Measuring Tokenomics: Statistical Modeling and Analysis

———

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)

eingereicht an der

Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

Min-Bin Lin

Präsidentin der Humboldt-Universität zu Berlin:
　Prof. Dr. Julia von Blumenthal

Dekan der Wirtschaftswissenschaftlichen Fakultät:
　Prof. Dr. Daniel Klapper

Gutachter:　　　　1. Prof. Dr. Wolfgang Karl Härdle
　　　　　　　　　2. Prof. Cathy Yi-Hsuan Chen Ph.D

Tag des Kolloquiums: 24. Mai 2023

"[A] counterfeit coin in the hands of a beggar [. . .] Might it not multiply into real coins? Could it not also lead him to prison? A tavern keeper, a baker, for example, was perhaps going to have him arrested as a counterfeiter or for passing counterfeit money. The counterfeit coin could just as well, perhaps, be the germ of several days' wealth for a poor little speculator. And so my fancy went its course, lending wings to my friend's mind and drawing all possible deductions from all possible hypotheses."

— Charles Baudelaire, *Counterfeit Money*

Acknowledgments

Life's journey is often akin to a bus ride, where people board and disembark at various intervals, reminding us that change is a constant. I, however, genuinely appreciate anyone who has been a part of it.

I would like to express my deepest gratitude to my principal supervisor, Professor Wolfgang Karl Härdle, who has been constantly encouraging and guiding me on my way to the degree. Although I have experienced many ups and downs with him, I have ultimately found it beneficial and memorable. I owe a debt of gratitude to Professor Cathy Yi-Hsuan Chen, my second supervisor, for her enduring encouragement and fresh insights that have enriched both my research and my interactions with others.

I am grateful for the opportunity to have collaborated with many brilliant researchers during my PhD studies, including Professor Christian Hafner, Professor Zhou Chao, Dr. Fabian Bocart, Dr. Kainat Khowaja, Dr. Rui Ren, Yifu Wang, and Bingling Wang, whose collective wisdom has enhanced my academic journey.

My time at the IRTG 1792 (International Research Training Group 1792) has been a source of great inspiration and learning, thanks to the opportunity to work with a remarkable team. I extend my deepest appreciation to my colleagues, including Dr. Xinwen Ni, Dr. Junjie Hu, Francis Liu, Ratmir Miftachov, Konstantin Häusler, Jovanka Lili Matic, Ilyas Agakishiev, Anna Shchekina, Marius Sterling, and others, whose support and insights have been invaluable. Dr. Raphael Reule deserves special recognition for his outstanding contributions to my academic journey. I am grateful for his unwavering support, wise counsel, and generous encouragement, which have been invaluable to me.

I want to express my heartfelt appreciation to my family and friends for their unconditional support and understanding during my PhD.

Abstract

The emergence of distributed ledger technologies, specifically blockchain, has revolutionized the ways individuals interact with one another by enabling a "trust-less trust" through the use of peer-to-peer networks, cryptography, and consensus algorithms. This technology eliminates the need for intermediaries and provides secure and transparent means of conducting transactions. However, despite the growing popularity of crypto assets and their associated economy "Tokenomics," the public still has a limited understanding of such a technology, and much of the discourse surrounding it remains speculative.

The primary objective of this thesis is to investigate the fundamental principles of cryptocurrencies (cryptos) and non-fungible tokens (NFTs) and establish a correlation between the technology and its impact on the economy from a statistical and economic perspective. To achieve this objective, Chapters 2 and 3 delve into the influence of blockchain technology on the economic and functional performance of cryptos, using econometric models and clustering techniques.

Chapter 3 introduces an empirical framework that provides an insight for coin creators and investors into the interplay between cryptonomics and blockchain functionality, as well as market dynamics. We demonstrate that the economic performance of cryptocurrencies can be affected by the design of their underlying blockchain technology, with Ethereum as an example. By applying a Beta-t-EGARCH model with long-run and short-run volatility components, a more clear resolution is obtained for the identification of the casual relationships between volatility and other variables.

Chapter 4 examines the partial correlations of Bitcoin returns across nine different centralized exchanges from a high-frequency dynamic network perspective. The proposed Multivariate Heterogeneous AutoRegression for Crypto Markets (MHAR-CM) provides reasonable covariance estimates that consider the peculiarities of crypto markets. The chapter uncovers the presence of spillover risk and counterparty risk among these exchanges and develops a portfolio considering partial correlations.

In Chapter 5, a hedonic regression approach is used to construct the DAI digital art index for the NFT art market. We emphasize the leveling of outliers' impact with a one-step robust regression Huberization and dynamic conditional score model (DCS). The DAI index allows us to cultivate comprehension and observe the macro trend of this brand-new art market.

iv

This thesis connects emerging technologies and the economy through statistical modeling and analysis. By providing empirical evidence, we can observe how blockchain technology is transforming our perception of money, art, and various other industries. Essentially, to navigate the uncertainties and disruptions brought on by this transformation, it is imperative to have a thorough comprehension of the nature and scope of the changes.

Keywords: *Tokenomics, blockchain technology, crypto assets, index construction, econometrics, network analysis.*

Zusammenfassung

Das Aufkommen der Distributed-Ledger-Technologien, insbesondere der Blockchain, hat die Art und Weise, wie Menschen miteinander interagieren, revolutioniert, indem es durch den Einsatz von Peer-to-Peer-Netzwerken, Kryptografie und Konsensalgorithmen „trustless trut" ermöglicht. Diese Technologie macht Zwischenhändler überflüssig und bietet ein sicheres und transparentes Mittel zur Durchführung von Transaktionen. Trotz der zunehmenden Beliebtheit von Krypto-Assets und den damit verbundenen „Tokenomics" hat die Öffentlichkeit immer noch kein umfangreiches Wissen über die Funktionsweisen dieser Technologie, und ein Großteil des Diskurses bleibt spekulativ.

Das Hauptziel dieser Arbeit ist, die grundlegenden Prinzipien von Krytowährungen (Cryptos) und Non-Fungible Tokens (NFTs) zu untersuchen sowie eine Korrelation zwischen der Technologie und ihren Auswirkungen auf die Wirtschaft aus statistischer und wirtschaftlicher Sicht herzustellen. Um dieses Ziel zu erreichen, wird in den Kapiteln 2 und 3 der Einfluss der Blockchain-Technologie auf Ökonomie und Funktionsweise von Kryptowährungen anhand ökonometrischer Modelle und Clustering-Techniken untersucht.

Kapitel 3 untersucht Kryptowirschaft und Blockchain-Funktionalität anhand empirischer Methoden, insbesondere für Coincreatoren und Investoren. Wir zeigen am Beispiel von Ethereum, dass die wirtschaftliche Leistung von Kryptowährungen durch die Gestaltung der ihnen zugrunde liegenden Blockchain-Technologie beeinflusst werden kann. Durch die Anwendung eines Beta-t-EGARCH-Modells mit langfristigen und kurzfristigen Volatilitätskomponenten wird eine klare Lösung für die Identifizierung der kausalen Beziehungen zwischen der Volatilität und anderen Variablen erzielt.

Kapitel 4 untersucht die partiellen Korrelationen von Bitcoin-Renditen über neun verschiedene Zentralbörsen aus der Perspektive eines hochfrequenten, dynamischen Netzwerks. Die vorgeschlagene *Multivariate Heterogeneous AutoRegression for Crypto Markets*(MHAR-CM) liefert Kovarianzschätzungen, die die Besonderheiten der Kryptomärkte berücksichtigen. Das Kapitel zeigt Spillover- und Third-Party-Risiken zwischen diesen Börsen und entwickelt ein Portfolio unter Berücksichtigung partieller Korrelationen.

Kapitel 5 verwendet eine Hedonische Bewertungsmethode, um den DAI Digital Art Index basierend auf dem NFT-Kunstmarkt zu konstruieren. Ein besonderer Fokus liegt auf der Nivellierung der Auswirkungen von Ausreißern mit einer einstufigen ro-

busten Regressions-Huberisierung (one-step robust regression Huberization) und einem *dynamic conditional score model* (DCS). Der DAI-Index ermöglicht ein besseres Verständnis für diesen neuen Kunstmarkt und die Beobachtung seiner makroökonomischen Entwicklungen.

Diese Arbeit verbindet neue Technologien und Wirtschaft durch statistische Modellierung und Analyse. Durch die Bereitstellung empirischer Belege können wir beobachten, wie Blockchain-Technologie unsere Wahrnehmung von Geld, Kunst und verschiedenen anderen Branchen verändert. Um die Unsicherheiten und Störungen, die dieser Wandel mit sich bringt, zu bewältigen, ist es unerlässlich, die Art und den Umfang der Veränderungen genau zu verstehen.

Schlagwörter: *Tokenomics, Blockchain-Technologie, Krypto-Assets, Indexkonstruktion, Ökonometrie, Netzwerkanalyse.*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Through the development of distributed ledger technologies (DLTs) – i.e. blockchain technology, we have been witnessing a tremendous burst of creative potential that has catalyzed exceptional levels of innovation. The inception of Ethereum (ETH) has sparked a groundbreaking shift in the application of blockchain technology, surpassing the boundaries of progress that had been achieved since the emergence of Bitcoin (BTC), especially with its advanced smart contract capabilities. This technology has introduced a new form of trust across a wide range of services, paving the way for a new era in the digital economy, *Tokenomics*, which has reformed and transformed the way we approach transactions and exchange of value. Originating in financial markets with cryptocurrencies (cryptos) and non-fungible tokens (NFTs), this technology has extended its reach to encompass supply chains, consumer and business-to-business services, as well as publicly held registers, leading to a greater degree of transparency and accountability. Although our economy is undergoing an unprecedented revolution, there is still a prevailing sense of hype and optimism. The thesis embarks on an exploration of the fundamentals of cryptos and NFTs, with the aim of establishing a connection between the technology and the impact on its economy, while also providing a statistical and economic perspective on this revolutionary development.

Chapter 2 delves into the relationship between the underlying blockchain mechanism of cryptos and their distributional characteristics of time series. In addition to price, we place particular emphasis on utilizing actual block size and block time as the key operational features of cryptocurrencies. We leverage a range of distributional characteristics, such as Fourier power spectrum, moments, quantiles, global optima, and measures of long-term dependencies, risk, and noise to summarize information from crypto time series. With the hypothesis that blockchain structure explains the distributional characteristics of cryptos, we employ characteristic-based spectral clustering to group selected cryptos into five distinct clusters. Scrutinizing these clusters reveals that they share similar mechanisms, including fork origin, difficulty adjustment frequency, and block size nature. The findings provide crypto creators and users with a more in-depth understanding of the connection between blockchain protocol design and the distributional characteristics of cryptos.

Building on the previous chapter, **Chapter 3** studies the relationship between blockchain

technology and the economic performance of cryptocurrencies, with a focus on Ethereum as a case study. Since its inception in Nakamoto, 2008, crypto conjures a new form of medium of exchange. However, the unregulated nature of the cryptocurrency market has made it difficult to transition from hype to hope in comparison to traditional money markets. To address this issue, this chapter proposes an empirical framework consisting of 33 time series factors divided into three major components: functional characteristics, market dynamics, and economic attributes – based on Abadi and Brunnermeier, 2018; Buterin, 2021. We examine the definition and evolution of each factor and use Ethereum as an example to investigate the causal relationship between these factors. We visualize the causal dynamics using a Granger causal network and utilize Beta-t-EGARCH (Harvey, 2013) to analyze the long-run and short-run volatility components of the causal dynamics. Our analysis shows that functional characteristics often serve as the root cause of economic attributes, and the large holders of ETH, known as ETH whales, may potentially benefit from the current inequality in distribution and significantly impact its economic attributes.

Since crypto trading commonly occurs through centralized exchanges, **Chapter 4** provides a network perspective on the influence of these exchanges. The volatility and counterparty risks associated with centralized exchanges make cross-exchange crypto trading a risky venture. To better understand the risk spillover across exchanges, we introduce the Multivariate Heterogeneous AutoRegression for Crypto Markets (MHAR-CM) based on Corsi, 2009 and use BTC as an example. MHAR-CM provides reasonable covariance estimates that consider the peculiarities of crypto markets, such as trading 24/7 and the long-memory effect on return variations (Dwyer, 2015). We use the monthly dependence coefficients of MHAR-CM to assess the influence of different exchanges within high-frequency partial correlation networks. The findings suggest that an exchange's scale determines its influence on others, and the interconnectedness among these exchanges is stronger during extreme events in the BTC market. Furthermore, the volatile eigenvector centralities of Futures Exchange Ltd (FTX) could be a meaningful indication of its bankruptcy. To mitigate these risks, we construct portfolios that incorporate network information into risk diversification, and we show that portfolios that consider the dynamics of partial correlations or eigenvector centralities offer a promising result in terms of risk measures.

The emergence of non-fungible tokens (NFTs) has brought about a new era for digital art, driven by blockchain and smart contracts. This marketplace provides artists and art collectors with more security, flexibility, publicity, and freedom to monetize their works. However, the novelty of the market has resulted in speculation and economic uncertainty, as the market is not yet well understood. In **Chapter 5**, we provide a comprehensive understanding of the NFT art market by constructing the Digital

Art Index (DAI), a price index based on hedonic regression of the top 10 liquid NFT art collections. To address the issue of outlying price observations, we propose two innovative procedures – Huberization and DCS-t filtering – to create a more robust index. By analyzing each artwork's time-variant and time-invariant characteristics, we can identify price determinants and assess the intrinsic value of the market.

In *The Work of Art in the Age of Mechanical Reproduction* (Benjamin, 1968), he explored how the emergence of photography challenged traditional notions of unique, time- and place-bound artworks. Today, in our rapidly digitalizing world, we can apply similar arguments to the uncertainties posed by technological innovation. To navigate such uncertainties, it's essential to first comprehend and then determine how we will face the changes and disruptions from an innovation. Stiglitz, 1991 noted that "The reason that the invisible hand often seems invisible is that it is often not there." The invisible hand of the market is not always present or effective, but perhaps technological advancements can revolutionize the way our economy is regulated. An "invisible hand" could emerge in the form of new technologies?

All the supplementary materials and source codes of this dissertation are available in the Q^2 ecosystem: Quantlet.com and Quantinar.com .

Bibliography

Abadi, J., & Brunnermeier, M. (2018). Blockchain economics.

Benjamin, W. (1968). The work of art in the age of mechanical reproduction, 1936 (H. Zohn, Trans.). In, *Illuminations: Essays and reflections* (pp. 217–252). Houghton Mifflin Harcourt.

Buterin, V. (2021). Why sharding is great: Demystifying the technical properties. https://vitalik.ca/general/2021/04/07/sharding.html

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Dwyer, G. P. (2015). The economics of bitcoin and similar private digital currencies. *Journal of Financial Stability*, *17*, 81–91.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: With applications to financial and economic time series* (Vol. 52). Cambridge University Press.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *white paper*.

Stiglitz, J. E. (1991). The invisible hand and modern welfare economics.

Chapter 2

Blockchain Mechanism & Distributional Characteristics of Cryptos

All the supplementary materials and source codes of this chapter are available in the Q^2 ecosystem: Blockchain_mechanism and .

2.1 Introduction

Cryptocurrency (crypto) is a digital asset designed to be as a *medium of exchange*, wherein individual coin ownership is recorded in a digital ledger or computerized database. Its creation of monetary units and verification of fund transactions are secured using encryption techniques and distributed across several nodes (devices) on a peer-to-peer network. Such technology-enhanced and privacy-preserving features make it potentially different from other existing financial instruments and has attracted the attention of many investors and researchers (Härdle et al., 2020). Many studies have investigated the similarity between a pool of cryptocurrencies in order to classify the important features of digital currencies. For example, Blau et al., 2020 has concluded that the top sixteen most active cryptocurrencies co-move with bitcoin. Researchers have also focused on describing the price behavior of cryptos using economic factors (Ciaian et al., 2016; Sovbetov, 2018). However, owing to the unique technology of cryptocurrencies, there still exists a gap between the creators of blockchain mechanism and users operating the financial market of the cryptocurrencies and through this research, we aim to take a step towards mitigating that gap.

We specialize our research on the following research questions. First, we characterize crypto behavior using distributional characteristics of time series data. Also, instead of using the prices alone, we use actual block time and block size to incorporate the operational features of cryptos. Second, we hypothesize that the blockchain

structure that the coin attaches plays a pivotal role in explaining the behavior. More explicitly, we investigate the extent to which blockchain structure leads to explaining the distributional characteristics. Using a characteristic based clustering coupled with spectral clustering technique, we group the selected cryptos into a number of clusters, and stratify the mechanisms that make the coins within the particular cluster showing the same behavior in price, actual block time, and actual block size, respectively.

When studying cryptocurrencies, many researchers only focus on crypto price and daily returns (Hou et al., 2020; Trimborn & Härdle, 2018). While price is important when cryptos are used as a medium of payment, it is definitely not the only measure for evaluation of cryptocurrencies. For example, many low price coins are highly traded and many coins that are not used as medium of payment have low prices, e.g., XPR and Dogecoin. Cryptos were introduced to serve various purposes, and the purpose of the coin does matter. This makes it necessary to use other time series while studying crypto markets. In this research, we propose to use actual block size and actual block time alongside price.

Actual block size is the average actual size "usage" of a single block of data storage for one day. Since a block comprises transaction data, it can represent the status of how a blockchain mechanism allocates transactions to a block. We consider it a measure of scalability of the system. A well-functioning blockchain should be able to level the transaction arrivals. Transaction distribution within a day for any crypto needs such balancing because it affects miner's rewards and hence the demand of the coin. An ideal block size would keep confirmation times from ballooning while keeping fees and security reasonable. Therefore, actual block size of cryptos can provide insight into the behavior of cryptos.

Actual block time, on the other hand, measures the consistency and performance of the system. It is defined as the mean time required in minutes for each day to create the next block. In other words, it is the average amount of time for the day a user has to wait, after broadcasting their transaction, to see this transaction appear on the blockchain. Think of crypto markets as a fast food franchise and miners as customers who have to wait a certain time to make the purchase. If the waiting time is shorter on certain days while on other instances, the customers have to wait much longer, there is a discrepancy in the system. Analogously, the time series of block time, which is the distribution of waiting time, can be seen as a service level of the whole system, and it is necessary to maintain as the users' expectation or target block time set by the system depend on it.

The idea of investigating the underlying blockchain mechanism, a cornerstone of crypto

technology, and its connection to the crypto behavior is still in its infancy. One of the first endeavors in explaining this relationship was made by L. Guo et al., 2018 who highlight that the fundamental characteristics of cryptocurrencies (e.g., algorithm and proof type) have a vital role in differentiating the performance of cryptocurrencies. They develop a spectral clustering methodology to group cryptos in a dynamic fashion, but their research is limited in the exploitation of blockchain characteristics. With a similar spirit, Iwamura et al., 2019 start by claiming that high fluctuation is a reflection of the lack of flexibility in the Bitcoin supply schedule. They further strengthen their arguments by considering the predetermined algorithm of cryptos (specifically, the proof of work) to explain the volatility in cryptocurrency market. Zimmerman, 2020 argue in their work that the higher congestion in blockchain technology leads to higher volatility in crypto prices. They claim that the limited settlement space in blockchain architecture makes users compete with one another, affecting the demand. In his model, the value of cryptos is governed by its demand, making the price sensitive to blockchain capacity.

These research results, albeit true, are limited to a particular set of cryptocurrency mechanisms and do not thoroughly explain the dynamics of cryptocurrencies. Also, most of the papers only use price as a proxy of behavior. We advance the previous findings by incorporating a rich set of underlying mechanisms and connecting them to multiple time series. We take a deep dive into eighteen cryptos with a variety of mechanisms- concluded in Garriga et al., 2020 – from a technical perspective to summarize their mechanism and algorithm designs using variables, such as consensus algorithm, type of hashing algorithm, difficulty adjustment frequency and so on.

We investigate a relationship between the underlying blockchain mechanism of cryptocurrencies and the distributional characteristics. Using a characteristic-based clustering technique, we cluster the selected coins into a number of clusters and scrutinize the compositions of fundamental characteristics in each group. We observe that the clusters obtained from these time series indeed share a common underlying mechanism. Through empirical evidence, we show that the cryptos forked from same origin and same consensus mechanism tend to become part of the same clustering group. Furthermore, the clusters obtained by the time series of block time have the same hashing algorithms and difficulty adjustment algorithms. Furthermore, a similar nature (static or dynamic) of block size was observed within clusters obtained by the time series of actual block size. We conclude with empirical evidence that the crypto behavior is actually linked with their blockchain protocol architectures.

The implications of this study are abundant. The creators of cryptocurrencies can manage the impact of blockchain underlying mechanisms on the corresponding distri-

butional characteristics, in a consideration of adoption rate of invented coins. From the users' perspective, they can make an optimal decision in which coins should be adopted while concerning the price fluctuation.

This chapter proceeds as follows. Section 2 discusses data source and the underlying mechanisms of the cryptos. Section 3 presents the methodology used for classifying characteristics of time series and the clustering algorithm. Section 4 provides an illustration of analysis results. Section 5 concludes and provides several avenues for future research.

2.2 Data description

According to CoinMarketCap, currently there are over 7,000 cryptocurrencies and their total market capitalization has surpassed USD$400 billion as of November 09, 2020. Most studies have focused on the mainstream coins (e.g., Bitcoin, Ethereum), and little has been investigated on the coins which have been introduced and featured with a diverse blockchain mechanisms and invented technologies. The work of X. Guo and Donev, 2020 is one of the exceptions. In this study, 18 cryptos with different sets of blockchain mechanisms have been examined –Bitcoin, Bitcoin Cash, Bitcoin Gold, Bitcoin SV, Blackcoin, Dash, Dogecoin, Ethereum, Ethereum Classic, Feathercoin, Litecoin, Monero, Novacoin, Peercoin, Reddcoin, Vertcoin, XRP (Ripple), and Zcash. We explore an interplay between distributional characteristics of cryptos and the blockchain mechanism. We discuss the key characteristics of blockchain mechanisms and the time series data in this section.

2.2.1 Underlying mechanism

Most of the cryptos nowadays apply blockchain-based systems, in which transactions are grouped into blocks and cryptographically interlinked to form a back-linked list of blocks containing transactions. The transactions are validated using the nodes within the crypto peer-to-peer network through a majority of consensus directed by algorithms instead of a central authority's approval. In such an operation process, many algorithmic mechanisms are required to govern the performance and outcome of a crypto system. Some key blockchain-based characteristics are discussed below:

Fork. It occurs as user base or developers conduct a fundamental or significant software change, see Figure 2.1. There are two types of forks – soft and hard forks. The former is an update to the protocol architecture, and then all the nodes are enforced to follow to proceed with the operations of a crypto. The latter one creates a duplicate of the origin blockchain and modifies the copy to meet the desired quality (e.g., safety, scalability).

In this case, a new crypto can be generated accordingly. For example, Peercoin network facilitates an alternative consensus mechanism –proof-of-stake (PoS) to Bitcoin's proof-of-work (PoW) system for reducing dependency on energy consumption from the mining process (King & Nadal, 2012).

Going beyond a digital currency, Ethereum establishes an open-ended decentralized platform for diverse applications such as decentralized applications (dapps) and smart contracts (Buterin, 2014).
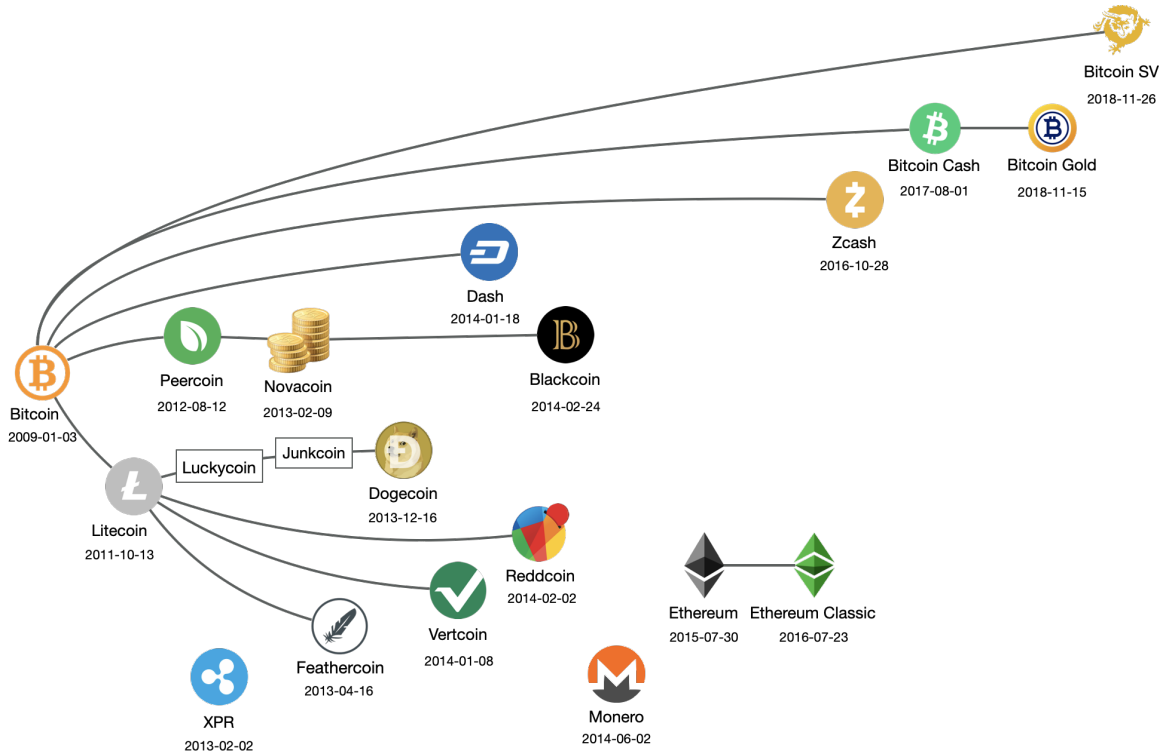
Figure 2.1: **Blockchain software forks in cryptocurrency**.

Consensus mechanism. In order to establish an agreement on a specific subset of the candidate transactions, the consensus mechanism provides a protocol for a large number of trust-less nodes in a decentralized blockchain network. For instance, PoW (Proof-of-Work, as adopted by e.g., Bitcoin, Litecoin) achieves consensus with a competition among miners on solving computational puzzles, which consume numerous computational resources; and PoS (Proof-of-Stake, as adopted by e.g., Peercoin, Blackcoin) randomly assigns a block creator (transaction validator) with probability proportional to their coins staked.

Hashing algorithm. It is a mathematical algorithm that encrypts a new transaction (or a new block) into a fixed length character string, known as hash value, and later interlinks this string with a given blockchain to ensure the security and immutability of a crypto. Various hashing algorithms are implemented in cryptos such as SHA-256,

Scrypt and Equihash. These provide different degree of complexity to blockchain operations.

Difficulty adjustment algorithm. It is an adaptive mechanism which periodically adjusts the difficulty toward hashrate to target an average time interval between blocks, known as target block time or target confirmation time. It regulates the creation rate of a block and maintains a certain number of outputs of a blockchain. Such a mechanism is commonly seen in a PoW framework. An example from Bitcoin is shown in Figure 2.2 where its difficulty adjustment algorithm, known as DAA, modifies the difficulty every 2016 blocks to meet the target block time of 10 minutes.
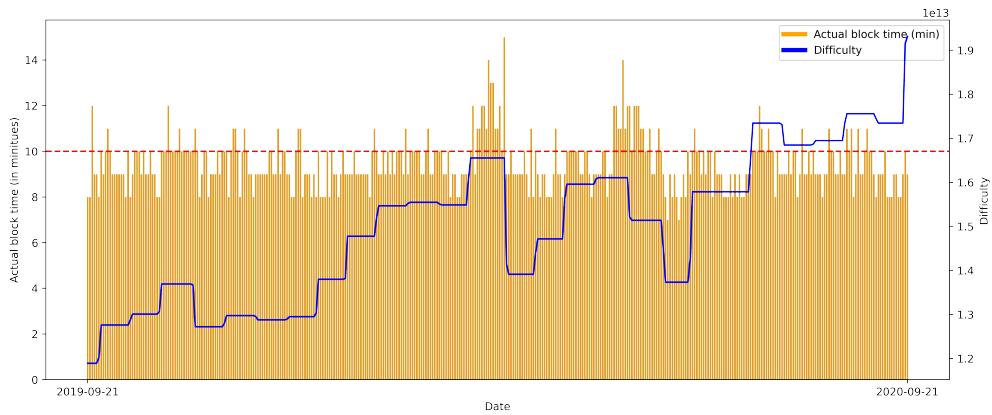
Figure 2.2: **Bitcoin's difficulty adjustment toward actual block time**.

2.2.2 Time series

The data applied in this paper are collected from Bitinfocharts. These time series are composed of data points observed daily from the genesis date of each crypto. The lengths of these time series are thus varied coin by coin, but as explained in the next section, we continue to use the whole time series for each coin.

Price. Much previous literature has been triggered by the substantial fluctuations in crypto prices. In this study, we investigate 18 crypto prices in USD on daily time series. Among these 18 cryptos, Bitcoin has been dominant and Reddcoin has the lowest price on balance, see Figure 2.3. We characterize these price time series in Table 2.A.1. Most of these coins (i.e., Bitcoin, Ethereum, Bitcoin Cash) have high fluctuations in price; while some coins (i.e., XRP, Blackcoin) tend to be steady.

Actual block time. It is the mean time required in minutes for each day to create the next block. In other words, it is the average amount of time for the day a user has to wait, after broadcasting their transaction, to see this transaction appear on the
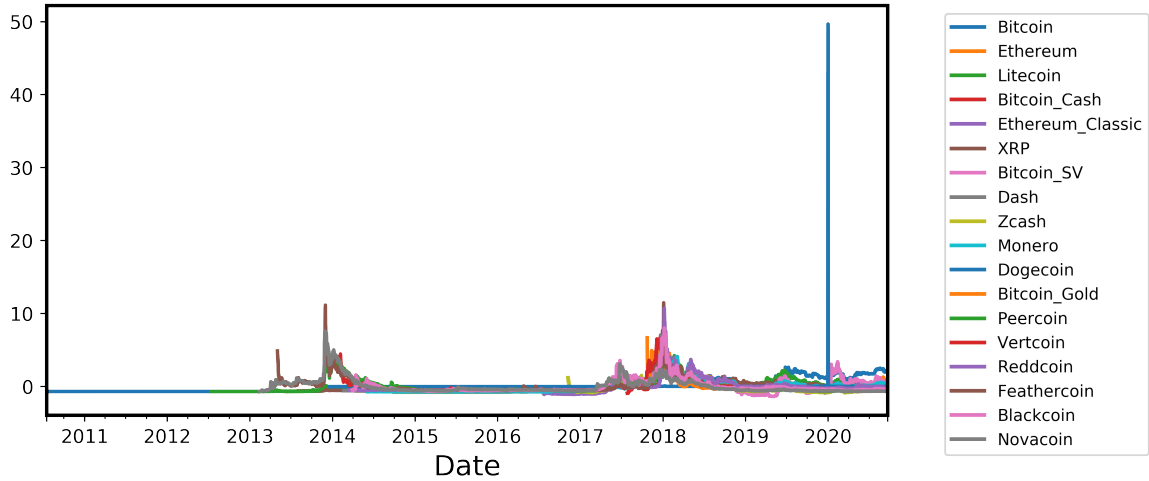
Figure 2.3: **Time series of prices of the 18 cryptos**.

blockchain. Some literature also refers to it as confirmation time. It can be considered as a service level indicator for cryptos which should be maintained by underlying mechanisms. Most of the coins discussed in this paper tend to have lower block time compared with Bitcoin, see Figure 3.A.2. Furthermore, many coins show outliers in observations and this can indicate that the extreme events appear in the blockchain system. The underlying mechanisms can be ineffective to accommodate the current system demand. The distributional characteristics for time series of actual block time are presented in Table 2.A.2. The data for XRP is missing, but its designed block time is around 5 seconds per transaction.

Actual block size. It is defined as the average actual size "usage" of a single block of data storage for one day. Since a block comprises transaction data, it can represent the status of how a cryptocurrency mechanism allocates transactions to a block. In this study, as introduced in Section 1, we consider it as an indicator for the stableness in scalability of a crypto. In Figure 2.5 shows that most of the cryptos under study have smaller block size usage than Bitcoin, except Bitcoin SV. The plot also depicts that almost all the coins have outliers. These outliers can lead to the imbalance in transaction fee and reward, which can influence the ecosystem of a crypto. The characteristics for block size time series are shown in Table 2.A.3. XRP does not have a typical blockchain structure; hence, there is no block size data in the study. The data for Peercoin is missing.

2.3   Methodology

In order to investigate the relationship between the underlying blockchain mechanism of cryptocurrencies and the distributional characteristics of cryptos as a proxy of behavior,
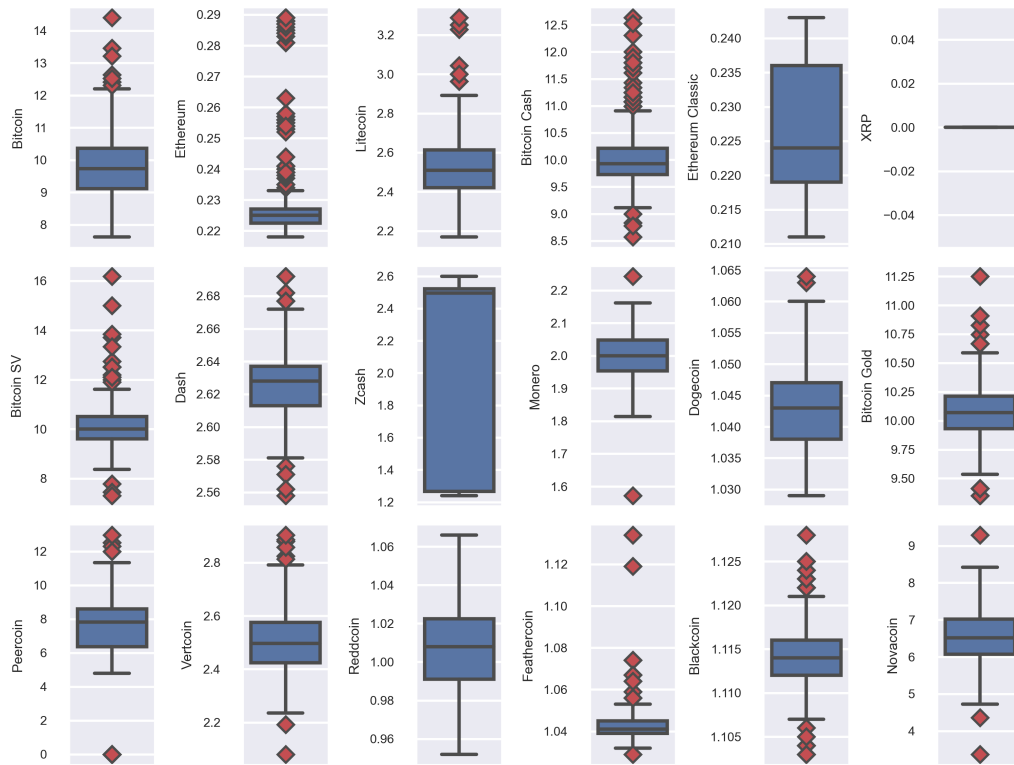
Figure 2.4: **Actual block time in minutes**.
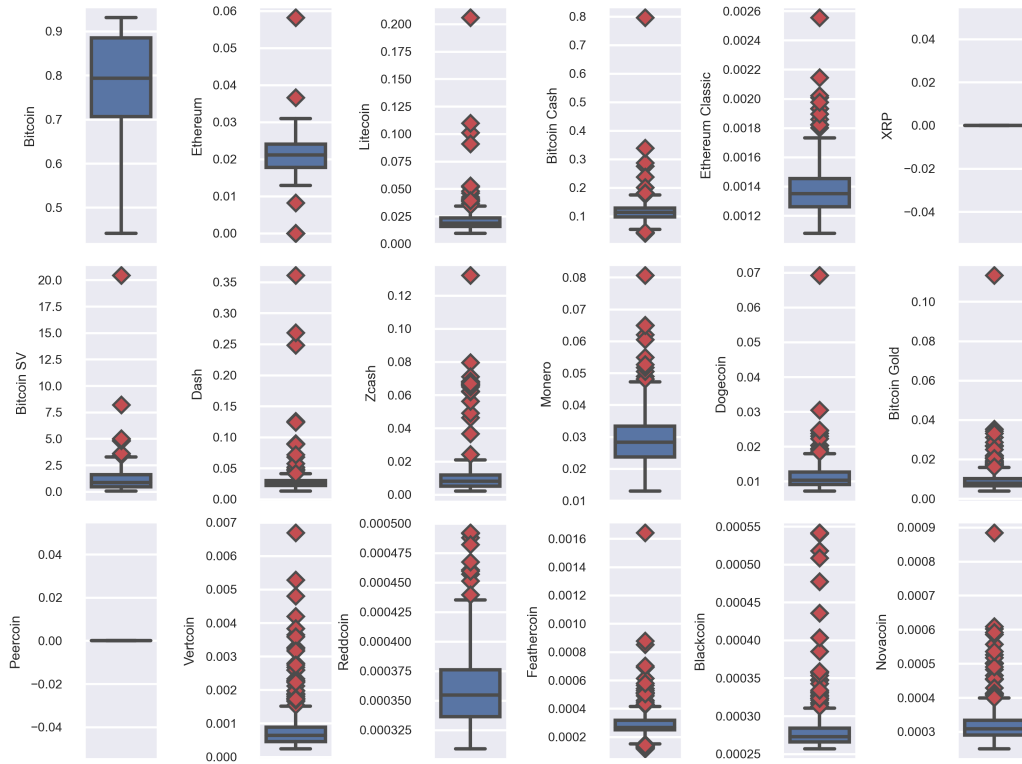
Figure 2.5: **Actual block size in megabytes**.

we aim to group them into several clusters and scrutinize the compositions of features in each group. These blockchain-based features manifest the underlying mechanism of how the cryptos operate transactions on their chains, and subsequently govern the price, actual block size and block time. As described in the previous section, we use the time series data of 18 different cryptos with a range of different mechanisms.

The time series data available for the cryptos is subject to numerous limitations. The most important one of them is that different coins were introduced at different time points, therefore, the data available for each coin has different lengths. For the clustering problems (Aghabozorgi et al., 2015), defining the distance metric between times series with various lengths is not conventional. For many analytical problems, this issue is easily tackled by truncating the time series to the shared sample period. We refrain from doing so because, in the analysis of cryptocurrency prices, the evolution of the data in time is highly crucial for an investigation of the short-term and long-term dynamics and therefore, truncating the time series would lead to loss of important information. Hence, we deal with the time series data of cryptos with different lengths and do not directly impose a distance metric on the input data points.

Furthermore, characterizing the behavior of a time series in terms of a single quantitative attribute (such as range-based volatility) has its limitations. The chosen attribute usually captures the dynamics of time series in one particular aspect, which may not be sufficient to encompass an entire behavior or introduces a biased assessment. This becomes particularly true in the problems of crypto classification and clustering where these attributes, used as a similarity measure, are very diverse, resulting in weak robustness in the results.

To cope with these limitations, we resort to the characteristic based clustering method proposed by Wang et al., 2005. It was recently applied by Pele et al., 2020 for classifying cryptos to distinguish them from traditional assets. This method recommends incorporating various global measures describing the structural characteristics of a time series for a clustering problem. These global measures are obtained by applying statistical operations that best represent the underlying characteristics. Also, by extracting a set of measures from the original time series, we simply bypass the issue of defining a distance metric. It's understood that the global measures are domain-specific. Employing a greedy search algorithm, Wang et al., 2005 selects the pivotal features in the clustering tasks. In our case, we import the experts' discretion on the choice of features as distributional characteristics which best represent the dynamics of cryptocurrencies.

2.3.1 Distributional characteristics

We choose a variety of measures for our analysis. Starting from the first four moments and quantiles that characterize the distribution and symmetry of the data, we include the statistics for concluding the global structure such as global optimum, as well as the measures for long-term dependencies, risk, and noise. The selected features are mean, standard deviation, skewness, kurtosis, maximum, minimum, first quartile, median, third quartile, 1% and 5% extreme quantiles as a measure of downside risk, linear trend, intercept, autocorrelation for long-term dependency, self-similarity using Hurst exponent and chaos using Lyapunov exponents.

Mean. It is the simplest attribute that can be driven from any data. Essentially, it is the most basic estimate of a value that any variable in our data will assume on any given day. We calculate this first moment by treating each value in the time series as a separate data point and averaging them.

Standard deviation. It is the most common measure of dispersion, or how spread out the data are around the mean. It is sometimes considered as a measure of stability, since the phenomenon of stability is usually manifested in the stability of the average of the process. Since stability largely accounts for the behavior of any time series, this statistics is one of the key measures for the volatility of crypto data in our analysis.

Skewness. The symmetric (or asymmetric) shape of a distribution is indicated by skewness. For example, a high, non-zero skewness coefficient for the crypto price data means that for many time points, the price deviated significantly from the average price. It is calculated with the adjusted Fisher-Pearson standardized moment coefficient G1.

Kurtosis. This fourth moment measures how heavily the tails of a distribution differ from the tails of a normal distribution. In other words, for crypto data, this statistic identifies whether the tails of the distribution of our time series contain extreme values. It is calculated with the adjusted Fisher-Pearson standardized moment coefficient G2.

Quantiles. Quantiles are another measure of distribution of data that provide information about how the values are distributed across the spectrum. For the time series data, such as that of prices of cryptocurrencies, these measures give the values that lie 25%, 50%, or 75% way up if the prices are sorted. The extreme quantiles of 1% and 5%, also known as Value-at-Risk(VaR99 and VaR95), are included in analysis for capturing downside risk.

Maximum and minimum. These simple, yet two of the most informative measures, provide information about the highest and the lowest value that the time series has

realized in its course.

Linear trend and intercept. These two features are usually identified as one of the two parameters of linear regression. For the time series data, the intercept simply gives the initial value, for example the price of crypto when it was first introduced. The linear trend, or slope, shows the general tendency of the data to increase or decrease during a long period of time. It is calculated by fitting a linear least-squares regression for the values of the time series versus the sequence from 0 to length of the time series minus one. This feature assumes the signal to be uniformly sampled. The time stamps are not used to fit the model.

Autocorrelation. It is serial correlation of a time series with a delayed version of itself. It measures the similarity between observations within a time series. For the example of crypto data, a large autocorrelation coefficient could be indicative of seasonal behavior within time series. It is calculated according to the formula

$$\frac{1}{(n-l)\sigma^2} \sum_{t=1}^{n-l} \left(X_t - \mu\right)\left(X_{t+l} - \mu\right)$$

where n is the length of the time series, X_i, σ^2 its variance and μ its mean. I denotes the lag. We fix the lag at 1.

Self-similarity. It is a measure of long-term memory of time series. It is an important attribute to include because it detects the persisting behavior of the cryptos. It is usually determined using the Hurst exponent. A Hurst exponent of 0.5 represents white noise. It becomes greater for time series that exhibit some positive dependency on previous values. For negative dependencies, it becomes less than 0.5.

Chaos. Chaos in the context of time series data means that the data has different functions in different situations. It is a measure for recognizing and quantifying the nature of underlying random behavior of the time series. We use Lyapunov exponents as a proxy to measure chaos in our time series, details of which can be found in Parlitz, 2016.

2.3.2   Frequency domain

We further extend the methodology by including the power spectrum of time series as an additional measure. The power spectrum is obtained in this work using Fast Fourier Transform (FFT). For computational ease, the discrete Fourier transform (DFT) has been formalized as a linear operator that maps the data points in a discrete input signal $X\left\{x_1, x_2, \cdots, x_n\right\}$ to the frequency domain $f = \{f_1, f_2, \cdots f_n\}$,

For a given time series X of n time points, sine, and cosine functions are used to get the coefficients $\omega_n = e^{-2\pi i}/\eta$ and the frequencies are calculated using the matrix multiplication:

$$
\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega_n & \omega_n^2 & \cdots & \omega_n^{n-1} \\ 1 & \omega_n^2 & \omega_n^4 & \cdots & \omega_n^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_n^{n-1} & \omega_n^{2(n-1)} & \cdots & \omega_n^{(n-1)^2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \tag{2.1}
$$

This matrix multiplication involves $\mathcal{O}(n^2)$ and makes DFT computationally expensive. FFT is a fast algorithm to compute DFT using only $\mathcal{O}(n \log n)$ operations (Brunton & Kutz, 2019). The power spectrum of this signal is the normalized squared magnitude of the f, and it indicates how much variance of the initial space each frequency explains (Brunton & Kutz, 2019). Including the power spectrum as a feature for characteristic based clustering allows capturing the variability in the time signal that is not explained by any other measure.

Accumulating all the aforementioned features in a vector gives in a reduced dimensional representation of the time series of each crypto. These vectors are then used to cluster the cryptos into groups using spectral clustering. Spectral clustering exploits the eigenvalues of the similarity matrix to cluster, and results in more balanced clusters than other techniques that were employed during the process. For details related to spectral clustering, the readers are recommended to follow the tutorial on spectral clustering by von Luxburg, 2006. The results of the above methodology are discussed in detail in the next section.

2.4 Empirical result

In this section, we showcase the result from the characteristic based clustering individually on the crypto price and operational features–which are constructed with price, block size "scalability" and block time "service level" time series. We explore the clustering results and classify them with the underlying mechanisms of the investigated 18 cryptos. The 18 cryptos are: Bitcoin, Bitcoin Cash, Bitcoin Gold, Bitcoin SV, Blackcoin, Dash, Dogecoin, Ethereum, Ethereum Classic, Feathercoin, Litecoin, Monero, Novacoin, Peercoin, Reddcoin, Vertcoin, XRP, and Zcash.

We calculate the characteristics for each of these cryptos for prices, block size and block time separately. The results of all other attributes except the FFT are summarized in

Tables 2.A.1, 2.A.2, 2.A.3 correspondingly. Note that the data for XRP is not available for the block size and block time, and for Peercoin block size is missing as described before in Section 3.

After calculating the attributes and FFT power spectrum described in Section 3, the feature space is 216 dimensional (200 dimensional vectors of power spectrum and 16 characteristics), visualization of which is not possible. We project the feature space into a three-dimensional space using principal component analysis (PCA), and the results of which are exhibited for an intuitive understanding. We discuss each of the clustering in detail below. Moreover, in order to avoid a monopoly outcome and sustain a certain level of interpretability, we impose the maximum number of the clusters to avoid a single coin case in each cluster.

2.4.1   Clustering with crypto prices

Table 2.A.1 shows that, as expected, Bitcoin has the highest average price and highest standard deviation, due to the high magnitude of its prices. The VaR99 and VaR95 for Bitcoin are, however, very low, showing a low downside risk of Bitcoin. On the contrary, Bitcoin Cash, Bitcoin SV, Bitcoin Gold and Zcash all show high value at risk. This could be due to low persistence of risk shocks (de Souza, 2019; Katsiampa et al., 2019). The high positive coefficients of self similarity for all the coins show high dependency on the previous time values. The high autocorrelation further confirms the presence of long-term dependencies of the time series. The Lyaponov exponent as a measure of chaos exceeds 0 for all the time series, which shows unstable dynamics throughout the prices of cryptos.

The characteristics of Dogecoin in Table 2.A.1 assume very low values, unlike any other coin because the prices of Dogecoin are very low, despite it being a popular coin. This can be due to high supply of the coin with no limit on the total number of coins created. The coin also has no technical innovations, which is considered as one of the reasons why the coin has such a small price. Hence, the uncontrolled underlying mechanism of the coin has a significant impact on the prices, despite the high trading volumes of the coin. The same can be concluded for XRP and Reddcoin, which also have a very high maximum supply that is reflected in their very low prices.

Using characteristic based clustering on price time series, we have the result with 5 clusters as below:

0. Bitcoin, Dash

1. Bitcoin SV, Zcash

2. Bitcoin Cash, Bitcoin Gold

3. Ethereum, Litecoin, XRP, Monero, Peercoin, Vertcoin, Reddcoin, Feathercoin, Blackcoin

4. Ethereum Classic, Dogecoin, Novacoin
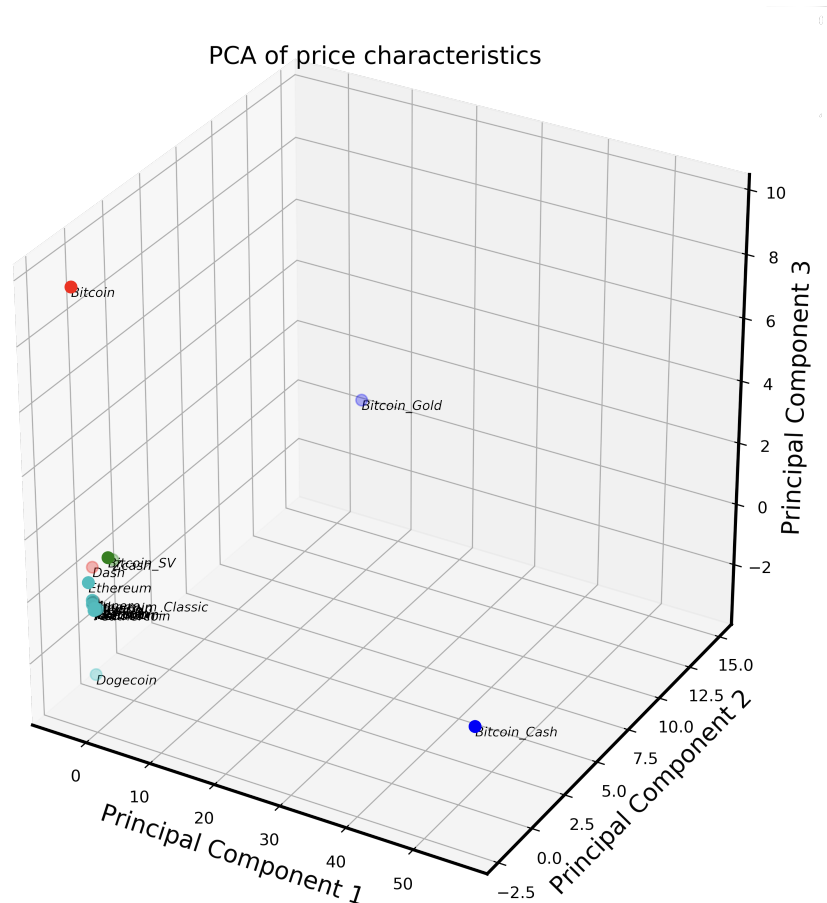
Figure 2.6: **Cluster visualization**. **0**, **1**, **2**, 3, **4** of cryptos based on the prices.

Most of the coins are close to each others in a three-dimensional space, see Figure 2.6. Except Dash, all the altcoins are in a different cluster than Bitcoin. Bitcoin Cash and Bitcoin Gold, which principally inherit the protocol architecture from Bitcoin, are clustered together, but not centered around with other coins. However, Bitcoin SV–which is a fork from Bitcoin Cash and mainly increases the designed block size to lower the transaction fee as a main software change–is not in the same cluster. This indicates that even as a crypto adopts a similar blockchain mechanism to the other crypto, it might have different price dynamics than its origin.

XRP, Monero, Peercoin, Reddcoin, and Blackcoin which apply significantly different blockchain protocols in their governance types and consensus mechanisms are in the same cluster. Specifically, XRP, Monero and Peercoin are private-based blockchain

which possesses a stronger moderator to control the entrants (users or investors) to their network. Peercoin, Reddcoin, and Blackcoin, instead of using PoW as their consensus mechanisms, employ PoS which does not depend on miners' effort to create a block. So, coin supply and demand can reach an equilibrium without the interference of miners, which leads to higher transaction costs. Moreover, the forks from Litecoin–Vertcoin, Reddcoin and Feathercoin are within the same cluster with Litecoin.

Ethereum Classic is, in fact, the version of Ethereum that existed before the hard fork of Ethereum resulting after the DAO attack, but it is not within the cluster with Ethereum.

2.4.2   Clustering with actual block time

The block time here is measured in minutes. Likewise, we apply the characteristic based clustering on the data and conclude them into 5 clusters as below.

0. Dogecoin, Feathercoin

1. Ethereum, Litecoin, Ethereum Classic, Dash, Zcash, Monero, Blackcoin

2. Bitcoin, Bitcoin Cash, Vertcoin

3. Bitcoin SV, Bitcoin Gold, Novacoin

4. Peercoin, Reddcoin

The result is correspondingly visualized in Figure 3.A.2. The figure indicates that Peercoin and Reddcoin lie far away from other coins (marked by cyan cluster). They are clustered in the same group because they both use PoS and their initial block takes the maximum time to be added, as shown by the maximum and intercept characteristics in Table 2.A.2. This shows that even though the coins have lower actual block time later (with low mean), their behavior is still the similar, resulting them in the same cluster. Also, the cryptos using PoS tend to lower the complexity of their hashing algorithms since it is not required for miners to spend computational effort on them. The difficulty adjustment algorithms of theirs are purely used as a mechanism for maintaining the certain service level for users without considering hashrate from miners. Their block time performance is relatively stable after the initialization. Here we emphasize that the initial price, block time and block size that are usually characterized by the underlying mechanism play a pivotal role in determining the price behavior of cryptos. This is why we did not truncate the time series, as mentioned in the Section 4.
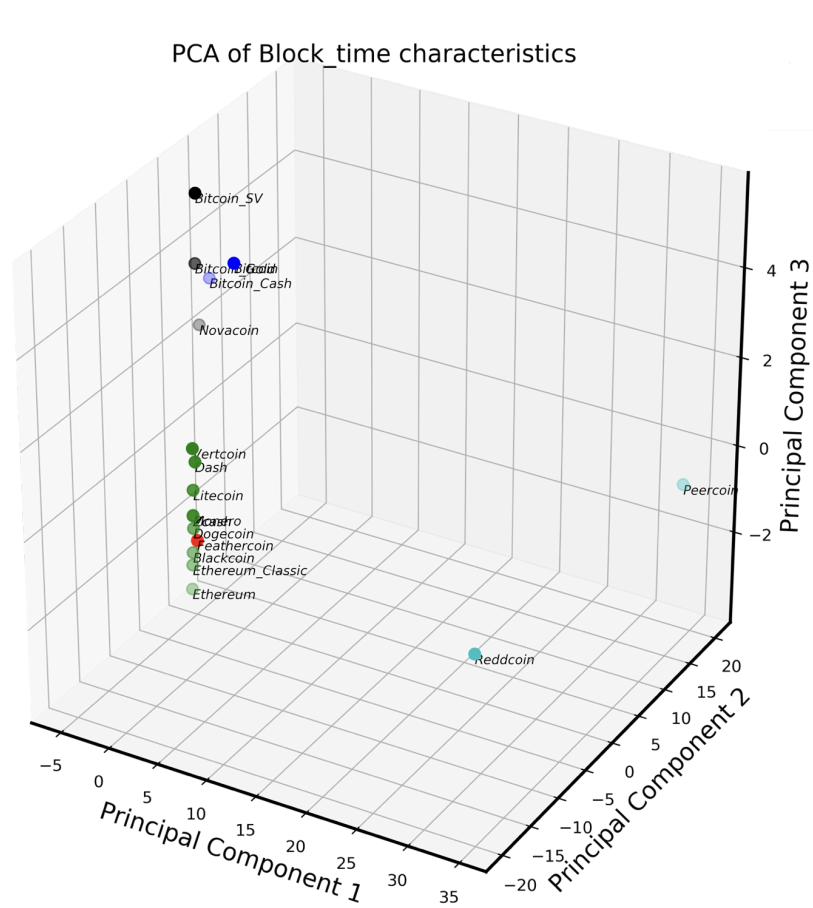
Figure 2.7: **Cluster visualization**. 0, 1, 2, 3, 4 of cryptos are based on block time.

Though Bitcoin, Bitcoin Gold, Bitcoin Cash and Bitcoin SV are not completely grouped into the same cluster, they are close to each others in the three-dimensional space see Figure 3.A.2. They apply the same hashing algorithm–SHA-256 and also with the same expected block time for their difficulty adjustment algorithms. Let's call attention to forks again. Dogecoin and Feathercoin are both forked from Litecoin with the Script-based hashing algorithm and difficulty adjustment frequency after large number of blocks–240 and 504 blocks. Litecoin is in a different cluster because the frequency is much higher as 2016 blocks. Given the cryptos forked from the same origin coins, their block time can be found in the same group, likewise Ethereum and Ethereum Classic.

2.4.3   Clustering with actual block size

As previously done for price and block time, we use the characteristics based clustering and grouped these cryptos into 5 clusters according to the characteristics of their time series. The block size here is measured in bytes for a better data representation. As stated before in Section 3, XRP and Peercoin data are missing due to the mechanism

design and incomplete data from the source, respectively. The clustering result is
shown as below and the corresponding visualization is in Figure 2.5.

0. Zcash, Bitcoin Gold, Reddcoin, Novacoin

1. Ethereum, Ethereum Classic, Dogecoin

2. Bitcoin Cash, Bitcoin SV

3. Bitcoin, Dash, Monero, Feathercoin

4. Litecoin, Vertcoin, Blackcoin
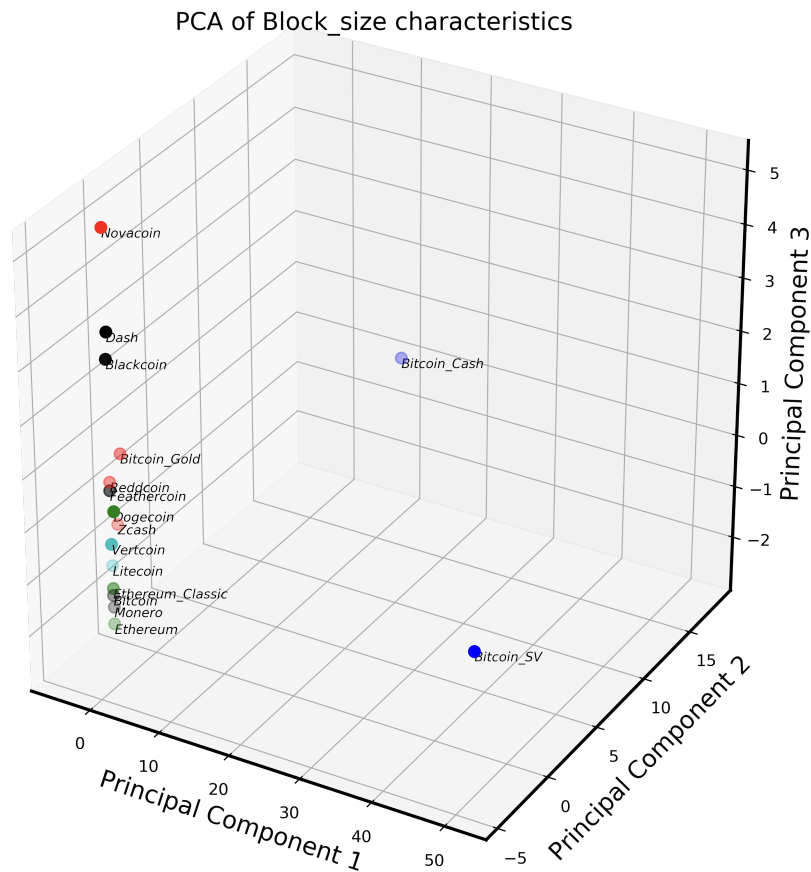
Figure 2.8: **Cluster visualization**. 0, 1, 2, 3, 4 of cryptos are based on block size.

The actual block size (usage) of these cryptos does rarely meet their designed block
size limit (capacity), except for Bitcoin that it nearly outstretches its limit, 1 megabyte,
see Table 2.A.3. In this case, it raises an issue: Can increasing crypto's block size
limit improve scalability? For example, Bitcoin SV enlarges dramatically its limit
to 128 megabytes, but it is out of the necessity for such a design. Likewise, Bitcoin
Cash, which Bitcoin SV forks from, has its limit as 32 megabytes. These two coins

are, therefore, clustered together. Moreover, instead of having a static block size limit, Ethereum and Ethereum Classic grouped in the same cluster apply block gas limit, which is the energy consumption limit for a block, to adaptively regulate its block size. Both Monero and Blackcoin have a dynamic mechanism to control the block size; however, it does not represent in the clustering result.

2.5 Conclusion

In this paper, we investigate the relationship between crypto behaviors and their underlying mechanisms. We specify the crypto behavior with their price and operational features defined by actual block time and block size. We calculate the distributional characteristics to define the behavior of time series. Using a characteristics based spectral clustering technique, we cluster the selected coins into clusters and scrutinize the blockchain mechanism in each group. We find that the underlying mechanism of cryptos is reflected in the clustering results. We observe that cryptos forked from the same origin and same consensus mechanism tend to become part of the same clustering group. Furthermore, the clusters obtained by the time series of block time have the same hashing algorithms and difficulty adjustment algorithms. Furthermore, a similar nature (static or dynamic) of block size was observed within clusters obtained by the time series of actual block size. We conclude with empirical evidence that the crypto behavior is indeed linked with their blockchain protocol architectures. As a result, cryptocurrency users and investors can have a better understanding and explanation of price and operational features through cryptocurrency mechanism. In the future research, we would elaborate the relation of price and operational features to underlying mechanism with an economic model and conduct relevant simulations. We would also like to investigate the impact of versions revisions on the dynamics of cryptos.

Bibliography

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering–a decade review. *Information Systems*, *53*, 16–38.

Blau, B., Griffith, T., & Whitby, R. (2020). Comovement in the Cryptocurrency Market. *Economics Bulletin*, *40*(1), 448–455.

Brunton, S. L., & Kutz, J. N. (2019). *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press. https://doi.org/10.1017/9781108380690

Buterin, V. (2014). A next-generation smart contract and decentralized application platform. *white paper*.

Ciaian, P., Rajcaniova, M., & d'Artis Kancs. (2016). The economics of bitcoin price formation. *Applied Economics*, *48*(19), 1799–1815. https://doi.org/10.1080/00036846.2015.1109038

de Souza, M. (2019). Var and persistence of risk shocks in cryptocurrencies market. *The Empirical Economics Letters*, *18*, 1–12.

Garriga, M., Dalla Palma, S., Arias, M., Derenzis, A., Pareschi, R., & Tamburri, D. (2020). Blockchain and cryptocurrencies: A classification and comparison of architecture drivers. *Concurrency and Computation Practice and Experience*. https://doi.org/10.1002/cpe.5992

Guo, L., Tao, Y., & Härdle, W. (2018). Understanding latent group structure of cryptocurrencies market: A dynamic network perspective. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3658206

Guo, X., & Donev, P. (2020). Bibliometrics and network analysis of cryptocurrency research. *Journal of Systems Science and Complexity*, 1–26.

Härdle, W. K., Harvey, C. R., & Reule, R. C. G. (2020). Understanding Cryptocurrencies*. *Journal of Financial Econometrics*, *18*(2), 181–208. https://doi.org/10.1093/jjfinec/nbz033

Hou, A. J., Wang, W., Chen, C. Y. H., & Härdle, W. K. (2020). Pricing Cryptocurrency Options*. *Journal of Financial Econometrics*, *18*(2), 250–279. https://doi.org/10.1093/jjfinec/nbaa006

Iwamura, M., Kitamura, Y., Matsumoto, T., & Saito, K. (2019). Can we stabilize the price of a cryptocurrency?: Understanding the design of bitcoin and its potential to compete with central bank money. *Hitotsubashi Journal of Economics*, *60*(1), 41–60.

Katsiampa, P., Corbet, S., & Lucey, B. (2019). High frequency volatility co-movements in cryptocurrency markets. *Journal of International Financial Markets, Institutions and Money*, *62*, 35–52. https://doi.org/https://doi.org/10.1016/j.intfin.2019.05.003

King, S., & Nadal, S. (2012). Ppcoin: Peer-to-peer crypto-currency with proof-of-stake. *self-published paper, August*, *19*, 1.

Parlitz, U. (2016). Estimating lyapunov exponents from time series. In *Chaos detection and predictability* (pp. 1–34). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-48410-4_1

Pele, D. T., Wesselhöfft, N., Härdle, W. K., Kolossiatis, M., & Yatracos, Y. (2020). A statistical classification of cryptocurrencies. *Available at SSRN 3548462*.

Sovbetov, Y. (2018). Factors influencing cryptocurrency prices: Evidence from bitcoin, ethereum, dash, litcoin, and monero. *Journal of Economics and Financial Analysis*, *2*(2), 1–27.

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

von Luxburg, U. (2006). A tutorial on spectral clustering. (149).

Wang, X., Smith-Miles, K., & Hyndman, R. (2005). Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery*, *13*, 335–364.

Zimmerman, P. (2020). Blockchain structure and cryptocurrency prices. *Bank of England Working Paper*.

Appendix

2.A   Time series characteristics

Table 2.A.1: **Price**.

Characteristic	Bitcoin	Ethereum	Litecoin	Bitcoin Cash	Ethereum Classic	XRP
mean	2659.127	178.966	34.394	537.723	9.381	0.192
standard_deviation	3798.466	222.452	48.645	509.244	7.827	0.302
skewness	1.338	1.950	2.389	2.322	1.491	4.193
kurtosis	0.672	4.654	7.272	6.157	2.239	29.471
maximum	19401.000	1356.000	352.799	3526.000	43.765	3.649
minimum	0.050	0.401	0.032	58.626	0.687	0.003
lowerquant	20.193	7.975	3.153	233.404	4.364	0.007
median	455.892	136.557	8.618	324.646	6.571	0.024
upperquant	5128.000	250.965	53.128	620.947	13.813	0.291
VaR99	0.062	0.578	0.040	107.426	0.809	0.004
VaR95	0.393	0.696	0.072	129.491	1.105	0.005
slope	2.781	0.163	0.032	-0.876	-0.002	0.000
intercept	0.050	2.820	0.033	63.765	0.892	0.006
autocorrelation	0.998	0.998	0.997	0.992	0.994	0.991
self_similarity	1.574	1.611	1.596	1.609	1.564	1.551
chaos	0.088	0.093	0.091	0.086	0.087	0.085

Characteristic	Bitcoin SV	Dash	Zcash	Monero	Dogecoin	Bitcoin Gold
mean	145.401	113.910	135.596	57.588	0.006	43.167
standard_deviation	66.784	187.915	125.654	75.569	0.193	70.420
skewness	0.678	3.126	1.756	2.145	49.692	2.879
kurtosis	0.079	11.777	3.208	5.300	2469.511	8.351
maximum	370.647	1436.000	728.159	439.391	9.608	513.293
minimum	52.683	0.516	23.940	0.233	0.000	5.093
lowerquant	87.323	3.950	50.251	1.100	0.000	9.710
median	135.217	66.508	72.251	44.090	0.001	15.869
upperquant	191.739	133.239	199.807	84.834	0.003	29.706
VaR99	53.377	0.711	27.767	0.272	0.000	5.357
VaR95	62.111	1.833	31.842	0.417	0.000	6.604
slope	0.218	0.083	-0.134	0.053	0.000	-0.147
intercept	111.700	1.380	286.297	1.911	0.000	513.293
autocorrelation	0.990	0.997	0.995	0.997	0.002	0.961
self_similarity	1.628	1.642	1.573	1.577	1.024	1.431
chaos	0.077	0.090	0.092	0.091	0.086	0.073

Characteristic	Peer coin	Vertcoin	Redd-coin	Feather-coin	Black-coin	Nova-coin
mean	1.004	0.670	0.001	0.062	0.095	2.185
standard_deviation	1.238	1.319	0.003	0.102	0.127	2.989
skewness	2.511	3.637	4.175	3.379	3.397	3.102
kurtosis	7.017	14.792	24.526	17.172	15.251	12.916
maximum	9.118	9.386	0.029	1.203	1.108	24.777
minimum	0.110	0.006	0.000	0.002	0.014	0.078
lowerquant	0.291	0.043	0.000	0.008	0.030	0.507
median	0.445	0.237	0.001	0.019	0.045	0.901
upperquant	1.275	0.626	0.001	0.072	0.088	3.301
VaR99	0.125	0.009	0.000	0.003	0.015	0.156
VaR95	0.168	0.015	0.000	0.004	0.020	0.187
slope	0.000	0.000	0.000	0.000	0.000	-0.001
intercept	0.382	6.315	0.000	0.559	0.035	0.078
autocorrelation	0.993	0.992	0.988	0.983	0.993	0.994
self_similarity	1.577	1.603	1.548	1.523	1.537	1.596
chaos	0.088	0.085	0.079	0.078	0.084	0.091

Table 2.A.2: **Block time**.

