A Quantitative Study of the Activity of Chainlink Transactions on the Ethereum Blockchain

July 21st, 2023

SMART CONTRACTS LAB

Abstract

This study examines the fundamental aspects of the Chainlink ecosystem on the Ethereum blockchain: The drivers of the LINK token price, the impact of the switch in consensus mechanism on transaction fees, and the degree of decentralization within the Chainlink network. Investigating the dynamics in digital currency prices shows that the number of oracle transactions processed on the blockchain does not indicate any price increase in the utility token LINK. Further, the analysis discovers that Ethereum's switch from PoW to PoS has resulted in a decrease in transaction fees. It also draws attention to broader implications beyond fees, such as those that affect network performance, market player behavior, security dynamics, and network congestion. Moreover, the extent of decentralization in the Chainlink network measured lower Gini coefficients and HHIs after the OCR model's implementation in early 2021, showing that decentralization has increased, as indicated by the data. It is crucial to remember that the results are restricted to the 84 observed node addresses and could not accurately reflect the complete Chainlink network. Furthermore, the absence of potentially important nodes should be taken into account.

Table of Contents

List of Figures & Tables

List of Abbreviations

BRM	Basic Request Model
DAO	Decentralized Autonomous Organization
DApps	Decentralized Applications
DDM	Decentralized Data Model
DeFi	Decentralized Finance
DON	Decentralized Oracle Network
EOA	Externally Operated Address
ETH	Ether
ETV	Ethereum Virtual Machine
HHI	Herfindahl-Hirschman Index
ICO	Initial Coin Offering
IEO	Initial Exchange Offering
OCR	Off-Chain Reporting
OLS	Ordinary Least Squares
PoS	Proof of Stake
PoW	Proof of Work
P2P	Peer to Peer
RSI	Relative Strength Index
SCL	Smart Contract Lab
SMA	Simple Moving Average
TPS	Transaction per Second
UZH	University of Zurich
VECM	Vector Error Correction Model

1. Introduction

Blockchain technology has enjoyed massive success in the past years. It can be considered a public ledger; all transactions are recorded on the blocks. To ensure user security and maintain ledger integrity, it employs techniques like asymmetric cryptography and distributed consensus algorithms. Some critical characteristics of blockchain technology are decentralization, immutability, persistence, and anonymity. Another factor is the absence of a single point of failure. These traits can significantly reduce costs, improve efficiency, and create transparency and trust. This allows for the elimination of the so-called middleman or intermediary. Thus, fields such as financial services, smart contracts, public services, and security services have made use of blockchain technology. (Zheng et al., 2020)

Smart contracts are self-executing lines of code, to put it simply. This derives a contract of shortcomings that are usually inherent in text-based ones. Man-made mistakes, like errors in the formulation of conditions or misunderstandings of the terms used, can be eliminated.

However, a persistent challenge hindering the widespread adoption of smart contracts is the lack of interoperability between off-chain data and on-chain decentralized applications and smart contracts. This is called “The Oracle Problem”. This problem consists of the inability of a blockchain to pull data from outside the network. Since smart contracts comprise around 90% of the use cases and generally need off-chain information, thus, creating a reliable bridge between the two ecosystems is essential. (Chainlink, 2022). Oracles are the most promising approach to overcoming this isolation and making these two ecosystems interoperable. At the same time, they serve as a possible single point of failure.

Although Chainlink is currently the biggest oracle operator in the market, there must be more literature regarding their activity. By analyzing oracle transactions of the past four years, this report aims to give transparency on some key aspects such as decentralization, the correlation between oracle transactions and the LINK price (Chainlink utility token), and token economics.

The study aims to analyze Chainlink oracle transactions on the Ethereum blockchain. The objective is to evaluate the performance and effectiveness of the oracle provider on the following metrics: Drivers of Digital Currency Prices, Token Economics, and Degree of Centralization. The thesis aims to analyze Chainlink oracle transactions on the Ethereum blockchain. The objective is to evaluate the performance and effectiveness of the oracle providers on the following metrics:

Drivers of Digital Currency Prices

The study aims to investigate the relationship between oracle transaction volume and the LINK token price. Specifically, it seeks to understand if there is a significant correlation between the two variables and, if so, to what extent. By examining this relationship, it aims to identify the factors driving LINK token price movements and provide insights for investors and stakeholders in the Chainlink ecosystem. Formally, the first hypothesis is:

> H_1: The number of oracle transactions processed on the blockchain does not correlate with the LINK token price.

Based on the hypothesis, Chainlink's oracle services are adopted and used because of factors other than the price of its native token, LINK. By testing this hypothesis, the study contributes to the broader academic discussion on cryptocurrency pricing and the role of network usage metrics in predicting asset values.

Token Economics

Ethereum recently altered its consensus algorithm from Proof-of-Work (PoW) to Proof-of-Stake (PoS). Sheth et al. (2019) claim that Proof-of-Stake may be the most promising consensus mechanism for cryptocurrencies and their underlying blockchains. Moreover, the transition to PoS has resulted in a significant reduction in energy consumption. According to Kapengut and Mizrach (2023), energy consumption has decreased by 99.98% since the transition. However, there may have been some economic implications for this transition related to network throughput and transaction fees. Thus, the second hypothesis states:

> H_2: The Ethereum transition from Proof-of-Work to Proof-of-Stake decreased Chainlink's oracle transaction fees (in ETH) as measured by the daily minimum transaction fees.

Degree of Decentralization

Since oracles are fundamental for integrating off-chain legacy systems with evolving on-chain ecosystems, ensuring the decentralization of oracle providers is particularly crucial. This decentralization enhances their resilience against potential manipulation and generally reduces the risk of being a single point of failure for the oracle. As part of the analysis, the degree of decentralization within the Chainlink ecosystem is examined. As a result, our third hypothesis is:

> H_3: The share of oracle providers within the Chainlink network is decentralized as measured by the Gini Coefficient and the Herfindahl Index.

Surprisingly enough, up until this date, no research measuring these metrics could be found. Most literature focuses on specific sectors such as FinTech or Decentralized Finance (DeFi). By testing the hypotheses, the study aims to give a better understanding of the Chainlink oracle network and to showcase its strengths and weaknesses, as well as the factors that lead to them. Furthermore, this report aims to build a stepping stone for future research on this topic and encourages an expansion of scope.

The report starts by giving a thorough literature review on oracles and their application. After providing a stable foundation, the three hypotheses and their backgrounds are explained. This directly leads to the methodology and data collection, followed by the testing of the hypotheses.

2. Literature Review

This section reviews existing literature and establishes the basic knowledge for understanding blockchain technology. Initially, Ethereum's smart contract platform is introduced, including its security, scalability, and other features that make it a popular choice among developers. Next, the features and benefits of Chainlink are discussed. This decentralized oracle network enables smart contracts to securely utilize off-chain data and services by leveraging its distributed infrastructure (Breidenbach et al., 2021). It has become increasingly popular to connect smart contracts to real-world data. Lastly, previous research on blockchain oracle transaction analysis is reviewed. Doing so provides context and background for our study and contributes to the existing body of knowledge within this field.

2.1 Ethereum

Vitalik Buterin first proposed Ethereum in 2013 as a decentralized, open-source, blockchain-based smart contract platform. It was officially launched in 2015, and similar to Bitcoin and other cryptocurrencies, Ethereum provides a secure, peer-to-peer transaction method without the need for intermediaries. Ethereum's native cryptocurrency is Ether (ETH), which is utilized to settle transaction fees and computational operations on the platform. While it may appear to be a digital currency, it is designed to be much more than that. Its main feature is its ability to enable the creation of decentralized applications (Dapps) and smart contracts without a central authority.

The Ethereum platform is built on the blockchain, which serves as a public ledger of all transactions and smart contract interactions. Using the Ethereum Virtual Machine (EVM), developers can write smart contracts in various programming languages, such as Solidity and Vyper. Ethereum is known for its ability to support complex and sophisticated smart contracts, such as those used in decentralized finance (DeFi), and its active developer community, which has developed numerous tools and libraries to facilitate smart contract development.

Ethereum is widely recognized as one of the most popular blockchain platforms for decentralized applications and smart contracts. As a result of its flexibility, scalability, and active developer community, it attracts a large number of developers. The following subsections will provide a more detailed discussion of smart contracts, Ethereum's security features, "The Merge", how a transaction works, and compare it to other blockchain platforms.

2.1.1 Smart Contracts

As early as the mid-1990s, smart contracts were conceptually introduced by computer scientist and cryptographer Nick Szabo. Szabo (1997) describes smart contracts as self-executing computer programs that enforce the terms of a contract automatically. Nevertheless, it was only with the advent of blockchain technology and the Ethereum platform, in particular, that smart contracts became feasible for practical use.

Unlike traditional contracts, smart contracts are autonomous agreements where conditions are embedded directly within the code itself. Once deployed, they are stored on a blockchain network, ensuring they cannot be altered. As with any emerging technology, smart contracts have their advantages and disadvantages. Through the use of smart contracts, it is possible to improve efficiency, transparency, and security while at the same time reducing costs and the need for intermediaries (Alshahrani et al., 2023; Zheng et al., 2020). Since smart contracts are derived from code, errors in the code can have significant consequences. If this occurs, the contract may not be executed correctly, which could result in legal or financial repercussions (Rizos, 2022). Moreover, due to the immutability of blockchain, a deployed smart contract or a verified transaction cannot be altered. Thus, once a smart contract is deployed, even if erroneous, it cannot be corrected, and severe financial losses may result (Andesta et al., 2019).

To terminate a smart contract on the blockchain network and redeem the contracts remaining Ether, it is necessary to implement a self-destruction function that can only be called from the address that initially deployed the smart contract. Upon discovering glitches in the code, It is common for developers to execute this function and to redeploy an updated version of the contract. Nonetheless, this feature may add to the complexity of the development process and present a potential attack vector for attackers. The use of self-destructive functions in smart contracts should therefore be carefully considered. (Chen, 2020)

Nevertheless, smart contracts are prevalent and offer many potential applications. They can be applied to various areas since they eliminate the need for an intermediary. Many financial processes can be improved using this innovation, including lending, borrowing, and trading (Bartoletti & Pompianu, 2017). In addition, smart contracts could be used in supply chain management to monitor the transit of goods and guarantee compliance with regulatory requirements (Saberi et al., 2019). Moreover, smart contracts may be used in voting systems to increase transparency and reduce election fraud (Russo et al., 2021).

Ethereum's smart contracts are typically written in Solidity, a high-level programming language created for the platform. Designed to be secure and easy to learn, it is a statically typed, contract-oriented language influenced by C++, Python, and JavaScript. Solidity supports various data types and provides built-in functions for everyday tasks. Nevertheless, its relative immaturity has led to several high-profile smart contract hacks and vulnerabilities. For instance, in 2016, a vulnerability in the code of the DAO, a smart contract built on the Ethereum blockchain, enabled an attacker to steal over $50 million worth of Ether tokens (Mehar et al., 2019). This incident resulted in a hard fork of the Ethereum blockchain, creating Ethereum Classic. Despite this, Solidity continues to be a popular programming language for creating smart contracts on Ethereum.

2.1.2 Security Features

Security is a critical concern for any smart contract platform, and Ethereum has implemented several features to mitigate potential vulnerabilities and attacks. Among the essential elements of the security of Ethereum was the Proof of Work (PoW) consensus mechanism, which required users to solve complex cryptographic puzzles to validate transactions and earn rewards

(Buterin, 2014). However, as of 2022, Ethereum has transitioned to a new consensus mechanism known as Proof of Stake (PoS). This transition has already been advocated by Buterin (2016) as early as 2016. Compared to PoW, the PoS architecture is considered more secure, less energy-intensive, and more suitable for implementing new scaling solutions (Nguyen et al., 2019). Proof-of-Stake requires that validators place a stake in their cryptocurrency to verify transactions. Those validators are then selected based on their stakes' size and reputation within the network.

Consequently, attackers will have more difficulty manipulating the network and executing fraudulent transactions (Nguyen et al., 2019). Furthermore, Ethereum offers developers a variety of tools and best practices to assist them in writing secure smart contracts, including the Solidity compiler, which checks for common vulnerabilities such as integer overflows and re-entrance attacks. As with any system, security breaches are always possible, and ongoing improvements to security features will be necessary as the platform develops.

2.1.3 The Merge

"The Merge" combines Ethereum's initial execution layer, known as the Mainnet since its inception, with the Beacon Chain, the newly introduced Proof-of-Stake consensus layer. This integration eradicated the requirement for energy-intensive mining and introduced the ability to secure the network through staked ETH. It marked a significant advancement toward achieving the Ethereum vision of enhanced scalability, security, and sustainability. (Ethereum, 2023). The switch did not influence the users' wallets, funds, or accounts, which meant the users required no action.

Furthermore, the energy needed to propose and validate a block remains constant regardless of its number of transactions. The notion of per-transaction energy expenditure suggests that fewer transactions would result in lower energy consumption and vice versa, which needs to be revised, according to Ethereum, 2023. This implies that if the gas fees for a new block decrease, the transaction fees accordingly decrease.

2.1.4 The Lifecycle of a Transaction

This short chapter focuses on the transaction lifecycle to give the reader a basic understanding: As a first step, a user initiates a transfer of assets from their wallet address through their wallet to another wallet or contract address. The transaction gets a transaction ID, also known as a transaction hash or tx hash, which is assigned to this specific transaction created. It acts as a reference number, determining the transaction's status and details on block explorers such as Etherscan. The transaction then is broadcasted to the network into a pool with other transactions. Miners pick out transactions from this pool to be included in a block on the blockchain; this step determines the time the transaction will need. Two factors determine the time, the network traffic and the transaction gas fee, which will be the next chapter's subject. (Amir, 2021)

2.1.5 Gas Fees

Due to Ethereum being a decentralized blockchain without a singular authority overseeing its operation, a mechanism must be needed to prevent the network from getting congested or spammed. Ethereum (2023) defined this mechanism by charging a sender of a transaction with a gas fee, which then is rewarded to the nodes that validate the transaction on the network (block rewards). The gas fee is in the form of the Ethereum native currency, Ether (ETH), and is also used to enable value transfers, rewarding mining efforts, and execution of smart contracts. The gas fee is automatically deducted from the user's ETH balance. The gas price must be multiplied by the gas limit to calculate the gas fee. These transactions are named “Type 0” (Choi, 2021).

Since August 2021, the gas fee mechanics for Ethereum have been changed. It underwent a major network enhancement known as the “London Hard Fork”. Five Ethereum improvement proposals are contained in this upgrade. The most important one for the gas fees is the EIP-1559, which aims to provide a better fee estimation and reduce variance in times of high demand. These transactions are named “Type 2” (Choi, 2021). These transactions include:

Base Fee: Set by the network and consequently removed from circulation. (Ethereum, 2023)

Max Priority Fee: Optionally decided by the user and often referred to as the miner top, directly compensation the miners. (Ethereum, 2023)

Max Fee Per Gas: The upper limit that the user is willing to pay for the inclusion of their transaction into a block. (Ethereum, 2023)

In general, one regular transaction needs 21,000 GWEI to be executed. Complex smart contracts often require higher fees, sometimes ten to twenty times higher (Donmez & Karaivanov, 2021). Furthermore, Chiu and Koeppl (2019) suggest that more urgent transactions or smaller block sizes drive up prices.

2.1.6 Comparison with Other Blockchains

Emerging blockchain platforms such as Polygon and Avalanche offer promising alternatives to Ethereum. The Polygon protocol, formerly the Matic Network protocol, is a layer two scaling system built on top of Ethereum, allowing for faster and cheaper transactions while maintaining Ethereum's security and ecosystem (*What Is Polygon? | Polygon Wiki*, 2023). Avalanche, meanwhile, is a separate blockchain platform that utilizes a consensus mechanism called Avalanche consensus to achieve higher throughput and faster finality than Ethereum's PoW or PoS (*Avalanche Dev Docs: Create Without Limits*, 2023). In contrast to the other two platforms, Ethereum's ecosystem is one of the most mature and widely adopted, with an active community of developers and a range of tools and resources available to develop smart contracts. Ethereum has also served as the primary platform for Dapps (Onica & Amariei, 2022) and nonfungible tokens (NFTs) (Nadini et al., 2021), representing a significant portion of the cryptocurrency market capitalization. Due to its limited transaction processing capacity, high gas fees, and congestion issues during periods of high demand, Ethereum has encountered scalability difficulties (Chen et al., 2020; Dağli et al., 2023; Urquhart, 2021). These challenges can be addressed with Polygon and Avalanche. The Polygon scaling solution enables thousands of transactions per second to be processed with minimal fees and fast confirmation times while also being interoperable with Ethereum. Avalanche's consensus mechanism allows for transactions of up to 4,500 per second with a finality time of less than one second, making it an attractive option for high-throughput applications. Even though Polygon and Avalanche are relatively new and have smaller developer communities than Ethereum, they offer compelling alternatives for those seeking more scalability and faster transaction processing.

2.2 Previous Research

First, an overview of the foundations for the study is provided, focusing on two predecessor studies with the same focus. These two papers are the only similar studies conducted on

Chainlink activity. Fortunately, more literature covers the drivers of digital currency prices, token economics, and the degree of decentralization.

2.2.1 Foundations for the Current Study

Fast-paced development in the blockchain industry can be taken for granted. However, the resulting research is often a step or two behind. This is the case for oracle activity analysis. In short, there has yet to be published literature discussing Chainlinks activity. Caldarelli (2022) suggests that only 15% of the public literature on blockchain involves oracles. One of the difficulties for this research is the constantly changing nature of, in this case, the oracle provider. Models on how the nodes are compensated for the data, or even how the data is aggregated and forwarded, have changed more than once. These problems make it impossible to find comparable and appropriate literature or data for this thesis. However, there are two theses by Ewerhart and Kante (2021) and Hof (2023).

