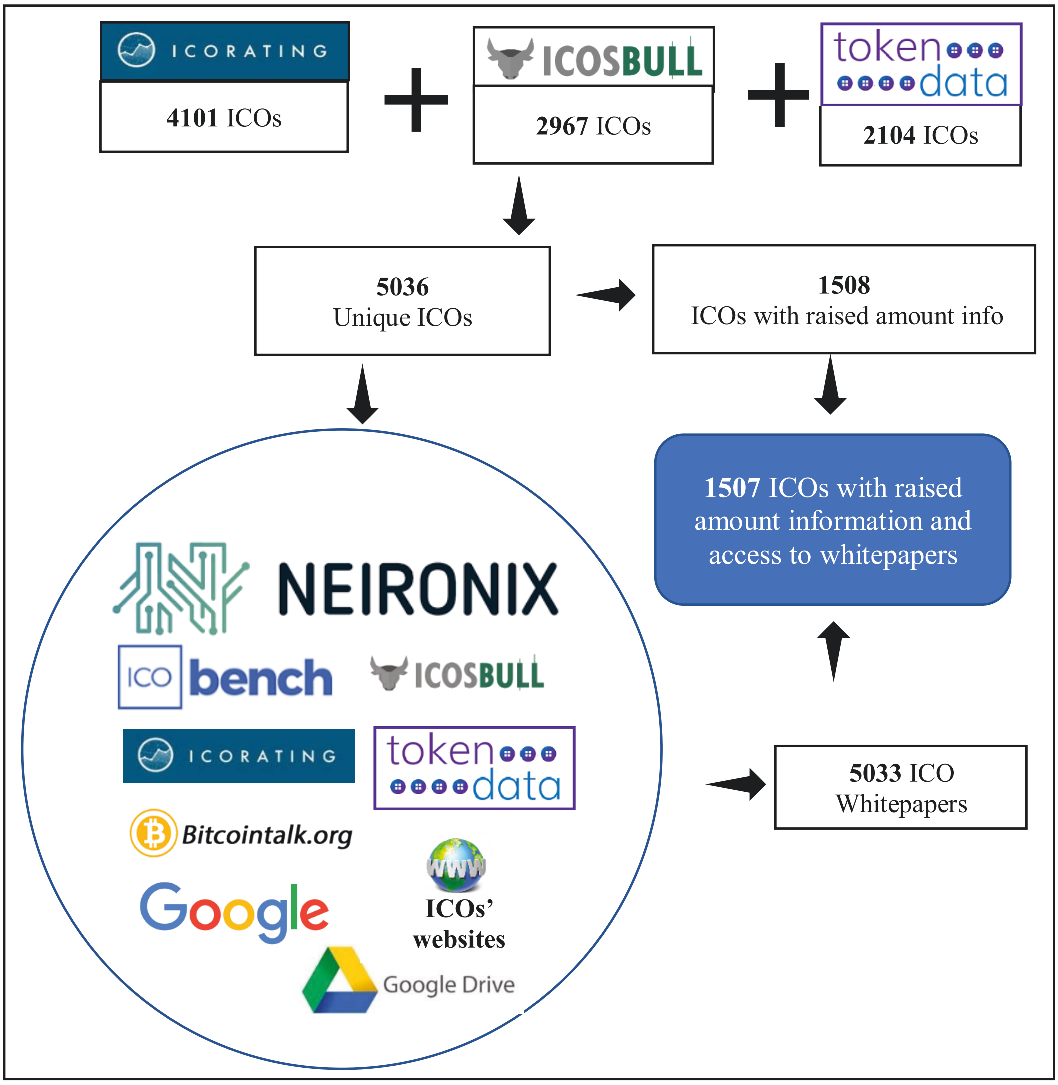
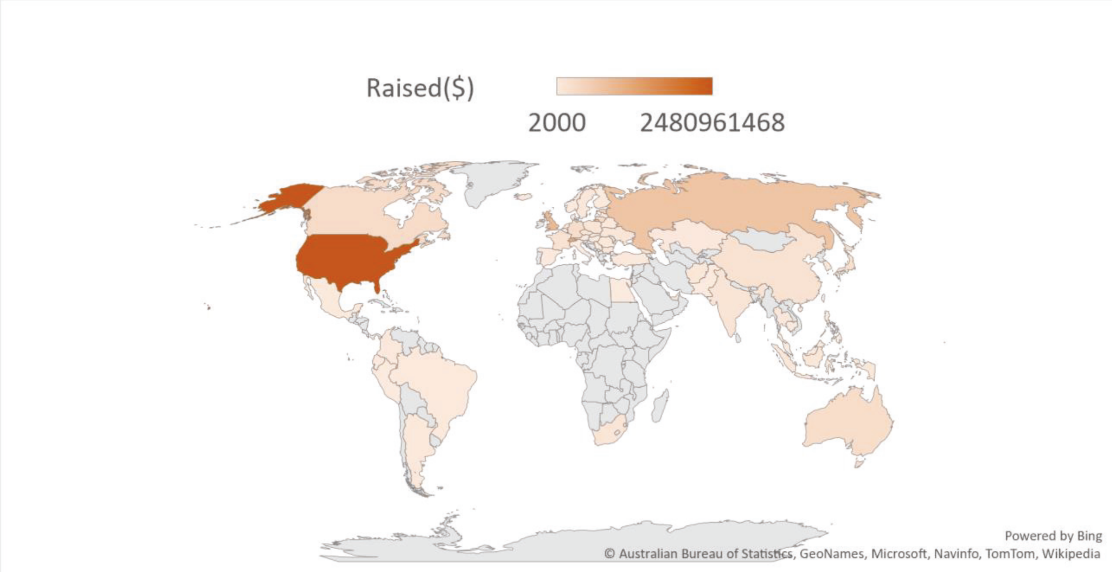


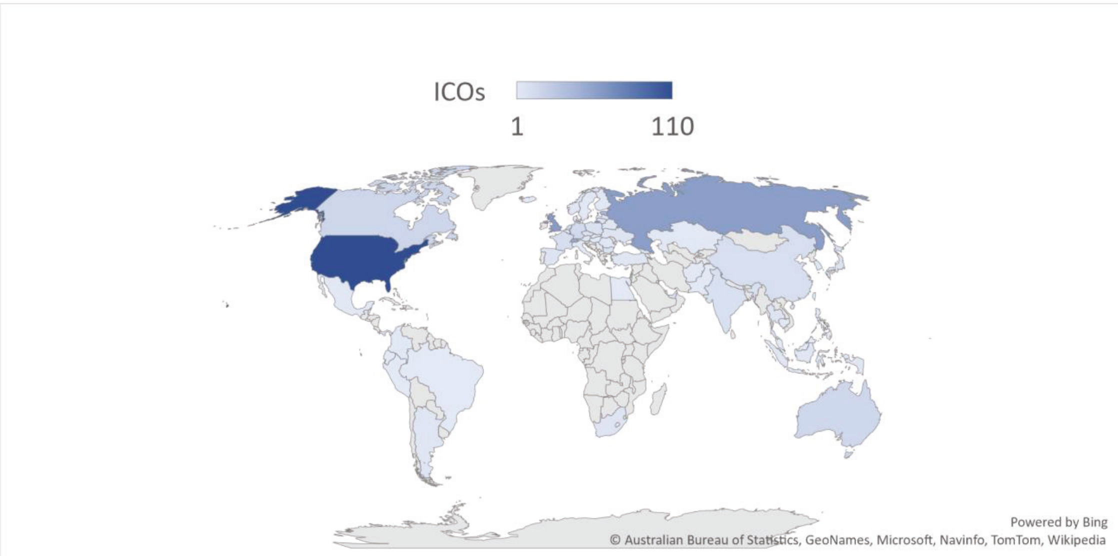
Figure 2 and **Figure 3** show the geographical representation of the amount raised by ICOs in US dollars and the number of ICOs launched during the 2014-2019 period respectively. Furthermore, **Figure 4** shows the pie chart of ICOs registered under the top 20 different categories.

Fig. 2. Geographical representation of amount raised by ICOs during 2014-2019 (This heat map is created using Microsoft Excel, it includes 1507 ICOs with the Raised amount available, for actual USD figures see **Appendix A.2**. and for the evolution of the funds raised using ICOs over time see **Appendix A.3.**)



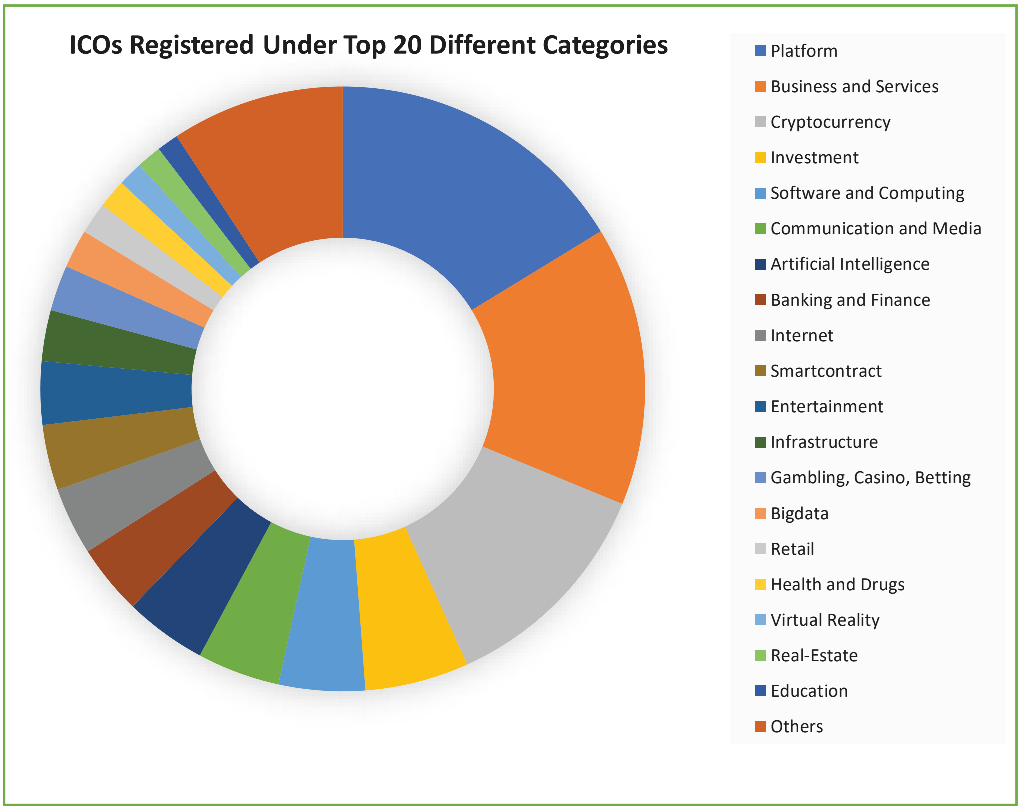
Note: This map excludes raised amount data for 765 ICOs where the country information is not disclosed.

Fig. 3. Geographical representation of the number of ICOs launched during 2014-2019 (This heat map is created using Microsoft Excel, it includes 1507 ICOs with the Raised amount available, for exact numbers of ICOs, see **Appendix A.2.)**



Note: This map excludes 765 ICOs where the country information is not disclosed.

Fig. 4. A pie chart representing ICOs registered under the top 20 categories
(Note: same ICO is registered under different categories, for detail see **Appendix A.4.**)



From data retrieval to tabulation of the results, we used various R packages which, along with their functions and usage, are described in **Appendix A.1**.

3.2. What variables could we identify after combining the information from whitepapers with other data sources?

3.2.1. How long, detailed and accessible is the whitepaper?

In total, we were able to identify 37 different variables for 1,507 ICOs of which we have information on the raised amount of funding. For each ICO published and completed during 2014-2019, we gathered project-related characteristics mainly from the particular ICO's whitepaper and also from different open source websites listing and rating ICOs (see **Figure 1**). We report the definitions for our variables in **Table 1** and the corresponding summary statistics of those variables in **Table 2**.

Table 1. Variable Names, Descriptions and Their Sources

S.No.	Variables	Full Name	Description	Sources
1	Name	Name of the ICO	ICOs published since 2014-2019	icorating.com, icobulls.com, tokendata.io
2	PL	Page Length	total number of pages in a whitepaper(pdf) file	icorating.com, icobench.com, icobulls.com, neironix.io
3	WC	Word Count	total number of words in a whitepaper corpus	Corpus, R SentimentAnalysis Package
4	S.GI	Overall Sentiment GI	psychological Harvard-IV dictionary (Overall= positive-negative)	Corpus, R SentimentAnalysis Package
5	N.GI	Negative Sentiment GI	psychological Harvard-IV dictionary (Negative)	Corpus, R SentimentAnalysis Package
6	P.GI	Positive Sentiment GI	psychological Harvard-IV dictionary (Positive)	Corpus, R SentimentAnalysis Package
7	S.HE	Overall Sentiment HE	Henry’s Business Communication dictionary (Overall= positive-negative)	Corpus, R SentimentAnalysis Package
8	N.HE	Negative Sentiment HE	Henry’s Business Communication dictionary (Negative)	Corpus, R SentimentAnalysis Package
9	P.HE	Positive Sentiment HE	Henry’s Business Communication dictionary (Positive)	Corpus, R SentimentAnalysis Package
10	S.LM	Overall Sentiment LM	LoughranMcDonald finance-specific dictionary (Overall=positive-negative)	Corpus, R SentimentAnalysis Package
11	N.LM	Negative Sentiment LM	LoughranMcDonald finance-specific dictionary (Negative)	Corpus, R SentimentAnalysis Package
12	P.LM	Positive Sentiment LM	LoughranMcDonald finance-specific dictionary (Positive)	Corpus, R SentimentAnalysis Package
13	RU.LM	Uncertain Sentiment LM	Sentiment uncertainty LoughranMcDonald finance-specific dictionary (neither Negative nor Positive)	Corpus, R SentimentAnalysis Package
14	S.QDAP	Overall Sentiment QDAP	QDAP Qualitative Data Analysis Program University of Pittsburgh (Overall=positive-negative)	Corpus, R SentimentAnalysis Package
15	N.QDAP	Negative Sentiment QDAP	QDAP sentiment polarity of text by grouping variables (Negative)	Corpus, R SentimentAnalysis Package
16	P.QDAP	Positive Sentiment QDAP	QDAP sentiment polarity of text by grouping variables (Positive)	Corpus, R SentimentAnalysis Package
17	S.Score	SimilarityScore	Jaccard Similarity Socer (25,331,089 pair comparision for similarity) 2nd highest match	Corpus, R Textreuse Package

Table 1 continue..

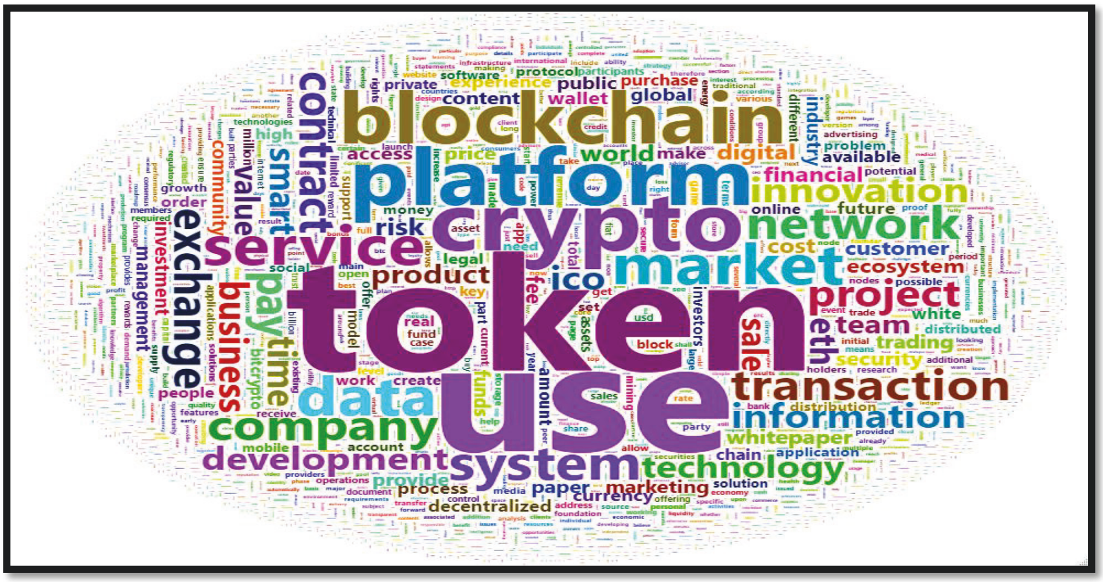
S.No.	Variables	Full Name	Description	Sources
18	Flesch	ReadabilityScoreType1	"ARI", "ARI.simple", "Bormuth", "Bormuth.GP", "Coleman", "Coleman.C2", "Coleman.Liau", "Coleman.Liau.grade", "Coleman.Liau.short", "Dale.Chall", "Dale.Chall.old", "Dale.Chall.PSK", "Danielson.Bryan", "Danielson.Bryan.2", "Dickes.Steiwer", "DRP", "ELF", "Farr.Jenkins.Paterson", "Flesch", "Flesch.PSK", "Flesch.Kincaid", "FOG", "FOG.PSK", "FOG.NRI", "FORCAST", "FORCAST.RGL", "Fucks", "Linsear.Write", "LIW", "nWS", "nWS.2", "nWS.3", "nWS.4", "RIX", "Scrabble", "SMOG", "SMOG.C", "SMOG.simple", "SMOG.de", "Spache", "Spache.old", "Strain", "Traenkle.Bailer", "Traenkle.Bailer.2", "Wheeler.Smith", "meanSentenceLength", "meanWordSyllables")	Corpus, Readability, KoRpus Package
19	Flesch.K	ReadabilityScoreType2		
20	RIX	ReadabilityScoreType3		Corpus, Readability, KoRpus Package
21	SMOG	ReadabilityScoreType4		Corpus, Readability, KoRpus Package
22	FOG	ReadabilityScoreType5	Readability score selected among these measures where R output has no any errors.	Corpus, Readability, KoRpus Package
23				
24	ARI	ReadabilityScoreType6		Corpus, Readability, KoRpus Package
25	Col	ReadabilityScoreType7		Corpus, Readability, KoRpus Package
26	HScore	HypeScore	Hype Score (Dummy, Rated(1) not rated (0))	icorating.com
27	RScore	RiskScore	Risk Score (Dummy, Rated(1) not rated (0))	icorating.com
28	NoC	NumberofCategory	ICO registered under different categories(1-14)	Icobulls.com
29	Team	TeamInformation	Team size	icorating.com
30	C.Info	CountryInformation	Country of issue information Dummy (1 given, 0 not given)	Icobulls.com
31	KYC	KnowYourCustomer	Known Your Customer Scor	icorating.com
32	Twitter	Twitter Information	Social media profile followers	icorating.com
33	Miles	Milestones	Milestone information in website/corpus Dummy (1 given, 0 not given)	Corpus, icorating.com
34	SoftCap	SoftCapital	Softcap information website/database/corpus Dummy (1 found, 0 not found)	Corpus, icorating.com
35	HardCap	HardCapital	Hardcap information website/database/corpus Dummy (1 found, 0 not found)	Corpus, icorating.com
36	NoT	Number of Tokens Info	Number of tokens Dummy (1 known, 0 unknown)	icorating.com
37	RaisedA	Raised Amount Info	Raised amount information in USD	icorating.com, icosbull.com, tokendata.io