Characteristic	Bitcoin	Ethereum	Litecoin	Bitcoin Cash	Ethereum Classic	XRP
mean	10.453	0.257	2.507	11.167	0.246	NA
standard_deviation	8.814	0.045	0.385	11.009	0.032	NA
skewness	21.779	3.098	5.003	11.597	5.144	NA
kurtosis	701.717	11.987	54.589	160.209	61.066	NA
maximum	360.000	0.509	8.521	205.714	0.800	NA
minimum	2.081	0.208	0.149	1.275	0.153	NA
lowerquant	8.623	0.235	2.357	9.664	0.235	NA
median	9.474	0.241	2.474	9.931	0.238	NA
upperquant	10.435	0.268	2.599	10.360	0.242	NA
VaR99	5.923	0.220	1.710	2.331	0.215	NA
VaR95	7.129	0.222	2.111	8.479	0.218	NA
slope	-0.001	0.000	0.000	-0.007	0.000	NA
intercept	102.857	0.208	0.149	160.000	0.208	NA
autocorrelation	0.494	0.981	0.705	0.395	0.818	NA
self_similarity	1.027	1.522	0.787	0.704	1.249	NA
chaos	0.012	0.070	0.012	0.003	0.068	NA

Characteristic	Bitcoin SV	Dash	Zcash	Monero	Dogecoin	Bitcoin Gold
mean	10.195	2.659	2.409	1.686	1.048	9.823
standard_deviation	1.639	0.805	0.345	0.541	0.043	0.741
skewness	12.504	19.831	-3.025	3.258	-9.220	-5.375
kurtosis	221.950	409.827	7.261	57.807	222.460	60.686
maximum	40.000	22.500	2.618	10.992	1.288	11.250
minimum	7.310	0.348	1.240	0.829	0.100	0.254
lowerquant	9.600	2.609	2.487	1.025	1.038	9.664
median	10.000	2.623	2.509	1.951	1.044	9.931
upperquant	10.511	2.637	2.531	2.020	1.050	10.141
VaR99	8.361	2.476	1.248	0.947	0.980	7.767
VaR95	9.034	2.571	1.258	0.984	1.031	8.623
slope	-0.001	0.000	0.000	0.001	0.000	0.001
intercept	40.000	0.348	2.286	1.627	0.100	0.254
autocorrelation	-0.115	0.707	0.982	0.805	0.787	0.378
self_similarity	0.367	0.811	1.121	0.922	1.044	0.494
chaos	0.023	0.003	0.010	0.001	0.011	-0.001

Characteristic	Peer-coin	Vert-coin	Redd-coin	Feather-coin	Black-coin	Nova-coin
mean	10.085	2.502	4.646	2.005	1.090	6.819
standard_deviation	47.070	0.180	68.175	6.443	0.105	2.295
skewness	30.324	-1.782	20.761	11.521	-4.368	24.326
kurtosis	919.356	30.015	434.280	157.793	18.525	891.281
maximum	1440.000	4.079	1440.000	130.909	1.335	96.000
minimum	1.377	0.151	0.646	0.148	0.442	0.451
lowerquant	7.742	2.412	0.986	1.042	1.111	6.154
median	8.372	2.500	1.007	1.048	1.114	6.606
upperquant	9.057	2.590	1.028	1.171	1.117	7.164
VaR99	5.464	2.144	0.935	1.034	0.551	4.364
VaR95	6.545	2.289	0.957	1.036	0.949	5.390
slope	-0.003	0.000	-0.010	-0.002	0.000	-0.001
intercept	1440.000	0.151	1440.000	0.291	1.309	1.765
autocorrelation	0.667	0.154	0.821	0.914	0.976	0.373
self_similarity	0.717	0.437	1.051	1.210	1.337	0.697
chaos	0.002	0.008	-0.001	0.032	0.006	0.009

Table 2.A.3: **Block size**.

Characteristic	Bitcoin	Ethereum	Litecoin	Bitcoin Cash	Ethereum Classic	XRP
mean	407162.152	14376.916	12909.684	138173.724	1297.638	NA
standard_deviation	363245.372	11337.562	15590.195	284058.956	340.581	NA
skewness	0.241	0.285	4.309	9.176	0.679	NA
kurtosis	-1.583	-0.819	31.780	109.791	2.106	NA
maximum	998092.000	58953.000	206020.000	4710539.000	3594.000	NA
minimum	134.000	575.164	134.000	4982.000	575.164	NA
lowerquant	21246.000	1627.750	4004.750	60455.500	1054.750	NA
median	310990.000	17024.000	7016.000	94775.000	1310.500	NA
upperquant	777369.500	23068.750	19366.500	122827.500	1492.250	NA
VaR99	134.548	658.423	561.630	15574.520	653.404	NA
VaR95	134.952	788.678	800.306	27169.700	775.052	NA
slope	266.541	17.464	8.806	-89.253	0.189	NA
intercept	204.000	643.886	199.000	385996.000	643.886	NA
autocorrelation	0.985	0.981	0.872	0.626	0.850	NA
self_similarity	1.067	1.310	1.148	1.074	1.131	NA
chaos	0.058	0.058	0.065	0.027	0.045	NA

Characteristic	Bitcoin SV	Dash	Zcash	Monero	Dogecoin	Bitcoin Gold
mean	1100149.254	12999.389	23802.102	39874.397	10523.242	25312.953
standard_deviation	1278250.457	26340.294	38911.209	47310.430	6607.125	67527.275
skewness	6.673	27.654	8.711	1.703	5.917	6.269
kurtosis	84.455	1040.743	117.847	4.063	68.981	45.828
maximum	20460199.000	1059232.000	687685.000	347816.000	116605.000	739259.000
minimum	5005.000	226.545	379.573	375.434	143.000	133.000
lowerquant	257789.500	3038.000	7189.500	3047.250	6775.000	6512.500
median	996071.500	9240.000	11670.000	20980.000	9510.000	9316.000
upperquant	1573243.000	19193.000	28242.000	62002.000	12022.000	14118.000
VaR99	6435.000	1312.960	2605.530	1058.990	3432.400	2727.870
VaR95	14660.750	1736.200	3103.900	1320.350	4491.000	3983.600
slope	2318.003	14.357	-25.267	26.939	1.018	-67.625
intercept	10871172.000	226.545	379.573	375.434	143.000	133.000
autocorrelation	0.377	0.298	0.836	0.958	0.798	0.618
self_similarity	1.004	0.947	1.138	1.214	1.070	1.015
chaos	0.009	0.018	0.030	0.041	0.021	-0.012

Characteristic	Peer-coin	Vert-coin	Redd-coin	Feather-coin	Black-coin	Nova-coin
mean	NA	2641.881	772.025	806.556	687.622	539.712
standard_deviation	NA	3611.409	634.442	1621.154	3441.373	1223.175
skewness	NA	3.420	3.613	10.605	28.388	38.218
kurtosis	NA	16.189	21.857	158.924	894.526	1712.453
maximum	NA	36709.000	7808.000	36789.000	120169.000	57527.000
minimum	NA	105.000	105.000	109.625	252.514	110.835
lowerquant	NA	682.104	388.361	359.746	286.296	360.352
median	NA	1149.000	526.043	460.827	386.251	436.181
upperquant	NA	3185.000	937.696	598.841	627.727	542.228
VaR99	NA	248.950	317.797	126.333	255.520	262.588
VaR95	NA	310.697	337.320	247.907	261.297	284.524
slope	NA	-0.586	-0.475	-0.739	-0.025	-0.204
intercept	NA	130.000	175.000	109.625	464.500	141.000
autocorrelation	NA	0.894	0.609	0.705	0.360	0.069
self_similarity	NA	1.129	1.007	1.063	0.959	0.951
chaos	NA	0.100	0.034	0.034	0.039	0.011

Chapter 3

Blockchain: An Invisible Hand for Crypto?

An Empirical Discussion on Ethereum

All the source codes of this chapter are available in QHAND.

3.1   Introduction

Cryptonomics offers a new marketplace for trading, liquidity, and price discover, and its growing valuation triggers a megatrend in which modern transaction technologies emerge and begin to supplant others. Under cryptographic protection, crypto (cryptocurrency) tokens allow any two willing parties to transact directly with each other without the need for a trusted third party. Compared to the conventional transaction vehicles, such technology achieves a higher degree of anonymity and security, and provides better flexibility in application infrastructure, i.e. layered structure of blockchain architecture. (Yang et al., 2019) Much research has been keen to define the financial role of cryptos as – e.g. a medium of exchange or speculative asset, safe heaven or risky hazard (Shahzad et al., 2019; White et al., 2020); and to differentiate from or to relate to classical assets in the sense of price and market value (Chuen et al., 2017; Mark et al., 2020). As cryptos are driven by distributed ledger technologies (DLT), i.e. blockchain (BC) technology, directed acyclic graphs (DAG), the interconnected mechanism behind cryptos – functional characteristics – provides significant features to their infrastructure and application (Cahyadi et al., 2021; Lin et al., 2021). However, the hinge between the economic attributes (i.e. return volatility) and underlying mechanisms has often been neglected. The regulation of blockchain on crypto is still met with skepticism.

Often, crypto market has been considered to be full of irrationality and froth (Aloosh & Ouzan, 2020; Kyriazis et al., 2020; Li et al., 2021), having unconditional volatility and being subject to sudden and massive price swings (Katsiampa et al., 2019). Investor sentiment (Gurdgiev & O'Loughlin, 2020; López-Cabarcos et al., 2021), in which

researchers employ textual analysis on social media posts, has often been related to the crypto price fluctuations. It raises an issue of data quality and sentiment bias, which might not be able to capture the full impact of the sentiment and market trend (Baker & Wurgler, 2007; McGurk et al., 2020). Instead, we include both functional characteristics (i.e. blockchain trilemma) that are the core driver of a crypto and market dynamics (i.e. transaction behaviors) in order to observe the endogenous and exogenous dynamics of the crypto system; conduct a comprehensive causal analysis. Studying the linkage among the functional characteristics, market dynamics, and economic attributes enables crypto users to better understand crypto's functioning mechanism and return evolution, in particular for coin creators.

Since this market exposes to the uncertainties and unexpected changes from the coin supply and demand, investor and user sentiments, government regulations, and media hype (Mai et al., 2018), there exists a need to depict the volatility dynamics of cryptos attentively. W. Zhang et al., 2018 examines the stylized facts of cryptos and concludes that the returns of cryptos display strong volatility clustering and leverage effects. Comparing with the classical assets, Pele et al., 2021 shows that cryptos have long tails of the log-return distribution and exhibit a synchronous evolution among others. Fakhfekh and Jeribi, 2020; Naimy et al., 2021 perform diverse GARCH (generalized autoregressive conditional heteroscedasticity) models, in which they emphasize the essence of catching asymmetries and persistence in crypto's volatility. Due to the existence of microstructure noise across crypto markets (Bouri et al., 2022; Dimpfl & Peter, 2021), it is hard to observe how the mechanism amplifies fundamental shocks volatility impact on returns and leads to boom and busts unconnected to fundamentals.

Underscoring such a noise of crypto returns on volatility, Conrad et al., 2018; Katsiampa, 2017; Walther et al., 2019 use the AR-CGARCH (autoregressive component) and GARCH-MIDAS (mixed data sampling) respectively to differentiate between short-run and long-run components of the conditional variance. Long-run volatility component displays long memory behavior and commonly driven by macroeconomic conditions, i.e., business cycle effect; on the other hand, short-run volatility component reflects a temporary increase in volatility after a large shock and driven by an asymmetric effect on volatility response, i.e., news effect, abnormal trading activities (Engle & Lee, 1999). Considering of fat tails, skewness and leverage, Beta-t-EGARCH model offers a better fit than comparable skewed-t GARCH models and superior out-of-sample predictive performance (A. Harvey & Lange, 2018) By extending to two-component volatility specification, it captures the persistence in returns and offers more interpretability by removing the interference from the short-run volatility component. In this paper, we apply the two component Beta-t-EGARCH to model the volatility components of Ethereum (ETH) returns as the economic attributes and examine their relationship

with ETH's functional characteristics and market dynamics via Granger causality.

This chapter proceeds as follows. Section 2 overviews the studies on the linkage between blockchain and crypto and then often an empirical framework to depict the dynamic between blockchain and cryptonomics. Based on the framework proposed in Sections 3, 4 and 5 detail the functional characteristic and market dynamics factors considered in this paper. Section 6 presents how we derive the economic attributes from ETH's log return time series. Section 7 elaborates our causal discovery method and discusses causal relationships among factors. Last, we conclude the study in Section 8.

3.2 Blockchain and cryptonomics

There exists a broad body of literature respectively on – crypto economics and finance, as well as blockchain architecture. As most of the cryptos are powered by blockchain technology, the construction of blockchains is of essence to the functionality of them. The inherent specifications of their mechanisms and algorithms determine their performance (i.e. transaction per second, block size, block time) and security (Bamakan et al., 2020; Ferdous et al., 2020). Yet, the investigation of blockchain technology as the underlying mechanism of crypto and its price behavior still happens to be a beginning field of research (Jiang et al., 2021; Pagnotta, 2022). Zimmerman, 2020 highlights the restricted settlement capacity in a blockchain can lead to a competition between users of the crypto and then produce a high volatility on price. Garratt and van Oordt, 2020 suggests a strong positive relationship between exchange rates and mining power for several proof-of-work (PoW) cryptocurrencies. On the blockchain network, traders' and transaction validators' (i.e. miners) decisions influence each other, and the evolution of Bitcoin (BTC) price and security are jointly determined (Pagnotta, 2022).

Thus far, most of the research works have been dedicated to constructing equilibrium-based models to depict how the blockchain interacts with the crypto prices, but have rarely offered empirical evidence to support their claims. Some empirical study investigations can be found in (Guo et al., 2022; Lin et al., 2021) in which they foreground the impact of fundamental blockchain characteristics (e.g., origin and fork, algorithm and proof type) on the performance of a crypto using statistical learning methods. In fact, crypto's underlying mechanism is built and governed by computer codes that are so-called "code is law" (Lessig, 2009). Lucchini et al., 2020 offers a fresh perspective that the network of developers and code development are in relation to the returns of Github-linked cryptos, but it cannot identify whether the mechanism drives the synchronization across these cryptos.

Abadi and Brunnermeier, 2018 view the design of blockchain as a digital record-keeping

problem which aims to establish consensus on an update to the ledger. They conclude the three properties of consensus algorithm – Fault-tolerance, resource efficiency, and full transferability – to understand the synergies and trade-offs, in light of *users' utility*. In contrast, Buterin, 2021's scalability trilemma – decentralization, scalability and security – underscores the *functional properties*. Often, it has been considered as a standpoint for blockchain architecture and its implementation.

From an *empirical viewpoint*, we propose a framework to depict the subtleties and intricacies of cryptos in this paper. There are three main components – functional characteristics, market dynamics, and economic attributes – under the two-level framework, as of blockchain ecosystem and cryptonomics. The first level, blockchain ecosystem, is the fundamental architecture of a blockchain which enables the functional trust among users and settles all the transactions. We consider it to be a micro-element that drives a crypto's functionality and economic behavior. The addition of multi-layered infrastructure (such as Oracle, the lightning network, and Plasma) offers greater flexibility and adaptability to blockchain technology. The second level, cryptonomics, is driven by market dynamics that boil down to the supply and demand of crypto. This is initiated by a coin's market and its applicability (extension market), i.e. the NFT (non-fungible token) market for Ethereum. Economic attributes are the result of a collaborative effort between functional characteristics and market dynamics. Due to the changes in supply and demand within the entire blockchain network, the three components are actively interacting with each other.

Market dynamics (i.e. market sentiment, transaction behaviors) are often considered to be the main driver of economic attributes. The influence of functional characteristics, which represent the underlying mechanism, is often overlooked. From Buterin, 2021's scalability trilemma, we propose a blockchain pyramid that highlights how the blockchain design can achieve a trustless trust of the entire cryptonomics. Developers at the top of the pyramid are the most indispensable role of functionality development, as the network of them generally guides the path of a crypto project, i.e. forks for algorithm updates or changes. In the bottom, a parallelogram combines the concept of the scalability trilemma with the component incentive, which these four aspects of blockchain design are calibrating with each other over time. The incentive component is designed to motivate both traders and transaction validators.

Following, we describe the three pillars of the proposed framework – functional characteristics, market dynamics, and economic attributes – and look at their relevant time series factors. The daily observations for each factor are plotted as a violin plot, which is a hybrid of a box plot and a kernel density plot, in the following sections. Violin plots display the range of the data with a box plot and the corresponding kernel density.
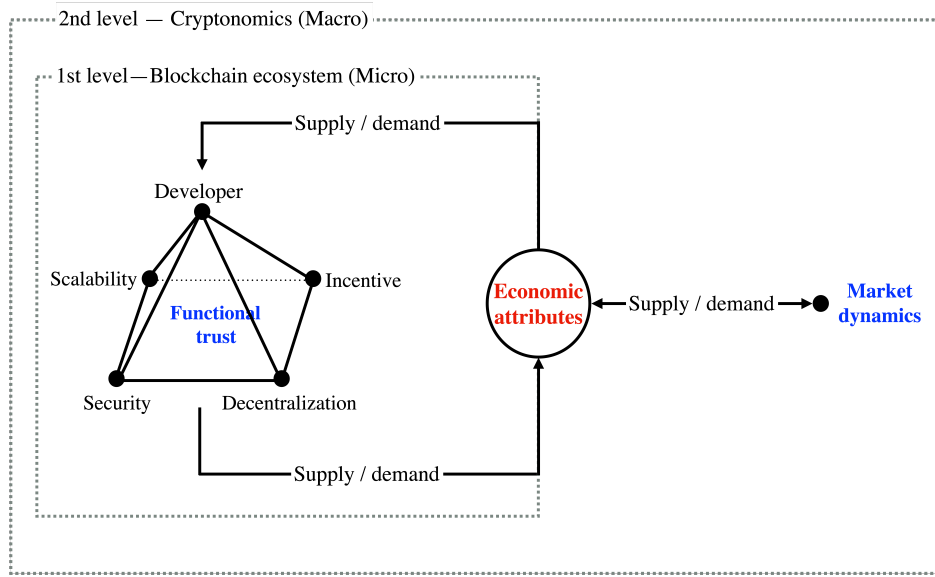
Figure 3.1: **Blockchain ecosystem and cryptonomics**.

Inside the box plot, there are two lines: the solid line represents the median and the dashed line represents the mean.

3.3 Functional characteristics

In this section, we discuss some functional characteristics of blockchain and discuss their dynamics in terms of algorithmic infrastructure. The characteristics are categorized by the proposed blockchain pyramid: developers, scalability, decentralization, security, and incentive.

3.3.1 Scalability

A blockchain network's capacity to handle a certain amount of transaction data is referred to as scalability. It is often influenced by endogenous factors, such as block time intervals and block size. Between 2014 and 2017, an interesting debate on the Bitcoin scalability problem emerged, known as Bitcoin's civil war (Bier, 2021). Large blockers advocated for the expansion of block size to avoid blockchain congestion and enhance Bitcoin's applicability, but this requires a hard-fork update, which could lead to Bitcoin losing its popularity. The other argument from the small blockers was that such an update might cause Bitcoin to split into two different coins due to incompatible nodes on the network, and thus cause a devaluation. The scalability of a blockchain can therefore be critical to crypto's technical performance and possibly return volatility.

Definition 1 (**Block utilization**). *The quotient of the actual usage of a block by its*
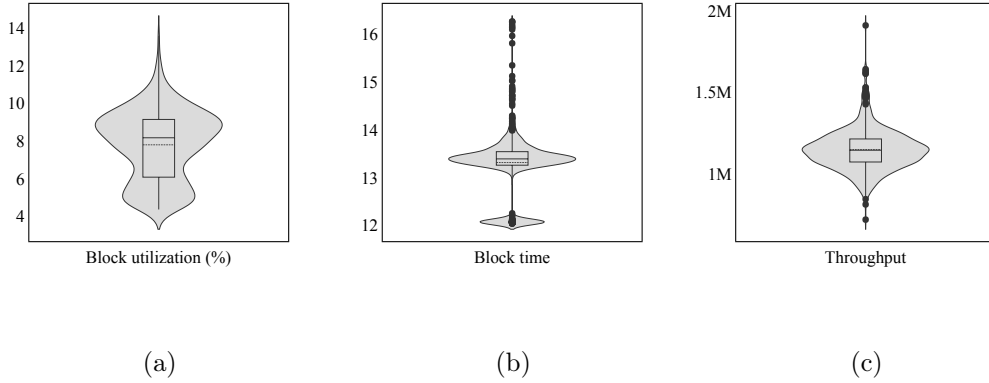
Figure 3.2: **Scalability factors**.

size target given by the blockchain system.

Block utilization refers to the congestion of a blockchain. A high utilization rate can result in a low scalability for transaction processing, as the network requires more time for information propagation. The denominator of it is the designed size of a block, for example, ETH's target block size is 15 million gas (around 0.937500 Mbyte with calldata gas cost 16 gases per block). It is worth noting that the block size of ETH is determined by the gas limit, which is the maximum amount of cost for a validator to undertake on a specific block, compared to using a predefined size limit. Therefore, the size of blocks can vary depending on network demand, up to a block limit of 30 million gas, which is twice the target block size. The nominator of block utilization here is the actual-used size of a block. The used size increases when demand is high, and vice versa. The time for generating a block, or the confirmation time for transactions, can be influenced by block utilization. Figure 3.A.1a illustrates the evolution of ETH's average block utilization daily. Since the first observation date on January 01, 2021, the block utilization has shown an upward trend. It starts decreasing as the deployment of Ethereum Improvement Proposal (EIP) -3675 – the Merge – begins. There is autocorrelation within a 95% confidence interval, with up to 6 lagged values, and the presence of a periodic signal, see Figure 3.A.1b. In Figure 3.2a, ETH does not show any congestion, that is, the utilization remains lower than 15%. It appears that the block utilization is not suffering from the impact of outliers.

Definition 2 (Block time). *The average time that it takes the miners or validators within a network to verify transactions within one block and append the block to the blockchain.*

Block time can also be considered a service level that provides a customer with a service within a given time period. Low block time provides high scalability. Multiple

factors, such as the consensus mechanisms used, block size limits, network sizes, and network demand (congestion), affect the duration of block generation. Especially high network demand leads to a long block time. **A blockchain sets its transaction fee in response to the likelihood of congestion in the network.** Due to the different underlying designs, each crypto has its block time. For example, BTC takes an average of 10 minutes to generate a block, whereas ETH only takes around 13 seconds. Figure 3.A.2a illustrates the evolution of ETH's average block time per day. Before the Merge, there is a periodic trend that goes upward then downward, despite the different deviation ranges. This can be attributed to ETH's difficulty adjustment algorithm (DAA) of PoW, which strives to maintain an average inter-block time between 12 and 14 seconds, refer to Section 3.3.3. While the block time is increasing, the DAA is reducing the difficulty requirement in order to grant validators a higher chance to find the hash value. The block time, after the Merge, is around 12 seconds on average. Since the block time is controlled by the DAA linearly based on the previous block, it is autocorrelated to the first two lags in Figure 3.A.2b. However, the daily data used cannot catch the dynamics of block generation, the autocorrelation is rather unobservable. Figure 3.2b shows that there are many outliers in block time and the two apparent peaks in the figure correspond to the periods before and after the Merge. The large outliers are linked to the difficulty bombs that DAA exponentially increases the difficulty level every 100,000 blocks.

Definition 3 (**Throughput**). *The number of transactions that are processed on a blockchain per second.*

Throughput is relevant to the blockchain mechanism and also the network demand. Higher throughput means the blockchain can process more transactions. The blockchain is capable of processing more transactions with high throughput. In particular, while network congestion appears, the blockchain mechanism should be able to allocate transactions and guarantee a stable throughput. However, in order to earn more fee in PoW, validators are often eager to include large value transactions and avoid small and scattered transactions. It influences the speed at which a blockchain can process transactions. On average, the daily transaction counts of the ETH are 1 million transactions per day and present some outliers in Figure 3.2c. Figure 3.A.3a shows that the transaction counts drop after the launch of the Merge. The spike on December 9, 2022, is a direct result of large-scale fiat withdrawals from exchanges that may be caused by FTXs bankruptcy. It displays autocorrelation, with up to 3 lagged values, see Figure 3.A.3b.

3.3.2 Decentralization

Decentralization is the delegation of control and decision-making from a central authority (individual, organization, or group) to a distributed network. It is the core of blockchain and cryptocurrency technologies that offers the freedom of being trust-less to any particular blockchain platforms or authorities, while establishing the trust-worthy of the whole system. The realization of decentralization in a network can be strenuous due to a skewed mining power and inherent scaling limits (Chu & Wang, 2018). Recent research on quantifying de-centrality can be found in (Barbereau et al., 2022; Gochhayat et al., 2020). Here, we mainly focus on the network users' dynamics.

(a) (b)
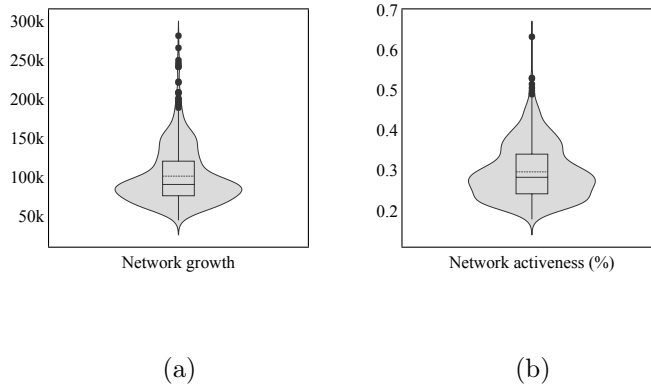
Figure 3.3: **Decentralization factors**.

Definition 4 (**Network growth**). *The amount of new addresses that join the blockchain within a day.*

Network growth shows the popularity of a blockchain and its adoption. In other words, the network is more decentralized by including unique addresses. Blockchain technology initially gained attention through its initial and most basic application, cryptocurrency. The increasing prevalence of NFTs and utility tokens has spread across different industrial sectors, such as creative industries, supply chain, and gaming, due to its high applicability and capability. It accelerates the adoption of blockchain technology, but high transaction fees can be a barrier. Consequently, many Layer-2 solutions have been created with lower fees, e.g. Polygon on Ethereum network. Figure 3.A.4a shows that ETH's network size has a significant growth and peaked in the May of the same year, which is one of the largest crypto crashes due to Elon Musk's unsupportive statement and new round of regulations by the Chinese government. The number of new addresses is stimulated by the launch of the Merge. Figure 3.A.4b shows that ETH's network growth presents only a very limited autocorrelation. There

are 100,000 unique addresses per day joining the ETH network during the observation period, and there are more upper outliers, see Figure 3.6a.

Definition 5 (Network activeness). *The quotient of the number of active addresses a day by the total number of unique addresses over the blockchain network.*

Despite network growth considered an indicator for decentralization, the number of daily network participants shows how active a blockchain network is. Considering the total network size, network activeness here discloses the entire network's involvement and daily adoption. A high network activeness also indicates that the market may experience an extreme event, such as large price drops or rises. The entire network frequently trades for buys and sales. This implies that network activeness can also be associated with price volatility and market sentiment. Figure 3.3b illustrates that there are around 0.3 unique addresses active daily on ETH's network. The network activeness has no autocorrelation, see Figure 3.A.5b. After the Merge, there is a spike around the end of 2022, as illustrated in Figure 3.A.5a.

3.3.3 Security

Blockchain technology utilizes cryptography in cloud computing to provide a high level of security and privacy. Scaling up of measures such as network size, protocols, applications, network elements, topological constraints, and functionality expectations can become a great challenge for the reinforcement and sustainability (Kiayias & Panagiotakos, 2015; R. Zhang et al., 2019). Provable security, scalability, and energy cost are often trade-offs, especially for PoW, which involves solving a computationally difficult problem. An unsteady block time can result. Instead, PoS has a fixed tempo for block generation and effectively avoids the massive computation efforts.[1] The computation efforts as proof can be restrained by a democratic roulette process in PoS. The following measures can be used to compare such a change from PoW to PoS.

Definition 6 (Difficulty adjustment). *The degree of difficulty involved in discovering new blocks through mining, i.e. number of hashes per second.*

The DAA is a crucial component for security as well as scalability in PoW. According to its design, the algorithm adjusts the mining difficulty in order to regulate the inter-block time around the target time. A high mining difficulty can result in a delay in validators

---

[1]For example, Casper, the ETH PoS protocol, divides time into 32 epochs, which consist of 32 slots that each last 12 seconds. Each slot is assigned to a block proposer to determine which transactions are included in the block. A minimum of 128 validators, as a committee, vote on the block. Each validator's deposit is used to weigh the votes, which are referred to as attestations. Upon receiving a two-third majority of validator votes, a block will be added to the blockchain. Validators are allowed to participate in one committee per epoch, but there may be multiple committees with the same size per slot.
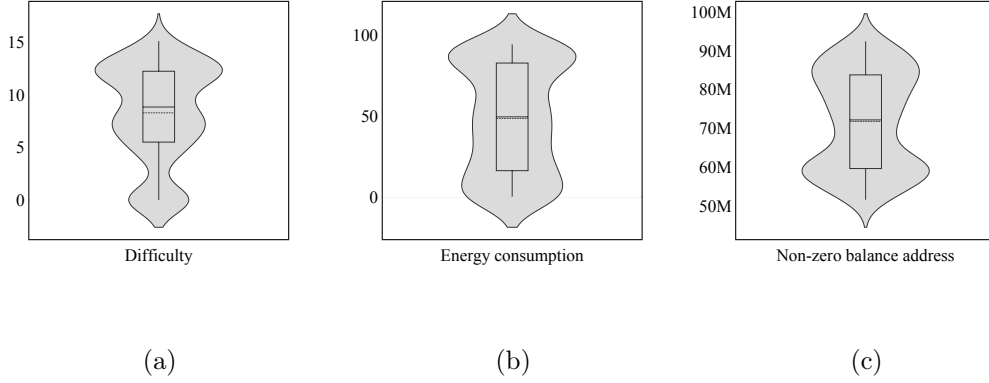
Figure 3.4: **Security factors**.

appending a block, whereas a low difficulty can lead to a danger that an adversary or fraudster could take to subvert a blockchain protocol. Below, we show an example of DAA in ETH's PoW, i.e. Metropolis fork determines the difficulty D_{i+1} of block $(i+1)$ by

$$D_{i+1} = \max[D_i + \frac{D_i}{C} \times \Delta, \min(D_i, D_0)], \quad i \in \mathbb{Z}_0^+$$

where the adjustment factor

$$\Delta = \begin{cases} \max(2 - \frac{\text{TS}_{i+1} - \text{TS}_i}{9}, -99), & \text{if any uncle block exists} \\ \max(1 - \frac{\text{TS}_{i+1} - \text{TS}_i}{9}, -99), & \text{otherwise.} \end{cases}$$

The bound divisor of the difficulty C is an arbitrary constant, e.g. $C = 1024$, which is equivalent to the difficulty of genesis block 0x0400 in hexadecimal. D_0 is the difficulty of the genesis block, or the minimum difficulty of any other block, and TS_i is the timestamp for block i. Uncle blocks denote blocks that are correctly mined, but are not appended to the blockchain.

Hashrate is also a relevant measure, here. It is the number of hashes or guesses per second that can be made on the network. A high hashrate strengthens the security of the blockchain, since more computational power is on the network, and it is more difficult to conduct malicious activities. In this paper, we do not include this measure as difficulty adjustment and hashrate significantly co-move with each other, especially for ETH's PoW that recalculates the difficulty each block rather than adjusting every 2016 blocks, as of BTC.

Figure 3.A.6a shows that ETH's difficulty continuously increases. Here, PH/s denotes one Petahash per second, equivalent to 10^{15} hashes per second. In order to migrate to PoS, ETH suddenly increases mining difficulty, known as a difficulty bomb. This

discourages miners from opting to stay with PoW. The continuous increase in difficulty in our observation period can be attributed to the upgrades of EIP-1559 London (2021), EIP-4345 Arrow Glacier (2021), and EIP-5133 Gray Glacier (2022). When beginning the Merge, the difficulty eventually reaches zero. The autocorrelation of difficulty adjustment only appears in the first and second lags, with a negative correlation in the second lag, see Figure 3.A.6b. Figure 3.4a appears that there are three peaks in the kernel, which may represent the periods between the issues of a difficulty bomb.

Definition 7 (**Energy consumption**). *The amount of electricity used in a blockchain network.*

The PoW protocol has an energy-intensive design. That is, the electrical energy consumption directly relates to the computational efforts on the network and is therefore an essential feature of PoW. Safety and security of the network depend on energy expenditure, which allows the network to maintain an honest record of transactions and a predetermined fixed credible monetary policy. It implies resistance to forgery, inflation, and theft. We therefore consider energy consumption to be a factor for security. It can also be seen as an incentive mechanism, since less energy consumption on the blockchain leads to lower transaction costs, which is more appealing to coin users.

Figure 3.A.7a is the estimated daily ETH energy consumption index of Digiconomist, 2021 –which estimates the total energy consumption of the ETH network. TWh/year represents one Terawatt hour a year, equivalent to one trillion watt-hours a year. Going into 2021, the ETH energy consumption had been on a dramatic increase. The Gray Glacier hard fork upgrade has delayed the planned difficulty bomb by another 700,000 blocks, which leads to a drift in both difficulty adjustment and energy consumption by the end of June 2022. The energy consumption drops at least 99.84% after changing from PoW to PoS, which is similar to the difficulty (De Vries, 2022). The autocorrelation structure in Figure 3.A.7b is similar to the one of difficulty adjustment in Figure 3.A.6b. The transition from PoW to PoS is also observed in the upper and lower parts of the violin plot, see Figure 3.4b.

Definition 8 (**Non-zero balance address**). *The total number of addresses with a non-zero balance on the blockchain.*

Non-zero balance addresses are those addresses where people have staked their coins on the network. The opposite are the ghost addresses, which have no transactions since and could be linked to a risk of undermining a blockchain protocol. Many blockchain protocols require the honesty of the majority of addresses on the network to reach consensus, so fraudsters may intend to manipulate these addresses. The increase in non-zero balance addresses is associated with market confidence and coin's widespread

adoption, which increases the security of a blockchain by removing the possibility of one that overpowers the majority.

The ETH's network maintains an average of 70 million addresses with a non-zero balance, see Figure 3.4c. Figure 3.A.8a demonstrates that the total number of non-zero balance addresses steadily increases, and it grows by nearly 20 million addresses over the last 12 months before going into 2023. The increasing adoption of NFTs is also one of the drivers that bring more users to the ETH's network. There is no autocorrelation, apart from the first lagged value in Figure 3.A.8b. It is an apparent time trend, implying non-stationarity or unit root.

3.3.4   Incentive mechanism

The provision for offering incentives is necessary to motivate validators to append globally agreed blocks by consensus for processing transactions, especially in a blockchain network under a PoW consensus mechanism that is highly costly for validators. Without sufficient participation of validators, a network might face security issues (i.e. adversaries from monopolizing the blockchain) and scalability issues (i.e. delays in transaction settlement). For example, Bitcoin offers incentives to its validators through mining rewards and transaction fees. Rewards and transaction fees are often used to compensate each other. Since the halving policy applies to BTC, rewards for validating transactions are reduced in half, which results in an increase in transaction fees paid by users by time in order to compensate for the loss in rewards of validators. Such an incentive mechanism reduces the possibility that people will use crypto as a medium of exchange for daily transactions. In this way, transaction fees can be viewed as an incentive for both users and validators.

| (a) | (b) | (c) |
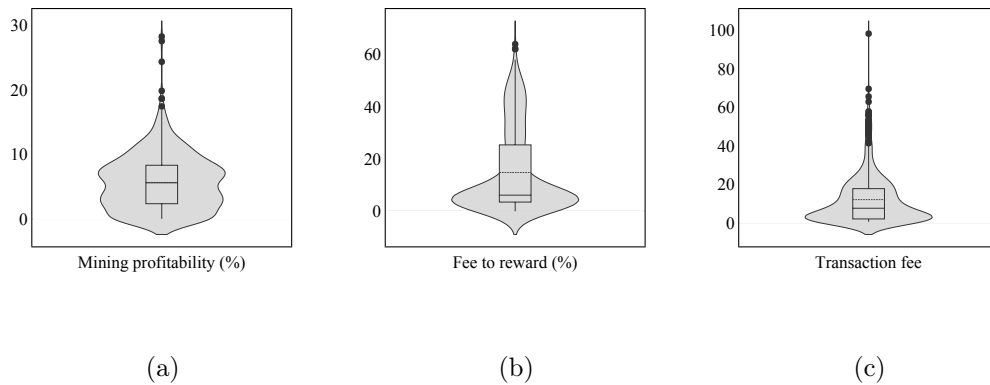
Figure 3.5: **Incentive factors**.

Definition 9 (**Mining profitability**). *The ratio derived validators' gain to the hashrate and price of the mining machines, and the costs incurred during validation.*

Mining profitability is highly dependent on the hardware used by validators and the current electricity costs, essentially for PoW. Mining hardware (i.e. the performance of graphics processing units, GPUs) directly associates with the hashrate, which is the number of hash can be computed for a second. For example, the GPU Nvidia GTX 1080 is rated the best on minerstat, which costs around US 600$ for one with 8-GB RAM. Electricity costs differ from the region the validators locate. Furthermore, mining hardware produces heat, which requires a cooling system to avoid machine breakdown.

In our observations, the average profit for the validators on ETH's network is around 10% of their costs, see Figure 3.5a. Due to ETH's transition from PoW to PoS, many difficulty bombs have been launched from 2021 to force users to accommodate to the upcoming protocol. The profitability of validators in Figure 3.A.9a significantly decreases as it requires more computation efforts. At the Merge, it eventually reaches zero. Figure 3.A.9b has a limited autocorrelation effect and a periodic pattern.

Definition 10 (**Fee to reward**). *The quotient of the transaction fee by the mining reward.*

The fee-to-reward ratio is used to assess the profitability of validation. Rewards and transaction fee are used to counterbalance each other, in particular for PoW. In order to stabilize the coin supply and demand, most PoW blockchains have a mechanism to regulate the number of coins available for rewards. Therefore, to speed up the transaction settlement, a user must offer an attractive transaction fee to validators, while the rewards are decreasing. A higher transaction fee encourages validators to include the transaction in their proposed lists of transactions, i.e. blocks. That is, users on the network are forced to compete among themselves. The fee bidding becomes a major obstacle for increasing the scalability and applicability of the coin (Malik et al., 2022).

In PoS, each validator is encouraged to stake their holding coins on the network and thus earn passive rewards. The staking rewards are based on how much a coin is validated and what rewards the network is offering over a given time period. A user's stake in coins determines the weight of her vote for block validation, and therefore more coins staked from different users will secure the blockchain by increasing the difficulty for one to monopolize the network. This design will reduce the energy-intensive computation and fee bidding in PoW. In the case of ETH, we can find that the proportion of transaction fees is generally much lower than the rewards, see Figure 3.5b. Figure 3.A.10a shows that the reward-to-fee ratio has declined significantly since October

2021. Following the EIP-1559 upgrade – named London, ETH introduced the burning of the base fee, which means a certain proportion of each transaction fee is removed from circulation and hence it is not given to validators. Base fees, also known as the market-clearing price, are adjusted by the gas limit and market demand. We see that the fee-to-reward ratio show autocorrelation, with up to three lags in Figure 3.A.10b.

Definition 11 (**Transaction fee**). *The average transaction fee paid to the network validators.*

As discussed previously, PoW greatly depends on the fee bidding process to encourage block validation. On the other hand, users broadcast their transactions while the average transaction fees are low, which indicates fewer demands on the network. Such a mechanism does not align the users and validators in terms of their utilities. Figure 3.5c shows many outliers in ETH's average transaction fee, which are mainly from the observations before the Merge. Since 2022 July, the average transaction fee has significantly decreased and eventually remained below 1 USD, see Figure 3.A.11a. Many industry development products, such as dApps, decentralized exchanges, smart contracts for enterprise applications, altcoins, shitcoins (i.e. meme tokens), stablecoins, and NFTs, that are backed by the Ethereum network, are benefited from the protocol upgrade. ETH is widely positioned by the market as a Swiss army knife that aims to achieve high functionality and applicability, distinguished from BTC. Furthermore, the average transaction fee presents an autocorrelation, which means that the current fee level is associated with the previous level.

3.3.5   Developer

In the blockchain pyramid of Figure 3.1, developers are the core of functional characteristics that determine the entire functionality of a blockchain. According to accessibility and authority, blockchains can be identified as public (e.g. ETH) and private (e.g. ripple). Private chains are closed networks that offer registered users the benefits of cryptography technology, but are not necessarily decentralized or distributed. Therefore, they are not the focus of this study.

Public chains, on the other hand, are generally open source collaborative projects that anyone can easily access. The architecture of a blockchain, including any hard and soft fork upgrades, is a collaborative design with a series of democratic decisions among voluntary developers. Developer involvement has a significant impact on both technical and economic performance of a crypto. Normally, their involvement can be observed through the project repository, such as Github. Below are two indicators that can help us assess developer involvement.

(a) (b)
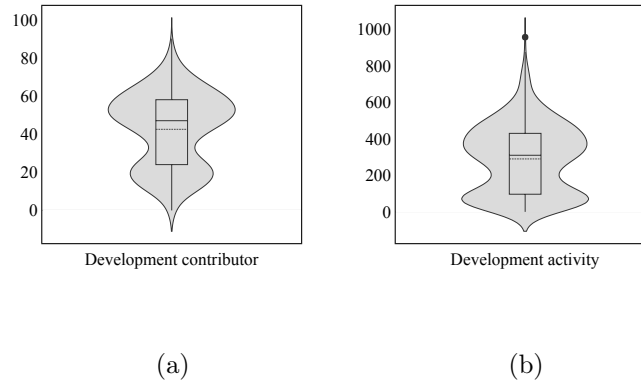
Figure 3.6: **Developer factors**.

Definition 12 (**Development contributor**). *The number of developers involved within a day for a development event.*

Developer engagement create and progress toward the development and adoption of the blockchain network. They design the front-end and back-end of a proposed blockchain, which then serves as a foundation for others to follow. Even though blockchains are being built, they are constantly developing new features and technology to improve functionality or activate upgrades. By facilitating applications that bring value to the end-users, developers attract more users, which in turn draws the attention of more developers.

Figure 3.8 indicates that the daily development contributors on the ETH network are around 50 developers. The majority of developers in the crypto space are on the Ethereum network, and the number is continuing to grow (Shen & Garg, 2022). In Figure 3.A.12a shows, the number of developers involved in ETH's development events is considerably dynamic. This may be related to the maintenance and upgrade in the Ethereum Improvement Proposals (EIPs), which are implemented regularly. This periodic pattern can be observed in Figure 3.A.12b.

Definition 13 (**Development activity**). *The number of development events in the project repository of the blockchain.*

An event is a specific activity that occurs in the blockchain project's repository and initiates a workflow run. Such events include every adjustment made to the repository, from code commits to new users joining the project. All comments on issues, forks, stars, etc. are excluded from the development activity here. Typically, a high number of development events may indicate a significant upgrade of the protocol.

ETH has an average of 300 daily events on its public repository on GitHub, see Figure

3.6b. Similar to the factor development contributor, the time series of the number of daily development events shows a periodic pattern, see Figure 3.A.13.

3.4   Market dynamics

This section focuses on the factors that connect users' trading behaviors, market dynamics and economic values. Several aspects can be used for characterizing market dynamics. Here, we examine dynamics via four dimensions. `utilities that coin offers`, referring to coin's market value, the realization of the monetary value of cryptoeconomics. The value of a coin is determined by its market supply and demand, which can be associated with the functions that coins aim to offer. For example, every utility token is used for one specific purpose (e.g. energy token, data token); shitcoin or memecoin offer no real utility and may be a scam; stablecoin is the coin that pegs to a reference asset (e.g. gold, fiat money). `medium of trading`, a channel for users to exchange coins. The common medium is either on-chain (i.e. decentralized exchange) or off-chain (i.e. centralized exchange). `trading behavior of users` in aggregation instead of at the individual level. `applications` addressing the use cases of one particular kind of coin, e.g. the NFT markets.

3.4.1   Coin circulation

Coin circulation is determined by the monetary policy designed and its adaption to current supply and demand. Historically speaking, we have considered a currency with price stability and a sufficiently large network of users to be functionally successful. The majority of crypto assets operate under a decentralized and algorithmic governance framework, which serves as an invisible hand to equilibrate the supply and demand of coins and then stabilize the price. Each crypto has its mechanisms for coin emissions from the ICO (initial coin offering) wallet and coin discards from current market circulation. Some coins have capped their total supply, which is the number of coins currently locked up in escrow of the circulating supply. For example, BTC will no longer release a new coin after reaching 21 million coins.

For ETH, the total coin supply is uncapped and is growing slowly, with an average of 9,000 coins per day during our observation period, see Figure 3.7. The ICO wallet has released some small amounts of coins, with the highest being around 35,000 coins a day and the lowest below one coin a day. The coin releases are labelled with red triangles indicating the release date, as each number of coins is comparably smaller than the total coin supply. Since the London upgrade on August 5, 2021, which introduced a new burn mechanism, ETH has burned many coins, with the highest being more than 2,800,000 coins a day and the lowest being around 3,000 coins a day. The burned coins

are marked with a blue area. Furthermore, there exist other monetary measures that we have not mentioned or observed, and they are cooperating with the adjustment of the coin supply over time. We, therefore, investigate the deviation between two consecutive days.
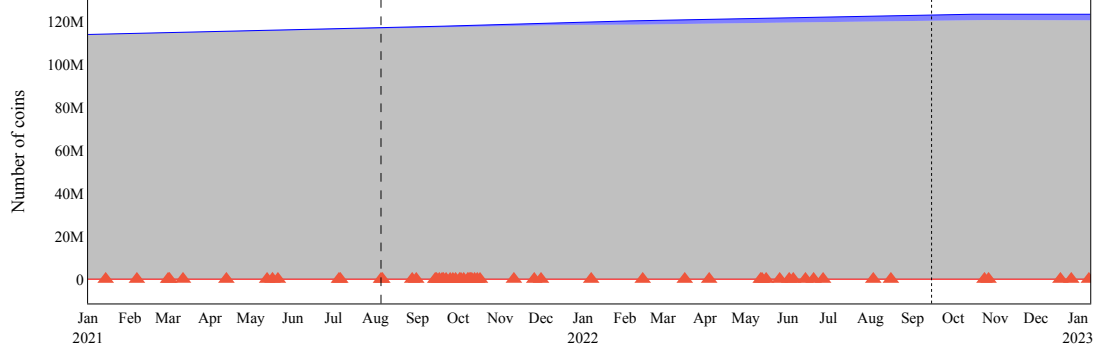
Figure 3.7: **The dynamics of ETH coin supply.** The gray area is the current coin supply, the blue area is the burned coins, and the red triangles around 0 are the times having any ICO released coins. The vertical dashed and dotted lines represent the times for London (EIP-1559) and Merge (EIP-3675) upgrades, respectively.

| (a) | (b) | (c) |
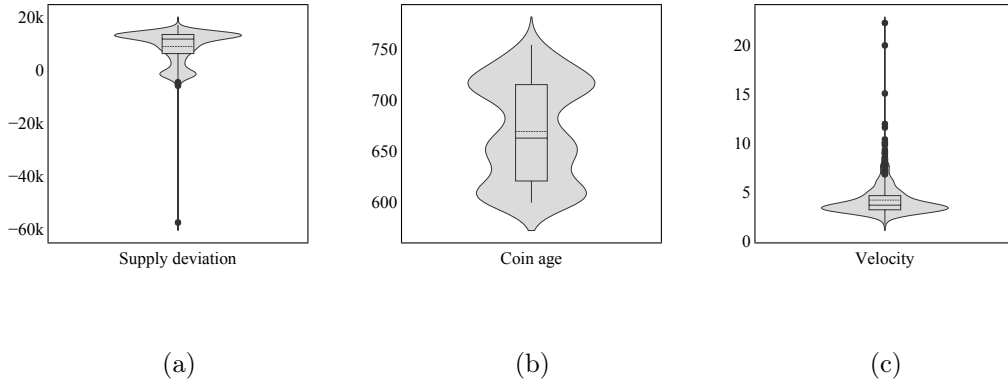
Figure 3.8: **Coin circulation factors**.

Definition 14 (**Supply deviation**). *The difference between the current day circulated coins and the previous day's.*

We are interested in how the monetary policy regulates supply, rather than investigating the total supply over time that remains relatively steady. Supply deviation reflects how coin release and burning, and other monetary measures on ETH influence the entire network. Figure 3.8a presents that the deviation per day is generally a positive sum, but

there is an outlier that reduces the coin supply by around 60,000 coins. Until December 2022, the total supply of ETH has been steadily increasing, see Figure 3.A.14a. The highest level of coin burning occurs in May of the same year, which corresponds to the lowest outlier. After the Merge, ETH's supply deviation significantly drops in October, and then it starts deviating around 0. Figure 3.A.14b shows that the supply deviation has autocorrelation, with up to 6 lags.

Definition 15 (**Coin age**). *The average amount of days that each coin has stayed in its current addresses.*

Coin age indicates how long a coin has been in a user's possession, which are often used as a way to prioritize its use in transactions or mining. It also indicates the liquidity of a coin. A long coin age indicates that users often trade in a long position and own the security, which means that most users consider such a crypto coin as an investment instrument. Conversely, a short coin age represents trading in a short position, which can be considered high speculation in the market or it has been used as a medium of exchange depending on price stability. This factor is related to the aggregate behavior of the entire network.

Figure 3.A.15a demonstrates that the average coin age of ETH continues to grow steadily, and that the entire market has an accumulation trend. The time series does not demonstrate significant autocorrelation in Figure 3.A.15b. The holding period is approximately 670 days, that is, there exist many long-term holders in the ETH network, see Figure 3.8b. Since PoS rewards staking, the coin age may continue to increase. Similar to energy for PoW, coin age becomes an expensive resource to accumulate in massive quantity, which can be a potential drawback for PoS.

Definition 16 (**Velocity**). *The average frequency that a coin gets transferred around addresses within a day.*

Compared to coin age measured in units of time, velocity considers the frequency of a coin's transfers within a day. It can be viewed as a turnover of a crypto. The higher the velocity the coin, the more flourishing its market, with more users and transactions. Attacking the drawback of staking, PoSV (Proof of Stake Velocity), Reddcoin consensus algorithm, emphasizes the concept of velocity (Ren, 2014). It encourages coin transfers among users by defining coin aging as an exponential decay function instead of a linear one. That is, a coin with a lower coin age gains its weight of the vote quickly, while one with a higher coin age gains its weight increasingly slowly.

Figure 3.8c shows that the average velocity of a day for ETH is around 4 transfers. The most frequent transfers of coins occur from April to July in 2021, as Yuga Labs – the creators of the NFT collection: Bored Ape Yacht Club (BAYC) – launched

its Metaverse project "Otherside" and sold 55,000 NFTs, see Figure 3.A.16a. In the following, ETH's velocity starts to decrease and remains around its mean. Furthermore, it exhibits autocorrelation, with up to 8 lagged values.

3.4.2 Market scale

In this section, we examine the total market value of a coin and the financial value of trading per day. Market scale can be linked to coin adoption as well as aggregate transaction behavior. The more users and transactions of a crypto, the higher its adoption. Financial value discussed here is measured in US dollars.

(a)   (b)
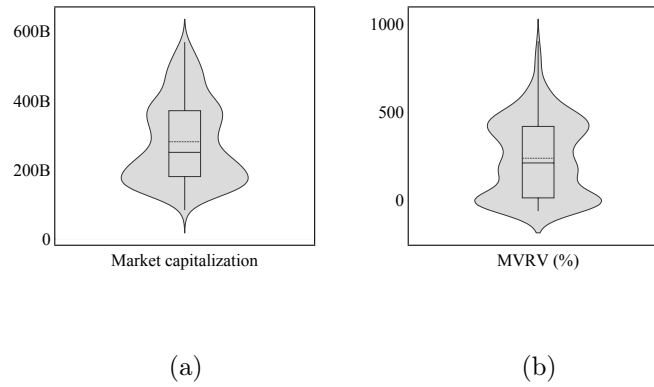
Figure 3.9: **Market scale factors**.

Definition 17 (Market capitalization). *The current market value of all unspent transaction outputs of a crypto that are circulating within the network.*

Unspent transaction output (UTXO) denotes a transaction output that can be used as input in a new transaction. Market capitalization (market cap) is simply derived by multiplying UTXOs, which are the current circulating supply, by the current price. It is the most common measure for the financial size of a crypto. Often, a stable market is considered to be one with a stable market cap, even during the turbulence.

The market cap of ETH reaches its all-time high of around 570 billions USD in November 2022, and it suffers from continuous downhills since then, see Figure 3.A.17. The continuous downhills can be observed in Figure 3.9a. There is a higher probability at the bottom of the violin plot. Over the observation period, ETH's market cap is not stable, with a high standard deviation of over 100 billion USD. After the Merge, it remains around 150 billions USD.

Instead of using market value, i.e. current price, realized capitalization (realized cap) uses realized value – which is the price at the time the coin last moved – to determine

the capitalization of a crypto. The last moved price is also the basis cost of obtaining a coin. It emphasizes the actual presence of coins in the blockchain network, while discounting the impact of lost and dormant coins, which are coins that have not been transacted for a long period of time. We illustrate ETH's market cap and realized cap in Figure 3.10; mark the difference between them in green as Market cap $\geq$ Realized cap and in red as Market cap $<$ Realized cap. The market cap of ETH was continuously higher than its realized cap by a large deviation before June 2022, and the market has an uptrend. That is, the market is aggregating profit and users are buying at low and selling at high. Later, as the market cap is lower than realized cap. The market begins to experience a downtrend and accumulates loss. Investors tend to trade at a lower price than ETH's current market value as they are bearing the market. Especially, two slight downward trends occur following the Merge. Ethereum market is recovering and receiving increasing attention from users as its protocol is updated.
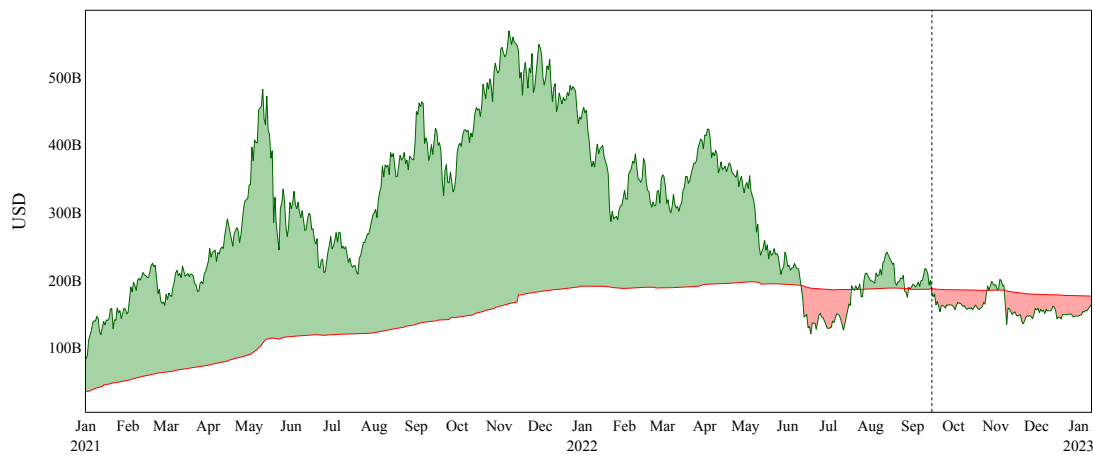
Figure 3.10: **Market cap** and **realized cap**. The colored area is Market cap $\geq$ Realized cap and Market cap $<$ Realized cap. The vertical dotted line is the date of the Merge upgrade.

Definition 18 (**MVRV**). *The ratio of a crypto's market capitalization to its realized capitalization.*

As previously discussed, the realized cap provides a meaningful insight into actual market evolution and profitability when compared to the market cap. MVRV (market-value-to-realized-value) concludes the concept of the comparison and discloses whether the crypto is undervalued or overvalued by the market. For example, if the MVRV is smaller than 1, then the crypto is undervalued and not profitable.

Figure 3.9b shows that ETH's market cap is generally higher than its realized cap, with an average of around 200% higher. In other words, ETH is a rather profitable

crypto. However, the MVRV has a significant downward trend and approaches below 0 after the Merge, see Figure 3.A.18. The market is still undervalued and not fully recovered. Figure 3.A.18b shows there is limited autocorrelation, with up to 2 lags.

3.4.3 Transaction pattern

In this section, we focus on users' collective trading behaviors. Depending on the current economic status and market value of a crypto, rational users will change their investment strategy to accommodate their risk tolerance and future needs for capital. Particularly, cryptocurrency prices suffer from high volatility, and they have often been considered an asset for speculation by the public. An investment strategy that is adaptive and market-focused is of importance. Consequently, a rapid change in the collective trading behaviors has often been discovered in crypto markets. Many research studies have indicated that there is a coordination mechanism for the trend of herding and feedback behaviors in crypto trading (Gurdgiev & O'Loughlin, 2020; King & Koutmos, 2021).

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
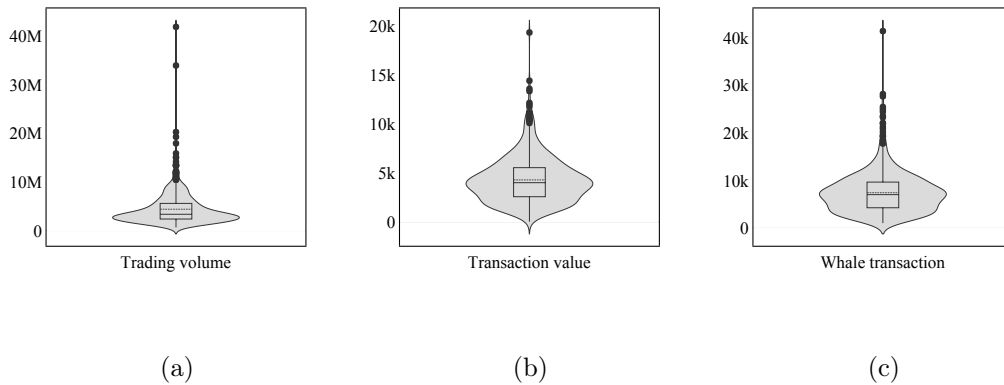
Figure 3.11: **Transaction pattern**.

Definition 19 (**Trading volume**). *The total dollar amount of transactions of a day.*

Trading volume provides an important link between momentum and value strategies (Lee & Swaminathan, 2000). For example, the magnitude and persistence of price momentum are a function of trading volume. As trading volume is an aggregate user behavior, it has been used to observe speculation and other collective trading behaviors, e.g. herding (Youssef, 2022); speculative trading (Bouoiyour et al., 2015); wash trading (Cong et al., 2022).

ETH has a daily trading volume of around 5 million USD, with a standard deviation of more than 3 million USD which is considerably volatile, see Figure 3.11a. ETH's

trading volume surges to an all-time high in May 2021 and significantly decreases as approaching the upgrade EIP-1559 (London) in August, which aims to reduce everlasting high transaction fee. Figure 3.A.19b demonstrates that the trading volume has autocorrelation of fifth order.

Definition 20 (**Transaction value**). *The average amount of transaction value per day.*

From the daily transaction count, we can determine the average transaction value from trading volume, which can help us distinguish what role a crypto plays toward users. With many small value transactions, it may be seen as a medium exchange instead of an asset for speculation. High-value transactions, however, often occur during speculation or specific events, such as hacking attacks, an exchange's bankruptcy, or governmental policy changes.

In Figure 3.11b, the largest outlier occurs on May 19, 2021, with the average transaction value around 19,373 USD during the crypto crash. This could indicate that users are actively avoiding future loss. The transaction value then starts to decrease, with an average of 4,200 USD, see Figure 3.A.20a Figure 3.A.20b shows that ETH's transaction value has autocorrelation, with up to 7 lagged values. This could be related to the herding and feedback trading behaviors.

Definition 21 (**Whale transaction**). *The amount of transactions a day that exceed 10^5 USD.*

Whale transactions often involve large coin holders and institutional users, who have a significant impact on coin supply and demand as well as market sentiment. Their coin movements are considered a major driver for price fluctuations, as a result, there are many crypto whale trackers, such as ClankApp and Whale Alert. Figure 3.12 shows that ETH's average transaction value evolves similarly with its whale transaction count, which indicates that these whales have a heavy influence on its average transaction value.

Whales are viewed as early adopters who exert market power by manipulating prices, i.e. pump and dump, and further triggering market informational cascades or herding. That is, the extreme imbalance of coin holding leads to information asymmetry and adverse selection in the market (Tiniç et al., 2020). Particularly, PoS may provide these whales with an opportunity to dominate the blockchain network, as they stake more coins and obtain higher weight of their votes. Furthermore, whale transactions can also crowd out low-value transactions, as they are able to offer higher transaction fees to validators, corresponding to Zimmerman, 2020's study. In Figure 3.12, the similarity between the whale transaction count and average transaction value may be
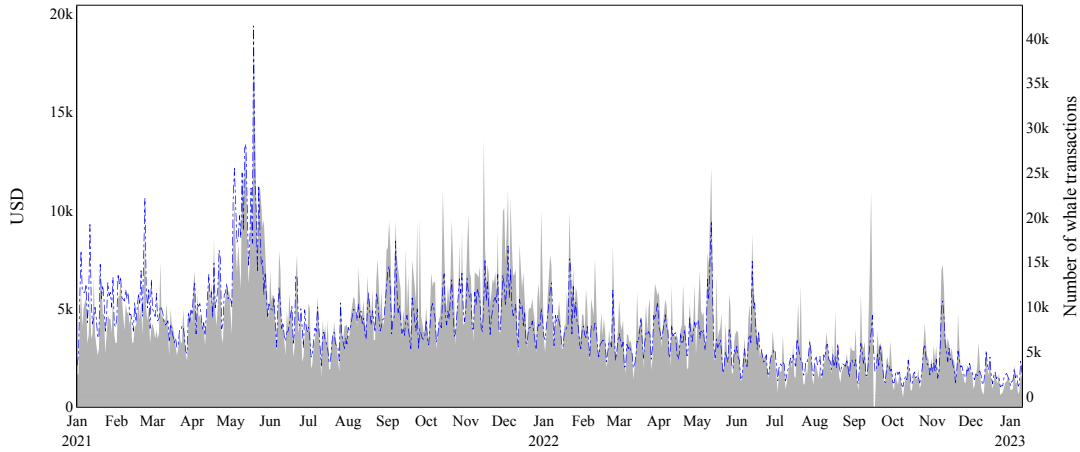
Figure 3.12: **Daily whale transaction count** and average transaction value.

explained by such a phenomenon. This effect of crowding out can prevent the crypto from being a medium of exchange.

Figure 3.A.21a shows that the transactions that exceed 10^5 USD pump up to an all-time high of around 41,000 transactions on May 21, 2021, and begin to drop to the level of around 7,500 transactions. The deviation of whale transaction count is large, i.e. large-valued transactions do not occur in a daily basis, see Figure 3.11c. We observe autocorrelation, with up to 7 lagged values in ETH's daily whale transaction count.

3.4.4 Exchange activity

There are currently over 600 crypto exchanges worldwide according to CoinGecko. These exchanges rely on clearing trades and off-chain scaling. They offer more advanced asset-management tools and diverse financial products, e.g. crypto derivatives, options, and futures, Users can trade between multiple cryptos and fiat currencies as well as other crypto-extended financial products with ease and flexibility. However, it may exacerbate the unregulated and volatile nature of cryptos with exchange third-party risk exposure, which may subject users to significant risk of loss. In November 2022, one of the largest exchanges worldwide, FTX, collapsed due to a shortage of liquidity. This triggered a chain effect among exchanges and resulted in an unstable market (Wang et al., 2022). Since exchanges are common in crypto markets, they have held many coins for users and reserves. Their activity therefore has the potential to influence the entire market.

Definition 22 (**Flow balance**)**.** *The difference between the amount of USD flows in and out of exchange wallets.*

(a)   (b)
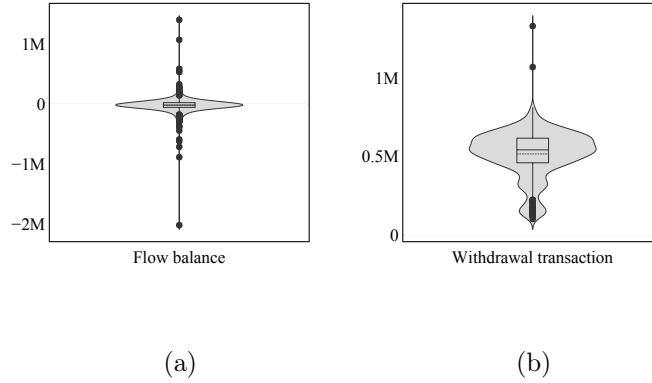
Figure 3.13: **Exchange activity**.