The first thesis by Ewerhart and Kante in 2021 examined the Chainlink oracle activities within the Ethereum blockchain. Transaction numbers, earnings, revenue, and digital currency trend influences on the Chainlink network were the main topics. They concluded that much of the public perception of actual transactional activity on Ethereum Blockchain involving Chainlink nodes might be unwarranted. The data suggest substantially lower transaction numbers and revenues than the ones publicly publicized by the Chainlink nodes. Furthermore, their data suggested that the hype surrounding the potential of Chainlink company and its affiliates to transform the oracle system had yet to materialize.

Hektor Hof reexamined the thesis two years later and discovered mistakes within the data set. The main problem consisted of considering only a fraction of the transactions, directly leading to faulty conclusions. The former dataset only included reward transactions within the Basic Request Model (BRM), including LINK rewards sent to oracle addresses, not node or admin wallet addresses. What caused this problem was the switch from BRM to Decentralized Data Model (DDM) and from oracle contracts to operator contracts. As a reminder, within the BRM, a data request is handled in a ratio of 1:1 between the consumer and the node. In contrast, DDM uses multiple nodes to update the data feeds.

Furthermore, the Off-Chain Reporting (OCR) aggregates the delivered data off-chain, and only then the node operator can access the updated data feeds. An operator contract brings more features, such as distributing funds to multiple addresses in a single transaction or multi-word responses. (*Operator | Chainlink Documentation*, 2022)

2.2.2 Drivers of Digital Currency Prices

Both digital currencies and traditional assets are used to store and exchange value. However, it is essential to note that digital currencies are decentralized and operate on a peer-to-peer network, meaning that they are subject to different regulations and have unique features, such as anonymity and security. As a result of these features, digital currencies are a subject of concern regarding the possibility of illicit activity involving them (Böhme et al., 2015). However, fast and low-cost transactions have increased their popularity (Almuraqab, 2020).

In much the same way as any other asset, the price of digital currencies is primarily determined by the forces of supply and demand. Kristoufek (2015) shows that the price will likely increase if there is a high demand for a particular digital currency and a limited supply. According to Sukamulja and Sikora (2018), In the long term, Bitcoin's supply does not influence its price fluctuation; however, it does exert an impact in the short term. Moreover, Kristoufek (2013) found a strong positive correlation between Bitcoin price and the number of searches for the term "Bitcoin" on Google, as well as a similar correlation with the number of views and edits of the Bitcoin Wikipedia page. The study suggests that Bitcoin prices are influenced by user interest and attention, similar to how other markets are affected by public opinion.

Contrary to traditional asset valuations, which typically discount future cash flows, Cong et al. (2020) developed a dynamic asset pricing model for cryptocurrencies. The model considers how much people want to purchase things with the token to determine the equilibrium value of tokens. Based on the authors' findings, the more people use the token, the more valuable it becomes since other people also wish to use it, and the adoption of the platform follows an S-curve pattern that starts slowly, rises in volatility, and eventually reaches its lowest point. The model offers a new perspective on the value of cryptocurrencies and has important implications for understanding the dynamics of digital platform adoption.

Meanwhile, digital currency markets are relatively new and unregulated, which makes them vulnerable to market manipulation. Large-scale traders can manipulate the price of digital currencies by buying or selling large amounts of them, which can cause a chain reaction in the market. Chen et al. (2019) revealed the phenomenon of market manipulation of digital currencies through mining the digital currency exchange trading network. The researchers examined the trading history of the Mt. Gox exchange as a case study. They found that one prominent trader named "Willy" purchased a significant amount of Bitcoin, leading to a sharp rise in its price.

It is important to note that not all digital currencies are native blockchain currencies, such as Bitcoin and Ethereum. Some digital currencies are regarded as utility tokens, such as LINK, the currency of Chainlink. According to Benedetti et al. (2023), utility tokens act as the exclusive mode of payment for blockchain-based services and products. Utility tokens, unlike native blockchain currencies, are based on the success of the underlying project or platform and the demand for the product or service they represent. Thus, factors driving the price of utility tokens may differ from those affecting native blockchain currencies. Rather than being solely influenced by broader market forces, Prat (2021) suggests that utility tokens' value is closely tied to their role as a means of access to platform services.

2.2.3 Token Economics

Token economics deals with designing and managing a digital currency's native tokens and the incentives and mechanisms that dictate their distribution, circulation, and value. As tokens are often the primary means of exchange and value transfer on blockchain networks, they are an essential technology component.

Ethereum's native token, Ether (ETH), was first distributed through an initial coin offering (ICO) in 2014. Since then, ETH has been distributed through mining rewards and other mechanisms, such as crowd sales, airdrops, and initial exchange offerings (IEOs). The popularity of ICOs has declined in recent years due to regulatory concerns and scams. According to Tiwari et al. (2020), 10% of ICO funds have been lost because of fraud.

Kapengut and Mizrach (2023) examine the impact of the Ethereum network’s transition to Proof-of-Stake. It has been found that the alteration has lowered energy consumption by 99.98% (Ethereum, 2023), and overall income of block rewards (in USD) has seen a reduction of 97%, although transaction fees (in ETH) have experienced a nearly 10% increase. The study also notes that miners have not transformed into validators. It is concluded that the transition to Proof of Stake has significantly impacted the network and competing platforms.

While Ethereum is widely known for its high transaction fees, one of the main reasons transaction fees have increased on the Ethereum network is the increasing demand for decentralized applications and Ethereum's limited capacity to handle large volumes of transactions. Due to the high demand for block space, transaction fees increase as more users and applications compete for block space (Donmez & Karaivanov, 2021). Ethereum's current scaling solutions have yet to meet the growing demand, leading to increased transaction fees. As a result of these challenges, some users have explored alternative smart contract platforms, such as Avalanche and Polygon, which offer faster transaction times and lower fees.

What is clear at this point is the decrease in energy consumption due to the switch from PoW to PoS, known as "The Merge" at Ethereum (De Vries, 2022), (Rieger et al., 2022). The Merge led to a decrease in energy consumption of over 99.8%. However, the switch did not lead to any changes regarding the ether price. Rieger et al. (2022) suggest that transaction fees for blockchains using PoS are lower than the ones using PoW. The limitations of their conducted studies lie in their small scope, which does not include Ethereum.

2.2.4 Degree of Decentralization

Identifying the degree of decentralization of blockchains can be challenging, as it involves assessing several factors that contribute to the overall level of decentralization. A network with more nodes, distributed throughout more than one location and operated by several parties, is generally considered more decentralized. On the other hand, if a small group of individuals or organizations hold a significant portion of tokens, this may indicate a low level of decentralization. Further, multiple data sources, such as numerous exchanges or APIs, may show greater decentralization. According to Zhang et al. (2022), decentralization has yet to be widely defined or measured.

Lin et al. (2021) state that oracles are crucial in blockchain decentralization. Moreover, the authors propose a method to measure decentralization in blockchain-based systems, specifically computing mining power distribution with metrics such as the Gini coefficient. Their results indicate that Bitcoin exhibits greater decentralization compared to Ethereum, whereas Ethereum is more stable.

Among two leading cryptocurrencies, Bitcoin and Ethereum, Gencer et al. (2018) present a study examining various decentralization metrics. The study uses the number of nodes, the distribution of nodes across the globe, and the concentration of mining power to assess the extent of decentralization. While both Bitcoin and Ethereum are decentralized, their degrees of decentralization differ. In terms of the number of nodes, Bitcoin has a more decentralized network, while Ethereum's network has a more geographically diverse distribution. Moreover, the study reveals that both networks are vulnerable to centralization because of the concentrated mining power in a few large pools. To put these results more into perspective, one must remember that during this time both blockchains mentioned above still used PoW as their consensus mechanism.

Finally, an article published by Kuśmierz and Overko in 2022 analyzes the wealth distribution of the wealthiest 50 to 100 nodes in different cryptocurrencies, focusing on coins and ERC20 tokens. It examines statistical metrics such as the approximated Zipf's law coefficient, Nakamoto coefficient, Shannon entropy, and for this study, especially relevant, the Gini coefficient to understand the distribution patterns. The study reveals that coins and ERC20 tokens have quantitatively different wealth distributions, with tokens being more centralized than coins. The Gini coefficient is higher for tokens than for coins. By analyzing transactions, they indicated a Gini coefficient of 0.8 for Chainlink between 2018 and 2022. It must be mentioned that they focused on wealth distribution, whereas this study focuses on income distribution.

3. Theory

The theory chapter outlines the theoretical framework of Chainlink and its main competitors. Its primary purpose is to extend its provided oracle network's architecture and highlight Chainlink as the leading oracle provider. This chapter is divided into two main subchapters. Whereas the first focuses on the oracle providers' general landscape, the second deepens into the subject of this study, Chainlink.

3.1 Landscape of Oracles

A first overview of the oracle provider landscape is given in this first subchapter. To highlight the different solutions to the oracle problem, the most important competitors in this field are introduced and compared. The introduction to Chainlink can be found in Chapter 3.2.

3.1.1 Band Protocol

Band Protocol was founded in 2017 and is the second most popular oracle after Chainlink. It uses its blockchain, Bandchain, using Cosmos technology which allows it to decrease costs and increase speed.

Validators are responsible for proposing and committing new blocks to the blockchain. The top 100 validators are selected by the number of BAND tokens they are staking in the consensus protocol. To ensure that the oracle service works appropriately, validators can see their stake slashed in the event of bad behavior. If they do not participate for excessive time to block proposals and commits, if they are unresponsive to data requests, or if they double sign, their stake will be slashed (Srinawakoon et al., 2019). The key advantages of Band Protocol are its flexibility and interoperability. It can be used on all blockchain platforms and provides a wide range of data such as prices, temperatures, sports scores, and many more.

In June 2023, Band Protocol used 73 validators, had an average request time of approximately six seconds, and got 31,000 requests per day on average (*Band Protocol - Cross-Chain Data Oracle,* 2023).

3.1.2 Witnet

Witnet was founded in 2017 and used its unique consensus mechanism, Proof of Eligibility (PoE). Nodes gain reputation points every time they provide correct data. These points are crucial since it is more likely to be selected for the subsequent request if the node has a good reputation. The Witnet network runs its blockchain and relies on the WIT token to incentivize autonomous nodes that resolve data requests. Witnesses are responsible for both validating transactions within the network and aggregating them into blocks, which are then periodically added to the blockchain. In 2023, Witnet is estimated to have over 10,000 nodes operating on its blockchain.

In the event of a dispute, the data requestor has the option to challenge the outcome by submitting a dispute transaction to the Witnet network. To ensure commitment, this transaction includes a monetary bond that will be forfeited if the dispute fails. At that stage, witnesses within the network can vote on the correct outcome. If a substantial majority agrees on the same result, the bond is returned to the party initially requesting the data (Sánchez De Pedro et al., 2017).

In May 2023, Witnet announced that they would switch from PoE to PoS for scalability reasons, becoming the first 100% PoS-based oracle. Switching their consensus mechanism to PoS will allow Witnet's oracle to operate faster (from 45 to 12 seconds, corresponding to -73.33%, to produce one block) (Rokowski, 2023).

3.1.3 DIA

DIA is a multi-chain oracle platform founded in 2020. It is currently available on more than 25 platforms.

DIA fetches daily a volume of 15 billion trades and proposes access to a comprehensive data library, including the price of more than 3,000 asset price feeds. (*Data Sources | DIA App*, 2023)

When users need data unavailable in the library, they can publish a bounty ticket (in DIA tokens) which will be paid to the data provider. The correctness of the data is first verified through a staking mechanism where anyone can challenge the data. The result is then stored in a public and immutable database and accessed on-chain and off-chain. (*Oracle Documentation - DIA Documentation*, 2023)

Unlike Chainlink or Band Protocol, DIA does not rely on a third party to source data but takes the crowdsourcing approach. Its end-to-end transparency is a crucial advantage.

3.1.4 Universal Market Access (UMA)

UMA's optimistic oracle was launched in 2018. UMA's optimistic oracle is a decentralized truth machine with human reasoning designed to connect real-world, subjective data to smart contracts. It allows smart contracts to securely request and receive any type of arbitrary data instead of just price data. The optimistic oracle functions as a decentralized truth machine flexible enough to handle ambiguity and expands what can be built in Web3. When data is requested, the proposer posts a bond as well as the proposed data. This is followed by a challenge period where anyone can dispute the data. If a proposal is disputed, the optimistic oracle funnels the data through a data verification system that relies upon UMA token holders to settle disputes. If the proposed data is determined to be wrong, the proposer loses its bond, and the disputer earns it (Universal Market Access, 2018).

UMA's oracle is called optimistic because disputes are designed to be unlikely. The true-unless-disputed verification process makes this oracle solution fast and cheap.

3.1.5 API3

Unlike the other oracles mentioned above, API3 tackles a different issue than the oracle problem. Launched in 2020, the API3 solution aims to target API connectivity. It promotes the development of "decentralized APIs" (dAPIs), where third-party service providers operate their oracle. By eliminating traditional oracles and allowing API providers to fetch data directly into the blockchain, API3 makes it possible to create data feeds cost-effectively and securely. This is made possible by Airnodes, which can be deployed by API providers in less than three minutes, without any knowledge of blockchain and with low maintenance. This service is priced on demand, meaning dAPIs start paying only after generating revenue (Benligiray et al., 2020). With no capital costs, this represents an attractive and scalable oracle solution.

3.1.6 Tellor

Tellor was founded in 2018 and used a hybrid consensus protocol to deliver real-time data to users at affordable rates. When a user submits a data request, they include a Tellor's token (TRB) reward as an incentive for the miners. If other users are interested in the same data, they can add a tip to the reward to make the query more attractive. Every ten minutes, the best-funded request is selected for miners to solve. When miners submit five values through PoW solutions, the median value is selected and saved on-chain. Anyone holding a TRB token can dispute the validity of a mined value by paying a dispute fee. Token holders vote and decide if the data is valid. If not, the miner loses its stake. If the data is correct, the miner earns the fee (tellor.io, 2023). Tellor's oracle is highly decentralized and guarantees high security and is thus meant for dApps valuing censorship resistance and decentralization.

3.1.7 NEST Protocol

NEST Protocol's oracle solution was introduced in 2018. It is a decentralized oracle system of price data in the chain that uses a "two-way quotation mechanism" to ensure that market costs are accurate. The core of the NEST oracle is the quote mining mechanism, a new type of mining that generates quotes and provides data for the Nest oracle. The quote miner is required to set up a NEST oracle quote contract and proceed to transfer 10 ETH and 2,000 USD assets to that particular contract. The NEST protocol operates a quotation mining process that makes it resistant to attacks. The data is validated through arbitrage, and the miner submitting the price quote receives NEST tokens as a reward (nestprotocol.org, 2022). This process allows the oracle to provide accurate prices since they are built directly into the blockchain. Nest Protocol's oracle is meant for DApps valuing decentralization and accuracy.

3.2 Chainlink

This subchapter gives an overview of Chainlink, the biggest oracle provider in the world. It accounts for around 80% of the oracle market capitalization (CoinMarketCap, 2023). Understanding the architecture of Chainlink is inevitable for understanding this thesis. Since the Basic Request Model is not applied anymore, it will not be discussed in detail. For further information on this topic, please refer to the Chainlink website.

3.2.1 The Purpose of Chainlink

According to Wang et al. (2019), smart contracts are characterized by having their source code published and verified on the blockchain, ensuring their immutability. Without centralized control and external authorities, the execution is enforced among anonymous, trustless individual nodes, making involving intermediaries unnecessary. Popovic et al. (2020) indicate that although self-executing code is nothing new, smart contracts enable business logic automation, minimize operational frictions and expenses, and increase business process efficiency. These characteristics have led to an increased interest in smart contracts throughout various industries, such as DeFi, insurance, and the Internet of Things. A connection between legacy systems and the on-chain world is needed to make use cases possible for these industries (Breidenbach et al., 2021)

As mentioned in the introduction, the purpose of an oracle is to bridge the on-chain and the off-chain world. In this case, the on-chain world is the blockchain, which cannot access data outside the network. For access to data from outside sources, oracles are required. By developing a network of decentralized oracle networks, whereas every single network implements a combination of different security techniques, Chainlink provides a solution to the oracle problem. The answer is subject to the same security and reliability standards as the blockchain (Breidenbach et al., 2021). Chainlink is based on the Ethereum blockchain but also works with other blockchains.

3.2.2 Chainlinks Architecture; Decentralized Oracle Networks

Previously, Chainlink made use of three different data feed models. The first one is called the Basic Request Model (BRM), in which one node deals directly with the request of smart contract consumers on Chainlink. A Decentralized Data Model (DDM) replaced this system by aggregating data from multiple oracles on-chain. This model has been further enhanced by the Decentralized Oracle Network (DON) and later Off-Chain Reporting (OCR). Nowadays, data aggregation is done off-chain through a Peer to Peer (P2P) network, allowing the oracles to communicate with each other. The following sections take a closer look at DONs, DDM, and OCR. Considering that the BRM has been abandoned, this section does not elaborate on it further.

The foundation of Chainlink 2.0 is a Decentralized Oracle Network (DON). A DON is maintained and controlled by a network of Chainlink nodes. It supports an infinite range of functions the committee chooses, which can be considered like the traditional consensus mechanism. A DON supports the main chain as a blockchain abstraction layer and provides highly efficient, decentralized, off-chain computing and off-chain resources for smart contracts. To summarize, DONs perform three basic functions: Storage, networking, and computation. (*Decentralized Data Model | Chainlink Documentation*, 2022).