Table 2. Summary Statistics

Variables	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(50)	Pctl(75)	Max
Word Count	1507	4337.591	2835.686	241	2411.5	3820	5645	41532
Page Length	1507	31.999	17.936	1	20	29	41	288
Flesch Readability Score	1507	34.461	11.503	-75.576	29.137	34.95	40.883	120.205
Flesch.K Readability Score	1507	14.629	3.462	-3.010	12.828	14.25	15.825	66.863
RIX Readability Score	1507	7.457	2.586	0.000	6.056	7.161	8.362	45.136
SMOG Readability Score	1507	15.613	1.936	3.129	14.430	15.467	16.606	32.506
FOG Readability Score	1507	17.861	3.643	0.800	15.957	17.49	19.150	71.530
ARI Readavility Score	1507	15.160	4.461	-6.300	12.934	14.71	16.570	87.841
Col Readability Score	1507	34.007	5.781	-38.450	30.775	33.44	36.692	90.550
Sentiment GI	1507	0.142	0.038	-0.013	0.124	0.143	0.166	0.240
Negative GI	1507	0.064	0.018	0.000	0.054	0.063	0.073	0.148
Positive GI	1507	0.206	0.040	0.0002	0.191	0.209	0.227	0.300
Sentiment HE	1507	0.013	0.008	-0.017	0.008	0.013	0.017	0.096
Negative HE	1507	0.006	0.004	0.000	0.003	0.005	0.008	0.030
Positive HE	1507	0.019	0.007	0.000	0.014	0.018	0.023	0.099
Sentiment LM	1507	-0.014	0.016	-0.088	-0.024	-0.014	-0.004	0.050
Negative LM	1507	0.045	0.015	0.000	0.036	0.045	0.055	0.107
Positive LM	1507	0.031	0.010	0.000	0.026	0.031	0.037	0.090
RUncertainty LM	1507	0.013	0.006	0.000	0.008	0.012	0.016	0.061
Sentiment QDAP	1507	0.100	0.029	-0.027	0.084	0.102	0.119	0.197
Negative QDAP	1507	0.036	0.013	0.000	0.028	0.035	0.044	0.109
Positive QDAP	1507	0.137	0.029	0.000	0.125	0.139	0.153	0.225
Raised Amount (1M, USD)	1507	20.265	122.517	0.0001	2.314	8.549	19.956	4234.276
Jaccard Similarity Score	1507	0.061	0.130	0.000	0.010	0.02	0.050	1.000
KYC Score	1507	1.656	1.767	0	0	1,655	3.4	5
Twitter Followers	1507	3897.20	16428.26	0	0	0	2103.20	318271
No. of Categories	1507	1.652	1.307	1	1	1	2	11
Team Size	1507	4.819	7.490	0	0	4	9	45
Dummy Variables	**N**	**Mean**	**St.Dev.**					
High Hype Score	1507	0.151	0.358					
Medium Hype Score	1507	0.280	0.449					
Low Hype Score	1507	0.248	0.432					
Hype Score Not Rated	1507	0.292	0.455					
High Risk Score	1507	0.111	0.315					
Medium Risk Score	1507	0.182	0.386					
Low Risk Score	1507	0.062	0.241					
Risk Score Not Rated	1507	0.639	0.481					
Country Disclosed	1507	0.493	0.500					
Road Map/ Milestone	1507	0.716	0.451					
SoftCap Disclose	1507	0.236	0.425					
HardCap Disclose	1507	0.309	0.462					
Number of Tokens Disclose	1507	0.310	0.463					

We list *Word Count* as a possible predictor variable affecting the raised amount as a detailed and inclusive whitepaper might have an extensive word count. It is important to understand whether or not a detailed whitepaper is sufficient to raise the amount needed to fund the business project of the ICO. Our data set reveals that there is one whitepaper with more than 41 thousand words and at the same time we observe that the minimum of words used in a whitepaper is 241 words only. On average, a whitepaper in our sample includes 4,338 words. In this regard, **Figure 5** displays a word cloud for the most frequently used words in these whitepapers. It shows that many ICO whitepapers are focused on token, usage, blockchain, and platform.

Fig. 5. Word cloud of 5033 ICO whitepapers 2014-2019
(This word cloud is created using the wordcloud2 package in R)



(**Note:** To see the frequency of the words, please see **Appendix A.5.**)

Similarly, the *Page Length* of an ICO may also provide some useful information on the project. Besides words, a page also consists of many graphical explanations. A lengthier whitepaper tends to have a large number of words and may also include many tables and graphs. Some ICO projects provide a light paper i.e., *'one-page white paper'* instead of or even on top of a regular whitepaper. There are many ICOs in our data sample that provides only the light paper version of their whitepaper. The lengthiest whitepaper in our data sample exhibits almost 300 pages, whereas the shortest one is just a single page of light paper. Mostly, the whitepapers are A4 and Letter size files in portable document format (pdf). However, some whitepapers are a single long pdf file with no page division. Our methodology treats these files as single-page files by default.

The readability score measures the reading difficulty of a document. The question arises whether it is difficult (easy) to raise funding if the whitepaper is difficult (easy) to read. We seek to answer this question by applying different readability measures in our studies[7]. For example, the Flesch Readability score for 1,507 whitepapers exhibiting information on raised funding has a mean of 34.461, which shows that, on average, whitepapers are difficult to read and best understood by college graduates (see **Appendix A.7.**). There is no limit on the lowest side of the Flesh readability score. Very complicated sentences can have negative scores. The lowest Flesh score in our sample is -75.58. Flesh-Kincaid Grade Level is a readability test designed for English texts. Note that the test focuses on polysyllabic words and long sentences. It measures reading difficulty related to the approximate US grade level.

Similarly, Automated Readability Index (ARI) works well with both English and Western European languages. It uses long words and long sentences to calculate a readability score. It indicates how difficult the page is to read. The score can be matched to an equivalent reading ability level. The mean ARI of 14 in our sample in **Table 2.** implies that on average the whitepaper can be read and understood by 14th grade (i.e., university degree) students. Furthermore, Coleman-Liau is another recognized readability test designed primarily for English texts. It focuses on long words and long sentences. Using this test, a score can only be calculated if the content exceeds 100 words. Since the lowest number of words in our data sample of whitepapers is 288, we have no difficulties in using this readability measure. The test produces an approximate representation of the US grade level needed to comprehend the text.

[7] To understand how the readability scores for different readability measures are calculated, see **Appendix A.6.**

Gunning Fog is a readability test for English texts only. It also focuses on complex polysyllabic words and long sentences. We also include the SMOG readability test which is developed just for English texts. This measure also primarily focuses on polysyllabic words. However, a score can only be calculated if the document has at least 30 sentences. Another readability measure is the Rate Index RIX created by the Australian teacher Jonathan Anderson. Anderson wanted to convert the formula to a grade level. The average score of 7.5 (7.2 and above = 12^{th} grade) in RIX tells us that the average whitepaper can be read by college-level students. Unlike other readability measures, RIX can be used for both English and non-English texts.

3.2.2. Identifying the sentiment in whitepapers using popular sentiment dictionaries

Another important novel element in our study is the sentiment hidden in the text of an ICO whitepaper. By applying four different sentiment dictionaries we seek to answer if positive/negative sentiment is associated with success/failure in raising capital for ICOs. We apply the *SentimentAnalysis* package in R statistical software, giving us positive, negative and overall sentiment scores for four different sentiment dictionaries applied to ICO whitepapers. In this regard, the *AnalyzeSentiment* function from the above-mentioned package gives sentiment scores for:

i. *Harvard-IV General Purpose Psychological Dictionary*
ii. *Henry's Finance Dictionary*
iii. *Loughran Mc Donald Finance-specific Dictionary*
iv. *QDAP (Quantitative Discourse Analysis Package) Dictionary*

Firstly, the Harvard-IV psychological dictionary, comprising a list of positive and negative words, is a general-purpose dictionary developed by Harvard University. It maps each whitepaper with counts on positive, negative and overall sentiment[8]. The mean result in the summary data statistics shows that the whitepapers have slightly more positive psychological sentiment than negative.

[8] For more please visit; http://www.wjh.harvard.edu/~inquirer/

Second, Henry (2008) studies the genre of earnings press releases along with quantitative analysis. He uses capital markets data to assess the investor impact of tone and other stylistic attributes. The study's results suggest that tone influences investors' reactions. Using the dictionary developed by the author, the R package gives us the sentiment scores as positive, negative and overall. Furthermore, we get a similar sentiment variation with this dictionary too, where the mean positive sentiment score is slightly larger than the mean of negative sentiment.

Third, Loughran and McDonald (2011) show that word lists developed for other disciplines misclassify common words in the financial text. Employing a large sample of 10-Ks during the 1994 to 2008 period, they find that almost three-fourths of the words identified as negative by the widely used Harvard Dictionary are words typically not considered negative in financial contexts. This is also noticeable in our summary statistics outlined in **Table 2.** Specifically, applying this dictionary, the negative sentiment score is higher than the positive sentiments for the ICO whitepapers, which gives us the negative overall sentiment. Furthermore, this measure also shows an equal positive risk for sentiment uncertainty[9]. It is interesting to see whether the dictionaries developed for the finance-specific areas can capture the sentiment of this new digital financial market.

Fourth, the Quantitative Discourse Analysis Package (QDAP) provides quantitative analysis of qualitative transcripts and therefore bridges the gap between quantitative and qualitative research approaches. It overlaps with natural language processing and text mining. We also get a similar pattern using QDAP, that is, on average, positive sentiment is higher than negative sentiments in the sample of ICO whitepapers.

---

[9] Overall sentiment score for Loughran and McDonald = -0.014 and sentiment uncertainty risk = 0.013 (see Table 2.)

3.3.3. On encountering Déjàvu: Is the whitepaper subject to plagiarism?

Next, we are interested in analyzing whether providing a similar whitepaper in terms of tokenized contents (words/sentences/paragraphs) is either beneficial or harmful for the projects in terms of attaining funding. Similarities in whitepapers can be beneficial to the investors as familiar words and sentences may increase confidence. On the other side, investors can accuse the project of plagiarism, if the content is highly similar. In our study, we use the *Jaccard similarity* (TextReuse R package) as a proxy for the similarity of whitepapers. The function jaccard_similarity provides the Jaccard measures of similarity for two sets. The coefficients will be numbers between 0 and 1. For the similarity coefficient, the higher the figure the more similar the two sets of whitepapers. We make over 25 million pairwise comparisons (5,033x5,033= 25,331,089) and get bi-directional comparisons. Note that we did not use the function *jaccard_bag_similarity* which provides only a one-directional comparison. We found that, on average, one ICO whitepaper is around 6% similar to all the other whitepapers. After excluding the self-matched pair there are still some ICO whitepapers fully copied from another project with a different name, which have a Jaccard similarity score of 1.

3.3.4. Media attention and social media

Another important variable that we consider is the *Hype Score,* which shows the level of interest in the project from potential investors. The values High, Medium and Low are calculated based on the number of users on the project pages on social media. Social media includes the websites *bitcoin forum* and *telegram*. The higher the value, the more people are interested in the project, which indicates a potentially high demand for the tokens. This score is available on icorating.com for free for the majority of planned, ongoing and completed ICOs. We use dummy variables to indicate the level of *Hype Score* for high, medium and low as well as for those ICOs with no rated Hype Score. The summary statistics reported in **Table 2.** show that 15.1% of ICOs have high, 28% have medium, 25% have low and 32.1% of ICOs have no rated hype score. This implies that the minority of the ICOs have a large number of social media followers.

Nowadays, social media like Telegram and Bitcoin forum are important tools for communication and Twitter is valuable for mass communication. The number of Twitter followers of a respective ICO project can indicate the popularity of the project. We consider the number of *Twitter Followers* as a potential variable that can be associated with the raised amount of funding. We mark ICOs with no Twitter account and assign them 0 followers. The summary statistics reported in **Table 2.** show that, on average, one ICO has around 4,000 Twitter followers.

3.3.5. Risk, disclosure, industrial sectors

Similar to the *Hyper Score,* the website icorating.com also provides the *Risk Score* of the project. It is used to assess risks of potential fraud, as well as the overall quality of project development. This variable determines the reliability of a project against aspects such as its team, the product, the existence of partners and so on. The value from low to high shows the risk of fraud from small to large. This is also a measure of the project's investment attractiveness. This score is available on icorating.com for free for the majority of planned, ongoing and completed ICOs. Again, we use dummy variables to indicate the level of a risk score for high, medium, low and also for those ICOs whose *Risk Score* is not rated. Our summary statistics reported in **Table 2.** show that 11% of ICOs have high, 18% have medium, 7% have low and 64% of ICOs have no rated risk score. Even ICOs exhibiting no rated risk score can have valuable information because these ICOs have a relatively low level of information disclosure.

Next, we extracted a Know Your Customer (KYC) score from all the ICOs with information on the raised amount from the popular ICO website ICObench.com. KYC is an essential method of verifying the identity of the teams and people. ICObench employs this procedure to verify the identity of the ICO/IEO team members and to verify experts and bloggers. Many bounty hunters participate in KYC campaigns in exchange for free tokens. However, a high KYC score is also a symbol of a higher level of trust in the project. The highest score of 5 shows that the project is highly reliable whereas a KYC campaign not undertaken (score=0) shows that there is a high risk that the project has fake teams and other stakeholders. The summary statistics reported in **Table 2.** indicate that our data sample exhibits a mean KYC score of 1.66, which is relatively low.

Next, in our list of variables, the *Number of Categories* is also a potentially important predictor variable. Investors might be interested in investing in a particular sector due to having prior knowledge in that specific industry. For instance, considering utility tokens, investors prefer to use platforms or services in some particular market segment. In this regard, it is important to note that one single ICO can be registered under multiple categories. As an example, our sample exhibits one ICO which has been registered in 11 different categories. **Figure 4** shows the top 20 different categories wherein ICOs have been registered. It is important to explore whether identifying an ICO as ICO related to some particular sector or general-purpose ICO provide information on the success in raising funding. Hypothetically, innovative business sectors related to platforms could be more appealing to the investors than finance sectors which can be considered traditional business sectors.

3.3.6. Where can we find you? The location of the project team

Another important variable that we consider is the team size. Some ICOs have a very large team size. For instance, the maximum number of people in a team in an ICO in our data sample is 45 (see **Table 2.**). The question arises whether a larger team is of support for increasing the amount of raised crowdfunding. Hypothetically, a larger group of developers could be considered more competent than a smaller group of developers. We also seek to answer this question by considering this variable in our analysis. ICOs with no team information disclosure is assigned 0 team members.

Further, we consider some important dummy variables in our regression model. Some ICO issuers do not disclose the issuers' home country. This information might be crucial to the investors as ICOs from certain countries have been revealed as scams due to poor domestic financial regulation. The Transparency Index of a country measuring the level of corruption and the gross domestic product as a measure of the level of economic activities might be an indication of the trust and prospect of an ICO. Only 50% of ICOs in our data sample have disclosed their home country. Hypothetically, concealing the country of origin could indicate a deliberate act of possible fraud. If the ICO has disclosed the country of issuance it receives a dummy variable of 1 and otherwise a variable of 0. The website icosbull.com provides information on the issuing country of past ICOs.

3.3.7. How does the ICO want to proceed and what does this progress require?

Next, we assign a dummy variable for those ICOs that have a clear *Roadmap/Milestones* mentioned in their whitepaper. A roadmap or milestone shows the prospect of the ICO and, hence, may serve as a strong quality signal. A clear plan might be the indication of a legitimate ICO. 72% of ICOs in our data sample exhibit a clear roadmap/milestone that is elaborately mentioned in the whitepaper (see **Table 2.**).

Moreover, we define the dummy variables *softcap* and *hardcap.* Specifically, the softcap is the lower limit and the hardcap is the upper limit of the required funding for an ICO. If a team receives funding exceeding the hardcap, it should be returned to the investors. Failing to do so is a red flag for a possible scam. Disclosure of required capital helps investors to monitor the progress of the ICO in collecting funds. Many ICO issuers do not disclose this information, meaning that investors do not have any prenotion of whether the project is going to be successful. Hiding this essential information could hypothetically help to generate a continuous flow of funds as the investors are unaware of whether the required capital limit is reached. Only around one-fourth of the ICOs have disclosed softcap information whereas around one-third have disclosed the hardcap information in our data sample. Again, it might be an indication of a scam if they are withholding the information about soft and hard capital. The website icosbull.com provides data on the softcap and hardcap (see the information on the financial view of the past ICOs). Similarly, icorating.com has information on the capital goal. We also used text mining to see if the softcap and hardcap information exists in the whitepaper. ICOs with this information disclosure are allocated a dummy variable of 1 and those without, receive a dummy variable of 0.

Finally, the *Number of Tokens Disclosed* is the last important dummy variable used in our analysis. This variable exhibits crucial information, that is, whether the token is heavily distributed among developers during the pre-sale phase or not. One can find the pre-distribution percentage by simply taking the ratio of pre-sale tokens and the total number of tokens. If the total number of tokens is not disclosed beforehand, it could be an indication of a scam (i.e., 'pump-and-dump'). In this regard, Grobys and Sapkota (2020) study potential determining factors for cryptocurrency default and they show that high levels of pre-mining could potentially be a get-rich-quick scheme on the part of the developers, rather than setting up the coin for long-term success and ultimately leading to default. Descriptive statistics (**Table 2.**) show that only one-third of the ICOs have disclosed information on the number of tokens.

3.4. Statistical model

In this paper, we focus on three major features associated with ICOs. The first is the level of information disclosures in terms of the availability of necessary information on the whitepaper itself. We address this feature by using variables such as Roadmap/Milestone, softcap, hardcap and disclosure of tokens numbers.

Second, this paper quantifies the qualitative aspects of whitepapers such as sentiment and readability. We address this feature by using four different sentiment measures following *(i) Harvard-IV General Purpose Psychological Dictionary, (ii) Henry's Finance Dictionary, (iii) Loughran Mc Donald Finance-specific Dictionary* and, *(iv) QDAP* in association with seven different readability indices, which are, *(i) Flesh, (ii) Flesh-Kincaid., (iii) ARI, (iv) Coleman-Liau, (v) Gunning Fog, (vi) SMOG*, and *(vii) RIX.*

Third, we also account for the characteristics of the ICO project found outside of the whitepapers such as social media followers, possible scams and KYC score. We apply the multiple linear regression model based on pooled ordinary least squares for the parameter estimation given by equations (1) to (4) for four sentiment dictionaries.

$$\begin{aligned}\ln\left(Raised_{i,j}\right) = \beta_0 &+ \beta_1 WC_i + \beta_2 PL_i + \beta_3 N.GI_i + \beta_4 P.GI_i + \beta_5 Read_{i,j} + \beta_6 Sim_i \\ &+ \beta_7 KYC_i + \beta_8 Twt_i + \beta_9 Team_i + \beta_{10} H.H_i + \beta_{11} M.H_i + \beta_{12} L.H_i \\ &+ \beta_{13} H.NR_i + \beta_{14} H.R_i + \beta_{15} M.R_i + \beta_{16} L.R_i + \beta_{17} R.NR_i + \beta_{18} CD_i \\ &+ \beta_{19} NoC_i + \beta_{20} RM.MS_i + \beta_{21} SC_i + \beta_{22} HC_i + \beta_{23} NoT_i + \in_i \end{aligned} \tag{1}$$

Equation (1) is the regression model employing the Harvard-IV General Purpose Psychological Dictionary. Moreover, $\ln\left(Raised_{i,j}\right)$ is the log of raised amount for each ICO i, for seven different readability measures j. The independent variables in this equation are WC_i (word count), PL_i (page length), $N.GI_i$ (negative Harvard psychological sentiment), $P.GI_i$ (positive Harvard psychological sentiment), $Read_{i,j}$ (seven different Readability measures from model (1) - model (7)), Sim_i (similarity score), KYC_i (know your customer score), Twt_i (Twitter followers), $Team_i$ (team size), $H.H_i$ (hype score dummy, high), $M.H_i$ (hyper score dummy, medium), $L.H_i$ (hype score dummy, low), $H.NR_i$ (hype score dummy, not rated), $H.R_i$ (risk score dummy, high), $M.R_i$ (risk score dummy, medium), $L.R_i$ (risk score dummy, low), $R.NR_i$ (risk score dummy, not rated), CD_i (country disclosure dummy), NoC_i (number of categories), $RM.MS_i$ (roadmap/milestone disclosure dummy), SC_i (softcap disclosure dummy), HC_i (hardcap disclosure dummy), NoT_i (number of tokens disclosure dummy).