Many users invest in cryptocurrencies by trading them on exchanges instead of doing on-chain trading that involves volatile transaction fees and complex operations on the blockchain. Moreover, small-valued transactions are often crowded out by whale transactions in PoW, as higher valued transactions offer high fees to the validators. This increases the popularity of exchanges, for example, the biggest exchange Binance has a daily trading volume over 25 billion USD, see CoinMarketCap for the up-to-date information. The flow balance of the exchange wallet provides a summary of user behavior, indicating whether the majority of users are depositing funds to participate in the market or withdrawing them.

During the observation period, the average exchange wallet for ETH has a balance of -14,000 USD, which indicates that users withdraw more than they deposit, see Figure 3.13a. Out of 740 days, there are 426 days where the withdrawal amount is larger than the deposit The largest amount of withdrawal in Figure 3.A.22a is on November 15, 2021 which is during the period that ETH's price reaches its all-time high. It comes with a spike after the merge and then drops down. The time series of flow balance shows really limited autocorrelation in Figure 3.A.22b.

Definition 23 (**Withdrawal transaction**). *The total number of transactions that involves the withdrawal addresses of exchanges.*

Withdrawal transaction count considers any transactions linking to exchange withdrawal addresses, regardless of transaction value. Without the influence of large-valued transactions, it can provide a better representation of collective user sentiment and activity within exchanges. ETH has around half a million withdrawal transactions on exchanges per day, and the amount deviates greatly with the standard deviation of around 150,000 transactions, see Figure 3.13b. It has been receiving increasing sell pressure from exchanges since April 2021. Especially after FTX's bankruptcy in November

2022, the withdrawal count surges to an all-time high of over a million transactions in December, see Figure 3.A.23a. ETH's withdrawal count shows autocorrelation with up to 4 lags, which indicates that selling from users has persistence.

3.4.5 Wealth distribution

Wealth distribution refers to the distribution of coins in a crypto network. It is a common measurement for wealth inequality. As previously discussed, whale transactions can have a significant impact on coin supply and demand, which could push a crypto to be an asset for speculation. Particularly under a PoS protocol, a large disparity in coin distribution leads to voting inequality. That is to say, it builds a barrier for small coin holders to participate in consensus making. Gupta and Gupta, 2018 points out that the rich have actually become richer, and that steps should be taken to curb such a wealth accumulation model in the network in the case of Bitcoin. Despite the functionality of a blockchain, fair wealth distribution also determines a crypto role in the market, such as a medium of exchange or speculative asset. Here, we take the supplies on exchange and non-exchange top holder in the discussion. The non-exchange top holders are the top 10 ETH addresses that have the highest balances of a day and are not owned by a known exchange. Figure ?? displays the proportions of ETH's held by exchanges and top holders over the total coin supply. Coins on exchanges start declining, while the ones on non-exchange top holders continue to increase. Overall, supply on exchange and non-exchange top holder accounts for around a third of the total supply.
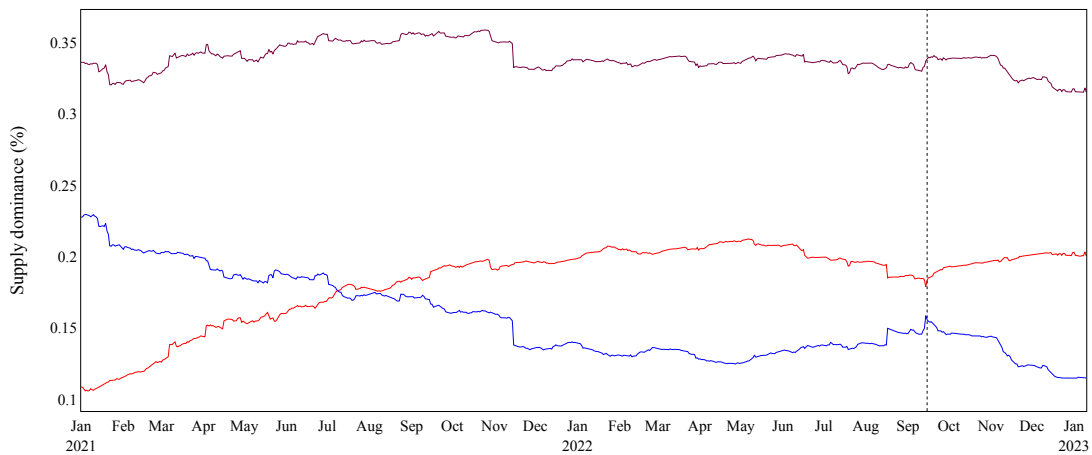
Figure 3.14: **Exchange dominance**, **non-exchange top holder dominance** and **aggregate (both) dominance**. The vertical dotted line is the date of the Merge upgrade.

Definition 24 (Supply on exchange). *The number of coins held by exchange addresses.*

(a)   (b)
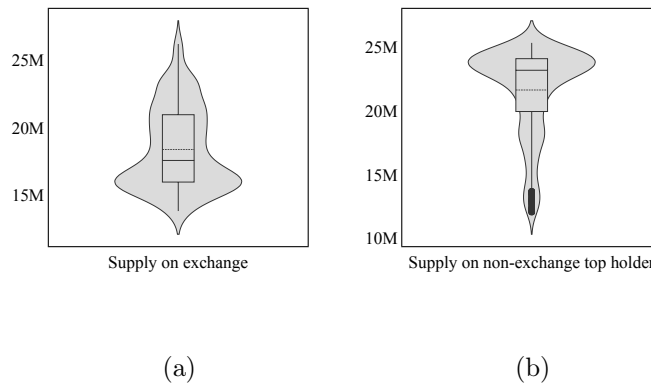
Figure 3.15: **Wealth distribution**.

Crypto exchange is a common channel for small coin holders who often value ease of use, a high variety of assets, deep liquidity, and customer support. Therefore, exchange wallets often include small coin holders' supply. ETH's supply on exchange is average 18 million coins, see Figure 3.15a. Figure 3.A.26a indicates that the supply on exchange continues to decline and peaks up while the Merge is happening. The supply on exchange is autocorrelated with only one lag in Figure 3.A.25b.

Definition 25 (**Supply on non-exchange top holder**). *The number of coins held by the top addresses, excluding exchange addresses.*

Non-exchange top holders are the top 10 addresses that have the most coins on the network, and exclude any exchange address. They are either an individual or an institution that does not serve as a platform for others' transaction settlements.

The top ETH holders by their balance can be found in Etherscan. The top 10 ETH holders continue to increase their holdings until May 2022, and then gradually reduce their holdings, see Figure 3.A.26a. After the Merge, they start buying ETH again. The supply on these holders is only autocorrelated by the first lag in Figure 3.A.26b. Figure 3.15b indicates that the top ETH holders remain holdings steadily around 20 to 25 million coins.

3.4.6   Social media

It is often noted that market sentiment is a metric for the influence of social media on an investor's overall attitude toward a particular market or an asset. Market sentiment varies depending on the text classification (i.e. positive, negative, neutral) for news, messages, comments, or other textual data from public media sources such as Twitter, Reddit, or Nasdaq news. By aggregating sentiment scores from each

textural data, it provides a straightforward narrative to explain market behaviors. Numerous research studies have indicated that the crypto market is susceptible to market sentiment or investor attention (Chen & Hafner, 2019; Rognone et al., 2020). There are, however, many challenges to interpretation, such as word ambiguity, irony, sarcasm, and multipolarity. Also, it is difficult to have a single source of data that can represent the market.

(a)   (b)
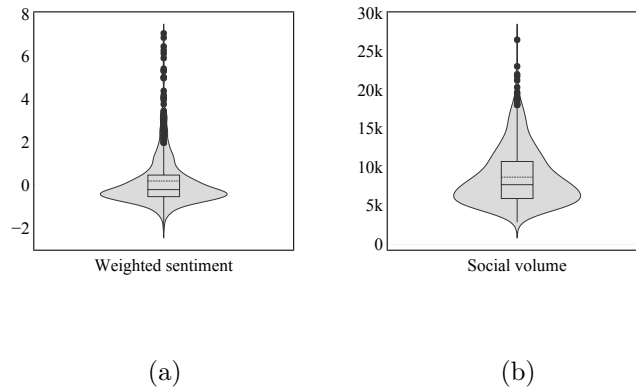
Figure 3.16: **Social media**.

Definition 26 (**Weighted sentiment**). *The difference between the positive and negative sentiments weighted by social volume over the observation period, excluding any duplicated messages.*

According to classification models, corpuses used, or aggregation methods, there are different measures to consider market sentiment. Here, we use weighted sentiment – sourced from Santiment – as an example, see Appendix 3.B for details.

ETH's wighted sentiment is generally neutral to positive with an average of 0 slightly higher than 0.in Figure 3.16a. During our observation period, it appears multiple times during the Merge upgrade and approaches its all-time high, see Figure 3.A.27a. ETH's PoS is viewed positively by users. The weighted sentiment has autocorrelation, with only up to 2 lagged values.

Definition 27 (**Social volume**). *The number of mentions of a coin on social media, i.e. Telegram, Twitter, Reddit, Bitcointalk.*

Alternative to market sentiment metrics, the frequency with which a crypto has been searched or mentioned is often regarded as the popularity of the crypto. Several common metrics used in the literature include Google trend, Twitter hashtags, and Reddit mentions.

Sourced from Santiment, the daily ETH social volume is around 8600 mentions, see Figure 3.16b. Any special event, such as releases of protocol upgrades, often induces spikes in social volume in Figure 3.A.28a. Figure 3.A.28b exhibits that ETH's social volume has autocorrelation until 5 days. It demonstrates higher autocorrelation than the weighted sentiment, since weighted sentiment discards the sentiment value between the interval $[-0.7, 0.7]$, which leads sentiment spillover to be omitted.

3.4.7 Applicability

In accordance with each crypto's blockchain, there exists a variety of extension applications that create an extension market. Applicability hinges on this blockchain to conduct transactions and support its functions. Examples include NFTs to Enjin (ENJ) and Ethereum, blockchain oracle to Chainlink (LINK), data exchange market to Ocean Protocol (OCEAN), and energy transition to Energy Web Token (EWT).

Since this study focuses on the Ethereum market, we use the NFT market for discussion below. NFTs are a digital token used as a proof of ownership and authenticity for both unique tangible and intangible assets such as artworks, domain names, in-game assets, and luxury goods. The first NFT standard, ERC-721, was launched by ETH. The following token standard, ERC-1155 Multi Token Standard, enables the efficient transfer of fungible and non-fungible tokens in a single transaction, which significantly reduces the transaction costs. Because ETH has more liquidity than most other cryptos, it is the most popular blockchain for NFT development and the main medium of exchange for NFTs.

(a)   (b)
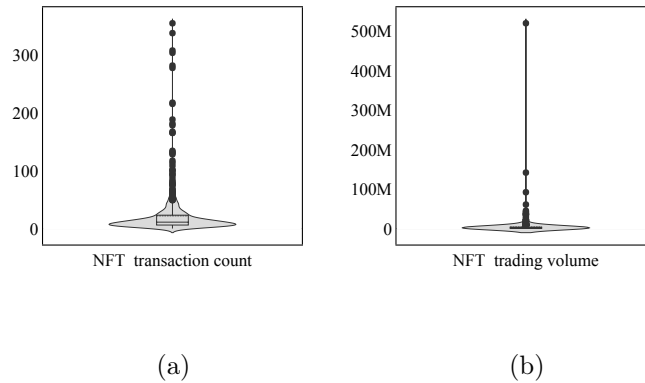
Figure 3.17: **Applicability**.

Definition 28 (**NFT transaction count**). *The total number of NFT transactions a day.*

NFT transactions are active in the year 2021, but they decline significantly later, see Figure 3.A.29a. Similar to the auction market, we observe a trading pattern that is irregular. The outliers in Figure 3.17a that frequently occur are also an indication. Thus, the NFT transaction count does not show significant autocorrelation in Figure 3.A.29b.

Definition 29 (**NFT trading volume**). *The total USD amount of NFT transactions of a day.*

The NFT trading volume here depends on the change in the value of the medium of exchange, i.e. ETH. A creator often considers the current price of ETH and transaction costs when determining the listing price of an NFT. Especially, ERC-721 is still widely used and does not facilitate a single transaction for multiple tokens. To sell an NFT, two transactions must be made: one for NFT trading and the other for ETH trading.

The NFT trading volume is on average 4 million USD a day, however, there are several large valued outliers in Figure 3.17b. The largest sales occur on October 28, 2021, which is the date that a CryptoPunk #9998 was sold for 124,457.07 ETH. Otherwise, the trading volume remains low, even as zero in a day, see Figure 3.A.30a. Because the market is so extreme, Figure 3.A.30b shows that there is no significant autocorrelation.

3.5   Economic attribute

Cryptocurrencies have been considered to be volatile, experiencing short-lived bursts of upward and downward movements over time. Volatility is time-varying, but often exhibits persistence, by which the current return has a large effect on the unconditional variance of many following periods. This is also known as a stylized fact – volatility clustering

Volatility persistence in the crypto markets is inevitably ruled by the blockchain characteristics, which yield various market microstructure noises, such as discrete price changes, gradual responses of prices toward market sentiment, and bid-ask bounces (Dimpfl & Peter, 2021). The microstructure noise may cause a spurious autocorrelation in high-frequency returns that worsens when estimating volatility using high-frequency data (Andersen et al., 2017). In order to understand the dynamics of the crypto return volatility, we opt for a volatility model that enables disentangling the desired persistent behavior from the part contaminated by microstructure noises. The Beta-t-EGARCH model (A. Harvey & Lange, 2018) is proposed and employed here in the hope of separating the long-run volatility component from the short-run part. In this paper, we discuss log returns, the two volatility components, and volatility (conditional standard deviation) derived from log returns as economic attributes. In the following, we review

the formulation of the model and review the characteristics of short and long-run volatility components.

3.5.1   Beta-t-EGARCH

It has been observed that the crypto market exhibits a high degree of volatility clustering and asymmetry, and is contaminated with outliers or jumps (Dutta & Bouri, 2022; W. Zhang et al., 2018). To account for the distributional characteristics of cryptos, we use an exponential specification of volatility, an exponential GARCH, EGARCH (Nelson, 1991). However, the EGARCH is not robust in the presence of outliers or jumps. One may need to adopt the dynamic conditional score (DCS) model in which each dynamic equation that drives a time-varying model parameter is updated by the conditional score of the log-likelihood with respect to the same time-varying parameter (Blazsek & Licht, 2020). The standard residual that updates each dynamic equation in the parameter driven model is replaced with a score function. Exponential DCS (dynamic conditional score) volatility model: Beta-t-EGARCH model (A. Harvey & Sucarrat, 2014) with $t = \{1, \ldots, T\}$ is written as

$$y_t = \exp(\lambda_{t|t-1})\varepsilon_t \tag{3.1}$$

Given the non-normality and fat tails of crypto returns, we assume ε_t has a Student t_ν-distribution with the degree of freedom $\nu > 2$. If $\nu \to \infty$, then the model converges to a Gaussian density function. $\exp(\lambda_{t|t-1})$ is the scale with a dynamic equation driven by the score u_t with respect to y_t. The dynamic equation for the log of scale is

$$\lambda_{t+1|t} = \delta + \phi\lambda_{t|t-1} + \kappa u_t \tag{3.2}$$

with conditional score

$$u_t = \frac{(\nu + 1)y_t^2}{\nu \exp(2\lambda_{t|t-1}) + y_t^2} - 1 \tag{3.3}$$

where $-1 \leq u_t \leq \nu$ and variance is the square of the scale, i.e. $\sigma_{t|t-1}^2 = \exp(\lambda_t)^2[\nu/(\nu - 2)]$. u_t is IID, and may expressed as $u_t = (\nu + 1)b_t - 1$ where, for finite degree of freedom $0 < \nu < \infty$,

$$b_t = \frac{y_t^2/\nu \exp(2\lambda_{t|t-1})}{1 + y_t^2/\nu \exp(2\lambda_{t|t-1})}$$

and $0 < b_t < 1$, distributed as beta$(1/2, \nu/2)$ over time, see A. C. Harvey, 2013, p. 99. So, $\mathsf{E}(u_t) = 0$ and $\sigma_u^2 = 2\nu/(\nu + 3)$. Incorporating the leverage effect that volatility tends to respond to price drops than to rises, the first-order model, Equation 3.2, is written with the sign of the observations as

$$\lambda_{t+1|t} = \delta + \phi\lambda_{t|t-1} + \kappa u_t + \kappa^* u_t^* \tag{3.4}$$

where κ^* the parameter to be estimated, which is normally non-negative, due to taking signum function of minus y_t, i.e. $u_t^* = \text{sgn}(-y_t)(u_t + 1)$. Given a positive return, the term $\kappa^* u_t^*$ enables $\lambda_{t|t-1}$ to respond asymmetrically to positive and negative values of y_t, e.g. $u_t^* < 0$, volatility can be downward sloping. With the restriction $\kappa \geq \kappa^* \geq 0$, one can impose that an increase in the absolute value of a standardized observation does not lead to a decrease in volatility.

To subtract the structure noise from the crypto returns instead of capturing long memory by a fractionally integrated process, we use two component setting, that is,

$$\lambda_{t|t-1} = \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1} \tag{3.5}$$

with long-run volatility component and short-run volatility component written as

$$\lambda_{1,t+1|t} = \phi_1 \lambda_{1,t|t-1} + \kappa_1 u_t$$

and

$$\lambda_{2,t+1|t} = \phi_2 \lambda_{2,t|t-1} + \kappa_2 u_t + \kappa^* u_t^*,$$

respectively. $\phi_1 \neq \phi_2$ imposes to ensure identifiability of the model. Having ϕ_1 close to one presents the persistence in long-run volatility component; while short-run volatility component has usually a higher κ_2 with lower ϕ_2. Here, the leverage effect, termed $\kappa^* u_t^*$, is only included in $\lambda_{2,t+1|t}$. In this case, the evolution of $\lambda_{1,t+1|t}$ is less susceptible to the effect of strongly negative returns and so can be suitable for capturing the ARCH-M effect, i.e. volatility clustering, persistence (Engle & Bollerslev, 1986). The skewness can be introduced by subtracting the expectation of ε_t, denoted as μ_ε from the error term in Equation 3.1, i.e. $\varepsilon_t^* = \varepsilon_t - \mu_\varepsilon$ where $\varepsilon_t^* \sim \text{skew-}t(0, \sigma_{\varepsilon*}^2, \nu, \gamma)$, $\nu > 2$, $\gamma \in (0, \infty)$, see A. Harvey and Sucarrat, 2014. If $\gamma = 0$, then we get a symmetric t-distribution. $\gamma > 1$ is the right-skewed; $\gamma < 1$ is the left-skewed.

3.5.2 Volatility component

Figure 3.18 compares ETH's log returns with the fitted conditional standard deviations of the one-component model and the two-component model, and the model coefficients are shown in Table 3.C.1. The log return series of the ETH is clearly characterized by time-varying volatility. The two-component model is more sensitive and volatile to the dynamics of log return series. Figure 3.19 demonstrates ETH's long-run and short-run volatility components. The long-run has a relatively lower frequency than the short-run.

Short-run volatility component captures return skewness risk driven by a strong
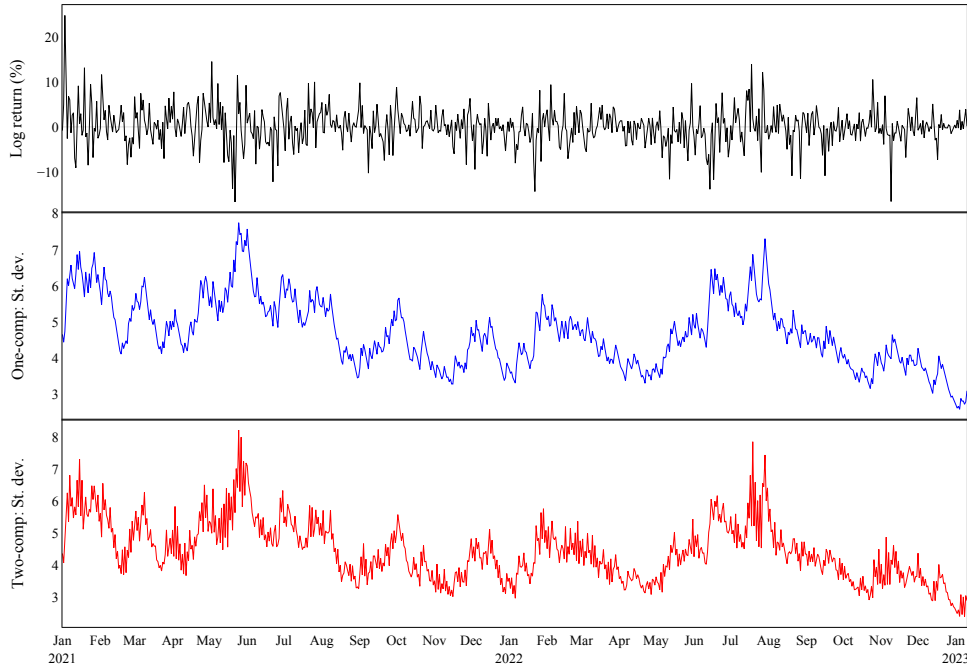
Figure 3.18: **ETH log returns (%), fitted conditional standard deviation of the one-component model, and fitted conditional standard deviations of the two-component model**.

leverage effect (Adrian & Rosenberg, 2008). Such skewness risk increases endogenously in pricing theories with financial constraints (Hong & Stein, 2003; Yuan, 2005). In the presence of different levels of microstructure noise, i.e. considering short-run volatility component, the permanent component – trend in time series can be concealed. Thus, it distorts price discovery and arises the market instability. As crypto investors highly depend on market sentiment (López-Cabarcos et al., 2021) and the market pervades information asymmetries and adverse selection, we expect that the influence from the short-run volatility component is higher than the classic financial assets.

Long-run volatility component preserves a higher level of persistence than the short-run. It often links to the business cycle or economic shifts in the market. Conrad et al., 2018; Fang et al., 2020 investigate the determinants of Bitcoin long-run volatility and associate it with the global economics and the other assets. Since cryptos are technology-driven, we expect that such exogenous influence from the global economics reflects on the endogenous factors in cryptos, i.e. the dynamics of blockchain characteristics. Also, we might relate the long-run volatility component to the deign and update of blockchain system.

Given above, we define a simple risk measure to see the degree of distortion from noise.
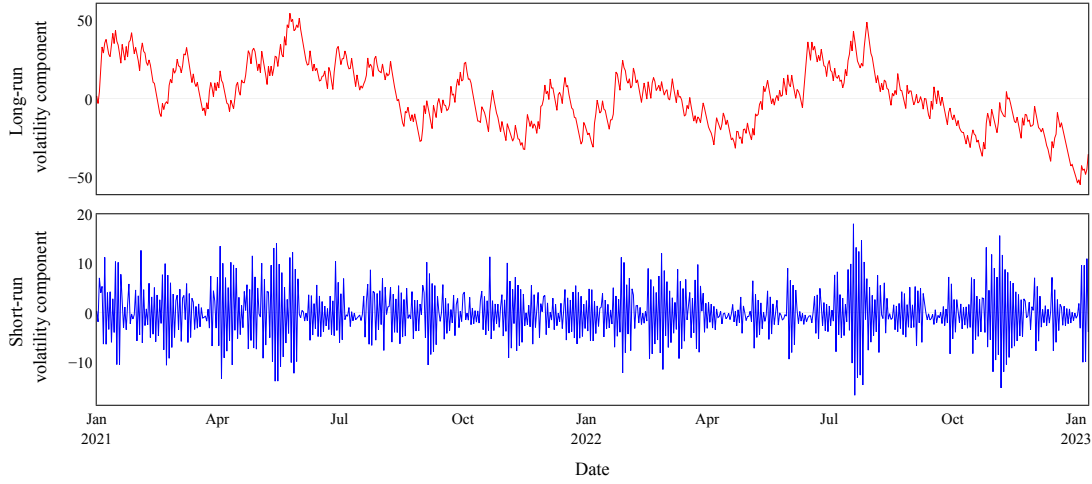
Figure 3.19: **Two compoents**.

We calculate **distortion level (DL)** for each time t.

$$\mathrm{DL}_t \stackrel{\text{def}}{=} \left[\frac{\exp(\lambda_{2,t})}{\exp(\lambda_{1,t}) + \exp(\lambda_{2,t})} \Big/ \frac{\exp(\lambda_{2,1})}{\exp(\lambda_{1,1}) + \exp(\lambda_{2,1})} \right] \times 100$$

where $\lambda_{1,t}$ and $\lambda_{2,t}$ are the long-run and short-run volatility components. DL is high if the crypto is highly influenced by (microstructure) noise; and otherwise. Figure 3.20 shows the distortion level of ETH. Our observations indicate that there is an increase in distortion level of ETH following a sudden drop of log return, that is, there may be more noises in the market. Moreover, despite a delay, the distortion level bears a resemblance to the 7-day moving average of log returns.

3.6 Data description

In this paper, we investigate whether the underlying blockchain functionally influences a cryptocurrency's economic performance, as well as whether aggregate market behaviors operationally direct this performance. We use Ethereum as an example. Particularly, it has been one of the major cryptos and has involved several interconnected protocol upgrades which have resulted in a variety of different dynamics in its ecosystem and cryptonomics, referred to Figure 3.1. The details of ETH upgrades are available in The history of Ethereum. ETH's empirical data used is sourced from Santiment API and BitInfocharts. The data for energy consumption proxy is retrieved from Ethereum Energy Consumption Index from Digiconomist. The period is between January 01, 2021, and January 10, 2023, with a daily frequency. There are 33 time series factors in conclusion, comprising 13 factors of functional characteristic, 16 factors of market
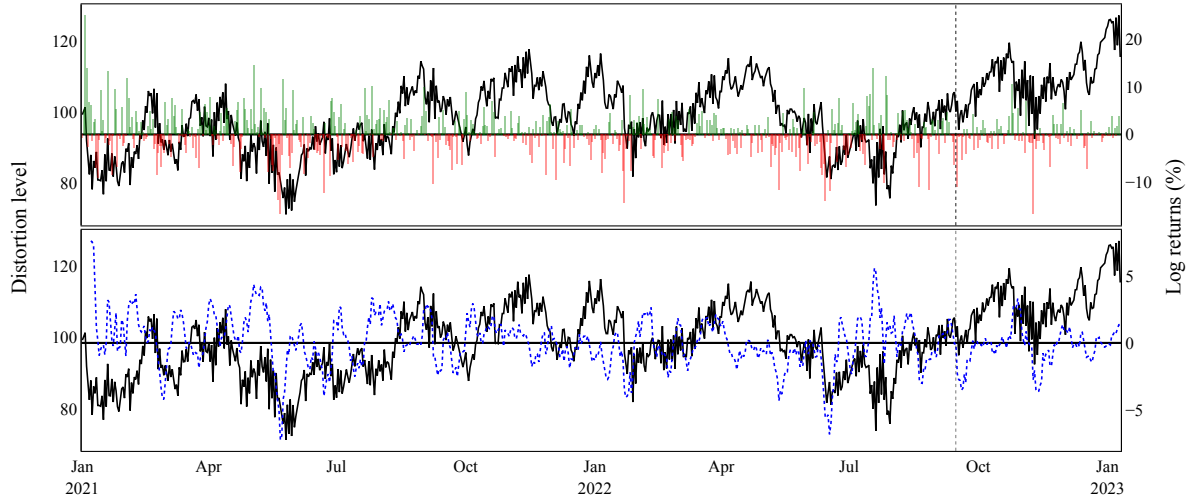
Figure 3.20: **Distortion level.** The bar chart in the upper panel shows positive log returns and negative log returns. The moving average of log returns over 7 days is in the lower panel in dotted line. The vertical dotted line is the date of the Merge upgrade.

dynamics, and 4 factors of economic attribute. We plot each time series factor and check its autocorrelation in Appendix 3.A. Table 3.D.1 summarizes their descriptive statistics, including mean, median, maximum, minimum, variance, skewness, and kurtosis.

3.7 Causal discovery

Causal discovery methods have undergone numerous developments, with an emphasis particularly on modern machine learning techniques and nonlinear causal dynamics (Imbens, 2022; Shojaie & Fox, 2022). Still, we apply the classic Granger causality (GC) test (Granger, 1969) because of its simplicity and wide adoption. Without any prior hypothesis, the GC test allows us to pinpoint directional influences of time series on one another. Thus, we are able to examine the cause and effect relationships among the multivariate time series defined as functional characteristic, market dynamics, and economic attribute of Ethereum.

3.7.1 Granger causality

Consider two time series $x^{(i)} = \{x_t^{(i)}\}_{t=1}^T$ and $x^{(j)} = \{x_t^{(j)}\}_{t=1}^T$ where $i, j = \{1, \ldots, n\}$ and $i \neq j$. The GC test examines whether including the past observations of $x^{(j)}$ can help to predict future values of the other series $x^{(i)}$, and vice versa. If so, we say $x^{(j)}$

Granger causes $x^{(i)}$, i.e. $x^{(j)} \overset{\text{GC}}{\Rightarrow} x^{(i)}$. The linear prediction model here is formulated as

$$x_t^{(i)} = c_0 + \sum_{\tau=1}^{p} \phi_\tau^{(i)} x_{t-\tau}^{(i)} + \sum_{\tau=1}^{p} \phi_\tau^{(j)} x_{t-\tau}^{(j)} + \eta_t \tag{3.6}$$

where $\phi_\tau^{(i)}$ and $\phi_\tau^{(j)}$ are coefficients; $\tau = \{1, \ldots, p\}$ the time lag; $\eta_t \overset{\text{iid}}{\sim} \mathcal{N}_n(0, \sigma^2)$ the error term; c_0 a constant. The hypothesis is to state that $H_0 : \phi_\tau^{(j)} = 0, \forall \tau$. That is, $\sigma^2(x^{(i)}|\mathcal{U}') = \sigma^2(x^{(i)}|\mathcal{U}' - x^{(j)\prime}) \succ x^{(j)} \overset{\text{GC}}{\nRightarrow} x^{(i)}$ where $\mathcal{U}' \subset \mathcal{U}$ is a set of prior values of the all causative variables; and $x^{(j)\prime} \subset x$ is the prior values of time series $x^{(j)}$. The GC test is performed pairwise on the 33 time series collected or computed.

As classic GC tests require the stationarity of time series variables, we accordingly conduct Augmented Dickey-Fuller (ADF) tests for stationarity check with a 95% confidence interval. Table 3.D.2 shows that the 13 factors – development contributor, development activity, velocity, flow balance, supply on non-exchange top holder, weighted sentiment, social volume, NFT transaction count, NFT trading volume, log return, long-run volatility component, short-run volatility component, and volatility fulfill stationarity. The rest of the factors are stationary at their first difference (Δ).

Instead of having a fixed lag order for all the pairs of variables, the maximum lag length p days for each pair in the GC test is determined by Bayes information criterion (BIC) of vector autoregression (VAR) models, such that $\hat{p} = \arg\min_p [\log |\hat{\Sigma}_p| + (k^2 p \log T)/T]$ where T the number of observations; k the dimension of the time series, i.e. $k = 2$ for bivariate VAR; $p = \{1, \ldots, T-1\}$ the estimated number of lags; $\hat{\Sigma}_p$ the estimated noise covariance matrix. BIC places a much higher penalty on the model and avoids overestimation while comparing to AIC (Akaike information criterion) (Gredenhoff & Karlsson, 1999; Ng & Perron, 2005). Table 3.D.3 shows the estimation result of the lag length for all pairwise combinations of factors. Based on the estimated lag lengths, Table 3.D.4 presents the Granger causality result for each pair under a 95% confidence interval.

3.7.2 Causal network

Given the pairwise GC test results, we define a granger causal network as a directed graph $G \overset{\text{def}}{=} (V, E)$ in which a finite and non-null set V of vertices are connected through a set $E \subseteq V \times V$ of edges. Each vertex $x^{(i)} \in V$ denotes a factor and $i = \{1, \ldots, n\}$, i.e. $m = 33$ here. Each edge $(x^{(j)}, x^{(i)}) \in E$ denotes the causal linkage between two factors and $j = \{1, \ldots, n\}$ where $j \neq i$. An edge $(x^{(j)}, x^{(i)})$ exists only if $x^{(j)} \overset{\text{GC}}{\Rightarrow} x^{(i)}$ and $p \neq 0$, that is, the Granger test for the pair of factors rejects the null hypothesis under a 95% confidence interval.

Ignoring the causal direction, Figure 3.21 shows ETH's causal network structured according to the empirical evaluation framework proposed as of Figure 3.1. Depending on the types of components, each factor is colored. There are causal relationships between the factors, with the exception of supply deviation. The changes in ETH's supply deviation are not followed by nor cause a significant change in the other factors in our observation period. In other words, the influence of monetary policies on ETH's supply may not be transmitted to its blockchain ecosystem and market dynamics. Policies such as the base fee burning in the London upgrade do not induce demand compensation for later settlement, or substitute with other assets. Moreover, MVRV and market capitalization have limited connections to the others within market dynamics, i.e. trading volume.

To account for the causal direction, Table 3.1 presents the in and out degree of each vertex. As for economic attributes, there are fewer connections outgoing from them to other factors than the incoming. Functional characteristics tend to have more out-degree connections than in-degree ones; whereas factors in market dynamics show the opposite. Economic attributes and market dynamics may be considered a response to blockchain mechanism.

3.7.3   Uncovering the invisible hand

Following, we discuss each economic attribute one by one. Each value on each edge indicates the lag length of days.

Log return (%). Figure 3.22 illustrates the in-degree and out-degree of ETH's log return. Log return is influenced by factors such as difficulty adjustment, energy consumption, mining profitability and fee-to-reward. These factors are significantly changed after the ETH's update of the consensus algorithm. Given such a precedence relationship between them, it appears that the transition from PoW to PoS does reflect on ETH log return. Development activity and development contributor designate how actively the blockchain repository is revised. Both factors have a higher lag order toward log return. It is a proof that ETH has a nature of high digitalization fueled by technology and algorithm advancement.

The linkage from coin age shows that the average holding days of a ETH is related to log return; while both block time and throughput considered as ETH's service level (Lin et al., 2021) have no connection to it. This suggests that ETH has not yet been considered a medium of exchange, since the change in ETH's return does not respond to how long a transaction is processed, but how long people hold it. Enlightening Ellingsen and Johannesson, 2009's study, investors demand no compensation for their
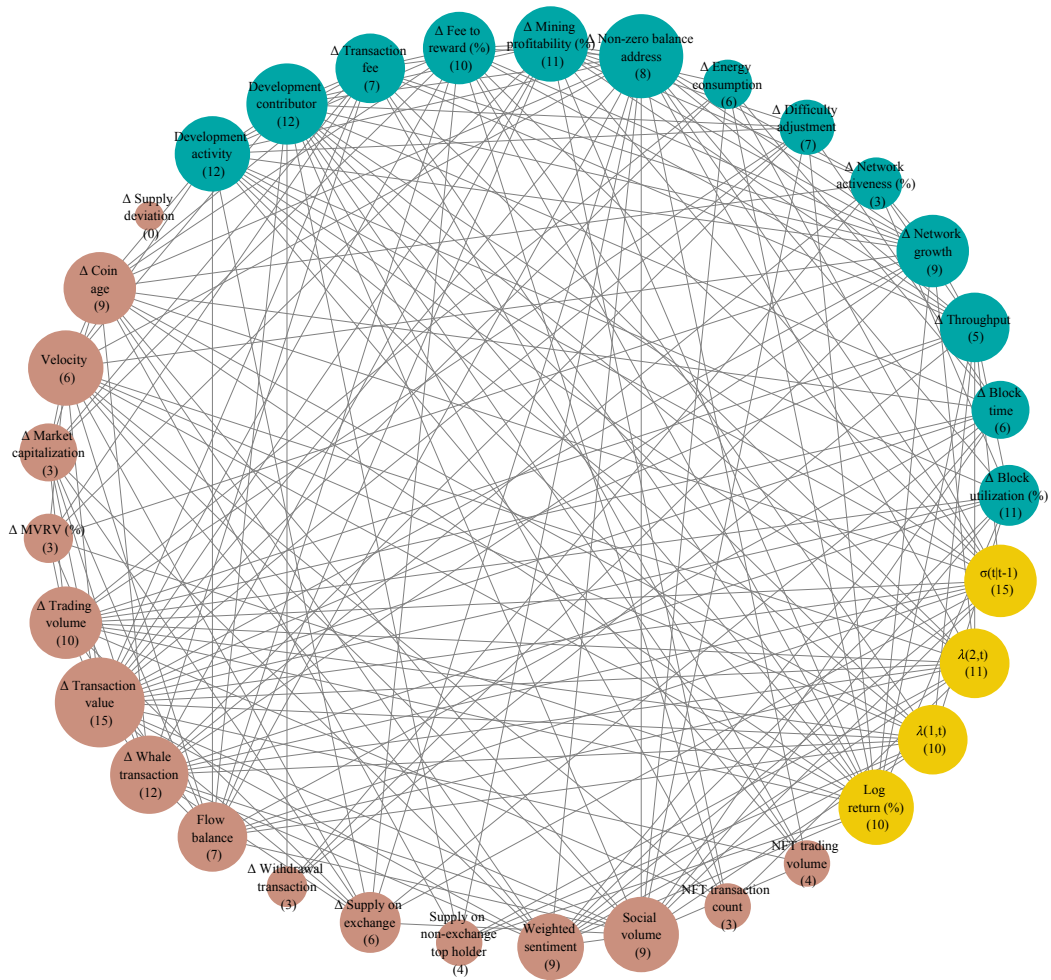
Figure 3.21: **Granger causal network.** Factors are colored according to the components they belong to – **functional characteristic**, **market dynamic**, and **economic attribute** The number in brackets represents the number of undirected connections between each factor and the other, e.g. block utilization (%) has 11 linkages with the other factors.

Table 3.1: **In-degree and out-degree**.

Factor	In-degree	Out-degree
Δ Block utilization (%)	9	6
Δ Block time	0	6
Δ Throughput	0	5
Δ Network growth	0	9
Δ Network activeness (%)	2	2
Δ Difficulty adjustment	1	6
Δ Energy consumption	1	5
Δ Non-zero balance address	0	8
Δ Mining profitability (%)	4	8
Δ Fee to reward (%)	3	7
Δ Transaction fee	2	5
Development contributor	4	8
Development activity	4	8
Δ Supply deviation	0	0
Δ Coin age	1	8
Velocity	2	4
Δ Market capitalization	2	1
Δ MVRV (%)	2	1
Δ Trading volume	5	6
Δ Transaction value	11	5
Δ Whale transaction	10	11
Flow balance	3	0
Δ Withdrawal transaction	3	3
Δ Supply on exchange	5	1
Supply on non-exchange top holder	1	3
Weighted sentiment	7	2
Social volume	6	3
NFT transaction count	2	1
NFT trading volume	4	0
Log return (%)	9	1
$\lambda_{1,t}$	10	0
$\lambda_{2,t}$	11	0
$\sigma^2_{t\mid t-1}$	10	6

time investments, whereas almost all subjects require compensation for equally costly monetary investments.

Supply on non-exchange top holder is the aggregate ETH's possessed by the top 10 highest balance addresses, known as whales. The transaction behaviors of these whales induce a change in log return, but supply on exchanges, where many small amount ETH holders transact, is not causally linked. The rich may have a significant influence on ETH's economy, which implies that there is an inequality in power.

Weighted sentiment has a causal link to log return. However, from Table 3.D.4 weighted sentiment has inflows from block time, non-zero balance address, coin age, velocity, transaction value, whale transaction, and $\sigma_{t|t-1}$ and outflows to block utilization and log return. The inflows to weighted sentiment suggest that there are factors that precede the sentiment happening. The outflows confirm that today's research into the connection between sentiment and crypto volatility is well-founded. A similar phenomenon is also found in social volume, as of Table 3.D.4. Last, users on ETH's network consider monetary gains before broadcasting transactions, as log return is directed to block utilization.

Long and short volatility components. The Beta-t-EGARCH model is employed to decompose volatility into two components - the long-run component, commonly associated with structural changes in the market, and the short-run component, which is linked to market noise. From a causation perspective, the two components have no outgoing connections. Both have no linkage to log return. Figure 3.23 demonstrates the incoming connections for the two components. Similar to ETH's log return, development contributor and development activity show a causal relationship to them. Interestingly, the long-run volatility component is directed by block utilization, whereas log return directs block utilization. Throughput change causally links to both long-run and short-run components. One possible explanation for this is herding and feedback trading, which starts from the short-run and then spreads to the long-run.

Reward-to-fee relationship is related to the long-run, which suggests that the change in rewards from transaction fees has a substantial impact on ETH, analog to the result of log return. However, long-run volatility component has no linkage from the factors – difficulty adjustment, energy consumption, and mining profitability. Velocity, which can be seen as how often a coin is traded between addresses in a day (i.e. turnover rate), has a connection to both components. Holden and Malani, 2022 argue that high velocity relates to an effective supply of coins, and that to increase velocity requires a reduction in blockchain operating costs, i.e. PoW to PoS.

Coin age directly connects to long-run as well as log return, which reflects that investors'
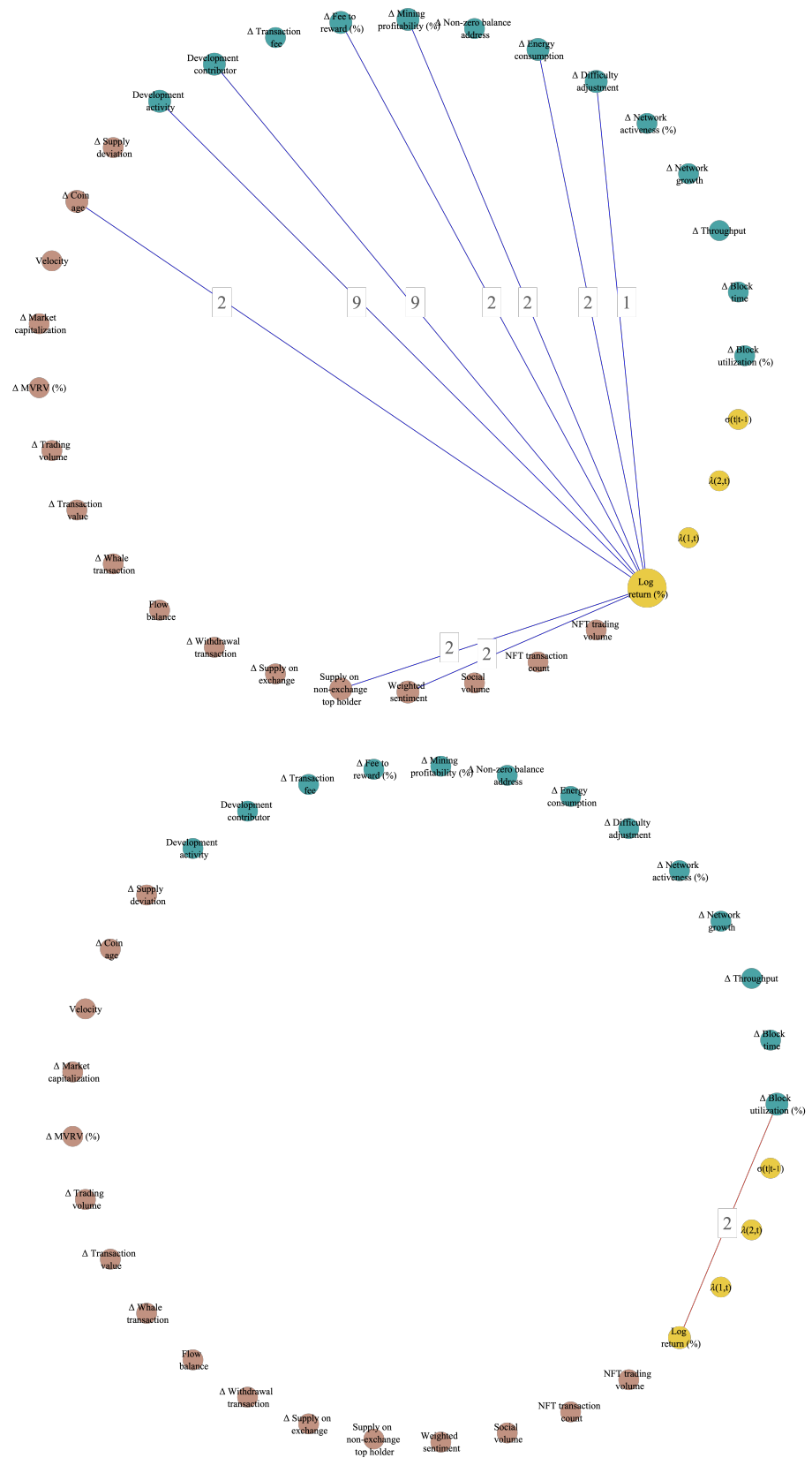
Figure 3.22: **In-degree** and **out-degree** of log return (%).

long and short positions influence ETH's economic performance persistently. Supply on non-exchange top holder and flow balance within exchanges show a connection to both long-run and short-run components. Flow balance, in particular, represents the investors' tendency toward buying or selling, which then directly impacts coin supply and demand. Traditionally, researchers associates trading volume to speculation (Tauchen & Pitts, 1983). It only demonstrates a causal link to the short-run volatility component, but not to the long-run and log return. Instead of weighted sentiment, social volume is connected to the long-run. Regarding applicability, only NFT transaction count directs to the short-run. In conclusion, both components appear to resemble the result of log return, and also provide a different resolution for describing causal connections.

Volatility. The result of volatility resembles the previous three cases discussed, in particular in its in-degree plot, see Figure 3.24. Volatility has a bidirectional relationship with two factors: mining profitability and fee-to-reward. Weighted sentiment outflows from it, while social volume inflows to it. Moreover, volatility directs to trading volume, trading value, and whale transaction. Two of the three factors are associated with one another, such that both trading volume and transaction value are induced by whale transactions. The whales at ETH's network react to the volatility and may not cause it. Instead, we see that flow balance and supply on exchange have in-degree connections with it.

3.8 Conclusion

Stiglitz, 1991 notes that "*The reason that the invisible hand often seems invisible is that it is often not there.*"

The argument highlights the concern regarding a self-regulated market where the equilibrium between supply and demand can be maintained while everyone is pursuing their interests. Jumping back to the cryptocurrency space, we have witnessed a functioning unregulated market for digital assets fueled by a cryptographic and algorithmic mechanism. Meanwhile, there is a community of people who recognize its value and democratically improve its applicability. Crypto's prices that commonly exhibit significant volatility and deviate substantially from the norms of the traditional money market can be challenging to rationalize or embrace. In this paper, we present an empirical framework for investigating the relationship among the three components of a crypto – economic attributes, functional characteristics, and market dynamics. We use Ethereum as a case study to elucidate the factors defined under these components and scrutinize their interplay and influence on one another through the application of Granger causality testing. Given the pairwise test results, we construct a Granger
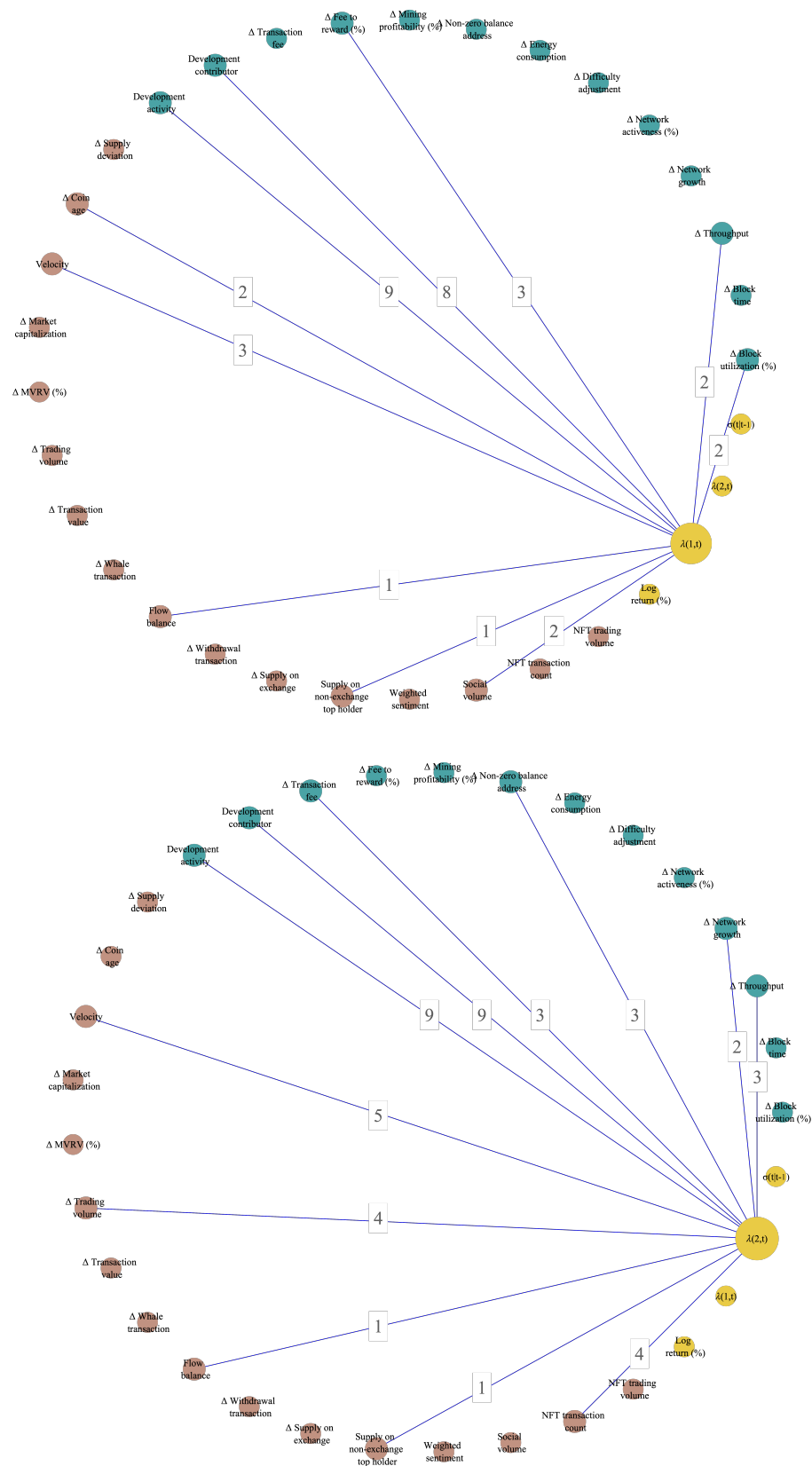
Figure 3.23: **In-degree** of long-run (upper) and short-run (lower) volatility components. There is no out-degree for both.
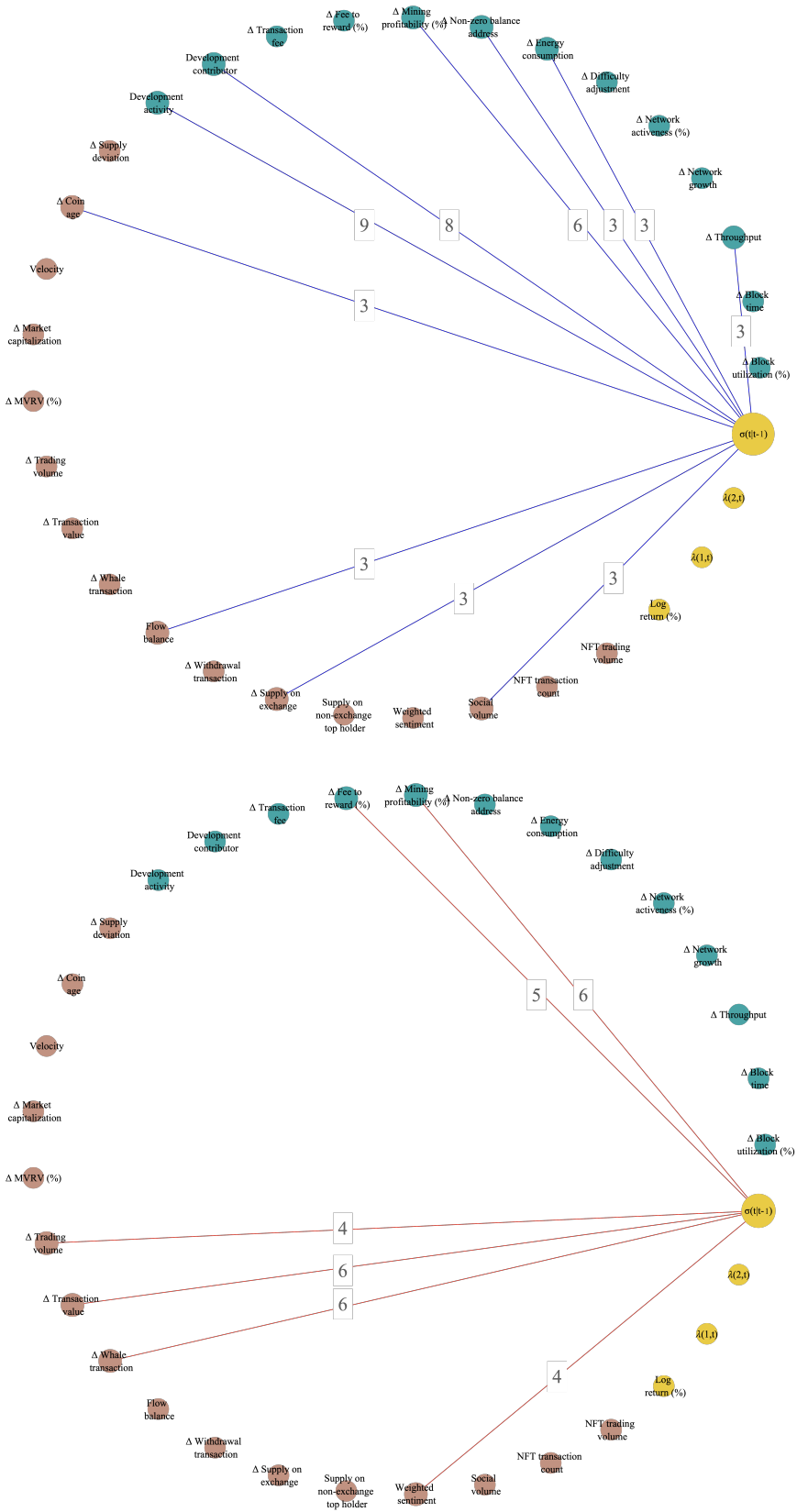
Figure 3.24: **In-degree** and **out-degree** of volatility.

causal network to better visualize their causal relationships.

Specifically, we use a Beta-t-EGARCH model to derive long-run and short-run volatility components, as well as volatility (conditional standard deviation) from daily log returns. These factors are considered as the economic attributes. Upon closer examination of ETH's economic attributes, we find that ETH's transition from PoW to PoS (i.e. energy consumption) influence on them. ETH has not yet been considered as a medium of exchange as users still demand no compensation for their time investment, given no connection from throughput and block time. Instead, our findings suggest that the economic attributes of ETH are linked to coin age, which implies that an individual's long and short positions are influential. Additionally, the concentration of wealth in the hands of large ETH holders, or whales, may have a disproportionate impact on ETH's performance, as evidenced by the links from whale transactions and the supply held by non-exchange top holders to economic attributes. In essence, changes in economic attributes and market dynamics often stem from alterations in functional characteristics. The results of long-run and short-run volatility components and volatility do resemble the one of log return. However, the analysis of these volatility components offers a greater potential for explaining the causal relationships among these factors in a better resolution.

The invisible hand appears to be at work in blockchain mechanism, regulating an otherwise unregulated market. Although often unseen, its influence is active and reflected.

Bibliography

Abadi, J., & Brunnermeier, M. (2018). Blockchain economics.

Adrian, T., & Rosenberg, J. (2008). Stock returns and volatility: Pricing the short-run and long-run components of market risk. *The journal of Finance*, *63*(6), 2997–3030.

Aloosh, A., & Ouzan, S. (2020). The psychology of cryptocurrency prices. *Finance Research Letters*, *33*, 101192.

Andersen, T. G., Cebiroglu, G., & Hautsch, N. (2017). Volatility, information feedback and market microstructure noise: A tale of two regimes.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of economic perspectives*, *21*(2), 129–152.

Bamakan, S. M. H., Motavali, A., & Bondarti, A. B. (2020). A survey of blockchain consensus algorithms performance evaluation criteria. *Expert Systems with Applications*, *154*, 113385.

Barbereau, T., Smethurst, R., Papageorgiou, O., Rieger, A., & Fridgen, G. (2022). Defi, not so decentralized: The measured distribution of voting rights. *Proceedings of the 55th Hawaii International Conference on System Sciences*.

Bier, J. (2021). The blocksize war: The battle over who controls bitcoin's protocol rules. *Independently Published*.

Blazsek, S., & Licht, A. (2020). Dynamic conditional score models: A review of their applications. *Applied Economics*, *52*(11), 1181–1199.

Bouoiyour, J., Selmi, R., & Tiwari, A. K. (2015). Is bitcoin business income or speculative foolery? new ideas through an improved frequency domain analysis. *Annals of Financial Economics*, *10*(01), 1550002.

Bouri, E., Kristoufek, L., Ahmad, T., & Shahzad, S. J. H. (2022). Microstructure noise and idiosyncratic volatility anomalies in cryptocurrencies. *Annals of Operations Research*, 1–27.

Buterin, V. (2021). Why sharding is great: Demystifying the technical properties. https://vitalik.ca/general/2021/04/07/sharding.html

Cahyadi, F. A., Owen, A. I., Ricardo, F., & Gunawan, A. A. (2021). Blockchain technology behind cryptocurrency and bitcoin for commercial transactions. *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*, *1*, 115–119.

Chen, C. Y.-H., & Hafner, C. M. (2019). Sentiment-induced bubbles in the cryptocurrency market. *Journal of Risk and Financial Management*, *12*(2), 53.

Chu, S., & Wang, S. (2018). The curses of blockchain decentralization. *arXiv preprint arXiv:1810.02937*.

Chuen, D. L. K., Guo, L., & Wang, Y. (2017). Cryptocurrency: A new investment opportunity? *The Journal of Alternative Investments*, *20*(3), 16–40.

Cong, L. W., Li, X., Tang, K., & Yang, Y. (2022). Crypto wash trading.

Conrad, C., Custovic, A., & Ghysels, E. (2018). Long-and short-term cryptocurrency volatility components: A garch-midas analysis. *Journal of Risk and Financial Management*, *11*(2), 23.

De Vries, A. (2022). Cryptocurrencies on the road to sustainability: Ethereum paving the way for bitcoin. *Patterns*, 100633.

Digiconomist. (2021). Ethereum energy consumption index.

Dimpfl, T., & Peter, F. J. (2021). Nothing but noise? price discovery across cryptocurrency exchanges. *Journal of Financial Markets*, *54*, 100584.

Dutta, A., & Bouri, E. (2022). Outliers and time-varying jumps in the cryptocurrency markets. *Journal of Risk and Financial Management*, *15*(3), 128.

Ellingsen, T., & Johannesson, M. (2009). Time is not money. *Journal of Economic Behavior & Organization*, *72*(1), 96–102.

Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews*, *5*(1), 1–50.

Engle, R. F., & Lee, G. (1999). A long-run and short-run component model of stock return volatility. *Cointegration, causality, and forecasting: A Festschrift in honour of Clive WJ Granger*, 475–497.

Fakhfekh, M., & Jeribi, A. (2020). Volatility dynamics of crypto-currencies' returns: Evidence from asymmetric and long memory garch models. *Research in International Business and Finance*, *51*, 101075.

Fang, T., Lee, T.-H., & Su, Z. (2020). Predicting the long-term stock market volatility: A garch-midas model with variable selection. *Journal of Empirical Finance*, *58*, 36–49.

Ferdous, M. S., Chowdhury, M. J. M., Hoque, M. A., & Colman, A. (2020). Blockchain consensus algorithms: A survey. *arXiv preprint arXiv:2001.07091*.

Garratt, R., & van Oordt, M. R. (2020). Why fixed costs matter for proof-of-work based cryptocurrencies. *Available at SSRN 3572400*.

Gochhayat, S. P., Shetty, S., Mukkamala, R., Foytik, P., Kamhoua, G. A., & Njilla, L. (2020). Measuring decentrality in blockchain based systems. *IEEE Access*, *8*, 178372–178390.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438.

Gredenhoff, M., & Karlsson, S. (1999). Lag-length selection in var-models using equal and unequal lag-length procedures. *Computational Statistics*, *14*, 171–187.

Guo, L., Härdle, W. K., & Tao, Y. (2022). A time-varying network for cryptocurrencies. *Journal of Business & Economic Statistics*, 1–20.

Gupta, M., & Gupta, P. (2018). Gini coefficient based wealth distribution in the bitcoin network: A case study. *Computing, Analytics and Networks: First International Conference, ICAN 2017, Chandigarh, India, October 27-28, 2017, Revised Selected Papers 1*, 192–202.

Gurdgiev, C., & O'Loughlin, D. (2020). Herding and anchoring in cryptocurrency markets: Investor reaction to fear and uncertainty. *Journal of Behavioral and Experimental Finance*, *25*, 100271.

Harvey, A., & Lange, R.-J. (2018). Modeling the interactions between volatility and returns using egarch-m. *Journal of Time Series Analysis*, *39*(6), 909–919.

Harvey, A., & Sucarrat, G. (2014). Egarch models with fat tails, skewness and leverage. *Computational Statistics & Data Analysis*, *76*, 320–338.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: With applications to financial and economic time series*. Cambridge University Press. https://doi.org/10.1017/CBO9781139540933

Holden, R., & Malani, A. (2022). An examination of velocity and initial coin offerings. *Management Science*, *68*(12), 9026–9041.

Hong, H., & Stein, J. C. (2003). Differences of opinion, short-sales constraints, and market crashes. *The Review of Financial Studies*, *16*(2), 487–525.

Imbens, G. W. (2022). Causality in econometrics: Choice vs chance. *Econometrica*, *90*(6), 2541–2566.

Jiang, S., Li, X., & Wang, S. (2021). Exploring evolution trends in cryptocurrency study: From underlying technology to economic applications. *Finance Research Letters*, *38*, 101532.

Katsiampa, P. (2017). Volatility estimation for bitcoin: A comparison of garch models. *Economics Letters*, *158*, 3–6.

Katsiampa, P., Corbet, S., & Lucey, B. (2019). High frequency volatility co-movements in cryptocurrency markets. *Journal of International Financial Markets, Institutions and Money*, *62*, 35–52.

Kiayias, A., & Panagiotakos, G. (2015). Speed-security tradeoffs in blockchain protocols. *Cryptology ePrint Archive*.

King, T., & Koutmos, D. (2021). Herding and feedback trading in cryptocurrency markets. *Annals of Operations Research*, *300*, 79–96.

Kyriazis, N., Papadamou, S., & Corbet, S. (2020). A systematic review of the bubble dynamics of cryptocurrency prices. *Research in International Business and Finance*, *54*, 101254.

Lee, C. M., & Swaminathan, B. (2000). Price momentum and trading volume. *the Journal of Finance*, *55*(5), 2017–2069.

Lessig, L. (2009). *Code: And other laws of cyberspace*. ReadHowYouWant. com.

Li, Y., Goodell, J. W., & Shen, D. (2021). Comparing search-engine and social-media attentions in finance research: Evidence from cryptocurrencies. *International Review of Economics & Finance*, *75*, 723–746.

Lin, M.-B., Khowaja, K., Chen, C. Y.-H., & Härdle, W. K. (2021). Blockchain mechanism and distributional characteristics of cryptos. *Lin MB, Khowaja K, Chen CYH, Härdle WK (2020) Blockchain mechanism and distributional characteristics of cryptos. Forthcoming in: Book Series: Advances in Quantitative Analysis of Finance & Accounting (AQAFA)*, *18*.

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

Lucchini, L., Alessandretti, L., Lepri, B., Gallo, A., & Baronchelli, A. (2020). From code to market: Network of developers and correlated returns of cryptocurrencies. *Science advances*, *6*(51), eabd2204.

Mai, F., Shan, Z., Bai, Q., Wang, X., & Chiang, R. H. (2018). How does social media impact bitcoin value? a test of the silent majority hypothesis. *Journal of management information systems*, *35*(1), 19–52.

Malik, N., Aseri, M., Singh, P. V., & Srinivasan, K. (2022). Why bitcoin will fail to scale? *Management Science*, *68*(10), 7323–7349.

Mark, M., Sila, J., & Weber, T. A. (2020). Quantifying endogeneity of cryptocurrency markets. *The European Journal of Finance*, 1–16.

McGurk, Z., Nowak, A., & Hall, J. C. (2020). Stock returns and investor sentiment: Textual analysis and social media. *Journal of Economics and Finance*, *44*(3), 458–485.

Naimy, V., Haddad, O., Fernández-Avilés, G., & El Khoury, R. (2021). The predictive capacity of garch-type models in measuring the volatility of crypto and world currencies. *PloS one*, *16*(1), e0245904.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Ng, S., & Perron, P. (2005). A note on the selection of time series models. *oxford Bulletin of Economics and statistics*, *67*(1), 115–134.

Pagnotta, E. S. (2022). Decentralizing money: Bitcoin prices and blockchain security. *The Review of Financial Studies*, *35*(2), 866–907.

Pele, D. T., Wesselhöfft, N., Härdle, W. K., Kolossiatis, M., & Yatracos, Y. G. (2021). Are cryptos becoming alternative assets? *The European Journal of Finance*, 1–42.

Ren, L. (2014). Proof of stake velocity: Building the social currency of the digital age. *Self-published white paper*.

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*, 101462.

Shahzad, S. J. H., Bouri, E., Roubaud, D., Kristoufek, L., & Lucey, B. (2019). Is bitcoin a better safe-haven investment than gold and commodities? *International Review of Financial Analysis*, *63*, 322–330.

Shen, M., & Garg, A. (2022). 2022 Electric Capital Developer Report. https://github.com/electric-capital/developer-reports/blob/master/dev_report_2022.pdf

Shojaie, A., & Fox, E. B. (2022). Granger causality: A review and recent advances. *Annual Review of Statistics and Its Application*, *9*, 289–319.

Stiglitz, J. E. (1991). The invisible hand and modern welfare economics.

Tauchen, G. E., & Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica: Journal of the Econometric Society*, 485–505.

Tiniç, M., Sensoy, A., Akyildirim, E., & Corbet, S. (2020). Adverse selection in cryptocurrency markets. *Journal of Financial Research*.

Walther, T., Klein, T., & Bouri, E. (2019). Exogenous drivers of bitcoin and cryptocurrency volatility–a mixed data sampling approach to forecasting. *Journal of International Financial Markets, Institutions and Money*, *63*, 101133.