Executables and adapters are the two key functionalities that enable a DON to realize its capabilities. Adapters are a generalization of the external adapters in Chainlink. Their primary function is to fetch data from sources. The executable fetches the data with the help of an adapter. Then the data is sent to the blockchain through an adapter. Executables are programs that run in a decentralized manner autonomously and continuously on a DON to perform deterministic operations. They do not store main chain assets. However, they have benefits such as high performance and the ability to conduct confidential computations. Adapters, as we envision them for DONs, are a generalization of the external adapters in Chainlink today. An Adapter is initiated by a code on the DON, and the executable can read and write to local storage or communicate with other executables. (Breidenbach et al., 2021)

3.2.3 Chainlinks Architecture; Data Feeds

Nowadays, most Chainlink users rely on smart contracts and data feeds. A data feed is a report containing the needed information or data based on off-chain sources. Data feeds are reports containing information or data obtained from off-chain sources. Different data types can be requested, such as index prices, weather, and election data. Using a DON system, the provided data feeds by Chainlink are enhanced in fields like scalability, confidentiality, agility, and security. The data feed includes three components: a consumer contract, a proxy contract, and an aggregator contract. Any contract that uses Chainlink data feeds to consume aggregated data is defined as a consumer contract. (Breidenbach et al., 2021)

On Chainlinks website about their data feeds, the process is described as follows:

To ensure proper execution, it is necessary to utilize the accurate AggregatorV3Interface and invoke one of its accessible functions. Proxy contracts function as on-chain intermediaries, directing to the aggregator responsible for a particular data feed. This arrangement allows for seamless upgrades of the underlying aggregator without disrupting the services provided to consumer contracts. An aggregator operates as a contract that receives periodic data updates from the oracle network and subsequently stores the aggregated data on-chain for DDM. This enables consumers to access and utilize the data within the same transaction

They must reference the correct `AggregatorV3Interface` and call one of the exposed functions. Proxy contracts are on-chain proxies that point to the aggregator for a specific data feed. This enables upgrading the underlying aggregator without interrupting service to consumer contracts. An aggregator is a contract that receives periodic data updates from the oracle network, the store aggregated data on-chain for DDM so that consumers can retrieve it and act upon it within the same transaction. As mentioned, this system has been moved off-chain (OCR). Finally, transmitters (basically oracle nodes) help optimize data transmission from the oracle network to the aggregator contract. All external transactions go through transmitters (Etherscan). (*Decentralized Data Model | Chainlink Documentation*, 2022).

In the first step, the end consumer uses an app based on a smart contract. The smart contract requests data from the Chainlink Functions Smart Contract, which then requests the required data from the DON. The DON works off-chain and rewards each publisher of data for their contribution. To enable the data feed aggregator contract to update the data, a minimum number of oracles is needed. Each oracle, part of the DON, is triggered to fetch data, which will then be checked by reaching a consensus on the final answer using Off-Chain Reporting 2.0 (OCR), and aggregated. (*Connect the World's APIs to Web3 With Chainlink Functions*, 2022)

3.2.4 Chainlinks Architecture; Off-Chain Reporting

One year following its development and thorough security audits, Chainlink introduced Off-Chain Reporting (OCR) to its mainnet. OCR's primary objective is to optimize data computation efficiency across Chainlink Oracles while significantly cutting down operational costs by up to 90%. This advancement leads to an expansion in the volume of real-world data accessible to smart contract applications, bringing substantial benefits to various sectors like DeFi, decentralized insurance, blockchain-based gaming, and more. With OCR, smart contract developers now have broader access to external data, covering a diverse range of assets, real-world events, and blockchains. This newfound accessibility empowers them to cater to new industries and explore unique use cases, unlocking new possibilities for innovation and growth. (*Off-Chain Reporting | Chainlink Documentation*, 2022)

OCR consists of all nodes communicating through a peer-to-peer (P2P) network. Breidenbach et al. (2021) explain the process as follows: A lightweight consensus algorithm runs during this process, requiring all nodes to report and sign their data. This method's advantage lies in reducing gas consumption by aggregating all the data into one transaction, which is then transmitted. The leader node consistently asks other nodes to supply new signed observations, which are then combined into a single report. After this step, the report is sent back to the nodes for verification. If the report passes the verification process, a signed copy is returned to the leader. The leader then brings together all the necessary data, including the quorum signatures, and shares it with all participants through broadcasting. These "leader" nodes attempt to transmit the final report to the aggregator contract based on a randomized schedule, as described in the OCR Protocol Paper on the Chainlink homepage. The aggregator ensures that a quorum of nodes has signed the reports and shares the median value with consumers along with a block timespan and an ID. To ensure a robust system without any single point of failure, all nodes endorse the final report. In case the designated node fails to get its transmission approved within a fixed period, a round-robin protocol activates, allowing other nodes to transmit their final reports. Ultimately, the main objective is to achieve a confirmed final report. Some of the benefits that accompany this process are the reduction of the overall network congestion from Chainlink oracle networks. (*Off-Chain Reporting | Chainlink Documentation*, 2022)

Additionally, gas costs for a single-node operator can be reduced. Moreover, because data feeds accommodate more nodes, node networks become more scalable. Finally, data feeds are more accurate since they can be updated timelier. (*Off-Chain Reporting | Chainlink Documentation*, 2022)

As mentioned by the predecessor of this thesis, Hof (2023), it is important to distinguish between the different addresses involved in this data transmission process. There are node addresses, which is the address for the Chainlink node wallet and is an externally owned address (EOA). The node, or more precisely, its wallet, must always be filled with native blockchain tokens to answer/process requests. The second type of address is the Oracle or Operator Contract Address, the address for the oracle/operator contract, which has been deployed on the blockchain. That is the address for contracts deployed on the blockchain like `Operator.sol` or `Oracle.sol`. It ought not to be funded with native blockchain tokens like ETH. When making API call requests, this is the contract the funds are directed through to operate with the Chainlink node.

Additionally, this is the address that Smart Contract developers point at when selecting a node for an API call. Finally, there is the Admin Wallet Address, the address to own the oracle/operator addresses (`Operator.sol` and `Oracle.sol`). This is important since that is the wallet address that receives the LINK rewards for providing the data (within the OCR model). The basic request model sends the so-called LINK rewards to the oracle address and not to the node or admin wallet address. (*Fulfilling Requests | Chainlink Documentation*, 2023)

When the BRM was still in use, an oracle operator needed to deploy oracle contracts. However, nowadays, operator contracts are used. Some crucial aspects involve the operator factory, forwarder, authorized receiver, authorized sender, and ownership.

An “operator factory” is a contract Chainlink provides to create new operators and forwarders for node operators. This makes it easier for clients to verify the authenticity of the deployed contract. Nevertheless, node operators typically use their own deployers to fit their needs.

Chainlink has introduced two new functionalities: multiple EOAs per node and forwarder contracts. The latter enables operators to oversee multiple EOAs and present them as a single address. Forwarder contracts act like a reverse proxy server, where the same address serves the user without knowing which server the traffic is coming from. At the same time, this enhances security by distinguishing between owner and authorized sender accounts. To do this, the node operator must add the forwarder contracts to a whitelist. The operator can create and manage these forwarder contracts from the operator factory through an operator contract linked to a secure address with either Multisig or private keys stored in a cold wallet (offline). As a result, the operator can utilize multiple EOAs and forwarders per node, making it possible to handle various job types on the same node. Using forwarders, infrastructure and maintenance costs can be reduced. (*Forwarder | Chainlink Documentation*, 2022)

`AuthorizedReceiver` is an abstract contract inherited by both operator and forwarder contracts. It enables the owners of forwarder contracts to designate authorized senders who can invoke the forward method. Meanwhile, owners of operator contracts can specify `AuthorizedSenders` who can call the `fulfillOracleRequest` and `fulfillOracleRequest2` methods. This ensures the accuracy of the received data and lets the operator and forwarder contracts utilize it (*Receiver | Chainlink Documentation*, 2022). An `AuthorziedSender` is a hot wallet that can trigger transactions on behalf of operator contracts. The contract owner defines and manages the senders. It can be changed to enhance security or in case a node has been compromised.

The owner does not have to be a Chainlink node. However, it is undoubtedly a Multisig or a cold wallet address. The owner address is inherited by the operator and forwarder contracts and through functions in those contracts. The owner or node operator can transfer and accept ownership to new owner addresses. (*Ownership | Chainlink Documentation*, 2022)

Typically, the operator contract should receive the LINK rewards, which can then be withdrawn to a designated address (the owner of the operator contract or an admin wallet address, usually not a Chainlink node). However, each node operator can establish distinct transaction-sending strategies, enabling them to employ multiple operators and exercise some discretion through the operator forwarder construct when collecting LINK rewards.

4. Methodology

The following subchapters focus on the research design, the data collection, and data analysis to give the reader an overview of the applied methodology. The chapter is concluded with limitations and assumptions.

4.1 Research Design

The research questions outlined in this paper require a quantitative approach. Thus, gathering data on Chainlink transactions was needed. Firstly, a big part of the transaction data could be downloaded from XBLOCK, while the remaining transactions had to be scraped from Etherscan, Ethereum's native blockchain explorer. The data collection part is elaborated further upon in the subsequent chapter.

A Python script was used for the data scraping through the Etherscan API. The script was run on multiple devices concurrently to speed up the data-gathering process. To aggregate the data chunks, a second script was implemented that ran in the background and combined the various data instances into a comprehensive database.

To facilitate the analysis of the dataset, statistical analysis was applied through the use of Python's built-in libraries.

4.2 Data Collection

Transaction data showing interactions with oracle contracts was needed to conduct a comparative study of the Chainlink oracle network. This was primarily obtained from XBLOCK, which conveniently archives Ethereum's on-chain data from the genesis block up to and including block 16,500,000. For the remaining portion of the analysis, which included around 500,000 blocks, the transaction data was manually extracted using a self-developed web scraper.

Three types of transactions were downloaded or scraped to access all the data we needed. This includes internal transactions, block transactions, and ERC20 Transactions.

Internal Transactions: They occur within a smart contract, representing interactions and operations within a specific smart contract. The execution of smart contract code triggers these transactions and can be found on Etherscan under internal transactions. (Amir, 2021)

Block Transactions: These are simple transfers of cryptocurrencies or value from one EOA (or wallet address) to another EOA within a blockchain network. They occur at the blockchain level and are recorded in blocks, forming the blockchain's transaction history. Accessing these transactions is possible for anyone. This is an integral part of the blockchain transaction ledger. (Amir, 2021)

ERC20 Transactions: These transactions are specific to the Ethereum blockchain and are used to transfer and interact with tokens following the ERC20 token standard. LINK follows this standard as well. Furthermore, these tokens represent fungible assets, meaning that each token is identical to another token of the same type and value. Anyone with access to the blockchain's transaction history can observe and verify them. (Amir, 2021)

The transactional data was extracted from the Ethereum blockchain using its native blockchain explorer, Etherscan. To do this, Python, a widely used programming language in data analytics and machine learning, was used to query the data and transform it into a format appropriate for analysis.

Multiple Python scripts were executed concurrently across several threads and machines to speed up data collection and preprocessing. As a safety measure, after processing a certain number of blocks, the extracted data was automatically stored in a pickle database—a built-in Python library that enables the serialization of Python objects. Simultaneously, another script retrieved the saved pickle instances from various computers and integrated them into a comprehensive database.

To address potential script failures due to internet connectivity issues, fallback functions designed to restart the scraping process several times before ultimately terminating the script were integrated. Moreover, a folder structure containing .txt files to store the most recent block processed for each address was used, ensuring no duplicates were generated when restarting a scraper for a specific address. This methodology provided considerable flexibility while maintaining the highest level of accuracy in the data collected and high resilience.

Once the data was collected, a series of preprocessing and data-cleaning steps to ensure the data was consistent and adjusted to errors was conducted. This included removing potential duplicates, imputing missing values, and standardizing variable names and formats.

Ultimately, the amount of data that could be downloaded from XBLOCK was considerably larger than the manually scraped transactions. Even though the same 84 Oracle Addresses were filtered for, the scraped dataset seemed insignificant but was still included in the analysis.

4.3 Data Analysis

The following chapter presents how the gathered data was analyzed.

4.3.1 Drivers of Digital Currency Prices

Two different time series were constructed to compute the correlation between the number of oracle transactions and the price of the underlying cryptocurrency. The first time series represents the progression of oracle transactions, while the second series captures the price evolution of the LINK utility token.

Before analyzing the correlation between the time series, it is essential to conduct a test for stationarity. A time series is deemed stationary if its mean, variance, and autocorrelation remain constant. Should the time series prove to be stationary, the Pearson correlation coefficient can be used to measure the correlation using the following formula:

$$Correlation = \frac{Cov(Number\ of\ Oracle\ Transactions, Token\ Price)}{\sigma_{Number\ of\ Oracle\ Transaction} * \sigma_{Token\ Price}} \quad \text{(X)}$$

Conversely, if the time series is non-stationary, the Johansen test to assess cointegration would be used. Cointegration suggests that non-stationary time series can be derived from a linear combination of stationary time series. Cointegration between two or more time series indicates a long-term equilibrium among them. To perform the Johansen test, it is necessary to verify non-stationarity and estimate a Vector Error Correction Model (VECM) (Dwyer, 2015). The VECM is then employed to analyze the cointegration relationship. The Johansen test can be executed using two distinct approaches, with the first emphasizing the trace and the second concentrating on the eigenvalues (Dwyer, 2015).

4.3.2 Token Economics

Part of the scope of the analysis was to examine the impact of the switch in the consensus mechanism in Ethereum on the transaction fees. The so-called “merge” that made Ethereum switch from the more energy-intensive PoW consensus mechanism to PoS occurred at block 15,537,393 on September 15th, 2022. (Nambiampurath, 2022)

To test whether the switch in consensus mechanism lowered transaction fees on the Ethereum blockchain, two time series were constructed, the first ranging until block 15,537,393 and the second from block 15,537,394 onwards. Following this, all the blocks daily were aggregated and computed the daily mean of the minimum transaction fees for all the blocks within that day. The minimum transaction fee was selected as it represents a lower bound for the validator's minimum requirement to include a transaction into a block. If a statistical difference is found between the two time series, it can be concluded that this minimal requirement changed and, thus, that the transaction fee structure changed after the switch to PoS. To determine the magnitude and direction of this effect, regression analysis such as ordinary least squares (OLS) with and without interaction terms and Lasso regression was used. By testing this hypothesis, insights into the economic implications of the Proof-of-Stake consensus mechanism for blockchain ecosystems and potential trade-offs and challenges in its adaptation can be showcased.

4.3.3 Degree of Decentralization

The computation was done using the Gini coefficient as outlined below:

First, the share of each oracle provider in the network was determined. The share is based on the number of oracles requests the node handles. The oracle providers were then ranked based on their share, from the largest to the smallest. Eventually, using the Lorenz curve and the following formula allows us to compute the Gini coefficient:

$$Gini - Coefficient = \frac{\sum_i^n \sum_j^n |Oracle\ Provider_i - Oracle\ Provider_j|}{2 * n^2 * \overline{Oracle\ Provider}}$$

Besides the Gini coefficient, the Herfindahl-Hirschman index (HHI) was calculated, which is particularly useful in analyzing market concentration by imposing a more significant penalty on larger market shares than the Gini coefficient. The Herfindahl index can be calculated using the following formula:

$$Herfindahl - Index\ = \sum_i^n (Market\ Share\ of\ Oracle\ Provider_i)^2$$

The computation of both the Gini coefficient and Herfindahl index allows for a comparative analysis of market dominance and thus provides more insight than just using one of the two metrics to assess market dominance.

4.4 Limitations and Assumptions

Three main difficulties impacted the scope of the research. The first was the option to properly filter on the block explorer to identify oracle contract addresses. Etherscan does not provide a specific search feature for locating the on-chain components for the oracle service. Consequently, identifying and extracting the relevant data for the analysis became increasingly time-consuming and complex. This means the study is currently limited to analyzing Chainlink because this is the only oracle with (previously public) access to a certain number of nodes. A list of node operators and access to the underlying blockchains data would be needed to extend the scope to other blockchains and oracles.

The second problem arose from the restrictions on how much data can be accessed per second through Etherscan API. This made the process of retrieving the relevant transaction data time-consuming. Even though only half a million blocks needed to be scraped for 84 node addresses, with 14 scripts simultaneously running, it took around one and a half months to gather the dataset.

The third problem was a practical one. Since the timeframe for the data analysis was four years, corresponding to 9,500,000 million blocks, the ERC20, internal transactions, and block transactions needed to be downloaded from XBLOCK. The resulting unfiltered database was around five to six terabytes. Unfortunately, hardware that could hold this amount of data at once was not accessible for this study. That meant the data needed to be downloaded on different devices, then filtered by the corresponding node operator addresses, and subsequently aggregated into complete datasets. Finally, working with the data, including filtering it for the relevant node addresses, requires a computer with enough capacity. Working with external hardware might help. The filtered data is still over half a terabyte.

Due to our limited capacity, no other oracles or blockchains were analyzed. Doing this would have required even more hardware capacity and potentially direct access to a blockchain node to query the data.

5. Results and Findings

In the following subchapters, the results and findings are thoroughly showcased and discussed.

5.1.1 Drivers of Digital Currency Prices

Examining the drivers of the LINK token price enabled comprehensively investigating the potential factors that might influence the price of a digital currency. The hypotheses were defined as follows:

H0: The number of oracle transactions processed on the blockchain does not correlate with the LINK token price.

H1: The number of oracle transactions processed on the blockchain does correlate with the LINK token price.

First, stationarity or non-stationarity of the implicated time series stated in the hypotheses needed to be declared, the daily count of oracle transactions processed, and the daily price of the LINK token, respectively. To assess stationarity, Augmented Dickey-Fuller (ADF) test was applied to identify a unit root's presence since a stationary time series' mean, variance, and autocorrelation remain constant over time.