In equation (2), our regression model accounts for Henry's Finance Dictionary, where, $N.HE_i$ is scores for Henry's negative finance sentiment and $P.HE_i$ is scores for Henry's positive finance sentiment.

$$\ln(Raised_{i,j}) = \beta_0 + \beta_1 WC_i + \beta_2 PL_i + \beta_3 N.HE_i + \beta_4 P.HE_i + \beta_5 Read_{i,j} + \beta_6 Sim_i + \beta_7 KYC_i + \beta_8 Twt_i + \beta_9 Team_i + \beta_{10} H.H_i + \beta_{11} M.H_i + \beta_{12} L.H_i + \beta_{13} H.NR_i + \beta_{14} H.R_i + \beta_{15} M.R_i + \beta_{16} L.R_i + \beta_{17} R.NR_i + \beta_{18} CD_i + \beta_{19} NoC_i + \beta_{20} RM.MS_i + \beta_{21} SC_i + \beta_{22} HC_i + \beta_{23} NoT_i + \in_i \quad (2)$$

Similarly, in equation (3), we employ the Loughran Mc Donald Finance-specific Dictionary to assess the whitepaper sentiment measures,

$$\ln(Raised_{i,j}) = \beta_0 + \beta_1 WC_i + \beta_2 PL_i + \beta_3 N.LM_i + \beta_4 P.LM_i + \beta_5 RU.LM_i + \beta_6 Read_{i,j} + \beta_7 Sim_i + \beta_8 KYC_i + \beta_9 Twt_i + \beta_{10} Team_i + \beta_{11} H.H_i + \beta_{12} M.H_i + \beta_{13} L.H_i + \beta_{14} H.NR_i + \beta_{15} H.R_i + \beta_{16} M.R_i + \beta_{17} L.R_i + \beta_{18} R.NR_i + \beta_{19} CD_i + \beta_{20} NoC_i + \beta_{21} RM.MS_i + \beta_{22} SC_i + \beta_{23} HC_i + \beta_{24} NoT_i + \in_i, \quad (3)$$

where, $N.LM_i$ defines the negative sentiment as measured via Loughran Mc Donald sentiment dictionary, $P.LM_i$ defines the corresponding positive sentiment, and $RU.LM_i$ defines the measure for the corresponding sentiment uncertainty risk.

Furthermore, we also account for the Qualitative Discourse Analysis Package (QDAP) for sentiment polarity in equation (4) below.

$$\ln(Raised_{i,j}) = \beta_0 + \beta_1 WC_i + \beta_2 PL_i + \beta_3 N.QDAP_i + \beta_4 P.QDAP_i + \beta_5 Read_{i,j} + \beta_6 Sim_i + \beta_7 KYC_i + \beta_8 Twt_i + \beta_9 Team_i + \beta_{10} H.H_i + \beta_{11} M.H_i + \beta_{12} L.H_i + \beta_{13} H.NR_i + \beta_{14} H.R_i + \beta_{15} M.R_i + \beta_{16} L.R_i + \beta_{17} R.NR_i + \beta_{18} CD_i + \beta_{19} NoC_i + \beta_{20} RM.MS_i + \beta_{21} SC_i + \beta_{22} HC_i + \beta_{23} NoT_i + \in_i \quad (4)$$

where, $N.QDAP_i$ is the score for the negative QDAP sentiment polarity and $P.QDAP_i$ is the score for the positive QDAP sentiment polarity.

4. RESULTS

4.1. Which country is the leading country in terms of raised funding or the number of launched ICOs?

We observe from **Figure 2.** that the U.S. acquired the largest amount of raised funding and is at the same time the leading country in terms of the number of launched ICOs. This is an interesting finding because in the IPO market a somewhat reverse picture was recently presented. For instance, in 2019, there were 404 IPOs in China but only 232 in the U.S. We clustered ICOs into 20 distinct industries. In this regard, it is noteworthy that one ICO can be in several industry sectors. From **Figure 2.** we observe that more than half of the ICOs produce products related to four sectors, which are platforms, business and services, cryptocurrencies and big data. In **Table 1.** we report the variables that we were able to identify using the information provided in whitepapers in association with various additional internet websites as illustrated in **Figure 1**.

4.2. Descriptive statistics

In **Table 2.** we report the descriptive statistics of our variables. For instance, from **Table 2.** we observe that based on N=1,507 whitepapers used in our analysis, a whitepaper has on average 4,338 words. The minimum number of words is 241 and the maximum number of words is 41,532. Next, let us consider the number of categories. We observe that the average ICO's product is associated with 1.65 categories. The minimum number of categories is one, whereas the maximum number of categories is as much as eleven. The raised amount shows a very high standard deviation among ICOs, where the minimum raised amount is 50 dollars and the maximum amount corresponds to 4 billion dollars. The average funding amount is 20 million dollars. However, we note that the distribution of the raised funding is highly skewed and some few projects raised billions of dollars[10].

[10] The maximum amount of raised funding is $4.2 billion.

4.3. What do we learn from analyzing our regression models?

4.3.1. Does sentiment have an impact on the success of ICOs?

We start our statistical analysis using a simple OLS regression incorporating the sentiment measured by the Harvard Sentiment Dictionary. The results are reported in **Table 3.** and show some interesting findings. First and most importantly, we find that only negative sentiment is significant. Specifically, the more negative the sentiment the larger the predicted amount of raised funding, whereas positive sentiment does not have any significant effects.

Table 3. Regression Result with Harvard Sentiment Dictionary (GI), Readability and Other ICO Characteristics

This table reports the OLS regression result with 7 different readability measures in the model (1) to (7) with Flesch, Flesch. K, RIX, SMOG, FOG, ARI, CoL readability scores respectively, incorporating the Harvard General Psychology Sentiment Dictionary. Negative.GI Sentiment is the negative psychological sentiment and Positive.GI Sentiment is the positive psychological sentiment. This table also accounts for social media and bitcoin talk forum hype as High, Medium, and Low Hype Score. Potential fraud risk is measured as High, Medium, and Low-Risk Score. Teams, Twitter followers, Token Economics and other whitepaper disclosures are the other explanatory variables.

	Dependent variable:						
	ln(Raised)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Word Count	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002
	(0.619)	(0.623)	(0.610)	(0.642)	(0.631)	(0.613)	(0.612)
Page Lengths	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(-0.461)	(-0.448)	(-0.458)	(-0.455)	(-0.448)	(-0.441)	(-0.473)
Negative.GI Sentiment	6.856**	6.741**	6.870**	6.800**	6.746**	6.764**	6.920**
	(2.522)	(2.481)	(2.534)	(2.507)	(2.485)	(2.497)	(2.556)
Positive.GI Sentiment	0.523	0.443	0.520	0.535	0.453	0.450	0.565
	(0.439)	(0.371)	(0.437)	(0.449)	(0.380)	(0.377)	(0.463)
Flesch Readbility Score	0.001						
	(0.307)						
Flesch.K Readbility Score		-0.009					
		(-0.665)					
RIX Readbility Score			-0.006				
			(-0.336)				
SMOG Readbility Score				-0.013			
				(-0.547)			
FOG Readbility Score					-0.009		
					(-0.681)		
ARI Readbility Score						-0.008	
						(-0.752)	
Col Readbility Score							0.001
							(0.181)
Jaccard Similarity Score	0.472	0.475	0.472	0.475	0.476	0.476	0.472
	(1.350)	(1.359)	(1.351)	(1.357)	(1.361)	(1.360)	(1.348)
High Hype Score	1.030***	1.032***	1.030***	1.030***	1.032***	1.032***	1.028***
	(3.498)	(3.504)	(3.499)	(3.498)	(3.504)	(3.506)	(3.491)
Medium Hype Score	0.754***	0.758***	0.755***	0.755***	0.758***	0.758***	0.752***
	(2.695)	(2.707)	(2.697)	(2.697)	(2.707)	(2.709)	(2.687)
Low Hype Score	0.184	0.185	0.184	0.183	0.185	0.186	0.183
	(0.664)	(0.666)	(0.664)	(0.661)	(0.667)	(0.670)	(0.661)
Hype Not Rated	0.637**	0.640**	0.638**	0.637**	0.640**	0.641**	0.636**
	(2.274)	(2.282)	(2.276)	(2.272)	(2.283)	(2.286)	(2.268)
High Risk Score	-0.175	-0.169	-0.169	-0.163	-0.168	-0.162	-0.178
	(-0.304)	(-0.293)	(-0.294)	(-0.282)	(-0.291)	(-0.280)	(-0.309)
Medium Risk Score	-0.340	-0.337	-0.335	-0.329	-0.337	-0.331	-0.340

	(-0.601)	(-0.598)	(-0.593)	(-0.583)	(-0.597)	(-0.586)	(-0.602)
Low Risk Score	0.046	0.048	0.051	0.056	0.047	0.054	0.045
	(0.078)	(0.082)	(0.087)	(0.095)	(0.081)	(0.093)	(0.077)
Risk Not Rated	-0.319	-0.315	-0.314	-0.308	-0.316	-0.309	-0.321
	(-0.567)	(-0.561)	(-0.558)	(-0.548)	(-0.562)	(-0.549)	(-0.570)
Number of Categories	-0.009	-0.009	-0.009	-0.009	-0.009	-0.009	-0.008
	(-0.228)	(-0.237)	(-0.225)	(-0.233)	(-0.237)	(-0.242)	(-0.219)
Team.Size	0.013*	0.013*	0.013*	0.013*	0.013*	0.013*	0.013*
	(1.869)	(1.863)	(1.870)	(1.867)	(1.862)	(1.865)	(1.872)
County Disclosed	-0.214	-0.219	-0.214	-0.216	-0.219	-0.219	-0.211
	(-1.147)	(-1.174)	(-1.149)	(-1.161)	(-1.174)	(-1.177)	(-1.133)
KYC Score	0.104**	0.106**	0.104**	0.105**	0.106**	0.106**	0.103*
	(1.976)	(2.005)	(1.979)	(1.992)	(2.005)	(2.007)	(1.961)
Twitter Followers	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**
	(2.347)	(2.368)	(2.352)	(2.348)	(2.369)	(2.376)	(2.333)
RoadMap/Milestone Stated	0.071	0.072	0.071	0.073	0.072	0.072	0.071
	(0.711)	(0.716)	(0.709)	(0.726)	(0.721)	(0.714)	(0.707)
SoftCap Given	-0.243	-0.241	-0.243	-0.242	-0.242	-0.241	-0.244
	(-1.577)	(-1.564)	(-1.573)	(-1.566)	(-1.567)	(-1.560)	(-1.585)
HardCap Given	0.098	0.098	0.097	0.098	0.098	0.097	0.097
	(0.671)	(0.672)	(0.668)	(0.676)	(0.673)	(0.669)	(0.669)
Number of Tokens Given	0.165	0.163	0.165	0.164	0.163	0.163	0.166
	(1.374)	(1.359)	(1.372)	(1.364)	(1.360)	(1.357)	(1.384)
Constant	14.428***	14.617***	14.509***	14.660***	14.638***	14.594***	14.414***
	(20.742)	(20.502)	(21.090)	(19.276)	(20.300)	(20.900)	(19.088)
N	1,507	1,507	1,507	1,507	1,507	1,507	1,507
R^2	0.056	0.056	0.056	0.056	0.056	0.056	0.056
Adjusted R^2	0.041	0.042	0.041	0.042	0.042	0.042	0.041
Residual Std. Error (df = 1483)	1.745	1.744	1.745	1.745	1.744	1.744	1.745
F Statistic (df = 23; 1483)	3.832***	3.848***	3.833***	3.841***	3.849***	3.854***	3.829***

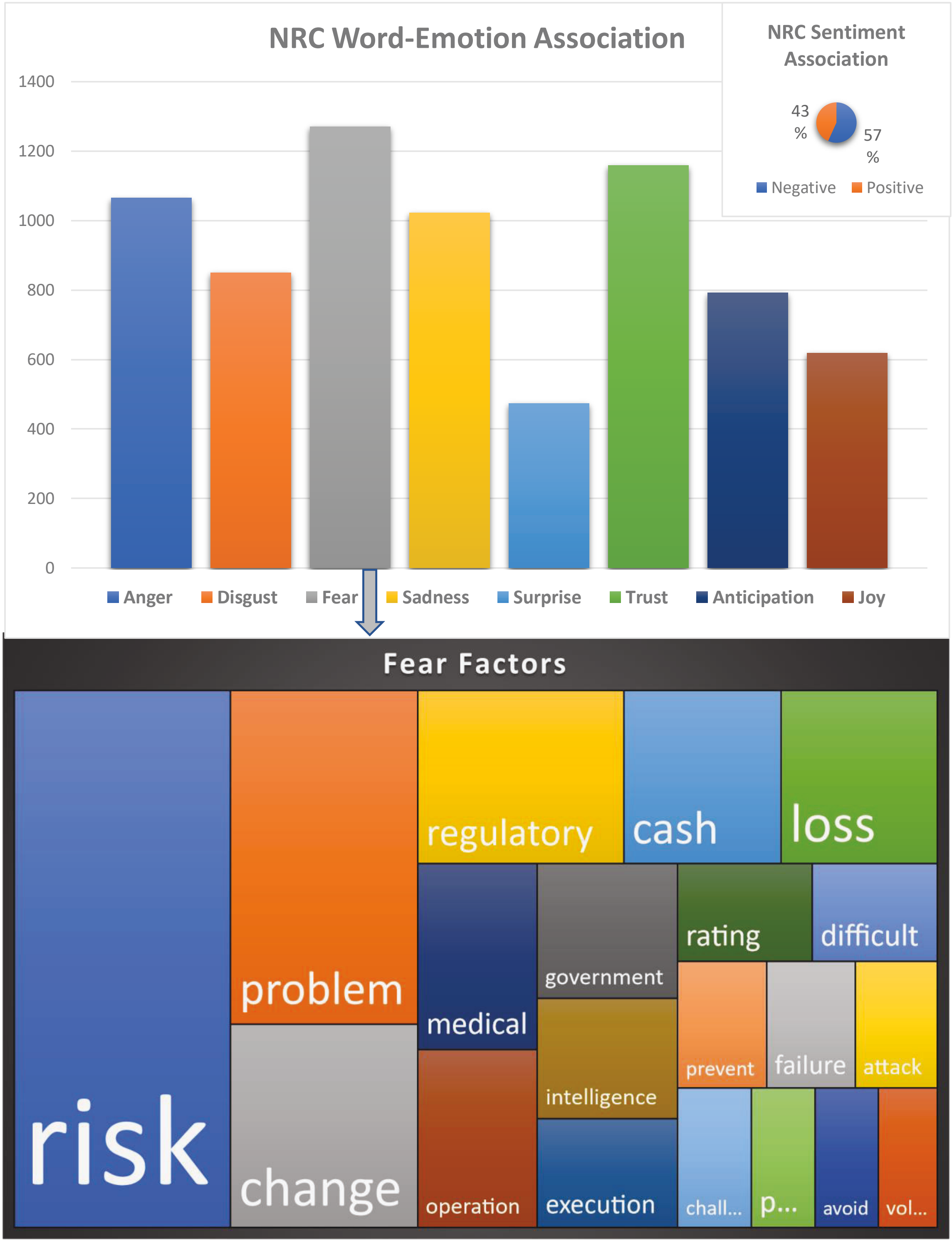
Note: Significance Levels

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
t-stats reported in the parenthesis.

Furthermore, to get a deeper understanding of what emotion is mostly associated with the negative or positive sentiment we followed the NRC Word-Emotion Association Lexicon by Mohammad and Turney (2013) which is also known as EmoLex[11]. It is a list of English words and their associations with eight basic emotions, which are anger, fear, anticipation, trust, surprise, sadness, joy, and disgust. Using EmoLex also enables us to capture negative and positive sentiments in the text. **Figure 6.** shows a histogram and a pie chart of sentiment and its corresponding emotion in the whitepapers of our data sample. We observe from **Figure 6.** that 57% of the overall text in the whitepaper has negative sentiment. Of this negative sentiment, fear is the most frequently identified emotion associated with the whitepapers. This might be the indication of an ICO marketing strategy, where customers are attracted by the trigger of fear. We further explored all the words associated with 'fear' in the NRC Lexicon and track them in the corpus of whitepapers and find that the ICO whitepapers (2014-2019) are selling '*fear of risk*', '*fear of change*', '*fear of problem*', '*fear of regulation*', '*fear of loss*' and other fears that trigger the negative emotions.

[11] see more at https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

Fig. 6. Unique Words of ICO Whitepapers (2014-2019) and NRC Word-Emotion Association



4.3.2. Is it important whether or not the whitepaper content is easy to read?

The regression result reported in **Table 4.** to **Table 6.** incorporates seven different models with seven different readability scores. Our findings strongly suggest that readability is irrelevant irrespective of how we measure it, whereas negative Harvard psychosocial sentiment is statistically significant across all model specifications.

Table 4. Regression Result with Henry's Finance Dictionary (HE), Readability and Other ICO Characteristics

This table reports the OLS regression result with 7 different readability measures in the model (1) to (7) with Flesch, Flesch. K, RIX, SMOG, FOG, ARI, and CoL readability scores respectively, incorporating Henry's Finance Sentiment Dictionary. Negative.HE Sentiment is the negative financial sentiment and Positive.HE Sentiment is the positive financial sentiment. This table also accounts for social media and bitcoin talk forum hype as High, Medium, and Low Hype Scores. Potential fraud risk is measured as a High, Medium, and Low-Risk Score. Teams, Twitter followers, Token Economics and other whitepaper disclosures are the other explanatory variables.