Wang, Y., Lu, W., Lin, M.-B., Ren, R., & Härdle, W. K. (2022). Cross-exchange crypto risk: A high-frequency dynamic network perspective. *Available at SSRN 4308825*.

White, R., Marinakis, Y., Islam, N., & Walsh, S. (2020). Is bitcoin a currency, a technology-based product, or something else? *Technological Forecasting and Social Change*, *151*, 119877.

Yang, W., Aghasian, E., Garg, S., Herbert, D., Disiuta, L., & Kang, B. (2019). A survey on blockchain-based internet service architecture: Requirements, challenges, trends, and future. *IEEE Access*, *7*, 75845–75872.

Youssef, M. (2022). What drives herding behavior in the cryptocurrency market? *Journal of Behavioral Finance*, *23*(2), 230–239.

Yuan, K. (2005). Asymmetric price movements and borrowing constraints: A rational expectations equilibrium model of crises, contagion, and confusion. *The Journal of Finance*, *60*(1), 379–411.

Zhang, R., Xue, R., & Liu, L. (2019). Security and privacy on blockchain. *ACM Computing Surveys (CSUR)*, *52*(3), 1–34.

Zhang, W., Wang, P., Li, X., & Shen, D. (2018). Some stylized facts of the cryptocurrency market. *Applied Economics*, *50*(55), 5950–5965.

Zimmerman, P. (2020). Blockchain structure and cryptocurrency prices. *Bank of England Working Paper*.

Appendix

3.A   Time series & ACF

The vertical dotted line on each figure refers to the transition date for ETH to upgrade the consensus mechanism from Proof-of-Work (PoW) to Proof-of-Stack (PoS) on September 15, 2022 – also known as the Merge (EIP-3675). The autocorrelation plot has a 95% confidence interval, as of grey area at the lower panel of each figure.

Functional characteristics
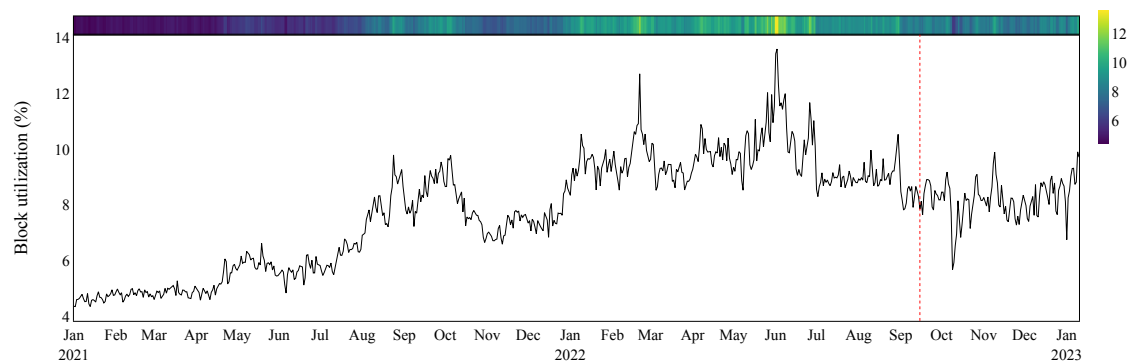
(a) Time series
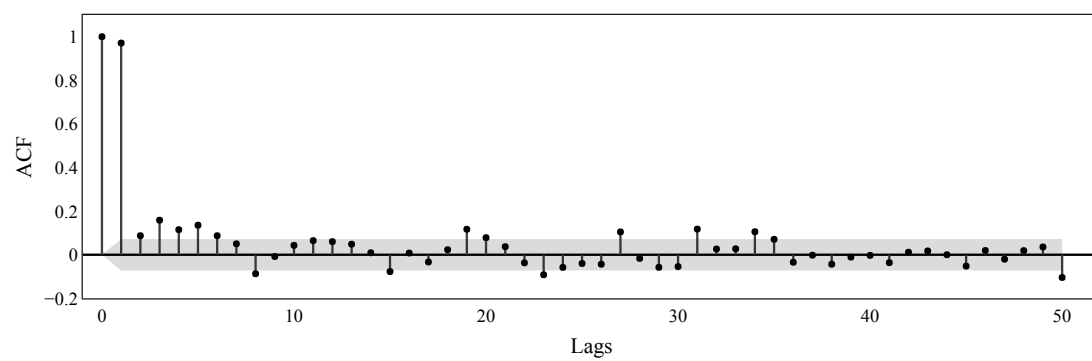
(b) Autocorrelation function

Figure 3.A.1: **Block utilization**.
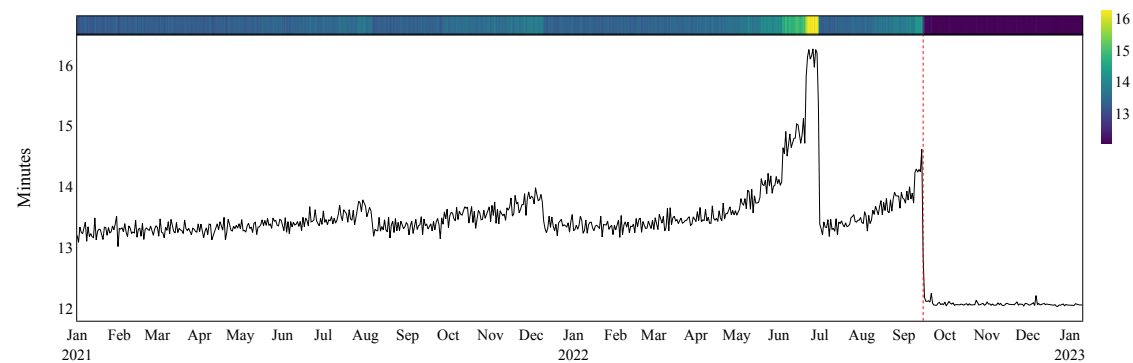
(a) Time series
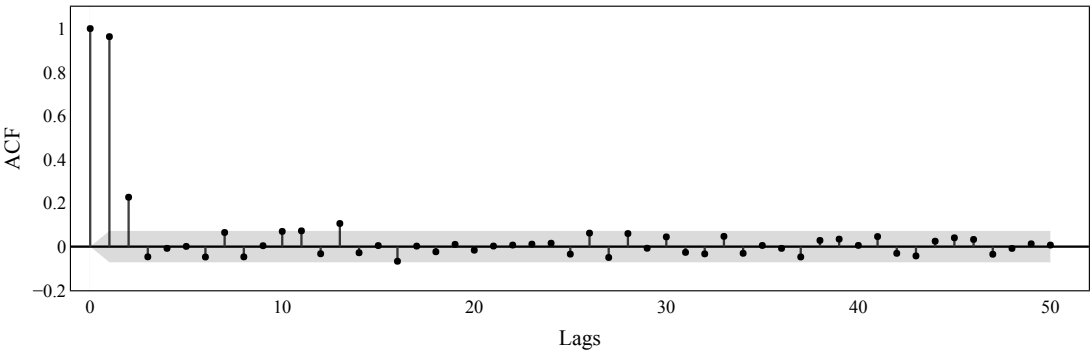
(b) Autocorrelation function

Figure 3.A.2: **Block time**.
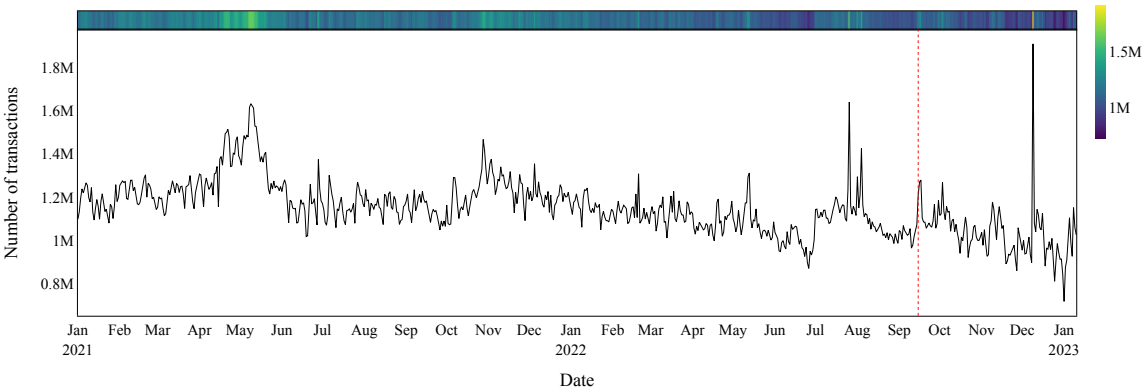
(a) Time series
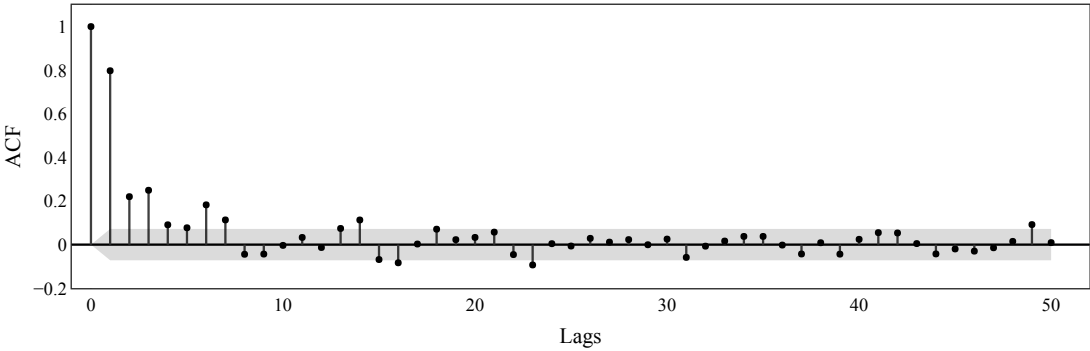
(b) Autocorrelation function

Figure 3.A.3: **Throughput**.
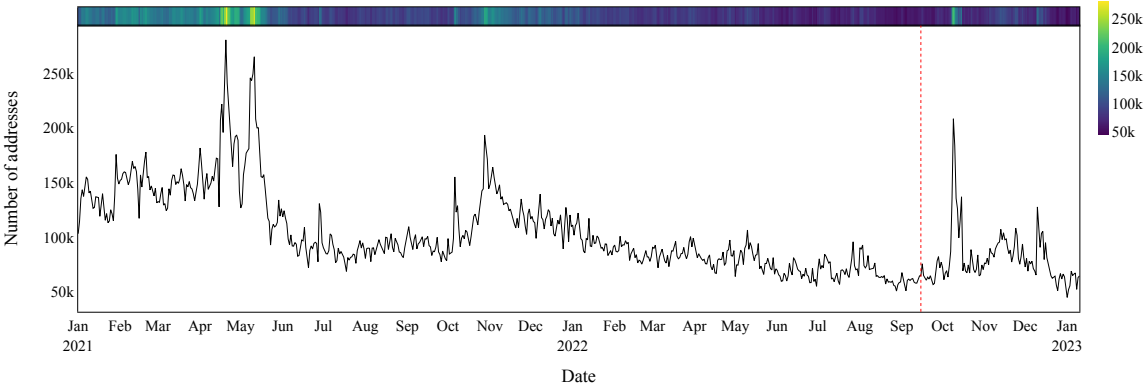
(a) Time series
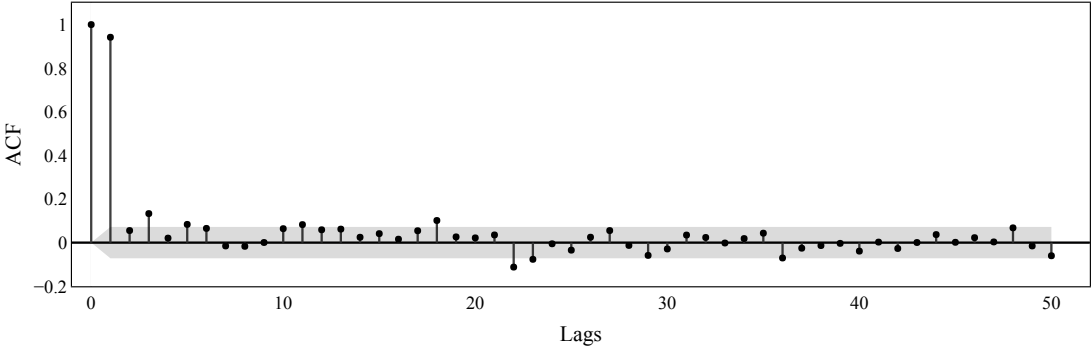
(b) Autocorrelation function

Figure 3.A.4: **Network growth**.
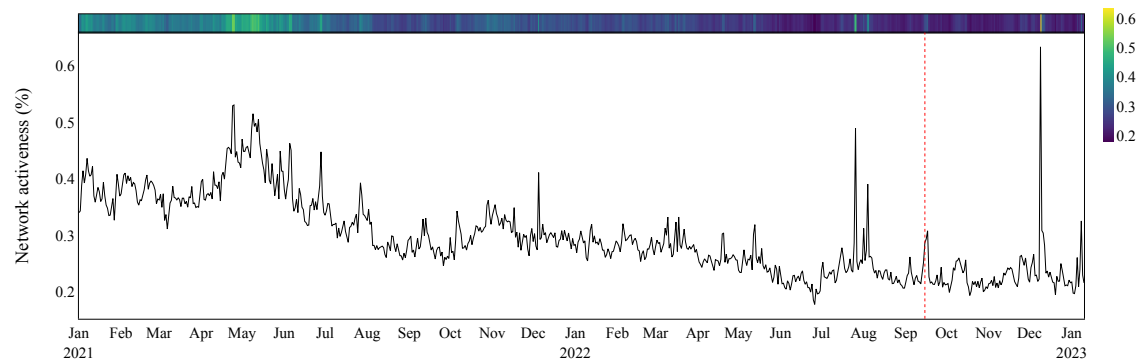
(a) Time series
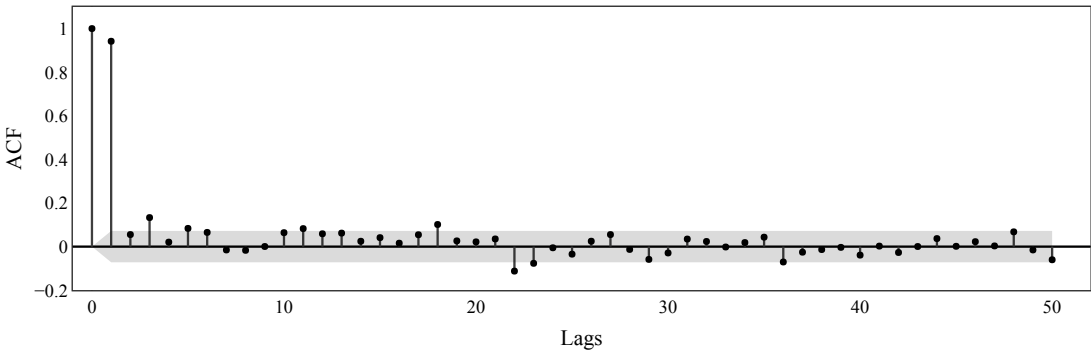
(b) Autocorrelation function

Figure 3.A.5: **Network activeness**.
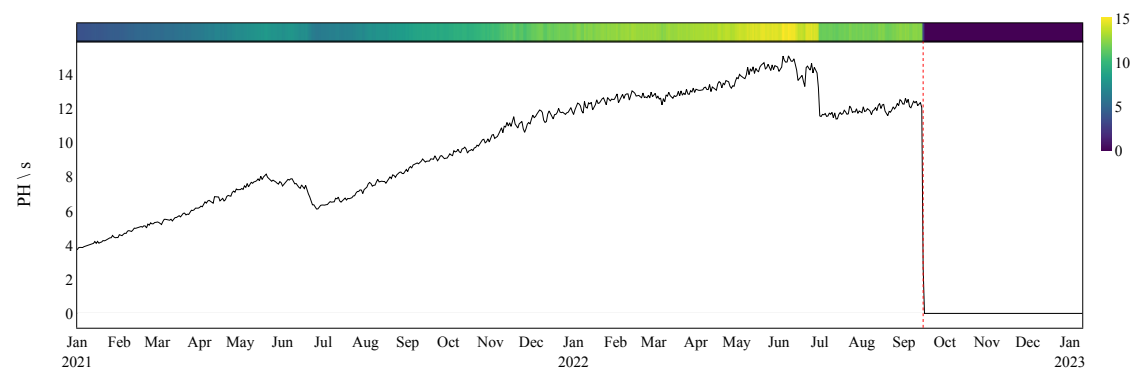
(a) Time series
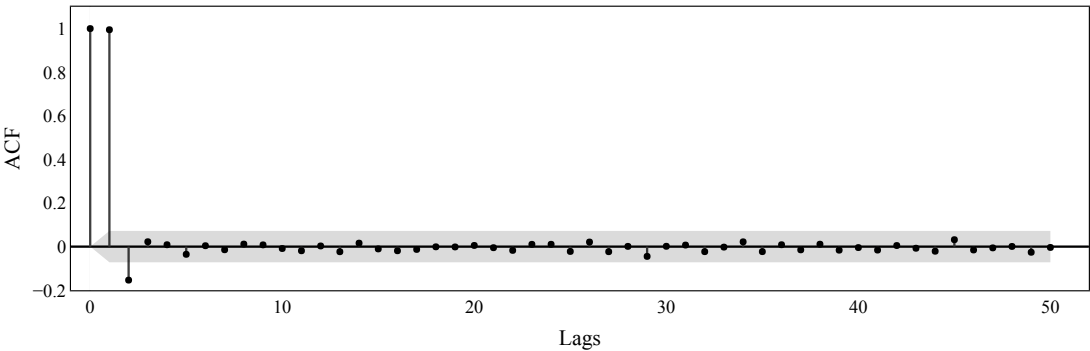
(b) Autocorrelation function

Figure 3.A.6: **Difficulty adjustment**.
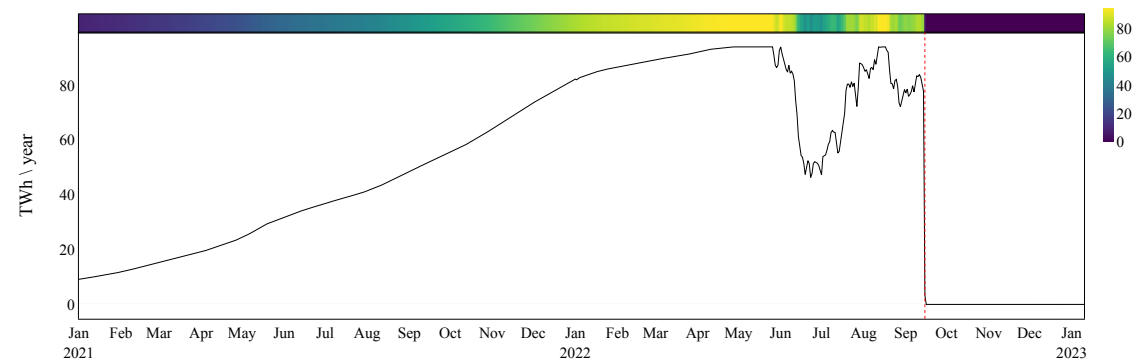
(a) Time series
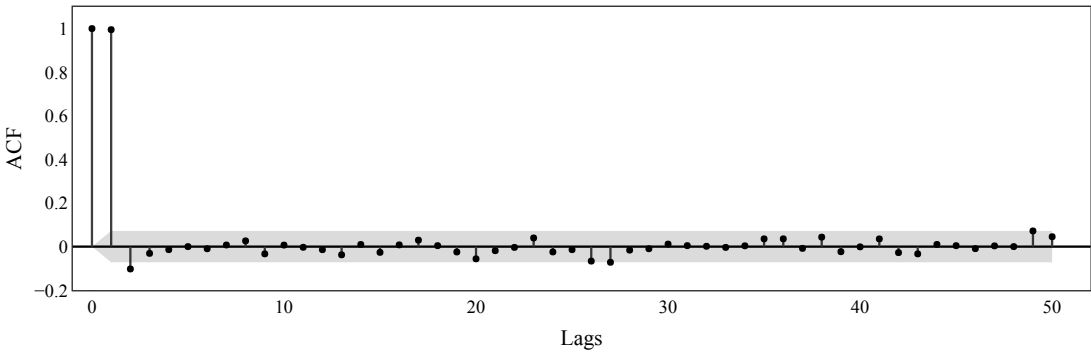
(b) Autocorrelation function

Figure 3.A.7: **Energy consumption**.
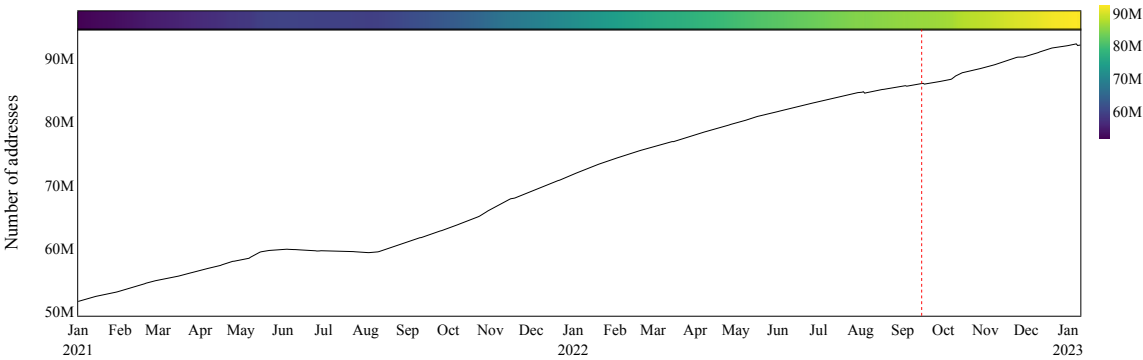
(a) Time series
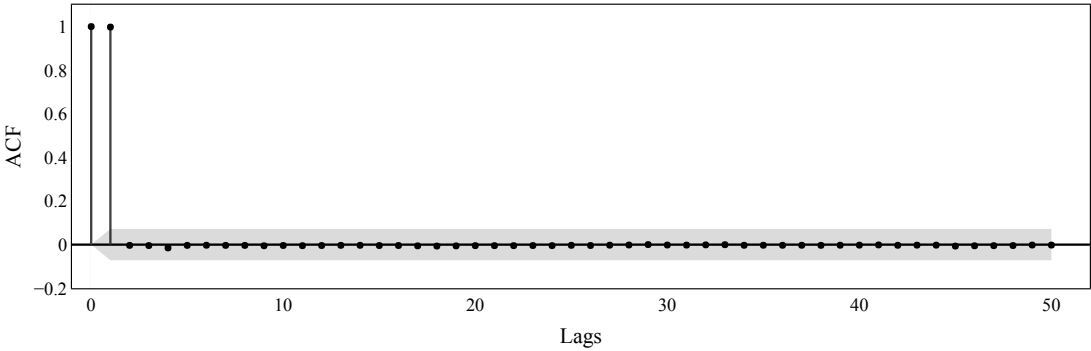
(b) Autocorrelation function

Figure 3.A.8: **Non-zero balance address**.
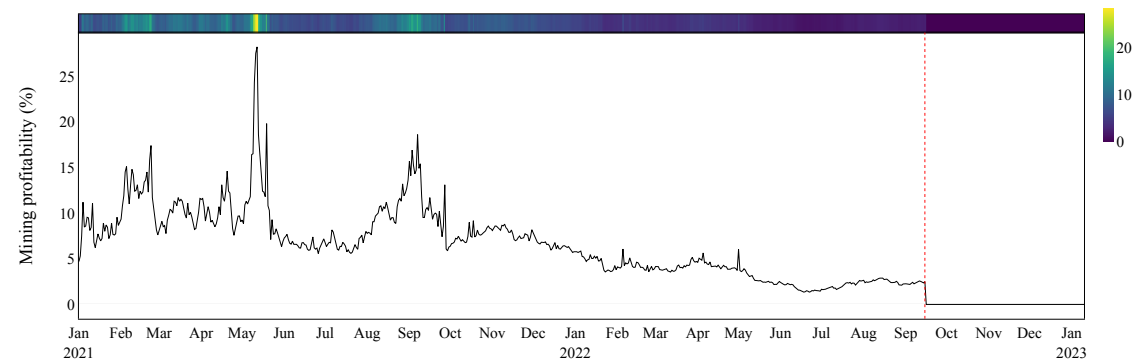
(a) Time series
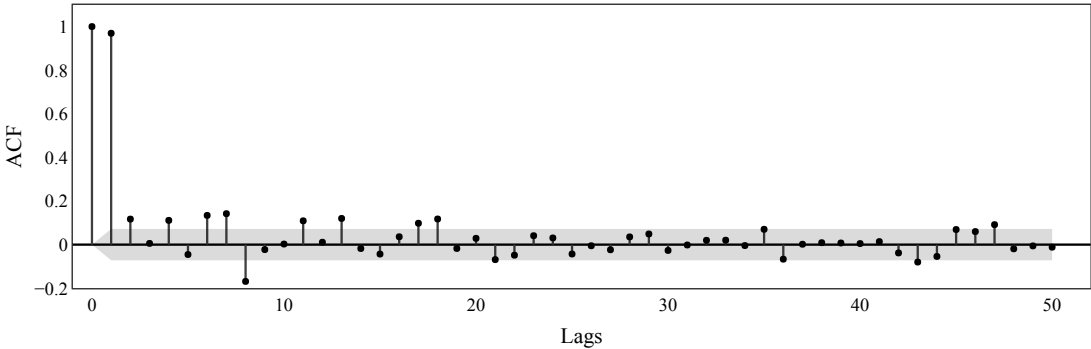
(b) Autocorrelation function

Figure 3.A.9: **Mining profitability**.
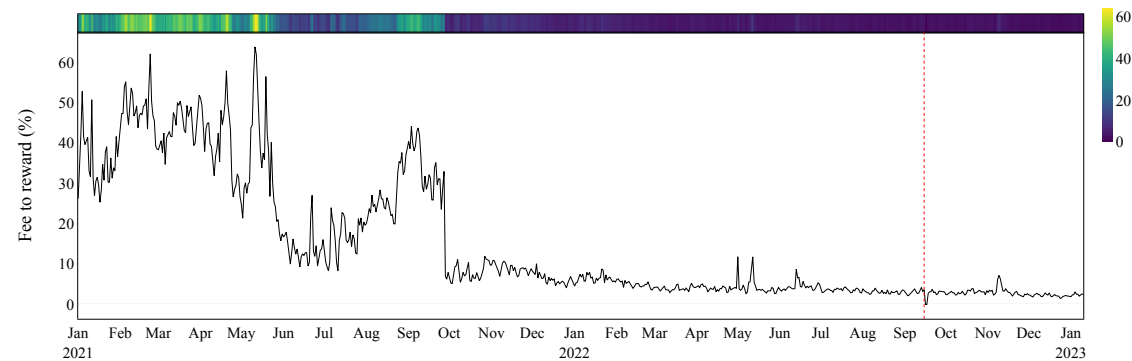
(a) Time series
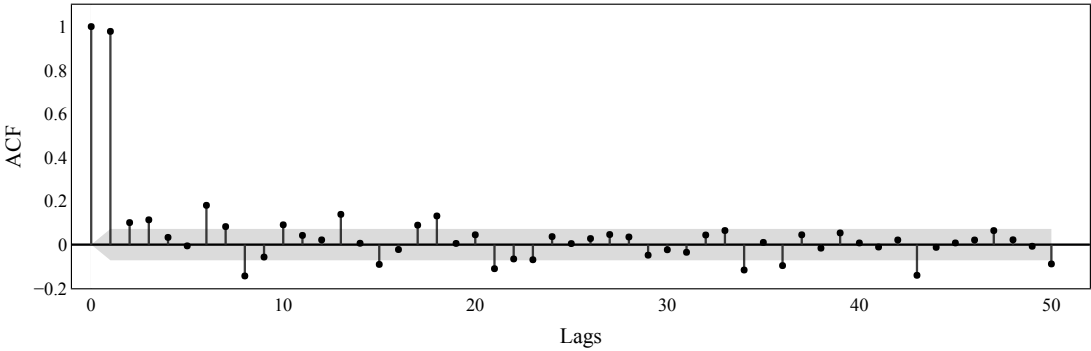
(b) Autocorrelation function

Figure 3.A.10: **Fee to reward**.
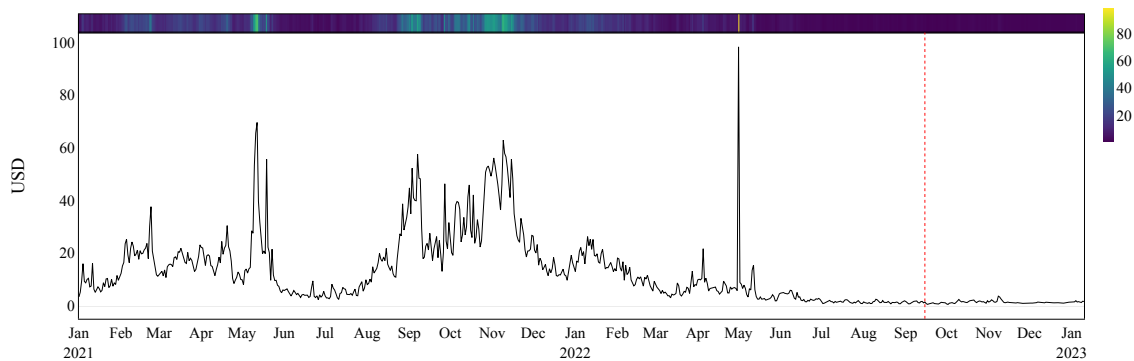
(a) Time series
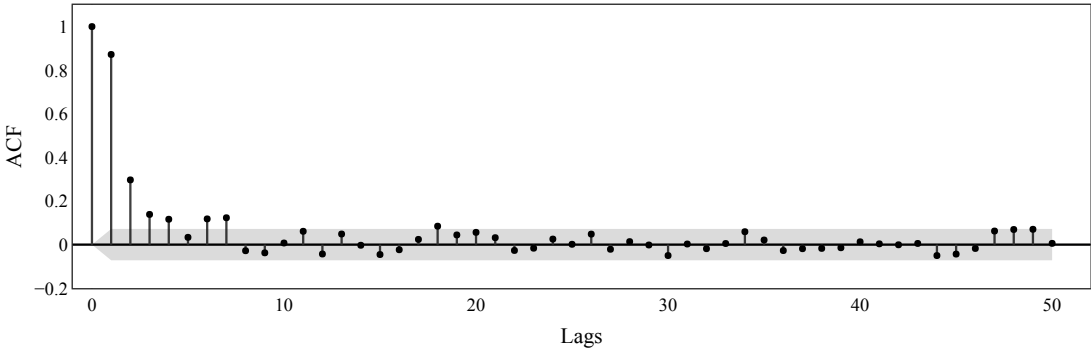
(b) Autocorrelation function

Figure 3.A.11: **Transaction fee**.
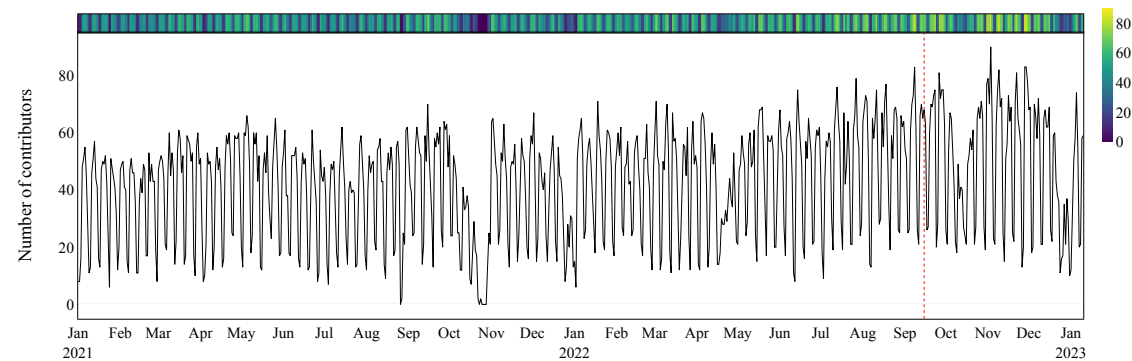
(a) Time series
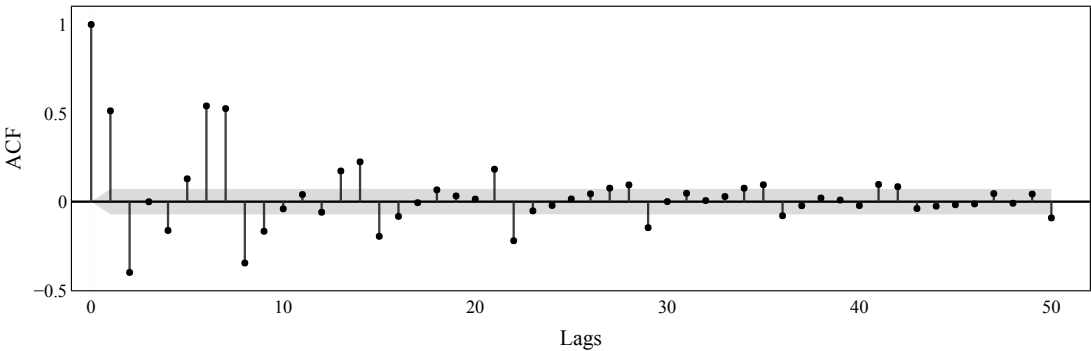
(b) Autocorrelation function

Figure 3.A.12: **Development contributor**.
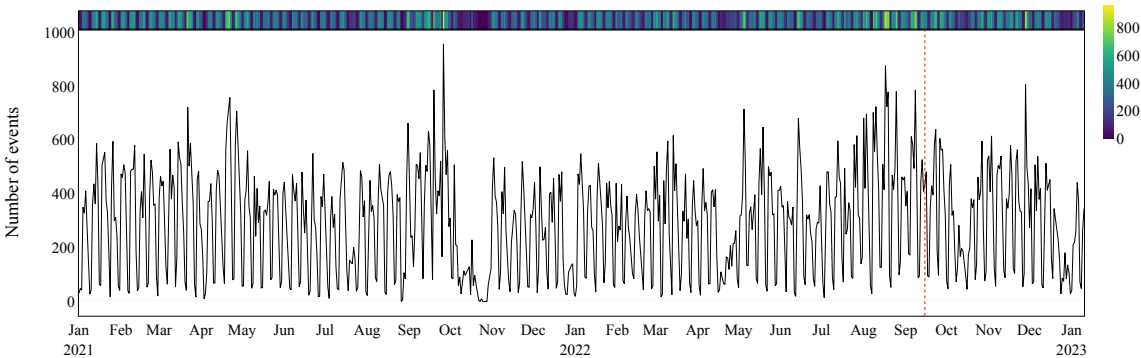
(a) Time series
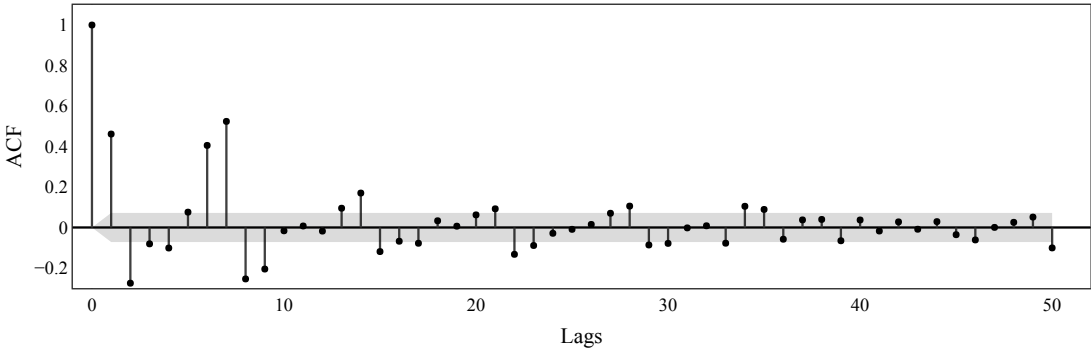
(b) Autocorrelation function

Figure 3.A.13: **Development activity**.

Market dynamics
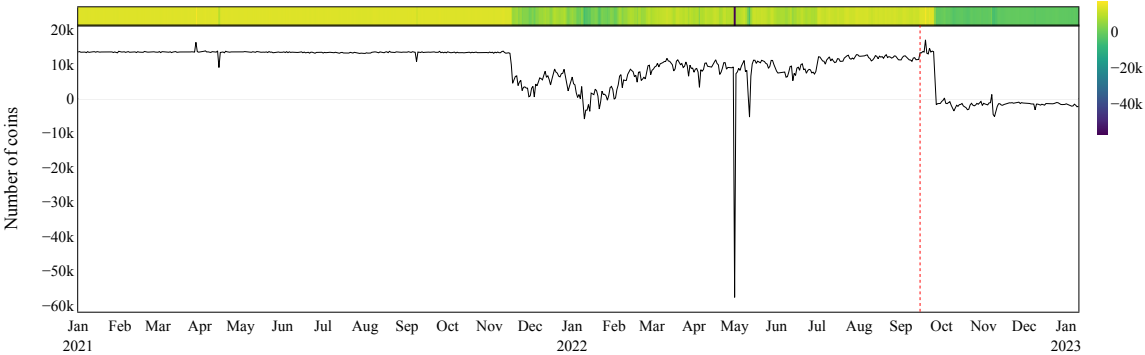
(a) Time series
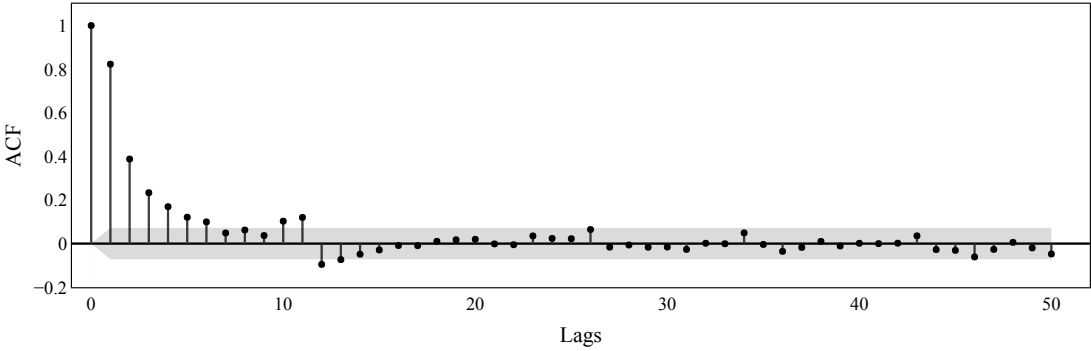
(b) Autocorrelation function

Figure 3.A.14: **Supply deviation**.
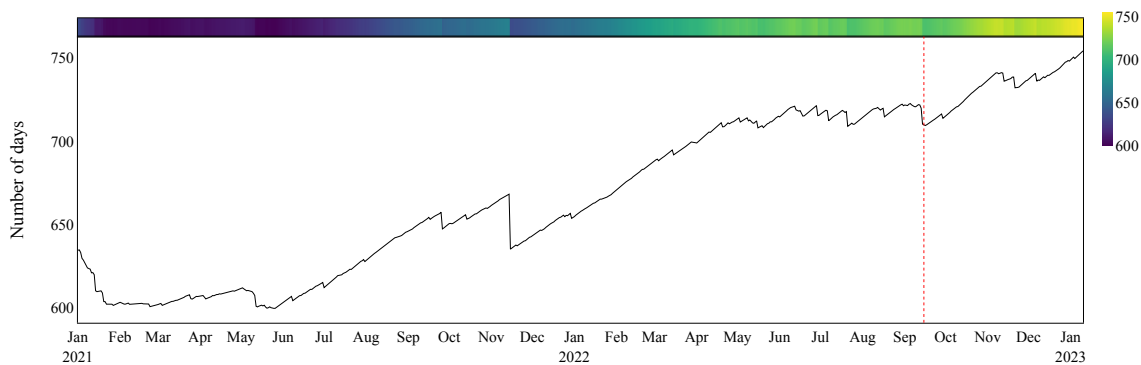
(a) Time series
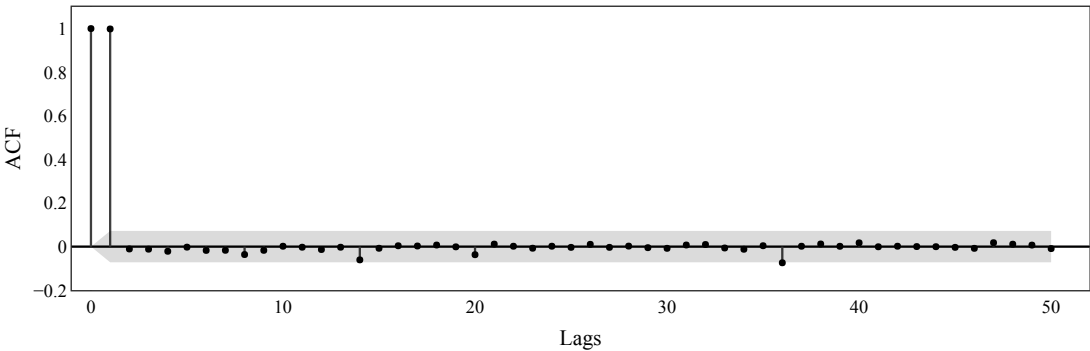
(b) Autocorrelation function

Figure 3.A.15: **Coin age**.
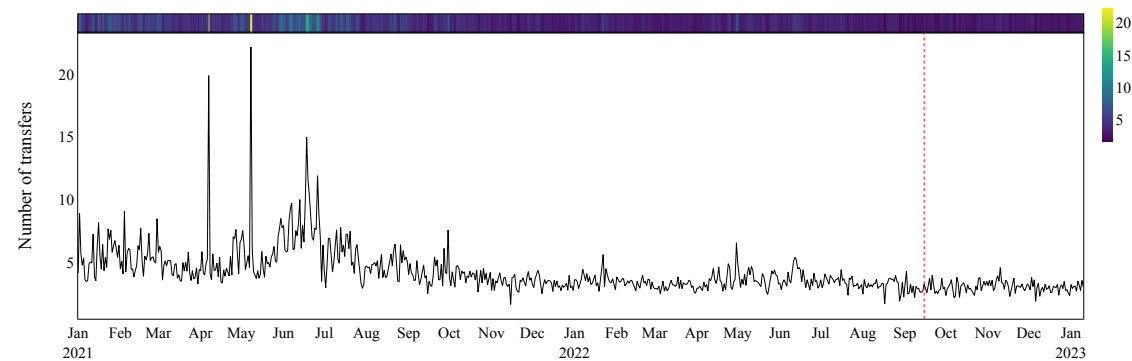
(a) Time series
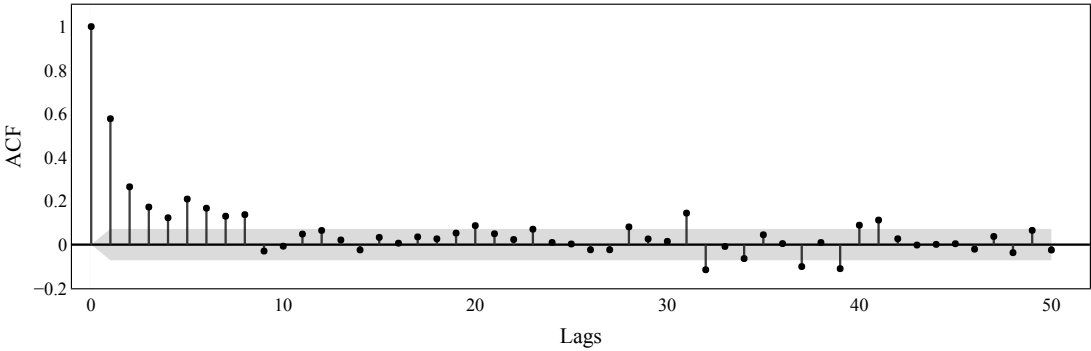
(b) Autocorrelation function

Figure 3.A.16: **Velocity**.

$=$
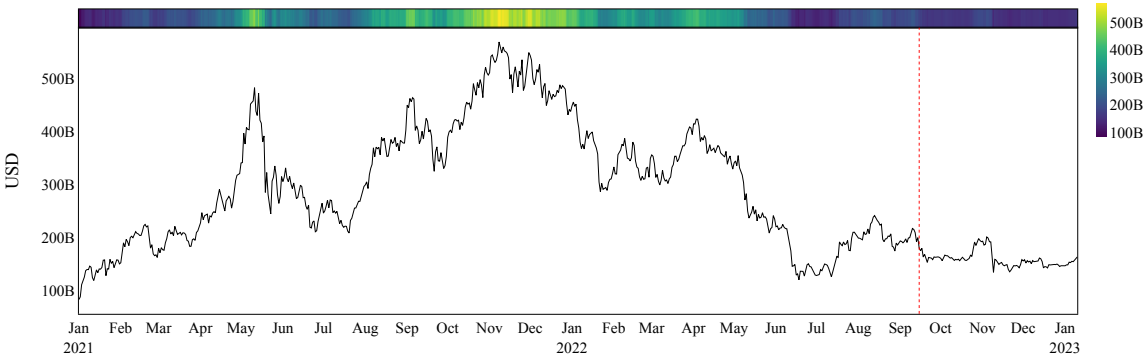
(a) Time series
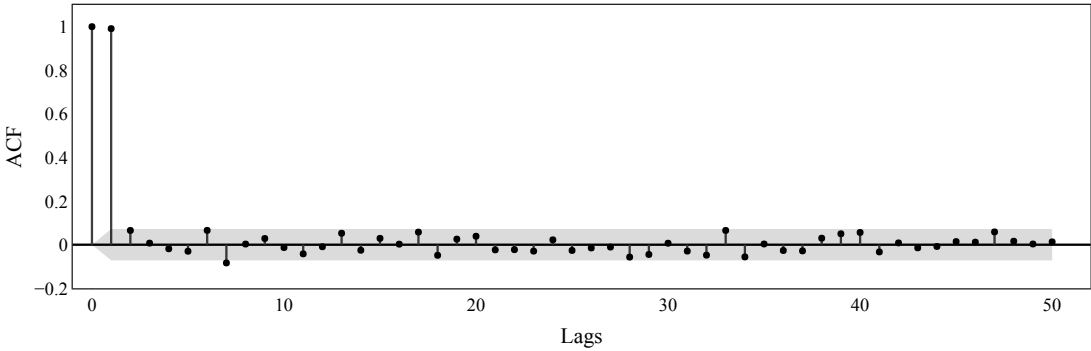
(b) Autocorrelation function

Figure 3.A.17: **Market capitalization**.
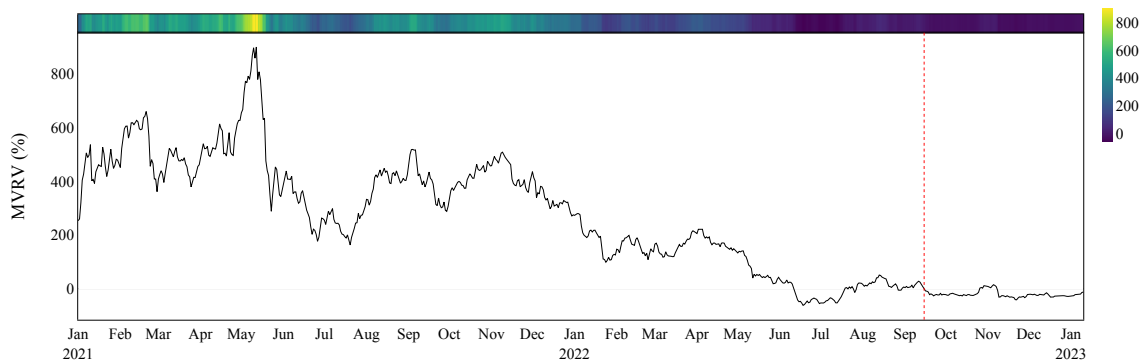
(a) Time series
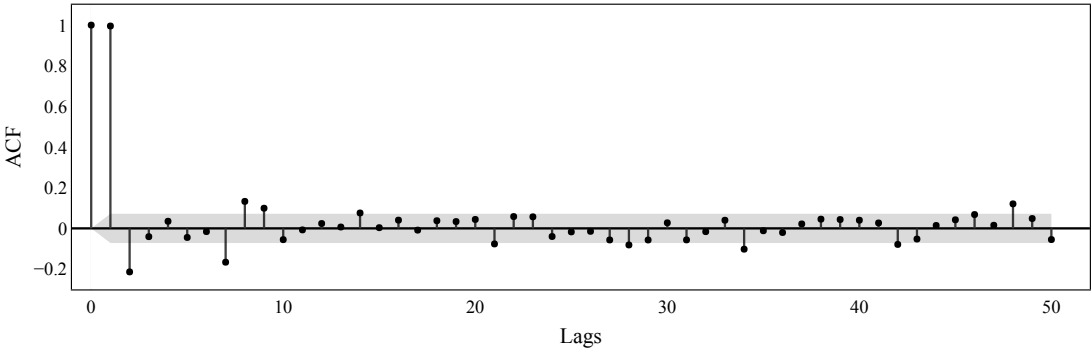
(b) Autocorrelation function

Figure 3.A.18: **MVRV**.
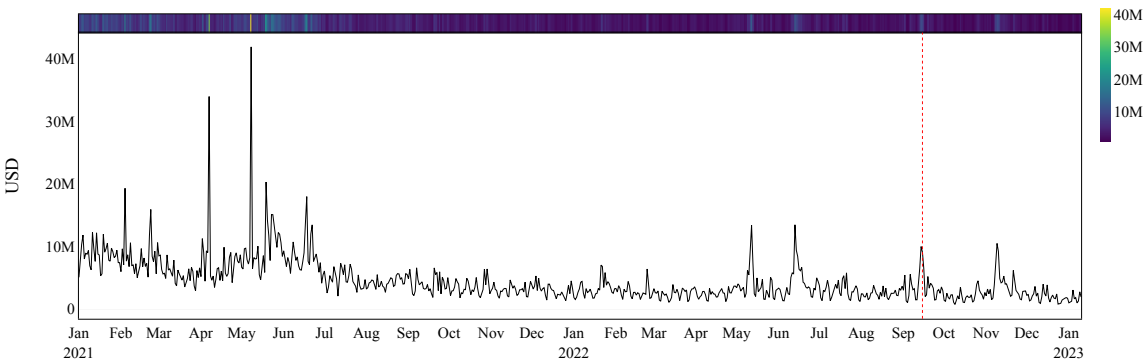
(a) Time series
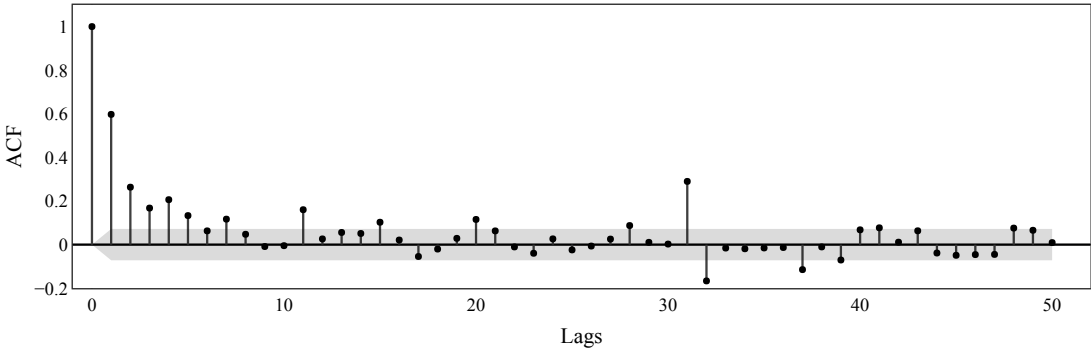
(b) Autocorrelation function

Figure 3.A.19: **Trading volume**.
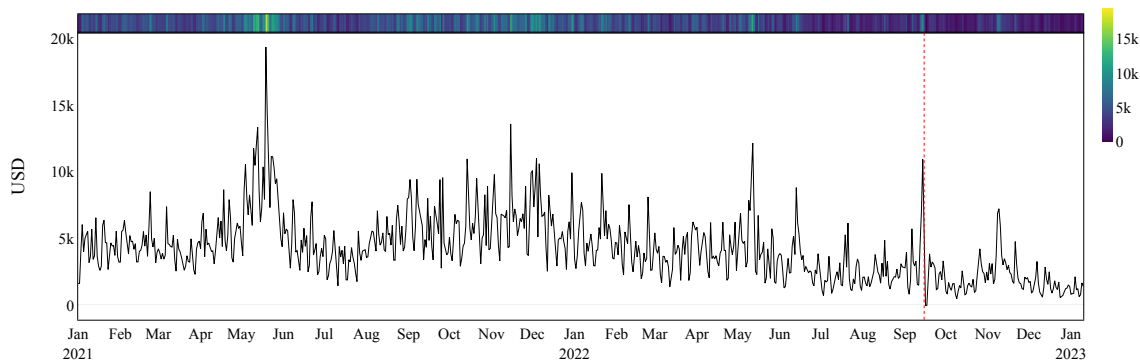
(a) Time series
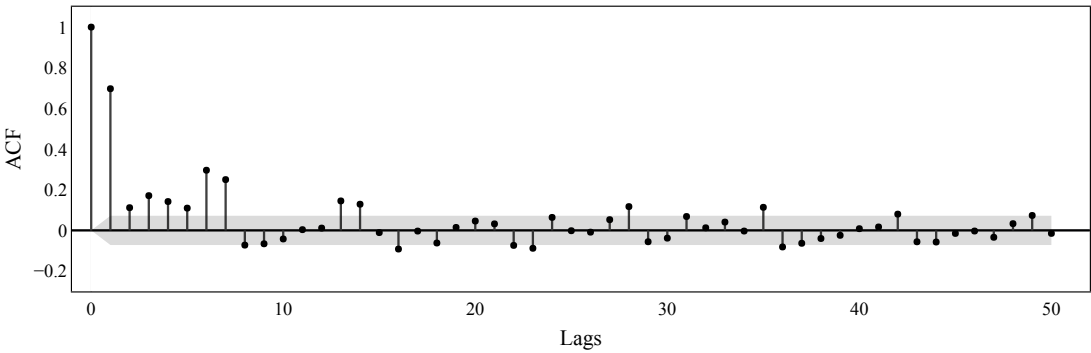
(b) Autocorrelation function

Figure 3.A.20: **Transaction value**.
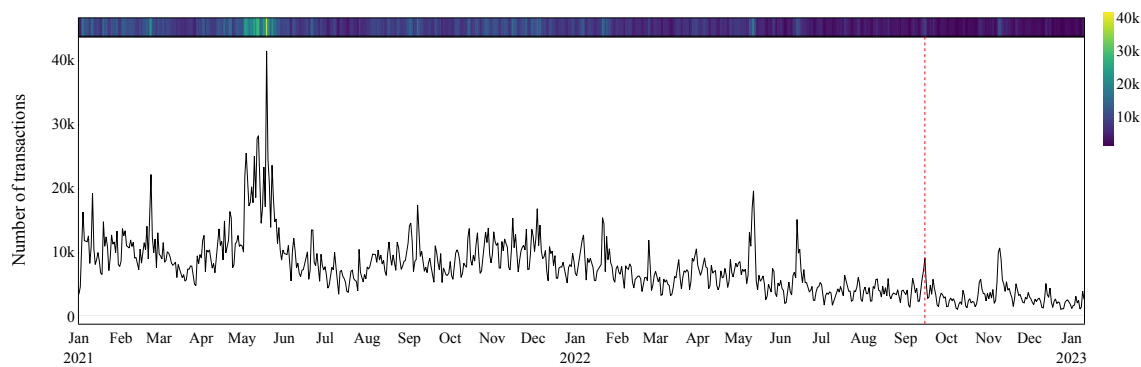
(a) Time series
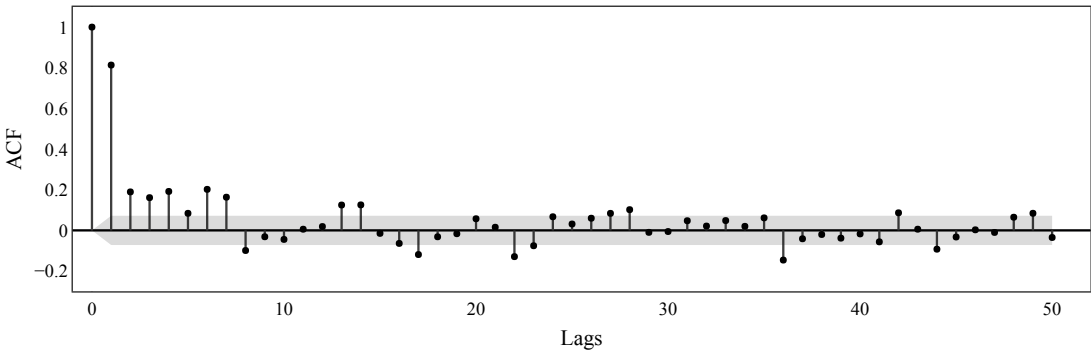
(b) Autocorrelation function

Figure 3.A.21: **Whale transaction**.
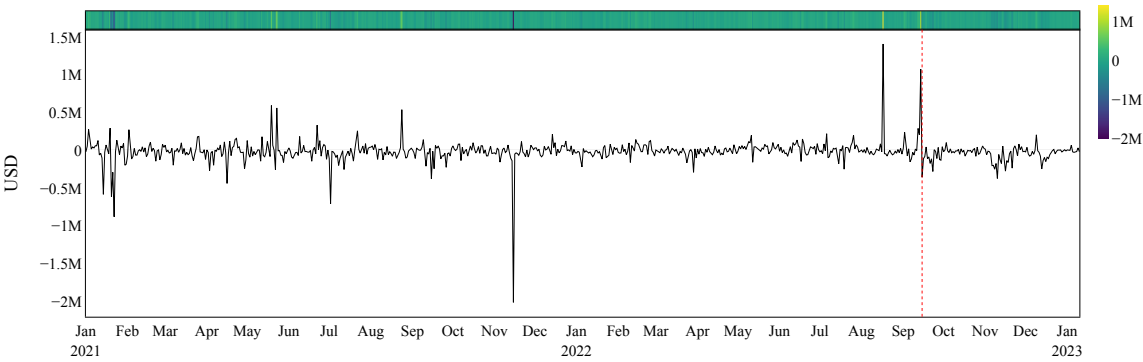
(a) Time series
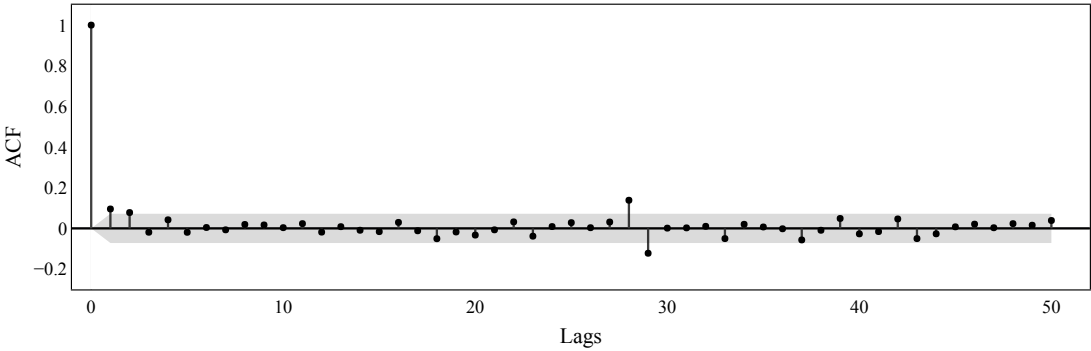
(b) Autocorrelation function

Figure 3.A.22: **Flow balance**.
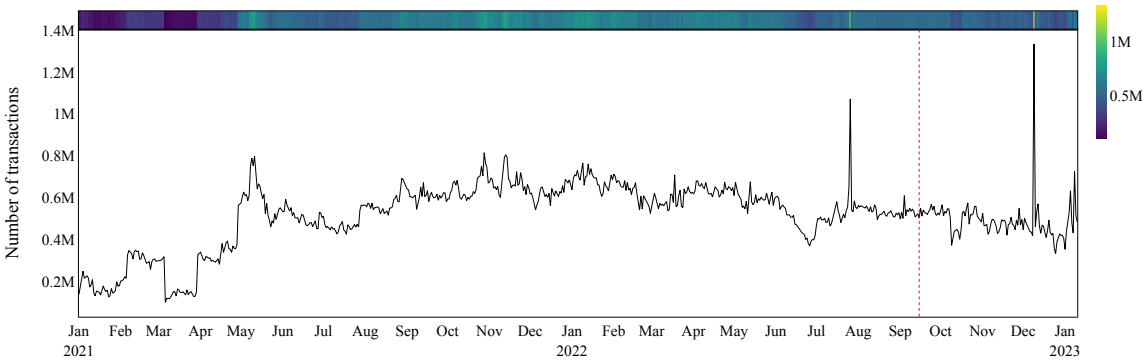
(a) Time series
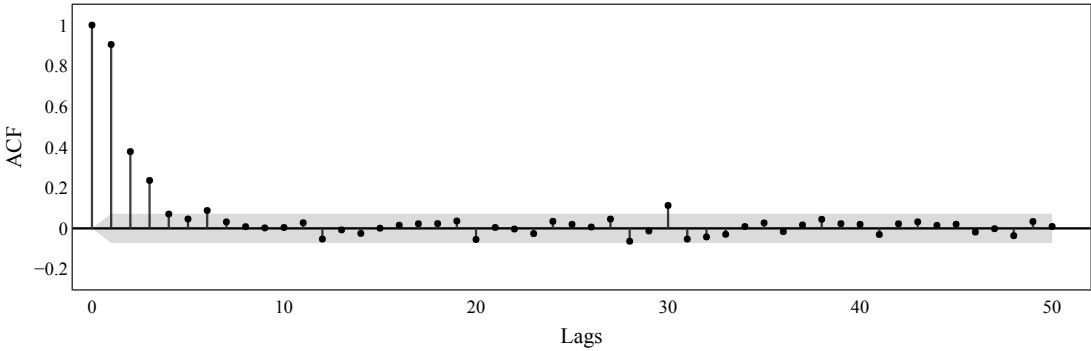
(b) Autocorrelation function

Figure 3.A.23: **Withdrawal transaction**.
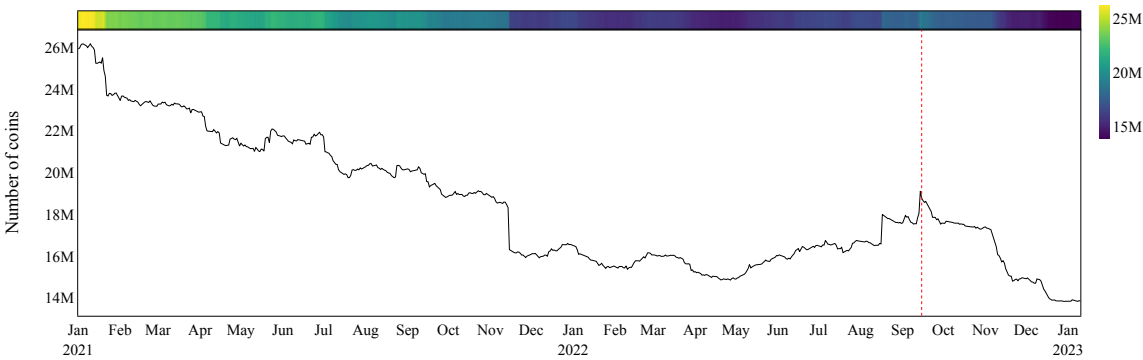
(a) Time series
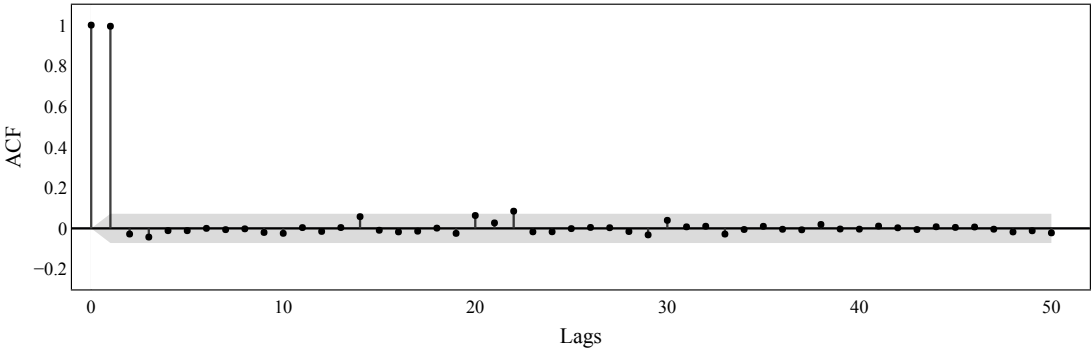
(b) Autocorrelation function

Figure 3.A.24: **Exchange supply**.
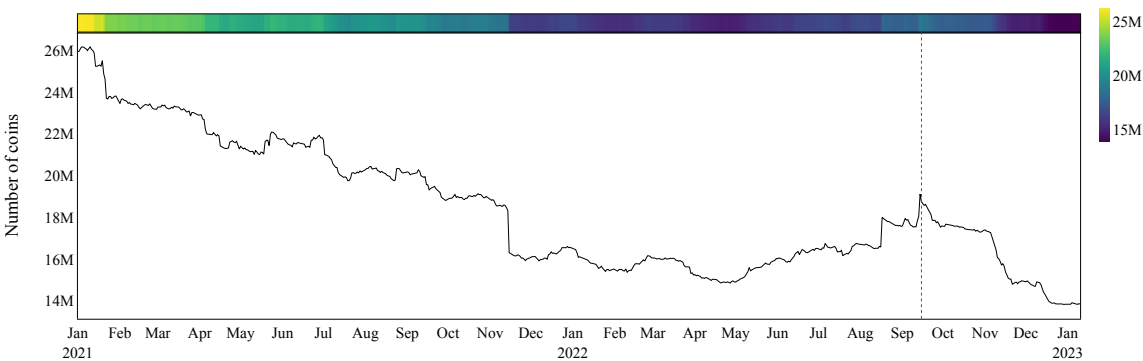
(a) Time series
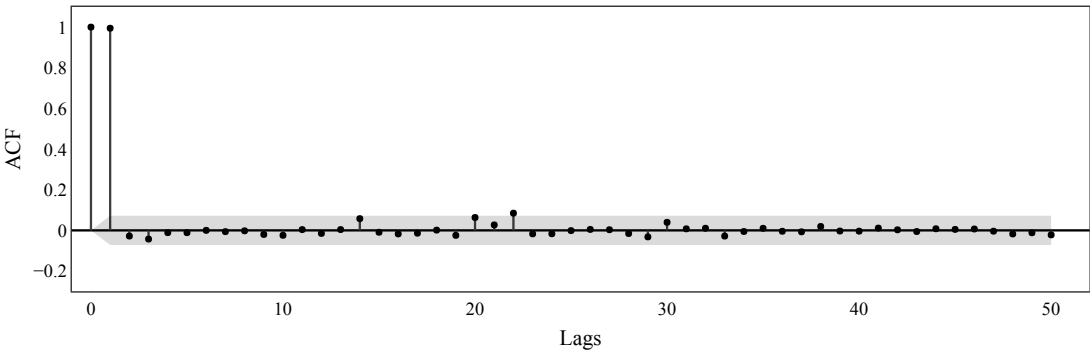
(b) Autocorrelation function