Figure 1 ADF Test (Daily Transaction Count)

Daily Transaction Count	Value
ADF Statistic	-3.4768362519259166
p-value	0.008602301982351367
Critical Values	{'1%': -3.4350161653396736, '5%': -2.863600780613854, '10%': -2.567867151504452}

Source: own illustration

The ADF for the daily transaction count is -3.4768. Therefore, the time series is stationary, according to the negative output. Substantial evidence is shown against non-stationarity due to the p-value of 0.0086 being below the usually accepted significance level of 0.05.

Figure 2 ADF Test (LINK Token Price)

```
LINK Token Price     Value
------------------   ------------------------------------------------------------------------------
ADF Statistic        -47.30389777193787
p-value              0.0
Critical Values      {'1%': -3.4336370214482437, '5%': -2.862992025899885, '10%': -2.5675429970585153}
```

Source: own illustration

The ADF Statistic for the daily change in the LINK token price is -47.30389777193787. This indicates substantial evidence against the non-stationarity of the time series. As the p-value is the likelihood of seeing the ADF Statistic under the null hypothesis (non-stationarity) and is 0.0, it further supports the stationary behavior of the daily change in the LINK token price. As a result of the stationarity of the two time series, the Pearson correlation coefficient can be used to measure the strength of the relationship.

Figure 3 Pearson Correlation Coefficient

```
[['Pearson correlation coefficient', -0.05780684080984266], ['', '']]
```

Source: own illustration

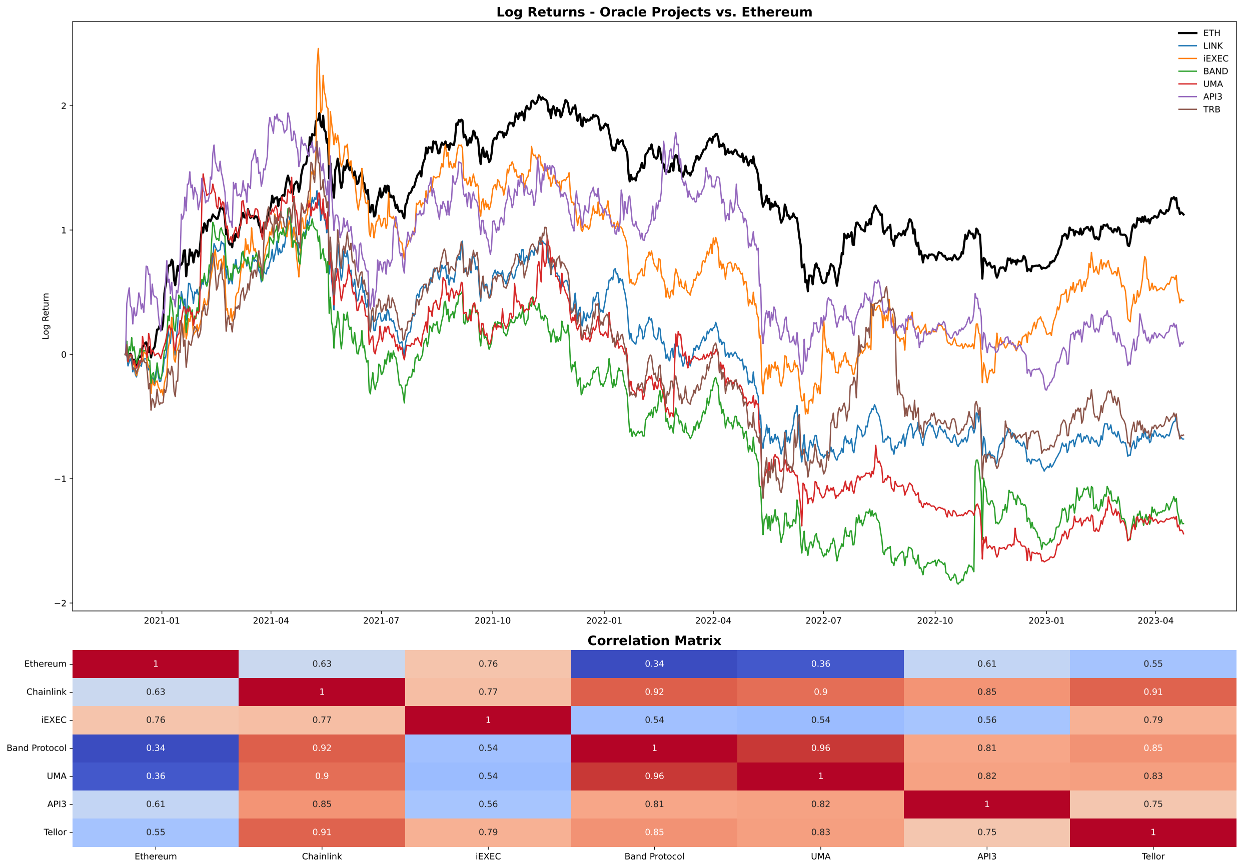
The Pearson correlation for the two time series study was -0.0578, implying a slightly negative connection between the number of oracle transactions on the blockchain and the LINK token price. This means that there is no clear visible linear relationship between the two variables. As LINK is a utility token, its value should be linked to the success of the underlying platform. Chainlink’s profitability is mainly driven by the number of oracle transactions performed on the platform. Hence it is dependent on the fees collected by those transactions. As a result, a positive correlation between the two time series would have been predicted. The following visualization illustrates the correlation between the oracle transactions and the LINK token. The correlation was computed in terms of changes in daily LINK closing prices and daily transaction counts.

Figure 4 Correlation between Daily Transaction Count and LINK Token Returns

Source: own illustration

According to the plot above, the number of transactions and the token price suggests an absence of a direct relationship. That does not rule out the potential of additional complicated variables impacting the relationship. Furthermore, there is significant variation in the LINK token price. Since digital currencies are often exposed to high volatility, identifying an isolated individual driver is challenging.

Figure 5 Log Returns - Oracle Projects vs. Ethereum and Correlation Matrix



Source: own illustration

Based on the graphic above, the correlation between the LINK token price and Ethereum (ETH) is 0.63, implying a moderately positive relationship between the two variables. This implies a connection between the increase in total transactions processed on Ethereum and the LINK token price. However, no clear association can be observed when examining oracle transactions in isolation.

Subsequently, to further elaborate on the drivers of the LINK token price, it would be beneficial to analyze additional data and consider other factors. This could involve examining the adoption rate of Chainlink's oracle services and the value and volume of data being processed through their oracles, and any announcements or developments related to Chainlink's ecosystem. As a result, additional study and inquiry are required to investigate additional variables that may influence the price of the LINK token. Some reflections are further exemplified in the discussion chapter.

5.1.2 Token Economics

Since the dataset provided means to investigate the impact of the Ethereum merge, where the consensus mechanism was switched from PoW to PoS, we assessed the effect of the merge on the transaction fees as represented by the minimum gas price per block. Besides limited scalability, high transaction fees are still considered one of the main obstacles to the widespread adoption of blockchain-based systems (Prewett et al., 2020), as such transaction fees are an important factor for Chainlink going forward. The hypotheses were the following:

H0: There is no significant difference in the minimum transaction fees and transaction volume before and after the switch from PoW to PoS on Ethereum.

H1: There is a significant difference in the minimum transaction fees and transaction volume before and after the switch from PoW to PoS on Ethereum.

In the first step, the minimum gas price per block was aggregated into a daily average and compared their means before and after the merge. The findings are highlighted in the subsequent figure.

Figure 6 Minimum Gas Price over Time

Source: own illustration

The analysis indicates that the null hypothesis can be rejected with a t-value of 16.35 on the 5% confidence level, thus suggesting that the switch in consensus mechanism lowered the average minimum transaction fees within the analyzed time series.

However, this comparative analysis of the means of the transaction fees falls short of capturing interaction effects that the change in consensus mechanism might have had on other variables, such as transaction count, which in turn could also influence the transaction fees. Therefore, a two-step approach was adopted in the analysis to address this potential bias. First, the impact of the merge was evaluated using an OLS regression model without including control variables. Subsequently, these control variables were incorporated into the model to observe any changes in its predictive power. This comparative analysis provides a more comprehensive understanding of the factors influencing minimum gas prices.

Figure 7 OLS Regression I

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            minGasPrice   R-squared:                       0.020
Model:                            OLS   Adj. R-squared:                  0.020
Method:                 Least Squares   F-statistic:                     31.02
Date:                Mon, 29 May 2023   Prob (F-statistic):           3.03e-08
Time:                        13:01:21   Log-Likelihood:                 22975.
No. Observations:                1487   AIC:                        -4.595e+04
Df Residuals:                    1485   BIC:                        -4.593e+04
Df Model:                           1
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
const         4.01e-08   1.28e-09     31.237      0.000    3.76e-08    4.26e-08
after_merge -2.381e-08   4.28e-09     -5.569      0.000   -3.22e-08   -1.54e-08
==============================================================================
Omnibus:                     1377.714   Durbin-Watson:                   0.449
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           122064.301
Skew:                           3.979   Prob(JB):                         0.00
Kurtosis:                      46.667   Cond. No.                         3.52
==============================================================================
```

Source: own illustration

In this study, an OLS regression model was deployed to examine the relationship between the minimum gas price and a binary variable that indicates whether the Ethereum merge had already occurred or not. The model's explanatory power, as represented by the R-squared value, was 0.02, indicating that merely 2% of the observed variation in the minimum gas price could be ascribed to the merge event. The coefficient of the binary variable after_merge is negative, which implies that the minimum transaction fees decreased after the merge. The t-value of the after_merge predictor suggests statistical significance on the 5% level. This means the null hypothesis that the minimum transaction fee remained unchanged after the merge can be rejected. However, given the low R-Squared value, a substantial proportion of the minimum gas price remains unexplained by this model. Moreover, the Durbin-Watson statistic of 0.662 indicates (positive) autocorrelation in the residuals, potentially violating a fundamental assumption of OLS and possibly undermining the model's credibility (CFI Team, 2020). Essentially a Durbin-Watson value of 2 would indicate no presence of autocorrelation, while a value below 2 indicates positive and a value above two negative autocorrelation.

To enhance the model’s robustness and explanatory power, a subsequent modification of the OLS regression model incorporated control variables found in the given dataset. The variables deemed suitable and bearing explanatory power for the gas fees are the following:

Volume Ethereum: A daily metric capturing the total transaction volume in ETH. This volume is typically incurred off-chain on centralized cryptocurrency exchanges such as Coinbase or Binance, as they account for most transaction volume in the blockchain sphere.

Volatility Ethereum Monthly: Derived from the daily closing price of Ethereum, this variable captures the monthly volatility, serving as a proxy for risk sentiment among market participants.

Close Ethereum: This variable denotes the daily closing price of ETH in USD.

Erc20TxCnt: This variable, aggregating the number of ERC20 transactions for a given block range, captures the daily amount of ERC20 transactions. In contrast to the volume, these transactions affect the network's throughput as they are settled on-chain through a mining/validation process.

Size: This variable captures the size of the blocks added to the blockchain.

Difficulty: Refers to how challenging it is in terms of computational resources to solve the mathematical puzzle that allows a miner to add a new block to the blockchain and cash in the block reward.

gasLimit: Captures a value defining the maximum amount of computational work (gas) that a single transaction or block can consume and thus serves as an upper bound for the fee incurred.

To account for the entirely different scales of the input variables, MinMax scaling was applied to control for the condition number. This procedure scales each feature to a range between zero and one, which helps mitigate the impact of outliers and improves the stability of algorithms that are sensitive to different input variables scales. The formula for min max-scaling an individual parameter y from a dataset X is as follows:

$$X_{scaled} = \frac{(y - X_{min})}{(X_{max} - y)}$$

The following summary table represents the results of the OLS model, controlling for the variables mentioned above.

Figure 8 OLS Regression II

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            minGasPrice   R-squared:                       0.377
Model:                            OLS   Adj. R-squared:                  0.373
Method:                 Least Squares   F-statistic:                     111.7
Date:                Mon, 29 May 2023   Prob (F-statistic):          6.95e-146
Time:                        14:56:56   Log-Likelihood:                -38320.
No. Observations:                1487   AIC:                         7.666e+04
Df Residuals:                    1478   BIC:                         7.671e+04
Df Model:                           8
Covariance Type:            nonrobust
===============================================================================================================
                                              coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------------------------
const                                     1.08e+10   4.11e+09      2.629      0.009    2.74e+09    1.89e+10
after_merge                             -1.232e+11   1.12e+10    -10.983      0.000   -1.45e+11   -1.01e+11
Volume_Ethereum_MinMaxScaled             5.826e+10   1.18e+10      4.943      0.000    3.51e+10    8.14e+10
Close_Ethereum_MinMaxScaled              6.039e+10   8.06e+09      7.490      0.000    4.46e+10    7.62e+10
Volatility_Ethereum_Monthly_MinMaxScaled -3.031e+10  6.96e+09     -4.354      0.000    -4.4e+10   -1.67e+10
erc20TxCnt_MinMaxScaled                  4.223e+10   9.32e+09      4.533      0.000     2.4e+10    6.05e+10
size_MinMaxScaled                        7.879e+10   1.89e+10      4.165      0.000    4.17e+10    1.16e+11
difficulty_MinMaxScaled                 -1.198e+11   1.51e+10     -7.932      0.000   -1.49e+11   -9.01e+10
gasLimit_MinMaxScaled                    4.425e+10   1.02e+10      4.342      0.000    2.43e+10    6.42e+10
==============================================================================
Omnibus:                     1879.878   Durbin-Watson:                   0.690
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           644402.938
Skew:                           6.329   Prob(JB):                         0.00
Kurtosis:                     104.195   Cond. No.                         34.1
==============================================================================
```

Source: own illustration

The R-squared value is now at 0.377, which indicates that approximately 37.7% of the variation in the minimum gas price can be explained by the model. This still leaves a significant portion of the variation in the dependent variable, minimum gas price, unexplained, suggesting that other factors not included in the model and our dataset affect the minimum gas price.

All predictors are statistically significant on the 5% level, suggesting that each of them has some influence on the minimum gas price. The after_merge binary variable still has a negative coefficient, which indicates that after the Ethereum merge, the daily minimum gas price decreased.

Volume, closing price, erc20 transaction count, size, and gas limit all have positive coefficients, suggesting that increases in these variables are associated with an increase in the minimum gas price. Intuitively these findings make sense, as a higher transaction count results in a more congested network, and thus market participants require a higher gas price to get their transactions validated. The impact of the volume on the gas price is analogous in the sense that a higher transaction volume results in more ETH tokens that might be routed from cold wallets onto exchanges to facilitate transactions which in turn drives up gas fees for settlement. Furthermore, a high closing price might imply that market participants are willing to embrace a higher overall fee, while the share of the fee to the transactional value might be unchanged. The positive impact of size on the gas Price makes intuitively sense, as a larger block size means stronger competition among market participants to get their transaction included in the given block, which results in higher selectivity of the miner and, thus, a larger minimum gas Price. Last but not least, it makes sense that the higher the gas limit, the higher the minimum gas Price, as the higher the gas limit, the more gas the transaction can consume before being added to a block, thus incrementing minimum gas prices.

The monthly volatility and the difficulty have a negative impact, suggesting that the higher the volatility and difficulty, the lower the minimum gas price. Several factors could drive the relationship between volatility and the minimum gas price:

Gas Price Management: During high price fluctuations, market participants might be more cautious in managing their transaction costs, choosing to submit transactions with lower gas prices to minimize costs.

Risk Aversion: In times of increased volatility, users might be more risk averse and thus less willing to engage in transactions involving high gas fees, reducing throughput and thus fees.

Miner Behavior: Miners may accept lower gas prices during times of high price fluctuations as they might be more interested in quickly processing transactions to capitalize on the potentially higher block rewards due to price swings.

The negative impact of the difficulty on the minimum gas price is intuitively harder to explain, as this suggests that the higher the difficulty the miner faces, the lower the minimum gas price chosen by market participants. A possible explanation could be that an increase in difficulty means a stable and healthy network, with an influx of miners competing to mine new blocks, as the difficulty is calculated proportionally to the network's overall hash rate and adjusted for every 1,000 blocks in Ethereum. This increased competition could lead miners to be willing to accept lower transaction fees to fill their blocks and thus maximize the rewards (Knapsack Principle). Nevertheless, the computation of the Gini Coefficients among miners before and after the switch in the consensus mechanism shows a similar value of around 0.7, and as such, it remains unclear whether the influx of validators has led to a more competitive validation process.

The after_merge variable plays a significant role in predicting minimum gas price, as its coefficient is statistically significant. Its negative sign indicates that the merge had a decreasing effect on the minimum gas price. However, the model still has some drawbacks.

Potential Limitations:

1. The model explains only 34.4% of the variation in the minimum gas price, suggesting that there are other factors with an impact on the minimum gas price.
2. The model assumes a linear relationship between the predictors and outcome, which might not be accurate as some predictors might have a nonlinear impact on the minimum transaction fee.
3. The MinMax scaling could distort the relationship between variables because it compresses all data instances in a narrow range.
4. The model is based on historical data, and the period after the merge was not subject to extensive and consistent price increases but instead home to consistently lowered prices. To account for momentum, we computed the Relative Strength Index (RSI) and a Simple Moving Average (SMA). However, both variables appeared to be statistically insignificant and thus did not bear any explanatory power for the minimum transaction fee.
5. The residuals are not normally distributed, as indicated by the Jarque-Bera test, violating the OLS normality assumption. This may result from omitted variables or nonlinear relationships between predictors and dependent variables. As such, the model may not provide the best estimates.
6. The Durban-Watson statistic indicates potential autocorrelation in the residuals. This means consecutive errors are not independent, violating another assumption of OLS.