	Dependent variable:						
	ln(Raised)						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Word Count	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002
	(0.785)	(0.785)	(0.771)	(0.812)	(0.796)	(0.770)	(0.773)
Page Lengths	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.003
	(-0.605	(-0.587)	(-0.606)	(-0.604)	(-0.588)	(-0.584)	(-0.640)
Negative.HE Sentiment	4.209	3.630	3.900	4.401	3.899	3.783	4.095
	(0.338)	(0.292)	(0.314)	(0.353)	(0.314)	(0.304)	(0.328)
Positive.HE Sentiment	3.688	3.358	3.818	3.762	3.395	3.454	3.896
	(0.598)	(0.544)	(0.620)	(0.611)	(0.550)	(0.560)	(0.631)
Flesch Readbility Score	0.003						
	(0.656)						
Flesch.K Readbility Score		-0.014					
		(-1.091)					
RIX Readbility Score			-0.011				
			(-0.627)				
SMOG Readbility Score				-0.020			
				(-0.844)			
FOG Readbility Score					-0.014		
					(-1.084)		
ARI Readbility Score						-0.011	
						(-1.096)	
Col Readbility Score							0.002
							(0.209)
Jaccard Similarity Score	0.468	0.473	0.468	0.471	0.474	0.472	0.466
	(1.333)	(1.349)	(1.335)	(1.342)	(1.350)	(1.347)	(1.329)
High Hype Score	1.044^{***}	1.046^{***}	1.045^{***}	1.043^{***}	1.046^{***}	1.046^{***}	1.041^{***}
	(3.537)	(3.543)	(3.540)	(3.534)	(3.543)	(3.546)	(3.526)
Medium Hype Score	0.777^{***}	0.781^{***}	0.778^{***}	0.776^{***}	0.780^{***}	0.781^{***}	0.773^{***}

	(2.769)	(2.784)	(2.772)	(2.768)	(2.783)	(2.784)	(2.756)
Low Hype Score	0.194	0.194	0.194	0.192	0.195	0.196	0.193
	(0.699)	(0.700)	(0.700)	(0.693)	(0.701)	(0.705)	(0.696)
Hype Not Rated	0.666**	0.668**	0.667**	0.665**	0.669**	0.670**	0.665**
	(2.371)	(2.381)	(2.377)	(2.367)	(2.382)	(2.385)	(2.367)
High Risk Score	-0.171	-0.165	-0.162	-0.155	-0.164	-0.156	-0.179
	(-0.297)	(-0.286)	(-0.280)	(-0.268)	(-0.284)	(-0.271)	(-0.311)
Medium Risk Score	-0.334	-0.334	-0.326	-0.320	-0.334	-0.325	-0.338
	(-0.590)	(-0.590)	(-0.576)	(-0.566)	(-0.590)	(-0.575)	(-0.597)
Low Risk Score	0.040	0.040	0.048	0.053	0.039	0.049	0.036
	(0.068)	(0.069)	(0.082)	(0.091)	(0.068)	(0.084)	(0.062)
Risk Not Rated	-0.324	-0.322	-0.317	-0.311	-0.323	-0.313	-0.331
	(-0.576)	(-0.571)	(-0.561)	(-0.551)	(-0.574)	(-0.556)	(-0.586)
Number of Categories	-0.008	-0.009	-0.008	-0.008	-0.008	-0.009	-0.007
	(-0.210)	(-0.220)	(-0.200)	(-0.209)	(-0.219)	(-0.222)	(-0.184)
Team.Size	0.012*	0.012*	0.012*	0.012*	0.012*	0.012*	0.012*
	(1.770)	(1.765)	(1.772)	(1.768)	(1.763)	(1.768)	(1.771)
County Disclosed	-0.205	-0.212	-0.205	-0.208	-0.212	-0.211	-0.199
	(-1.099)	(-1.134)	(-1.097)	(-1.114)	(-1.134)	(-1.132)	(-1.070)
KYC Score	0.104**	0.106**	0.104**	0.105**	0.106**	0.106**	0.102*
	(1.962)	(2.001)	(1.963)	(1.977)	(1.999)	(1.996)	(1.929)
Twitter Followers	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**
	(2.435)	(2.463)	(2.446)	(2.429)	(2.463)	(2.469)	(2.409)
RoadMap/Milestone Stated	0.072	0.073	0.072	0.074	0.073	0.072	0.071
	(0.720)	(0.723)	(0.714)	(0.737)	(0.730)	(0.718)	(0.706)
SoftCap Given	-0.238	-0.235	-0.237	-0.236	-0.236	-0.235	-0.241
	(-1.540)	(-1.520)	(-1.536)	(-1.529)	(-1.525)	(-1.518)	(-1.558)
HardCap Given	0.106	0.106	0.105	0.107	0.106	0.105	0.106
	(0.729)	(0.725)	(0.722)	(0.733)	(0.726)	(0.720)	(0.724)
Number of Tokens Given	0.149	0.147	0.149	0.148	0.147	0.147	0.152
	(1.241)	(1.222)	(1.239)	(1.234)	(1.225)	(1.224)	(1.260)
Constant	14.812***	15.116***	14.976***	15.195***	15.145***	15.065***	14.857***
	(22.123)	(22.436)	(22.840)	(20.898)	(22.212)	(22.735)	(20.775)
N	1,507	1,507	1,507	1,507	1,507	1,507	1,507
R^2	0.051	0.052	0.051	0.052	0.052	0.052	0.051
Adjusted R^2	0.037	0.037	0.037	0.037	0.037	0.037	0.036
Residual Std. Error (df = 1483)	1.749	1.749	1.749	1.749	1.749	1.749	1.749
F Statistic (df = 23; 1483)	3.490***	3.524***	3.488***	3.502***	3.524***	3.525***	3.472***

Significance Levels

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
t-stats reported in the parenthesis.

Table 5. Regression Model with Loughran Mc Donald Finance-specific Dictionary (LM), Readability and Other ICO Characteristics.

This table reports the OLS regression result with 7 different readability measures in the model (1) to (7) with Flesch, Flesch. K, RIX, SMOG, FOG, ARI, and CoL readability scores respectively, incorporating the Loughran Mc Donald Finance-specific Dictionary. Negative.LM Sentiment is the negative financial sentiment, Positive.LM Sentiment is the positive financial sentiment, and Uncertain.LM Sentiment is the uncertainty in the sentiment. This table also accounts for social media and bitcoin talk forum hype as High, Medium, and Low Hype Scores. Potential fraud risk is measured as a High, Medium, and Low-Risk Score. Teams, Twitter followers, Token Economics and other whitepaper disclosures are the other explanatory variables.

	Dependent variable:						
	ln(Raised)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Word Count	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002
	(0.797)	(0.796)	(0.782)	(0.822)	(0.806)	(0.781)	(0.786)
Page Lengths	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	(-0.655)	(-0.642)	(-0.655)	(-0.654)	(-0.642)	(-0.637)	(-0.686)
Negative.LM Sentiment	-2.648	-2.840	-2.586	-2.717	-2.841	-2.766	-2.391
	(-0.778)	(-0.833)	(-0.761)	(-0.798)	(-0.833)	(-0.813)	(-0.706)
Positive.LM Sentiment	5.638	5.285	5.666	5.642	5.318	5.372	5.827
	(1.203)	(1.125)	(1.209)	(1.204)	(1.132)	(1.145)	(1.234)
Uncertain.LM Sentiment	3.845	3.615	3.708	4.074	3.810	3.650	3.750
	(0.488)	(0.459)	(0.471)	(0.516)	(0.484)	(0.464)	(0.476)
Flesch Readbility Score	0.003						
	(0.702)						
Flesch.K Readbility Score		-0.015					
		(-1.109)					
RIX Readbility Score			-0.012				
			(-0.660)				
SMOG Readbility Score				-0.021			
				(-0.886)			
FOG Readbility Score					-0.014		
					(-1.105)		
ARI Readbility Score						-0.011	
						(-1.110)	
Col Readbility Score							0.002
							(0.283)
Jaccard Similarity Score	0.435	0.442	0.436	0.438	0.442	0.441	0.434
	(1.239)	(1.257)	(1.240)	(1.247)	(1.258)	(1.255)	(1.234)
High Hype Score	1.049^{***}	1.051^{***}	1.049^{***}	1.048^{***}	1.051^{***}	1.052^{***}	1.044^{***}
	(3.550)	(3.559)	(3.552)	(3.547)	(3.559)	(3.561)	(3.534)
Medium Hype Score	0.787^{***}	0.792^{***}	0.788^{***}	0.787^{***}	0.792^{***}	0.792^{***}	0.782^{***}
	(2.805)	(2.822)	(2.807)	(2.804)	(2.821)	(2.821)	(2.786)
Low Hype Score	0.205	0.206	0.205	0.204	0.206	0.207	0.203
	(0.739)	(0.742)	(0.739)	(0.734)	(0.743)	(0.746)	(0.732)
Hype Not Rated	0.678^{**}	0.681^{**}	0.679^{**}	0.677^{**}	0.681^{**}	0.682^{**}	0.675^{**}
	(2.411)	(2.423)	(2.416)	(2.409)	(2.424)	(2.426)	(2.402)

High Risk Score	-0.182	-0.174	-0.173	-0.165	-0.173	-0.166	-0.191
	(-0.315)	(-0.302)	(-0.299)	(-0.285)	(-0.300)	(-0.288)	(-0.331)
Medium Risk Score	-0.342	-0.341	-0.334	-0.327	-0.340	-0.333	-0.346
	(-0.604)	(-0.602)	(-0.590)	(-0.577)	(-0.602)	(-0.587)	(-0.611)
Low Risk Score	0.034	0.035	0.042	0.048	0.035	0.044	0.029
	(0.057)	(0.060)	(0.071)	(0.082)	(0.059)	(0.075)	(0.050)
Risk Not Rated	-0.334	-0.331	-0.327	-0.320	-0.332	-0.323	-0.341
	(-0.593)	(-0.587)	(-0.579)	(-0.567)	(-0.589)	(-0.573)	(-0.605)
Number of Categories	-0.009	-0.009	-0.008	-0.009	-0.009	-0.009	-0.008
	(-0.229)	(-0.238)	(-0.219)	(-0.226)	(-0.236)	(-0.240)	(-0.204)
Team.Size	0.012*	0.012*	0.012*	0.012*	0.012*	0.012*	0.012*
	(1.773)	(1.766)	(1.775)	(1.771)	(1.764)	(1.770)	(1.776)
County Disclosed	-0.195	-0.202	-0.195	-0.198	-0.202	-0.201	-0.189
	(-1.045)	(-1.080)	(-1.043)	(-1.059)	(-1.080)	(-1.077)	(-1.015)
KYC Score	0.101*	0.103*	0.101*	0.102*	0.103*	0.103*	0.099*
	(1.914)	(1.952)	(1.916)	(1.930)	(1.951)	(1.947)	(1.880)
Twitter Followers	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**
	(2.472)	(2.501)	(2.481)	(2.466)	(2.501)	(2.506)	(2.440)
RoadMap/Milestone Stated	0.072	0.072	0.071	0.074	0.073	0.072	0.070
	(0.713)	(0.717)	(0.707)	(0.731)	(0.724)	(0.712)	(0.699)
SoftCap Given	-0.231	-0.228	-0.230	-0.229	-0.229	-0.228	-0.233
	(-1.490)	(-1.476)	(-1.486)	(-1.478)	(-1.479)	(-1.473)	(-1.506)
HardCap Given	0.100	0.100	0.099	0.101	0.100	0.099	0.099
	(0.685)	(0.684)	(0.678)	(0.687)	(0.684)	(0.679)	(0.679)
Number of Tokens Given	0.148	0.146	0.147	0.147	0.146	0.146	0.150
	(1.227)	(1.210)	(1.225)	(1.220)	(1.212)	(1.211)	(1.247)
Constant	14.800***	15.129***	14.975***	15.210***	15.159***	15.072***	14.820***
	(21.876)	(21.956)	(22.435)	(20.483)	(21.733)	(22.301)	(20.392)
N	1,507	1,507	1,507	1,507	1,507	1,507	1,507
R^2	0.052	0.053	0.052	0.052	0.053	0.053	0.052
Adjusted R^2	0.037	0.037	0.037	0.037	0.037	0.037	0.037
Residual Std. Error (df = 1482)	1.749	1.748	1.749	1.749	1.748	1.748	1.749
F Statistic (df = 24; 1482)	3.400***	3.433***	3.398***	3.413***	3.432***	3.433***	3.382***

Note: Significance Levels

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
t-stats reported in the parenthesis.

Table 6. Regression Model with QDAP Sentiment Dictionary, Readability and Other ICO Characteristics.

This table reports the OLS regression result with 7 different readability measures in the model (1) to (7) with Flesch, Flesch. K, RIX, SMOG, FOG, ARI, and CoL readability scores respectively, incorporating the Qualitative Discourse Analysis Package Dictionary. Negative.QDAP Sentiment is the negative discourse sentiment and Positive.QDAP Sentiment is the positive discourse sentiment. This table also accounts for social media and bitcoin talk forum hype as High, Medium, and Low Hype Scores. Potential fraud risk is measured as a High, Medium, and Low-Risk Score. Teams, Twitter followers, Token Economics and other whitepaper disclosures are the other explanatory variables.

	Dependent variable:						
	ln(Raised)						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Word Count	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002
	(0.818)	(0.814)	(0.804)	(0.845)	(0.825)	(0.802)	(0.812)
Page Lengths	-0.003	-0.002	-0.003	-0.003	-0.002	-0.002	-0.003
	(-0.612)	(-0.598)	(-0.612)	(-0.610)	(-0.599)	(-0.594)	(-0.640)
Negative.QDAP Sentiment	1.504	1.377	1.459	1.499	1.415	1.409	1.560
	(0.425)	(0.388)	(0.412)	(0.423)	(0.399)	(0.398)	(0.440)
Positive.QDAP Sentiment	2.205	2.052	2.214	2.209	2.070	2.080	2.351
	(1.363)	(1.260)	(1.369)	(1.367)	(1.273)	(1.280)	(1.437)
Flesch Readbility Score	0.002						
	(0.597)						
Flesch.K Readbility Score		-0.012					
		(-0.917)					
RIX Readbility Score			-0.010				
			(-0.556)				
SMOG Readbility Score				-0.019			
				(-0.786)			
FOG Readbility Score					-0.012		
					(-0.928)		
ARI Readbility Score						-0.010	
						(-0.949)	
Col Readbility Score							0.003
							(0.415)
Jaccard Similarity Score	0.466	0.471	0.466	0.469	0.471	0.470	0.465
	(1.330)	(1.343)	(1.331)	(1.339)	(1.345)	(1.342)	(1.327)
High Hype Score	1.033***	1.035***	1.033***	1.032***	1.035***	1.035***	1.029***
	(3.499)	(3.506)	(3.501)	(3.496)	(3.505)	(3.508)	(3.484)
Medium Hype Score	0.764***	0.768***	0.764***	0.763***	0.767***	0.768***	0.759***
	(2.721)	(2.736)	(2.723)	(2.720)	(2.735)	(2.736)	(2.704)
Low Hype Score	0.182	0.183	0.182	0.180	0.183	0.184	0.180
	(0.654)	(0.658)	(0.655)	(0.649)	(0.658)	(0.662)	(0.646)
Hype Not Rated	0.649**	0.652**	0.651**	0.648**	0.652**	0.653**	0.646**
	(2.311)	(2.322)	(2.316)	(2.308)	(2.323)	(2.325)	(2.299)
High Risk Score	-0.178	-0.173	-0.170	-0.163	-0.172	-0.165	-0.183
	(-0.308)	(-0.301)	(-0.295)	(-0.281)	(-0.298)	(-0.286)	(-0.317)

Medium Risk Score	-0.342	-0.342	-0.336	-0.329	-0.342	-0.335	-0.342
	(-0.605)	(-0.606)	(-0.594)	(-0.582)	(-0.605)	(-0.592)	(-0.604)
Low Risk Score	0.028	0.029	0.035	0.041	0.028	0.036	0.028
	(0.048)	(0.049)	(0.060)	(0.070)	(0.048)	(0.062)	(0.048)
Risk Not Rated	-0.330	-0.329	-0.324	-0.317	-0.329	-0.321	-0.332
	(-0.586)	(-0.583)	(-0.574)	(-0.563)	(-0.585)	(-0.570)	(-0.590)
Number of Categories	-0.009	-0.009	-0.009	-0.009	-0.009	-0.010	-0.009
	(-0.237)	(-0.242)	(-0.229)	(-0.237)	(-0.242)	(-0.245)	(-0.224)
Team.Size	0.012^{*}	0.012^{*}	0.012^{*}	0.012^{*}	0.012^{*}	0.012^{*}	0.012^{*}
	(1.785)	(1.780)	(1.787)	(1.783)	(1.778)	(1.783)	(1.788)
County Disclosed	-0.191	-0.197	-0.190	-0.194	-0.197	-0.197	-0.185
	(-1.022)	(-1.054)	(-1.020)	(-1.036)	(-1.054)	(-1.053)	(-0.991)
KYC Score	0.101^{*}	0.103^{*}	0.101^{*}	0.101^{*}	0.102^{*}	0.102^{*}	0.099^{*}
	(1.903)	(1.937)	(1.903)	(1.918)	(1.936)	(1.934)	(1.868)
Twitter Followers	0.00001^{**}	0.00001^{**}	0.00001^{**}	0.00001^{**}	0.00001^{**}	0.00001^{**}	0.00001^{**}
	(2.403)	(2.426)	(2.411)	(2.397)	(2.426)	(2.432)	(2.374)
RoadMap/Milestone Stated	0.072	0.072	0.071	0.074	0.073	0.072	0.071
	(0.716)	(0.718)	(0.710)	(0.733)	(0.724)	(0.715)	(0.707)
SoftCap Given	-0.230	-0.228	-0.229	-0.228	-0.229	-0.228	-0.232
	(-1.490)	(-1.477)	(-1.485)	(-1.479)	(-1.480)	(-1.474)	(-1.503)
HardCap Given	0.101	0.101	0.100	0.102	0.101	0.100	0.100
	(0.693)	(0.691)	(0.686)	(0.696)	(0.692)	(0.687)	(0.689)
Number of Tokens Given	0.148	0.146	0.148	0.147	0.146	0.146	0.149
	(1.228)	(1.215)	(1.227)	(1.221)	(1.216)	(1.215)	(1.243)
Constant	14.575^{***}	14.856^{***}	14.725^{***}	14.931^{***}	14.882^{***}	14.813^{***}	14.533^{***}
	(21.117)	(21.143)	(21.656)	(19.888)	(20.947)	(21.489)	(19.438)
Observations	1,507	1,507	1,507	1,507	1,507	1,507	1,507
R^2	0.053	0.053	0.053	0.053	0.053	0.053	0.052
Adjusted R^2	0.038	0.038	0.038	0.038	0.038	0.038	0.038
Residual Std. Error (df = 1483)	1.748	1.748	1.748	1.748	1.748	1.748	1.748
F Statistic (df = 23; 1483)	3.578^{***}	3.600^{***}	3.576^{***}	3.590^{***}	3.601^{***}	3.603^{***}	3.570^{***}

Significance Levels

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
t-stats reported in the parenthesis.

Another interesting and somewhat surprising finding is that neither the number of words nor the length or the similarity is associated with raised funding. This suggests that investors do not critically review the whitepaper. If investors paid more attention to scrutinizing whitepapers, the risk of deception could be decreased. This is an important issue because the vast majority of ICOs are scams.[12] Our results also show that the more social media attention an ICO receives, the higher the predicted amount of raised funding. This is another issue which may suggest that ICO investors are guided by emotional experiences rather than critical reasoning. This view may be additionally substantiated by the insignificance of the risk scores, as risk scores do not have an impact on the amount of raised funding. Risky investments should be priced differently from less risky investments, but we do not find evidence to support this in our study. One important finding about social media followers, which is captured by the variable 'Hype Score', shows that not rated ICOs tend to be more successful in raising funds than those with a lower number of social media followers. Specifically, ICOs with no social media account are more successful in raising funds than ICOs with very few social media followers. Investors may perceive the number of social media followers as a proxy for the popularity of a project. A project with a smaller number of followers may imply that the project is less popular among other investors. Therefore, the ICOs need to have excellent social media marketing strategies from the very beginning. On the other hand, avoiding the usage of social media channels for marketing an ICO project may result in investors searching for other factors, such as *team members*, to assess the quality of the project.

Considering that Howell el al. (2020) document that ICO success is associated with disclosure, credible commitment to the project, and quality signals, it is surprising that our evidence gives a somewhat mixed picture: Specifically, county disclosure, road map or explicitly stated softcap

[12] The Satis Research Group report of 11 July 2018 investigated approximately 1,500 ICOs whereof 78% were *identified scams*, corresponding to a monetary equivalent in terms of US dollar of $1.3B. On the other hand, slightly more than $8B (~70% of ICO fundraising) was allocated to those that moved on to trade on an exchange. Even though the vast majority of funding was funneled to ICOs that proceeded to trade, about 1,170 out of 1,500 projects were revealed as fraud. The most well-known ICO scams are Pincoin, Arisebank and Savedroid, that illicitly obtained $660M, $600M, and $50M, respectively.

or hardcap are all quality signals but our findings indicate that none of these is associated with ICO success in terms of raised funding. On the other hand, it is not surprising that our results show that Twitter followers and signature campaigns, as measured by the KYC score, have an effect. Both variables are positively correlated with raised funding. Again, our results suggest that ICO investors base their decisions on attention signals and are attracted to ICOs that are frequently advertised on social media.