Figure 3.A.25: **Supply on Exchange**.
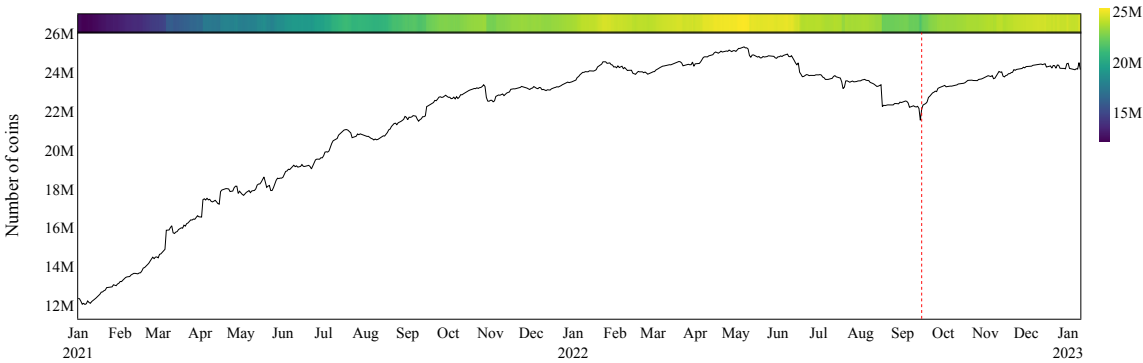
(a) Time series
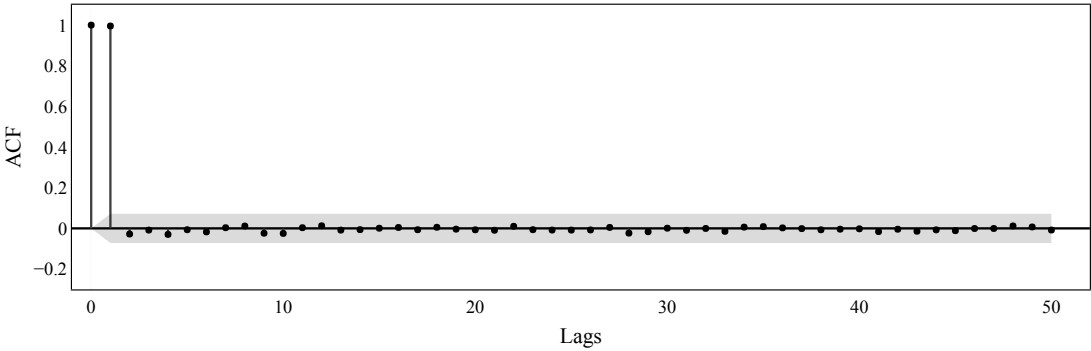
(b) Autocorrelation function

Figure 3.A.26: **Supply on non-exchange top holder**.
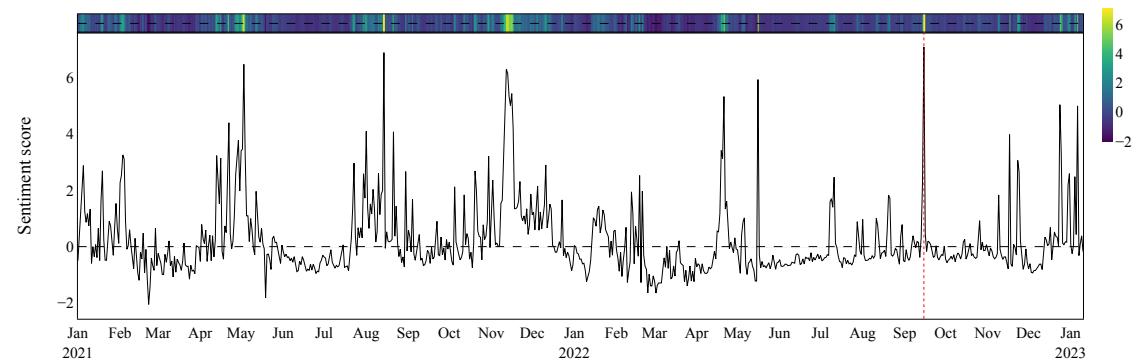
(a) Time series
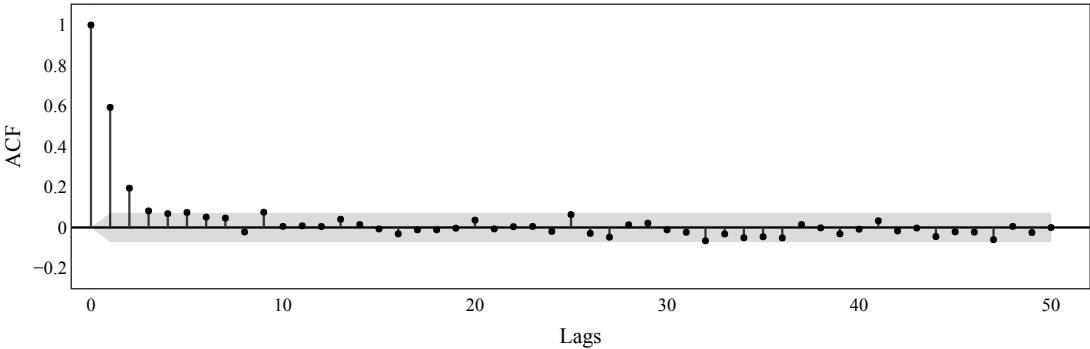
(b) Autocorrelation function

Figure 3.A.27: **Sentiment**.
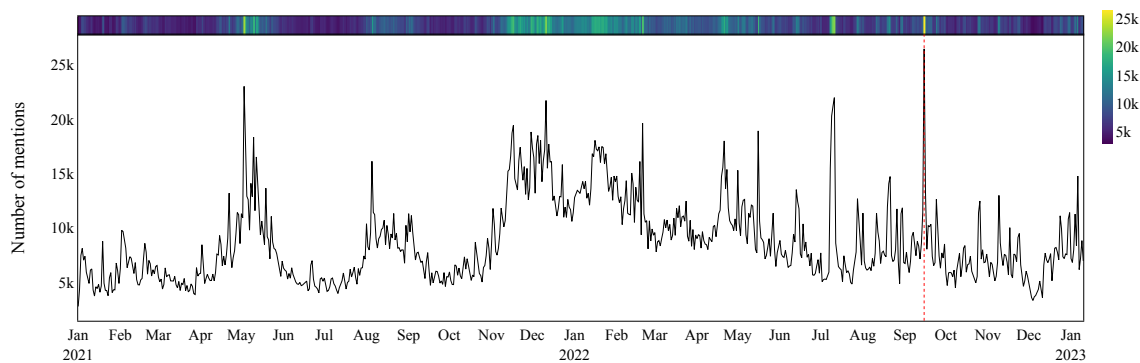
(a) Time series
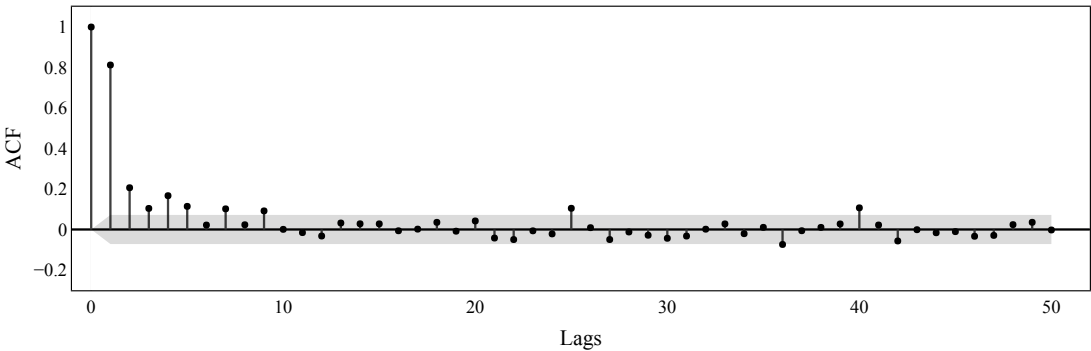
(b) Autocorrelation function

Figure 3.A.28: **Social volume**.
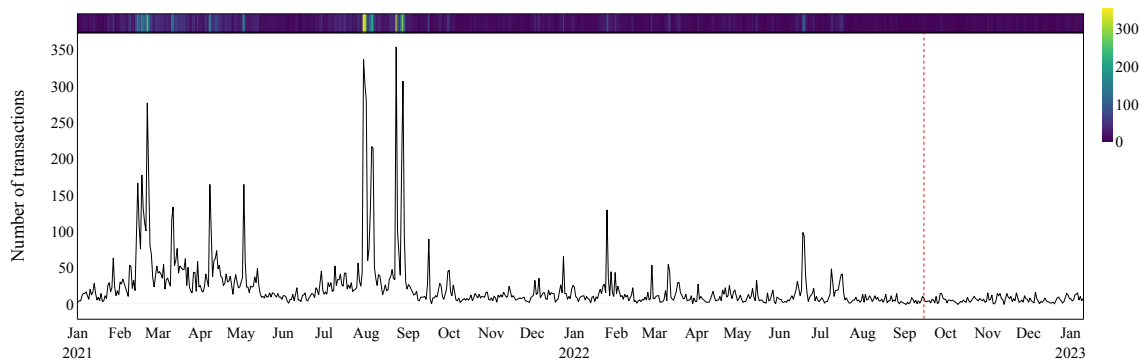
(a) Time series
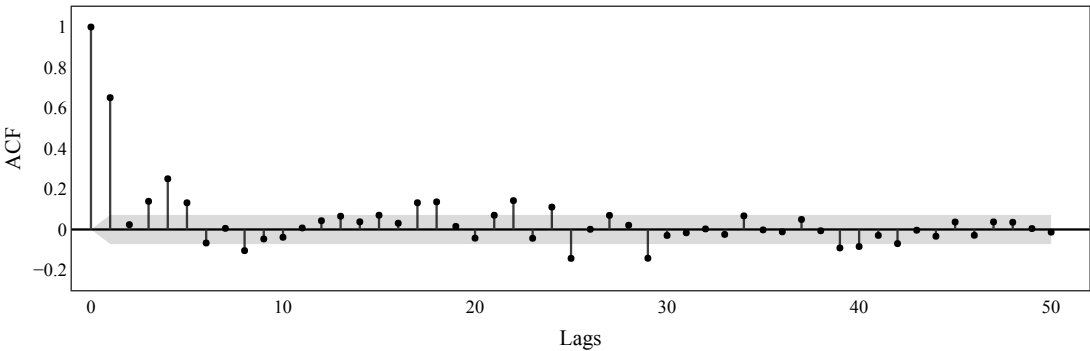
(b) Autocorrelation function

Figure 3.A.29: **NFT transaction count**.
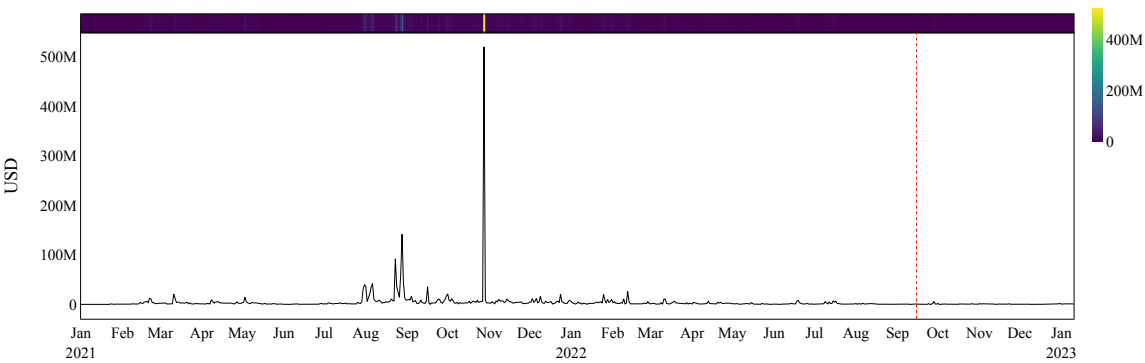
(a) Time series
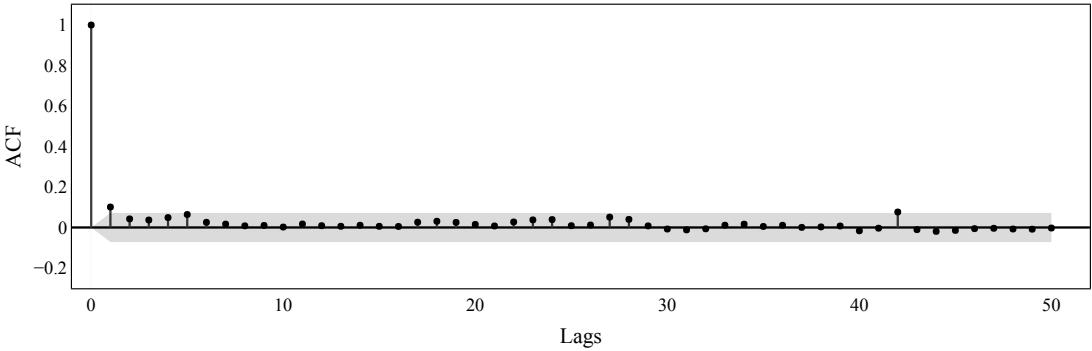
(b) Autocorrelation function

Figure 3.A.30: **NFT trading volume**.

3.B   Sentiment measure

Weighted sentiment is computed by

$$\text{WS}_\tau \overset{\text{def}}{=} \frac{x_\tau - \bar{x}_d}{s_d}$$

where $t - d \leq \tau \leq t$; $\bar{x}_d = d^{-1} \sum_{\tau=t-d}^{t}(x_\tau)$ is the expected sentiment value between times $t - d$ and t; and $s_d = [(d-1)^{-1} \sum_{\tau=t-d}^{t}(x_\tau - \bar{x}_d)^2]^{1/2}$ is the standard deviation. For each time τ, the sentiment value is calculated by

$$x_\tau \overset{\text{def}}{=} \underbrace{[\sum_{\mathbf{m}_i^+ \in \mathbb{M}_\tau^+} \tilde{\psi}(\mathbf{m}_i^+) + \sum_{\mathbf{m}_j^- \in \mathbb{M}_\tau^-} \tilde{\psi}(\mathbf{m}_j^-)]}_{\text{Sentiment balance}} \times \underbrace{(|\mathbb{M}_\tau^+| + |\mathbb{M}_\tau^-|)}_{\text{Unique social volume}}$$

with

$$\psi(\mathbf{m}) = \begin{cases} \tilde{\psi}(\mathbf{m}), & \text{if } |\tilde{\psi}(\mathbf{m})| > \delta \\ 0, & \text{otherwise} \end{cases}$$

where $\mathbf{m}^+ \in \mathbb{M}_\tau^+$ is the positive message from all the positive messages mentioning a given crypto's name at time τ; while $\mathbf{m}^- \in \mathbb{M}_\tau^-$ is the negative message from all the negative messages mentioning a given crypto's name at time τ. $|\mathbb{M}_\tau^+|$ and $|\mathbb{M}_\tau^-|$ are the cardinal numbers of the sets $\mathbb{M}_\tau^+$ and $\mathbb{M}_\tau^-$, which eliminate all redundant messages. $\psi : \mathbb{M} \mapsto [-1, 1]$ is the sentiment function, which gives a real number between -1 and 1 for any message in $\mathbb{M} = \mathbb{M}^+ \cup \mathbb{M}^-$. The threshold δ is set as 0.7 in our dataset, that is, only as a positive sentiment with the value greater than 0.7 and a negative sentiment with the value smaller than -0.7 are included.

3.C Beta-t-EGARCH coefficient

The degree of freedom ν for skew-t distribution is estimated to be 3.168, the presence of fat tails. Skewness γ is slightly larger than 1 showing a right-skewed distribution.

Coefficients	Estimate	Std. error
ω	-3.65752140	0.08433820
ϕ_1	0.95971465	0.02490888
ϕ_2	-0.88026440	0.10086900
κ_1	0.05349386	0.01712863
κ_2	0.01305982	0.01337889
κ^*	-0.01639700	0.01563484
ν	3.16800000	0.42584400
γ	1.00407467	0.03047742
Log-likelihood	1818.83334300	
BIC	-3.73201200	

Table 3.C.1: **Coefficient summary**.

∎

3.D Data summary

Table 3.D.1: **Descriptive statistics.**

Component	Element	Factor	Mean	Median	Maximum	Minimum	Variance	Skewness	Kurtosis
	Scalability	Block utilization (%)	7.791972e+00	8.167307e+00	1.359648e+01	4.359253e+00	3.466998e+00	-0.170707	-0.721345
		Block time	1.331142e+01	1.338938e+01	1.626935e-01	1.203343e-01	4.543463e-01	0.307097	3.425811
		Throughput	1.148436e+06	1.144629e+06	1.908621e+06	7.199180e+05	1.560044e+10	0.852928	3.111308
	Decentralization	Network growth	1.013485e+05	9.070650e+04	2.807940e+05	4.487400e+04	1.306324e+09	1.391067	2.459052
		Network activeness (%)	2.962851e-01	2.827283e-01	6.334347e-01	1.779499e-01	4.683612e-03	0.887137	0.723661
		Difficulty adjustment	8.280864e+00	8.833732e+00	1.510118e-01	0.000000e+00	2.119161e+01	-0.560528	-0.826029
Functional characteristic	Security	Energy consumption	4.821559e+01	4.924531e+01	9.397526e+01	9.043988e-03	1.118302e+03	-0.081735	-1.458111
		Non-zero balance address	7.171021e+07	7.213546e+07	9.237356e+07	5.159098e+07	1.576989e+14	0.043382	-1.439485
		Mining profitability (%)	5.639973e+00	5.580000e+00	2.820000e+01	0.000000e+00	1.819690e+01	0.849518	1.724565
	Incentive mechanism	Fee to reward (%)	1.478565e-01	6.160500e+00	6.382500e-01	0.000000e+00	2.454886e-02	1.179886	0.002879
		Transaction fee	1.199509e+01	7.541574e+00	9.825444e-01	4.980105e-01	1.658707e+02	1.934279	5.080463
	Developer	Development contributor	4.250676e+01	4.700000e+01	9.000000e+01	0.000000e+00	3.737956e+02	-0.260828	-0.980086
		Development activity	2.899324e+02	3.100000e+02	9.570000e+02	0.000000e+00	3.548168e+04	0.260661	-0.637048
	Coin circulation	Supply deviation	8.945249e+03	1.173224e+04	1.706540e+04	-5.771768e+04	3.791802e+07	-2.476391	17.519222
		Coin age	6.694368e+02	6.629122e+02	7.542113e-02	5.996747e-02	2.330320e+03	-0.011826	-1.453108
		Velocity	4.132064e+00	3.615477e+00	2.219884e-01	1.625331e+00	2.968097e+00	3.894708	28.483895
	Market scale	Market capitalization	2.809029e+11	2.502320e+11	5.690943e+11	8.331854e+10	1.359256e+22	0.536666	-0.809102
		MVRV (%)	2.370659e+02	2.101643e+02	9.004998e+02	-6.253587e+01	4.760226e+04	0.358990	-0.789135
		Trading volume	4.387203e+06	3.339703e+06	4.189358e+07	6.610565e+05	1.112076e+13	3.848375	30.729226
	Transaction pattern	Transaction value	4.277390e+03	3.994500e+03	1.937300e+04	0.000000e+00	5.573850e+06	1.218155	3.144587
		Whale transaction	7.433551e+03	7.037500e+03	4.137900e+04	1.060000e+03	1.873066e+07	1.770198	7.293474
Market dynamics	Exchange activity	Flow balance	-1.394385e+04	-1.098193e+04	1.408993e+06	-2.012725e+06	2.076103e+10	-2.167267	70.048201
		Withdrawal transaction	5.160178e+05	5.421365e+05	1.334386e+06	9.993200e+04	2.282865e+10	-0.655756	1.685254
	Wealth distribution	Supply on exchange	1.837773e+07	1.757217e+07	2.617872e+07	1.383429e+07	9.241638e+12	0.600394	-0.682259
		Supply on non-exchange top holder	2.163031e+07	2.318779e+07	2.533451e+07	1.206038e+07	1.207065e+13	-1.301094	0.618359
	Social media	Weighted sentiment	8.652309e+03	7.673000e+03	2.648900e+04	2.789000e+03	1.337572e+07	1.190855	1.406290
		Social volume	1.816305e-01	-2.213505e-01	7.092642e+00	-2.049853e+00	1.560701e+00	2.280214	6.762146
	Applicability	NFT transaction count	2.179865e+01	1.100000e+01	3.540000e+02	0.000000e+00	1.357165e+03	5.287051	34.708478
		NFT trading volume	4.035312e+06	1.419584e+06	5.194883e+08	0.000000e+00	4.251589e+14	21.788295	533.071658
Economic attribute		Log return (%)	0.078055	-0.079531	24.965541	-16.760681	16.719916	0.128502	3.597929
		$\lambda_{1,t}$	2.212877	2.170235	54.520946	-54.345805	386.79043	-0.016531	-0.321923
		$\lambda_{2,t}$	0.022689	-0.024264	18.170428	-16.681751	29.091235	0.103182	0.070790
		$\sigma_{t|t-1}$	4.438963	4.343693	8.207351	2.385711	0.872233	0.654901	0.50809

Table 3.D.2: **ADF test**[a] **results**.

Factor	x Test statistic	p-value	Δx Test statistic	p-value
Block utilization (%)	-1.728	**0.417**	-8.534	0.000
Block time	-1.834	**0.364**	-10.060	0.000
Throughput	-2.453	**0.127**	-8.390	0.000
Network growth	-1.676	**0.443**	-7.840	0.000
Network activeness (%)	-2.226	**0.197**	-13.741	0.000
Difficulty adjustment	-1.067	**0.728**	-17.924	0.000
Energy consumption	-1.240	**0.656**	-23.978	0.000
Non-zero balance address	-0.317	**0.923**	-4.200	0.001
Mining profitability (%)	-1.189	**0.678**	-9.668	0.000
Fee to reward (%)	-1.616	**0.475**	-6.772	0.000
Transaction fee	-2.759	**0.064**	-16.359	0.000
Development contributor	-3.304	0.015	-7.767	0.000
Development activity	-3.976	0.002	-8.152	0.000
Supply deviation	-1.570	**0.499**	-9.882	0.000
Coin age	0.896	**0.993**	-25.791	0.000
Velocity	-3.071	0.029	-18.414	0.000
Market capitalization	-1.792	**0.385**	-10.354	0.000
MVRV (%)	-1.667	**0.448**	-10.164	0.000
Trading volume	-2.217	**0.200**	-9.923	0.000
Transaction value	-2.479	**0.121**	-9.358	0.000
Whale transaction	-2.679	**0.078**	-6.199	0.000
Flow balance	-16.869	0.000	-11.206	0.000
Withdrawal transaction	-2.828	**0.054**	-16.567	0.000
Supply on exchange	-2.214	**0.201**	-16.710	0.000
Supply on non-exchange top holder	-4.251	0.001	-25.373	0.000
Weighted sentiment	-3.291	0.015	-14.058	0.000
Social volume	-5.317	0.000	-14.174	0.000
NFT transaction count	-3.220	0.019	-11.238	0.000
NFT trading volume	-24.533	0.000	-10.803	0.000
Log return (%)	-22.773	0.000	-10.569	0.000
$\lambda_{1,t}$	-3.409	0.011	-29.807	0.000
$\lambda_{2,t}$	-24.405	0.000	-10.764	0.000
$\sigma_{t|t-1}$	-3.527	0.007	-18.786	0.000

[a] H_0 : Series is non-stationary or series has a unit root; H_1 : Series is stationary or series has no unit root.

[b] 95% confidence interval.

Causal discovery

Table 3.D.3: Lag value for Granger causality test.

$x^{(i)} \setminus x^{(j)}$	Δ Block utilization (%)	Δ Block time	Δ Throughput	Δ Network growth	Δ Network activeness (%)	Δ Difficulty adjustment	Δ Energy consumption	Δ Non-zero balance address	Δ Mining profitability (%)	Δ Fee to reward (%)	Δ Transaction fee
Δ Block utilization (%)	0	1	5	2	4	1	1	3	3	5	3
Δ Block time	1	0	0	0	0	0	0	0	0	0	0
Δ Throughput	5	2	0	0	0	0	0	0	0	0	0
Δ Network growth	2	1	2	0	0	0	0	0	0	0	0
Δ Network activeness (%)	4	2	6	5	0	0	0	0	0	0	0
Δ Difficulty adjustment	1	1	2	1	2	0	0	0	0	0	0
Δ Energy consumption	1	1	2	1	2	1	0	0	0	0	0
Δ Non-zero balance address	3	3	3	3	3	3	3	0	0	0	0
Δ Mining profitability (%)	3	1	6	3	5	1	1	3	0	0	0
Δ Fee to reward (%)	5	1	5	2	5	1	1	3	3	0	0
Δ Transaction fee	3	1	3	2	3	1	1	3	3	3	0
Development contributor	9	9	8	9	8	9	8	9	8	15	9
Development activity	9	9	9	9	9	9	9	9	9	11	9
Δ Supply deviation	4	3	5	3	4	3	3	3	3	5	5
Δ Coin age	1	1	2	1	2	1	1	3	1	2	2
Velocity	5	2	6	5	5	2	2	5	7	12	6
Δ Market capitalization	1	1	2	1	2	1	1	3	1	2	1
Δ MVRV (%)	1	1	2	2	3	1	1	3	7	11	3
Δ Trading volume	4	3	5	4	4	4	3	4	4	11	4
Δ Transaction value	6	6	6	6	6	6	6	6	6	6	6
Δ Whale transaction	5	3	6	6	6	3	3	6	6	6	6
Flow balance	2	1	2	2	2	1	1	3	3	3	2
Δ Withdrawal transaction	2	2	2	2	2	2	2	3	3	2	2
Δ Supply on exchange	1	1	2	2	2	1	1	3	1	3	2
Supply on non-exchange top holder	1	1	2	1	2	2	2	3	1	1	1
Weighted sentiment	4	2	5	2	4	2	2	4	3	5	2
Social volume	3	2	2	2	2	2	2	3	3	2	2
NFT transaction count	4	1	5	4	4	1	1	4	5	5	4
NFT trading volume	1	1	2	1	2	1	1	3	1	1	1
Log return (%)	2	2	2	2	2	2	2	3	2	2	2
$\lambda_{1,t}$	2	1	2	1	3	1	1	3	1	3	3
$\lambda_{2,t}$	4	1	3	2	3	1	1	3	1	3	3
$\sigma_{t|t-1}$	4	3	3	3	3	3	3	3	6	5	3

Table 3.D.3: **Lag value for Granger causality test (cont'd).**

$x^{(i)} \setminus x^{(j)}$	Development contributor	Development activity	Δ Supply deviation	Δ Coin age	Velocity	Δ Market capitalization	Δ MVRV (%)	Δ Trading volume	Δ Transaction value	Δ Whale transaction	Flow balance
Δ Block utilization (%)	9	9	4	1	5	1	1	4	6	5	2
Δ Block time	0	0	0	0	0	0	0	0	0	0	0
Δ Throughput	0	0	0	0	0	0	0	0	0	0	0
Δ Network growth	0	0	0	0	0	0	0	0	0	0	0
Δ Network activeness (%)	0	0	0	0	0	0	0	0	0	0	0
Δ Difficulty adjustment	0	0	0	0	0	0	0	0	0	0	0
Δ Energy consumption	0	0	0	0	0	0	0	0	0	0	0
Δ Non-zero balance address	0	0	0	0	0	0	0	0	0	0	0
Δ Mining profitability (%)	0	0	0	0	0	0	0	0	0	0	0
Δ Fee to reward (%)	0	0	0	0	0	0	0	0	0	0	0
Δ Transaction fee	0	0	0	0	0	0	0	0	0	0	0
Development contributor	8	0	4	1	0	0	0	0	0	0	0
Development activity	0	8	9	1	0	0	0	0	0	0	0
Δ Supply deviation	9	9	5	3	0	0	0	0	0	0	0
Δ Coin age	8	9	3	3	0	0	0	0	0	0	0
Velocity	9	9	5	3	0	0	0	0	0	0	0
Δ Market capitalization	8	9	3	1	6	0	0	0	6	0	0
Δ MVRV (%)	10	10	3	1	11	2	0	0	6	0	0
Δ Trading volume	10	10	6	4	7	6	10	0	6	0	0
Δ Transaction value	8	8	6	6	6	6	6	6	0	0	0
Δ Whale transaction	8	8	5	3	7	6	7	10	6	0	0
Flow balance	9	9	3	1	2	1	1	3	6	3	0
Δ Withdrawal transaction	8	9	3	2	5	2	2	3	6	5	2
Δ Supply on exchange	9	9	3	1	2	1	1	3	6	3	2
Supply on non-exchange top holder	9	9	4	2	2	2	1	4	6	3	1
Weighted sentiment	9	9	3	2	6	2	2	4	6	5	2
Social volume	9	9	3	2	6	2	2	4	6	6	2
NFT transaction count	9	9	4	1	5	1	4	4	6	6	1
NFT trading volume	9	9	3	19	2	1	1	3	6	3	1
Log return (%)	8	9	3	2	3	3	2	5	6	6	2
$\lambda_{1,t}$	8	9	3	1	6	1	2	4	6	3	1
$\lambda_{2,t}$	9	9	3	1	5	2	1	4	6	6	1
$\sigma_{t|t-1}$	8	9	4	3	5	3	3	4	6	6	3

Table 3.D.3: Lag value for Granger causality test (cont'd).

$x^{(i)} \setminus x^{(j)}$	Δ Withdrawal transaction	Δ Supply on exchange	Supply on non-exchange top holder	Weighted sentiment	Social volume	NFT transaction count	NFT trading volume	Log return (%)	$\lambda_{1,t}$	$\lambda_{2,t}$	$\sigma_{t\|t-1}$
Δ Block utilization (%)	2	1	1	4	3	4	1	2	2	4	4
Δ Block time	0	0	0	0	0	0	0	0	0	0	3
Δ Throughput	0	0	0	0	0	0	0	0	0	0	3
Δ Network growth	0	0	0	0	0	0	0	0	0	0	3
Δ Network activeness (%)	0	0	0	0	0	0	0	0	0	0	3
Δ Difficulty adjustment	0	0	0	0	0	0	0	0	0	0	3
Δ Energy consumption	0	0	0	0	0	0	0	0	0	0	3
Δ Non-zero balance address	0	0	0	0	0	0	0	0	0	0	3
Δ Mining profitability (%)	0	0	0	0	0	0	0	0	0	0	6
Δ Fee to reward (%)	0	0	0	0	0	0	0	0	0	0	5
Δ Transaction fee	0	0	0	0	0	0	0	0	0	0	3
Development contributor	0	0	0	0	0	0	0	0	0	0	8
Development activity	0	0	0	0	0	0	0	0	0	0	9
Δ Supply deviation	0	0	0	0	0	0	0	0	0	0	4
Δ Coin age	0	0	0	0	0	0	0	0	0	0	3
Velocity	0	0	0	0	0	0	0	0	0	0	5
Δ Market capitalization	0	0	0	0	0	0	0	0	0	0	3
Δ MVRV (%)	0	0	0	0	0	0	0	0	0	0	3
Δ Trading volume	0	0	0	0	0	0	0	0	0	0	4
Δ Transaction value	0	0	0	0	0	0	0	0	0	0	6
Δ Whale transaction	0	0	0	0	0	0	0	0	0	0	6
Flow balance	0	0	0	0	0	0	0	0	0	0	3
Δ Withdrawal transaction	0	0	0	0	0	0	0	0	0	0	3
Δ Supply on exchange	2	1	0	0	0	0	0	0	0	0	3
Supply on non-exchange top holder	2	2	2	0	0	0	0	0	0	0	3
Weighted sentiment	2	2	2	2	2	4	2	2	4	3	4
Social volume	4	1	2	2	2	1	1	2	2	2	3
NFT transaction count	2	1	1	4	4	0	1	2	4	4	4
NFT trading volume	2	2	2	2	2	1	0	0	3	3	3
Log return (%)	2	1	1	2	2	4	2	0	0	2	4
$\lambda_{1,t}$	2	1	1	4	2	1	1	2	0	1	3
$\lambda_{2,t}$	2	1	1	3	2	4	1	2	1	0	3
$\sigma_{t\|t-1}$	3	3	3	4	3	4	3	4	3	3	0

Table 3.D.4: GC test[a] result.

$x^{(i)} \setminus x^{(j)}$	Δ Block utilization (%)	Δ Block time	Δ Throughput	Δ Network growth	Δ Network activeness (%)	Δ Difficulty adjustment	Δ Energy consumption	Δ Non-zero balance address	Δ Mining profitability (%)	Δ Fee to reward (%)	Δ Transaction fee
Δ Block utilization (%)	1.000	0.412	0.001	0.055	0.002	0.539	0.651	0.012	0.916	0.458	0.278
Δ Block time	0.335	1.000	0.373	0.485	0.219	0.000	0.000	0.187	0.251	0.505	0.602
Δ Throughput	0.202	0.714	1.000	0.003	0.056	0.148	0.148	0.013	0.013	0.000	0.002
Δ Network growth	0.450	0.533	0.003	1.000	0.002	0.478	0.355	0.000	0.063	0.015	0.031
Δ Network activeness (%)	0.015	0.854	0.485	0.001	1.000	0.128	0.450	0.009	0.050	0.054	0.073
Δ Difficulty adjustment	0.605	0.009	0.674	0.286	0.698	1.000	0.000	0.213	0.843	0.959	0.414
Δ Energy consumption	0.889	0.000	0.253	0.903	0.067	0.062	1.000	0.250	0.747	0.918	0.907
Δ Non-zero balance address	0.185	0.097	0.000	0.000	0.000	0.000	0.000	1.000	0.022	0.012	0.224
Δ Mining profitability (%)	0.309	0.164	0.766	0.000	0.530	0.010	0.025	0.174	1.000	0.014	0.007
Δ Fee to reward (%)	0.408	0.566	0.470	0.004	0.540	0.135	0.375	0.251	0.008	1.000	0.236
Δ Transaction fee	0.629	0.532	0.105	0.000	0.856	0.484	0.911	0.062	0.000	0.236	1.000
Development contributor	0.000	0.008	0.000	0.000	0.302	0.169	0.113	0.005	0.007	0.000	0.006
Development activity	0.000	0.129	0.000	0.002	0.271	0.177	0.123	0.005	0.061	0.000	0.276
Δ Supply deviation	0.249	0.502	0.116	0.166	0.980	0.215	0.764	0.789	0.868	0.690	0.331
Δ Coin age	0.986	0.593	0.559	0.548	0.443	0.637	0.807	0.075	0.433	0.969	0.304
Velocity	0.560	0.133	0.197	0.164	0.073	0.862	0.530	0.000	0.000	0.000	0.113
Δ Market capitalization	0.838	0.877	0.335	0.340	0.183	0.401	0.563	0.174	0.331	0.246	0.457
Δ MVRV (%)	0.445	0.637	0.112	0.002	0.152	0.555	0.610	0.083	0.002	0.000	0.409
Δ Trading volume	0.003	0.010	0.088	0.394	0.147	0.010	0.075	0.388	0.124	0.078	0.230
Δ Transaction value	0.003	0.049	0.108	0.024	0.489	0.000	0.004	0.148	0.005	0.020	0.004
Δ Whale transaction	0.016	0.054	0.100	0.018	0.456	0.033	0.348	0.595	0.002	0.025	0.025
Flow balance	0.715	0.748	0.344	0.444	0.569	0.388	0.963	0.256	0.000	0.244	0.227
Δ Withdrawal transaction	0.463	0.450	0.010	0.387	0.007	0.308	0.890	0.582	0.218	0.126	0.169
Δ Supply on exchange	0.653	0.821	0.330	0.497	0.603	0.418	0.950	0.214	0.954	0.007	0.257
Supply on non-exchange top holder	0.850	0.876	0.673	0.071	0.568	0.055	0.338	0.014	0.336	0.055	0.568
Weighted sentiment	0.266	0.000	0.404	0.411	0.640	0.000	0.000	0.015	0.061	0.209	0.113
Social volume	0.180	0.090	0.730	0.195	0.677	0.001	0.001	0.138	0.341	0.198	0.227
NFT transaction count	0.238	0.804	0.374	0.746	0.450	0.456	0.678	0.227	0.905	0.602	0.714
NFT trading volume	0.804	0.383	0.920	0.580	0.828	0.752	0.736	0.041	0.551	0.247	0.050
Log return (%)	0.054	0.564	0.868	0.127	0.084	0.007	0.006	0.117	0.028	0.007	0.107
$\lambda_{1,t}$	0.027	0.991	0.017	0.729	0.159	0.770	0.548	0.529	0.332	0.001	0.295
$\lambda_{2,t}$	0.084	0.248	0.030	0.000	0.070	0.246	0.148	0.005	0.000	0.000	0.001
$\sigma_{t|t-1}$	0.225	0.685	0.025	0.722	0.064	0.598	0.011	0.013	0.005	0.239	0.937

[a] $H_0: x^{(j)} \overset{GC}{\nRightarrow} x^{(i)}$; $H_1: x^{(j)} \overset{GC}{\Rightarrow} x^{(i)}$; $i \neq j$.

[b] 95% confidence interval.

Table 3.D.4: GC test[a] result (cont'd).

$x^{(i)} \setminus x^{(j)}$	Development contributor	Development activity	Δ Supply deviation	Δ Coin age	Velocity	Δ Market capitalization	Δ MVRV (%)	Δ Trading volume	Δ Transaction value	Δ Whale transaction	Flow balance
Δ Block utilization (%)	0.000	0.000	0.258	0.457	0.571	0.455	0.062	0.034	0.001	0.008	0.192
Δ Block time	0.033	0.152	0.656	0.011	0.251	0.197	0.653	0.198	0.014	0.134	0.015
Δ Throughput	0.000	0.000	0.277	0.184	0.165	0.303	0.097	0.035	0.000	0.000	0.458
Δ Network growth	0.000	0.000	0.195	0.881	0.015	0.120	0.285	0.076	0.033	0.045	0.281
Δ Network activeness (%)	0.180	0.415	0.530	0.020	0.324	0.610	0.806	0.162	0.196	0.017	0.366
Δ Difficulty adjustment	0.031	0.015	0.415	0.000	0.558	0.980	0.855	0.005	0.000	0.539	0.000
Δ Energy consumption	0.121	0.168	0.902	0.000	0.470	0.152	0.185	0.042	0.003	0.499	0.000
Δ Non-zero balance address	0.000	0.001	0.477	0.014	0.001	0.064	0.206	0.627	0.000	0.099	0.001
Δ Mining profitability (%)	0.001	0.000	0.459	0.384	0.000	0.000	0.000	0.000	0.000	0.000	0.211
Δ Fee to reward (%)	0.000	0.000	0.359	0.001	0.000	0.001	0.000	0.000	0.001	0.000	0.266
Δ Transaction fee	0.000	0.001	0.000	0.010	0.060	0.002	0.000	0.205	0.000	0.002	0.166
Development contributor	1.000	0.000	0.059	0.214	0.014	0.052	0.075	0.000	0.000	0.000	0.042
Development activity	0.000	1.000	0.061	0.390	0.055	0.231	0.447	0.000	0.221	0.000	0.009
Δ Supply deviation	0.143	0.268	1.000	0.619	0.079	0.230	0.909	0.520	0.771	0.156	0.613
Δ Coin age	0.065	0.011	0.335	1.000	0.558	0.998	0.843	0.271	0.001	0.657	0.383
Velocity	0.154	0.018	0.645	0.037	1.000	0.000	0.001	0.000	0.005	0.000	0.013
Δ Market capitalization	0.262	0.267	0.458	0.014	0.015	1.000	0.232	0.016	0.053	0.002	0.006
Δ MVRV (%)	0.654	0.573	0.871	0.055	0.000	0.000	1.000	0.000	0.000	0.014	0.056
Δ Trading volume	0.000	0.000	0.259	0.159	0.000	0.000	0.002	1.000	0.000	0.000	0.582
Δ Transaction value	0.000	0.000	0.138	0.000	0.182	0.001	0.000	0.036	1.000	0.000	0.249
Δ Whale transaction	0.000	0.000	0.119	0.005	0.084	0.000	0.000	0.000	0.001	1.000	0.188
Flow balance	0.157	0.053	0.796	0.732	0.887	0.954	0.246	0.030	0.077	0.004	1.000
Δ Withdrawal transaction	0.018	0.050	0.519	0.282	0.229	0.115	0.061	0.402	0.625	0.295	0.873
Δ Supply on exchange	0.138	0.026	0.827	0.913	0.956	0.955	0.234	0.010	0.048	0.007	0.570
Supply on non-exchange top holder	0.195	0.450	0.378	0.105	0.359	0.441	0.093	0.261	0.568	0.630	0.347
Weighted sentiment	0.053	0.051	0.062	0.033	0.041	0.723	0.737	0.073	0.014	0.017	0.388
Social volume	0.043	0.204	0.092	0.003	0.758	0.222	0.867	0.059	0.000	0.025	0.253
NFT transaction count	0.043	0.324	0.959	0.882	0.000	0.661	0.847	0.023	0.184	0.264	0.898
NFT trading volume	0.000	0.045	0.988	0.845	0.758	0.309	0.450	0.236	0.059	0.242	0.565
Log return (%)	0.009	0.001	0.371	0.042	0.094	0.000	0.000	0.000	0.000	0.000	0.288
$\lambda_{1,t}$	0.001	0.000	0.548	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.025
$\lambda_{2,t}$	0.015	0.007	0.221	0.096	0.023	0.000	0.000	0.001	0.000	0.000	0.026
$\sigma_{t|t-1}$	0.015	0.001	0.612	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.001

[a] $H_0 : x^{(j)} \overset{GC}{\nRightarrow} x^{(i)}$; $H_1 : x^{(j)} \overset{GC}{\Rightarrow} x^{(i)}$; $i \neq j$.

[b] 95% confidence interval.

Table 3.D.4: **GC test**[a] **result (cont'd).**

$x^{(i)} \setminus x^{(j)}$	Δ Withdrawal transaction	Δ Supply on exchange	Supply on non-exchange top holder	Weighted sentiment	Social volume	NFT transaction count	NFT trading volume	Log return (%)	$\lambda_{1,t}$	$\lambda_{2,t}$	$\sigma_{t\mid t-1}$
Δ Block utilization (%)	0.076	0.170	0.898	0.014	0.000	0.124	0.664	0.020	0.635	0.288	0.171
Δ Block time	0.885	0.011	0.987	0.151	0.042	0.275	0.806	0.148	0.769	0.629	0.656
Δ Throughput	0.005	0.478	0.465	0.230	0.363	0.610	0.749	0.006	0.343	0.594	0.663
Δ Network growth	0.000	0.370	0.812	0.064	0.629	0.894	0.547	0.027	0.891	0.303	0.678
Δ Network activeness (%)	0.017	0.819	0.576	0.368	0.853	0.767	0.615	0.352	0.120	0.412	0.295
Δ Difficulty adjustment	0.927	0.000	0.000	0.354	0.123	0.331	0.563	0.060	0.693	0.685	0.584
Δ Energy consumption	0.218	0.000	0.069	0.000	0.296	0.641	0.754	0.000	0.228	0.527	0.162
Δ Non-zero balance address	0.000	0.001	0.000	0.000	0.001	0.628	0.390	0.000	0.000	0.397	0.000
Δ Mining profitability (%)	0.057	0.496	0.577	0.103	0.001	0.012	0.508	0.000	0.934	0.746	0.040
Δ Fee to reward (%)	0.539	0.882	0.942	0.000	0.577	0.033	0.622	0.005	0.114	0.197	0.002
Δ Transaction fee	0.248	0.405	0.062	0.062	0.942	0.129	0.548	0.021	0.342	0.644	0.771
Δ Supply deviation	0.953	0.625	0.742	0.510	0.731	0.854	0.950	0.858	0.319	0.903	0.363
Development contributor	0.138	0.060	0.010	0.138	0.134	0.586	0.013	0.107	0.261	0.079	0.207
Development activity	0.063	0.015	0.060	0.688	0.096	0.061	0.000	0.204	0.102	0.107	0.155
Δ Coin age	0.601	0.002	0.481	0.819	0.005	0.631	0.000	0.268	0.268	0.058	0.147
Velocity	0.870	0.000	0.004	0.000	0.002	0.020	0.554	0.001	0.000	0.000	0.000
Δ Market capitalization	0.841	0.073	0.013	0.033	0.318	0.280	0.193	0.501	0.133	0.364	0.080
Δ MVRV (%)	0.728	0.114	0.225	0.147	0.259	0.628	0.155	0.000	0.438	0.173	0.211
Δ Trading volume	0.587	0.842	0.391	0.232	0.288	0.672	0.549	0.001	0.025	0.014	0.002
Δ Transaction value	0.455	0.220	0.128	0.071	0.180	0.140	0.001	0.000	0.030	0.020	0.026
Δ Whale transaction	0.320	0.128	0.378	0.088	0.114	0.063	0.149	0.000	0.012	0.058	0.002
Flow balance	0.694	0.378	0.347	0.634	0.004	0.239	0.409	0.165	0.662	0.012	0.726
Δ Withdrawal transaction	1.000	0.823	0.576	0.456	0.787	0.899	0.930	0.059	0.579	0.073	0.357
Δ Supply on exchange	0.743	1.000	0.091	0.869	0.003	0.179	0.355	0.138	0.630	0.748	0.823
Supply on non-exchange top holder	0.556	0.406	1.000	0.132	0.537	0.280	0.000	0.449	0.660	0.283	0.076
Weighted sentiment	0.438	0.361	0.869	1.000	0.001	0.350	0.455	0.575	0.006	0.006	0.003
Social volume	0.444	0.259	0.003	0.291	1.000	0.520	0.008	0.202	0.102	0.679	0.122
NFT transaction count	0.504	0.572	0.280	0.184	0.150	1.000	0.214	0.609	0.077	0.937	0.137
NFT trading volume	0.696	0.507	0.000	0.759	0.289	0.064	1.000	0.790	0.300	0.812	0.283
Log return (%)	0.272	0.066	0.449	0.000	0.202	0.202	0.220	1.000	0.137	0.329	0.149
$\lambda_{1,t}$	0.388	0.007	0.660	0.000	0.001	0.485	0.295	0.137	1.000	0.157	0.131
$\lambda_{2,t}$	0.622	0.055	0.748	0.000	0.000	0.011	0.940	0.468	0.137	1.000	0.050
$\sigma_{t\mid t-1}$	0.170	0.008	0.823	0.000	0.000	0.174	0.533	0.000	0.000	0.000	1.000

[a] $H_0 : x^{(i)} \overset{GC}{\nRightarrow} x^{(j)}$; $H_1 : x^{(i)} \overset{GC}{\Rightarrow} x^{(j)}$; $i \neq j$.

[b] 95% confidence interval.

Chapter 4

Cross-exchange Crypto Risk:
A High-frequency Dynamic Network Perspective

All the supplementary materials and source codes of this chapter are available in the Q^2 ecosystem: network_BTC_exchanges and .

4.1 Introduction

During the COVID-19 pandemic, cryptocurrency (crypto), as an alternative risk diversifier for typical global commodities such as gold and oil, attracted considerable attention from investors. Cryptos have increasingly become common currencies between economies, facilitating more trade. A large body of literature (Guesmi et al., 2019; Huang et al., 2022; Vukovic et al., 2021) has pointed to the potential diversification benefits of cryptos. Cryptos have thus far demonstrated rather different price evolution and risk profiles – i.e., price stability and resistance to manipulative trading – compared to stock and commodity markets (Alexander & Heck, 2020); see Figure 4.1. Dwyer, 2015 and Petukhina et al., 2021 indicate that the rise of high-frequency, 24/7 trading on the computerized markets of cryptos and reputational equilibrium from the central mechanism – blockchain technology – offer no intervention from intermediaries and redirect investors' trading behaviors, e.g., algorithmic trading. While considerable research has been conducted on cross sections of cryptos, limited research has been conducted on cross sections of their exchanges. To capture the high-frequency nature of crypto trading and examine the fast-changing relationship among exchanges, we consider hourly returns on Bitcoin (BTC) across crypto exchanges using a dynamic network representation. Furthermore, we demonstrate a dynamic portfolio construction that considers the exchanges' relationship dynamics.

There currently exist many crypto exchanges worldwide; due to their varying levels
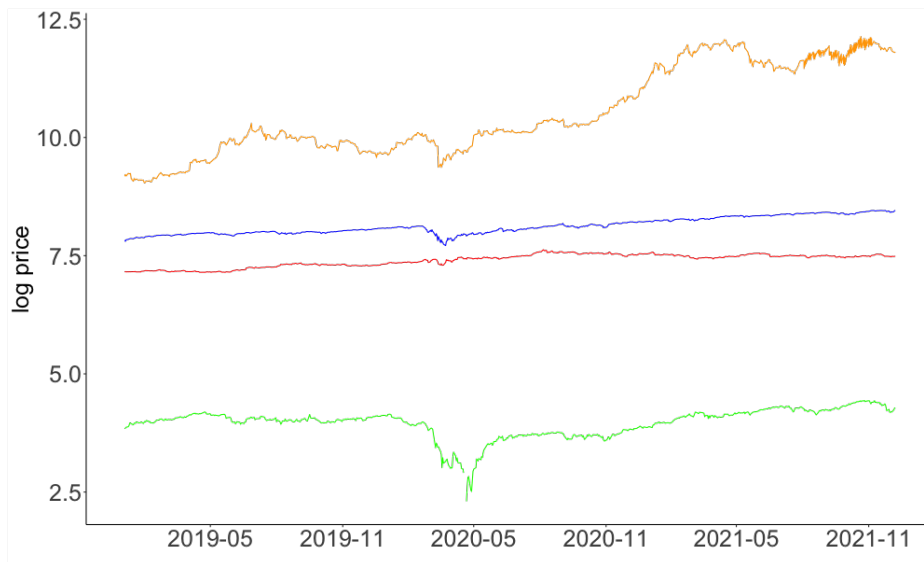
Figure 4.1: **CRIX, S&P500, gold and oil log prices in USD.** The time series are listed from top to bottom. CRIX is a cryptocurrency index that traces the evolution of the crypto market.

of trading volume and liquidity – supply and demand – they have different exchange prices. In particular, many of these exchanges are centralized and are operated by privately owned companies as intermediaries between buyers and sellers. There is often third-party risk for such exchanges because of government interference. For example, FTX, one of the largest exchanges worldwide, collapsed in November 2022 due to a shortage of liquidity (Kelley, 2022). Shortly thereafter, the BlockFi exchange filed for bankruptcies as a result of the collateral damage from FTX. That is, BlockFi had a significant risk exposure to FTX. It is therefore critical to study the correlation risk that was induced by the time-varying correlations among these exchanges (Buraschi et al., 2010; Krishnan et al., 2009).

In Figure 4.2, the BTC returns across nine different exchanges demonstrate common upward and downward patterns; however, the return-fluctuation ranges are quite different. A closer examination of any certain exchange pair as an example in Figure 4.2 reveals that there exists a deviation from the 45 degree, which indicates that the prices are more aligned as the price discovery occurs – so known as arbitrage opportunity (Makarov & Schoar, 2020). Given such price deviation, investors can seek short-term profit and extend it to a longer time period by creating a dynamic portfolio, e.g., using a trading bot. A cross-sectional analysis and operation of crypto arbitrage can be found in the study by Borri and Shakhnov, 2022. However, such exchange arbitrage is limited by the fund transfer time on a certain blockchain and a cryptocurrency exchange's processing time. If a blockchain network is congested (often the case with the BTC network) or an exchange takes too long to process transactions, timely arbitrage will fail, resulting in loss of profits. To counter such potentially debilitating factors, most

institutional quantitative investors store a large amount of various cryptocurrencies in various exchanges to facilitate quick capitalization of arbitrage opportunities. Such a strategy, however, could also result in losses arising from funds sitting idle; thus, investors are encouraged to seek an appropriate balance between risk and opportunity. We thus focus on the cross-exchange risk and include the correlation risk induced by the changing correlations among these exchanges.
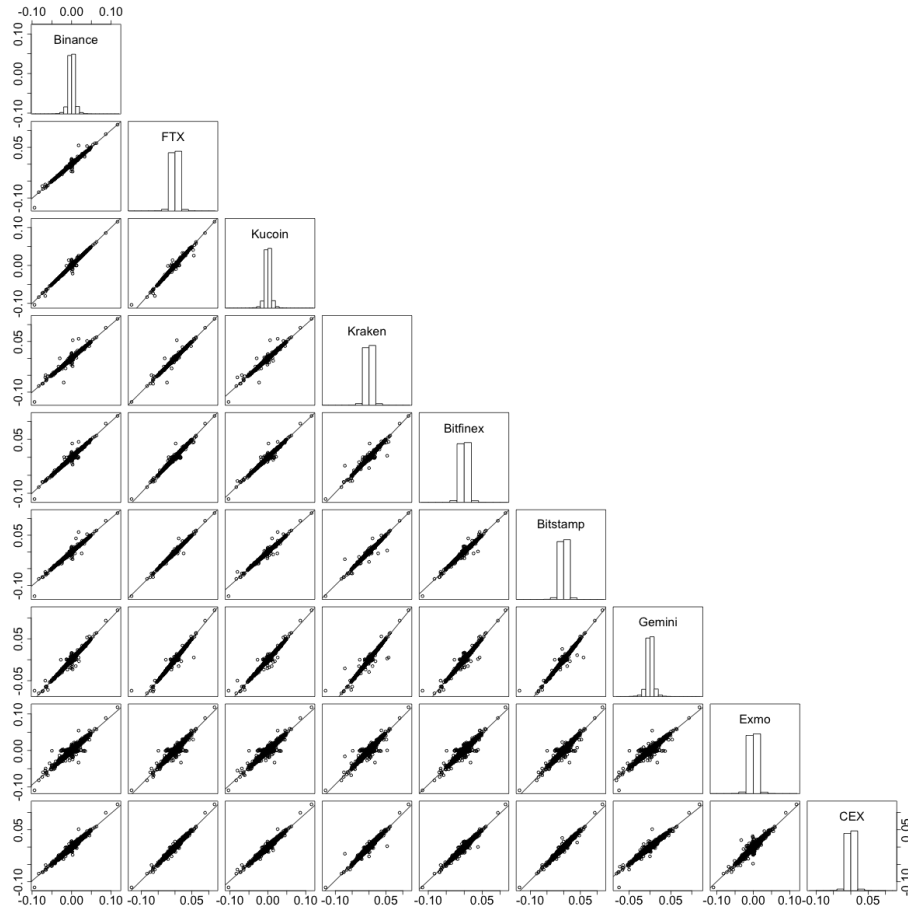
Figure 4.2: **Paired Bitcoin returns among 9 different exchanges.**

We examine correlation risk and quantify its influence via partial correlations, which allow us to model pairwise interactions between two variables after accounting for other variables. In other words, we can disentangle the impact of other exchanges' price movements on BTC. The implementation of partial correlations and realization of the corresponding network models have been discussed in the literature. Andrieş et al., 2022 treat partial-correlation networks as a better measure to capture market structure and systemic risks than traditionally used correlation networks due to their consideration of direct relationships. Highlighting the impact of the COVID-19 pandemic, So et al., 2021 assess systemic risk via partial-correlation networks and show that partial correlations can better reflect the increase in the network connectedness during the pandemic period than correlations. If a pair of exchanges is correlated with other elements in the market,

then the correlation between them may introduce spurious information. For example, this would be typical of exchanges that trade common cryptos. Generally, partial correlations extract core information and control the influence of other exchanges. That is, we avoid the common influence from BTC itself and focus on the pairwise exchange relationships.

Currently, there are over 12,000 different cryptos across more than 600 exchanges on CoinGecko. Given the exuberance of the crypto market and its adequate data, we have a deeper understanding of its unique characteristics, i.e., the perceived opportunities for short gains and impacts of blockchain architecture (Lin et al., 2021). Some stylized facts about cryptos can be found in Vidal-Tomás, 2021. For example, the return on BTC is relatively volatile and, consequently, it is important to consider a high-frequency setting (e.g., hourly) to depict its rapid and dynamic price evolution. A quantitative model that is economically motivated by the interplay between short-term traders is the heterogeneous autoregression (HAR) model of Corsi, 2009. It matches the long-memory property and is numerically attractive and intuitive as it appears as a simplified regression-based procedure. Indeed, it approximates the persistence of volatility time series under an additive cascade structure from the short to the long term. This idea of additive components in time series can be justified in terms of differences in agents' risk profiles, institutional structures, temporal horizons, etc. Based on HAR, we accommodate the stylized facts about cryptos and construct a dynamic, multivariate, volatility model that allows long-range dependence, known as Multivariate Heterogeneous AutoRegression for Crypto Markets (MHAR-CM). Last, we accordingly construct a dynamic portfolio that considers correlation risk based on the centrality measures of high-frequency, dynamic, partial-correlation networks to conduct a robust and adaptive trading strategy specifically for cross-exchange trading on BTC.

The chapter proceeds as follows. Section 2 describes the methodology developed. Section 3 conducts an empirical study and analyzes partial correlations and dynamic networks. Section 4 is dedicated to the implementation of correlation risk in asset allocation. We conclude with findings in Section 5.

4.2   Methodology

In this section, we present, in detail, the construction of partial correlation networks across crypto exchanges and their embedding in the HAR. The estimation procedures for partial correlations are presented in Section 4.2.5. Essentially, the design of the networks aims to capture the dynamic trading patterns of cryptos. We update the partial-correlation network in high, time-frequency resolution to reflect intraday changes

within the partial-correlation structure.

4.2.1 Partial correlation

Partial correlations help to measure the conditional strength of a linear relationship between two random variables, given a set of other random effects. It is particularly essential to specify and disclose dependence among crypto exchanges, as one can later distinguish their joint dynamics. The partial correlation, ρ_t^{ij}, at hour t ($t \in [0, T]$) between crypto exchanges, i and j, is defined as follows:

$$\rho_t^{ij} = -\frac{\Omega_t^{ij}}{\sqrt{\Omega_t^{ii}\Omega_t^{jj}}}, \tag{4.1}$$

where $i, j \in \{1, \ldots, q\}$. Ω_t^{ij} denotes the (i, j)-element of the precision matrix, Ω_t, which equals the inverse of the covariance matrix, Σ_t, i.e., $\Omega_t = \Sigma_t^{-1}$. The partial correlation, ρ_t^{ij}, measures the relationship between the crypto exchanges, i and j, with indirect correlations removed. Alternatively, we can implement a multivariate regression to account for confounding factors; however, this provides the influence size rather than a numerical value that captures the strength of the relationship.

Here, Σ_t denotes the covariance matrix across exchanges at time t. The estimation of Σ_t is crucial, particularly for network analysis (Brownlees et al., 2018), portfolio selection, and risk management (Cai et al., 2020). We estimate Σ_t via the realized covariance matrix, Σ_t^{RC}, with the element equal to the sum of the products of high-frequency (e.g., 5-minute) returns within a given period (Barndorff-Nielsen & Shephard, 2004).

To ensure the positive semi-definiteness of the covariance-matrix forecast, Chiriac and Voev, 2011 applied the Cholesky decomposition to Σ_t, that is, $\Sigma_t = P_t^\top P_t$, where P_t is the Cholesky factor. They then constructed dynamics following the half-vectorization of the Cholesky factor and obtained forecasts of the covariance matrix through reverse transformation. Then, $X_t = \text{vech}(P_t)$, where $\text{vech}(\cdot)$ is a half-vectorization operator. Here, X_t is a $q(q + 1)/2$-dimensional vector. However, as Heiden, 2015 points out, the ordering of the variables in the Cholesky decomposition induces significant differences in forecast performance; see 4.A for an empirical illustration. To avoid the issue of ordering and to ensure positive semi-definiteness, Bauer and Vorkink, 2011; Hafner and Wang, 2021 estimate the covariance matrix in the matrix-log space. In the following, we discuss the matrix-logarithm transformation in detail.

4.2.2   The matrix-logarithm transformation

The matrix exponential and logarithm functions (Chiu et al., 1996) are useful for our analysis because they offer no parameter constraints to ensure proper covariance positive-definiteness. The matrix exponential function performs a power-series expansion on a real, symmetric matrix, S, i.e.,

$$\Sigma = \text{expm}(S) \overset{\text{def}}{=} \sum_0^\infty \frac{1}{n!} S^n. \tag{4.2}$$

This automatically guarantees a real, positive semi-definite covariance matrix, Σ. Meanwhile, S can be obtained by the inverse of the matrix exponential function, i.e., $S : S = \text{logm}(\Sigma)$, with logm the matrix-logarithm function. Accordingly, we can obtain the forecasts of the conditional covariance matrix on BTC returns following the three-step procedure below: First, for each hour, t, we use high-frequency data to construct the $q \times q$ realized covariance matrix, Σ_t^{RC}. Second, using the matrix-logarithm function on Σ_t^{RC}, we obtain S_t and a $p \times 1$ vector, s_t, with $p = \frac{1}{2}q(q+1)$, by stacking the elements, i.e.,

$$s_t = \text{vech}(S_t). \tag{4.3}$$

The vector, s_t, forms the basis for the subsequent modeling and application. In a later section, we present a factor model that allows the use of both lagged values of s_t and other variables to forecast volatility. Last, using the inverse of the vech function, we construct a $q \times q$ symmetric matrix, $\widehat{S}_t$, of the fitted values at each time t from s_t. Applying the matrix exponential function, we obtain the following:

$$\widehat{\Sigma_t} = \text{expm}\,\widehat{S}_t. \tag{4.4}$$

It yields a positive semi-definite matrix, $\widehat{\Sigma}_t$, which is our estimate of the conditional covariance matrix at time t. In practice, the matrix exponential and matrix logarithm transformations are implemented via the scaling and squaring and inverse scaling and squaring methods in Higham, 2008, respectively.

4.2.3   MHAR-CM

Regarding the stylized facts about BTC and other cryptos (e.g., trading 24/7 and the long-memory effect on returns), we extend the HAR long-memory model for realized volatility (Corsi, 2009). We refer to such a multivariate and high-frequency extension as MHAR-CM. The HAR-type model considers the log-realized volatility as a linear function of the log-realized volatility of the last day, week, and month to reflect trader's preferences on different horizons in the past, which is easy to implement. It succeeds
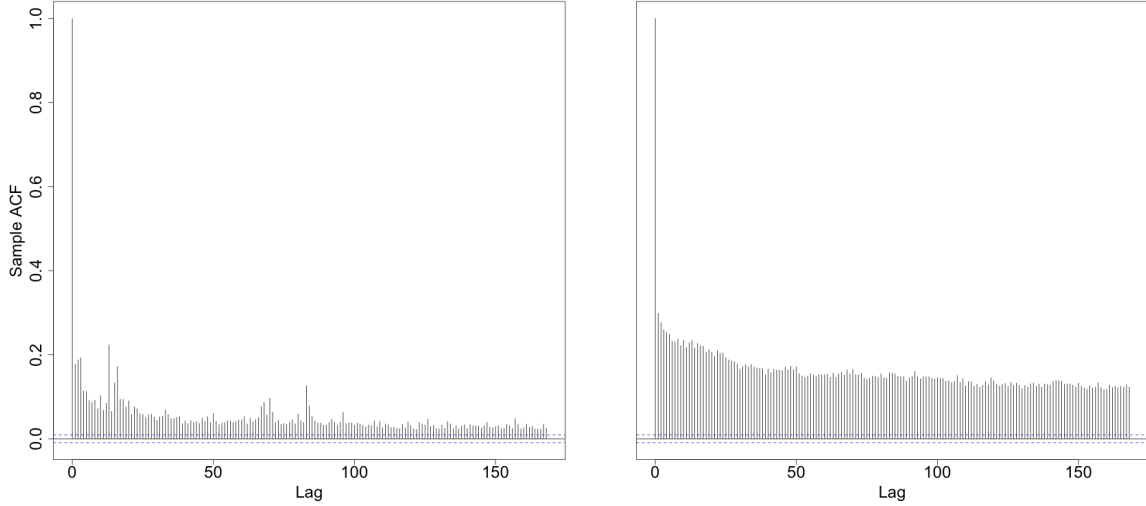
Figure 4.3: **ACFs of squared (left) and absolute (right) BTC hourly log returns in Binance.**

in reproducing the main empirical features of financial returns, e.g., long memory, fat tails, and self-similarity in a tractable and parsimonious manner. Accordingly, Bauer and Vorkink, 2011 offer a multivariate extension of HAR to model half-vectorized s_t as follows: $s_{t+1}^{(1)} = c + \beta^{(1)} s_t^{(1)} + \beta^{(5)} s_t^{(5)} + \beta^{(22)} s_t^{(22)} + \varepsilon_t$, where 1, 5, and 22 represent the frequencies of a day, a week ($= 5$ days), and a month ($= 22$ days), respectively; c is a $m \times 1$ vector of constants; and $\beta^{(\cdot)}$ are scalar parameters. $s_t^{(\cdot)}$ are averages of lagged daily volatility, i.e., $s_t^{(5)} = 1/5 \sum_{i=0}^{4} s_{t-i}$.

The most common implementation of HAR-type modeling uses daily data. It does not reflect the rapid changes in the crypto market. Moreover, the lagged volatilities here are not suitable for cryptos, as they are traded 24/7 and demonstrate a different trading pattern from that of traditional financial assets. Consequently, it motivates a more general method, MHAR-CM:

$$s_{t+1/24}^{(1/2)} = c + \beta^{(1/2)} s_t^{(1/2)} + \beta^{(1)} s_t^{(1)} + \beta^{(7)} s_t^{(7)} + \beta^{(30)} s_t^{(30)} + \varepsilon_t, \tag{4.5}$$

where the $s_{t+1/24}^{(1/2)}$ is the 1-hour-ahead forecast and $s_t^{(1/2)}$ is the half-day (12-hour) estimate.