The primary insights from comparing the two OLS models suggest that including additional predictors in the second model has significantly improved the explanatory power, as measured by the r-squared value. The after_merge binary variable has retained its significance and negative relationship with the minimum gas price, even after considering other variables. This provides more substantial evidence for the impact of the Ethereum merge event on the minimum gas price.

Nevertheless, despite these improvements, a large portion of the variability in the minimum gas price still needs to be captured. This suggests that other factors not included in our dataset influence the minimum gas price.

As such, interaction terms were added in the next step, which increased the model's explanatory power to 57.2% but also introduced high multicollinearity, as indicated by the condition number.

Figure 9 OLS Explanatory Power

OLS Regression Results

Dep. Variable:	minGasPrice	R-squared:	0.572
Model:	OLS	Adj. R-squared:	0.559
Method:	Least Squares	F-statistic:	45.94
Date:	Wed, 21 Jun 2023	Prob (F-statistic):	1.04e-232
Time:	13:38:14	Log-Likelihood:	-38041.
No. Observations:	1487	AIC:	7.617e+04
Df Residuals:	1444	BIC:	7.640e+04
Df Model:	42		
Covariance Type:	nonrobust		

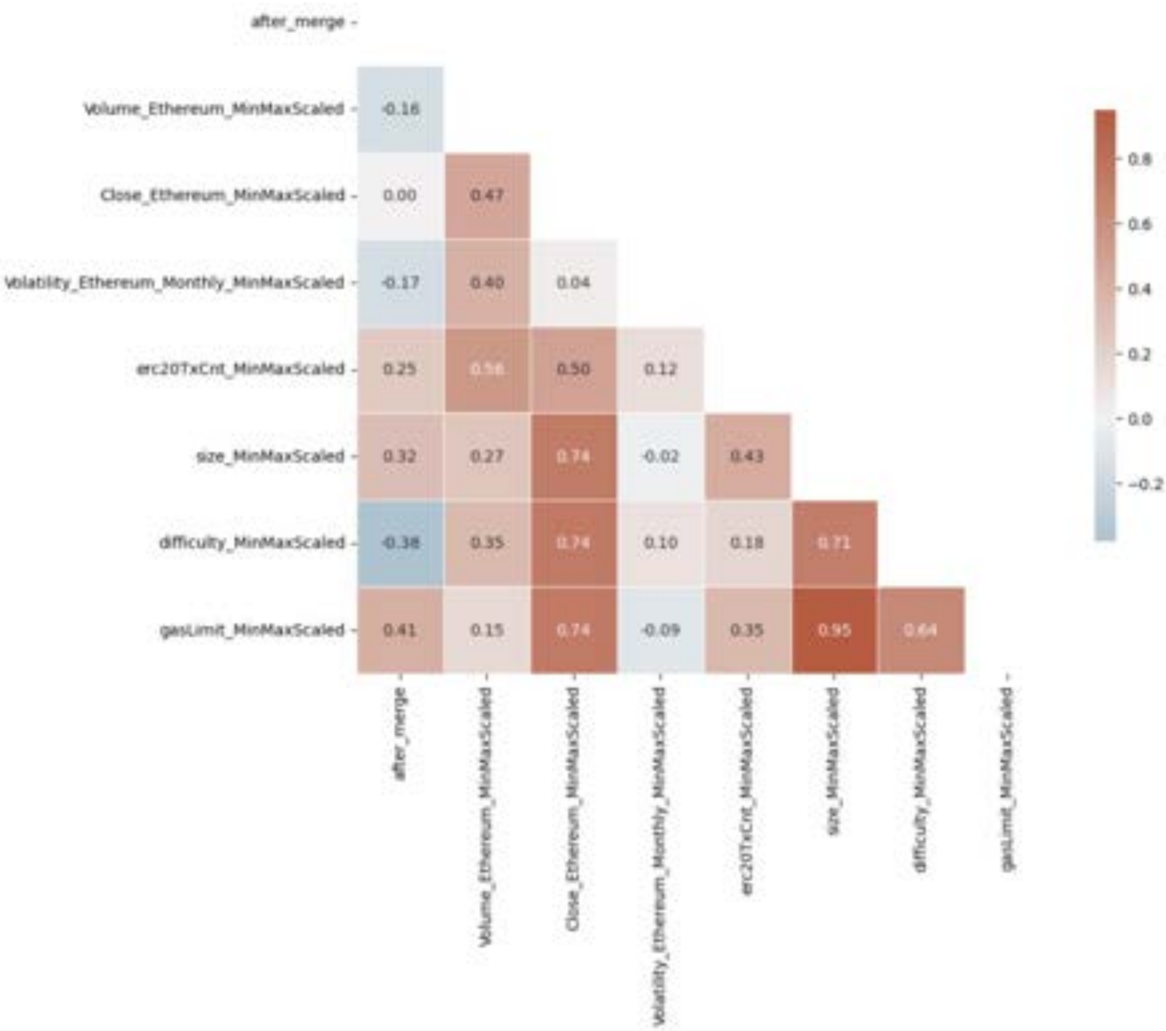
	coef	std err	t	P>|t|	[0.025	0.975]
const	4.858e+21	1.05e+23	0.046	0.963	-2.02e+23	2.11e+23
after_merge	-1.551e+14	3.53e+14	-0.440	0.660	-8.47e+14	5.37e+14
Volume_Ethereum_MinMaxScaled	1.343e+10	2.99e+10	0.449	0.653	-4.52e+10	7.21e+10
Close_Ethereum_MinMaxScaled	3.192e+11	4.43e+10	7.197	0.000	2.32e+11	4.06e+11
Volatility_Ethereum_Monthly_MinMaxScaled	-6.42e+10	1.75e+10	-3.663	0.000	-9.86e+10	-2.98e+10
erc20TxCnt_MinMaxScaled	2.392e+10	3.34e+10	0.717	0.474	-4.16e+10	8.94e+10
size_MinMaxScaled	6.394e+10	4.63e+10	1.382	0.167	-2.68e+10	1.55e+11
difficulty_MinMaxScaled	2.753e+11	5.45e+10	5.050	0.000	1.68e+11	3.82e+11
gasLimit_MinMaxScaled	-2.322e+11	3.26e+10	-7.121	0.000	-2.96e+11	-1.68e+11
const^2	-4.858e+21	1.05e+23	-0.046	0.963	-2.11e+23	2.02e+23
const after_merge	-1.551e+14	3.53e+14	-0.440	0.660	-8.47e+14	5.37e+14
const Volume_Ethereum_MinMaxScaled	1.344e+10	2.99e+10	0.449	0.653	-4.52e+10	7.21e+10
const Close_Ethereum_MinMaxScaled	3.192e+11	4.44e+10	7.197	0.000	2.32e+11	4.06e+11
const Volatility_Ethereum_Monthly_MinMaxScaled	-6.421e+10	1.75e+10	-3.665	0.000	-9.86e+10	-2.98e+10
const erc20TxCnt_MinMaxScaled	2.392e+10	3.34e+10	0.716	0.474	-4.16e+10	8.94e+10
const size_MinMaxScaled	6.394e+10	4.63e+10	1.382	0.167	-2.68e+10	1.55e+11
const difficulty_MinMaxScaled	2.753e+11	5.45e+10	5.050	0.000	1.68e+11	3.82e+11
const gasLimit_MinMaxScaled	-2.322e+11	3.26e+10	-7.121	0.000	-2.96e+11	-1.68e+11
after_merge^2	-1.551e+14	3.53e+14	-0.440	0.660	-8.47e+14	5.37e+14
after_merge Volume_Ethereum_MinMaxScaled	4.018e+11	1.47e+11	2.740	0.006	1.14e+11	6.89e+11
after_merge Close_Ethereum_MinMaxScaled	3.026e+10	1.41e+11	0.215	0.830	-2.46e+11	3.06e+11
after_merge Volatility_Ethereum_Monthly_MinMaxScaled	-4.02e+10	9.47e+10	-0.424	0.671	-2.26e+11	1.46e+11
after_merge erc20TxCnt_MinMaxScaled	-1.242e+12	1.42e+11	-8.742	0.000	-1.52e+12	-9.63e+11
after_merge size_MinMaxScaled	-1.314e+12	2.21e+11	-5.957	0.000	-1.75e+12	-8.81e+11
after_merge difficulty_MinMaxScaled	-413.3225	939.399	-0.440	0.660	-2256.055	1429.410
after_merge gasLimit_MinMaxScaled	4.681e+14	1.06e+15	0.441	0.660	-1.62e+15	2.55e+15
Volume_Ethereum_MinMaxScaled^2	-4.177e+10	7.71e+10	-0.542	0.588	-1.93e+11	1.09e+11
Volume_Ethereum_MinMaxScaled Close_Ethereum_MinMaxScaled	-1.303e+10	1.03e+11	-0.127	0.899	-2.15e+11	1.89e+11
Volume_Ethereum_MinMaxScaled Volatility_Ethereum_Monthly_MinMaxScaled	-2.548e+11	8.97e+10	-2.842	0.005	-4.31e+11	-7.89e+10
Volume_Ethereum_MinMaxScaled erc20TxCnt_MinMaxScaled	3.451e+10	1.39e+11	0.248	0.804	-2.38e+11	3.07e+11
Volume_Ethereum_MinMaxScaled size_MinMaxScaled	3.399e+11	3e+11	1.132	0.258	-2.49e+11	9.29e+11
Volume_Ethereum_MinMaxScaled difficulty_MinMaxScaled	7.856e+11	1.91e+11	4.105	0.000	4.1e+11	1.16e+12
Volume_Ethereum_MinMaxScaled gasLimit_MinMaxScaled	-5.9e+11	1.42e+11	-4.141	0.000	-8.7e+11	-3.11e+11
Close_Ethereum_MinMaxScaled^2	-2.037e+10	5.65e+10	-0.361	0.718	-1.31e+11	9.04e+10
Close_Ethereum_MinMaxScaled Volatility_Ethereum_Monthly_MinMaxScaled	-4.034e+11	7.91e+10	-5.103	0.000	-5.58e+11	-2.48e+11
Close_Ethereum_MinMaxScaled erc20TxCnt_MinMaxScaled	-4.666e+11	1.14e+11	-4.085	0.000	-6.91e+11	-2.43e+11
Close_Ethereum_MinMaxScaled size_MinMaxScaled	-6.736e+11	1.69e+11	-3.984	0.000	-1.01e+12	-3.42e+11
Close_Ethereum_MinMaxScaled difficulty_MinMaxScaled	2.665e+11	1.13e+11	2.361	0.018	4.51e+10	4.88e+11
Close_Ethereum_MinMaxScaled gasLimit_MinMaxScaled	9.934e+10	1.05e+11	0.944	0.345	-1.07e+11	3.06e+11
Volatility_Ethereum_Monthly_MinMaxScaled^2	2.112e+10	2.46e+10	0.857	0.391	-2.72e+10	6.94e+10
Volatility_Ethereum_Monthly_MinMaxScaled erc20TxCnt_MinMaxScaled	3.268e+11	7.4e+10	4.414	0.000	1.82e+11	4.72e+11
Volatility_Ethereum_Monthly_MinMaxScaled size_MinMaxScaled	2.488e+11	2.09e+11	1.188	0.235	-1.62e+11	6.6e+11
Volatility_Ethereum_Monthly_MinMaxScaled difficulty_MinMaxScaled	1.041e+11	1.34e+11	0.776	0.438	-1.59e+11	3.68e+11
Volatility_Ethereum_Monthly_MinMaxScaled gasLimit_MinMaxScaled	-6.269e+10	1.12e+11	-0.557	0.577	-2.83e+11	1.58e+11
erc20TxCnt_MinMaxScaled^2	8.173e+10	7.07e+10	1.156	0.248	-5.69e+10	2.2e+11
erc20TxCnt_MinMaxScaled size_MinMaxScaled	2.613e+11	1.9e+11	1.373	0.170	-1.12e+11	6.35e+11
erc20TxCnt_MinMaxScaled difficulty_MinMaxScaled	-1.744e+12	1.85e+11	-9.439	0.000	-2.11e+12	-1.38e+12
erc20TxCnt_MinMaxScaled gasLimit_MinMaxScaled	1.009e+12	1.38e+11	7.296	0.000	7.38e+11	1.28e+12
size_MinMaxScaled^2	3.802e+10	1.71e+11	0.223	0.824	-2.97e+11	3.73e+11
size_MinMaxScaled difficulty_MinMaxScaled	-1.504e+12	2.77e+11	-5.439	0.000	-2.05e+12	-9.62e+11
size_MinMaxScaled gasLimit_MinMaxScaled	1.149e+12	2.09e+11	5.506	0.000	7.4e+11	1.56e+12
difficulty_MinMaxScaled^2	5.788e+09	1.1e+11	0.053	0.958	-2.1e+11	2.22e+11
difficulty_MinMaxScaled gasLimit_MinMaxScaled	8.132e+11	1.38e+11	5.899	0.000	5.43e+11	1.08e+12
gasLimit_MinMaxScaled^2	-7.27e+11	9.61e+10	-7.563	0.000	-9.16e+11	-5.38e+11

Omnibus:	2150.286	Durbin-Watson:	0.987
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1467303.488
Skew:	7.971	Prob(JB):	0.00
Kurtosis:	156.062	Cond. No.	1.52e+16

source: own illustration

Despite its high condition number, the results derived from the OLS model suggested a meaningful contribution from certain variables. However, the high condition number indicated the presence of multicollinearity, which could potentially compromise the validity of the model's interpretations. A high level of multicollinearity might lead to an OLS estimator that is very imprecise and thus has a high variance. Greene (2002) suggests that multicollinearity causes a problem if the condition number surpasses a value of 20. In this model with interaction terms, the condition number is far above this threshold, and as such, the model must be refined to obtain a more precise estimator. The following heatmap confirms and visualizes the presence of a high correlation among the variables.

Figure 10 Correlation Heatmap



Source: own illustration

To address the issue of multicollinearity, a Lasso regression was implemented. Lasso, unlike OLS, deals effectively with multicollinearity by integrating a penalty term, which can diminish some coefficients towards zero, thereby performing an implicit feature selection. This characteristic is particularly advantageous in the presence of correlated predictors.

The Lasso model resulted in a comparable R-squared value (0.543) to the OLS model and confirmed the directions of the effect of the after_merge variable seen in the initial OLS model. These observations affirm that the influential features selected by the Lasso model are not due to multicollinearity but genuinely contribute to the model's explanatory power.

Figure 11 Lasso Regression

```
R-squared for Lasso: 0.5431931420445104
                                                         Lasso Coefficients
const                                                          0.000000e+00
after_merge                                                   -1.100265e+11
Volume_Ethereum_MinMaxScaled                                   3.068704e+11
Close_Ethereum_MinMaxScaled                                    3.783150e+11
Volatility_Ethereum_Monthly_MinMaxScaled                      -1.549499e+11
erc20TxCnt_MinMaxScaled                                       -9.523750e+10
size_MinMaxScaled                                              9.645366e+10
difficulty_MinMaxScaled                                        2.348018e+10
gasLimit_MinMaxScaled                                         -3.658410e+10
const^2                                                        0.000000e+00
const after_merge                                              2.504848e+10
const Volume_Ethereum_MinMaxScaled                            -2.761547e+11
const Close_Ethereum_MinMaxScaled                              5.280610e+10
const Volatility_Ethereum_Monthly_MinMaxScaled                 9.356898e+10
const erc20TxCnt_MinMaxScaled                                  1.850346e+10
const size_MinMaxScaled                                        2.379109e+10
const difficulty_MinMaxScaled                                  2.831896e+10
const gasLimit_MinMaxScaled                                   -4.353076e+10
after_merge^2                                                  1.711015e+10
after_merge Volume_Ethereum_MinMaxScaled                      -7.522865e+10
after_merge Close_Ethereum_MinMaxScaled                        5.772150e+10
after_merge Volatility_Ethereum_Monthly_MinMaxS...            -7.865603e+10
after_merge erc20TxCnt_MinMaxScaled                           -4.223524e+11
after_merge size_MinMaxScaled                                  1.469606e+11
after_merge difficulty_MinMaxScaled                            0.000000e+00
after_merge gasLimit_MinMaxScaled                              1.785184e+11
Volume_Ethereum_MinMaxScaled^2                                -9.083905e+10
Volume_Ethereum_MinMaxScaled Close_Ethereum_Min...             5.498249e+10
Volume_Ethereum_MinMaxScaled Volatility_Ethereu...            -2.912740e+11
Volume_Ethereum_MinMaxScaled erc20TxCnt_MinMaxS...             1.083710e+11
Volume_Ethereum_MinMaxScaled size_MinMaxScaled                 9.913044e+11
Volume_Ethereum_MinMaxScaled difficulty_MinMaxS...             1.693907e+11
Volume_Ethereum_MinMaxScaled gasLimit_MinMaxScaled            -5.690606e+11
Close_Ethereum_MinMaxScaled^2                                 -4.570013e+10
Close_Ethereum_MinMaxScaled Volatility_Ethereum...            -3.419661e+11
Close_Ethereum_MinMaxScaled erc20TxCnt_MinMaxSc...            -5.334039e+11
Close_Ethereum_MinMaxScaled size_MinMaxScaled                 -4.311789e+11
Close_Ethereum_MinMaxScaled difficulty_MinMaxSc...             2.824790e+11
Close_Ethereum_MinMaxScaled gasLimit_MinMaxScaled              1.846492e+11
Volatility_Ethereum_Monthly_MinMaxScaled^2                    -8.918155e+09
Volatility_Ethereum_Monthly_MinMaxScaled erc20T...             2.181251e+11
Volatility_Ethereum_Monthly_MinMaxScaled size_M...             4.727682e+11
Volatility_Ethereum_Monthly_MinMaxScaled diffic...             4.547856e+10
Volatility_Ethereum_Monthly_MinMaxScaled gasLim...            -1.451846e+11
erc20TxCnt_MinMaxScaled^2                                      1.577668e+11
erc20TxCnt_MinMaxScaled size_MinMaxScaled                     -9.381291e+10
erc20TxCnt_MinMaxScaled difficulty_MinMaxScaled               -6.756794e+11
erc20TxCnt_MinMaxScaled gasLimit_MinMaxScaled                  4.617973e+11
size_MinMaxScaled^2                                           -1.657908e+11
size_MinMaxScaled difficulty_MinMaxScaled                     -4.585644e+10
size_MinMaxScaled gasLimit_MinMaxScaled                        4.328942e+09
difficulty_MinMaxScaled^2                                     -7.131879e+10
difficulty_MinMaxScaled gasLimit_MinMaxScaled                  1.956838e+11
gasLimit_MinMaxScaled^2                                       -1.581695e+11
```

Source: own illustration

5.1.3 Degree of Decentralization

The following chapter addresses one of the critical characteristics of Chainlink and blockchain technology in general. Namely, the degree of decentralization, which would ideally be high to prevent a single party from having most of the influence and the possibility of delivering incorrect data. It is divided into two subchapters focusing on the Gini coefficient and the Herfindahl-Hirschman index.