Next, we use the sentiment measured by Henry's Business Communication Dictionary, Loughran Mc Donald's Finance-specific Dictionary, and the QDAP Sentiment Polarity Dictionary. For each mode, we use various readability measures. The results are reported in **Table 5., Table 6., and Table 7.** The main difference between the results reported in **Table 5., Table 6., and Table 7.** as opposed to those reported in **Table 3.** is that the sentiment measured by those sentiment dictionaries is statistically insignificant. However, the statistical significances of all those other variables as discussed earlier do not change and, impressively, the point estimates are virtually the same. This result suggests that the sentiment incorporated in whitepapers cannot be cached by Henry's Business Communication Dictionary, Loughran Mc Donald's Finance-specific Dictionary or the QDAP Sentiment Polarity Dictionary. This is an interesting issue because the Loughran Mc Donald Finance-specific Dictionary has been exclusively created because of the inability of standard sentiment dictionaries to measure sentiment in finance-specific contexts.

4.4. Additional robustness check implementing Artificial Neural Networks (ANN).

A textural analysis is either Lexicon-based or Machine Learning based.[13] As the main tool, we followed Lexicon-based (i.e. Readability Measure and Dictionary-based). The Machine Learning based textural analysis includes Naïve Bayes, Support Vector Machine, Semantic Analysis and,

[13] How the choices of our approach are in line with Textural Analysis is illustrated in **Appendix A.8.**

Neural networks. Recently, Artificial Neural networks (ANN) is gaining more attention in the field of big data and machine learning. ANN is becoming the first choice for the researcher who is stepping into the field of Deep Learning. As an additional robustness check, we also implemented ANN to see which sentiment dictionary is the best fit for the linear model. We followed the min-max normalization method to scale our dataset. Except for the dummy variables and negative sentiment (*although they are negative sentiments, the sentiments weights are non-negative values*) and positive sentiment scores under each dictionary, we also scaled *Word Count, Page Length, Number of Categories, Team Size, Twitter Followers* and *Raised Amount*. In addition to these variables, we decided to use the *RIX, SMOG,* and *FOG* readability scores as they have non-negative values (see **Table 2.**). The *min-max* normalization process scales the variables between 0 and 1 which feeds the machine uniform sets of variables in the dataset.

$$x_{scaled} = \frac{x - min(x)}{max(x) - min(x)} \tag{5}$$

where x represents a single feature or a variable vector, $min(x)$ is the minimum value amongst variable vector x, and $max(x)$ is the maximum amongst variable vector x.

We apply a neural network package (*neuralnet*) available in R statistical software[14]. Furthermore, the calculation of generalized weights is implemented. The data frame is amended by a mean response, the mean of all responses corresponding to the same covariate vector. To obtain an overview of the results of the neural network and the generalized linear model objects, the covariate matrix is bound to the output of the neural network and the fitted values of the generalized linear model object.

[14] This package utilizes the training of neural networks using the backpropagation, resilient backpropagation with or without weight backtracking. The package allows flexible settings through custom-choice of error and activation function. See more at; https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf

We begin the ANN with 75% training and 25% test data implementing 3 hidden layers without any hidden neurons to each model incorporating different sentiment dictionaries and one readability score RIX (see **Appendix A.9. – A.12.**). The implemented neural network setup gives the lowest mean squared errors (MSE) to the test data that incorporates the Harvard GI sentiment dictionary in the linear model, the MSE on different ANN setups are given in panel a of **Table 7.** Similarly, **Figure 7**. shows the ANN model plot fitting on four different sentiment dictionaries with 75% training and 25% test data with 3 hidden layers. QDAP dictionary also has an equivalent MSE value, whereas Henry's and Loughran McDonald finance specific dictionaries have the first and the second highest MSE among all four. This might be the indication that finance-specific sentiment dictionaries do not accurately capture fintech-specific sentiments.

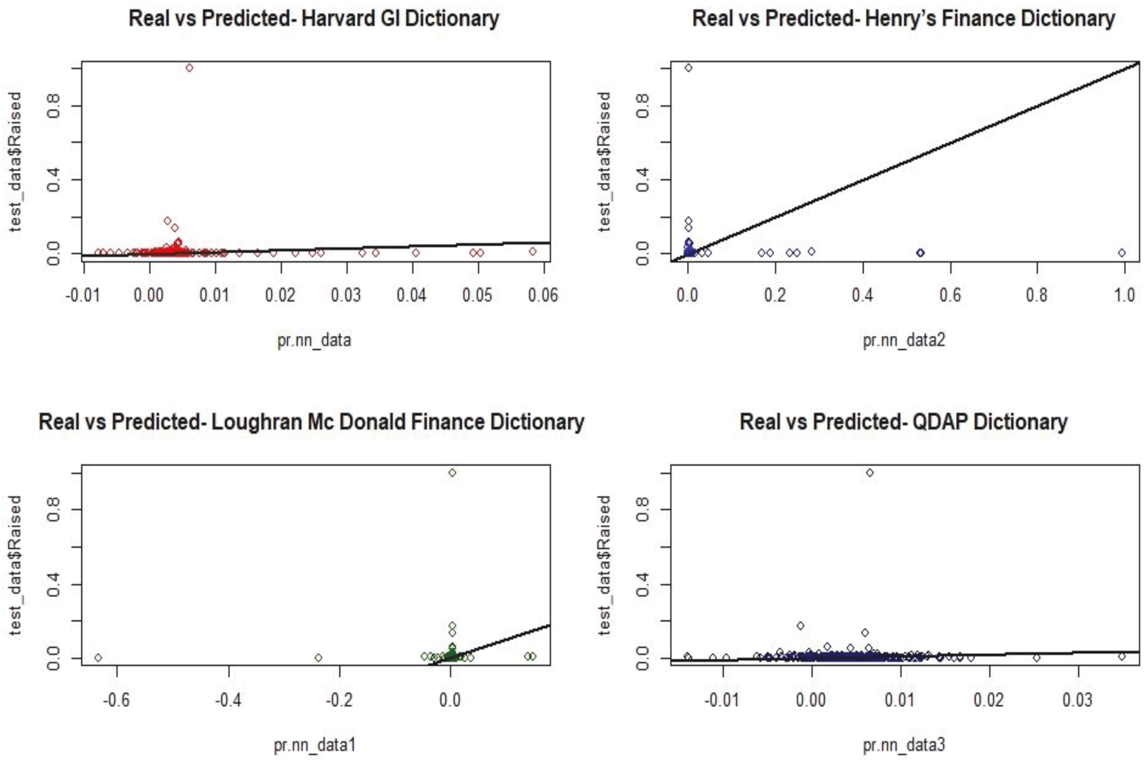
Table 7. Various Artificial Neural Networks setups and Mean Squared Errors (MSE).

This table reports the mean squared errors under various ANN setups implementing three different readability measures incorporating four different sentiment dictionaries and other ICO characteristics.

ANN	Descriptions (Training%/Test%/HiddenLayers/Neurons)	MSE.GI	MSE.HE	MSE.LM	MSE.QDAP
	Panel a	Readability Index (RIX) Readability Score			
1	75/25/3/0	0.00282	0.00757	0.00413	0.00284
2	90/10/3/0	0.00025	0.00029	0.00026	0.00026
3	75/25/10/5	0.00280	0.00280	0.00280	0.00280
4	90/10/10/5	0.00025	0.00026	0.00026	0.00027
	Average MSE	**0.00153**	**0.00273**	**0.00186**	**0.00154**
	Panel b	**SMOG Readability Score**			
5	75/25/3/0	0.00280	0.00280	0.00280	0.00280
6	90/10/3/0	0.00025	0.00025	0.00024	0.00024
7	75/25/10/5	0.00280	0.00280	0.00280	0.00290
8	90/10/10/5	0.00023	0.00025	0.00027	0.00025
	Average MSE	**0.00152**	**0.00153**	**0.00153**	**0.00155**
	Panel c	**Gunning Fog (FOG) Readability Score**			
9	75/25/3/0	0.00280	0.00300	0.00280	0.00280
10	90/10/3/0	0.00025	0.00024	0.00024	0.00024
11	75/25/10/5	0.00280	0.00290	0.00290	0.00290
12	90/10/10/5	0.00024	0.00025	0.00027	0.00025
	Average MSE	**0.00152**	**0.00160**	**0.00155**	**0.00155**

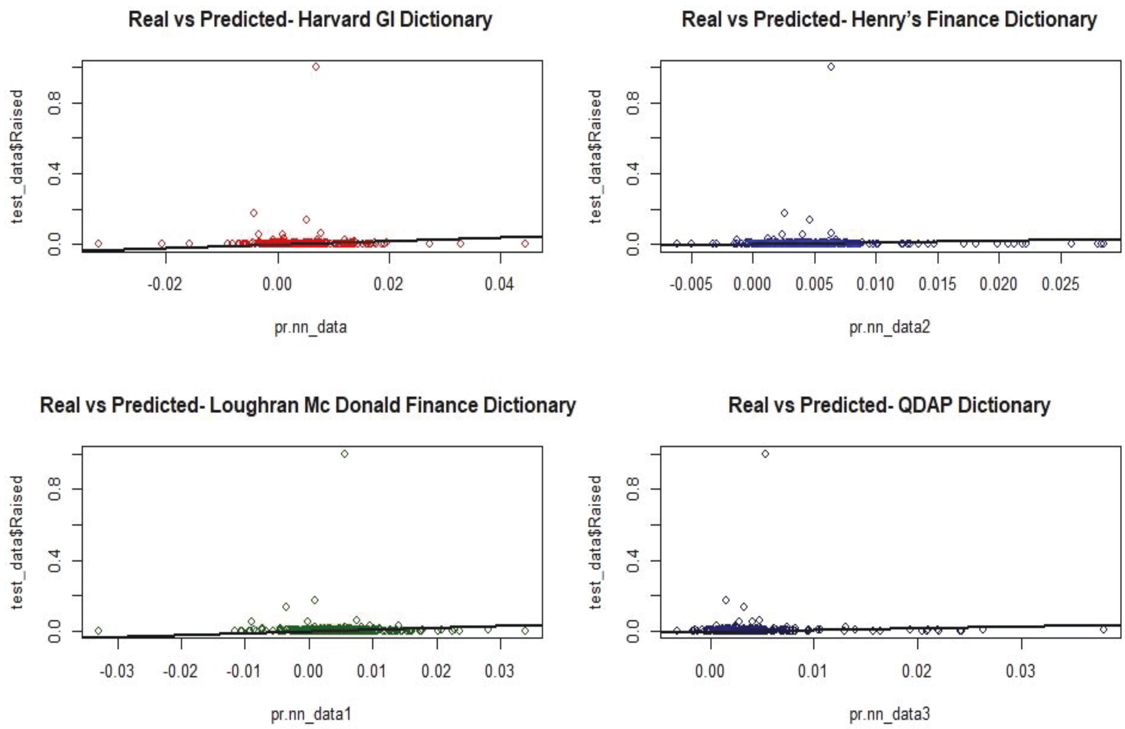
Note:
MSE.GI (Mean Squared Error, Harvard GI Dictionary)
MSE.HE (Mean Squared Error, Henry's Finance Dictionary)
MSE.LM (Mean Squared Error, Loughran Mc Donald Finance-specific Dictionary)
MSE.QDAP (Mean Squared Error, Qualitative Discourse Analysis Package)

Fig. 7. ANN Model Fitting on Four Different Sentiment Dictionaries (*75% training and 25% test data with 3 hidden layers, RIX***)**



The predictive power of ANN is subject to change based on the proportion of training and test data as well as the number of hidden layers and neurons. Keeping the training and testing proportion the same as previous, this time we increased the hidden layers to 10 from 3 and added 5 neurons to the ANN setup. Surprisingly, each model incorporating different sentiment dictionaries and RIX readability score gives the same mean squared errors to all the models for the test data. The fitted model plot for the test data is given in **Figure 8**.

Fig. 8. ANN Model Fitting on Four Different Sentiment Dictionaries (*75% training and 25% test data with 10 hidden layers, 5 neurons, RIX***)**



Implementing ANN with a higher number of hidden layers makes it difficult to decide on which sentiment dictionary best suits the model as mean squared errors are lower and equal for all. Instead of adding more hidden layers and hidden neurons, we increased the proportion of training data to 90%. Adding three hidden layers and no hidden neurons, again the liner model implementing ANN gives the lowest mean squared errors to the test data with a model that incorporates the Harvard GI sentiment dictionary. **Figure 9.** shows the ANN model plot fitting on four different sentiment dictionaries with 90% training and 10% test data with 3 hidden layers. This time, QDAP and Loughran McDonald dictionaries got the same accuracy, whereas Henry's finance-specific dictionary is still giving the highest MSE value. Furthermore, we again increased the hidden layers to 10 from 3 and added 5 hidden neurons, the test results in **Figure 10.** show the lowest MSE for Harvard GI and the highest MSE for QDAP and the same MSE for Henry's and Loughran McDonald dictionaries.

Fig. 9. ANN Model Fitting on Four Different Sentiment Dictionaries (*90% training and 10% test data with 3 hidden layers, and RIX readability score***)**

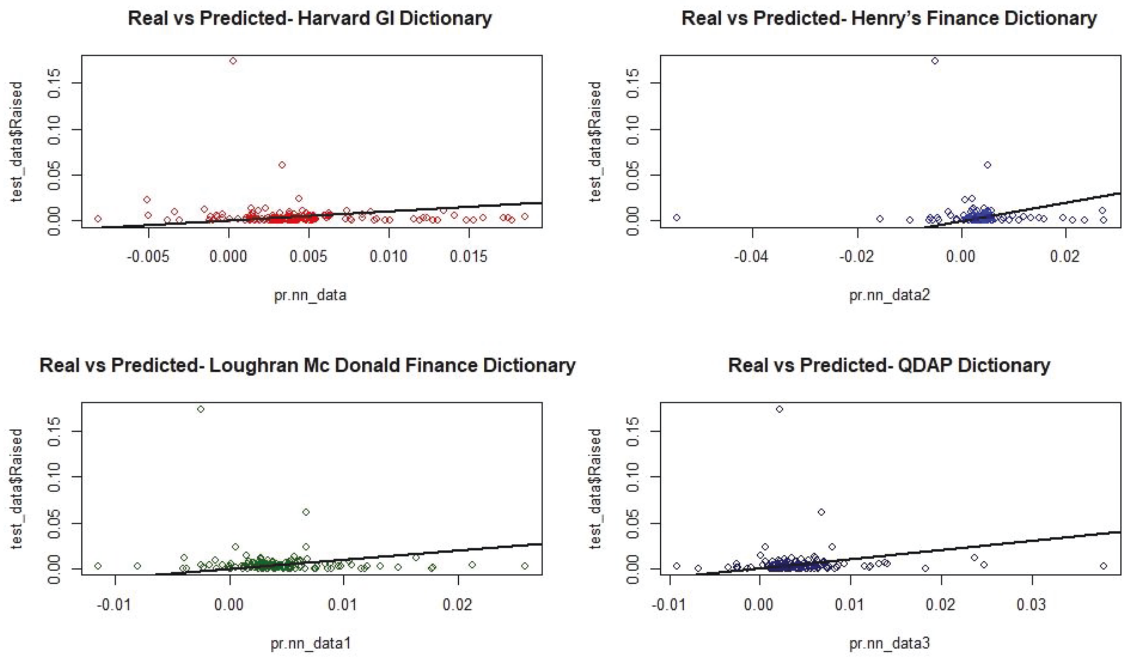
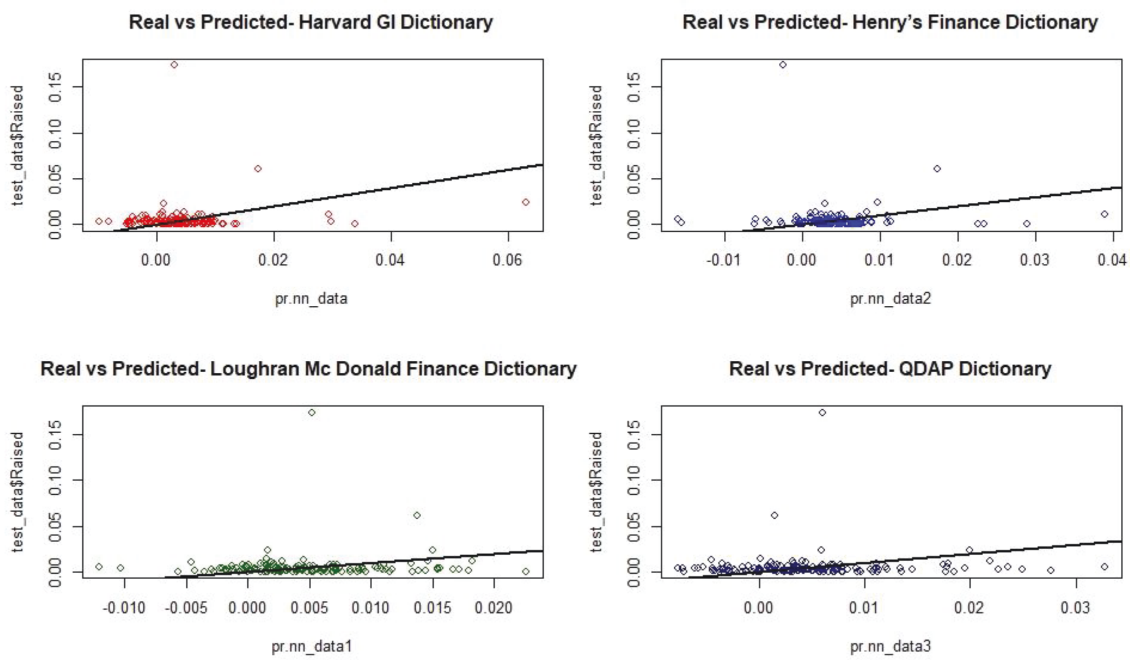


Figure 10: ANN Model Fitting on Four Different Sentiment Dictionaries (*90% training and 10% test data with 10 hidden layers and 5 neurons, and RIX readability score***)**



This confirms that the Artificial Neural Network Incorporating Harvard Sentiment Dictionary is a better fitting model. However, the predictive power of the model is subject to change based on different factors like; the number of variables used, the number of hidden layers, the ratio of training and testing data set, etc.

One could argue that the result is subject to change based on the choice of readability score measure. Therefore, we followed the same ANN setups following two other readability measures, SMOG and FOG as they also have non-negative values and the same scaling methods can be implemented. We got the lowest and almost the same MSE on average for the Harvard GI sentiment dictionary incorporating SMOG and FOG readability measures, which are reported in panel b and panel c in **Table 7**. The plot fitting accuracy on four different sentiment dictionaries with SMOG and FOG under different ANN setups is given in **Appendix (A.13- A.20)**. The ANN setups using *RIX*, *SMOG*, and *FOG* also show that our regression model is best fitted with the Harvard GI sentiment dictionary, thus the result is robust.

5. CONCLUSION

Extending earlier studies by retrieving the entire population of ICOs that have been launched in the 2014–2019 period, we found 1,507 ICOs that exhibit data on the amount of raised funding. By searching for data on ICO characteristics on various websites in association with textual analysis of those whitepapers, we identified 37 potential variables that could serve as factors associated with the success of ICOs. Contrary to earlier studies, our findings indicate that quality signals such as the number of tokens and/or softcap/hardcap, do not appear to predict ICO success. Also, the readability of a whitepaper, which may serve as an additional indicator of quality, is not associated with ICO success. We hypothesize that a rational investor would intensively deal with an ICO whitepaper and assess a project's quality based on quality and risk assessments. We do not find such evidence either as risk scores are not associated with ICO success.