4.2.4 Network and its centrality

Network-centrality measures are particularly salient in financial markets as they provide a way to understand the relative importance of financial institutions and thus help explain the propagation of shocks in the system. To calculate the centrality measures, we consider the partial-correlation matrix among exchanges in the form of an adjacency

matrix, which reflects the direct associations on the network, namely, $A_t = \{a_t^{i,j}\}_{i,j=1}^q$, where

$$a_t^{i,j} = \begin{cases} \rho_t^{i,j}, & i \neq j \quad \text{and} \quad |\rho_t^{i,j}| > \alpha \\ 0, & \text{otherwise} \end{cases} \tag{4.6}$$

α is the partial-correlation threshold. This procedure is supposed to reduce spurious connections, i.e., given $\alpha \neq 0$, there exists a disconnection. However, while thresholding is a good way to reduce the possibility of too much influence of spurious connections, it leads to confounds related to arbitrary choice of the calculation routines. There is no clear consensus on the statistical thresholds for the confirmation of a significant effect. Boginski et al., 2005 suggest $\alpha \geq 0.2$, which results in the degree (number of neighbors of a certain node) distribution resembling a power law. In this study, we consider partial correlations instead of correlations so that spurious connections are less likely to occur. We therefore choose a smaller threshold, $\alpha = 0.1$, for degree centrality. Note that A_t – viewed as the network – synthesizes the partial correlation structure of high-frequency returns changing by time t. We thus can dynamically describe the interdependencies without any indirect correlations. In particular, we can eliminate the correlation effect induced by the same types of assets, i.e., BTC traded across exchanges.

To describe the topology of networks and essentially quantify interconnectedness, we consider two commonly used centrality measures – degree and eigenvector centralities (Costenbader & Valente, 2003; Olmo, 2021; Peralta & Zareei, 2016). Degree centrality captures the total connectedness – the number of edges connected to a node – in a network and is conceptually the simplest. The degree centrality for exchange i is obtained as follows:

$$D_t^i = \sum_{j=1}^q \mathbf{1}(a_t^{i,j}), \tag{4.7}$$

where $\mathbf{1}(a_t^{i,j})$ is the indicator variable that equals to a if $a_t^{i,j} \neq 0$. We set $\alpha = 0.1$ for degree centrality as it preserves the major connections and omits the minor linkages.

Meanwhile, eigenvector centrality measures the transitive influence of nodes. Given the concept that high-scoring nodes contribute more to a certain node than connections with low scores, we can access the per-node influence. Avoiding losing information, α here is set as 0. The centrality of exchange i is defined as follows:

$$E_t^i = \lambda_t^{-1} \sum_{j=1}^q a_t^{i,j} E_t^j, \tag{4.8}$$

with λ_t the maximum eigenvalue for A_t. In the following, we use both centrality measures to examine the evolution of crypto exchange networks.

4.2.5   Overall procedure

We first calculate the covariance matrix, $\widehat{\Sigma}_t$, based on data acquired at time t. The matrix log-transformation yields $\widehat{S}_t = \text{logm}\,\widehat{\Sigma}_t$; then, by half-vectorization, we obtain $\widehat{s}_t = \text{vech}(\widehat{S}_t)$. The MHAR-CM framework,

$$s_{t+1/24}^{(1/2)} = \widehat{c} + \widehat{\beta}^{(1/2)}\widehat{s}_t^{(1/2)} + \widehat{\beta}^{(1)}\widehat{s}_t^{(1)} + \widehat{\beta}^{(7)}\widehat{s}_t^{(7)} + \widehat{\beta}^{(30)}\widehat{s}_t^{(30)} + \widehat{\varepsilon}_t, \tag{4.9}$$

produces a predicted value for the covariance. Last, we compute the covariance matrix, $\widehat{\Sigma}_{t+1/24}^{(1/2)}$, and the precision estimate, $\widehat{\Omega}_{t+1/24}^{\text{MHAR-CM}}$, by reverse transformation. Here, we forecast partial correlations as follows:

$$\rho_{t+1/24(ij)}^{\text{MHAR-CM}} = -\frac{\Omega_{t+1/24(ij)}^{\text{MHAR-CM}}}{\sqrt{\Omega_{t+1/24(ii)}^{\text{MHAR-CM}}\Omega_{t+1/24(jj)}^{\text{MHAR-CM}}}}. \tag{4.10}$$

Thus, we can construct the dynamic partial-correlation networks based on (4.6) and accordingly measure the centrality.

4.3   Empirical study

In this section, we investigate the network dynamics of the crypto exchanges, observing the returns on BTC across the $q = 9$ different exchanges.

4.3.1   Data and descriptive statistics

The dataset includes the hourly BTC closing prices of nine crypto exchanges – i.e., Kraken, Gemini, Bitstamp, FTX, Bitfinex, Exmo, CEX, Binance, and Kucoin – from August 03 00:00, 2020 to September 30 23:00, 2022. Each exchange comprises $18,936$ observations. The hourly BTC data for Gemini, Bitstamp, FTX, Bitfinex, Exmo, CEX Binance, and Kucoin are sourced from CryptoDataDownload and listed in Table 4.1. The Kraken data are collected from Kraken's REST API, which is currently updated to the end of 2022 Q3. Due to Kraken's data availability, the period for this empirical study ends in September 2022.

Among these exchanges, Kraken, Gemini, Bitstamp, FTX, Bitfinex, Exmo, and CEX are allowed to trade in USD. The remaining Binance and Kucoin only trade in different cryptos, without any fiat currency. They mainly use the stable coin, Tether (USDT), or their native coins as the fundamental trading medium. USDT, pegged to the USD and backed by Tether's liquidity reserves, has been highly traded and has retained a value close to that of the USD; see 4.B. For Binance and Kucoin, we thus use USDT

to exchange BTC. The hourly return is calculated as follows:

$$R_{i,t} = \log(p_{i,t}) - \log(p_{i,t-1/24}), \tag{4.11}$$

where $p_{i,t}$ is the hourly closing price of BTC on exchange i at time t. We exclude the observations in the period from August 2020 03 00:00 to November 02 11:00, i.e., the burn-in period for the network dynamics in (4.9). The rolling-window size for network construction is 30 days (i.e., we have 30×24 observations in each window), and the out-of-sample period is between November 02 12:00, 2020 and September 30 23:00, 2022.

Table 4.1: **BTC exchanges**.

Name	Country/Region	Trading Volume
Binance	Asia Pacific	71,887,560,454
FTX	Asia Pacific	11,260,464,103
Kucoin	Asia Pacific	995,236,044
Kraken	US UK	718,113,308
Bitfinex	US UK	662,990,566
Bitstamp	US UK	200,757,579
Gemini	US UK	75,406,613
Exmo	EU Russia	52,279,596
CEX	US UK	6,649,905

Note: We follow the country-region classification in CryptoDataDownload. In the last column, we report the trading volume on the last day in the sample period (September 30, 2022) sourced from nomics.

Table 4.2 shows the descriptive statistics for the Bitcoin returns for each exchange. We find that each exchange presents a different return behavior. Although the mean returns are similar, their medians differ considerably.

Table 4.2: **Summary statistics for BTC returns**.

Name	Currency	Med($\times 10^{-5}$)	Mean($\times 10^{-5}$)	Var($\times 10^{-5}$)	Skewness	Kurtosis
Binance	USDT	7.4761	2.8128	6.4535	-0.2345	11.6728
FTX	USD	8.3356	2.9416	6.6207	-0.3175	13.8636
Kucoin	USDT	7.4185	2.9396	6.4971	-0.2972	12.2049
Kraken	USD	8.9746	2.9416	6.6135	-0.3196	13.9767
Bitfinex	USD	7.4368	2.9215	6.4834	-0.2667	12.7466
Bitstamp	USD	1.1059	2.9440	6.6399	-0.3415	14.3515
Gemini	USD	6.9679	2.9409	6.5343	-0.1430	11.4551
Exmo	USD	5.1523	4.2416	6.0422	-0.2093	13.3215
CEX	USD	9.9799	2.9534	6.3578	-0.2181	13.9856

Note: Mean, var, and skewness denote the mean, variance, and skewness of BTC returns, respectively.

4.3.2   Model coefficient

Figure 4.4 illustrates the evolution of coefficient estimates in MHAR-CM. Compared with the longer time coefficients – $\beta^{(7)}$ and $\beta^{(30)}$ – the half-day and one-day coefficients, $\beta^{(1/2)}$ and $\beta^{(1)}$, are inconsiderable. This provides evidence of a long-memory effect in the realized covariance, and thus demonstrates the necessity of a HAR-type setting for BTC returns. Closely examining $\beta^{(7)}$ and $\beta^{(30)}$, there exist many upward and downward spikes, which relate to the strong influence from the covariance between returns. Interestingly, these spikes often relate to some extreme events in the BTC market. For example, the rise in $\beta^{(7)}$ and $\beta^{(30)}$ relates to the large-scale BTC withdrawals from the exchanges on 2020-10-26. The BTC plunged 30% on 2021-05-19, which induced an inflow to the exchanges for BTC; correspondingly, it is reflected in the coefficients. From November 2021, the BTC price starts reaching its historical highest peak – USD $69,000$; and it leads to a continuous growth in $\beta^{(30)}$. Later, in July 2022, BTC prices surge after a series of declines that correspondingly can be observed in $\beta^{(30)}$.

In conclusion, in the case of extreme events, the exchanges become more dependent, and such a dependence decays slowly. This is consistent with the finding by Brunnermeier and Sannikov, 2014 that in extreme event regimes, asset prices become more correlated, as well as with that by Geraci et al., 2018 that the relationship between short selling and price changes strengthens.

Focusing on each β, we plot them against non-negative and negative BTC returns in FTX as an instance for comparison in Figure 4.5. We find that negative shocks to returns are often followed by an increase in β. A high increase in $\beta^{(30)}$ shows that the long-memory effect of these negative shocks is rather strong; see the last panel of Figure 4.5. This is consistent with the discussion in Assaf et al., 2022 and can also be related to the leverage effect that the impact of shocks is asymmetric. A possible explanation here is that the cumulative risk leads to the larger $\beta^{(7)}$ and $\beta^{(30)}$ as market participants' expectations influence the market movement and propagate among investors, i.e., a self-fulfilling prophecy. The matching dynamics between the extremes in $\beta^{(7)}$ and $\beta^{(30)}$ and the BTC market dynamics (i.e., extreme events and return behaviors) is evidence that the dynamic model designed in this study captures the BTC market changes.

4.3.3   Analysis of partial correlation

To explore the evolution of the partial correlations among these exchanges, we use Binance and FTX as an example and illustrate their partial correlations with the other 8 exchanges as a series of monthly heatplots for different periods – i.e., February 2021, May 2021, December 2021, and June 2022; see Figures 4.6 and 4.7. During May 2021
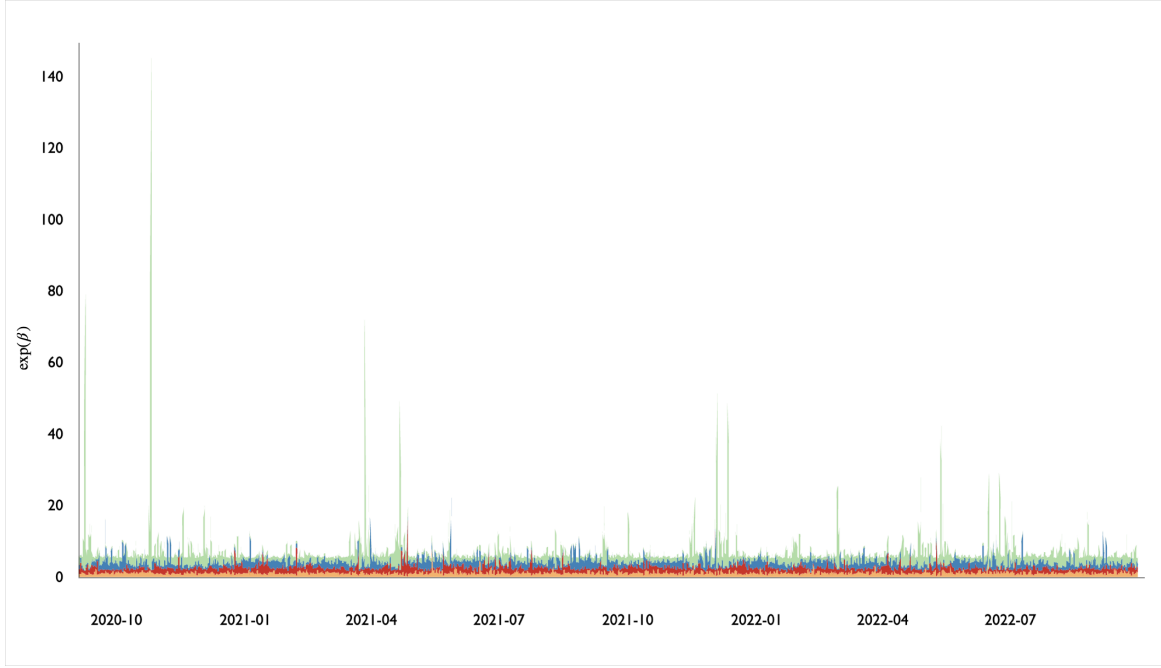
Figure 4.4: **Streamgraph for coefficients** $\beta^{(1/2)}$, $\beta^{(1)}$, $\beta^{(7)}$, **and** $\beta^{(30)}$. We take exponents of the coefficients for visualization. $\beta^{(7)}$ and $\beta^{(30)}$ present larger variations.

and December 2021, there exists an extreme event condition, as discussed in Section 4.3.2. Note that the green here illustrates that their partial correlations are nearly zero. The depth of the blue and red colors of the boxes indicates that the pair of exchanges has a higher linkage at that moment. We first investigate the – Hour of the Day" and "Day of the Month effects – (Baur et al., 2019; Kinateder & Papavassiliou, 2021) on the x- and y-axes, respectively. There is no solid evidence that supports the presence of the Hour of the Day and Day of the Month effects on the partial correlations. This might be attributed to the fact that the market never sleeps and traders worldwide are active for the whole day (Vidal-Tomás, 2021); in particular, trading bots are popular in the crypto market. This result shows a development in trading behavior from Petukhina et al., 2021 that crypto trading is firmly in the hands of humans. Examining the heatplots by rows, there is a sequence of boxes with a similar degree of blue or red, which indicates the presence of a long-memory effect.

Considering Binance's heatplots relative to the other 8 exchanges in Figure 4.6, Kucoin and FTX present a strong positive correlation. This may reflect their firm sizes; see Table 4.1. Note that both Binance and FTX favor USDT, which causes a strong dependence; see Table 4.2. Kraken, however, often has a negative partial correlation with Binance. During the extreme events, in the second and third panels, Binance's partial correlations with the other firms are relatively vigorous. In the case of FTX, it has some negative partial correlations with the Exmo exchange. In May 2021, FTX's linkage to the other firms, except Binance, is evident. Overall, each exchange has a
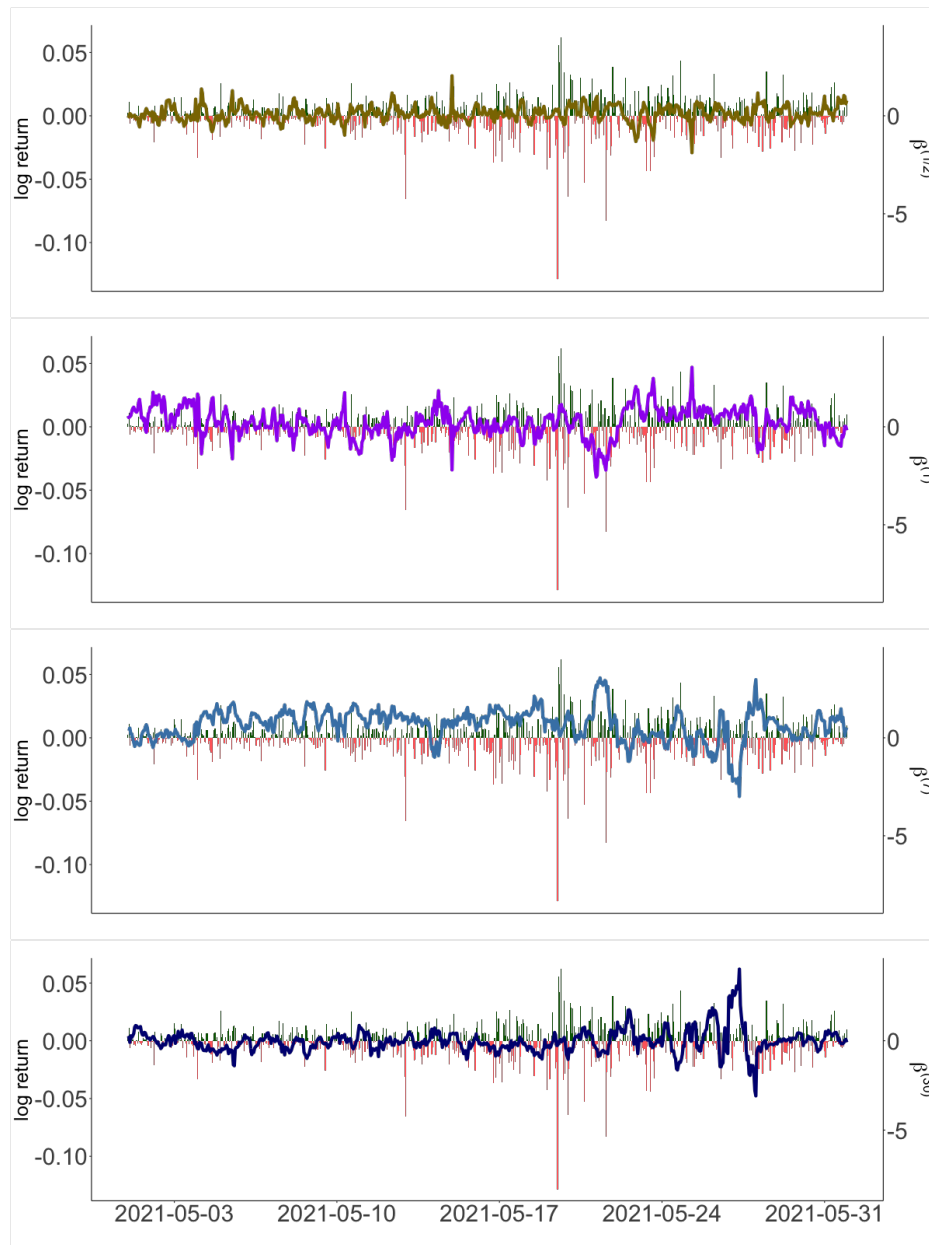
Figure 4.5: **Coefficients** $\beta^{(1/2)}$, $\beta^{(1)}$, $\beta^{(7)}$, **and** $\beta^{(30)}$ **with nonnegative and negative log BTC returns on FTX (background).** The returns are from 2021-05-01 to 2021-05-31, including 2021-05-19, when the BTC surged 30%.

different degree of dependence on the others; moreover, during an extreme event, the dependence can increase. This indicates that the BTC market is sensitive to negative shocks, which leads to a challenge regarding risk diversification via portfolios. The current result is consistent with the finding by Acemoglu et al., 2015 that, when the price drops lower than a certain point, dense interconnections serve as a mechanism for the propagation of shocks, which causes a more fragile market as a whole. As shown in the color-changing heatplots, the relationships between these exchanges are unstable, which reflects a rapidly shifting condition of the market and confirms the necessity of examining the network in a high-frequency fashion.

4.3.4   Network dynamics

We examine the network connectedness using the MHAR-CM model for the 9 centralized exchanges, as well as the persistence and spillover of risk. To depict the evolution of the networks and further compare them, we plot the networks for the periods discussed in Section 4.3.3. In Figures 4.8 - 4.11, each panel on the left is the partial-correlation heatmap, while the corresponding network is on the right. On the networks, the thicker the line, the stronger the paired exchanges' connectedness. We summarize the network centralities for these four periods in 4.D.

During extreme events when large BTC drops occur, as in Figures 4.9 and 4.10, most of the exchanges are connected to one another. Thus, each exchange has a rather high degree of centrality, which refers to an exchange's many direct linkages to others. Regarding eigenvector centrality, the exchanges with higher trading volumes have higher values, e.g., Binance, while those with lower trading volumes obtain lower values, e.g., Exmo. That is, a large-scale exchange is more influential in a time of crisis than under normal conditions, as in Figures 4.8 and 4.11; dense linkages among firms induce a higher risk spillover in the market. Note that a firm with a low trading volume or in a more isolated position in the market is less influential during a crisis; however, it does not mean that such a firm is riskier itself. Exmo and CEX retain the linkages with the others; however, both have lower eigenvector centrality values. Thus, these two exchanges may act as risk receivers and accumulate the risk from the other exchanges during a crisis.

In summary, we report the descriptive centrality statistics for each exchange in the period from 2020-09-02 12:00 to 2022-09-30 23:00 in Table 4.3. The results for both centrality measures in terms of rankings are consistent. Overall, Kraken, which offers the most fiat trading pairs (i.e., AUD, EUR, USD, and GBP) occupies a prominent position. Exmo and CEX are in lower ranked positions, i.e., they are less influential within the network. In addition, FTX has a relatively high variation in eigenvector
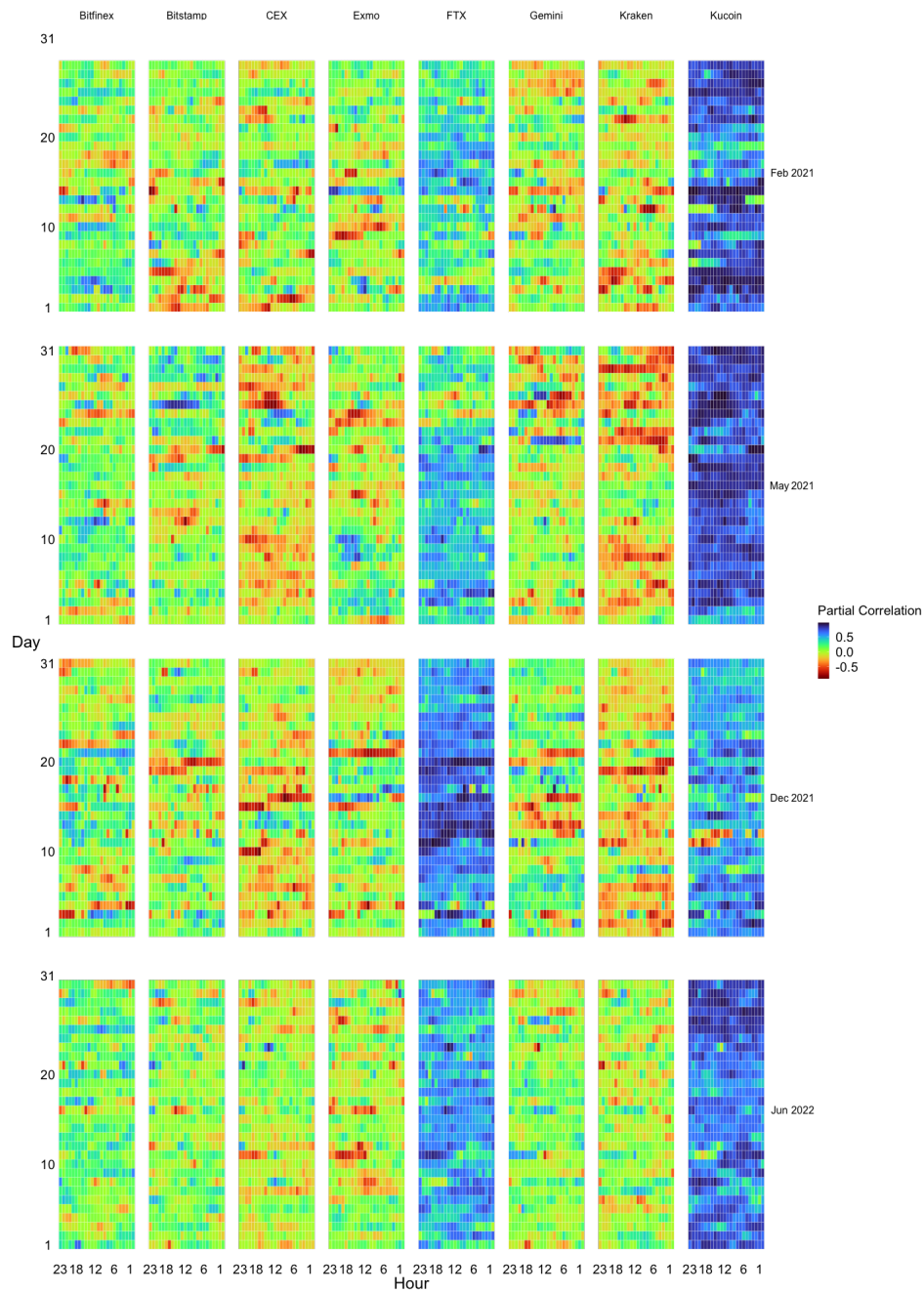
Figure 4.6: **Heatplots of partial correlations of Binance with FTX, Kucoin, Kraken, Bitfinex, Bitstamp, Gemini, Exmo, and CEX.**
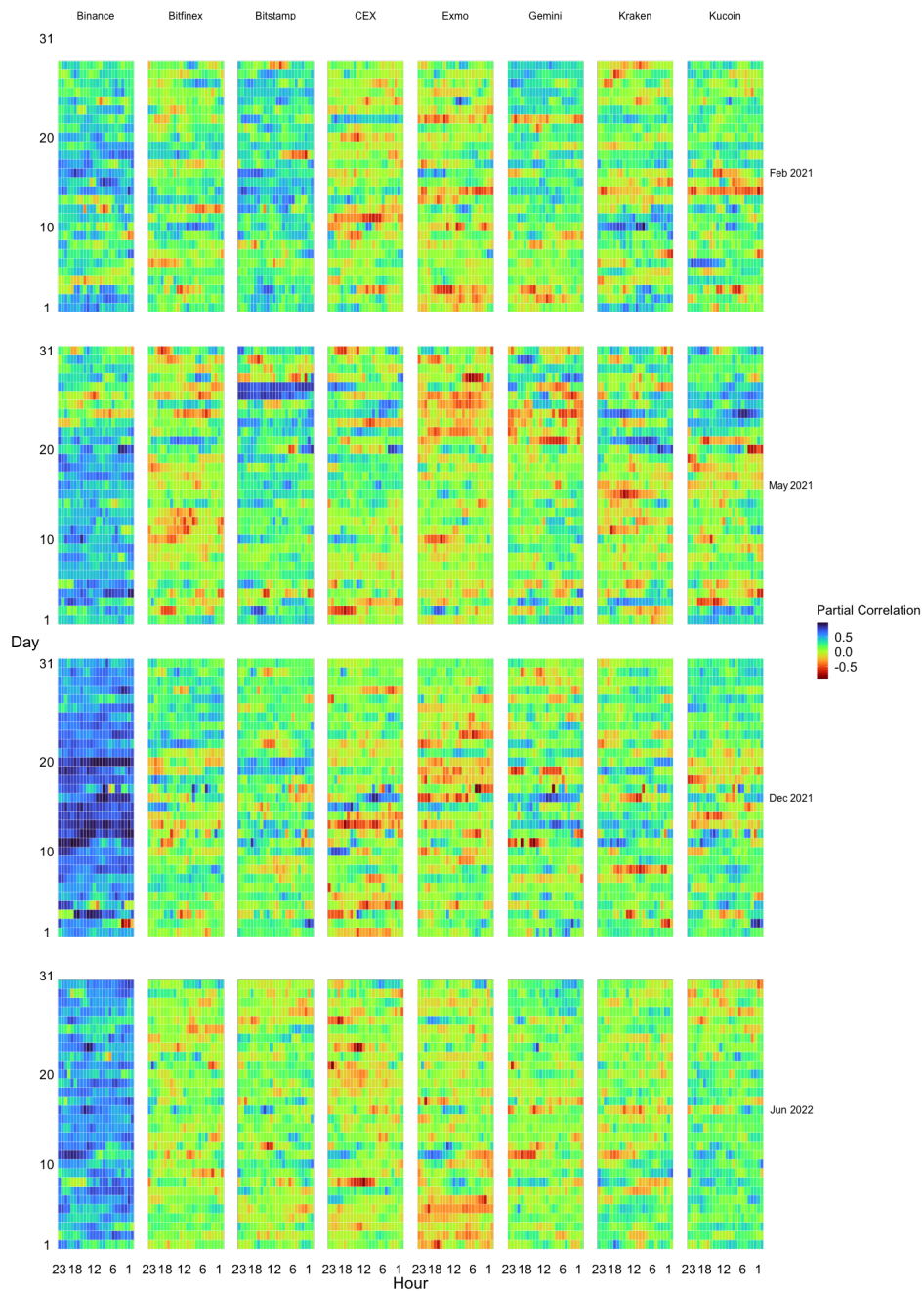
Figure 4.7: **Heatplots of partial correlations of FTX with Binance, Kucoin, Kraken, Bitfinex, Bitstamp, Gemini, Exmo, and CEX.**
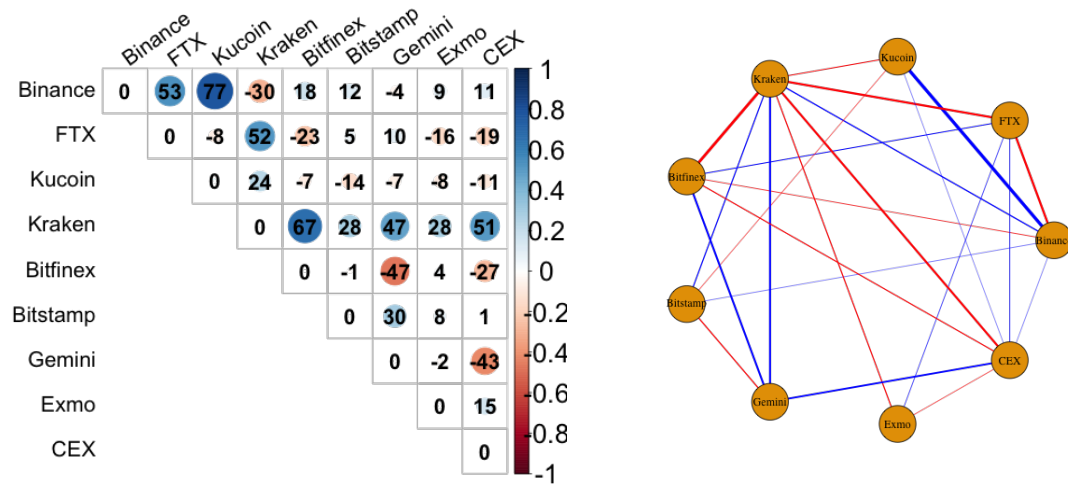
Figure 4.8: **Partial-correlation network on 2021-02-12 05:00:00.**
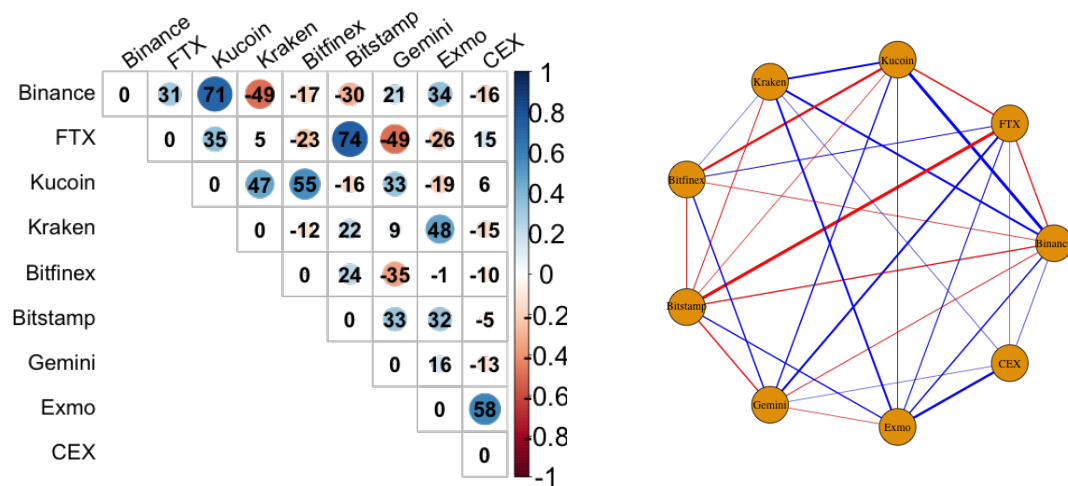
Figure 4.9: **Partial-correlation network on 2021-05-27 09:00:00.**
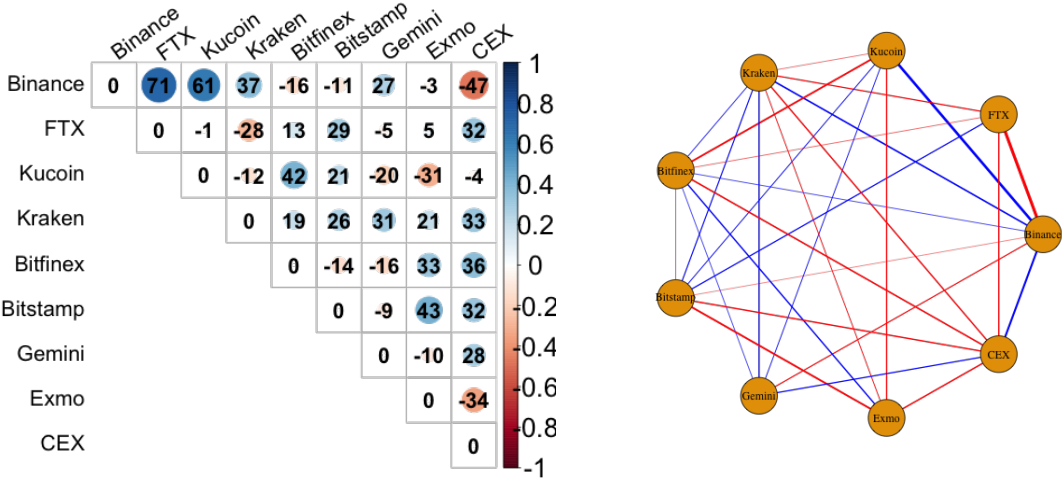
Figure 4.10: **Partial-correlation network on 2021-12-04 04:00:00.**
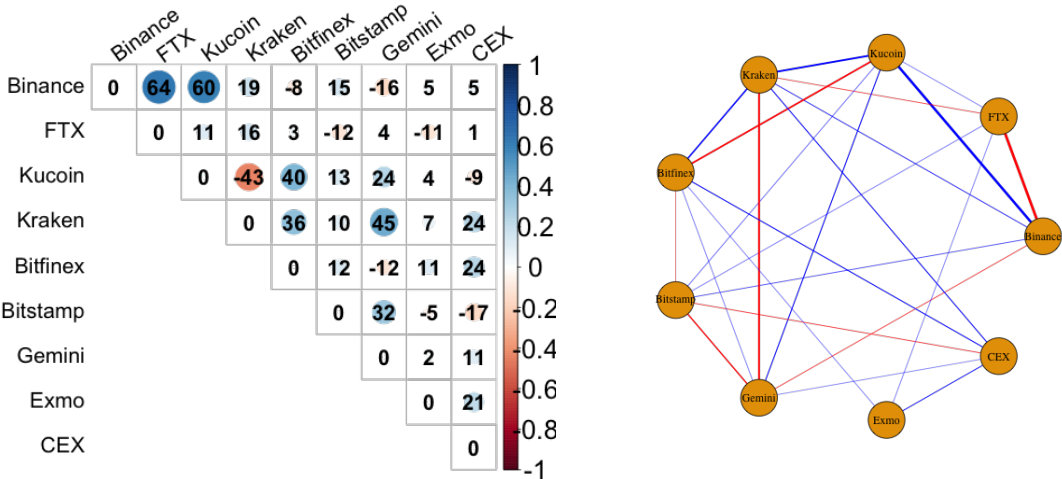
Figure 4.11: **Partial-correlation network on 2022-06-04 09:00:00.**

centrality.

Table 4.3: **Centrality of each exchange**.

Exchange	Degree			Eigenvector		
	Mean	Median	Std.	Mean	Median	Std.
Binance	4.175	4	1.424	0.516	0.511	0.296
FTX	5.207	6	1.557	0.515	0.487	0.323
Kucoin	4.215	5	1.481	0.462	0.452	0.271
Kraken	5.951	7	2.332	0.782	1.000	0.321
Bitfinex	3.628	4	1.933	0.505	0.509	0.297
Bitstamp	4.657	5	2.156	0.441	0.432	0.293
Gemini	5.302	6	2.284	0.502	0.500	0.318
Exmo	1.373	1	1.501	0.326	0.358	0.314
CEX	1.374	1	1.381	0.343	0.379	0.298

4.3.5   FTX: a node with volatile centrality

BTC and other cryptos generally purport to offer a trustless environment for users and investors through the decentralized architecture of distributed ledger technologies, e.g., blockchain. The idea of centralized exchanges for cryptos negates decentralization and invokes third-party risk. Centralized exchanges are subject to varying degrees of operational and regulatory risks based on their corporate management and the jurisdictions in which they operate. Based on how they source liquidity (e.g., in-house reserves and liquidity provider), exchanges present different dependencies on others. That is, each exchange experiences different levels of risk exposure. The collapse of FTX, one of the world's largest crypto exchanges, on November 11, 2022, is crucial evidence here. This highlights the significance of this study.

The wildfire that was ignited by FTX's bankruptcy spread throughout the market and led to an 120% uptick in average daily trading volume (Ng, 2022). It benefited Binance, which now dominates with a 64% market share across the Top 10 crypto exchanges. Due to data limitations, we could not access FTX's BTC prices around the time of its bankruptcy. Based on past dynamic partial correlation networks, we provide some remarks on FTX below. We rank each exchange by its monthly centrality aggregated by mean in Figure 4.12. The higher an exchange's centrality value, the higher its rank. In terms of degree-centrality rankings, FTX is in a rather volatile position. Based on eigenvector centrality, the three firms with the highest trading volumes – Binance, Kucoin, and FTX – remain influential over time. That is, their risk spillovers are significant to the market. We illustrate the centrality rankings aggregated by daily mean in 4.C, in which each firm's influential position within the network over time can easily be observed. Specifically, we plot each firm's eigenvector centrality in

the last year of our sample period as a monthly boxplot in Figure 4.13. The largest trading-volume exchange, Binance, has a rather stable eigenvector centrality, with its median around 1, while FTX's and Kucoin's substantially fluctuate over time. The remaining firms' centralities are less disperse.
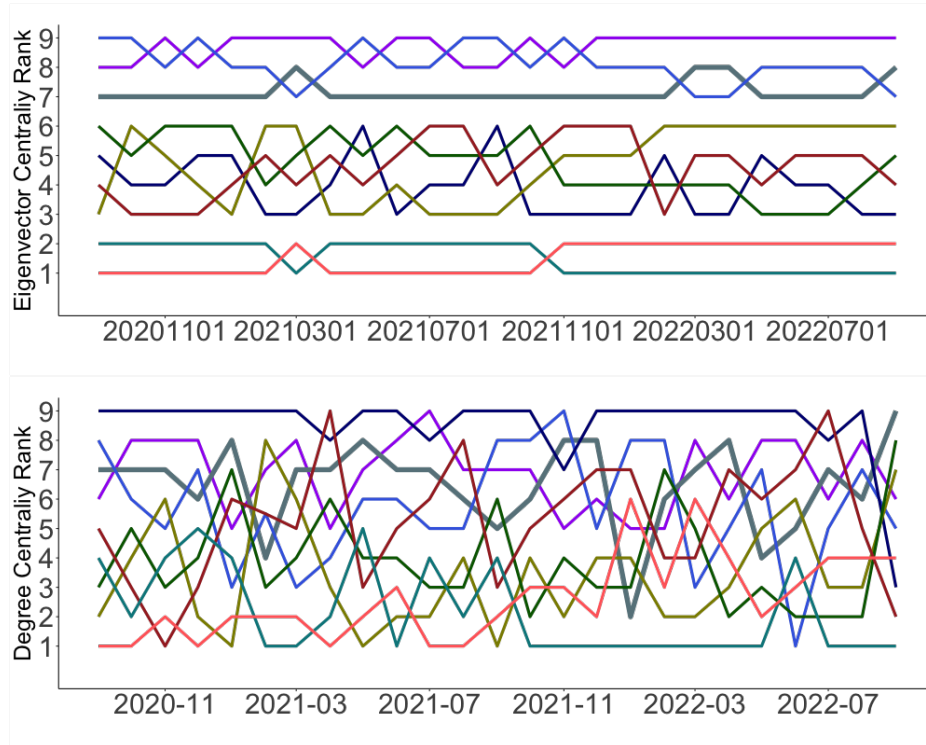
Figure 4.12: **Monthly centrality rank of Binance, FTX, Kucoin, Kraken, Bitfinex, Bitstamp, Gemini, Exmo, CEX.**

Issue of centralized decentralization. We discuss how centralized exchanges differ from decentralized ones below. Centralized crypto exchanges rely on clearing trades and off-chain scaling. Barbon and Ranaldo, 2021 point out that crypto prices in centralized exchanges are more efficient than those in decentralized ones; however, there exist significant risks and latency associated with delegated custody. Centralized exchanges operate under a trading infrastructure and execution rules, similar to traditional asset exchanges, to maintain liquidity provision and the price discovery process. Thus, a centralized exchange is easily adopted and offers advanced asset-management tools and financial products while retaining some comparable characteristics to traditional asset exchanges – i.e., control on investors' funds, non-anonymity, fragile cybersecurity, and sever downtime – as a single point of failure. The unregulated and volatile nature of cryptos, compounded by exchange counterparty risk exposure, subjects investors to significant risk of loss. One can possibly observe an arbitrage opportunity here; however, centralized exchanges may act as the Frankenstein, with high spillover effects on such compounded risks, especially during a crisis.
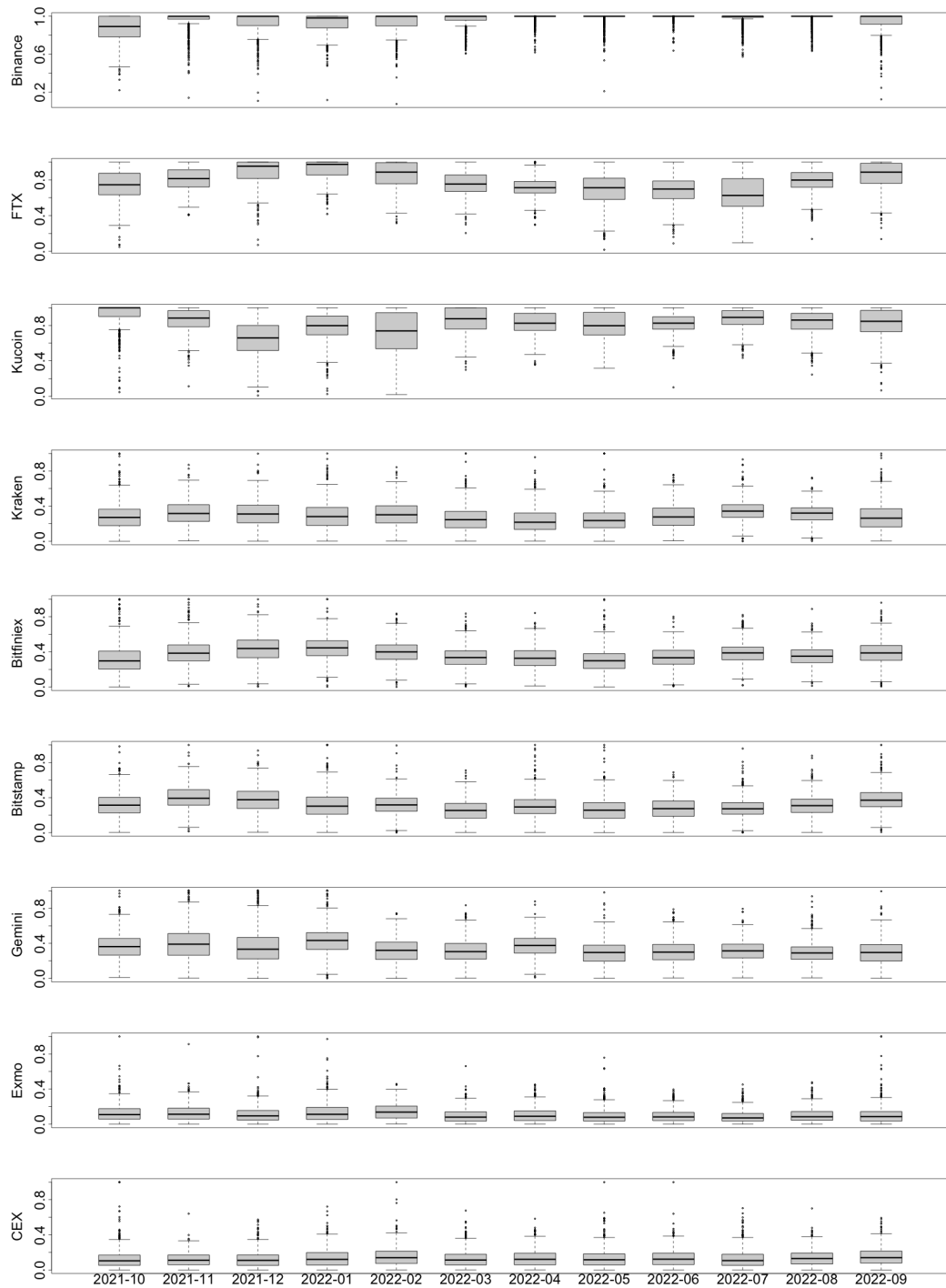
Figure 4.13: **Eigenvector centralities for Binance, FTX, Kucoin, Kraken, Bitfinex, Bitstamp, Gemini, Exmo, and CEX from September 2021 to September 2022.**

4.4   Portfolio construction

Because it is difficult to achieve cross-exchange arbitrage even if investors store abundant cryptocurrencies in various exchanges, this study incorporates network knowledge into risk diversification via portfolio allocation on exchanges. We consider the following portfolio-allocation strategies in this section: The first is based on portfolio-risk considerations, a hierarchical risk parity (HRP) portfolio with two variants to determine the weights in this study – using HRP_MHAR_CM and realized partial correlations (HRP_RCor). There exists rich literature on risk diversification for asset allocation. HRP (De Prado, 2016) is among the most frequently used methods in practice. This approach overcomes the limitations of other risk-based optimization approaches, such as global minimum variance, equal-risk contribution, and risk budgeting, by generating portfolios on an ill-conditioned or even singular matrix. In addition, it has been considered to be a robust approach in that it is less susceptible to noise (Jaeger et al., 2021). The next is the network-based strategy – inverse eigenvector centrality portfolio (IECP) – that uses eigenvector centrality to assign weights to each exchange. We do not consider it for the portfolio construction to ensure complete network information as there is a threshold setting for degree centrality, i.e., $\alpha = 0.1$. The last is an equally weighted (EW) portfolio, whose weights here are assigned by $1/q$ and is used as a benchmark for evaluation. Based on the burn-in period in Section 4.3.1, we exclude the observations in the evaluation. The back-testing period, therefore, is from November 02 12:00, 2020 to March 31 23:00, 2022.

4.4.1   HRP

HRP is a risk-based portfolio-optimization approach that diversifies portfolios without imposing a positive-definite return covariance matrix. Classical HRP is based on variance and covariance matrices. However, the assets considered may share a common cause (e.g., BTC prices on the exchanges) and be correlated, resulting in inference of spurious relationships. Extending this classical HRP to partial correlations that encode the core associations between two random variables excludes a common causality. Thus, one can access all the effective information when optimizing a cross-exchange portfolio. The HRP algorithm has three main steps: tree clustering, quasi-diagonalization, and recursive bisection. We explain each step in detail in the following.

Step 1- Hierarchical Tree Clustering. We group the assets into different hierarchical clusters following the procedures below.

(a) Estimate the $q \times q$ partial-correlation matrix by (4.10) $\widehat{\rho}_{t+1/2(ij)}^{\mathrm{MHAR\text{-}CM}}$ and $\widehat{\rho}_{t+1/2}^{\mathrm{MHAR\text{-}CM}} = \{\widehat{\rho}_{t+1/2(ij)}^{\mathrm{MHAR\text{-}CM}}\}_{i,j=1}^{q}$.

(b) Transform the partial-correlation matrix to a distance matrix, D, where, for d : $(X_i, X_j) \subset B \to R \in [0, 1]$ and each element $d_{i,j} = d[X_i, X_j] = \sqrt{\frac{1}{2}(1 - \widehat{\rho}_{t+1/2(ij)}^{\text{MHAR-CM}})}$.

(c) Compute a new distance index, $\tilde{d}$, by taking the pairwise Euclidean distances among the columns in D; the augmented distance matrix is given by

$$\tilde{d} = \tilde{d}[d_i, d_j] = \sqrt{\sum_{n=1}^{q}(d_{n,i} - d_{n,j})^2},$$

where $\tilde{d}_{i,j} : (d_i, d_j) \subset B \to R \in [0, \sqrt{q}]$. Note that for two exchanges, i and j, $D_{i,j}$ denotes the distance between them and $\tilde{d}_{i,j}$ denotes the closeness in similarity of $\{i, j\}$ relative to the remaining assets in the portfolio. More precisely, a lower $\tilde{d}_{i,j}$ indicates that the assets, i and j, are similarly correlated to the others in the portfolio.

(d) Recursively form the clusters of assets via Sub-step (c). The set of clusters is donated by U, with the first cluster formed, (i^*, j^*), defined as follows:

$$U[1] = (i^*, j^*) = \text{argmin}(i, j)_{i \neq j} \tilde{d}_{i,j}.$$

(e) Update the distance matrix, d, by computing other assets' distances from the newly formed cluster, $U(1)$, using single-linkage clustering. For any asset, i, outside $U(1)$, its distance to $U(1)$ is updated using $d_{i,U[1]} = \min\{\tilde{d}_{i,j}\}_{j \in U[1]}$.

Thus, the HRP algorithm recursively forms clusters and updates the distance matrix until there exists only one cluster with all the assets.

Step 2- Quasi Diagonalization / Matrix Seriation. Following the previous step, we re-sort the columns and rows of the covariance matrix, placing similar assets together. Specifically, we arrange the larger covariances along the diagonal of the matrix and the smaller ones around this diagonal. The off-diagonal elements are not completely zero. Such a matrix is the so-called quasi-diagonal covariance matrix.

Step 3- Recursive Bisection. Last, we assign the actual portfolio weights to each exchange in the portfolio.

(a) Assign a unit weight to all assets, i.e., $W_i = 1 \quad \forall i = 1, \cdots, q$.

(b) Bisect each cluster into two sub-clusters in a top-down order, such that each cluster has left and right sub-clusters.

(c) Calculate the variance for each sub-cluster using $V_{1,2} = w^\top \Sigma w$, where $w = \frac{\text{diag}[\Omega]}{\text{tr}[\text{diag}[\Omega]]}$ and Σ and Ω are the covariance and inverse covariance matrices, respectively. Because we are dealing with a quasi-diagonal matrix, the algorithm

uses the portfolio's property that the inverse-variance allocation is optimal for a diagonal covariance matrix. Thus, we adopt the inverse-variance allocation weights while computing the variances for the sub-clusters.

(d) Calculate the weighting factor based on the quasi-diagonalized covariance matrix as $\alpha_1 = 1 - \frac{V_1}{V_1+V_2}$, such that $0 \le \alpha_i \le 1$ and $\alpha_2 = 1 - \alpha_1$.

(e) Update the weights, w_1 and w_2, for both sub-clusters using $w_1' = \alpha_1 w_1$ and $w_2' = \alpha_2 w_2$.

(f) Repeat Sup-steps (b) - (e) and terminate when there is one asset for each cluster. We then assign the corresponding weight to each asset in the portfolio.

As each weight is assigned in a top-down order, only the assets within a cluster are included instead of all the assets in the portfolio; see Vỳrost et al., 2019. The above steps are for HRP_MHAR_CM. For HRP_RCor, we follow the same steps above, except that we substitute $\widehat{\rho}_{t+1/24(ij)}^{\mathrm{MHAR\text{-}CM}}$ with

$$\rho_{t+1/24(ij)}^{\mathrm{RCor}} = -\frac{(\Sigma_t^{\mathrm{RC}})_{(ij)}^{-1}}{\sqrt{(\Sigma_t^{\mathrm{RC}})_{(ii)}^{-1}(\Sigma_t^{\mathrm{RC}})_{(ii)}^{-1}}},$$

where $(\Sigma_t^{\mathrm{RC}})_{(ij)}^{-1}$ is the (i,j) th element in the inverse realized covariance matrix, $(\Sigma_t^{\mathrm{RC}})^{-1}$.

4.4.2   Network-based Strategy

In this section, we develop an asset-allocation strategy, IECP, based on the network topology of the exchanges, specifically through eigenvector centrality. Peralta and Zareei, 2016 show that eigenvector centrality presents a negative relationship to optimal portfolio weights and point out that considering the underlying structure of the financial-market network is an effective tool in enhancing the portfolio-selection process. A related argument can also be found in (Olmo, 2021). The idea behind this approach is simple: the more volatile and centralized an exchange, the less weight it is allocated. The allocated weight for each exchange, i, at time t is given by (Jaeger & Marinelli, 2022):

$$w_{i,t}^{\mathrm{IECP}} = \frac{(\widehat{\sigma}_t^{i,i} E_t^i)^{-1}}{\sum_{n=1}^{q} (\widehat{\sigma}_t^{n,n} E_t^n)^{-1}},$$

where E_t^i is the eigenvector centrality of exchange i at time t defined in Section 4.2.4, and $\widehat{\sigma}_t^{i,i}$ is the (i,i) th element in the estimated covariance matrix, $\widehat{\Sigma}_t$. Exchanges that are strongly embedded in a correlation-based network greatly affect the market. Their inclusion in a portfolio undermines the benefit of diversification, resulting in larger variances, Sharpe ratios, or volatilities, depending on the specific portfolio objective.

4.4.3 Portfolio evaluation

We compute the hourly updated daily return, $R_{i,t} = \log(p_{i,t}) - \log(p_{i,t-1}), t \in [0, T]$, for exchange i and portfolio return, $R_{p,t} = \sum_{i=1}^{N} w_{i,t} R_{i,t}$, due to the confirmation time for cross-exchange trading considered by Borri and Shakhnov, 2022. In this study, we consider a variety of performance measures to evaluate each portfolio. Note that the measures in the second panel in Table 4.4 consider the downside risk.

Table 4.4: **Portfolio performance**.

	σ_p ($\times 10^{-3}$)	Sharpe ($\times 10^{-3}$)	δ_0 ($\times 10^{-3}$)	Sortino ($\times 10^{-3}$)	MDD ($\times 10^{-3}$)	Calmar ($\times 10^{-3}$)
EW	38.960	*19.300	27.800	*27.060	1000.000	*6585.77
HRP_MHAR_CM	8.085	4.000	5.809	5.561	790.439	357.972
HRP_RCor	8.093	3.770	5.815	5.250	797.038	335.538
IECP	*7.995	4.077	*5.733	5.686	*781.235	365.526

Note: * means the corresponding portfolio performs best in the specific measure (the column). σ_p and δ_0 denote the standard deviation and semi deviation of portfolio returns, respectively. See 4.E for details.

Table 4.4 shows that HRP_MHAR_CM outperforms HRP_RCor in all the risk measures, including the standard deviation, σ_p, assuming a symmetrical return distribution, the semi deviation, δ_0, as a downside risk measure, and maximum drawdown (MDD). The deficiencies of the Sharpe ratio due to its symmetry property and inability to consider the risk of low-probability events are well known. Thus, the two other alternatives, the Sortino and Calmar ratios, are included. The risk-adjusted measures – the Sharpe, Sortino, and Calmar ratios – reinforce the conclusion, which provides empirical evidence that the method proposed captures the exchange-specific risk as well as the dynamic network structure. Although the EW portfolio yields higher risk-adjusted ratios, this occurs at the cost of a large MDD. In practice, an EW portfolio may encounter liquidity problems and cannot finance itself in the long term. The IECP achieves the least risk among the portfolios considered, in terms of σ_p, δ_0, and MDD. It outperforms the two HRP portfolios in risk-adjusted returns such as the Sharpe, Sortino, and Calmar ratios. In other words, considering eigenvector centrality assists in risk diversification.

Specifically, Figure 4.14 illustrates the updates of FTX's weights within a week using the four different strategies in the upper panel and their corresponding eigenvector centralities in the lower panel. The EW portfolio maintains equal weights over time. For the IECP, the weights resemble the dynamics of the eigenvector centralities, whereas those in the two HRP portfolio variants do not. HRP_MHAR_CM changes more frequently than HRP_Rcor. HRP_Rcor uses realized partial correlations, and thus is less responsive to a new observation and does not promptly reflect the evolution of covariances. HRP_MHAR_CM estimates each covariance matrix under an additive

cascade structure from the short to the long term. Consequently, it responds to such an evolution.
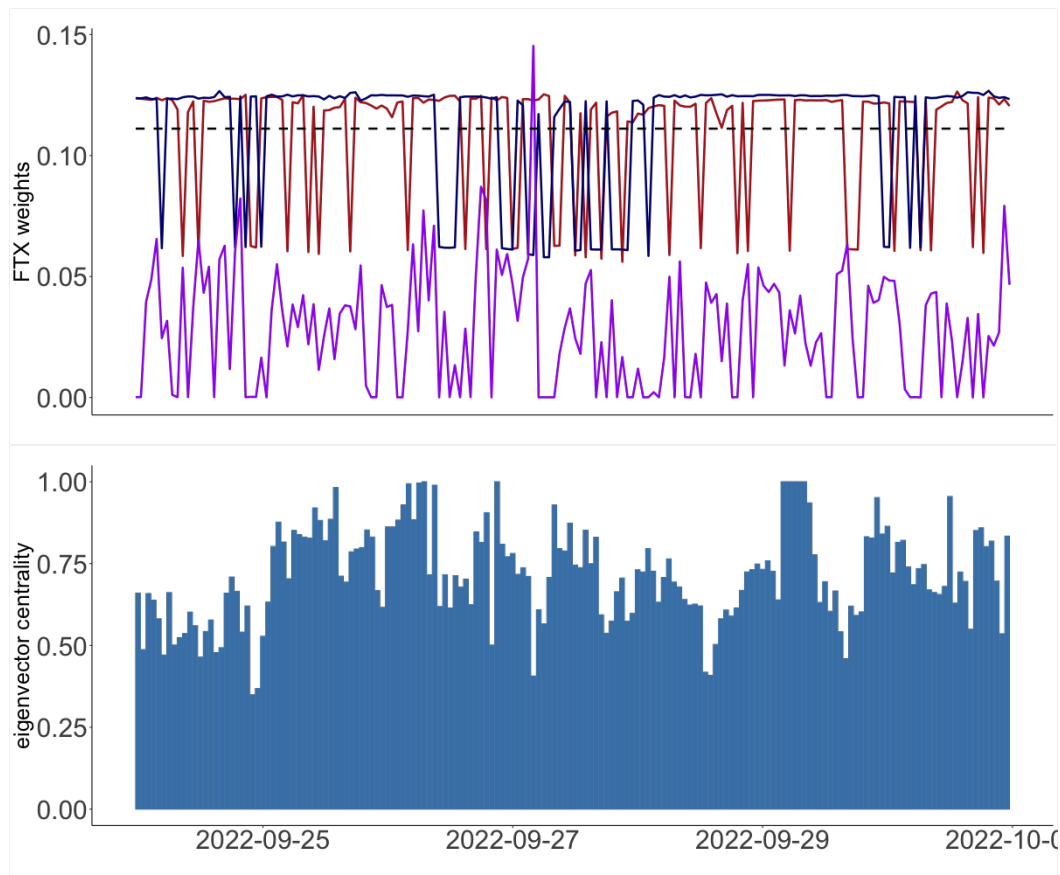
Figure 4.14: **Weights of FTX in HRP_MHAR_CM, HRP_Rcor, IECP and EW (dashed); and eigenvector centralities of FTX.**

To highlight the two HRP portfolio variants, we illustrate the dendrograms, showing their portfolio structures; see 4.F. On the dendrograms, we find that HRP_MHAR_CM is relatively distinct from the classic HRP_RCor approach in terms of the Euclidean distance. That is, using partial correlations for HRP can better cluster the assets, offering a more identifiable distance matrix. Last, there exists a large body of literature on asset allocation for multiple cryptos as well as cryptos with traditional investment assets, e.g., stocks. This study, in comparison, focuses on the risk within cross-centralized exchanges. To better discuss such risk and limit the influence from heterogeneous cryptos, we consequently focus on one common crypto – BTC – for all the exchanges. However, such a setting cannot guarantee a promising result. As discussed in Borri, 2019, a portfolio of multiple coins improves risk-adjusted and conditional returns and offers hedging properties for investors. This study provides a distinctive perspective on how to achieve risk diversification, i.e., among exchanges rather than among assets, which complements the stream of studies that constructs portfolios on different cryptos.

4.5 Conclusion

As the growing popularity of crypto-asset investment sparks an increased demand for trading platforms, many marketplaces are springing up. Avoiding the volatile gas fee and complex operation of cross-chain trading, centralized exchanges enable a straightforward trading mechanism for investors that closely resembles traditional asset exchanges. However, the mechanism encounters a compounding threat of crypto volatility and counterparty risk. Using MHAR-CM designed for the crypto market, we consider the stylized facts of BTC, i.e., trading 24/7, long-memory effect, and high-frequency data, and find that the monthly coefficients depict the evolution of returns, especially during a crisis. In light of dynamic partial-correlation networks, a firm's scale and the economic condition of the crypto market highly relate to each exchange's connectedness with the others. In extreme situations (e.g., large price drops and hacks on exchanges), there are high and persistent partial correlations that decay more slowly than those under normal conditions. Given the degree and eigenvector centralities over time, we better position each firm on the network. Higher trading-volume exchanges are more influential; that is, their risks can be easily propagated to others. Considering the example of FTX, the influence of its bankruptcy can definitely trigger a chain effect among exchanges, i.e., increased partial correlations with other exchanges. We also find that, in the networks, FTX was very often volatile. In conclusion, this study highlights the spillover risk and its persistence effect across the centralized exchanges. It is necessary to diversify risk and consider an exchange's market position while conducting cross-exchange trading on cryptos, considering the potential loss induced by processing delays or unused funds. Due to the accuracy in the measurement of correlation structure among the exchanges, the use of partial correlations in portfolio construction further offers a better diversification of risk. In particular, HRP_MHAR_CM outperforms the realized partial-correlation counterpart for downside risk-adjusted measures. Utilizing the eigenvector centrality of partial-correlation networks in weight allocation, the IECP achieves a better result in both risk-adjusted returns and risk measures. The EW portfolio performs the best in all risk-adjusted returns while it is far worse than the others in risk measures, implying considerable risk exposure for an investor adopting such a strategy across exchanges, which may lead to enormous future loss.

In this study, we only include the centralized exchanges. Future studies may consider decentralized exchanges and investigate the return dynamics of both types of exchanges.

Bibliography

Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, *105*(2), 564–608.

Alexander, C., & Heck, D. F. (2020). Price discovery in bitcoin: The impact of unregulated markets. *Journal of Financial Stability*, *50*, 100776.

Andrieş, A. M., Ongena, S., Sprincean, N., & Tunaru, R. (2022). Risk spillovers and interconnectedness between systemically important institutions. *Journal of Financial Stability*, *58*, 100963.

Assaf, A., Bhandari, A., Charif, H., & Demir, E. (2022). Multivariate long memory structure in the cryptocurrency market: The impact of covid-19. *International Review of Financial Analysis*, *82*, 102132.

Barbon, A., & Ranaldo, A. (2021). On the quality of cryptocurrency markets: Centralized versus decentralized exchanges. *arXiv preprint arXiv:2112.07386*.

Barndorff-Nielsen, O. E., & Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, *72*(3), 885–925.

Bauer, G. H., & Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, *160*(1), 93–101.

Baur, D. G., Cahill, D., Godfrey, K., & Liu, Z. F. (2019). Bitcoin time-of-day, day-of-week and month-of-year effects in returns and trading volume. *Finance Research Letters*, *31*, 78–92.

Boginski, V., Butenko, S., & Pardalos, P. M. (2005). Statistical analysis of financial networks. *Computational statistics & data analysis*, *48*(2), 431–443.

Borri, N. (2019). Conditional tail-risk in cryptocurrency markets. *Journal of Empirical Finance*, *50*, 1–19.

Borri, N., & Shakhnov, K. (2022). The cross-section of cryptocurrency returns. *The Review of Asset Pricing Studies*, *12*(3), 667–705.

Brownlees, C., Nualart, E., & Sun, Y. (2018). Realized networks. *Journal of Applied Econometrics*, *33*(7), 986–1006.

Brunnermeier, M. K., & Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, *104*(2), 379–421. https://doi.org/10.1257/aer.104.2.379

Buraschi, A., Porchia, P., & Trojani, F. (2010). Correlation risk and optimal portfolio choice. *The Journal of Finance*, *65*(1), 393–420.

Cai, T. T., Hu, J., Li, Y., & Zheng, X. (2020). High-dimensional minimum variance portfolio estimation based on high-frequency data. *Journal of Econometrics*, *214*(2), 482–494.

Chiriac, R., & Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, *26*(6), 922–947.

Chiu, T. Y., Leonard, T., & Tsui, K.-W. (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association*, *91*(433), 198–210.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Costenbader, E., & Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Social networks*, *25*(4), 283–307.

De Prado, M. L. (2016). Building diversified portfolios that outperform out of sample. *The Journal of Portfolio Management*, *42*(4), 59–69.

Dwyer, G. P. (2015). The economics of bitcoin and similar private digital currencies. *Journal of Financial Stability*, *17*, 81–91.

Geraci, M. V., Garbaravičius, T., & Veredas, D. (2018). Short selling in extreme events. *Journal of Financial Stability*, *39*, 90–103.

Guesmi, K., Saadi, S., Abid, I., & Ftiti, Z. (2019). Portfolio diversification with virtual currency: Evidence from bitcoin. *International Review of Financial Analysis*, *63*, 431–437.

Hafner, C. M., & Wang, L. (2021). A dynamic conditional score model for the log correlation matrix. *Journal of Econometrics*.

Heiden, M. (2015). Pitfalls of the cholesky decomposition for forecasting multivariate volatility. *Available at SSRN 2686482*.

Higham, N. J. (2008). *Functions of matrices: Theory and computation*. SIAM.

Huang, X., Han, W., Newton, D., Platanakis, E., Stafylas, D., & Sutcliffe, C. (2022). The diversification benefits of cryptocurrency asset categories and estimation risk: Pre and post covid-19. *The European Journal of Finance*, 1–26.

Jaeger, M., Krügel, S., Marinelli, D., Papenbrock, J., & Schwendner, P. (2021). Interpretable machine learning for diversified portfolio construction. *The Journal of Financial Data Science*, *3*(3), 31–51.

Jaeger, M., & Marinelli, D. (2022). Network diversification for a robust portfolio allocation. *Available at SSRN*.