5.1.3.1 Gini Coefficient

The Gini coefficient, or the Gini index, is an indicator of statistical variability intended to represent income, wealth, or consumption inequality within a group or usually within nations (The World Bank, 2023). The concept is frequently used in economics. It is based on the Lorenz curve, which plots the total node transactions on the y-axis delivered by the nodes cumulatively on the x-axis. A perfect 45-degree line would mean perfect equality and a Gini coefficient of 0. This model requires all the values to be positive, which in our case, is also given. An alternative interpretation of the Gini coefficient is half of the relative mean absolute difference, in line with the definition derived from the Lorenz curve. The relative mean absolute difference is calculated by dividing the mean absolute difference by the average, providing a scaled measure.

A Gini coefficient of 0 signifies complete equality, where all income or wealth values are identical. Conversely, a Gini coefficient of 1 represents the utmost inequality among the values. To illustrate, if every person has an equal income, the Gini coefficient would be 0. Conversely, a Gini coefficient of 1 suggests that within a given group, one individual possesses all the income, wealth, or consumption, while everyone else possesses none. In the case of an oracle, a Gini coefficient of 0 would mean all the nodes deliver and receive the same number of transactions. Accordingly, a Gini coefficient with a value of 1 would mean that only one single node delivers all the data and receives all the rewards.

Not only is decentralization one of the pillars of blockchain, but also an integral part of solving the oracle problem. Chainlink aims to solve the named problem by providing a decentralized solution with the same standards as the blockchain itself. Thus, a low value for the Gini coefficient can be considered optimal for Chainlink.

The Gini coefficient is a relative measure, not an absolute one, which directly leads to the limitations of this model and the limited assumptions that can be made. It might be possible to have a low Gini coefficient with a lot of low income and a few very high earners. To solve this problem, the Herfindahl-Hirschman index is determined in the following chapter.

There is no universal cut-off for the different Gini coefficient values, although some standards have gained wide use. For example, UNICEF (2018) defines the following cut-off values:

0.2 – 0.3: corresponds with relative equality.

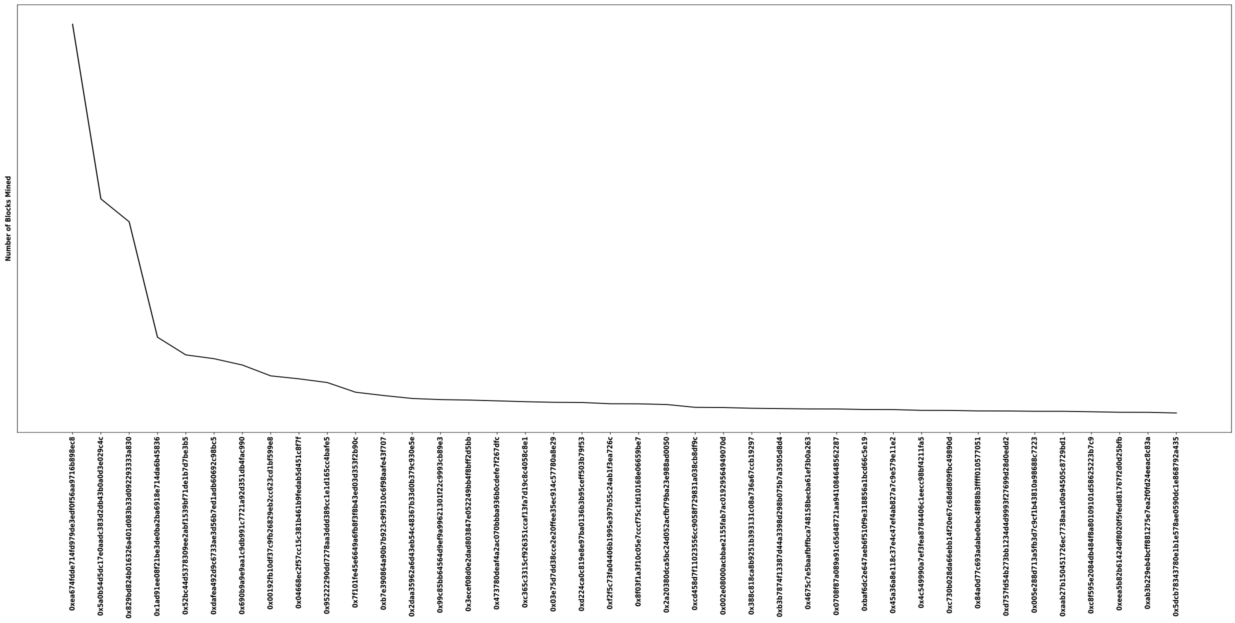
0.3 – 0.4: corresponds with relatively reasonable equality.

0.4 – 0.5: corresponds with high inequality.

> 0.5: corresponds with severe inequality.

To further elaborate, most European countries have a Gini Index between 0.22 and 0.38, with Switzerland being at 0.33 in 2018 (CIA, 2018). In stark contrast, the Gini index for Ethereum nodes before the switch to PoS was around 0.85 and approximately 0.84 for Bitcoin nodes. As can be seen in Figure 12, the curve takes a steep dive in the number of blocks mined, which indicates high inequality. For

Figure 12 Visualization of the number of blocks mined and the biggest miners (by



address)

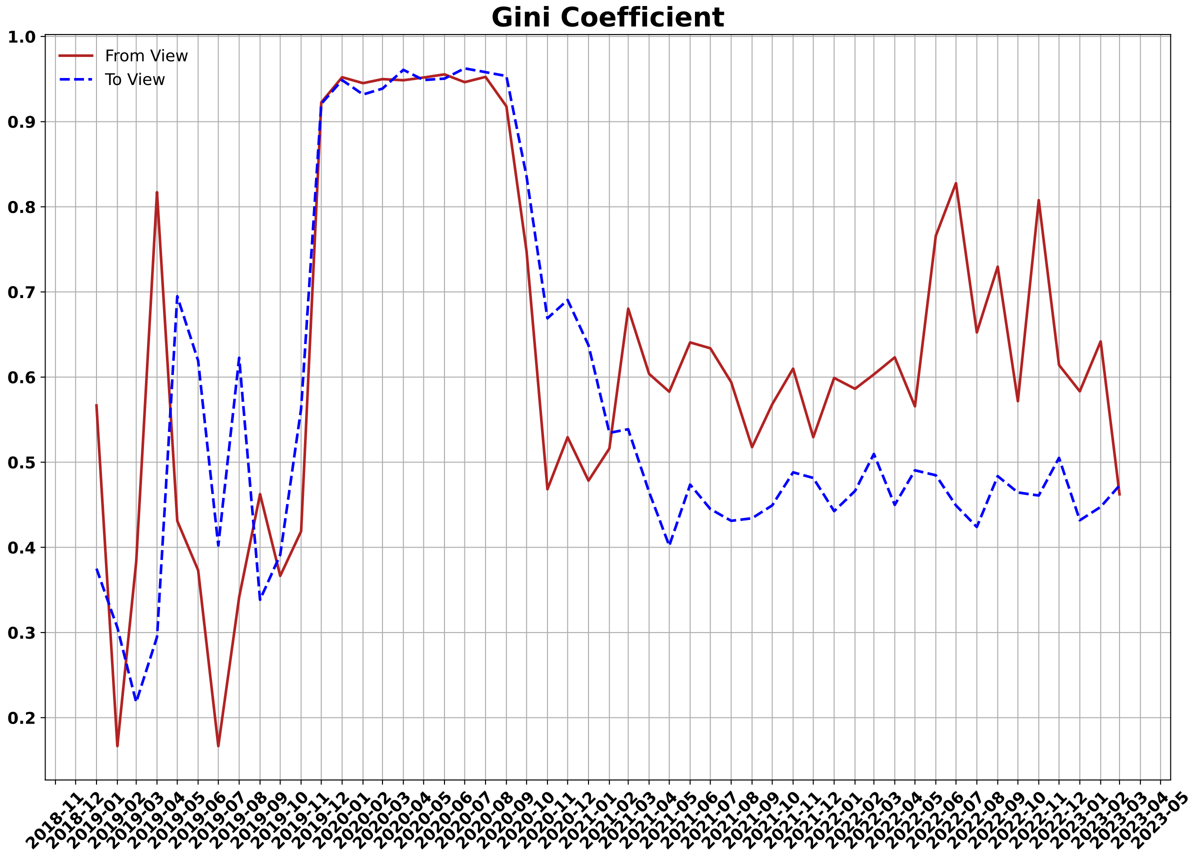
Source: own illustration

miners, the inequality was even more remarkable, according to David (2018). Furthermore, PoS does not lead to a lower Gini Coefficient than PoW. In general, there is evidence indicating high Gini coefficients for Tokens and Coins. The below plot shows high inequality among the largest 40 miners.

The visualization is close to following a hockey stick graph, indicating a high level of inequality among the blocks produced. To quantify the inequality, we can compute the Gini Coefficient for the Miners. During the PoW period of Ethereum, the computation of this study yields a value of approximately 0.7, while the switch to PoS slightly increased the Gini Coefficient to 0.73. These findings are in line with the study of David (2018).

In the following figure, the Gini coefficient for the two directions of transactions is showcased.

Figure 13 Gini Coefficient over Time



Source: own illustration

The graph above visualizes the two directions the transactions can be divided into. As the name suggests, the *From View* (the red line) aggregates the node operator addresses based on the originating transaction, i.e., the sender of the transaction. In contrast, the *To View* (blue line) aggregates the node operator addresses based on the recipient of the transaction. The recipient of the transaction is not necessarily the one that delivered the data, which is essential for analyzing the Gini coefficient. This is the case for the DDM and the OCR models.

As can be conducted from the graph above, the Gini coefficient for both views shows a high rate of volatility in the chosen timeframe. Between early 2019 and the last quarter of 2019, the lowest point for the *To View* is 0.22, and the highest is 0.7. For the *From View*, respectively 0.15 and 0.82. The end of this period marks a sharp increase in both Gini coefficients.

Between October 2019 and August 2020, both coefficients remained stable at around 0.95, followed by a drastic decrease until the end of 2020. What led to this highly unequal state has yet to be found. There are multiple possibilities, which are discussed in the discussion section of this report.

The switch from PoW to PoS at the beginning of 2021 was the starting point for less volatile trends, with both coefficients remaining over 0.4. Whereas the *From View* coefficient remains steady between 0.4 and 0.5, the *To View* coefficient remains volatile, with the lowest point being slightly above 0.5 and the highest just above 0.8. Even though there is no standardized cut-off point for the Gini coefficient, anything above 0.5 is considered unequal, which has been the case for both views since 2021.

One possible reason for the change to a more visible trend in early 2021 is the switch from the BRM to DDM/OCR. As explained in the theory chapter of this report, the BRM system works like a direct request to one node. This makes the relative amount of data delivered in a timeframe dependent on real-world happenings, such as rising interest in crypto, rising attention of DeFi, or new regulatory challenges or opportunities. The responsible nodes were to deliver more data through rising requests about certain topics, thus showing more transactions. Since early 2021, the DDM and OCR models have led to many different nodes delivering the data and aggregating everything off-chain. One single node is chosen to deliver the aggregated and signed data, which receives the according rewards again.

5.1.3.2 Herfindahl-Hirschman Index

To get a better understanding of the degree of decentralization, a second parameter, namely the Herfindahl-Hirschman index (HHI), is calculated for the *From View* and the *To View*. The Herfindahl-Hirschman index (HHI) is a metric used to assess the relative size of firms within an industry and serves as an indicator of competition among them. In this case, nodes and the number of transactions can be observed. Calculating the HHI involves squaring the market share of each firm in the industry and summing these values. Typically, this calculation is based on the market shares of the 50 largest firms. The resulting figure reflects the average market share, weighted according to each firm's market share. It ranges from 0 to 1, with higher values indicating a market with a small number of large firms and lower values suggesting greater competition, which means a market with many small firms. In the context of this report, 0 means there are lots of nodes, all having small amounts of transactions. Accordingly, one equals a few nodes being responsible for large numbers of transactions. The advantage of the

HHI over measures like the concentration ratio is that it assigns greater significance to larger firms.

Whereas the Gini coefficient measures the degree of inequality in a distribution, the HHI measures the diversity or concentration within a market. This gives more insight into the concentration and participation of the nodes within the Chainlink oracle. Two following cases can be distinguished:

HHI < Gini coefficient:

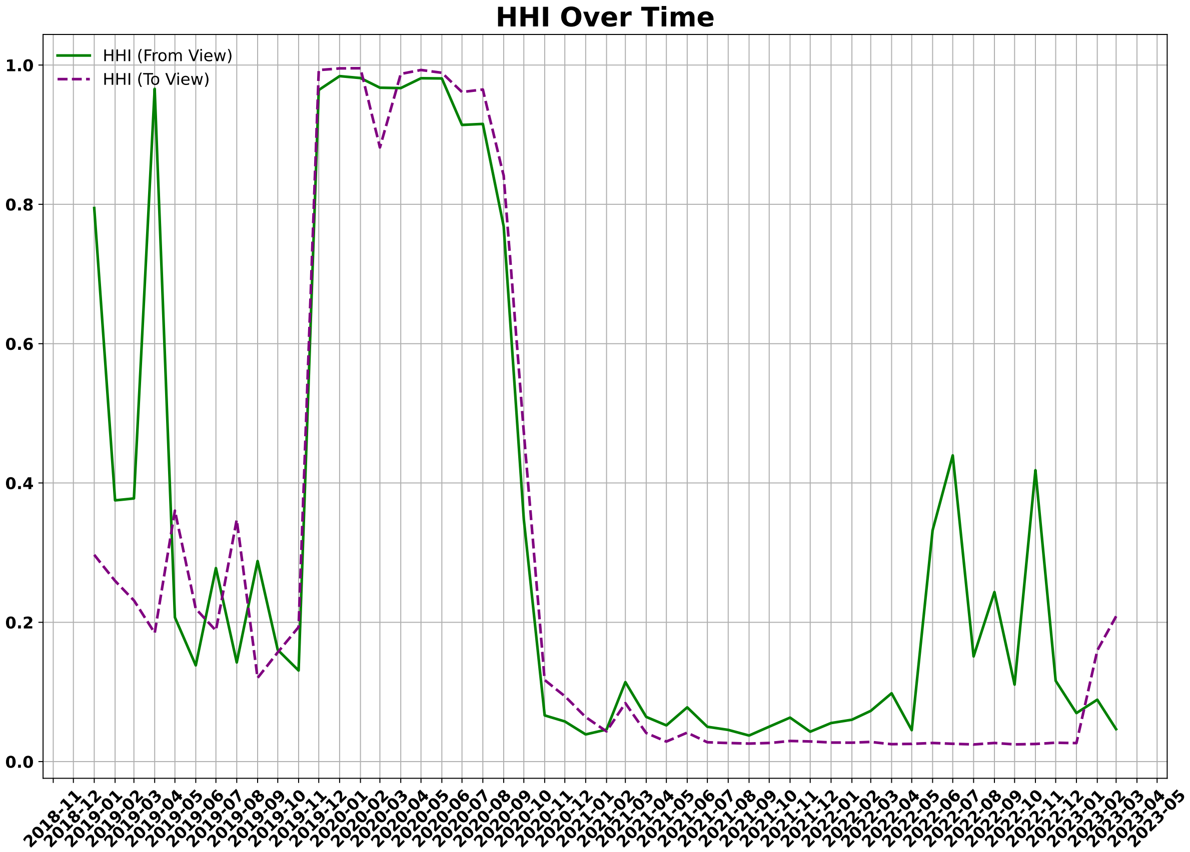
Even though inequality in the distribution of transactions is visible, the market is relatively less concentrated. That means a more diverse set of entities is participating in the transactions, even though some major dominant players exist. These dominant players have incomes of similar size. For example, two out of ten players have a much higher income than the rest. At the same time, these two have identical incomes.

HHI > Gini coefficient:

A small number of nodes is responsible for most transactions. That implies a higher concentration of transaction activity within this small number of nodes, while the general distribution might not be as unequal. In an extreme case, a small Gini index and a high HHI translate to many nodes, all delivering small amounts of data, but the difference in income can be significant within this subset. For example, two out of ten players have a much higher income than the rest. The player with the highest income earns ten times more than the second-placed, who still earns more than the third.