Interestingly, our results provide strong evidence that ICO investors are mainly guided by their emotional experience when investing in the ICO market. Specifically, we find that negative sentiment in ICO whitepapers is positively associated with the amount of raised funding. This result suggests that negative emotions are an important factor in acquiring funding via ICOs. Moreover, the number of followers on Twitter and the attention that an ICO attract influence ICO success. Specifically, the more followers on Twitter an ICO have the higher the amount of raised funding which may be an indication of herding behavior. Since this behavior is also characterized by a desire to stay continually connected with what others are doing, we argue that the significance of the number of Twitter followers, signature campaigns and attention scores are clear indications of the significance of this phenomenon. Future research is strongly encouraged to elaborate more on this important issue.

The question arises which type of fear impacts investors' demand for tokens? Our findings indicate that investors' behavior in the ICO market is mainly driven by fears associated with 'risk', 'problem', 'change', and 'regulation', among others. Concerning fear associated with 'risk', for instance, people face nowadays (i) risk of inflation due to extremely low-interest rates in association with quantitative easing, (ii) a risk of global warming due to pollution, (iii) risk of cyberattacks due to lacks in technological advance, among others. Our findings show that projects that successfully communicate in their whitepapers how they address those risks are the successful ICOs in terms of acquiring higher amounts of funding.

Finally, Loughran and McDonald (2011) show that word lists developed for other disciplines misclassify common words in the financial text. However, research shows that the overall sentiment accuracy of the Loughran and McDonald dictionary is around 60% even in financial contexts. This indicates that a borrowed dictionary from a different discipline is likely to misjudge the sentiment exponentially. Capturing the true sentiment from the ICO whitepapers plays a significant role in risk management, this paper has important implications for investors willing to finance the project(s) related to blockchain, more specifically by investing in ICOs. Furthermore, analyzing whitepaper sentiments with non-FinTech dictionaries might be one limitation of this study. We could also observe that the sentiment captured by both of the finance-specific dictionaries did not provide any significant results in our analysis. The artificial neural network analysis also favors the Harvard GI psychological sentiment dictionary and confirms that our result is robust. We argue that there can be two different reasons for this phenomenon, that is, either the finance-specific sentiment is of no significance to investors or these dictionaries did not capture the true sentiment in FinTech-related contexts such as ICO. Therefore, we argue that there is an absolute necessity for a FinTech-specific sentiment dictionary that accurately captures the sentiment in the contexts of the new digital financial markets. This is, however, left for future research. Moreover, as the market for ICO is not free from scams and frauds, it would be interesting to see future research on how the scammers are misleading the investors via whitepaper sentiments.

REFERENCES

Adhami, S., Giudici, G. and Martinazzi, S., 2018. Why do businesses go crypto? An empirical analysis of initial coin offerings. *Journal of Economics and Business*, *100*, pp.64-75.

Alessia, D., Ferri, F., Grifoni, P. and Guzzo, T., 2015. Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, *125*(3).

Altarriba, J., 2006. 9. Cognitive Approaches to the Study of Emotion-Laden and Emotion Words in Monolingual and Bilingual Memory. In *Bilingual minds* (pp. 232-256). Multilingual Matters.

Amsden, R. and Schweizer, D., 2018. Are Blockchain Crowdsales the New'Gold Rush'? Success Determinants of Initial Coin Offerings. *SSRN (April 16, 2018)*.

Anooshian, L.J. and Hertel, P.T., 1994. Emotionality in free recall: Language specificity in bilingual memory. *Cognition & Emotion*, *8*(6), pp.503-514.

Bajo, E. and Raimondo, C., 2017. Media sentiment and IPO underpricing. *Journal of Corporate Finance*, *46*, pp.139-153.

Baker, M. and Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. *The journal of Finance*, *61*(4), pp.1645-1680.

Baker, M. and Wurgler, J., 2007. Investor sentiment in the stock market. *Journal of economic perspectives*, *21*(2), pp.129-152.

Baker, S.R., Bloom, N. and Davis, S.J., 2016. Measuring economic policy uncertainty. *The quarterly journal of economics*, *131*(4), pp.1593-1636.

Benedetti, H. and Kostovetsky, L., 2021. Digital tulips? Returns to investors in initial coin offerings. *Journal of Corporate Finance*, *66*, p.101786.

Boreiko, D. and Risteski, D., 2020. Serial and large investors in initial coin offerings. *Small Business Economics*, pp.1-19.

Catalini, C. and Gans, J.S., 2018. *Initial coin offerings and the value of crypto tokens* (No. w24418). National Bureau of Economic Research.

Chen, P., Lin, J., Chen, B., Lu, C. and Guo, T., 2015. Processing emotional words in two languages with one brain: ERP and fMRI evidence from Chinese–English bilinguals. *Cortex*, *71*, pp.34-48.

Chod, J. and Lyandres, E., 2021. A theory of icos: Diversification, agency, and information asymmetry. *Management Science*.

Domingo, R.S., Piñeiro-Chousa, J. and López-Cabarcos, M.Á., 2020. What factors drive returns on initial coin offerings?. *Technological Forecasting and Social Change*, *153*, p.119915.

Fisch, C., 2019. Initial coin offerings (ICOs) to finance new ventures. *Journal of Business Venturing*, *34*(1), pp.1-22.

Giudici, G., Moncayo, G.G. and Martinazzi, S., 2020. The role of advisors' centrality in the success of Initial Coin Offerings. *Journal of Economics and Business*, *112*, p.105932.

Grobys, K. and Sapkota, N., 2020. Predicting cryptocurrency defaults. *Applied Economics, 52(46)*, pp.5060-5076.

Guldiken, O., Tupper, C., Nair, A. and Yu, H., 2017. The impact of media coverage on IPO stock performance. *Journal of Business Research*, *72*, pp.24-32.

Guo, L., Shi, F. and Tu, J., 2016. Textual analysis and machine leaning: Crack unstructured data in finance and accounting. *The Journal of Finance and Data Science*, *2*(3), pp.153-170.

He, W., Wu, H., Yan, G., Akula, V. and Shen, J., 2015. A novel social media competitive analytics framework with sentiment benchmarks. *Information & Management*, *52*(7), pp.801-812.

Henry, E., 2008. Are investors influenced by how earnings press releases are written?. *The Journal of Business Communication (1973)*, *45*(4), pp.363-407.

Hornuf, L., Kück, T. and Schwienbacher, A., 2021. Initial coin offerings, information disclosure, and fraud. *Small Business Economics*, pp.1-19.

Howell, S.T., Niessner, M. and Yermack, D., 2020. Initial coin offerings: Financing growth with cryptocurrency token sales. *The Review of Financial Studies*, *33*(9), pp.3925-3974.

Huang, D., Jiang, F., Tu, J. and Zhou, G., 2015. Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies*, *28*(3), pp.791-837.

Huang, W., Meoli, M. and Vismara, S., 2020. The geography of initial coin offerings. *Small Business Economics*, *55*(1), pp.77-102.

Jovanovic, B. and Rousseau, P.L., 2001. Why wait? A century of life before IPO. *American Economic Review*, *91*(2), pp.336-341.

Kaal, W.A., 2018. Initial Coin Offerings: The top 25 jurisdictions and their comparative regulatory responses (as of May 2018). *Stan. J. Blockchain L. & Pol'y*, *1*, p.41.

Kissler, J., Herbert, C., Peyk, P. and Junghofer, M., 2007. Buzzwords: early cortical responses to emotional words during reading. *Psychological Science*, *18*(6), pp.475-480.

Li, J. and Mann, W., 2018. Initial coin offerings and platform building. *SSRN Electronic Journal*.

Li, X., Xie, H., Chen, L., Wang, J. and Deng, X., 2014. News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, *69*, pp.14-23.

Loughran, T. and McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of finance*, *66*(1), pp.35-65.

Loughran, T. and McDonald, B., 2013. IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, *109*(2), pp.307-326.

Lyandres, E., Palazzo, B. and Rabetti, D., 2019. Do tokens behave like securities? An anatomy of initial coin offerings. *SSRN Electronic Journal*.

Meyer, A. and Ante, L., 2020. Effects of initial coin offering characteristics on cross-listing returns. *Digital Finance*, *2*(3), pp.259-283.

Mohammad, S.M. and Turney, P.D., 2013. Nrc emotion lexicon. *National Research Council, Canada*, 2.

Momtaz, P., Drobetz, W. and Schroeder, H., 2018. Investor sentiment and initial coin offerings. *The Journal of Alternative Investments*, *21*(4).

Momtaz, P.P., 2020. Initial coin offerings. *Plos one*, *15*(5), p.e0233018.

Muhammad, A., Wiratunga, N. and Lothian, R., 2016. Contextual sentiment analysis for social media genres. *Knowledge-based systems*, *108*, pp.92-101.

Pavlenko, A., 2008. Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and cognition*, *11*(2), pp.147-164.

Pröllochs, N., Feuerriegel, S. and Neumann, D., 2015, May. Generating Domain-Specific Dictionaries using Bayesian Learning. In *ECIS*.

Qadan, M., 2019. Risk appetite, idiosyncratic volatility and expected returns. *International Review of Financial Analysis*, *65*, p.101372.

Roosenboom, P., van der Kolk, T. and de Jong, A., 2020. What determines success in initial coin offerings?. *Venture Capital*, *22*(2), pp.161-183.

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

Tsukioka, Y., Yanagi, J. and Takada, T., 2018. Investor sentiment extracted from internet stock message boards and IPO puzzles. *International Review of Economics & Finance*, *56*, pp.205-217.

Wu, C., Zhang, J. and Yuan, Z., 2021. Exploring Affective Priming Effect of Emotion-Label Words and Emotion-Laden Words: An Event-Related Potential Study. *Brain Sciences*, *11*(5), p.553.

Yap, M.J. and Seow, C.S., 2014. The influence of emotion on lexical processing: Insights from RT distributional analysis. *Psychonomic bulletin & review*, *21*(2), pp.526-533.

Yekrangi, M. and Abdolvand, N., 2021. Financial markets sentiment analysis: Developing a specialized Lexicon. *Journal of Intelligent Information Systems*, *57*(1), pp.127-146.

Young, L. and Soroka, S., 2012. Affective news: The automated coding of sentiment in political texts. *Political Communication*, *29*(2), pp.205-231.

Yu, G.N., 2019. Factors of success of Initial coin offering. Empirical evidence from 2016-2019. *Финансы: теория и практика*, *23*(5).

Zetzsche, D.A., Buckley, R.P., Arner, D.W. and Föhr, L., 2017. The ICO Gold Rush: It's a scam, it's a bubble, it's a super challenge for regulators. *University of Luxembourg Law Working Paper*, (11), pp.17-83.

Zhang, D., He, W., Wang, T., Luo, W., Zhu, X., Gu, R., Li, H. and Luo, Y.J., 2014. Three stages of emotional word processing: an ERP study with rapid serial visual presentation. *Social cognitive and affective neuroscience*, *9*(12), pp.1897-1903.

Zhang, S., Aerts, W., Lu, L. and Pan, H., 2019. Readability of token whitepaper and ICO first-day return. *Economics Letters*, *180*, pp.58-61.

Zhao, R. and Hou, W., 2018. Socially responsible initial coin offerings. Proceedings of the *European Financial Management Association*, 2019 Annual Meetings, University of Azores, Ponta Delgada, Island of S. Miguel, Portugal.

Zhao, W., Chen, L., Zhou, C. and Luo, W., 2018. Neural correlates of emotion processing in word detection task. *Frontiers in psychology*, *9*, p.832.

APPENDICES

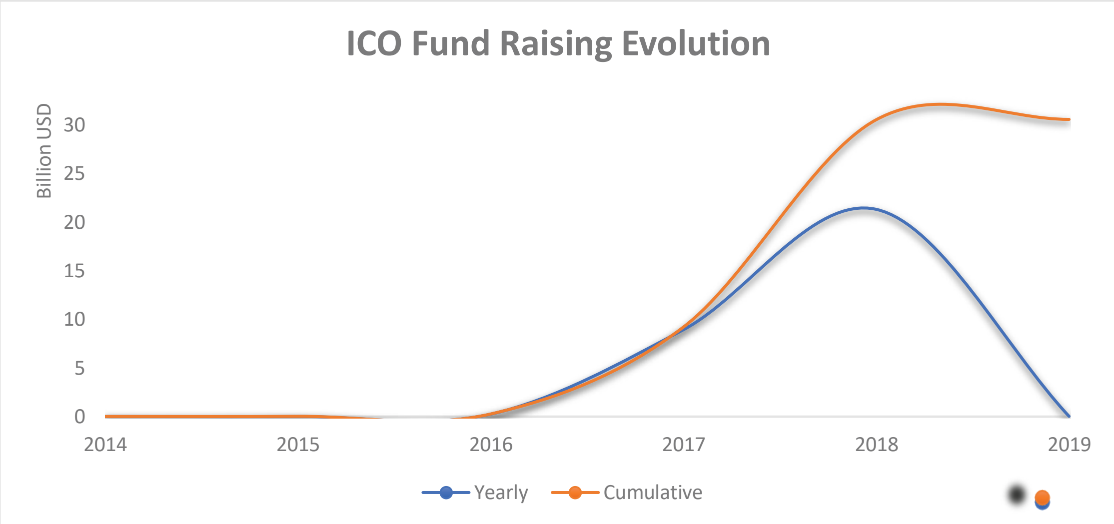
Appendix A.1. R packages and functions and their usage in the paper

S.No.	R Package, function	Usage
1	rvest, xml2	to web scrap the scattered data
2	pdftools, VCorpus	to read pdf files into R and create a corpus
3	tm, tm_map, DocumentTermMatrix, TermDocumentMatrix	cleaning the corpus, creating term and document matrix
4	wordcloud, wordcloud2	to create the word cloud
5	SentimentAnalysis, alalyzesentiment	to get the various sentiment scores of each document corpus
6	textreuse, TextReuseCorpus, pairwise_compare, jaccard_similarity	to get the pairwise Jaccard similarity scores for each document corpus
7	KoRpus, textstat_readability	to get the different readability scores for each document corpus
8	stargazer	to display R outputs into tables
9	neuralnet	to perform artificial neural network

Appendix A.2. The number of ICOs and amount raised by different countries around the globe during the 2014-2019 Period.

S.No.	Country	Raised($)	ICOs	S.No.	Country	Raised($)	ICOs	S.No.	Country	Raised($)	ICOs
1	Non-disclosed	18270321760	765	28	Netherlands	74679825	13	55	Brazil	19629000	1
2	United States	2480961468	110	29	South Korea	70666256	3	56	Ecuador	16314065	1
3	Switzerland	1211340383	49	30	Belize	64416465	9	57	Spain	14959000	5
4	Singapore	1088358207	65	31	Afghanistan	62000000	2	58	Marshall Islands	13044000	2
5	United Kingdom	898694814	59	32	Indonesia	61159000	3	59	Latvia	12648000	6
6	Russia	700688304	58	33	Bulgaria	59031295	7	60	CostaRica	12017000	2
7	Cayman Islands	685258522	13	34	Sweden	49653398	2	61	Panama	11000000	2
8	Estonia	532220166	42	35	Czech Republic	47623000	6	62	Philippines	10864000	1
9	Gibraltar	311165807	15	36	Seychelles	46560000	8	63	Luxembourg	10643815	2
10	HongKong	284419350	20	37	Iceland	45949800	1	64	Cambodia	10561000	2
11	Canada	282268020	17	38	Argentina	44316000	4	65	Bahamas	10000000	1
12	Lithuania	272378963	5	39	Turkey	42868731	3	66	American Samoa	6507000	1
13	Germany	257943033	14	40	Israel	42784000	7	67	Finland	6125000	2
14	Australia	250441332	16	41	Latin America	41710592	1	68	Malaysia	5074000	1
15	Georgia	186652722	9	42	Dubai	41706262	2	69	Norway	3751000	3
16	Japan	186462000	6	43	Thailand	39495517	5	70	Belgium	3609000	1
17	Malta	158261610	12	44	Taiwan	39175060	2	71	Egypt	2877000	1
18	United Arab Emirates	157640000	9	45	Slovakia	37378000	1	72	Serbia	2820000	1
19	France	155191896	10	46	Romania	31513000	3	73	Denmark	1022593	2
20	China	155057000	10	47	Kazakhstan	30280000	2	74	Jersey	1000000	1
21	India	119564600	9	48	Liechtenstein	28682516	2	75	SaintVincent	1000000	1
22	South Africa	112582843	5	49	Austria	28362000	5	76	Italy	708000	2
23	British Virgin Islands	102469132	9	50	Isle of Man	27515417	3	77	Saint Kitts and Nevis	305000	1
24	Poland	101885080	8	51	Mexico	24000000	2	78	Hungary	249000	1
25	Ukraine	88855172	5	52	Belarus	22874000	3	79	Pakistan	51000	1
26	Slovenia	86815000	11	53	Korea	20665996	2	80	Peru	15000	1
27	Cyprus	79269000	8	54	Colombia	40930000	2	81	Andorra	2000	1
									Total	30559957787	1508

Appendix A.3. Evolution of ICO fundraising over time.