Kelley, L. (2022). The Latest on FTX's Bankruptcy, BlockFi and More. https://www.nytimes.com/article/ftx-bankruptcy-crypto-collapse.html

Kinateder, H., & Papavassiliou, V. G. (2021). Calendar effects in bitcoin returns and volatility. *Finance Research Letters*, *38*, 101420.

Krishnan, C., Petkova, R., & Ritchken, P. (2009). Correlation risk. *Journal of Empirical Finance*, *16*(3), 353–367.

Lin, M.-B., Khowaja, K., Chen, C. Y.-H., & Härdle, W. K. (2021). Blockchain mechanism and distributional characteristics of cryptos. *Lin MB, Khowaja K, Chen CYH, Härdle WK (2020) Blockchain mechanism and distributional characteris-*

tics of cryptos. Forthcoming in: Book Series: Advances in Quantitative Analysis of Finance & Accounting (AQAFA), 18.

Makarov, I., & Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics, 135*(2), 293–319.

Ng, J. (2022). Market share of biggest crypto exchanges, by daily trading volume. https://www.coingecko.com/research/publications/centralized-exchanges-market-share

Olmo, J. (2021). Optimal portfolio allocation and asset centrality revisited. *Quantitative Finance, 21*(9), 1475–1490.

Peralta, G., & Zareei, A. (2016). A network approach to portfolio selection. *Journal of Empirical Finance, 38*, 157–180.

Petukhina, A. A., Reule, R. C., & Härdle, W. K. (2021). Rise of the machines? intraday high-frequency trading patterns of cryptocurrencies. *The European Journal of Finance, 27*(1-2), 8–30. https://doi.org/10.1080/1351847X.2020.1789684

So, M. K., Chu, A. M., & Chan, T. W. (2021). Impacts of the covid-19 pandemic on financial market connectedness. *Finance Research Letters, 38*, 101864.

Sortino, F. A., & Price, L. N. (1994). Performance measurement in a downside risk framework. *the Journal of Investing, 3*(3), 59–64.

Vidal-Tomás, D. (2021). An investigation of cryptocurrency data: The market that never sleeps. *Quantitative Finance, 21*(12), 2007–2024.

Vukovic, D., Maiti, M., Grubisic, Z., Grigorieva, E. M., & Frömmel, M. (2021). Covid-19 pandemic: Is the crypto market a safe haven? the impact of the first wave. *Sustainability, 13*(15), 8578.

Vÿrost, T., Lyócsa, Š., & Baumöhl, E. (2019). Network-based asset allocation strategies. *The North American Journal of Economics and Finance, 47*, 516–536.

Appendix

4.A Pitfall of the Cholesky decomposition on ordering

Here, we showcase the issue of the Cholesky decomposition on variable ordering. A partial-correlation forecast based on the Cholesky decomposition in the last time snapshot of the sample period is provided in Figure 4.A.1. On the left, the exchanges are ordered based on their trading volumes in Table 4.1; and on the right, they are ordered alphabetically. Sorting by trading volumes in descending order and by alphabetical order yield different forecasts. For instance, the partial correlation between Binance and Kucoin in the left panel is 0.72, while it is 0.98 on the right, resulting in a different network structure. In contrast, based on the matrix-logarithm transformation in 4.A.2, the two sorting methods generate an identical result.
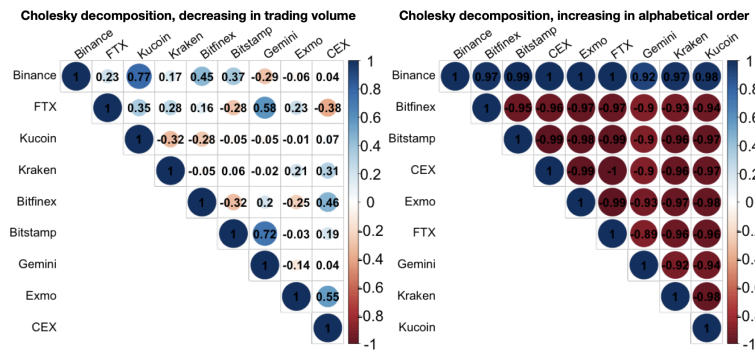
Figure 4.A.1: **Partial correlations based on the Cholesky decomposition on 2022-09-30 23:00:00.**
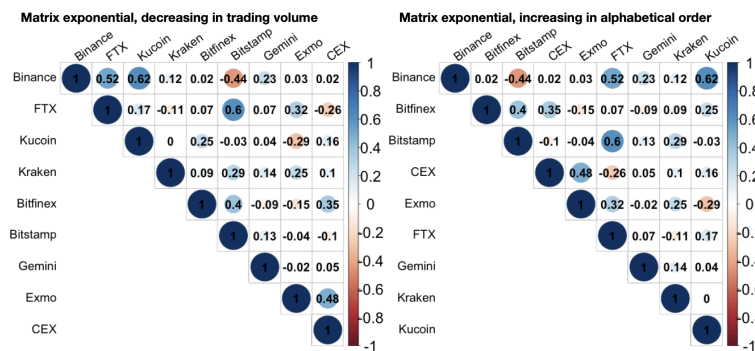
Figure 4.A.2: **Partial correlations based on matrix logarithm on 2022-09-30 23:00:00.**

4.B   USDT and USD close price
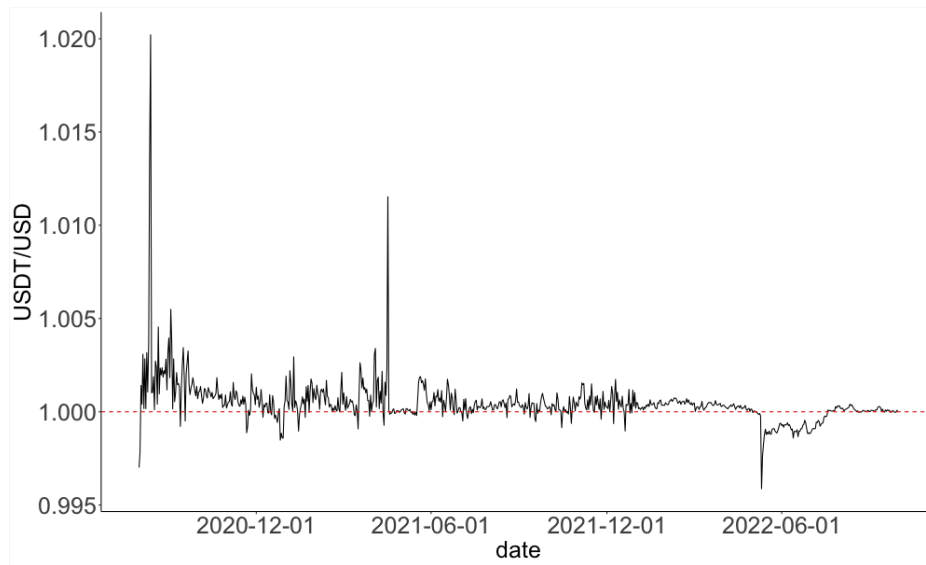
Figure 4.B.1: **Daily USDT (solid) and USD (dashed) close prices.**

4.C Daily aggregated centrality ranks
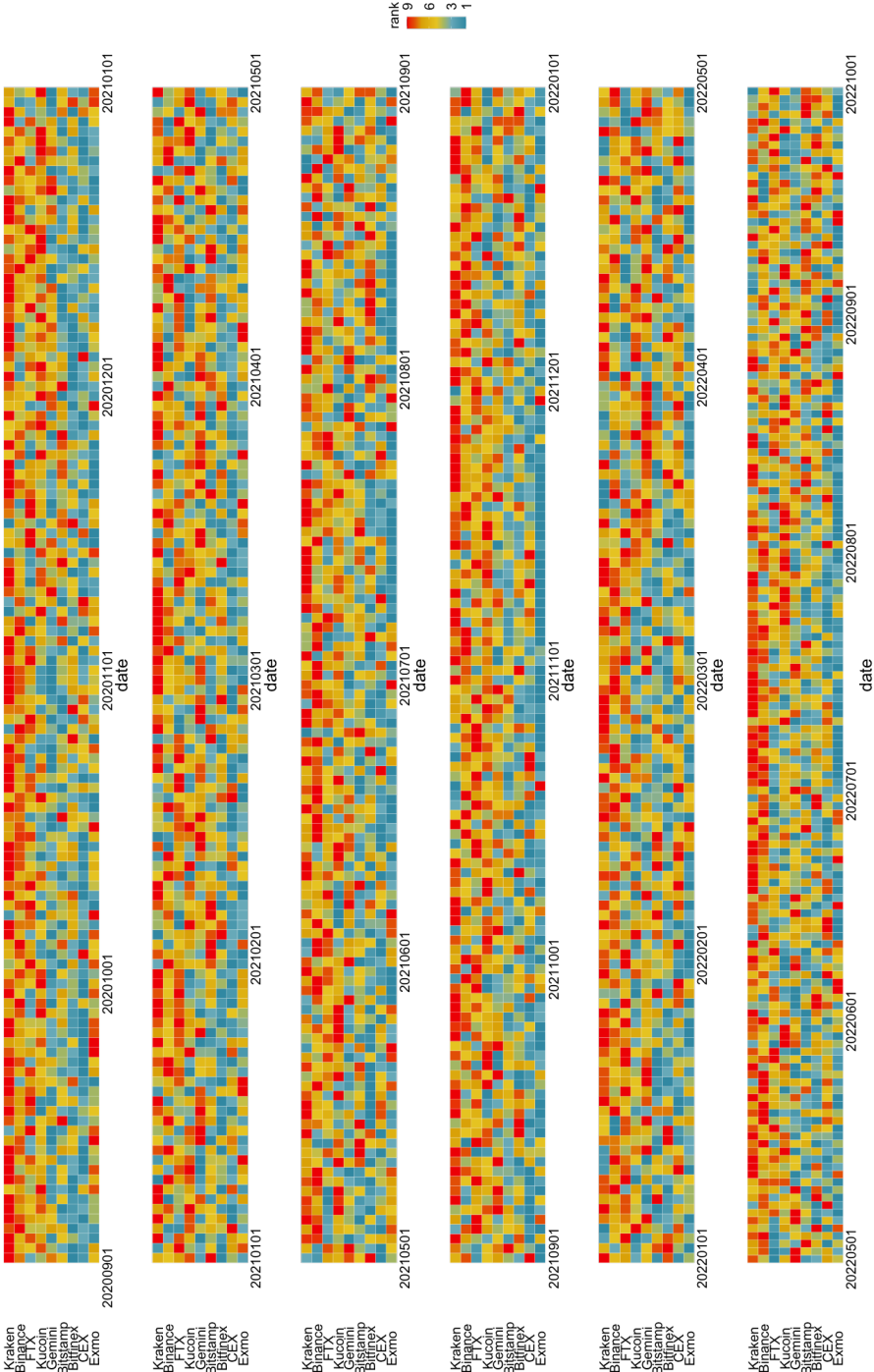
Figure 4.C.1: **Daily degree-centrality ranks aggregated by mean.** The higher the centrality value, the higher the rank.
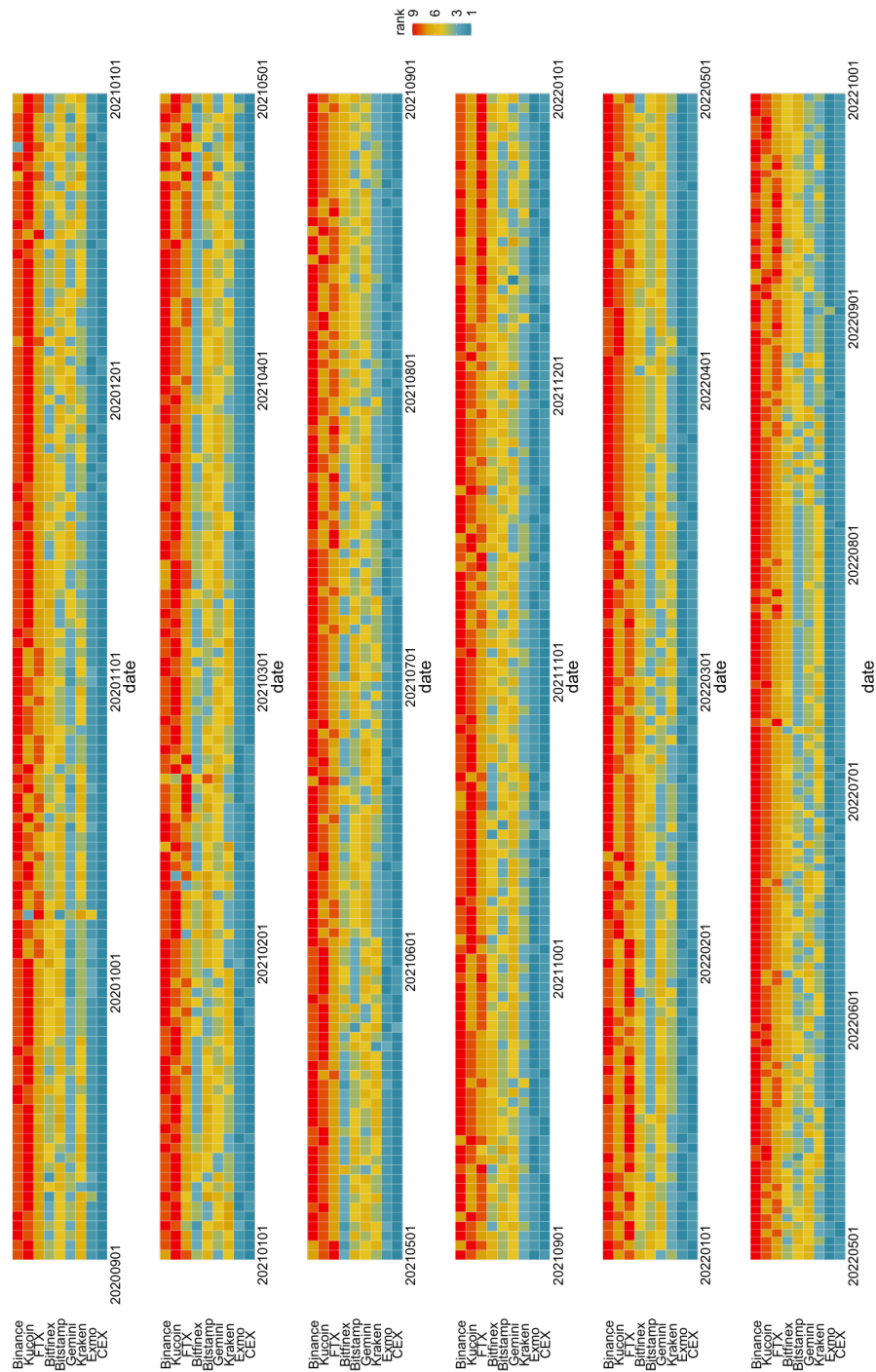
Figure 4.C.2: **Daily eigenvector-centrality ranks aggregated by mean.** The higher the centrality value, the higher the rank.

4.D Centrality for different periods

Table 4.D.1: Centrality of each exchange for Figures 4.8 - 4.11.

Exchange	2021-02-12 05:00:00		2021-05-27 09:00:00		2021-12-04 04:00:00		2022-06-04 08:00:00	
	Degree	Eigenvector	Degree	Eigenvector	Degree	Eigenvector	Degree	Eigenvector
Binance	5	0.046	8	0.826	7	1.000	5	1.000
FTX	6	0.487	7	0.532	5	0.809	5	0.724
Kucoin	6	0.499	7	1.000	6	0.696	6	0.838
Kraken	7	1.000	6	0.093	8	0.376	6	0.198
Bitfinex	4	0.619	6	0.228	8	0.301	6	0.444
Bitstamp	7	0.839	7	0.148	7	0.417	6	0.260
Gemini	8	0.315	7	0.233	5	0.243	6	0.169
Exmo	7	0.270	7	0.038	5	0.074	3	0.007
CEX	6	0.266	5	0.080	7	0.195	5	0.123

4.E   Performance measure

To underscore the confirmation time for cross-exchange trading, the regularly (i.e., hourly, daily) updated return for exchange i is computed using $R_{i,t} = \log(p_{i,t}) - \log(p_{i,t-1})$, while the portfolio return is given by $R_{p,t} = \sum_{i=1}^{N} w_{i,t} R_{i,t}$. Below, we list the performance measures considered.

⊡ Standard deviation: The dispersion of the portfolio return, $R_{p,t}$, from its mean, $\mathsf{E}(R_{p,t})$. It is given by

$$\sigma_p = \sqrt{\sum_{t=1}^{T} (R_{p,t} - \mathsf{E}(R_{p,t}))^2}.$$

⊡ Sharpe ratio:
$$\text{Sharpe} = \mathsf{E}(R_p - R_f)/\sigma_p,$$

where R_p, σ_p and R_f denote the mean, standard deviation of the portfolio return, $R_{p,t}$, and risk-free rate, respectively. $\mathsf{E}(R_p - R_f)$ is the expected value of the excess of the asset return over the risk-free rate.

Because there is no acknowledged high-frequency risk-free rate in the crypto market, we set $R_f = 0$. The Sharpe ratio relies on the assumption that the distribution of returns is normal; thus, the second-order moment can depict the risk. The standard deviation includes variations above the average returns. Most people prefer these and are only concerned about the below-average returns. Moreover, if the portfolio returns have non-normal distributions, comparing portfolios based on the Sharpe ratio is often insufficient, as it ignores investors' positive preferences for odd moments and aversion to even moments.

The following performance measures are adjusted by downside tail or extreme risks.

⊡ Semi deviation: Semi deviation eliminates positive returns when calculating risk. It measures the variability of the (under) performance below 0. It is given by

$$\delta_0 = \sqrt{\sum_{t=1}^{T} \frac{\min(R_{p,t}, 0)^2}{T}}.$$

⊡ Sortino ratio: The Sortino ratio (Sortino and Price, 1994) differs from the Sharpe ratio as it only penalizes deviations that fall under our target rate of return, which refers to the negative returns here, and it is computed as follows:

$$\text{Sortino} = \mathsf{E}(R_p - R_f)/\delta_0.$$

⊡ MDD: The decline from a historical peak in return is defined as follows:

$$\text{MDD} = \max_{\tau \in (0,T)} \{ \max_{t \in (0,\tau)} R_{p,t} - R_{p,\tau} \},$$

where $R_{p,t} and R_{p,\tau}$ denote the portfolio returns at times t and τ, respectively.

⊡ Calmar ratio: Whereas the Sortino ratio considers all the downside volatility over a given time, the Calmar ratio considers the maximum drawdown for that time:

$$\text{Calmar} = \frac{\text{Scale}}{T} \times \Sigma_{t=1}^{T}(R_{p,t})/\text{MDD},$$

where Scale $= 24 \times 365$ is the number of observations in a year.

4.F   Hierarchical cluster

For each figure in this section, the dendrograms compare the two HRP variants –
HRP_MHAR_CM (on left) and HRP_RCor (on right). If the two variants generate
an equal cluster, the linkage between the same assets is a horizontal line. Furthermore,
we can observe the updating of asset allocation by time in the figure here.
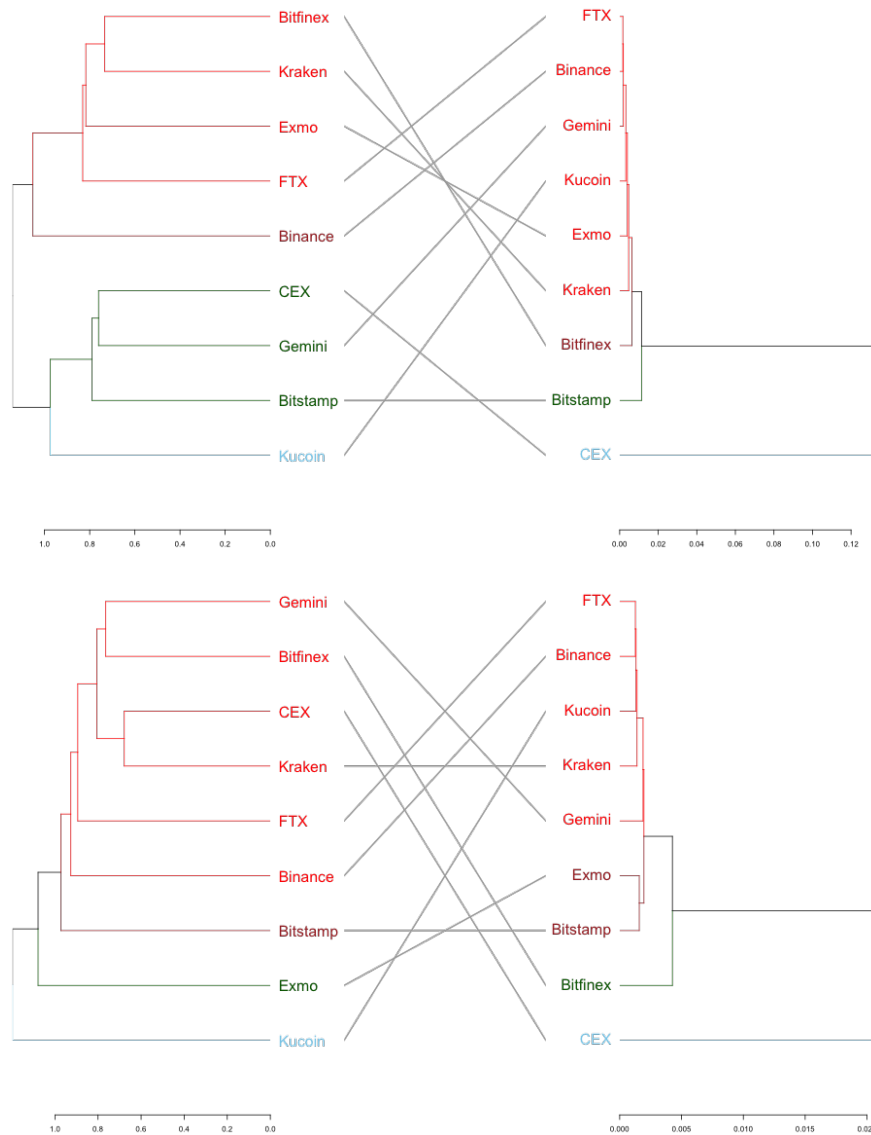

Figure 4.F.1: **Dendrograms for the two HRP variants** – 2021-05-27 03:00:00 (top) and
2021-05-27 09:00:00 (bottom).

Chapter 5

DAI Digital Art Index

All the supplementary materials and source codes of this chapter are available in the Q^2 ecosystem: ￮DAI_digital_art_index and ￮.

5.1   Introduction

Non-fungible token (NFT), a digital token as a proof of ownership and authenticity for assets, offers a futuristic possibility for art trading and thus impacts the market. Its market capitalization booms up to US\$ 41 billion in 2021, which is a 30 % growth by last year; meanwhile, the conventional art market shrinks by 22% in sales from 2019 (Chainalysis, 2022; Dailey, 2022). A preliminary comparison of quarterly sales of four different art markets worldwide is illustrated in Figure 5.1, sourced from Artnet.com [1]. We observe a surge of NFT art since the COVID-19 pandemic in 2020; while it receives a decline later in the Q4 of 2021. Other art markets also get boosted on sales since the pandemic.

Driven by blockchain (BC) technology and smart contracts (SC), NFTs inherit their functional features – i.e. network decentralization, privacy and encryption, immutability of data, and programmability in contracts. NFT digital art market has been seen as a rising opportunity for the creators from professionals to amateurs to interact with under-served populations. It furnishes more equitable access to wealth and the inclusion potential to artworks. Yet, such a market is still under development and not fully understood.

In this paper, we investigate the top 10 NFT art collections ranked by transaction volumes on the marketplace – Opensea which is currently the most traded NFT

---

[1]Post war: Works produced between 1945 and 1970, i.e. Yves Klein, Rothko, Gerhard Richter; Contemporary: Works produced after postwar from the late 20th to early 21st, i.e. Andy Warhol, Yayoi Kusama; Ultra-contemporary: Works produced by artists born from 1975 to the present, i.e. JR.
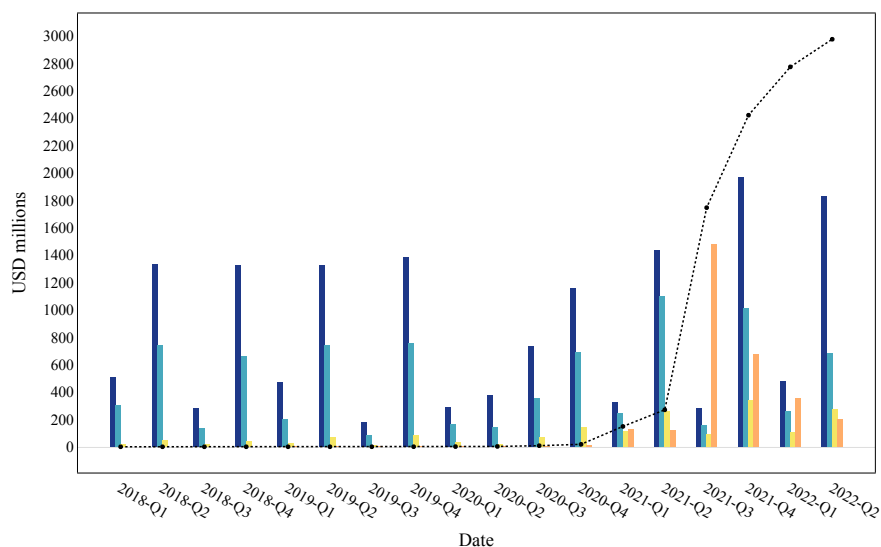
Figure 5.1: **Quarterly sales in USD millions of post war, contemporary, ultra-contemporary, NFT art markets, plotted in order**. The dotted line is the cumulative sales on NFT.

marketplace. The top 10 collections have consisted of a total of 92,763 digital art pieces and represent around 84% of total market trading volume In order to outline the general trend of the NFT art market and develop a broad understanding of its risk and return potential, we construct a characteristic-free price index – Digital Art Index (DAI) and then conclude its price movement and determinants. We apply hedonic regression – which is derived from data that controls for differences in the characteristics of assets in various samples – and use the Kalman filtering (KF) with Huberization or DCS-t filtering to smooth the index time-series. Our results can be used for observing and evaluating this new market and for the comparison with asset classes, e.g. cryptocurrencies (see CRIX).

The construction of an art price index is different from building indices for stocks or similar assets, which are based on prices of identical shares. For the art market, heterogeneous objects are to be considered. McAndrew, 2010 indicates that the essence of art indices are: representativeness, liquidity, and capacity, i.e. the potential and values of sales. Due to the features of art market – high heterogeneity among artworks and infrequent trading behaviors (i.e. unbalanced panel data), the determination of changes in market value is difficult to ascertain (Beckert & Rössel, 2013; Ginsburgh et al., 2006). In addition, we encounter the presence of outliers in pricing, see Figure 5.2. An average (arithmetic mean or median) price index might thus not be sufficient. Most art indices (e.g. Artnet.com, Sothebys Mei Moses) are based on data reflecting the artwork's price at a certain time, a function of the fixed characteristics of work

– time-invariant characteristics – and elements that vary over time – time-variant characteristics. For the NFT digital art, we collect the metadata and historical prices of each work and characterize it by its time-variant characteristics (i.e. number of sales, owner address) and time-invariant characteristics (i.e. collection name, NFT scheme).
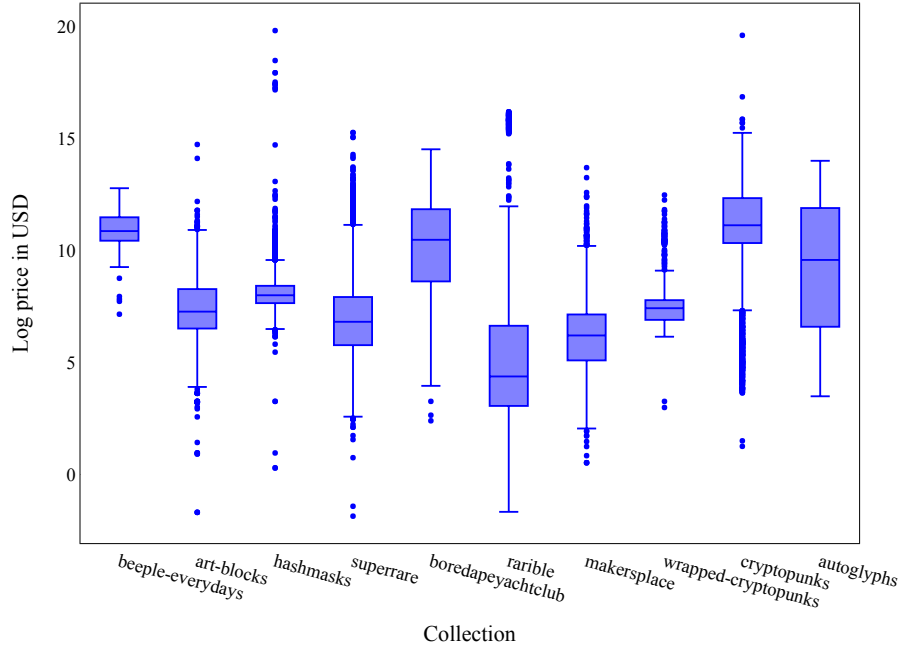
Figure 5.2: **The presence of outliers in NFT digital art pricing**.

Thus far, there is a limited amount of literature that has been published to comprehend the economics of NFT art market and underpin its role to the conventional art market. Kräussl and Tugnetti, 2022 overview NFT pricing and valuation methods, and identify NFTs having a similar investment profile – high yield and high risk – with classic collectibles. Dowling, 2022a; Goldberg et al., 2021 investigate the pricing of virtual properties in Decentraland. Nadini et al., 2021 study the behaviors of NFT traders from a perspective of network theory and offer brief statistics. Ante, 2021; Dowling, 2022b look into co-integration relationship and time - varying co-movements among the selected cryptos and NFTs. Borri et al., 2022 construct a market index based on repeat-sales regression (RSR) and compare it to the conventional assets and cryptos. Due to the lack of a representative market index, most of the studies are nevertheless restricted to individual NFT assets. In order to mitigate such a research gap and facilitate the realization of market research, we develop a hedonic regression-based price index. Based on a comprehensive dataset collected in this study, we include each work's characteristics and historical prices. We apply two alternative methods for the robustness of price index – Huberization and dynamic conditional score (DCS) model

for parameter filtering in order to reduce the influence of immanent outliers.

The paper proceeds as follows. Section 2 portrays the selling processes for NFT and conventional art markets and highlights their stylized facts. Section 3 exposes the commonly used estimation methods for constructing art indices. Section 4 elaborates the methodologies used in this study. Section 5 presents the data attributes and data processing. We demonstrate the result in Section 6 and offer a discussion for interpretation in Section 7 Finally, Section 8 concludes our findings.

5.2   NFT art market

Since the emergence of BC technology, there have been many adoptions in different sectors across all applications – crypto utility tokens in particular (Kumar et al., 2021). NFTs are different from utility tokens. They offer the possibility for digital assets to be owned and traded, certifying ownership and authenticity. We define NFT as – *a cryptographically secured and unique digital identifier minted through smart contracts on a BC that cannot be replicated, subdivided or substituted.* It can be used to represent both tangible and intangible assets and to facilitate a programmable smart contract between seller and buyer. So, an artwork is traded directly among creators and buyers in an automated and flexible manner. Such transactions are secured and transparent. An outlook for the NFT digital art market can be found in Allen et al., 2022; Pawelzik and Thies, 2022.

On the other hand, in the conventional art market artists and creators are generally required to seek for a dealer or gallery to access potential buyers, e.g. independent collectors, museums, or auction houses (Velthuis, 2003; Zorloni, 2005). Before a work of art enters a marketplace, it involves a series of valuation and curation processes. Involving third parties, it can be costly and time-consuming. McEachern, 2016 offers review literature of the contemporary art market and its economics. In the following, we overview both the NFT and conventional art markets.

5.2.1   Selling process

Figure 5.3 illustrates the selling processes for the NFT and conventional art markets. The arrow lines flow with the transfer of an artwork. The primary market is where each artwork issues when it is first made available for sale by a creator; the secondary market contains every later sale. In the NFT art market, a creator can be either institutional (e.g. Larva Labs, Dapper Labs) or independent. Most of the artworks are traded in the secondary market via a marketplace (e.g. Opensea, Rarible) except the works from the institutes that support the architecture of a BC. Art creators generally

preserve the copyright of artworks and are able to reproduce a series of similar works. Also, they can decide the listing price – the first sale price – and control the bidding process with buyers. Due to the programmability of NFTs, a creator can later demand a royalty, a preset percentage of the sale price ranging from 5-10% typically. The crypto payment and the transfer of NFT are processed automatically and simultaneously in a single transaction between two mutually distrusting parties on a BC network. NFT consequently establishes diverse contented marketplaces and enables a vibrant and sustainable revenue stream for creators.

In a conventional art market, creators depend on a centralization dealer / gallery to proceed with the selling process. The listing price is decided via a valuation process and a negotiation between creators and dealers. After the first sale, the artwork is traded in different channels (Towse & Hernández, 2020; Zorloni, 2005); however, each channel might be independent of each other. Moreover, there are a few causal linkages between different marketplaces across the world and no feedback relationships in between (Le Fur, 2020). The transaction history and pricing are less transparent in comparison to the NFT art market, where all the transaction history is secured by BC technology. Resale royalty right, also called droit de suite in the conventional market, might be difficult due to the lack of traceability in sales history.
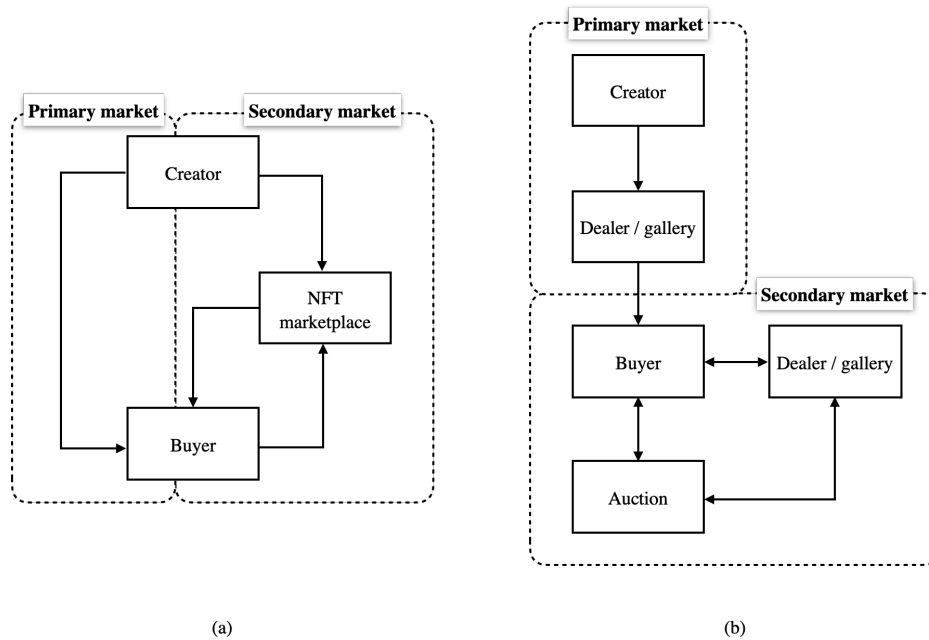

Figure 5.3: **Selling processes for (a) NFT and (b) conventional art markets**.

5.2.2   Economics of art market

Art has been considered as one of the investment instruments that offers lower volatility and lower correlation with other assets, thus being used for portfolio diversification (Campbell, 2008; Mei & Moses, 2002a; Worthington & Higgs, 2004). In terms of the performance in returns, Bocart et al., 2020; Whitaker and Kräussl, 2020 indicate that contemporary art and gallery portfolios perform almost as well as the S&P 500, and the majority of art markets do not load significantly on momentum or liquidity factors. Penasse and Renneboog, 2021 show that the art market is subject to frequent booms and busts, so it is not applicable to rational consumer models and exhibits an opportunity for spectacular profits. Goetzmann et al., 2011; Renneboog and Spaenjers, 2013 direct attention to income inequality and art market sentiment and show their correlation to price trends. A wide body of literature analyzes price determinants via – the work of art (size, material, genre), the artist (age, sex, place of residence), and the sales channel (location, affiliation) (Garay, 2021; Li et al., 2021; Rengers & Velthuis, 2002). This can be summarized on the main stylized facts of the conventional art market – i.e. *infrequent trading, price inequality, illiquidity, centralized patrons, less transparency in price formulation, multiple independent marketplaces.*

The exchanges and transfers of NFT art, compared with the conventional artworks, proceed safely and with (nearly) no intermediaries. Cryptos are essentially the means of payment for NFTs – e.g. Ethereum (ETH), Solana (SOL), and therefore, crypto pricing affects NFT asset pricing and the market (Ante, 2021). Dowling, 2022b remarks that NFT assets (i.e. Decentraland, Cryptopunks) have a limited volatility spillover from cryptos. Urom et al., 2022 find that the connectedness and volatility of returns between NFT market and other conventional markets (e.g. stock, commodity, crypto) are higher during extreme market conditions; NFT markets usually is a net return shock receiver. In terms of liquidity, Wilkoff and Yildiz, 2022 conclude that there is a positive relationship between ETH returns and NFT illiquidity. In other words, while ETH returns are high, investors are unwilling to buy NFTs. In Figure 5.4, we observe that the downwards and upwards of total transaction volume of NFT art are often corresponding to dropping and rising of CRIX index proposed by Trimborn and Härdle, 2018, an index for tracking the evolution of cryptocurrency market.

Similar to the conventional art, NFT artworks are traded infrequently and there exists an imbalance in asset pricing. In the report of Vasan et al., 2022, they indicate that market first-mover creators, in general, own an advantage in pricing and sales, and a network effect (i.e. from twitter followers) is observed. Rarity and uniqueness of collectibles induce market demand and thus offer a higher price premium, known as rarity premium (Koford & Tschoegl, 1998). For instance, the on-chain art generative
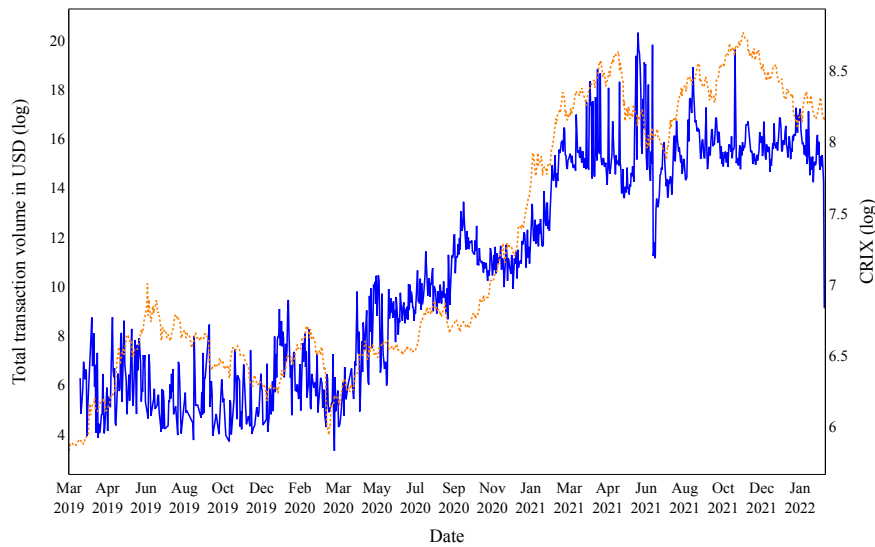
Figure 5.4: **Time series of total transaction volume of NFT art in USD (solid) and CRIX (dotted).**

collection – Autoglyphs, due to its limited number of issues, has consistent pricing and resale rates in Figures 5.2 and 5.5. Mekacher et al., 2022 quantify rarity based on natural language descriptions (i.e. traits) in the metadata of NFT art and show that high rarity induces high returns. Moreover, in the conventional art market, curation processes play a critical role in the valuation of a work. Applying such a mechanism to the digital art market, ArtBlocks, for example, offers a curation board for evaluating generative artworks from their scripts.

We observe there exist many adoptions from the conventional on the NFT art market. Transforming and evolving towards digitalization, the NFT art market has provided many distinctive technical properties – backed by BC technology and smart contracts – that attract different groups of art collectors (i.e. crypto investors) and consequently influence trading behaviors. In this study, we dedicate to categorizing the stylized facts for such a brand-new market and empirically compare it to other markets.

5.3 Towards DAI, an Index for Digital Art

Two estimation methods – Repeat-sales regression (RSR) and hedonic regression (HR) are commonly applied to construct price indices for heterogeneous goods that comprise various characteristics. In the absence of individual characteristics, RSR uses the prices traded within two distinct moments. Bailey et al., 1963 employs the RSR for real

estate price index construction and it is formulated as:

$$\underbrace{\log\left(\frac{p^i_{t+s}}{p^i_t}\right)}_{\text{Relative price}} = \delta_{t+s}d^i_{t+s} + u^i_{t+s} \tag{5.1}$$

where p^i_t is the observed price of object $i \in \{1, \ldots, N\}$ in period $t \in \{1, \ldots, T\}$ with $T \geq t+s > t$; and t and $t+s$ are the periods for the first sale and the second sale, respectively. δ_t is the index; $d^i_t \in \{0,1\}$ is the time dummy; and error term $u \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_u{}^2)$. One can see δ_t here as the average return at period t in the portfolio. RSR might suffer though from sample biases when the data is rather sparse. Moreover, it only considers relative prices (returns), so that the individual objects having more than two sales are not included.

Hedonic regression model regresses prices on various characteristics of assets transacted and can adapt the changing characteristics over time. Though the model has been proposed since the applications of Waugh, 1928 and Court, 1939, later Griliches, 1971; Griliches, 1961 and Rosen, 1974 have justified it as a quality-adjusted price index and HR model has started widely being implemented in e.g. house pricing, art pricing. A general form for an HR model is written as:

$$\log(p^i_t) = \sum_{\tau=1}^{t} \beta_\tau d^i_\tau + \sum_{k=1}^{m} \gamma_k v^i_k + \sum_{\tau=1}^{t}\sum_{j=1}^{n} \theta^j_\tau w^{ij}_\tau + \varepsilon^i_t \tag{5.2}$$

where p^i_t is the observed price of object $i \in \{1, \ldots, N\}$ in period $t \in \{1, \ldots, T\}$; β_τ is the estimated coefficient of time dummy variable $d^i_\tau \in \{0,1\}$ with $\tau \in \{1, \ldots, T\}$; and ε^i_t is the error term with $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon{}^2)$. p^i_t is a function of m time-invariant characteristics v^i_k with $k \in \{1, \ldots, m\}$ and of n time-variant characteristics w^{ij}_τ at period τ with $j \in \{1, \ldots, n\}$; and γ_k and θ^j_τ are the corresponding coefficients. Characteristics depend on the type of objects studied. We might encounter omitted bias and misspecification challenges (Wallace & Meese, 1997).

We can easily find both the RSR and HR concepts coincide if there are no time-variant characteristics in the hedonic factors. Then, Equation 5.2 is written as:

$$\log(p^i_t) = \sum_{\tau=1}^{t} \beta_\tau d^i_\tau + \sum_{k=1}^{m} \gamma_k v^i_k + \varepsilon^i_t \tag{5.3}$$

We, thus, get the difference between the first sale log price p^i_t and the second sale p^i_{t+s} in an HR fashion as below. Note that d is being non-zero at time t and $t+s$ at which

the two observable p occur succeeding.

$$\log(p_{t+s}^i) - \log(p_t^i) = (\sum_{\tau=1}^{t+s} \beta_\tau d_\tau^i + \sum_{k=1}^{m} \gamma_k v_k^i + \varepsilon_{t+s}^i) - (\sum_{\tau=1}^{t} \beta_\tau d_\tau^i + \sum_{k=1}^{m} \gamma_k v_k^i + \varepsilon_t^i)$$
$$= \beta_{t+s} d_{t+s}^i + (\varepsilon_{t+s}^i - \varepsilon_t^i) \tag{5.4}$$

Similarly to other collectibles (e.g. equipment, traditional artworks, real estate), the NFT art market is rather non-liquid and often has low frequency in sales. Figure 5.5 presents that most of the collections listed have the medians of number of sales around 1. As discussed in Shiller, 2008, if either the fraction of sales or the relative price of new assets varies a lot, the presence of new sales along with existing sales is of essence. Hence, RSR concept does not fit the NFT artworks with only a single sale since it cannot represent a precise resolution of the market. One may argue that incorporating both new and existing asset prices is an issue (Shiller, 2008), i.e. the supply of new NFTs varies over time as market conditions change. The NFT art market is emerging and uneven in assets' prices and sales; therefore, it is indeed less stable and alters with the market conditions.
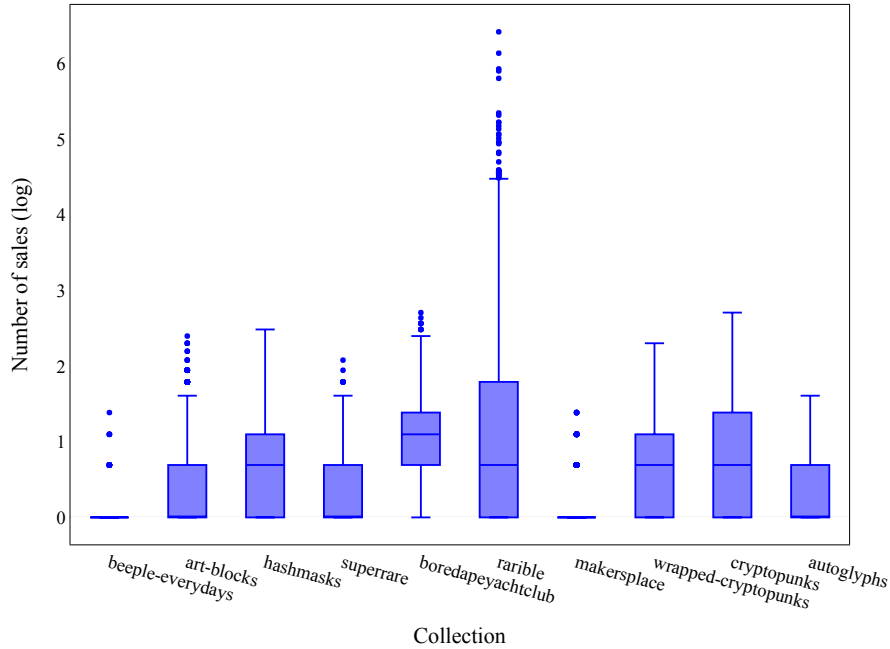

Figure 5.5: **Low frequent sales in the NFT art market**.

5.4   Methodology

After going through the preliminary of the methodology, in this section we explain in detail the regression model and the construction of art index. We explicitly consider the hedonic price index model and refine it by emphasizing outlying and influential data and issues on biased sampling.

5.4.1   Hedonic regression

To consider the characteristics of NFT art and understand its pricing fundamentally, we apply a hedonic regression model and extend it from the work of Bocart and Hafner, 2015. Due to the infrequent trading, we have to deal with an unbalanced panel data structure, i.e. panel data where the individual time series have unequal length of observations. Writing $y_t = (y_t^1, \ldots, y_t^{N_t})^\top$ where y_t^i is the log price of the i-th sale at time $t \in \{1, \ldots, T\}$ with $i \in \{1, \ldots, N_t\}$; $\beta_t = (\beta_t^1, \ldots, \beta_t^{N_t})^\top$; and $\varepsilon_t = (\varepsilon_t^1, \ldots, \varepsilon_t^{N_t})^\top$ with $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon{}^2)$. Note that i here refers to the sales instead of an object. Equation 5.2 is equivalently expressed as:

$$y_t = d_t \beta_t + X_t^\top \alpha + \varepsilon_t \tag{5.5}$$

where $d_t = (1, \ldots, 1)^\top$ is a $(N_t \times 1)$ vector. $X_t = (v_k, w_t^j)$ is of dimension $(N_t \times K)$ and contains K explanatory variables including time-variant and time-invariant characteristics at time t; and $\alpha = (\alpha^1, \ldots, \alpha^K)^\top$ is the corresponding vector parameter. Note that N_t changes with time t.

Once we obtain the parameter α, interpreted as implicit prices of the various characteristics, β_t is estimated:

$$\hat{\beta}_t^{OLS} = \arg\min_{\beta_t \in \mathbb{R}} \left\{ \|y_t - d_t \beta_t - X_t^\top \alpha\|_2^2 \right\} \tag{5.6}$$

Since the model marginalizes over X_t, one creates the characteristic-free price index $\hat{\beta}_t^{OLS}$.

Incorporating with lasso (the least absolute shrinkage and selection operator) for model regularization, we subject Equation 5.5 to $\|\alpha\|_1 < b$ and the lasso determined coefficient $\hat{\beta}_t^{lasso}$ in the equivalent Lagrangian form:

$$\hat{\beta}_t^{lasso} = \arg\min_{\beta_t \in \mathbb{R}, \alpha \in \mathbb{R}^K} \left\{ \|(y_t - d_t \beta_t - X_t^\top \alpha)\|_\rho + \lambda \|\alpha\|_1 \right\} \tag{5.7}$$

where b is a pre-defined parameter that determines the degree of regularization; ρ is a certain loss function, i.e. Huber loss, and λ is regularization parameters. The result of

lasso is mainly for the use of price determinants, emphasizing the characteristics of artworks.

Given the estimate $\hat{\beta}_t$, the price index with base 100 in $t = 1$ is defined as:

$$\text{Index}_t \stackrel{\text{def}}{=} 100 \frac{\exp(\hat{\beta}_t)}{\exp(\hat{\beta}_1)} \tag{5.8}$$

5.4.2 Selection bias correction

We apply Heckman sample selection model, known as Tobit-II model – see Heckman, 1976, 1979 – to address the potential problem of sample selection bias, especially omitted variable bias (Collins et al., 2009; Malpezzi et al., 2003). In our case, we only observe a sub sample of the NFT transactions that have been successfully transacted but other transaction event types (i.e. bid entered, cancelled) cannot be included in the models of Section 4.1 due to unobserved dependent variables (i.e. price). The Heckman model utilizes the regression covariates to approximate selection, transforms the prediction, and then includes that transformation in the model.

Assume two latent variables $y_{t,i}^{\mathbf{o}}{}^*$ and $y_{t,i}^{\mathbf{s}}{}^*$ where the time $t \in \{1, \dots, T\}$ and the sale $i \in \{1, \dots, N_t\}$. $y_t^{\mathbf{o}*} = (y_{t,1}^{\mathbf{o}}{}^*, \dots, y_{t,N_t}^{\mathbf{o}}{}^*)^\top$ links the covariates of interest to the outcomes and $y_t^{\mathbf{s}*} = (y_{t,1}^{\mathbf{s}}{}^*, \dots, y_{t,N_t}^{\mathbf{s}}{}^*)^\top$ is a vector of the selection propensity variables, as seen below:

$$y_t^{\mathbf{o}*} = x_t^{\mathbf{o}\top} \beta_t^{\mathbf{o}} + \varepsilon_t^{\mathbf{o}} \tag{5.9}$$

$$y_t^{\mathbf{s}*} = x_t^{\mathbf{s}\top} \beta_t^{\mathbf{s}} + \varepsilon_t^{\mathbf{s}} \tag{5.10}$$

where $x_t^{\mathbf{o}}$ and $x_t^{\mathbf{s}}$ are matrices of explanatory variables. Here, assume $x^{\mathbf{s}} \subseteq x^{\mathbf{o}}$, i.e. predictor variables using for the main outcomes of interest $y^{\mathbf{o}*}$ are used to predict the selection $y^{\mathbf{s}*}$. $\varepsilon_t^{\mathbf{o}}$ and $\varepsilon_t^{\mathbf{s}}$ are error terms. $\beta_t^{\mathbf{o}}$ and $\beta_t^{\mathbf{s}}$ are coefficient parameters. $y_{t,i}^{\mathbf{o}}$ is only observed if the selection propensity variable is positive such as:

$$y_{t,i}^{\mathbf{s}} = \begin{cases} 1 & \text{if } y_{t,i}^{\mathbf{s}}{}^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.11}$$

Given selection to the main sample, the expected value of the outcome in Equation 5.9 is given by:

$$\mathsf{E}[y_t^{\mathbf{o}}|x_t^{\mathbf{s}}, \varepsilon_t^{\mathbf{s}}] = x_t^{\mathbf{o}\top} \beta_t^{\mathbf{o}} + \mathsf{E}[\varepsilon_t^{\mathbf{o}}|\varepsilon_t^{\mathbf{s}} \geq -x_t^{\mathbf{s}\top} \beta_t^{\mathbf{s}}] \tag{5.12}$$

Estimating the model above by OLS gives generally biased results, hence assume $\varepsilon_t^{\mathbf{o}}$ and $\varepsilon_t^{\mathbf{s}}$ are mean independent and follow a bivariate normal distribution, i.e. $\begin{pmatrix} \varepsilon^{\mathbf{o}} \\ \varepsilon^{\mathbf{s}} \end{pmatrix} \sim$

$\mathcal{N}\left(\binom{0}{0}, \begin{pmatrix} 1 & \rho \\ \rho & \sigma^2 \end{pmatrix}\right)$. Accordingly, we can write:

$$
\begin{aligned}
y_t^{\mathbf{o}} &= \beta_t^{\mathbf{o}} x_t^{\mathbf{o}\top} + \mathsf{E}[\varepsilon_t^{\mathbf{o}} | \varepsilon_t^{\mathbf{s}} \geq -x_t^{\mathbf{s}\top} \beta_t^{\mathbf{s}}] + u_t \\
&\overset{\text{def}}{=} \beta_t^{\mathbf{o}} x_t^{\mathbf{o}\top} + \rho \sigma \lambda(\beta_t^{\mathbf{s}} x_t^{\mathbf{s}\top}) + u_t
\end{aligned}
\tag{5.13}
$$

where $\lambda(\cdot) = \varphi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio (IMR), being $\varphi(\cdot)$ the standard normal probability density function (pdf) and $\Phi(\cdot)$ the complementary cumulative distribution function (cdf). u_t is an error term which is independent of both $x_t^{\mathbf{s}}$ and $x_t^{\mathbf{o}}$. $\beta_t^{\mathbf{s}}$ is estimated via a probit model, assuming $\mathsf{P}(y_t^{\mathbf{s}} = 1 | x_t^{\mathbf{s}}) = \Phi(x_t^{\mathbf{s}\top} \beta_t^{\mathbf{s}})$ and performing MLE, and thereafter, we access the estimates of $\lambda(\beta_t^{\mathbf{s}} x_t^{\mathbf{s}\top})$. Since $\rho \neq 0$ the unknown multiplicator $\rho\sigma$ is estimated by OLS regressing $y_t^{\mathbf{o}}$ on $x_t^{\mathbf{o}}$ given the selection sample and $\lambda(\hat{\beta}_t^{\mathbf{s}} x_t^{\mathbf{s}\top})$. Further details on the consistency of such estimators are discussed in Miranda and Rabe-Hesketh, 2006.

5.4.3 Huberization

Proceeding with the OLS regression, we observe that there exist many outliers in the residuals $\hat{\varepsilon}_t$, see Figure 5.6. This is also visible in Figure 5.7 where the OLS residuals are displayed via box plots over time in comparison with the number of transactions per day. The number of transactions of NFT artworks induces the outlying observations in residuals; the mean of residuals gets stablized by time.
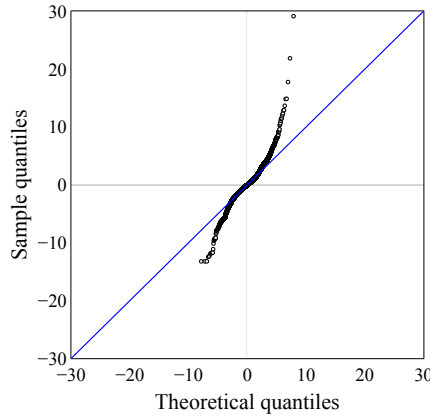
Figure 5.6: **QQ plot for OLS standard residuals against standard normal distribution**.

To alleviate such an impact of outliers and offer a robust estimate of β, we introduce a correction procedure – Huberization to reflect the actual outlyingness. Härdle, 1984 uses an influence curve, Huber's ψ-function from Huber, 1964, to bent down the influence of outlying observations. We construct pseudo observations $\tilde{y}_t = (\tilde{y}_t^1, \ldots, \tilde{y}_t^{N_t})^\top$ as a
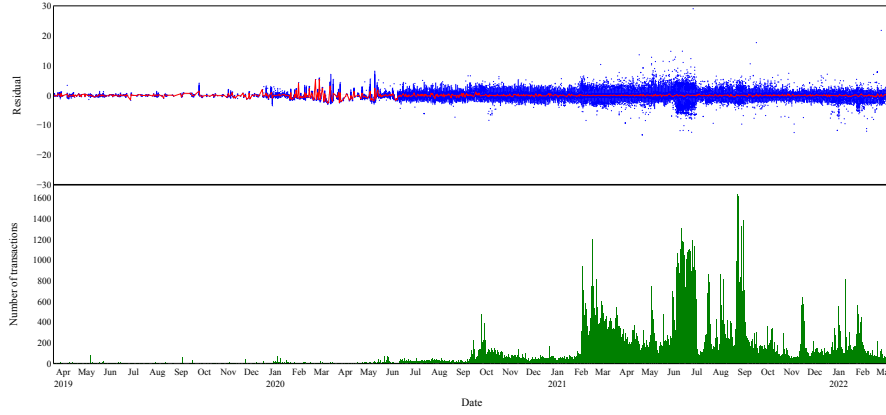
Figure 5.7: **OLS residual box plots** over time with **mean of residuals** and **number of transactions per day**.

one-step correction of y_t:

$$\tilde{y}_t = y_t + \tilde{\varepsilon}_t$$
$$= d_t \beta_t + X_t^\top \alpha + \tilde{\varepsilon}_t \tag{5.14}$$

where y_t is estimated using $\hat{y}_t = d_t \hat{\beta}_t + X_t^\top \alpha$ from Equations 5.6 or 5.7. $\tilde{\varepsilon}_t = (\tilde{\varepsilon}_t^1, \ldots, \tilde{\varepsilon}_t^{N_t})^\top$ is a vector of the bounded one-step residuals:

$$\tilde{\varepsilon}_t^i = \frac{\psi(\varepsilon_t^i)}{\Delta} \tag{5.15}$$

where the scaling parameter $\Delta = \mathsf{E}[\psi'(\varepsilon_t^i)]$ is estimated via

$$\hat{\Delta} = T^{-1} \sum_{t=1}^{T} [n_t]^{-1} \sum_{i=1}^{n_t} [\psi'_\tau(\hat{\varepsilon}_t^i)]$$

from OLS or lasso. The ψ-function is simply the derivative of the ρ-function.

$$\rho_\tau(\varepsilon_t^i) = \begin{cases} \frac{1}{2}(\varepsilon_t^i)^2 & \forall |\varepsilon_t^i| \leq \tau \\ \tau(|\varepsilon_t^i| - \frac{1}{2}\tau) & \text{otherwise} \end{cases} \tag{5.16}$$

where $\tau \in \mathbb{R}^+$ is a hyperparameter that controls the effect of outliers. The huberized estimation of β is obtained: $\hat{\beta}_t^* = \arg\min_{\beta_t \in \mathbb{R}} \left\{ \|\tilde{y}_t - d_t \beta_t - X_t^\top \alpha\|_2^2 \right\}$.

5.4.4 Kalman filtering

To efficiently obtain estimates of the unobserved components and to refine prior estimates, we treat the parameter β_t of Equation 5.5 in a state space form and

apply Kalman filtering. Extending the classic random-effects models and imposing exogeneity of the regressors with respect to the time component such that $\beta_t \sim \mathcal{N}(0, \sigma_\beta^2)$ with $\mathsf{E}[\beta_t] = 0$, we rewrite (5.5) as $y_t = X_t^\top \alpha + \eta_t$ with a composite error term $\eta_t = (\eta_t^1, \ldots, \eta_t^{N_t})^\top$ and

$$\eta_t = d_t \beta_t + \varepsilon_t \tag{5.17}$$

where $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon{}^2)$. σ_ε is interpreted as idiosyncratic volatility such that it reflects the variation around the predicted price using the market index β_t and characteristics X_t.

Given the recursive nature and normality assumption of error terms, we fit an autoregressive order one process, AR(1), of β_t with a random walk:

$$\beta_t = \phi \beta_{t-1} + \xi_t \tag{5.18}$$

with $|\phi| \leq 1$ and $\xi \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\xi{}^2)$. As $\phi = 1$, ξ_t is the excess over the average price – interpreted as returns of a market portfolio – and σ_ξ is hence seen as market volatility.

Given the linear Gaussian state space representation in 5.17 and 5.18, we now use Kalman filtering to access the underlying β_t via the parameter estimation of σ_ε and σ_ξ. We use MLE to calibrate these parameters. Denote the unkown parameter vector by $\theta = (\sigma_\varepsilon, \sigma_\xi)$ within the parameter space $\Theta = \{\theta : \sigma_\varepsilon^2 > 0, \sigma_\xi^2 > 0\}$. ϕ can be estimated jointly with σ_ε and σ_ξ while the AR(1) process of Equation 5.18 is stationary, i.e. $|\phi| < 1$. Here, we explicitly feature the unit root process in Equation 5.18 as $\phi = 1$.

Denote conditional mean and variance by $\eta_{t|t-1}$ and $\Sigma_\eta(t|t-1)$, respectively. Let $e_t(\theta) = \eta_t - \eta_{t|t-1}$ and $\Sigma_t(\theta) = \Sigma_\eta(t|t-1)$. We therefore define the log-likelihood function of θ as:

$$\mathcal{L}(\theta) = \frac{1}{2} \sum_{t=1}^{T} \left\{ \log |\Sigma_t(\theta)| + e_t(\theta)^\top \Sigma_t(\theta)^{-1} e_t(\theta) \right\} \tag{5.19}$$

and the maximum likelihood estimator is:

$$\hat{\theta} = \arg \max_{\theta \in \Theta = \mathbb{R}_+^2} \mathcal{L}(\theta) \tag{5.20}$$

Obtained $\hat{\theta}$ through MLE we predict $\beta_{t|t-1}$ and $\eta_{t|t-1}$ and correct the prediction with $\sigma_\beta(t|t-1)$ and $\Sigma_\eta(t|t-1)$. Given the full sample information $(t = 1, \ldots, T)$, we therefore infer the underlying β_t.

5.4.5 DCS-t filtering

The Kalman filtering is optimal in a linear Gaussian state space model. In the presence of outliers, however, it may overweight the impact of these outliers on the changes in the index β_t. In consequence, we proposed to huberize residuals. Instead of Huberization, an alternative is based on dynamic conditional score (DCS) models introduced by Creal et al., 2011, 2013 and Harvey, 2013. The idea is to replace the one step ahead prediction error in the updating equation of the Kalman filtering by a likelihood score with respect to the index. For heavy-tailed distributions, this simply downweighs the impact of extremes. For example, consider a multivariate t-distribution with ν degrees of freedom. Assuming a random walk for β_t , the model becomes:

$$
\begin{aligned}
\eta_t &= d_t\beta_t + v_t, \quad v_t \sim t_\nu(0, \sigma_v^2 I_{n_t}) &\text{(5.21)} \\
\beta_t &= \beta_{t-1} + \kappa u_{t-1} &\text{(5.22)} \\
u_t &= \frac{1}{w_t}\frac{\nu + n_t}{\nu}\frac{1}{\sigma_v^2}d_t^\top e_t &\text{(5.23)} \\
w_t &= 1 + \frac{1}{\nu\sigma_v^2}e_t^\top e_t &\text{(5.24)}
\end{aligned}
$$

Here, u_t is the score w.r.t. β_t, i.e.

$$u_t = \frac{\partial \log f_t}{\partial \beta_t}$$

where f_t is the conditional density of η_t,

$$\log f_t(\eta_t; \theta) = \log \Gamma(\frac{\nu + n_t}{2}) - \log \Gamma(\frac{\nu}{2}) - \frac{n_t}{2}\log(\pi\nu\sigma_v^2) - \frac{\nu + n_t}{2}\log w_t.$$

Figure 5.8 depicts the score u_t as a function of e_t for the case $n_t = \sigma_v = 1$, $\nu = 4$ and $\nu = 10$. We see that the downweighting is stronger the heavier the tails of the distribution. Note that in the Gaussian case ($\nu \to \infty$), the score reduces to $u_t = \frac{1}{\sigma_v^2}d_t^\top e_t$, which is analogous to Kalman filtering, and in which case there is no downweighting of outliers. Estimation of the model parameters is again performed by maximizing numerically the log likelihood function

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \log f_t(\eta_t; \theta)$$

with respect to $\theta = (\sigma_v^2, \kappa, \nu)^\top$.
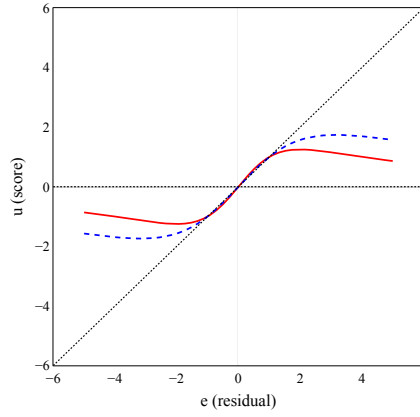
Figure 5.8: **Score u_t with respect to β_t as a function of the one-step prediction error for a multivariate t distribution with $\nu = 4$ (solid) and $\nu = 10$ (dashed) degrees of freedom**.

5.5   Data description

Our data are collected using the API from OpenSea which is the first and largest NFT marketplace. The data are available at the Blockchain Research Center (BRC). To consider the representative constituents, to not explode in sample size, we select the top 10 NFT art collections ranked by the total transaction volumes, which is 83.71% of the NFT art market volume in 2021 July. The data include two data frames – artwork metadata and transaction histories, which are then merged for the purpose of Hedonic regression, by the unique id of each work. Despite the continuous variables, i.e. number of sales and punk accessory, all the other variables are one-hot encoded. Table 5.A.1 shows the attributes of each variable.

The metadata describe the properties of each work, inclusive of the collection slug, contract type, scheme name, creator, year of creation and traits (i.e. characteristics of artworks). Among which creator ids and addresses are only available in the collections MakersPlace and SuperRare because the other collections considered in this study are algorithm generated works. We assign the number of works listed to the creators, i.e. the creators with only 1 piece, 2 to 100 pieces, 100 to 1,000 pieces, and more than 1000 pieces.