For this study, the discrepancy between the Gini coefficient and the HHI indicates that while there may be some concentration of transactions among a few entities, there is also broader participation and diversity in the market.

Figure 14 Herfindahl-Hirschman Index (HHI) over Time



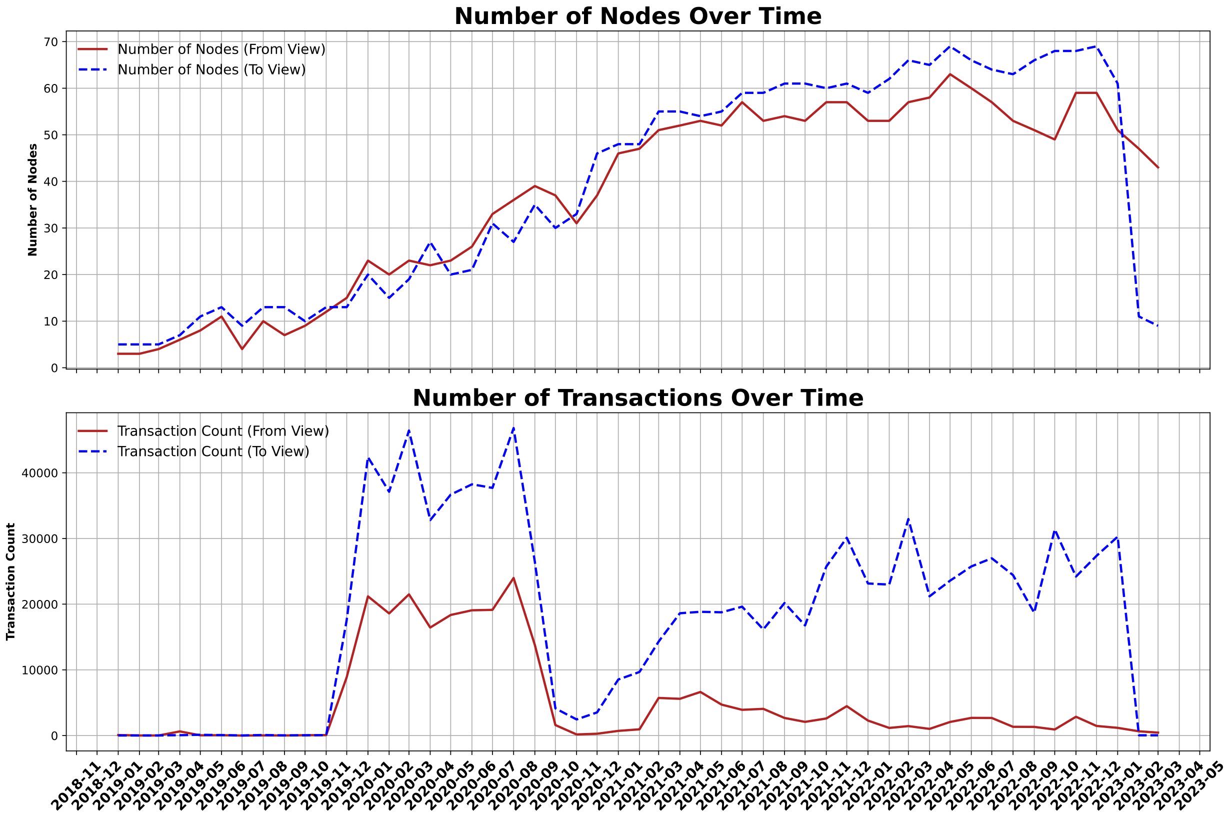
Source: own illustration

As seen above, the HHI shows similar trends to the Gini coefficients until the beginning of 2021. From the end of 2020, the HHIs for both Views remain low, with the *To View* keeping a constant index at around 0.05, which indicates many nodes, each responsible for a small number of transactions. A similar, rather volatile course can be observed for the *From View*, even though it remains below the corresponding Gini coefficient.

The similar course of both indexes indicates similar trends, which is sensible. Moreover, the Gini coefficient suggests severe inequality since 2021 for both views, but with the help of the HHI, the inequality seems not to be as pronounced as previously indicated. Even though only a small number of nodes is responsible for the transactions, this small number seems to possess equal income. Simultaneously, the time between October 2019 and August 2020 shows a similar value as the Gini coefficients.

When comparing both graphs, the beginning of 2021 marks a turning point within Chainlink. To get further insight into the degree of decentralization, the total number of nodes over time and the number of transactions over time are showcased.

Figure 15 Number of Nodes and Transactions over Time

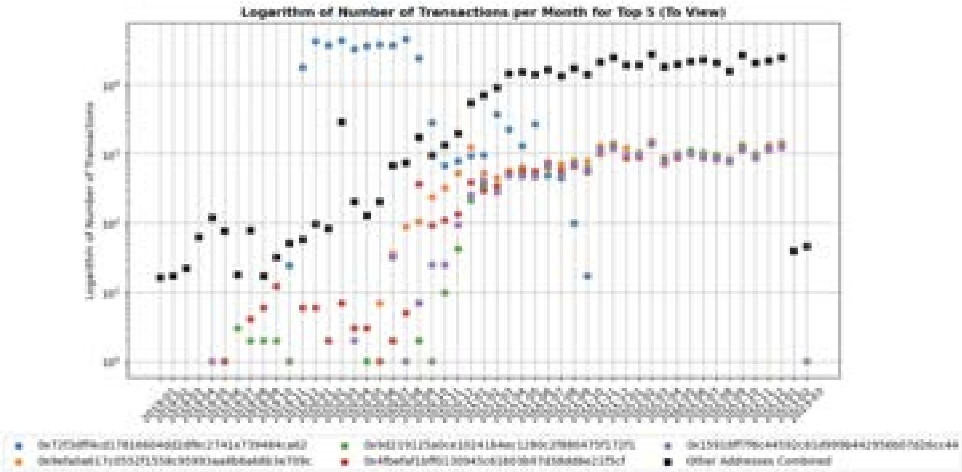


Source: own illustration

The large number of transactions between October 2019 and August 2020 coincides with the high values indicated by the Gini coefficient and the HHI. Furthermore, the transaction count for the *To View* showed signs of growth for the following 14 months, followed by a sharp decline at the beginning of 2023. Meanwhile, the transaction count for the *From Views* remained stable and consistently below the *To View*. By the beginning of 2023, *To View's* transaction count sharply declined to almost 0 transactions. The same thing happened to the nodes for the *To View*.

To find out, what caused the peak between October 2019 and 2020, the transactions were screened for the most active nodes. This clearly shows the indication of the Gini and the HHI that very few, in this case, a single node, is responsible for most of the transactions. The following two graphics show the transaction distribution of the most prominent five nodes (colored circles) in comparison to the remaining nodes (black squares). Note that the scale is logarithmic. A significant imbalance during the period of very high transactions can be observed. In the *To View*, it can be seen that the imbalance weakened in February 2020 – as captured by the HHI but not by Gini. Figure 16 shows once again a few big nodes having equal amounts of transactions while being responsible for a large number of the overall transactions i. e., a high Gini coefficient, and a low HHI.

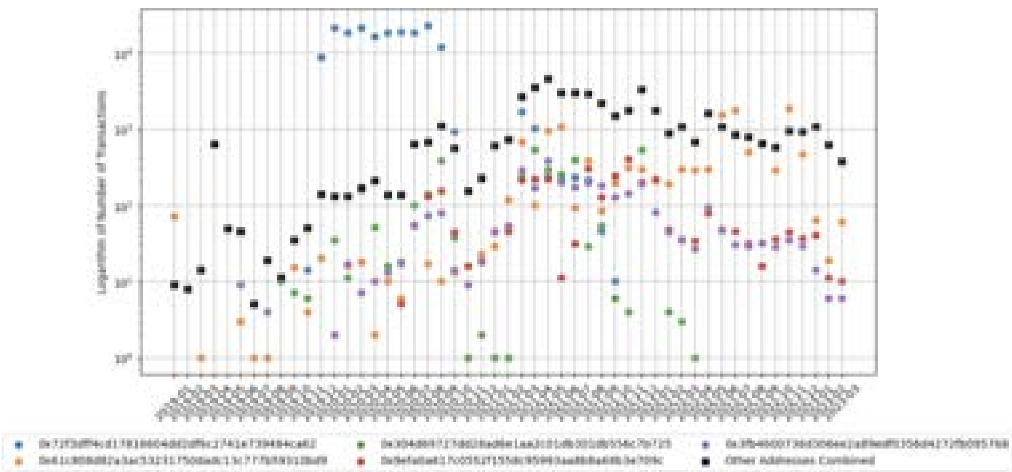
Figure 16 Logarithm of Number of Transactions per Month for Top 5 (To View)



Source: own illustration

The difference between both views can currently not be explained and might serve as a stepping stone for future research.

Figure 17 Logarithm of Number of Transactions per Month for Top 5 (From View)



Source: own illustration

6. Discussion

Drivers of Digital Currency Prices

The market for digital currencies is impacted by a multitude of factors, including investor attitude, current market trends, regulatory changes, and especially the corporate crises that have recently taken place. These drivers have the potential to significantly affect the price of cryptocurrencies, including the LINK token. Speculation and market sentiment regarding Chainlink's prospects, collaborations, technological improvements, and competition in the oracle field may also influence price swings.

Digital currencies are highly volatile and often driven by either euphoria or panic. The random walk theory indicates that asset price movements are random and that past prices cannot reliably anticipate future prices. According to Malkiel and Fama (1970), the random walk model combined the assumption that prices fully reflect all available information and that price fluctuations are independent. However, assuming that prices are random would imply that the market for digital currencies is efficient.

A further approach to elaborate on the price dynamics of digital currencies is through the noise trader model. De Long et al. (1990) established a model that incorporates market participants' biased misperceptions. Those so-called noise traders can be characterized by the fact that they tend to overestimate or underestimate the expected return of a risky asset by a variable ρ_t, in a given period. This leads to inflated or deflated prices due to the excessively bullish or bearish behavior driving asset prices.

Token Economics

This study investigated the impact of Ethereum's from PoW to PoS on transaction fees, explicitly examining changes in the average minimum gas price per day. The given hypothesis suggested that the change in consensus mode resulted in lower transaction fees for market participants.

A two-step analytical approach was employed, first utilizing an OLS regression model with no control variables and then incorporating potential influencing factors such as volatility and difficulty. The OLS regression model confirmed the negative impact of the Ethereum merge on the minimum transaction fees. However, with an R-squared value of 0.02, it was obvious that the transition event accounted for merely 2% of the observed variation in transaction fees.

However, it is crucial to keep in mind that this analysis, focusing solely on transaction fees, is expected to oversimplify the broader consequences of the transition from PoW to PoS on the Ethereum blockchain. The shift in consensus mechanisms has likely influenced a wide array of network attributes beyond transaction fees.

Firstly, the transition might affect the overall throughput of the network. With the switch to PoS, the Ethereum network could potentially increase its capacity for transactions per second (TPS), thereby allowing increased scalability. Scalability is essential for accommodating the growing user base and Dapps, which in turn could drive further changes in transaction fees.

Secondly, this shift could imply a change in the behavior of market participants. For instance, PoS requires less energy consumption than PoW, which could attract environmentally conscious investors and developers to the Ethereum ecosystem. This transition aligns with the increasing emphasis on sustainable practices in the digital world, which could enhance Ethereum's reputation and increase its user base, affecting transaction fees. Furthermore, the switch might influence miners' incentives as they transition from mining to staking, changing the dynamic of transaction validation and potentially the distribution of block rewards in Ether. The entry barriers to running a validator node are expected to be lower as opposed to miners during the PoW consensus periods. As a result of the increased competition, the validators might settle for a lower share in rewards and thus drive down transaction fees.

Thirdly, the consensus mechanism change may impact the security dynamics of the Ethereum network. PoS is believed to provide better security guarantees than PoW, mainly because it is more costly for an attacker to acquire 51% of the staking power than 51% of the hashing power. This enhanced security may further stimulate confidence and growth in the network, leading to additional effects on transaction fees.

Lastly, changes in network congestion could also be observed post-transition. A faster and more efficient network could decrease network congestion, resulting in lower gas prices and improved user experience. Simultaneously, the increased capacity might attract more DApps and transactions, potentially causing a rise in congestion again.

In summary, the impact of the consensus mechanism switch extends far beyond transaction fees, influencing numerous aspects of the Ethereum blockchain ecosystem. This makes it essential to interpret the results of this study within the broader context of these potential changes. Therefore, control variables related to transaction volume, price volatility, closing price, number of ERC20 transactions, block size, blockchain difficulty, and gas limit were included, which improved the model's R-squared to 0.377, meaning these variables accounted for 37.7% of the variation in transaction fees. Despite these enhancements, a significant portion of the variation in minimum gas prices remained unexplained.

Interactions between variables were introduced to refine the model further, increasing the explanatory power to 57.2% but raising concerns about multicollinearity. To address this, a Lasso regression was employed. It effectively dealt with multicollinearity through its implicit feature selection and reinforced previous findings, confirming the impact of the Ethereum merge on minimum gas prices.

This confirms hypothesis H1, suggesting that the transition to PoS led to an average reduction in transaction fees. However, as shown in the visualization of the minimum transaction fees (Figure 9), fees continue to exhibit significant volatility, potentially influenced by external market factors like adverse news events or regulatory changes.

Going forward, these results could be influenced by additional variables not considered in the models, such as network congestion or changes in market behavior. Hence, future research may aim to uncover these unknown factors to increase the model's explanatory power. Also, additional investigation into the performance of this model under different market conditions, such as significant and consistent price increases, would be a valuable contribution to the current understanding.

The findings of this study of lowered transaction fees after the transition to PoS contradict with current findings by Kapengut and Mizrach (2023) that the switch in consensus mechanism from PoW to PoS resulted in higher transaction fees on the Ethereum blockchain. The methodology of this study diverges from that of Kapengut and Mizrah (2023) in a few key aspects. Firstly, this analysis is based on the minimum transaction fee, which serves as a proxy for the lowest requirement of a validator to include a transaction in a block, while Kapengut and Mizrah (2023) focused on the median transaction fee. Furthermore, different time series were utilized, and last but not least, the calculations were done based on a block level as opposed to the transaction level of Kapengut and Mizrah (2023).

Degree of Decentralization

Since switching from BRM to DDM/OCR, Chainlink has demonstrated higher decentralization, as seen by the HHIs and Gini coefficients for the *From View* and the *To View*, respectively. According to Chainlink, the node that transmits the data and receives the reward must be completely random.

Our results for the Gini Coefficient differ from existing results. According to Kuśmierz and Overko (2022), the Gini coefficient for the LINK token remained stable at around 0.8 between 2018 and 2022. It is essential to emphasize the difference in wealth and income distribution since most studies focus on the former. In contrast to their results, we show that the Gini coefficient has been relatively stable at around 0.4-0.5 since 2021, accompanied by an objectively low HHI, which indicates good diversity for the most active nodes. Another difference between existing studies and this is the fact that others show a stable Gini coefficient, but here we have highly volatile values between 2019 and 2021.

Even though a Gini index of 0.4 to 0.5 is considered highly inequal for an economy, in the blockchain world, it is comparably low. For blockchain mining, the centralization is exceptionally high due to the mining process being dependent on expensive resources like electricity or hardware and software capacity. This leads to mining pools responsible for a large majority of blocks mined. For PoS-based blockchains, the probability that a node gets selected for validating a new block depends on its "wealth". This further leads to a high level of centralization.

The reason for the peak between October 2019 and August 2020, when both values reflect nearly complete centralization, is still undetermined. Even if the peak fades after a month, there are more transactions at this peak than usual, which could be attributed to several factors, such as increasing adoption, new partnerships, or new technology. Another reason would be a significant relationship with organizations like Google. Since the Gini coefficient and the HHI peaked and stayed at nearly 1, and the transaction number was huge, one or a couple of nodes must have been responsible for many transactions. To identify the respective nodes, it had to be checked for the activity of all analyzed nodes. It turns out that one single node was responsible for nearly all transactions. Due to blockchain's secure and anonymous nature, the type of data it delivers or to whom it belongs is impossible to find out. Furthermore, it is not clear how well implemented the OCR model at Chainlink really is.

The data suggests that the OCR model at Chainlink led to the desired result of more decentralization. This is highly valuable for delivering accurate and correct data. However, it has been criticized that the security provided by keeping the whole oracle on-chain could be lost due to OCR. On one side, OCR is subject to less transparency, while on the other, it leads to more decentralization and lower fees. It must be mentioned that the list of nodes considered for this study is not publicly available anymore, and thus it is not guaranteed to be complete or correct. By that means, the results are limited to the observed 84 nodes.

7. Conclusion

In this study, three key aspects of Chainlink's ecosystem were examined: the drivers of digital currency prices, the transition from PoW to PoS, and the degree of decentralization.

The analysis of digital currency prices, mainly focusing on the LINK token, revealed that various factors drive price fluctuations. These drivers include investor sentiment, market trends, regulatory changes, corporate crises, and speculations about Chainlink's prospects and technological advancements. Driven by euphoria or panic, the market for digital currencies is highly volatile, making price movements challenging to predict.

The transition from PoW to PoS in Ethereum has shown an impact on transaction fees, with the switch resulting in lower minimum transaction fees. However, the broader consequences of this transition extend beyond transaction fees, including effects on network throughput, the behavior of market participants, security dynamics, and network congestion. Not only lower fees but also lower entry barriers and lower energy consumption could have a positive impact on the users. In summary, the switch from PoW to PoS led to lower gas fees, but other factors need to be studied in the future.

The study also analyzed the degree of decentralization in Chainlink. The results showed higher decentralization, as indicated by lower Gini coefficients and HHIs for the *From View* and *To View* after the switch to the OCR model in early 2021. The OCR model overall seems to have led to improved decentralization while offering accurate and diverse data delivery. However, concerns about security, traceability, and transparency have been raised.