Appendix A.4. Number of ICOs under each category (same ICO has been listed under various categories)

S.No.	Category	ICOs	S.No.	Category	ICOs
1	Platform	1373	24	Manufacturing	58
2	Business and Services	1260	25	Blockchain	55
3	Cryptocurrency	1015	26	Charity	49
4	Investment	469	27	Payments and Wallets	38
5	Software and Computing	389	28	Legal	33
6	Communication and Media	373	29	Art	33
7	Artificial Intelligence	370	30	Electronics	31
8	Banking and Finance	316	31	Identity and Security	22
9	Internet	306	32	Content	21
10	Smartcontract	295	33	Mining	16
11	Entertainment	285	34	Gaming Industry	16
12	Infrastructure	230	35	Commerce	16
13	Gambling, Casino, Betting	206	36	Marketplace	10
14	Bigdata	174	37	Advertising	9
15	Retail	143	38	Logistics	8
16	Health and Drugs	130	39	Augumented Reality	8
17	Virtual Reality	112	40	Utilities	7
18	Real-Estate	112	41	Jobs	6
19	Education	97	42	Asset Management	6
20	Other	88	43	Venture Capital	5
21	Tourism	87	44	Internet of Things	5
22	Energy	76	45	Funding	5
23	Sports	70	46	Transportation	3

Appendix A.5. Top 200 words used in ICO whitepapers (2014-2019)

S.No.	word	freq	S.No.	word	freq	S.No.	word	freq	S.No.	word	freq
1	token	299563	51	provide	32093	101	growth	20656	151	rate	15589
2	use	289016	52	global	32034	102	mining	20179	152	fiat	15571
3	crypto	155546	53	management	31577	103	current	20131	153	internet	15554
4	platform	147976	54	public	31453	104	fund	19801	154	research	15527
5	blockchain	134098	55	future	30518	105	get	19788	155	reward	15383
6	market	105919	56	process	30440	106	open	19686	156	full	15329
7	service	102761	57	experience	30185	107	made	18876	157	features	15264
8	data	102193	58	currency	30167	108	asset	18791	158	take	15135
9	system	89070	59	decentralized	29644	109	year	18605	159	launch	15050
10	network	83274	60	make	29635	110	possible	18550	160	technologies	15027
11	company	81228	61	white	29047	111	sales	18469	161	share	15005
12	transaction	80592	62	amount	28825	112	potential	18414	162	proof	14859
13	contract	77877	63	available	28764	113	applications	18383	163	start	14838
14	exchange	75775	64	purchase	28084	114	receive	18351	164	capital	14831
15	pay	72459	65	chain	27543	115	level	18264	165	help	14736
16	project	69882	66	high	27320	116	case	18052	166	parties	14724
17	innovation	65650	67	legal	27187	117	game	17929	167	required	14714
18	information	63264	68	order	27167	118	participants	17926	168	ensure	14706
19	business	62271	69	investors	26836	119	usd	17840	169	additional	14659
20	time	61990	70	real	26805	120	limited	17776	170	means	14567
21	ico	61572	71	people	26617	121	terms	17674	171	provides	14419
22	smart	60384	72	bitcrypto	25895	122	supply	17409	172	trust	14392
23	development	60287	73	online	25621	123	holders	17258	173	control	14377
24	eth	58331	74	create	25338	124	allows	17182	174	offering	14359
25	technology	58321	75	work	25245	125	party	17166	175	bank	14323
26	sale	55197	76	assets	24588	126	storage	17140	176	document	14320
27	value	53084	77	private	24534	127	media	17102	177	event	14319
28	product	48643	78	total	24285	128	large	17091	178	rewards	14299
29	team	48108	79	key	24241	129	page	16918	179	members	14254
30	risk	44313	80	model	24190	130	rights	16879	180	period	14252
31	world	41952	81	application	24084	131	power	16733	181	created	14236
32	digital	39818	82	money	24028	132	allow	16588	182	website	14153
33	cost	37971	83	problem	23958	133	increase	16488	183	change	14133
34	financial	37912	84	distribution	23773	134	nodes	16225	184	node	14124
35	ecosystem	36681	85	set	23679	135	main	16207	185	peer	14041
36	customer	36035	86	protocol	23666	136	solutions	16195	186	securities	13997
37	whitepaper	35879	87	need	23448	137	various	16177	187	currencies	13982
38	fee	35877	88	social	23266	138	form	16165	188	partners	13906
39	security	35685	89	account	22977	139	secure	16132	189	related	13872
40	trading	34244	90	distributed	22461	140	technical	16097	190	source	13828
41	access	34103	91	part	22434	141	operations	15989	191	provided	13759
42	price	33844	92	support	22177	142	existing	15975	192	marketplace	13712
43	community	33830	93	software	21878	143	energy	15960	193	version	13704
44	investment	33820	94	solution	21723	144	trade	15890	194	foundation	13689
45	industry	33243	95	mobile	21719	145	address	15886	195	demand	13688
46	funds	32650	96	different	21576	146	revenue	15859	196	group	13688
47	paper	32640	97	block	21409	147	initial	15830	197	businesses	13624
48	content	32631	98	offer	21404	148	billion	15738	198	developers	13607
49	marketing	32603	99	app	21379	149	buy	15644	199	regulatory	13566
50	wallet	32244	100	million	21197	150	advertising	15603	200	making	13513

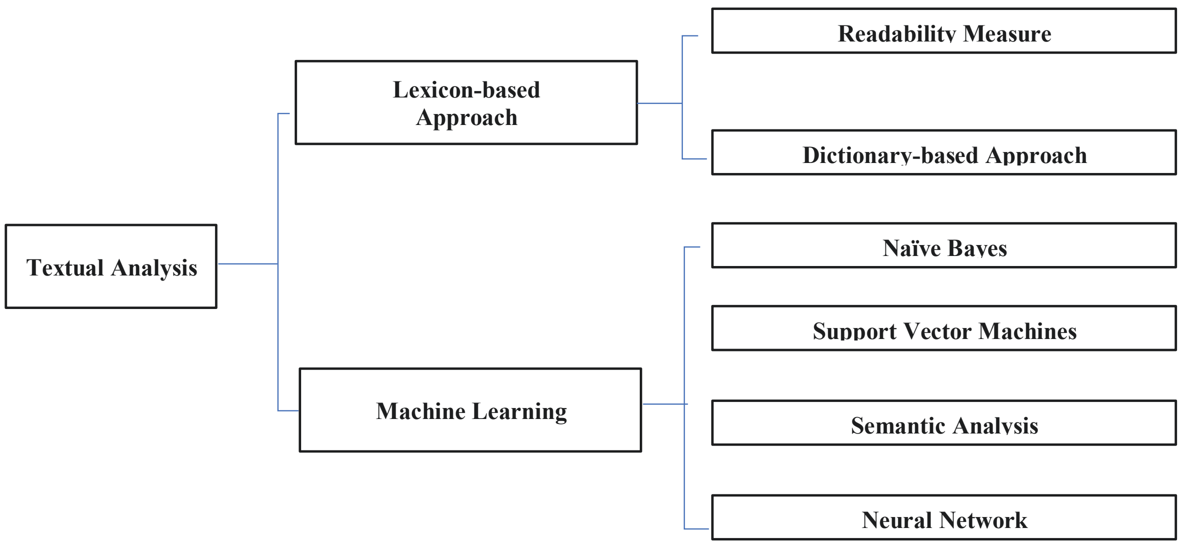
Appendix A.6. Readability and their measurements

Readability	Formula
Flesch	206.835 - (1.015 ∗ number of words/number of sentences) - (84.6/number of syllables/number of words)
Flesch–Kincaid	0.39 ∗ (number of words/number of sentences) + 11.8 ∗ (number of syllables / number of words) – 15.59
COL	0.0588 ∗ (Average number of letters per 100 words) – 0.296 ∗ (Average number of sentences per 100 words) – 15.8
RIX	(Number of words with 7 characters or more) / (number of sentences)
FOG	((Average number of words per sentence) + (number of words of 3 syllables or more)) ∗ 0.4
ARI	4.71 ∗ (number of characters / number of words) + 0.5 ∗ (number of words / number of sentences) - 21.43
SMOG	1.043 ∗ sqrt (30 ∗ number of words with more than two syllables / number of sentences) + 3.1291

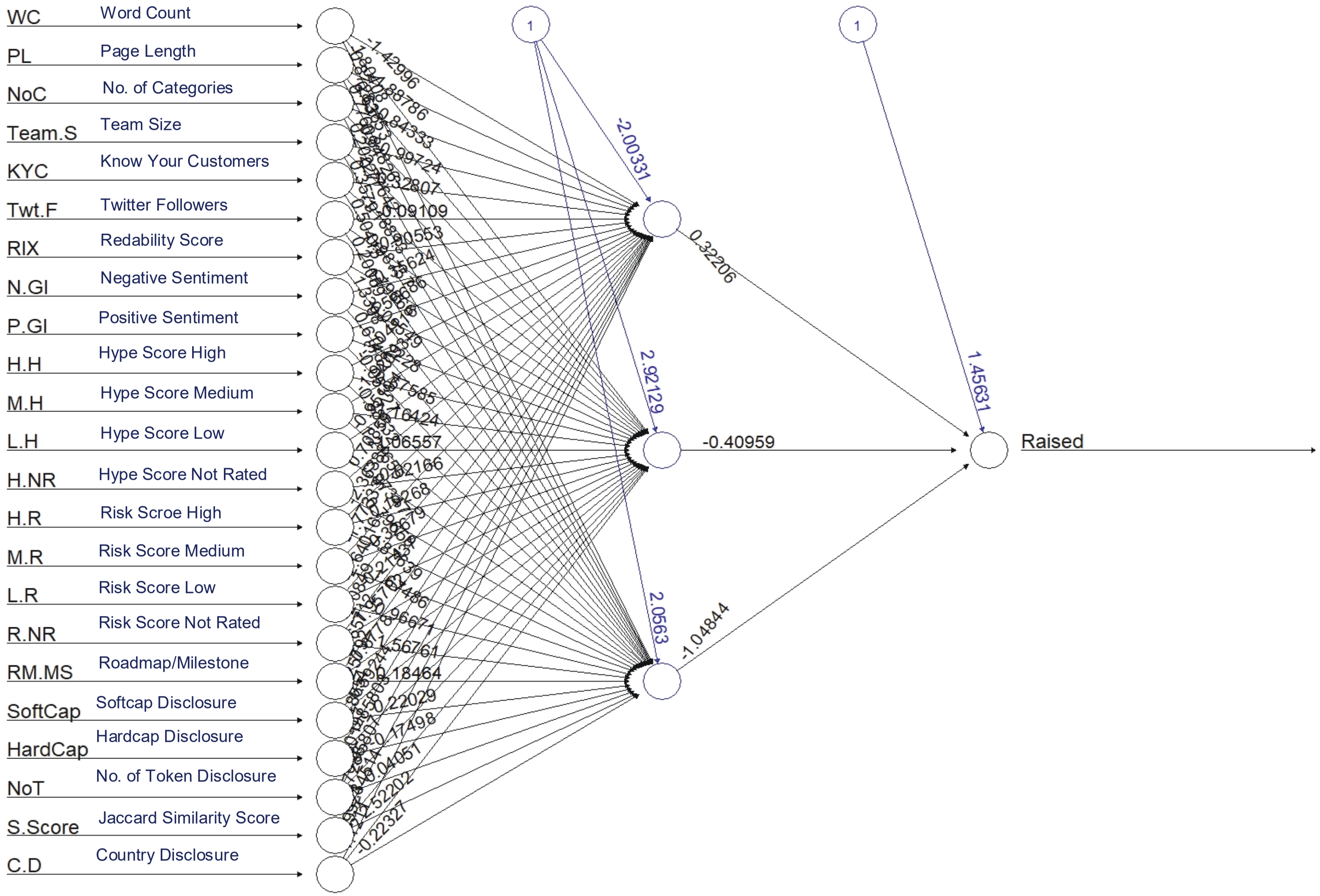
Appendix A.7. Flesh/Flesh-Kincaid readability measurement

Score	Notes
90-100	very easy to read, easily understood by an average 11-year-old student
80-90	easy to read
70-80	fairly easy to read
60-70	easily understood by 13- to 15-year-old students
50-60	fairly difficult to read
30-50	difficult to read, best understood by college graduates
0-30	very difficult to read, best understood by university graduates

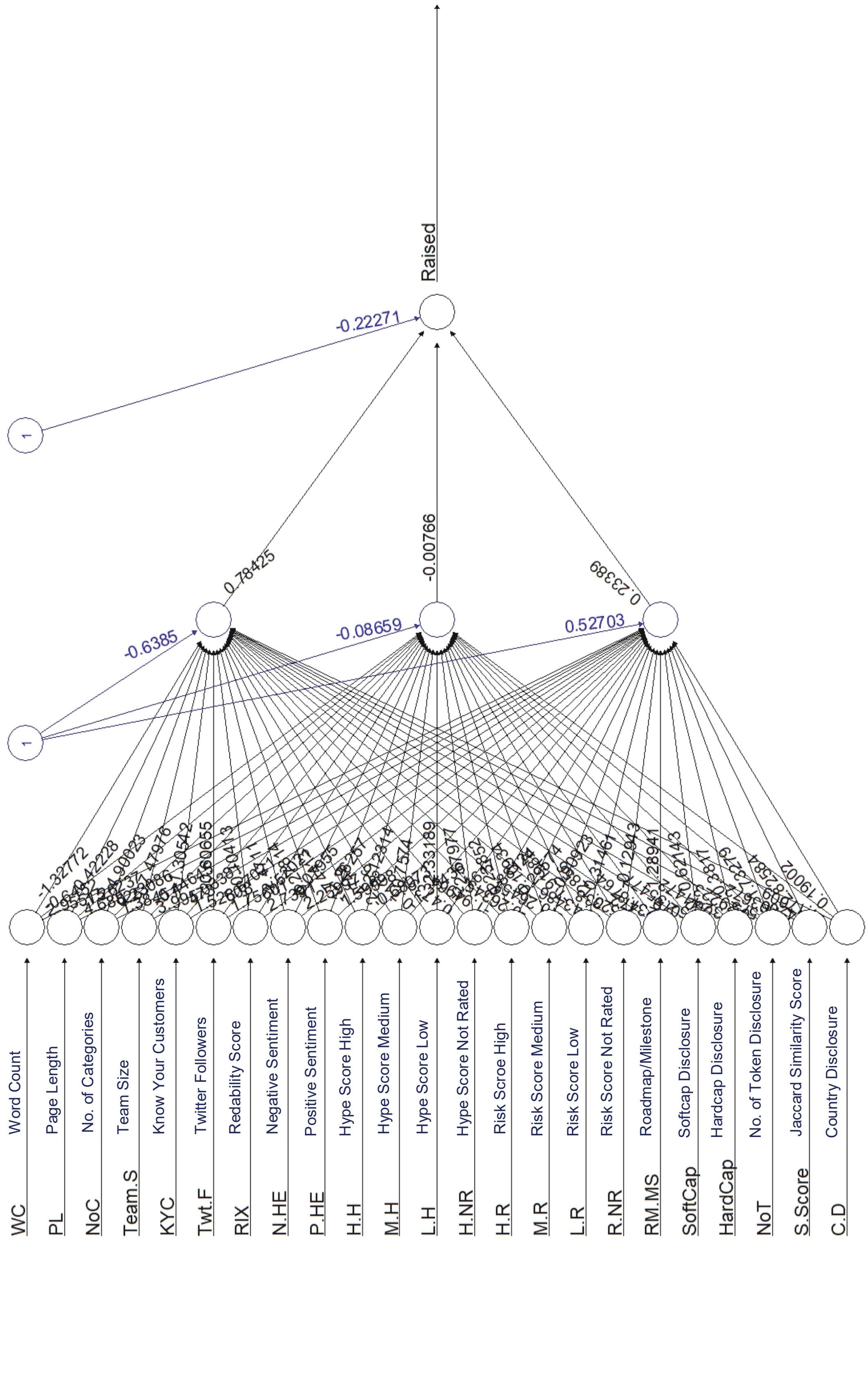
Appendix A.8. A general method for textual analysis (source: Guo et al., 2016)



Appendix A.9. ANN incorporating Harvard GI sentiment dictionary (*75% training and 25% test data with 3 hidden layers, RIX*)

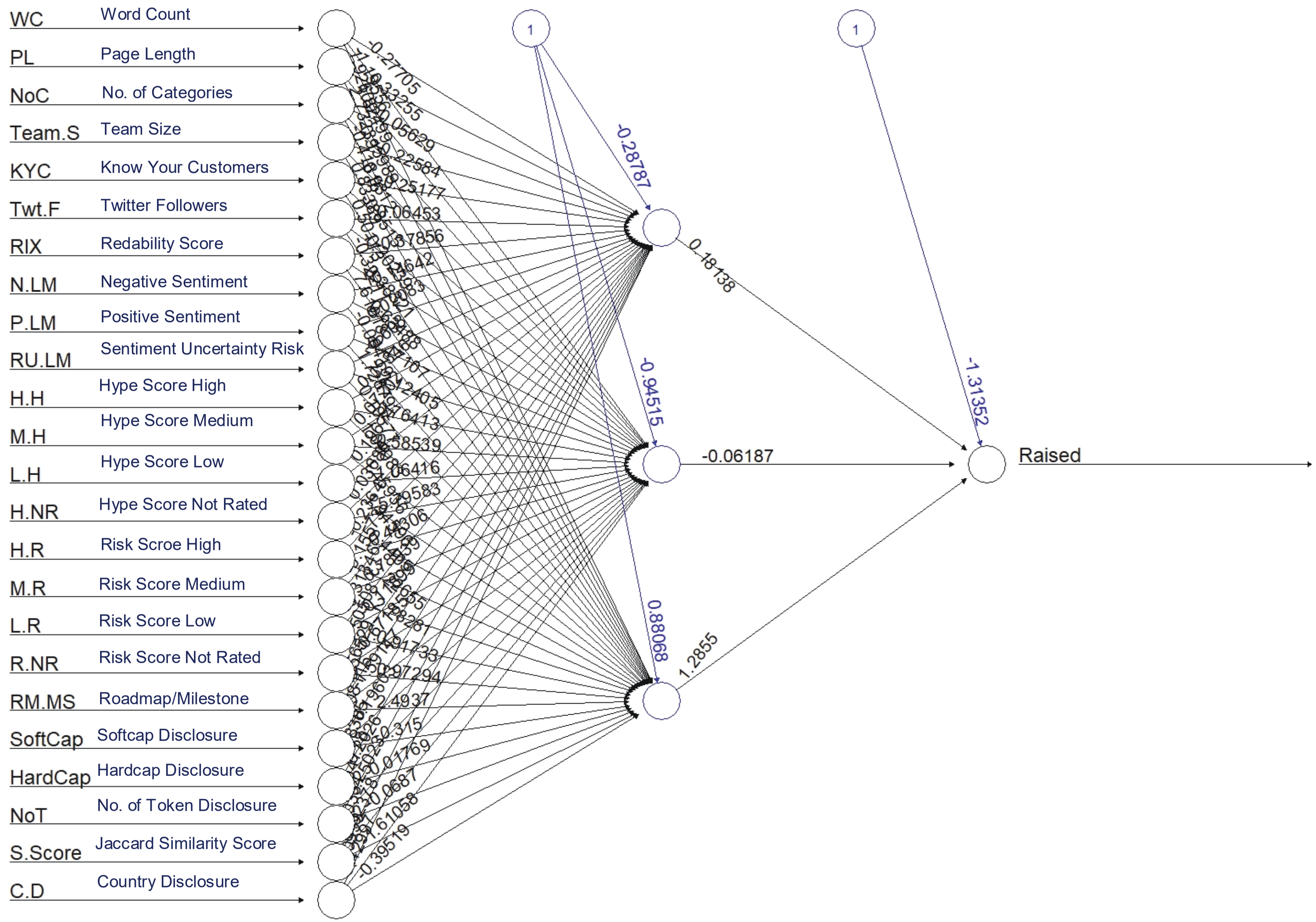


Appendix A.10. ANN incorporating Henry's finance dictionary *(75% training and 25% test data with 3 hidden layers, RIX)*