Collection specific properties are included as interactive dummies, for example, 'art_blocks_subcollection' in Art Blocks, 'bdap_Background' in Bored Ape Yacht Club. For those collections with too large set of words in traits – e.g. accessories for CryptoPunks – and with unstructured textual descriptions – e.g. SuperRare, we adopt the natural language processing (NLP) techniques to avoid the explosion in feature space. CryptoPunks are in 5 types and each work has a set of accessories – such

as headgears, glasses, and earrings – as of the total 92 categories. The traits within these categories are converted into the scarcity index using the concept of TF-IDF (term frequency inverse document frequency) in Jones, 1972. The traits of the p-th punk's sales are denoted by T_p where $p \in (1, 2, \ldots, P)$, i.e. $P = 5466$. The TF of a word w in trait T_p is computed by:

$$\text{TF}(w, T_p) = \frac{f_{w, T_p}}{\sum_{w' \in T_i} f_{w', T_p}}$$

where f_{w, T_p} is the raw count of a word in a given trait.

The IDF of the word is:

$$\text{IDF}(w, T_p) = \log \frac{P}{1 + \sum_{p=1}^{P} \mathbf{I}(w, T_p)}.$$

The scarcity index of T_p is defined as the sum of the product of TF and IDF of all the words within a trait as:

$$\text{SI}_{T_p} = \sum_{w \in T_p} \text{TF}(w, T_p) * \text{IDF}(w, T_p).$$

Having a consistent structure of CryptoPunks' traits, the scarcity of each punk is measured by summing the scarcity index of each accessory equipped. SI_{T_p} of each accessory is the importance of an accessory for a given punk, which is adjusted by the rarity of this accessory owned by the existing punks.

In SuperRare each artwork has a singe tag (i.e. a word) or a set of heterogeneous tags in its trait. The number of tags within these traits varies from the others. Most of these tags only appear once. We, thus, apply the pre-trained word embedding model – Global vectors for word representation (GloVe) from Pennington et al., 2014 to handle these inconsistent traits. Based on global word-word co-occurrence statistics from the corpus, the training result shows a linear substructure in word vector space. It implies that the similarity between words can be measured by the linear distance of two resulting word vectors, and these vectors are additive. We create GloVe index for each SuperRare trait T_s, $p \in (1, 2, \ldots, S)$,

$$\text{GI}_{T_s} = \frac{1}{f_{w, T_s}} \sum_{w \in T_s} \text{GloVe}_{w, T_s},$$

where f_{w, T_s} denotes the raw count of a tag within a given trait; and GloVe_{w, T_s} is the output of the pre-trained GloVe model with the input of word w, trained on Wikipedia corpuses in a 50 dimensional vector. Taking the weighted sum of the GloVe vectors for

each word, we obtain the GloVe index for each artwork. We use this index to consider the similarity among works in SuperRare. To further reduce the dimensionality, we apply k-means clustering on the GloVe indices and yield the dummy variables to represent the resulting 30 clusters.

Last, artworks' transaction histories among all the collections include the categorical variables (e.g. auction type, payment token, seller and winner addresses) and the continuous variables (e.g. number of historical sales). Categorical variables are one hot encoded, and continuous variables are aggregated into different intervals by the frequencies.

5.6   Result

As pointed out previously, we observe the frequent presence of outliers, and it significantly influences the stability and robustness of indices. In this paper, we have proposed two alternative procedures (in thick lines) – to resolve the overweighted impact from these outliers, see Figure 5.9. We first discuss how KF and DCS-t filtering perform in our empirical study and offer a simulation to show their difference in estimation performance. Later, we conduct the residual analysis to demonstrate how huberizing before applying KF assists to diminish the impact of outliers. Diametrically, considering the evolution of conditional distribution of an observation via score function, DCS-t filtering reduces such an impact in a dynamic fashion.
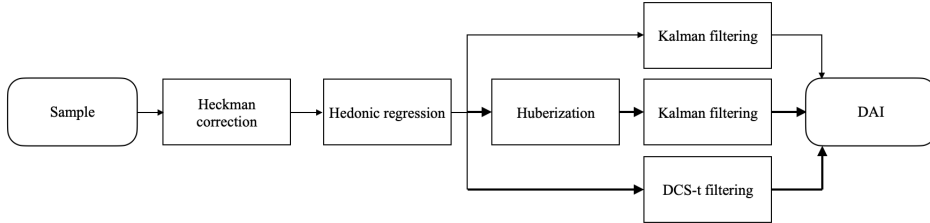
Figure 5.9: **Proposed procedures to handle the impact of outliers**.

5.6.1   KF and DCS-t

A substantial body of literature points out that the score-driven filter (i.e. DCS-t filtering) appears to be adaptive towards the estimation of time-varying parameters that depend on past observations in a wide class of nonlinear models (Artemova et al., 2022; Harvey, 2022). Related to parameter-driven models, DCS-t filtering replaces the one-step ahead prediction error ξ with the score of the Student-t density u. Figure 5.10 shows the update of score u in our empirical study of the NFT art market. It, hence, can deal with the intricate dynamics of time series – i.e. outlier errors, changing conditional variances, or non-negative variables (Harvey, 2013). In contrast, due to the

assumption of an accurate model of the state and prior distribution of noises, classic KF might fail to react to a sudden change of an observation. In Figure 5.11, the empirical result shows that $\hat{\beta}_{t|t}^{\text{KF}}$ is volatile during the first half observation period and afterward becomes smooth; whereas, $\hat{\beta}_{t|t-1}^{\text{DCS-t}}$ behaves in the opposite manner.
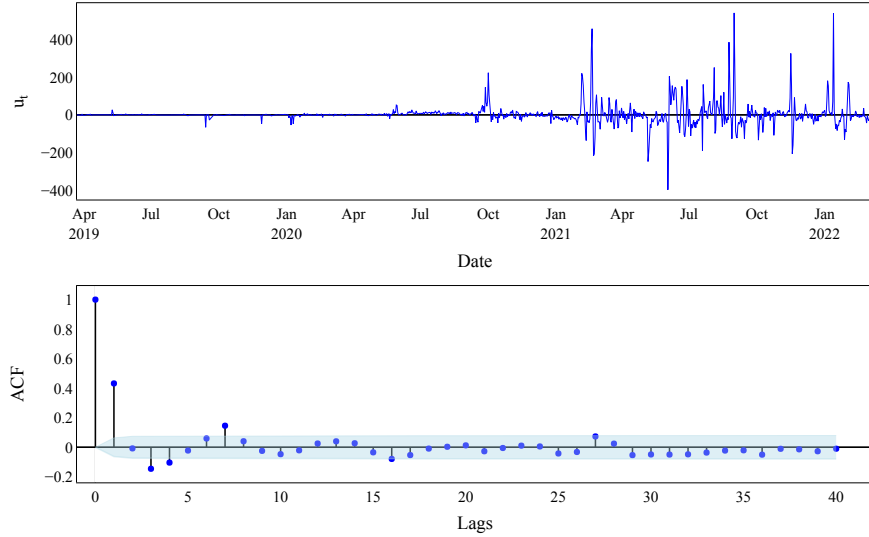
Figure 5.10: **The update and ACF of score u.**
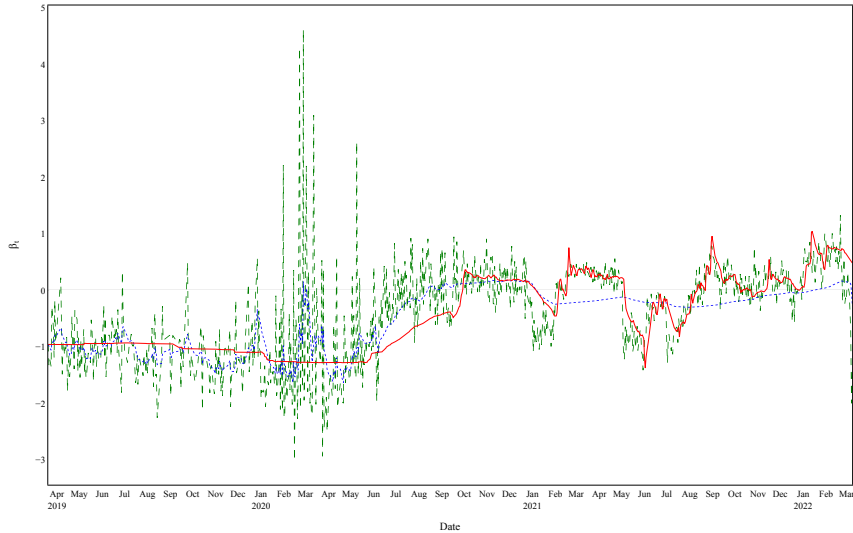

Figure 5.11: **KF predicted** $\hat{\beta}_{t|t-1}^{\textbf{KF}}$ **(dashed), KF corrected** $\hat{\beta}_{t|t}^{\textbf{KF}}$ **(dotted), and DCS-t predicted** $\hat{\beta}_{t|t-1}^{\textbf{DCS-t}}$ **(solid).**

Offering a better resolution to compare them, we conduct simulation with an univariate state space model wherein the state variable follows an AR(1) process, see 5.17 and

5.18. At each time t there exist N_t observations and $t \in \{1, \ldots, 500\}$. Parameters $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon{}^2)$ and $\xi \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\xi{}^2)$ are the error terms for the observable and state variables, respectively. An example with $N_t = [5, 20]$, $\sigma_\varepsilon = 0.1$ and $\sigma_\xi = 0.2$ is shown in Figure 5.12. In Table 5.1, the KF achieves a better outcome than the DCS-t in terms of mean squared errors (MSE) if number of observations are fixed at each time t. Under random observations the DCS-t surpasses the KF while having higher standard deviations in the model. According to the simulation result, we find that the DCS-t adapts better a volatile time series.

Table 5.1: **DF and DCS-t simulation results**.

N_t	σ_ε	σ_ξ	MSE	
			KF	**DCS-t**
5	0.01	0.02	2.00e-04	2.64e-04
5	0.1	0.2	1.03e-02	2.56e-02
50	0.01	0.02	2.58e-03	4.17e-03
50	0.1	0.2	1.95e-01	2.19e-02
[5, 20]	0.01	0.02	7.66e-04	2.33e-04
[5, 20]	0.1	0.2	7.44e-02	2.75e-02
[20, 100]	0.01	0.02	2.72e-03	2.34e-04
[20, 100]	0.1	0.2	2.16e-01	2.45e-02
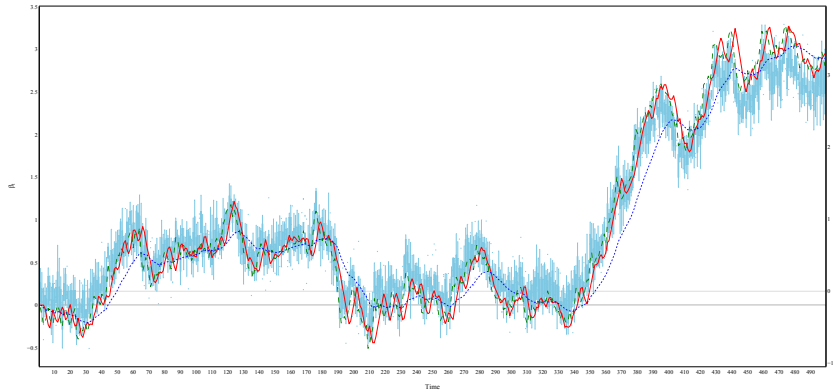
Figure 5.12: β_t (dashed), $\hat{\beta}_{t|t}^{\textbf{KF}}$ (dotted), $\hat{\beta}_{t|t-1}^{\textbf{DCS-t}}$ (solid) and residual box plot over time (background) with $N_t = [5, 20]$, $\sigma_\varepsilon = 0.1$ and $\sigma_\xi = 0.2$.

As discussed previously, the implementation of Huberization before applying Kalman filtering effectively avoids deterioration caused by outliers. Huber, 1964 shows that minimizing the Huber loss can be interpreted as MLE, i.e. its asymptotic variance $\mathsf{E}_F \psi^2(\varepsilon) / [\mathsf{E}_F \psi'(\varepsilon)]^2 \geq 1/\mathsf{E}_F[(f'(\varepsilon)/f(\varepsilon))^2]$ has the equality when $\psi \propto -f'/f$ where $\varepsilon \sim F$ with density f. That is, once we obtain the optimal threshold τ over time, it offers

robust estimates of β_t. However, an optimizing update like this can be computationally costly.

5.6.2 Residual analysis

The OLS estimation of DAI exhibits a fat-and heavy-tail behavior on residuals, see Figure 5.6. In order to achieve robustness and account for the influence of outliers, Huberization is proposed as a one-step robust regression. Given $\tau = q_{0.1}(\hat{\varepsilon})$, after Huberization the scale of residuals has shrunk significantly in Figure 5.13b, comparing to Figure 5.13a. Figure 5.14 illustrates that Huberization truncates the residuals over time, and the influence of outlying observations is significantly downweighted.

(a) After Heckman correction   (b) After Huberization with $\tau = q_{0.1}(\hat{\varepsilon})$
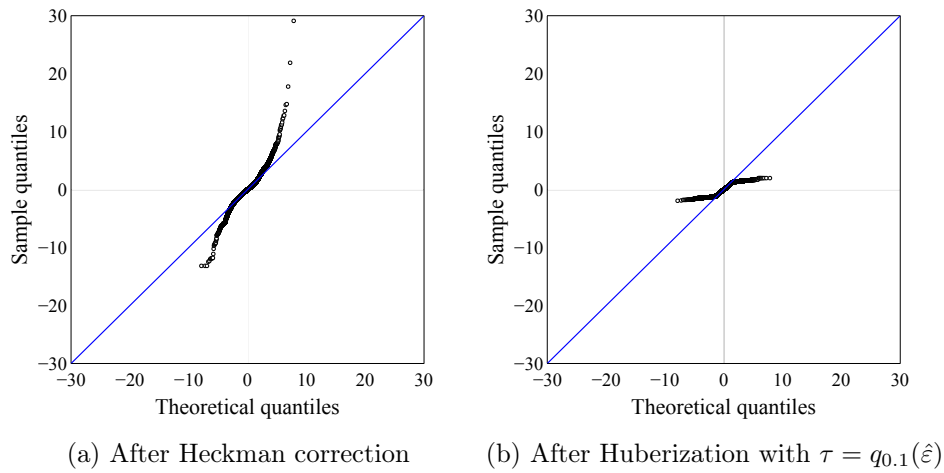
Figure 5.13: **QQ plots for OLS standard residuals against standard normal distribution after Heckman correction and after Huberization**.
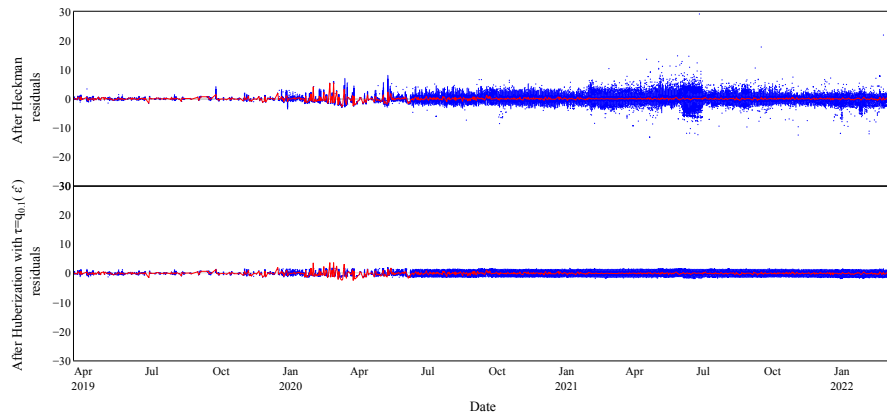

Figure 5.14: **Residual box plots** for after Heckman correction (upper panel) and after Huberization (lower panel) over time with mean of residuals.

Since τ here is a hyperparameter in Equation 5.16, we select three different quantile

levels of residuals after Heckman correction – $q_{0.1}(\hat{\varepsilon})$, $q_{0.05}(\hat{\varepsilon})$, and $q_{0.01}(\hat{\varepsilon})$ – to examine their sensitiveness to outliers, being against standard normal distribution. Figure 5.15 shows that Huberizing with the three quantile levels curve off in the extremities, which indicates fat tails are still observed. Given $\tau = q_{0.01}(\hat{\varepsilon})$ within theoretical quantiles around $[-3, 3]$ $\tau = q_{0.01}(\hat{\varepsilon})$ it demonstrates a wider scope of fitting, however, in the right extremities it has a spike where residuals aligned nearly horizontally. Both $\tau = q_{0.1}(\hat{\varepsilon})$ and $\tau = q_{0.05}(\hat{\varepsilon})$ have a slight jump on the right tail. All in all, $\tau = q_{0.01}(\hat{\varepsilon})$ resembles better the normal distribution compared with the others.


(a) $\tau = q_{0.1}(\hat{\varepsilon})$ (b) $\tau = q_{0.05}(\hat{\varepsilon})$ (c) $\tau = q_{0.01}(\hat{\varepsilon})$
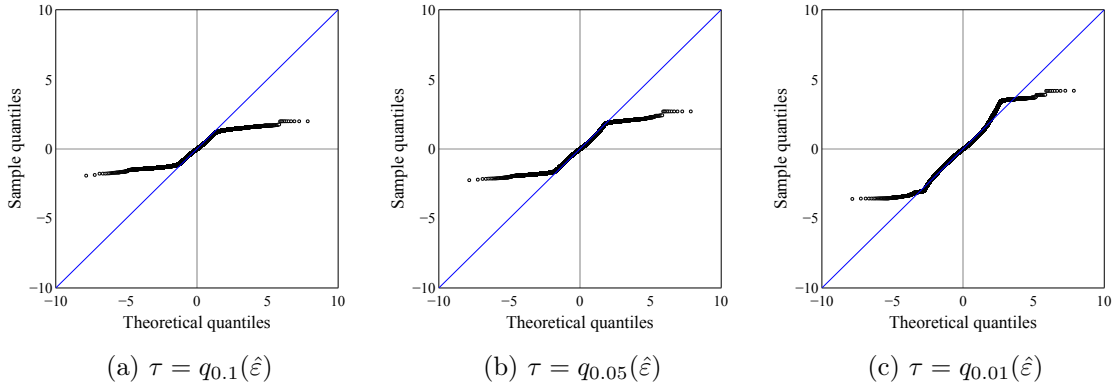
Figure 5.15: **QQ plots for OLS residuals against standard normal distribution after Huberization with different quantile levels**.

As an alternative to using Huberization and Kalman filtering, we propose DCS-t filtering to adaptively accommodate the updates of conditional distribution. Considering the presence of outliers or heavy tailed distributions, one takes a multivariate t-distribution, as in Equation 5.21. Here, we plot the QQ plot for the OLS residuals after Heckman correction against the standard t-distribution with two settings of degrees of freedom ν in Figure 5.16. As discussed in Figure 5.8 previously, having a heavier tail in the t-distribution (i.e. $nu = 4$) is more robust than the normal distribution. However, we observe some gaps in the density in the extremities. The larger is the degree of freedom in Figure 5.16b, the more is it approaches to the case of normal distribution as in Figure 5.13b.

5.6.3   DAI variants

In this section we showcase the price indices computed by both proposed procedures in different parameter settings.

Figure 5.17 illustrates the indices using Huberization with different settings on τ and filtering. The Heckman corrected index is highly influenced by extremes, i.e. the time 2020 March to April exhibiting a relatively sharp spike. Huberization with $\tau = q_{0.1}(\hat{\varepsilon})$

(a) $\nu = 4$ (b) $\nu = 10$
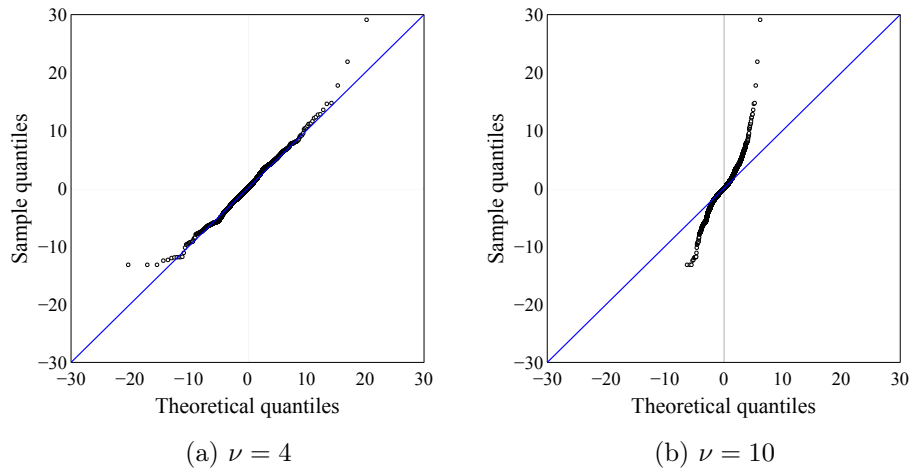
Figure 5.16: **QQ plot for after Heckman against standard t-distribution with two different ν.**

is slightly sensitive, compared to the DAI variants with $\tau = q_{0.05}(\hat{\varepsilon})$ and $q_{0.01}(\hat{\varepsilon})$. It is natural that with the rise of the sales in the NFT art market, see Figures 5.1 and 5.7, the DAI develops smoothly in its four variants.
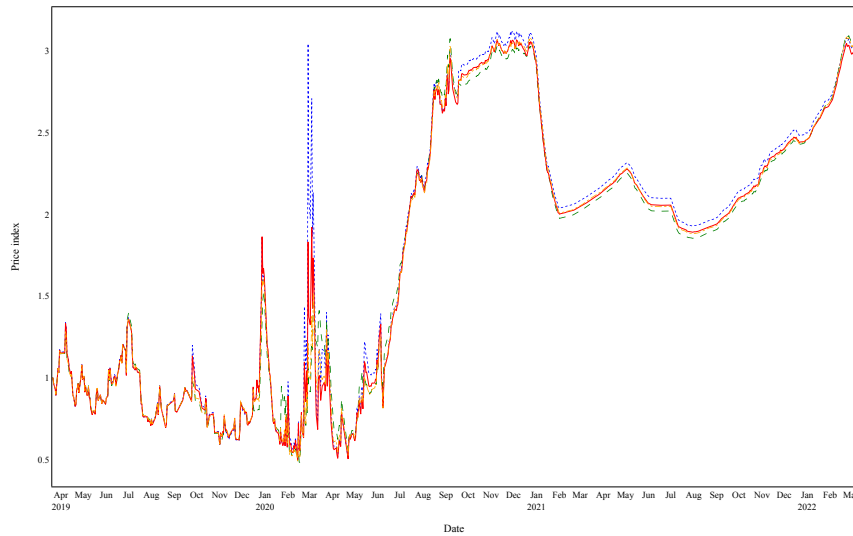
Figure 5.17: **DAI** − **after Heckman correction** (dotted), **after Huberization with $\tau = q_{0.1}(\hat{\varepsilon})$** (dashed), **after Huberization with $\tau = q_{0.05}(\hat{\varepsilon})$** (dash-dotted), and **after Huberization with $\tau = q_{0.01}(\hat{\varepsilon})$** (solid).

Offering an alternative to Huberization and filtering, in Figure 5.18 we compare the DAI variants using DCS-t filtering given optimized parameters and infinite degree of freedom to huberizing with $q_{0.01}(\hat{\varepsilon})$. After the forth quarter of 2020 the indices applying DCS-t filtering are more reactive. Setting $\nu \sim \infty$, the t-distribution in Equation 5.21

ideally should approach the normal distribution and smooth the index; however, it remains a certain degree of sensitiveness.
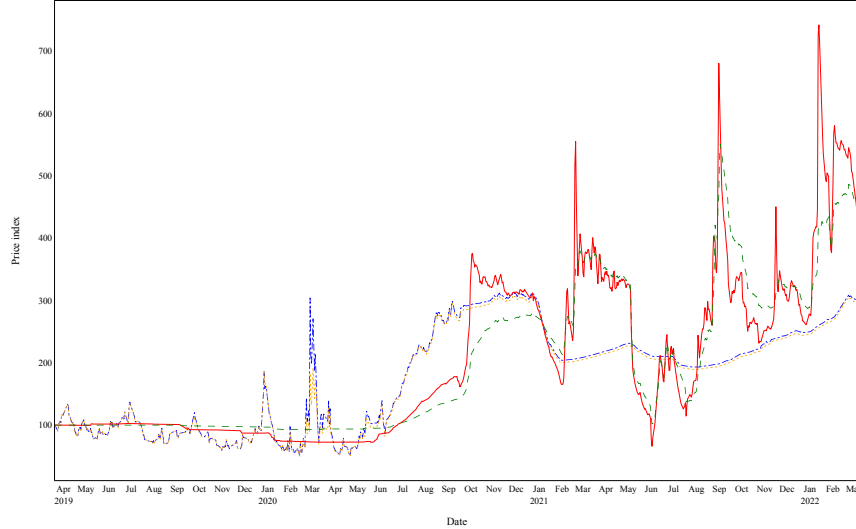

Figure 5.18: **DAI variants – after Heckman correction (dash-dotted), after Huber-ization with $\tau = q_{0.01}(\hat{\varepsilon})$ (dotted) DCS-t filtering (solid), and DCS-t filtering with $\nu \sim \infty$ (dashed).**

As has been shown previously in Section 5.6.2, both procedures provide robustness. Huberization in a one-step robust regression, crops off the influence from the outliers at once. Thereafter, we use Kalman filtering to approach the unobserved components. Instead, DCS-t filtering fits the data with a heavy-tailed distribution and gradually adjusts by one-step updates of conditional distributions. In this case, the alternative procedures are not that alternative, but rather complimentary. However, as the NFT art market is rapidly changing, including the dynamics of conditional distribution may be more adaptive and straightforward. In the following, we therefore only consider the DAI variants with DCS-t filtering.

5.7   The DAI

The works of NFT digital art are transacted by the means of cryptocurrencies, in particular ETH. In order to examine the dynamics between DAI and other relevant crypto variables like the CRIX, ETH or trading volume (ETH vol.), we explore their causal relationships. We showcase the price determinants for collected NFT artworks, concluding the result from Lasso as a variant. Focusing on art markets, we discuss DAI with the quarterly sales of – Post war, contemporary and ultra-contemporary markets.

5.7.1 Causal inference

Evaluate as the returns, $(p_t^i - p_{t-1}^i)/p_{t-1}^i$ where $i \in \{\text{DAI, CRIX, ETH, ETH vol.}\}$. Given two time series $y = \{y_t\}_{t=1}^T$ and $x = \{x_t\}_{t=1}^T$, we test whether including one can help to predict the other in (any) unidirectional (i.e. $x \Rightarrow y$, $y \Rightarrow x$) or bidirectional (i.e. $x \Leftrightarrow y$) fashion. That is, in 5.25 and 5.26 $\eta_t > \varepsilon_t \succ x \Rightarrow y$. This implies $\sigma^2(y|\mathcal{U}') < \sigma^2(y|\mathcal{U}' - x')$ where $\mathcal{U}' \subset \mathcal{U}$ is a set of prior values of the all causative variables; and $x' \subset x$ is the prior values of time series x. We examine each pair with different number of time lags, τ. There is no evidence that CRIX, ETH, and ETH vol. significantly cause DAI, as well as vice versa, see Table 5.2. Consequently, NFT art market can possibly act as a financial instrument for risk diversification towards cryptos as their causal influence on DAI is restricted.

$$y_t = \sum_{j=1}^{\tau} a_j y_{t-j} + \eta_t \tag{5.25}$$

$$y_t = \sum_{j=1}^{\tau} a_j y_{t-j} + \sum_{j=1}^{\tau} b_j x_{t-j} + \varepsilon_t \tag{5.26}$$

Table 5.2: **Pairwise Granger causality test results**.

Number of lags τ	1		2		3		4	
Null hypothesis H_0	F-test	p-value	F-test	p-value	F-test	p-value	F-test	p-value
CRIX $\nRightarrow$ DAI	0.0019	0.9649	0.2195	0.8030	0.1465	0.9319	0.5682	0.6858
DAI $\nRightarrow$ CRIX	0.0299	0.8628	0.1481	0.8624	0.2746	0.8437	0.2032	0.9366
ETH $\nRightarrow$ DAI	0.0690	0.7928	0.0795	0.9236	0.4170	0.7409	0.5071	0.7305
DAI $\nRightarrow$ ETH	0.0369	0.8477	0.0359	0.9647	0.2721	0.8456	0.3202	0.8645
ETH vol. $\nRightarrow$ DAI	1.0194	0.3131	0.4592	0.6320	0.4014	0.7521	0.5235	0.7185
DAI $\nRightarrow$ ETH vol.	0.3906	0.5323	0.7010	0.4965	0.7884	0.5007	0.6632	0.6178

5.7.2 Price determinant

We examine the regression results and investigate price determinants of NFT artworks by applying Hedonic Lasso regression to the pseudo observations $\tilde{y}_t$ from Huberization and the hedonic factors in Table 5.A.1. Having the penalty parameter $\lambda = 0.001$ in Lasso, we show the active variables and their coefficients in Tables B.1 and B.2. If a variable receives a positive coefficient, it is marked in orange; otherwise, it is in blue. In order to better explain price determination among these works, we categorize the active variables into two parts – external factors and internal factors.

External factors

External factors are relevant to transaction information which describe the trading details such as transaction time, inflow and outflow addresses, and number of sales occurred.

First, we see that the auction type – English auction – has a positive effect on price. English auction is an open-outcry, ascending dynamic bids. It is more common and straightforward for sellers and buyers. In terms of seller experience, a seller with an address having less than 100 transaction records has the upmost influence on price. An inexperienced seller may still have an opportunity to have a high sale price. That is, the market is still open to new entrants. Conversely, buyers who have experienced between 100 and 500 sales have a significant negative impact on price; whereas less experienced ones may contribute to a higher deal price. Those artworks owned by experienced owners or professional collectors who have between 100 and 1000 transaction records are likely to be sold at higher prices; while the opposite is also valid. Consequently, this may be due to the bidding skill and market sensibility of experienced buyers.

Moreover, we observe that trading with commonly used cryptos (i.e. ETH, WETH) and stable coins (i.e. DAI) may induce higher price. As NFT scheme ERC721 is the most common standard used in the market, it presents a positive impact here. The contract addresses are referred to the launching platforms – i.e. 0x8c9f364bf7a56ed058fc63ef81c6cf09c 833e656 links to the auction house, SuperRare. The influence, therefore, is relevant to the collection slugs, which we discuss later in the internal factors. Last, looking at the number of sales of works, we find it has a positive but limited influence.

Internal factors

These factors are corresponding to the intrinsic values of artworks. They present, i.e. at which collection or auction house a work has been created or published, by whom and when it is created, and what visual properties it has.

One of our major findings from the collection slugs is that artworks (e.g. in Rarible, Marketplace) created by independent artists are, in general, priced lower than institutional initiated works (e.g. in Autoglyphs by Lava Lab). This result corresponds to Figure 5.2 at which the medians of log price for Rarible and Marketplace are lower than the others. This may explain why in the external factors of Figure B.1 frequency of creator addresses (i.e. number of works created) and creator id negatively impact the price as the institutional initiated works are more in favor. Interestingly, these institutional initiated works within the same collection tend to have a similar visual

composition; they are generated by an algorithm. Also, the finding might be against the pricing mechanism in the conventional art market (Renneboog & Spaenjers, 2013). For example, the reputation of an artist or the authenticity evidence (the signature of the artist) usually raise the price. Instead, the price of an NFT artwork is highly influenced by its publisher and collection. We find that the works created during the last 3 years induce a lower price. This might be due to the oversupply and uneven distribution of artworks among collections, similarly to the tragedy of the commons in Hardin, 1998.

Spotting the interactive variables that describe the properties of each work in each collection, we conclude as follows:

Art Blocks is a collection of demand-oriented generative works. Buyers select their desired style and later receive a corresponding randomly generated work from an algorithm. The artwork can be a static image, 3D model, or an interactive animation. Consequently, the selection of style is crucial within this collection, see in art_blocks_subcollection of Table B.2. For instance, the styles in Archetypes and Ringers give a positive influence on price. In addition, the works from Art Blocks section – Playground, which includes curated artists' experimental projects impact positively In contrast, the works in the section Factory, which has no curation and valuation process, have a negative influence.

Bored Ape Yacht Club (BAYC) is a collection of 10,000 unique Bored Ape NFTs, as well as a membership of an online community which grants access to members-only benefits. For this collection, features such as background, earring, fur, hat, eyes, clothes and mouth of an ape all matter in the pricing of a work. Apes with bolder and more scarce features – i.e. trippy fur, king's crown, beam or laser eyes, black suit, grin multi-colored mouth – raise the price. This result might relate to the buyers' preference for unique features that may work as status symbols within the online community.

Hashmasks are created by over 70 artists worldwide. The collection includes 16,384 digital portraits. Item of 'Shadow Monkey'; masks of 'Abstract', 'Animal' and 'Steampunk'; characters of 'Golden Robot', 'Mystical' and 'Puppet'; skin color of 'Freak'; eye color 'Heterochromatic'; backgrounds of 'Doodle' and 'Pixel' have a positive impact on price. However, Hashmasks works having an indistinguishable appearance or features (e.g. no items, male and female characters) lead to a lower price.

CryptoPunks is a collection with limited 10,000 unique collectible characters. It is one of the earliest adopters of an NFT with the ERC-721 standard. For this collection, the scarcity of attributes plays a significant role in pricing. The variable 'punk_accessory' is based on the scarcity index for the accessories of punks (see Section 5.5) and has

a significantly positive impact. The most common punk type 'Male' has a negative impact.

For **SuperRare**, we cluster these tags into groups based on their similarity (see Section 5.5) and we find that these tags have a significant influence on price. Last, none of the patterns used in **Autoglyphs** are significant in their pricing. On a final note, we see that the scarcity of NFT artworks' internal factors is of importance in their price determination. Such a result resembles the conventional art market.

5.7.3   DAI and other art markets

In this section, we intend to interpret DAI and provide a narrative to understand its evolution through the emerging of the NFT art market. Applying Hedonic regression, we discount the implicit price given by the characteristics of an artwork; therefore, DAI, seen as a premium (characteristic-free price) of NFT artworks, represents better market trends. First, we examine the return of DAI and find that it becomes more volatile since 2021 as the growing in popularity of NFTs. Figure 5.19 presents a limited autocorrelation. As suggested by Goetzmann, 1995 for the conventional art market, there should exist the possibility of persistent trends – which can be considered as a measure of market efficiency captured by serial dependency in returns. That is, the NFT art market might be rather inefficient and immature and consequently suffers from an instantaneous uncertainty in terms of resale values. A tendency like this produces a price risk for investors.
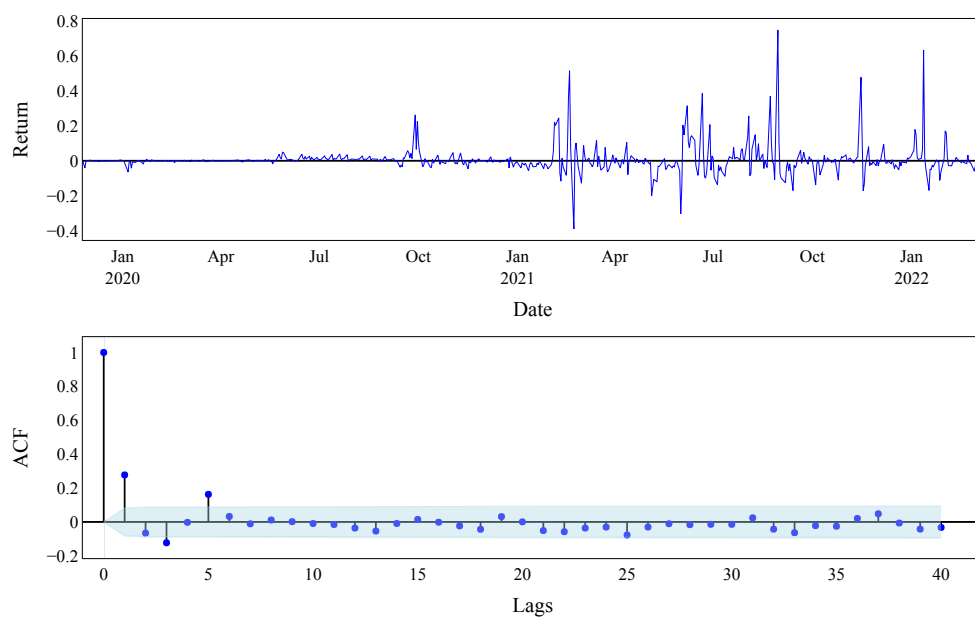

Figure 5.19: **Returns and autocorrelation function for DAI with** 95% **CIs.**

Due to the limitation of data availability for the other markets, we are not able to access their corresponding indices directly. Here, we illustrate DAI with the sales shares of different art markets – post war, contemporary, ultra-contemporary, and NFT art markets – plotting in the background of the index in Figure 5.20. During the outbreak of the COVID-19 pandemic in 2020, all the art markets receive a recession and these markets recover shortly after. As discussed by Mei and Moses, 2002b, during an economic recession, declines in art prices offer an opportunity to investors and these artworks might outperform the classic investment assets, i.e. stocks, equities, bonds.
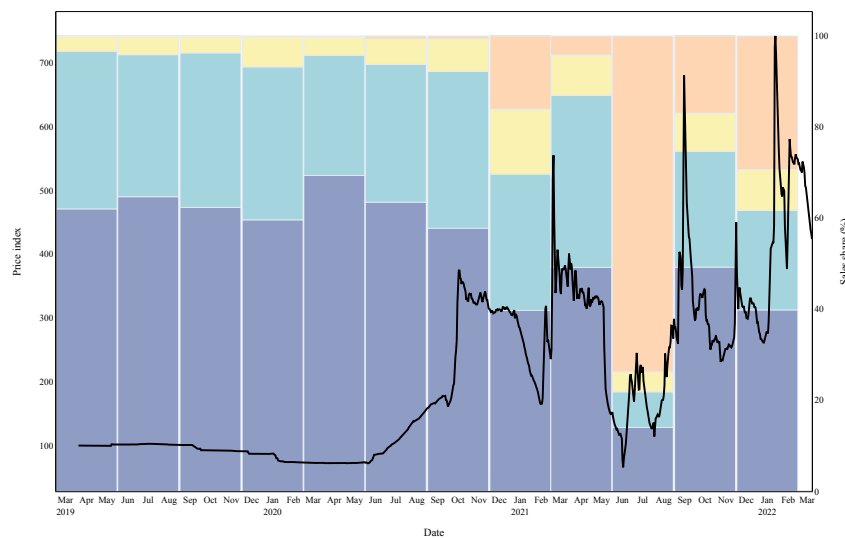

Figure 5.20: **DAI (price index) and the sales shares of** – **post war**, **contemporary**, **ultra-contemporary**, **NFT** **art markets (background)**.

The growing of NFT art sales brings up DAI since the end of 2020. In the mid of 2021 it starts suppressing the other markets, but DAI does not reflect this trend. The sales of NFT art market are not fully equivalent to the performance of the market on price or return. That is, DAI is essential to reflect this market. After the mid of 2021, we see several spikes on DAI and the sales of these four markets remain relatively steady. It is hard to distinguish how the NFT art would affect the other markets, yet we do not see it as a substitute or complementary to the conventional art.

5.8 Conclusion

The discussion regarding NFT art often comes to the point – *What is art? and what is the definition of aesthetics?* Experiencing the emerging of photography, Benjamin, 1968 describes the unique existence of artworks within a limited time period at a specific place, becomes an untenable category in the era of art reproduction. NFT

artworks are digitalized and highly accessible through the internet and replicable despite the ownership of tokens. While the aesthetics of digital artworks can be perfectly reproduced by simply copying and pasting, we might need to approach NFT art in a different manner. This can be seen in consumer and market behaviors that contrast to the conventional art market (Belk et al., 2022). This points out the essence of our study that we not only offer an index – DAI for such a unique market, but also seek to highlight the transformation of the overall art industry.

In this study, in order to construct a robust price index for heterogeneous digital artworks, we include the time-variant and time-invariant characteristics of each work through Hedonic regression. Avoiding the common issue for such a model: Selection bias, we apply Heckman correction to refine our samples. As outlying observations significantly impact the index, two alternative procedures – Huberization with Kalman filter and DCS-t filtering are proposed to robustify the results. Consequently, we see Huberization as a one-step robust regression that can be applied promptly and efficiently. Meanwhile, DCS-t filtering is a stepwise method which includes the changes of conditional distribution of an observation in the KF's state space formulation. Both procedures offer a degree of robustness to the indices, but under the fast changing environment of NFT art market we consider the index variant from the DCS-t filtering more adaptive and straightforward to the case.

Through the causality tests, we find that NFT art is an option for risk diversification to cryptos. Also, the return of DAI presents limited autocorrelation, which is different from the conventional art market. Looking into price determinants observed from Lasso, we conclude the possible price formulation with the external and internal factors. The external factors (i.e. payment tokens, frequency of seller addresses) influence much on price determination. For the internal factors, the traits of NFT artworks have limited impact and collection slugs are relatively influential. In addition, institutional initiated collections (i.e. CryptoPunks by Lava Lab) tend to have better pricing than the works from an independent creator. Whether the NFT art acts as a positive or negative stimulus to the other art markets is still in vague due to the inefficiency and immaturity of the market. This requires a follow-up study on its later development and applications. As one of the first indices of such an emerging market, we do not aim to define its role within the art or finance contexts. Instead, DAI is an index helping us to foster understanding and witness the market evolution. For future studies, it can be meaningful to look at the NFT art from the aesthetic perspective – e.g. visual autocorrelation (Özdilek, 2013) and visual link & knowledge discovery (Castellano et al., 2021) – to identify the value of images towards pricing.

Bibliography

Allen, S., Juels, A., Khaire, M., Kell, T., & Shrivastava, S. (2022). NFTs for art and collectables: Primer and outlook.

Ante, L. (2021). The non-fungible token (NFT) market and its relationship with bitcoin and ethereum. *Available at SSRN 3861106*.

Artemova, M., Blasques, F., van Brummelen, J., & Koopman, S. J. (2022). Score-driven models: Methodology and theory. In *Oxford research encyclopedia of economics and finance*.

Bailey, M. J., Muth, R. F., & Nourse, H. O. (1963). A regression method for real estate price index construction. *Journal of the American Statistical Association*, *58*(304), 933–942.

Beckert, J., & Rössel, J. (2013). The price of art: Uncertainty and reputation in the art field. *European societies*, *15*(2), 178–195.

Belk, R., Humayun, M., & Brouard, M. (2022). Money, possessions, and ownership in the metaverse: Nfts, cryptocurrencies, web3 and wild markets. *Journal of Business Research*, *153*, 198–205.

Benjamin, W. (1968). The work of art in the age of mechanical reproduction, 1936 (H. Zohn, Trans.). In, *Illuminations: Essays and reflections* (pp. 217–252). Houghton Mifflin Harcourt.

Bocart, F. Y., Ghysels, E., & Hafner, C. M. (2020). Monthly art market returns. *Journal of Risk and Financial Management*, *13*(5), 100.

Bocart, F. Y., & Hafner, C. M. (2015). Volatility of price indices for heterogeneous goods with applications to the fine art market. *Journal of Applied Econometrics*, *30*(2), 291–312.

Borri, N., Liu, Y., & Tsyvinski, A. (2022). The economics of non-fungible tokens. *Available at SSRN*.

Campbell, R. (2008). Art as a financial investment. *The Journal of alternative investments*, *10*(4), 64–81.

Castellano, G., Lella, E., & Vessio, G. (2021). Visual link retrieval and knowledge discovery in painting datasets. *Multimedia Tools and Applications*, *80*(5), 6599–6616.

Chainalysis. (2022). The chainalysis 2021 NFT market report: Everything you need to know about the NFT market and its most successful collectors. https://go.chainalysis.com/nft-market-report.html

Collins, A., Scorcu, A., & Zanola, R. (2009). Reconsidering hedonic art price indexes. *Economics Letters*, *104*(2), 57–60.

Court, A. (1939). *Hedonic price indexes with automotive examples*.

Creal, D., Koopman, S. J., & Lucas, A. (2011). A dynamic multivariate heavy-tailed
        model for time-varying volatilities and correlations. *Journal of Business &
        Economic Statistics, 29*(4), 552–563.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models
        with applications. *Journal of Applied Econometrics, 28*(5), 777–795.

Dailey, N. (2022). NFTs ballooned to a $41 billion market in 2021 and are catching up to
        the total size of the global fine art market. https://markets.businessinsider.com/
        news/currencies/nft-market-41-billion-nearing-fine-art-market-size-2022-1

Dowling, M. (2022a). Fertile land: Pricing non-fungible tokens. *Finance Research
        Letters, 44*, 102096.

Dowling, M. (2022b). Is non-fungible token pricing driven by cryptocurrencies? *Finance
        Research Letters, 44*, 102097.

Garay, U. (2021). Determinants of art prices and performance by movements: Long-run
        evidence from an emerging market. *Journal of Business Research, 127*, 413–426.

Ginsburgh, V., Mei, J., & Moses, M. (2006). The computation of prices indices. *Handbook
        of the Economics of Art and Culture, 1*, 947–979.

Goetzmann, W. N. (1995). The informational efficiency of the art market. *Managerial
        Finance*.

Goetzmann, W. N., Renneboog, L., & Spaenjers, C. (2011). Art and money. *American
        Economic Review, 101*(3), 222–26.

Goldberg, M., Kugler, P., & Schär, F. (2021). The economics of blockchain-based
        virtual worlds: A hedonic regression model for virtual land. *Available at SSRN
        3932189*.

Griliches, Z. (1971). Introduction: Hedonic price indexes revisited.

Griliches, Z. (1961). Hedonic price indexes for automobiles: An econometric of quality
        change.

Hardin, G. (1998). Extensions of" the tragedy of the commons". *Science, 280*(5364),
        682–683.

Härdle, W. (1984). Robust regression function estimation. *Journal of Multivariate
        Analysis, 14*(2), 169–180.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: With applications
        to financial and economic time series* (Vol. 52). Cambridge University Press.

Harvey, A. C. (2022). Score-driven time series models. *Annual Review of Statistics and
        Its Application, 9*, 321–342.

Heckman, J. J. (1976). The common structure of statistical models of truncation,
        sample selection and limited dependent variables and a simple estimator for
        such models. In *Annals of economic and social measurement, volume 5, number
        4* (pp. 475–492). NBER.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153–161.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*(1), 73–101, 29.

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*.

Koford, K., & Tschoegl, A. E. (1998). The market value of rarity. *Journal of Economic Behavior & Organization*, *34*(3), 445–457.

Kräussl, R., & Tugnetti, A. (2022). Non-fungible tokens (NFTs): A review of pricing determinants, applications and opportunities. *Applications and Opportunities (May 17, 2022)*.

Kumar, V., Ramachandran, D., & Kumar, B. (2021). Influence of new-age technologies on marketing: A research agenda. *Journal of Business Research*, *125*, 864–877.

Le Fur, E. (2020). Dynamics of the global fine art market prices. *The Quarterly Review of Economics and Finance*, *76*, 167–180.

Li, Y., Ma, M. X., & Renneboog, L. (2021). Pricing art and the art of pricing: On returns and risk in art auction markets. *European Financial Management*.

Malpezzi, S., et al. (2003). Hedonic pricing models: A selective and applied review. *Housing economics and public policy*, *1*, 67–89.

McAndrew, C. (2010). *Fine art and high finance: Expert advice on the economics of ownership* (Vol. 36). John Wiley & Sons.

McEachern, W. A. (2016). *Economics: A contemporary introduction*. Cengage Learning.

Mei, J., & Moses, M. (2002a). Art as an investment and the underperformance of masterpieces. *American Economic Review*, *92*(5), 1656–1668.

Mei, J., & Moses, M. (2002b). Economic activity and painting performance.

Mekacher, A., Bracci, A., Nadini, M., Martino, M., Alessandretti, L., Aiello, L. M., & Baronchelli, A. (2022). How rarity shapes the NFT market. *arXiv preprint arXiv:2204.10243*.

Miranda, A., & Rabe-Hesketh, S. (2006). Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. *The stata journal*, *6*(3), 285–308.

Nadini, M., Alessandretti, L., Di Giacinto, F., Martino, M., Aiello, L. M., & Baronchelli, A. (2021). Mapping the NFT revolution: Market trends, trade networks, and visual features. *Scientific reports*, *11*(1), 1–11.

Özdilek, Ü. (2013). Visual autocorrelation of prices. *Journal of Cultural Economics*, *37*(2), 203–223.

Pawelzik, L., & Thies, F. (2022). Selling digital art for millions-a qualitative analysis of NFT art marketplaces.

Penasse, J., & Renneboog, L. (2021). Speculative trading and bubbles: Evidence from the art market. *Management Science.*

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543.

Rengers, M., & Velthuis, O. (2002). Determinants of prices for contemporary art in dutch galleries, 1992–1998. *Journal of cultural economics*, *26*(1), 1–28.

Renneboog, L., & Spaenjers, C. (2013). Buying beauty: On prices and returns in the art market. *Management Science*, *59*(1), 36–53.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of political economy*, *82*(1), 34–55.

Shiller, R. J. (2008). Derivatives markets for home prices.

Towse, R., & Hernández, T. N. (2020). *Handbook of cultural economics*. Edward Elgar Publishing.

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

Urom, C., Ndubuisi, G., & Guesmi, K. (2022). Quantile return and volatility connectedness among non-fungible tokens (nfts) and (un)conventional assets. (2022-017). https://EconPapers.repec.org/RePEc:unm:unumer:2022017

Vasan, K., Janosov, M., & Barabási, A.-L. (2022). Quantifying NFT-driven networks in crypto art. *Scientific reports*, *12*(1), 1–11.

Velthuis, O. (2003). Symbolic meanings of prices: Constructing the value of contemporary art in amsterdam and new york galleries. *Theory and society*, *32*(2), 181–215.

Wallace, N. E., & Meese, R. A. (1997). The construction of residential housing price indices: A comparison of repeat-sales, hedonic-regression, and hybrid approaches. *The Journal of Real Estate Finance and Economics*, *14*(1), 51–73.

Waugh, F. V. (1928). Quality factors influencing vegetable prices. *Journal of farm economics*, *10*(2), 185–196.

Whitaker, A., & Kräussl, R. (2020). Fractional equity, blockchain, and the future of creative work. *Management Science*, *66*(10), 4594–4611.

Wilkoff, S., & Yildiz, S. (2022). The behavior and determinants of illiquidity in the non-fungible tokens (NFTs) market. *Available at SSRN 4097200.*

Worthington, A. C., & Higgs, H. (2004). Art as an investment: Risk, return and portfolio diversification in major painting markets. *Accounting & Finance*, *44*(2), 257–271.

Zorloni, A. (2005). Structure of the contemporary art market and the profile of italian artists. *International Journal of Arts Management*, 61–71.

Appendix

5.A Hedonic factors

Table 5.A.1: Data attributes.

Type of variables	auction_type	collection_slug	event_type	payment_tocken	to_account	seller_address	winner_address	contract_address
	dutch	art-blocks	cancelled	ABST	$(1000, \infty]$	$(1000, \infty]$	$(500, \infty]$	0x0ba51d9c015a7544e3560081ceb16ffe222d64f
	english	autoglyphs	successful	ASH	$(100, 1000]$	$(500, 1000]$	$(100, 500]$	0x131aebbfe55bca0c9eaad4ea24d386c5c082dd58
	min_price	beeple-everydays		ATRI	$[2,100]$	$[100,500]$	$[2,100]$	0x1f0678d1238921dc99309360668b16a17098aa2a
		boredapeyachtclub		DAI (crypto)	$= 1$	$[2,100]$	$= 1$	0x2947f98c42597966a0ec25e92843c09ac17fbaa7
		cryptopunks		ETH		$= 1$		0x41a322b28d0ff35404e2cbc676f0320d8c8850d
		hashmasks		KLTR				0x7a6425c9b3f55221bfa5d71df710a2fb805083f9b
Variable		makersplace		LUX				0x7be8076f4ea4a4ad080075c2508e481d6c946d12b
		rarible		NCT				0x7e3abde9d9e80fa2d1a02c89e0eae911233cde35
		superrare		RARI				0x7268357a8c2552623316e2562d90e642b b538e5
		wrapped-cryptopunks		TRSH				0x8c9f364bf7a56ed058fc63ef81c6cf09c833e656
				USDC				0x93f2a75d77162885f637f256da95e99ea28aafbe
				WETH				0xa5af48b105ddf2fa73cbaac61d420ea31b3c2a07
				WHALE				0xb47e3cd837ddf8e4c57f05d70ab865de6e193bbb
				nan				0xccd4ec7b66fbc029c116ba9ffb3e59351c20b5b06
								0xf2ee97405593bc7b62756825b03311694a48fedec7
								nan

Type of variables	asset_contract_type	schema_name	owner_address	creator_adress	art_blocks_section
	non-fungible	CRYPTOPUNKS	$(1000, \infty]$	$(2000, \infty]$	curated
	semi-fungible	ERC1155	$(500, 1000]$	$(1000, 2000]$	factory
Variable		ERC721	$[100,500]$	$[100,1000]$	nan
			$[2,100]$	$[2,100]$	playground
			$= 1$	$= 1$	

Continued on next page

Table 5.A.1: **Data attributes** (cont'd).

Type of variables	art_blocks_subcollection	autoglyphs_trait	bdap_Background	bdap_Earring	bdap_Fur	bdap_Hat
	All 27-Bit Digitals	Symbol Scheme: +|	Aquamarine	Cross	Black	Army Hat
	All 720 Minutes	Symbol Scheme: $\triangleleft$	Army Green	Diamond Stud	Blue	Baby's Bonnet
	All Aerial Views	Symbol Scheme: O	Blue	Gold Hoop	Brown	Bandana Blue
	All AlgoRhythms	Symbol Scheme: O|-	Gray	Gold Stud	Cheetah	Bayc Flipped Brim
	All Algobots	Symbol Scheme: X/$\triangle$	New Punk Blue	Silver Hoop	Cream	Bayc Hat Black
	All Apparitions	Symbol Scheme: ||	Orange	Silver Stud	Dark Brown	Bayc Hat Red
	All Archetypes	Symbol Scheme: ||O	Purple		Death Bot	Beanie
	All Bubble Blobbys	Symbol Scheme: ||-+	Yellow		Dmt	Bowler
	All CENTURYs	Symbol Scheme: $\rangle$			Golden Brown	Bunny Ears
	All Chromie Squiggles	Symbol Scheme: -/			Gray	Commie Hat
	All Construction Tokens				Noise	Cowboy Hat
	All Cryptoblots				Pink	Faux Hawk
	All Dreams				Red	Fez
	All Dynamic Slices				Robot	Fisherman's Hat
	All Elementals				Solid Gold	Girl's Hair Pink
	All Elevated Deconstructions				Tan	Girl's Hair Short
	All Fidenzas				Trippy	Halo
	All Frammentis				White	Horns
Variable	All Genesis				Zombie	Irish Boho
	All HyperHashs					King's Crown
	All Ignitions					Laurel Wreath
	All Inspirals					Party Hat 1
	All NimBuds					Party Hat 2
	All Ringers					Police Motorcycle Helmet
	All Singularitys					Prussian Helmet
	All Spectrons					Sm Hat
	All Subscapes					Safari
	All Synapses					Sea Captain's Hat
	All The Blocks of Arts					Seaman's Hat
	All Unigrids					Short Mohawk
	All Watercolor Dreams					Spinner Hat
	nan					Stuntman Helmet
						Sushi Chef Headband
						Trippy Captain's Hat
						Vietnam Era Helmet
						Ww2 Pilot Helm

Table 5.A.1: **Data attributes** (cont'd).

Types of variables	bdap_Eyes	bdap_Clothes	bdap_Mouth	hash_Item	hash_Mask	hash_Character
	3d	Admirals Coat	Bored	Book	Abstract	Female
	Angry	Bandolier	Bored Bubblegum	Bottle	African	Golden Robot
	Blindfold	Bayc T Black	Bored Cigar	Golden Toilet Paper	Animal	Male
	Bloodshot	Bayc T Red	Bored Cigarette	Mirror	Aztec	Mystical
	Blue Beams	Biker Vest	Bored Dagger	No Item	Basic	Puppet
	Bored	Black Holes T	Bored Kazoo	Shadow Monkey	Chinese	Robot
	Closed	Black Suit	Bored Party Horn	Toilet Paper	Crayon	
	Coins	Black T	Bored Pipe		Doodle	
	Crazy	Blue Dress	Bored Pizza		Hawaiian	
	Cyborg	Bone Necklace	Bored Unshaven		Indian	
	Eyepatch	Bone Tee	Bored Unshaven Bubblegum		Mexican	
	Heart	Caveman Pelt	Bored Unshaven Cigar		Pixel	
	Holographic	Cowboy Shirt	Bored Unshaven Cigarette		Steampunk	
	Hypnotized	Guayabera	Bored Unshaven Dagger		Street	
	Laser Eyes	Hawaiian	Bored Unshaven Kazoo		Unique	
	Robot	Hip Hop	Bored Unshaven Party horn		Unmasked	
	Sad	Kings Robe	Bored Unshaven Pipe			
	Scumbag	Lab Coat	Bored Unshaven Pizza			
	Sleepy	Leather Jacket	Discomfort			
	Sunglasses	Leather Punk Jacket	Dumbfounded			
	Wide Eyed	Lumberjack Shirt	Grin			
Variable	X Eyes	Navy Striped Tee	Grin Diamond Grill			
	Zombie	Pimp Coat	Grin Gold Grill			
		Prison Jumpsuit	Grin Multicolored			
		Prom Dress	Jovial			
		Puffy Vest	Phoneme ooo			
		Rainbow Suspenders	Phoneme L			
		Sailor Shirt	Phoneme Oh			
		Service	Phoneme Vuh			
		Sleeveless Logo T	Phoneme Wah			
		Sleeveless T	Rage			
		Smoking Jacket	Small Grin			
		Space Suit	Tongue Out			
		Striped Tee				
		Stunt Jacket				
		Tanktop				
		Tie Dye				
		Toga				
		Tuxedo Tee				
		Tweed Suit				
		Vietnam Jacket				
		Wool Turtleneck				
		Work Vest				

Table 5.A.1: **Data attributes** (cont'd).

Types of variables	hash_Skin Color	hash_Eye Color	hash_Background	hash_Glyph	creators	super_tag	punk_type	created_year
	Blue	Blue	Book	Egyptian Hieroglyph	$(1000, \infty]$	cluster_0	Alien	2018
	Dark	Dark	Davinci	Greek Symbol	$(100, 1000]$	cluster_1	Ape	2019
	Freak	Freak	Doodle	Mannequin	$[2, 100]$	cluster_10	Female	2020
	Gold	Glass	Expressionist	Planetary	$= 1$	cluster_11	Male	2021
	Gray	Green	Mystery Night	Zodiac Sign		cluster_12	Zombie	
	Light	Heterochromatic	Pixel			cluster_13		
	Mystical	Mystical	Street Art			cluster_14		
	Steel	Painted	Waves			cluster_15		
	Transparent					cluster_16		
	Wood					cluster_17		
						cluster_18		
						cluster_19		
						cluster_2		
						cluster_20		
						cluster_21		
Variable						cluster_22		
						cluster_23		
						cluster_24		
						cluster_25		
						cluster_26		
						cluster_27		
						cluster_28		
						cluster_29		
						cluster_3		
						cluster_4		
						cluster_5		
						cluster_6		
						cluster_7		
						cluster_8		
						cluster_9		

5.B　Price determinants

Table B.1: **External factors**.

Type of variables	Variable	Coeff.
auction_type	english	0.5364
event_type	successful	-1.2119
payment_tocken	DAI (crypto)	4.5703
	ETH	1.6079
	WETH	1.3497
frequency of 'seller_address'	(500, 1000]	0.3941
	(100, 500]	0.8425
	[2, 100]	1.0586
	= 1	0.8854
frequency of 'winner_address'	(100, 500]	-0.1205
	= 1	0.2120
contract_address	0x2947f98c42597966a0ec25e92843c09ac17fbaa7	0.3141
	0x7be8076f4ea4a4ad08075c2508e481d6c946d12b	-0.2917
	0x7e3abde9d9e80fa2d1a02c89e0eae91b233cde35	0.3086
	0x8c9f364bf7a56ed058fc63ef81c6cf09c833e656	0.5100
	0xcd4ec7b66fbc029c116ba9ffb3e59351c20b5b06	0.0714
schema_name	ERC721	0.6775
frequency of 'owner_address'	(500, 1000]	0.5347
	(100, 500]	0.2085
	=1	-0.3033
frequency of 'creator_address'	(1000, 2000]	-0.7539
	(100, 1000]	-1.1146
	[2, 100]	-1.1746
	=1	-0.6828

Table B.1: External factors (cont'd)

Type of variables	Variable	Coeff.
frequency of 'creator_id'	(100, 1000]	-0.1855
num_sales		0.0023

Table B.2: **Internal factors**.

Type of variables	Variable	Coeff.
collection_slug	autoglyphs	2.9271
	beeple-everydays	2.4190
	boredapeyachtclub	1.2886
	cryptopunks	1.4028
	hashmasks	0.4400
	makersplace	-0.7214
	rarible	-2.8638
	superrare	0.1091
	wrapped-cryptopunks	1.6581
art_blocks_section	factory	-1.3547
	playground	0.1033

Continued on next page

Table B.2: **Internal factors** (cont'd).

Type of variables	Variable	Coeff.
art_blocks_subcollection	All 27-Bit Digitals	-1.8213
	All 720 Minutes	-0.6192
	All Aerial Views	-3.0553
	All AlgoRhythms	-1.2175
	All Algobots	-0.1761
	All Apparitions	-2.2533
	All Archetypes	0.3477
	All Bubble Blobbys	-1.2125
	All CENTURYs	-1.2927
	All Chromie Squiggles	-1.6814
	All Cryptoblots	-2.3973
	All Dreams	-1.1674
	All Elementals	-1.1412
	All Fidenzas	-0.3117
	All Frammentis	-1.0715
	All Inspirals	-1.5496
	All Ringers	1.0678
	All Subscapes	-0.4786
	All Synapses	-1.4979
	All The Blocks of Arts	-1.6330
	All Watercolor Dreams	-2.1511
bdap_Background	Aquamarine	0.0337
	Gray	0.0022
bdap_Earring	Cross	0.0136
	Gold Stud	0.0142
	Silver Stud	0.0305

Continued on next page

Table B.2: **Internal factors** (cont'd).

Type of variables	Variable	Coeff.
bdap_Fur	Black	-0.0437
	Brown	-0.0451
	Cheetah	0.0293
	Cream	-0.0210
	Dark Brown	-0.0066
	Death Bot	0.1723
	Solid Gold	0.2578
	Trippy	0.8243
bdap_Hat	Bayc Hat Black	0.0133
	Bayc Hat Red	0.0009
	Fisherman's Hat	-0.0586
	King's Crown	0.4068
	Seaman's Hat	-0.0071
	Trippy Captain's Hat	0.0239
bdap_Eyes	3d	0.0133
	Blue Beams	0.3318
	Closed	-0.1473
	Crazy	-0.0127
	Laser Eyes	0.2793
	Sleepy	-0.0187
	Wide Eyed	-0.0647

<div align="center">Continued on next page</div>

Table B.2: **Internal factors** (cont'd).

Type of variables	Variable	Coeff.
bdap_Clothes	Black Suit	0.2845
	Black T	-0.0118
	Guayabera	-0.0083
	Hip Hop	0.0042
	Pimp Coat	0.0093
	Tie Dye	0.0450
bdap_Mouth	Bored	-0.0795
	Bored Cigarette	-0.0063
	Bored Unshaven	-0.1011
	Bored Unshaven Cigarette	-0.0240
	Dumbfounded	-0.0428
	Grin	-0.0147
	Grin Multicolored	0.0836
hash_Item	Golden Toilet Paper	0.2682
	No Item	-0.6320
	Shadow Monkey	0.1354
hash_Mask	Abstract	0.0270
	Animal	0.0388
	Basic	0.0156
	Doodle	-0.0598
	Indian	-0.0169
	Steampunk	0.0755

Table B.2: **Internal factors** (cont'd).

Type of variables	Variable	Coeff.
hash_Character	Female	-0.0339
	Golden Robot	0.9744
	Male	-0.0508
	Mystical	1.0136
	Puppet	0.1841
hash_Skin Color	Dark	-0.0212
	Freak	0.1429
hash_Eye Color	Blue	-0.3567
	Dark	-0.4657
	Green	-0.3512
	Heterochromatic	0.0739
hash_Background	Doodle	-0.0065
	Pixel	0.0492
super_tag	cluster_10	0.0369
	cluster_18	-0.3642
	cluster_2	-0.0648
	cluster_22	0.0207
	cluster_23	0.0113
punk_type	Male	-0.1262
created_year	2019	-2.8126
	2020	-1.5785
	2021	-1.0743
punk_accessory		0.0011

Declaration

I hereby declare that I completed this work without any improper help from a third party and without using any aids apart from those cited. All ideas derived directly or indirectly from other sources are identified as such. The results of Chapter 2 are based the publication with Kainat Khowaja, Cathy Yi-Hsuan Chen and Wolfgang Karl Härdle, which appeared in Advances in Quantitative Analysis of Finance & Accounting (AQAFA) in 2021. The results of Chapter 3 are based on joint work with Cathy Yi-Hsuan Chen. Chapter 4 is based on joint work with Yifu Wang, Wanbo Lu, Rui Ren, and Wolfgang Karl Härdle. Finally, Chapter 5 is based on a collaboration with Bingling Wang, Fabian Bocart, Christian Hafner, and Wolfgang Karl Härdle.

I testify through my signature that all information that I have provided about resources used in the writing of my doctoral thesis, about the resources and support provided to me as well as in earlier assessments of my doctoral thesis corresponds in every aspect to the truth.

Berlin, den 17.03.2023   Min-Bin Lin

Declaration on Contribution

Cumulative dissertation – declaration on co-authors, own contribution, and publication status.

Chapter	Paper title	Names of co-authors	Declaration of own contribution	Publication status (when/where)
2	Blockchain Mechanism & Distributional Characteristics of Cryptos	Kainat Khowaja, Cathy Yi-Hsuan Chen, Wolfgang Karl Härdle	0.6	2021 / Advances in Quantitative Analysis of Finance & Accounting (AQAFA), Vol. 18
3	Blockchain: An Invisible Hand for Crypto? An Empirical Discussion on Ethereum	Cathy Yi-Hsuan Chen	1.0	Working paper
4	Cross-exchange Crypto Risk: A High-frequency Dynamic Network Perspective	Yifu Wang, Wanbo Lu, Rui Ren, Wolfgang Karl Härdle	0.5	Under review by Journal of Business Economics & Statistics
5	DAI Digital Art Index	Binling Wang, Fabian Bocart, Christian Hafner, Wolfgang Karl Härdle	0.9	Under review by Journal of Financial Econometrics

Note: Chapter 1 is the introduction for the thesis.