It is essential to acknowledge that the results are limited to the observed 84 nodes and may not fully represent the entire Chainlink network. Moreover, some potentially meaningful nodes might be missing. Further research is needed to explore additional variables and market conditions that may impact the outcomes and to deepen the understanding of the dynamics within the blockchain ecosystems. As a suggestion for further research, we propose to follow through with similar research questions for other blockchains. A comparison with other oracle service providers was beyond the scope of this study, but we highly suggest doing it in case access to node addresses can be obtained.

In conclusion, the study sheds light on the complex factors affecting digital currency prices and the impact of consensus mechanism changes on transaction fees and decentralization in the Chainlink network. However, given the evolving nature of the blockchain space, including oracles, continuous research and analysis are essential to gain comprehensive insights into these complex systems.

8. Authors

Tui Haukanomm

Management

Kevin Bründler

Publications

Noé Fischer

Operations

Adrian Stock

Management

Camille Zuber

BD/CRM

9. References

Almuraqab, N. a. S. (2020). Predicting determinants of the intention to use digital currency in the UAE: An empirical study. *EJISDC: The Electronic Journal on Information Systems in Developing Countries*, *86*(3). https://doi.org/10.1002/isd2.12125

Alshahrani, N. M., Kiah, M. L. M., Zaidan, B. B., Alamoodi, A. H., & Saif, A. (2023). A Review of Smart Contract Blockchain Based on Multi-Criteria Analysis: Challenges and Motivations. *Computers, Materials & Continua*, *75*(2), 2833–2858. https://doi.org/10.32604/cmc.2023.036138

Amir, R. (2021). Understanding an Ethereum Transaction. *Etherscan Information Center*. https://info.etherscan.com/understanding-an-ethereum-transaction/

Andesta, E., Faghih, F., & Fooladgar, M. (2019). Testing Smart Contracts Gets Smarter. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1912.04780

Avalanche Dev Docs: Create Without Limits. (2023). Retrieved July 14, 2023, from https://docs.avax.network/overview/getting-started/avalanche-consensus

Band Protocol - Cross-Chain Data Oracle. (2023). Retrieved June 25, 2023, from https://www.bandprotocol.com/

Bartoletti, M., & Pompianu, L. (2017). An Empirical Analysis of Smart Contracts: Platforms, Applications, and Design Patterns. In *Springer eBooks* (pp. 494–509). Springer Nature. https://doi.org/10.1007/978-3-319-70278-0_31

Benedetti, H., Caceres, C., & Abarzúa, L. Á. (2023). Utility Tokens. In *Emerald Publishing Limited eBooks* (pp. 79–92). Emerald (MCB UP). https://doi.org/10.1108/978-1-80455-320-620221006

Benligiray, B., Milìc, S., Vänttinen, H., & api3.org. (2020). *Decentralized APIs for Web 3.0*. https://whitepaper.io/document/736/api3-whitepaper

Böhme, R., Christin, N., Edelman, B., & Moore, T. M. (2015). Bitcoin: Economics, Technology, and Governance. *Journal of Economic Perspectives*, *29*(2), 213–238. https://doi.org/10.1257/jep.29.2.213

Breidenbach, L., Cachin, C., Chan, B., Coventry, A., Ellis, S., Juels, A., Koushanfar, F., Miller, A., Magauran, B., Moroz, D., Nazarov, S., Topliceanu, A., Tramèr, F., & Zhang, F. (2021). Chainlink 2.0: Next Steps in the Evolution of Decentralized Oracle Networks. In *research.chain.link*. Chainlink. Retrieved April 19, 2023, from https://research.chain.link/whitepaper-v2.pdf?_ga=2.227020635.718264878.1681824766-466859478.1681318250

Buterin, V. (2014). *Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform*. Retrieved April 20, 2023, from https://ethereum.org/669c9e2e2027310b6b3cdce6e1c52962/Ethereum_Whitepaper_-_Buterin_2014.pdf

Buterin, V. (2016, December 30). A Proof of Stake Design Philosophy. *Medium*. Retrieved April 21, 2023, from https://medium.com/@VitalikButerin/a-proof-of-stake-design-philosophy-506585978d51

Caldarelli, G. (2022). Overview of Blockchain Oracle Research. *Future Internet, 14*(6), 175. https://doi.org/10.3390/fi14060175

CFI Team. (2022). Durbin Watson Statistic. *Corporate Finance Institute*. https://corporatefinanceinstitute.com/resources/data-science/durbin-watson-statistic/

Chainlink. (2022). Chainlink Achieves Major Scalability Upgrade With Mainnet Launch of Off-Chain Reporting (OCR). *Chainlink Blog*. https://blog.chain.link/off-chain-reporting-live-on-mainnet/

Chen, J. (2020). Finding Ethereum smart contracts security issues by comparing history versions. In *Automated Software Engineering*. Springer Nature (Netherlands). https://doi.org/10.1145/3324884.3418923

Chen, T., Zihao, L., Zhu, Y., Chen, J., Luo, X., Lui, J. C. S., Lin, X., & Zhang, X. (2020). Understanding Ethereum via Graph Analysis. *ACM Transactions on Internet Technology*, *20*(2), 1–32. https://doi.org/10.1145/3381036

Chen, W., Wu, J., Zheng, Z., Chen, C., & Zhou, Y. (2019). Market Manipulation of Bitcoin: Evidence from Mining the Mt. Gox Transaction Network. In *arXiv (Cornell University)*. Cornell University. https://doi.org/10.1109/infocom.2019.8737364

Chiu, J., & Koeppl, T. (2019). *Blockchain-Based Settlement for Asset Trading*. https://econpapers.repec.org/article/ouprfinst/v_3a32_3ay_3a2019_3ai_3a5_3ap_3a1716-1753..htm

Choi, K. (2021). What is Gas Fee. *Etherscan Information Center*. https://info.etherscan.com/what-is-gas-fee/

CIA. (2023, June 27). *Gini Index coefficient - distribution of family income - The World Factbook*. Retrieved June 27, 2023, from https://www.cia.gov/the-world-factbook/field/gini-index-coefficient-distribution-of-family-income/country-comparison

CoinMarketCap. (2023, April 25). *Top Oracles Tokens by Market Capitalization | CoinMarketCap*. Retrieved April 25, 2023, from https://coinmarketcap.com/view/oracles/

Cong, L. W., Li, Y., & Wang, N. (2020). Tokenomics: Dynamic Adoption and Valuation. *Review of Financial Studies*, *34*(3), 1105–1155. https://doi.org/10.1093/rfs/hhaa089

Connect the World's APIs to Web3 With Chainlink Functions. (2022). chain.link. Retrieved April 25, 2023, from https://chain.link/functions

Dağli, İ., PehliVan, C., & Özbay, F. (2023). The nexus between non-fungible tokens (NFTS) and cryptocurrencies. *Human Systems Management*, 1–13. https://doi.org/10.3233/hsm-220196

Data Sources | DIA App. (2023). Retrieved June 11, 2023, from https://app.diadata.org/source/defi

David. (2018, October 23). *Gini Coefficients Of Cryptocurrencies*. Retrieved June 27, 2023, from https://blog.dshr.org/2018/10/gini-coefficients-of-cryptocurrencies.html

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, *98*(4), 703–738. https://doi.org/10.1086/261703

De Vries, A. (2022). Cryptocurrencies on the road to sustainability: Ethereum paving the way for Bitcoin. *Patterns*, *4*(1), 100633. https://doi.org/10.1016/j.patter.2022.100633

Decentralized Data Model | Chainlink Documentation. (2022). Chainlink Documentation. Retrieved April 25, 2023, from https://docs.chain.link/architecture-overview/architecture-decentralized-model?parent=gettingStarted

Donmez, A., & Karaivanov, A. (2021). Transaction fee economics in the Ethereum blockchain. *Economic Inquiry*, *60*(1), 265–292. https://doi.org/10.1111/ecin.13025

Dwyer, G. (2015). *The economics of Bitcoin and similar private digital currencies*. https://econpapers.repec.org/article/eeefinsta/v_3a17_3ay_3a2015_3ai_3ac_3ap_3a81-91.htm

Ethereum. (2023, May 25). *Learnhub | ethereum.org*. ethereum.org. Retrieved May 29, 2023, from https://ethereum.org/en/learn/

Forwarder | Chainlink Documentation. (2022). Chainlink Documentation. Retrieved April 25, 2023, from https://docs.chain.link/chainlink-nodes/contracts/forwarder

Fulfilling Requests | Chainlink Documentation. (2023). Chainlink Documentation. Retrieved April 25, 2023, from https://docs.chain.link/chainlink-nodes/v1/fulfilling-requests/

Gencer, A. E., Basu, S., Eyal, I., Van Renesse, R., & Sirer, E. G. (2018). Decentralization in Bitcoin and Ethereum Networks. In *Springer eBooks* (pp. 439–457). Springer Nature. https://doi.org/10.1007/978-3-662-58387-6_24

Greene, W.H. (2002) Econometric Analysis. 5th Edition, Prentice Hall, Upper Saddle River, 802. - References - Scientific Research Publishing. (o. D.). Abgerufen am 19. Juli 2023, von https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/referencespapers.aspx?referenceid=1243286

Kapengut, E., & Mizrach, B. (2023). An Event Study of the Ethereum Transition to Proof-of-Stake. *Commodities*, *2*(2), 96–110. https://doi.org/10.3390/commodities2020006

Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, *3*(1). https://doi.org/10.1038/srep03415

Kristoufek, L. (2015). What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis. *PLOS ONE*, *10*(4), e0123923. https://doi.org/10.1371/journal.pone.0123923

Kuśmierz, B., & Overko, R. (2022). How centralized is decentralized? Comparison of wealth distribution in coins and tokens. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2207.01340

Lin, Q., Li, C., Zhao, X., & Chen, X. (2021). Measuring Decentralization in Bitcoin and Ethereum using Multiple Metrics and Granularities. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2101.10699

Malkiel, B. G., & Fama, E. F. (1970). EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK*. *Journal of Finance*, *25*(2), 383–417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

Mehar, M., Shier, C. L., Giambattista, A., Gong, E., Fletcher, G., Sanayhie, R., Kim, H. M., & Laskowski, M. (2019). Understanding a Revolutionary and Flawed Grand Experiment in Blockchain. *Journal of Cases on Information Technology*, *21*(1), 19–32. https://doi.org/10.4018/jcit.2019010102

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). https://doi.org/10.1038/s41598-021-00053-8

Nambiampurath, R. (2022). Ethereum Finally Completes The Merge. *Investopedia*. https://www.investopedia.com/ethereum-completes-the-merge-6666337

nestprotocol.org. (2022, August 31). *NEST: Decentralized Probabilistic Virtual Machine Mode*. https://whitepaper.io/document/0/nest-protocol-whitepaper

Nguyen, C. T., Hoang, D. T., Nguyen, D. N., Niyato, D., Nguyen, H. N., & Dutkiewicz, E. (2019). Proof-of-Stake Consensus Mechanisms for Future Blockchain Networks: Fundamentals, Applications and Opportunities. *IEEE Access*, *7*, 85727–85745. https://doi.org/10.1109/access.2019.2925010

Off-Chain Reporting | Chainlink Documentation. (2022). Chainlink Documentation. Retrieved April 25, 2023, from https://docs.chain.link/architecture-overview/off-chain-reporting?parent=gettingStarted

Onica, E., & Amariei, C. (2022). Using SGX for Meta-Transactions Support in Ethereum DApps. In *Lecture Notes in Computer Science* (pp. 21–28). Springer Science+Business Media. https://doi.org/10.1007/978-3-031-16092-9_2

Operator | Chainlink Documentation. (2022). Chainlink Documentation. Retrieved April 25, 2023, from https://docs.chain.link/chainlink-nodes/contracts/operator

Oracle Documentation - DIA Documentation. (2023). Retrieved June 9, 2023, from https://docs.diadata.org/documentation/oracle-documentation

Ownership | Chainlink Documentation. (2022). Chainlink Documentation. Retrieved April 25, 2023, from https://docs.chain.link/chainlink-nodes/contracts/ownership

Popovic, D. S., Avis, C., Byrne, M., Cheung, C. C., Donovan, M. E., Flynn, Y., Fothergill, C., Hosseinzadeh, Z., Lim, Z., & Shah, J. (2020). Understanding blockchain for insurance use cases. *British Actuarial Journal*, *25*. https://doi.org/10.1017/s1357321720000148

Prat, J. (2021, January 4). *Fundamental Pricing of Utility Tokens*. https://hal.science/hal-03096267/

Prewett, K., Prescott, G. L., & Phillips, K. T. (2020). Blockchain adoption is inevitable—Barriers and risks remain. *Journal of Corporate Accounting & Finance*, *31*(2), 21–28. https://doi.org/10.1002/jcaf.22415

Price Feed Contract Addresses | Chainlink Documentation. (2022). Chainlink Documentation. Retrieved April 23, 2023, from https://docs.chain.link/data-feeds/price-feeds/addresses/

Receiver | Chainlink Documentation. (2022). Chainlink Documentation. Retrieved April 25, 2023, from https://docs.chain.link/chainlink-nodes/contracts/receiver

Rieger, A., Roth, T., Sedlmeir, J., & Fridgen, G. (2022). We need a broader debate on the sustainability of blockchain. *Joule*, *6*(6), 1137–1141. https://doi.org/10.1016/j.joule.2022.04.013

Rizos, E. (2022). A Contract Law Approach for the Treatment of Smart Contracts' 'Bugs.' *European Review of Private Law*, *30*(Issue 5), 775–802. https://doi.org/10.54648/erpl2022037

Rokowski. (2023, May 26). *Announcing Witnet 2.0, including huge protocol enhancements like Proof of Stake, tokenomic updates, and tooling upgrades.* Medium. https://medium.com/witnet/announcing-witnet-2-0-48b04bd75918

Russo, A., Anta, A. F., Vasco, M. I. G., & Pietro, R. (2021). Chirotonia: A Scalable and Secure e-Voting Framework based on Blockchains and Linkable Ring Signatures. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2111.02257

Saberi, S., Kouhizadeh, M., Sarkis, J., & Shen, L. (2019). Blockchain technology and its relationships to sustainable supply chain management. *International Journal of Production Research*, *57*(7), 2117–2135. https://doi.org/10.1080/00207543.2018.1533261

Sánchez De Pedro, A., Levi, D., Cuende, L. I., & Witnet. (2017). Witnet: A Decentralized Oracle Network Protocol. In *Whitepaper*.

Sheth, N., Prajapati, P., Shaikh, A., & Shah, P. D. (2019). Casper: Modification of Bitcoin Using Proof of Stake. *Smart Innovation, Systems and Technologies*. https://doi.org/10.1007/978-981-13-1742-2_8

Srinawakoon, S., Suriyakarn, S., & Band Protocol. (2019). BAND: Decentralized Data Governance. In *Whitepaper*. https://whitepaper.io/document/589/band-protocol-whitepaper

Sukamulja, S., & Sikora, C. O. (2018). THE NEW ERA OF FINANCIAL INNOVATION: THE DETERMINANTS OF BITCOIN'S PRICE. *Journal of Indonesian Economy and Business*, *33*(1), 46. https://doi.org/10.22146/jieb.30646

Szabo, N. (1997). Formalizing and Securing Relationships on Public Networks. *First Monday*, *2*(9). https://doi.org/10.5210/fm.v2i9.548

tellor.io. (2023). tellor: a community, an oracle, unstoppable. In *Whitepaper*. https://tellor.io/whitepaper/

The World Bank. (2023). *Glossary | DataBank*. Retrived July 16, 2023, from https://databank.worldbank.org/metadataglossary/gender-statistics/series/SI.POV.GINI

Tiwari, M., Gepp, A., & Kumar, K. (2020). The future of raising finance - a new opportunity to commit fraud: a review of initial coin offering (ICOs) scams. *Crime Law and Social Change*, *73*(4), 417–441. https://doi.org/10.1007/s10611-019-09873-2

UNICEF. (2018). *Figure 2.7 National Gini index, 2003–2017*. Retrieved July 16, 2023, from https://www.unicef.cn/en/figure-27-national-gini-index-20032017#:~:text=Although%20there%20are%20no%20internationally,high%20income%20disparity%2C%20above%200.5

Universal Market Access. (2018). UMA: A Decentralized Financial Contract Platform. In *Whitepaper*. https://whitepaper.io/document/620/uma-whitepaper

Urquhart, A. (2021). Under the hood of the Ethereum blockchain. *Finance Research Letters*, *47*, 102628. https://doi.org/10.1016/j.frl.2021.102628

Wang, S., Ouyang, L., Yuan, Y., Ni, X., Han, X., & Wang, F.-Y. (2019, November 1). *Blockchain-Enabled Smart Contracts: architecture, applications, and future trends*. IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/8643084/

What is Polygon? | Polygon Wiki. (2023, March 6). https://wiki.polygon.technology/docs/home/polygon-basics/what-is-polygon

Zhang, L., Ma, X., & Liu, Y. (2022). SoK: Blockchain Decentralization. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2205.04256

Zheng, P., Zheng, Z., Wu, J., & Dai, H. (2020). XBlock-ETH: Extracting and Exploring Blockchain Data From Ethereum. *IEEE Open Journal of the Computer Society*, *1*, 95–106. https://doi.org/10.1109/ojcs.2020.2990458

Zheng, Z., Xie, S., Dai, H., Chen, W., Chen, X., Weng, J., & Imran, M. (2020). An overview on smart contracts: Challenges, advances and platforms. *Future Generation Computer Systems*, *105*, 475–491. https://doi.org/10.1016/j.future.2019.12.019