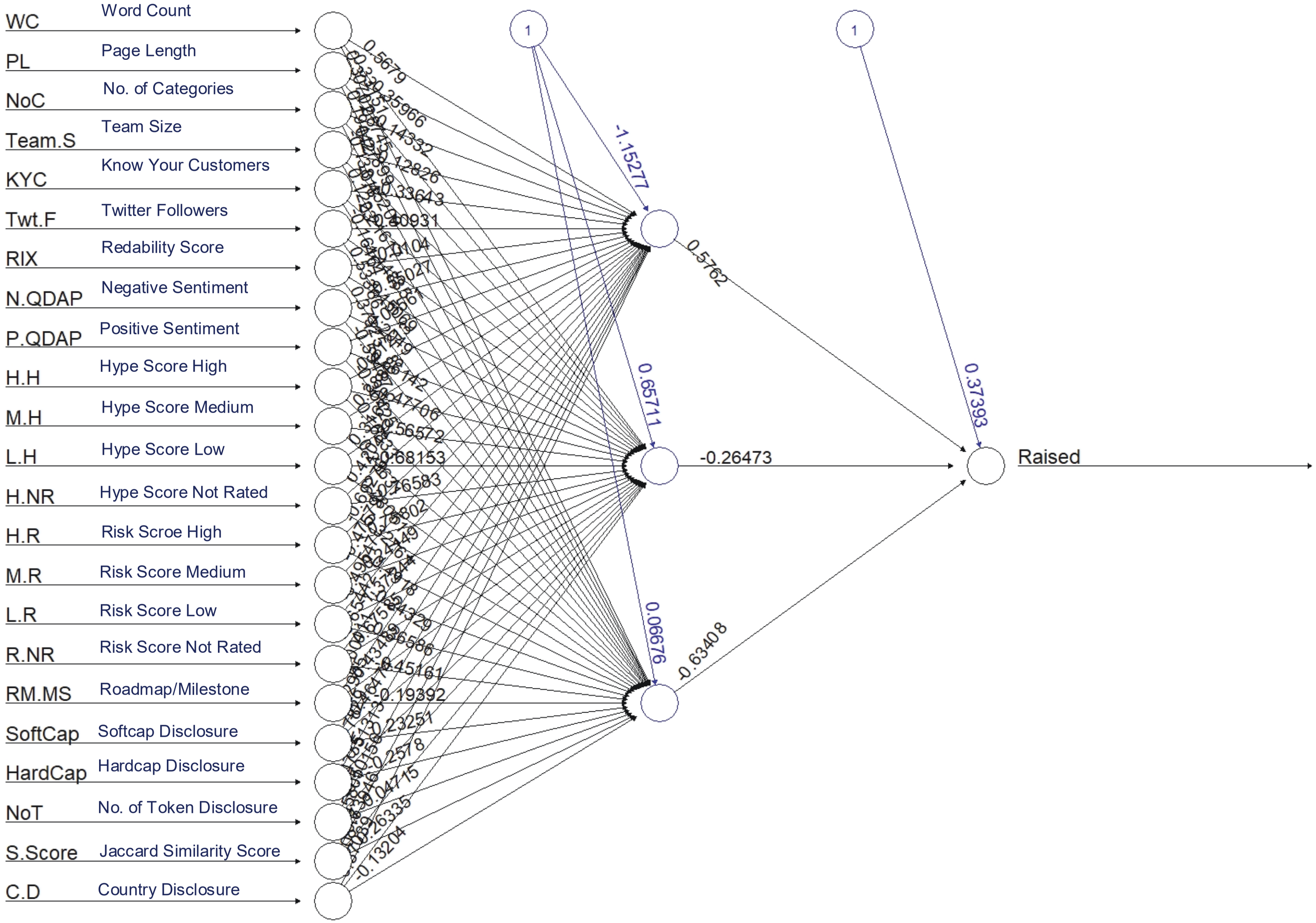


62

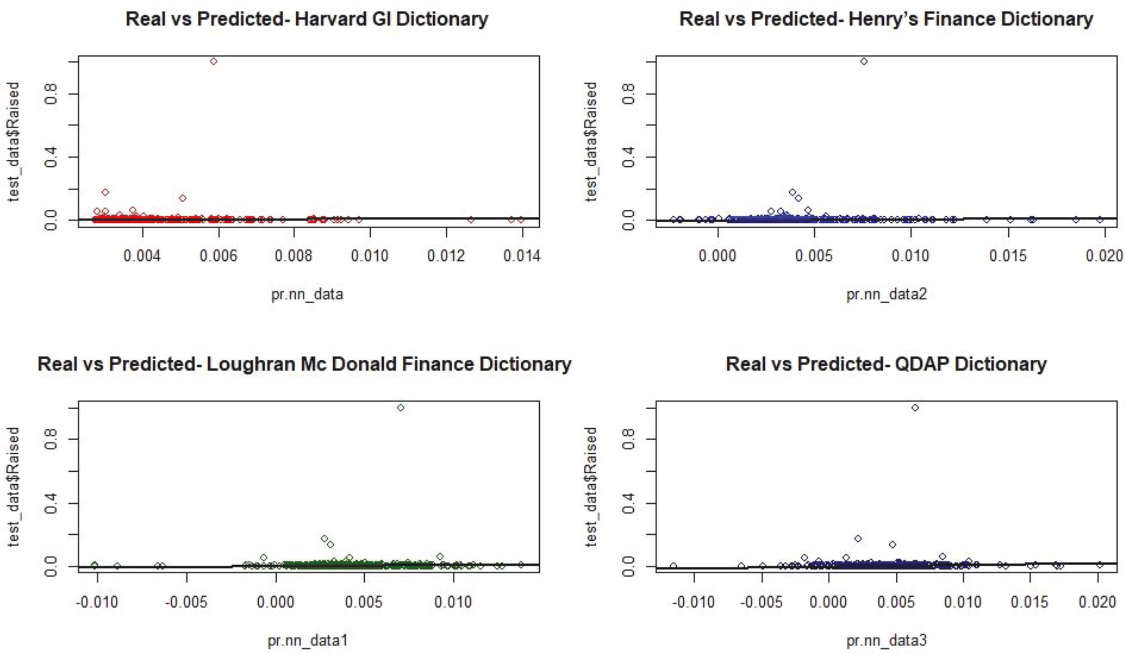
Appendix A.11. ANN incorporating Loughran McDonald finance dictionary (*75% training and 25% test data with 3 hidden layers, RIX*)



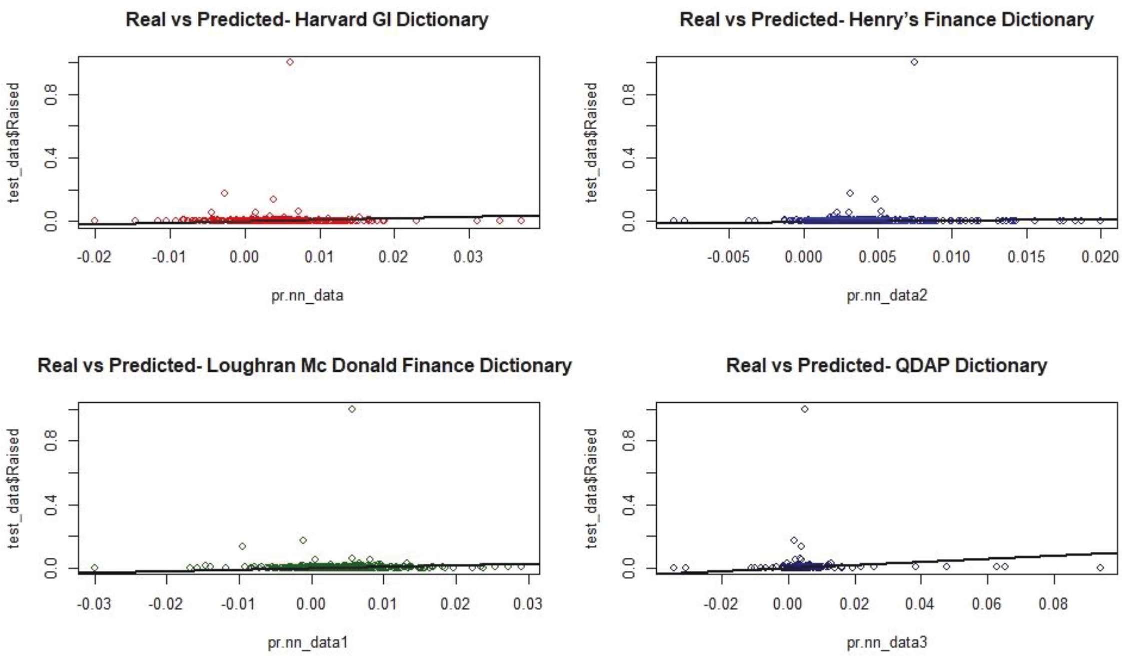
Appendix A.12. ANN incorporating QDAP sentiment dictionary (*75% training and 25% test data with 3 hidden layers, RIX*)



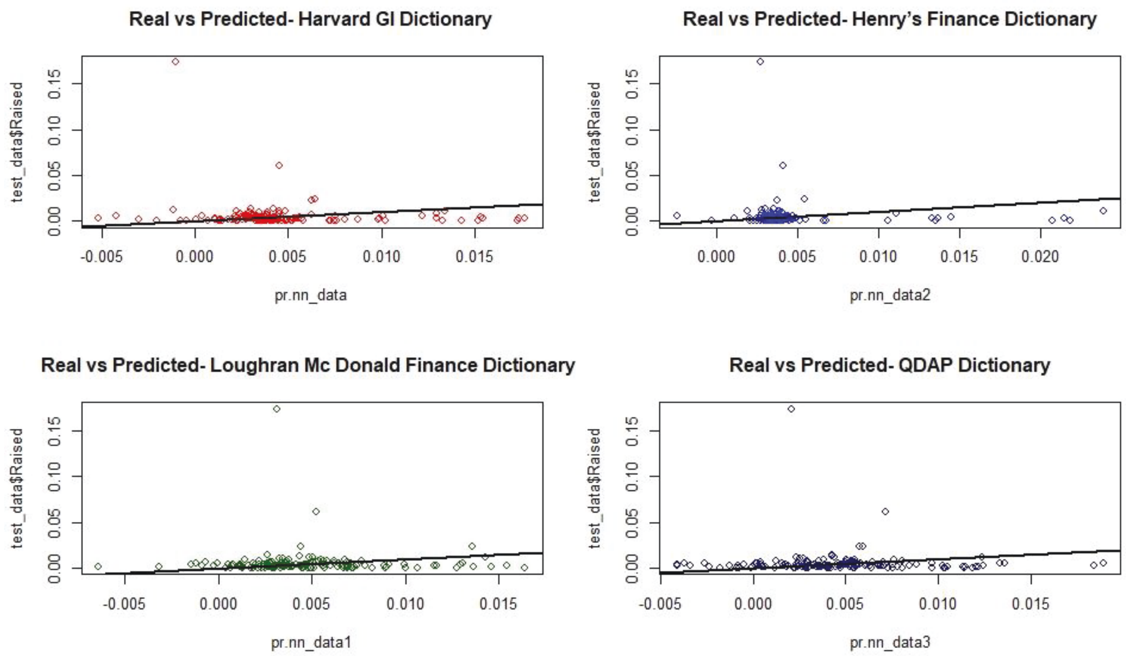
Appendix A.13. ANN model fitting on four different sentiment dictionaries (*75% training and 25% test data with 3 hidden layers, and SMOG readability score*)



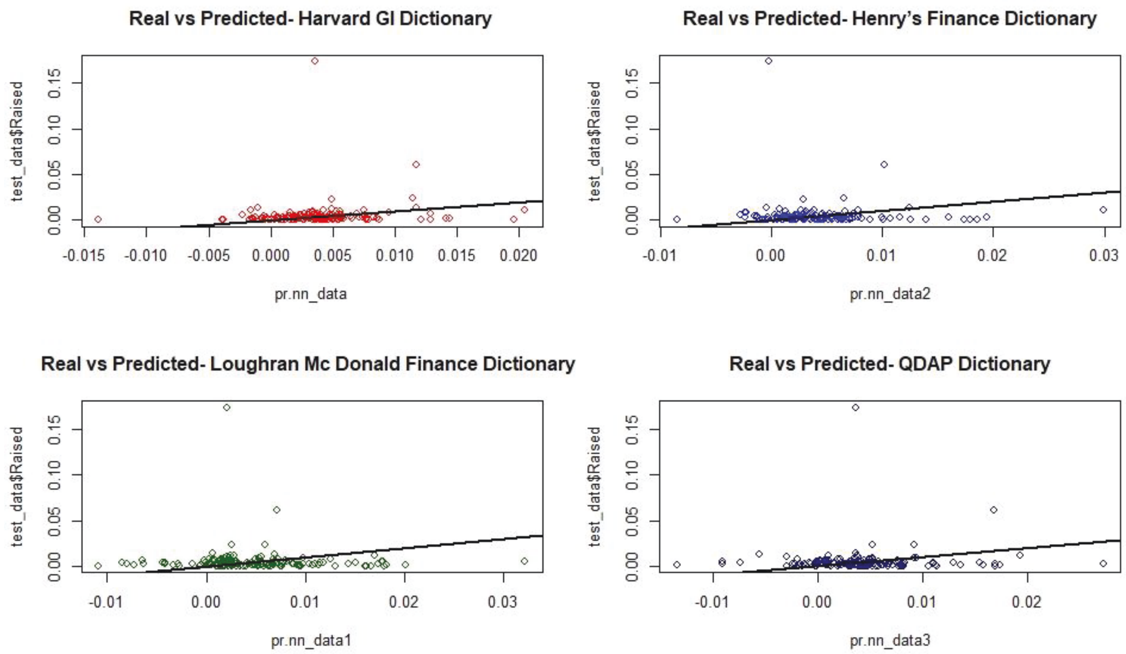
Appendix A.14. ANN model fitting on four different sentiment dictionaries (*75% training and 25% test data with 10 hidden layers and 5 neurons, and SMOG readability score*)



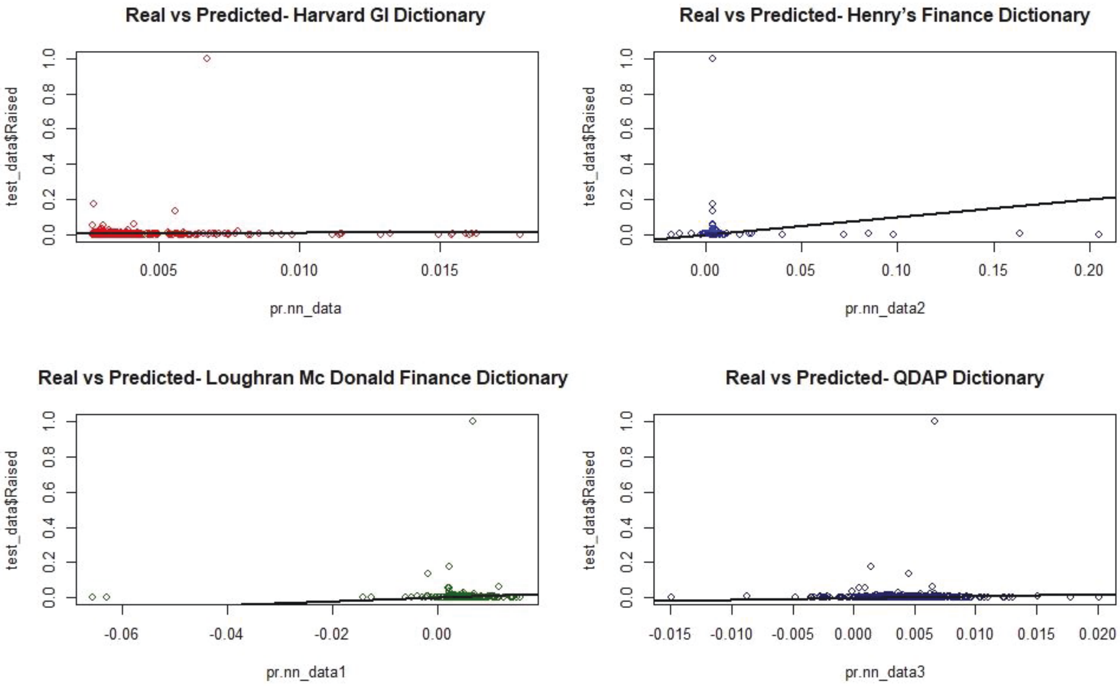
Appendix A.15. ANN model fitting on four different sentiment dictionaries (*90% training and 10% test data with 3 hidden layers, and SMOG readability score*)



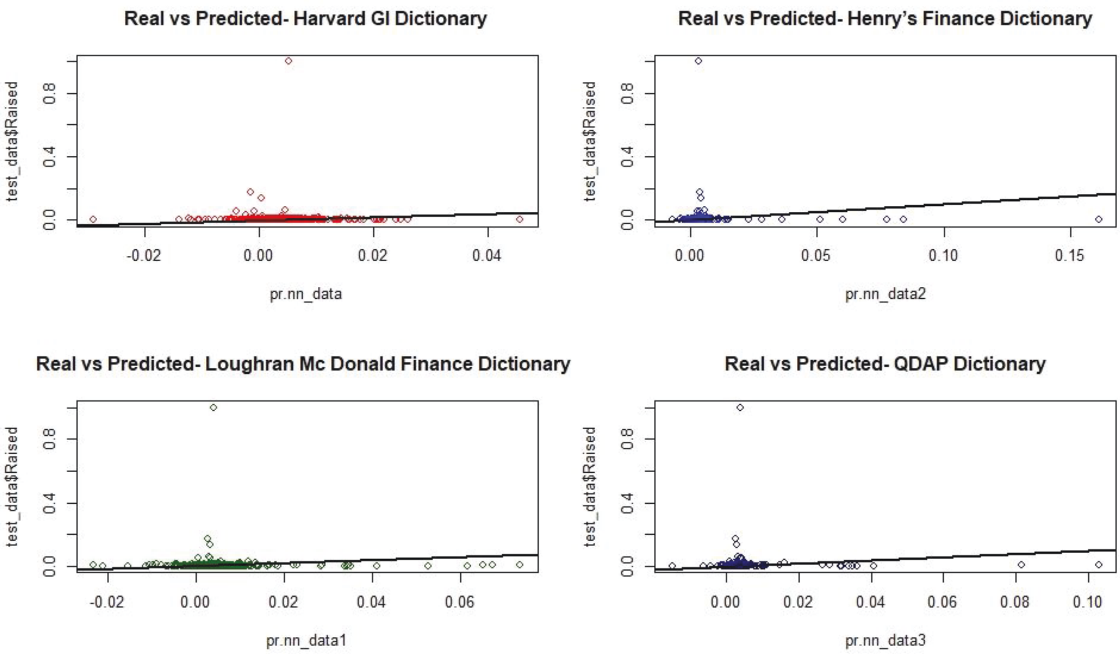
Appendix A.16. ANN model fitting on four different sentiment dictionaries (*90% training and 10% test data with 10 hidden layers and 5 neurons, and SMOG readability score*)



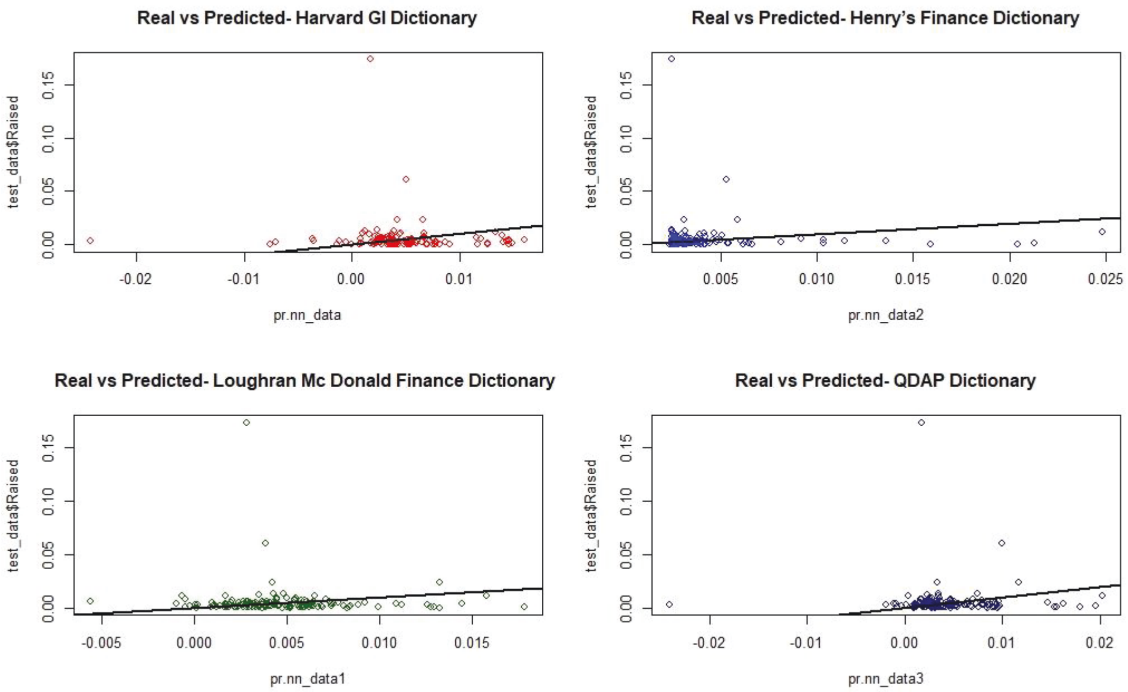
Appendix A.17. ANN model fitting on four different sentiment dictionaries (*75% training and 25% test data with 3 hidden layers, and FOG readability score*)



Appendix A.18. ANN model fitting on four different sentiment dictionaries (*75% training and 25% test data with 10 hidden layers and 5 neurons, and FOG readability score*)



Appendix A.19. ANN model fitting on four different sentiment dictionaries (*90% training and 10% test data with 3 hidden layers, and FOG readability score*)



Appendix A.20. ANN model fitting on four different sentiment dictionaries (*90% training and 10% test data with 10 hidden layers and 5 neurons, and FOG readability score*)