INSTITUTO
UNIVERSITÁRIO
DE LISBOA

The Brilliancy of The Trivial Many: A comprehensive analysis of retail trader’s positions on decentralized exchanges

Rodrigo Alves Pinto Silveira Pinheiro

Master in Finance

Supervisor:
PhD, José Joaquim Dias Curto, Full Professor, Iscte-Iul

Eng. Frederico Pascoal Gaião Martins Monteiro, Computer Science and Engineering, Instituto Superior Técnico

July, 2024

BUSINESS
SCHOOL

Department of Finance

The Brilliancy of The Trivial Many: A comprehensive analysis of retail trader's positions on decentralized exchanges

Rodrigo Alves Pinto Silveira Pinheiro

Master in Finance

Supervisor:
PhD, José Joaquim Dias Curto, Full Professor, Iscte-Iul

Eng. Frederico Pascoal Gaião Martins Monteiro, Computer Science and Engineering, Instituto Superior Técnico

July, 2024

To the ones who, like me, wish to belong in the fifth quintile.

Acknowledgment

First of all, I would like to thank the insights given by Professor José Dias Curto, that were fundamental to accomplish the intended result for this thesis.

Secondly, I want to thank to Frederico and Pedro, two great engineers who truly belong in the fifth quintile. Over the past 2 years, they have constantly given up their time to help me learn and develop new skills. To them, I own a debt of gratitude.

Third, I want thank to the 3 Comma Capital team for the provided insights on the topic, specially Samuel for helping me over the longs discussions about the topic and results.

I want to acknowledge my grandparents for being an example of success through hard work. I also want to thank my parents, and my younger brother, for the continuous support in my professional and personal life. Last but not least, I also want to thank Carolina for her patience, love, and support over the years.

Resumo

A ampla adoção de métodos quantitativos tem revolucionado o processo de tomada de decisões nos mercados financeiros, na medida em que os dados de múltiplas fontes são utilizados para extrair informações e gerar *insights* valiosos. Este estudo foca-se nos dados retirados das posições dos investidores de retalho. Embora estes sejam geralmente menos qualificados e frequentemente não lucrativos, investigações anteriores sugerem que, a curto prazo, o desequilíbrio das suas ordens está positivamente correlacionado com os retornos do mercado. A literatura atual atribui a razão da limitação destes investidores em obter lucros, a vários enviesamentos comportamentais que influenciam negativamente as tomadas de decisões. Nesses, destaco o excesso de negociação e a dificuldade na análise de informação.

O objetivo deste trabalho é contribuir com investigação para o mercado de ativos digitais, fornecendo dados sobre o comportamento dos investidores e demonstrando como este difere de mercados mais maduros. Ao mesmo tempo, este trabalho enfatiza a importância da gestão de dados em mercados financeiros. Para tal, foi realizado um estudo abrangente sobre os seguintes aspetos: tamanho da posição, alavancagem, alinhamento direcional e frequência de negociação, utilizando dados sobre o fluxo de ordens de uma exchange descentralizada, com um foco específico nas criptomoedas mais negociadas, tais como Bitcoin e Ethereum.

Este estudo concluiu que a maior parte do volume de negociação foi realizado por investidores que assumem posições menores e que, em média, mantêm essas posições por menos tempo. As taxas de negociação constituem uma parte significativa dos retornos, pois os mesmos são cumulativos dos investidores situados nas extremidades da distribuição, tendendo a ser relativamente piores. Em média, as posições desses investidores demoram 10 vezes mais tempo que as dos investidores com desempenhos mais próximos da média, além de usarem 28% mais alavancagem. Estas observações sobre o comportamento dos investidores foram utilizadas para desenvolver uma estratégia de investimento através de aprovisionamento de liquidez, visando obter retornos ajustados ao risco, de forma eficiente e consistente.

Palavras-chave: Finanças Descentralizadas, Blockchain, Python, Criptomoedas, Dados On-Chain, Investidores de Retalho, Tick Data

Abstract

The broader adoption of quantitative methods has deeply changed the decision-making process in financial markets as data from multiple sources is being used to extract information and produce valuable insights. This work focuses on retail traders' position data. Although less proficient and usually not profitable, research indicates that, in the short term, the imbalance in their orders is positively correlated with market returns. Current literature attributes their difficulty in making a profit to various behavioural biases, such as over-trading and information processing, that affect their decision-making.

The goal behind this work is to contribute with research to the digital assets market, through inputs about traders' behaviour and how it differs from more mature markets while emphasizing the importance of data management in today's financial markets. To accomplish this, a comprehensive study was performed on aspects such as position size, leverage, directional alignment, and trading frequency, using order flow and tick data from an on-chain decentralised exchange, with a specific focus on prominent traded cryptocurrencies: such as Bitcoin and Ethereum.

This work concluded that the biggest share of trading volume was done by traders taking smaller positions that, on average, take less time. Trading fees take a big portion of the returns, as the cumulative returns of traders who fall into one of the distribution's tails tend to be relatively worse. On average, these traders' positions take 10x more time than those with performances closer to the average, while also taking positions with 28% more leverage. These inputs about traders' behaviour were used to achieve a liquidity provision approach with good risk-adjusted returns.

Keywords: Decentralised Finance, Blockchain, Python, Cryptocurrencies, On-Chain Data, Retail Traders, Tick Data

Contents

List of Figures

List of Tables

Glossary

Alchemy: Alchemy is a provider of API access to the Ethereum and IPFS networks.. 24
alpha: The concept of a strategy generating returns.. 1

big data: Large collection of structured, and unstructured data that grows over time.. 1

cryptocurrencies: Digital currencies that are issued on the blockchain.. 1, 8, 10, 12

hedge fund: Actively managed capital that is deployed over a different set of investment strategies, with the goal of building. 5

Infura: Infura is a provider of API access to the Ethereum and IPFS networks.. 24

onchain: On the blockchain.. 12, 16, 23
order-flow: The listings of open and closed orders in a market.. 1, 5

Robinhood: Trading and investing platform for stocks, fx, cryptocurrencies, and derivatives.. 5

stablecoin: Cryptocurrencies that are pegged to the value of a traditional currency like the U.S. Dollar.. 16, 36

Acronyms

AMM: Automated Market-Maker. 13, 15

BTC: Bitcoin. v, vii, 1, 2, 8–12, 15, 16, 21, 25–27, 30–33, 35, 38, 39

CEX: Centralised Exchange. 15

DAG: Directed Acyclic Graphs. 11
dApps: Decentralised Applications. 13
DeFi: Decentralised Finance. 12, 13, 15
DEX: Decentralised Exchange. 16

ETF: Exchange Traded Fund. 12, 31, 39
ETH: Ethereum. v, vii, 1, 2, 9–13, 16, 21, 25–27, 30–33, 35, 38, 39

GME: Gamestop. 5
GMX: Decentralised Perpetuals Exchange. 2, 16, 17, 22–24, 28, 32–34, 37, 38

L2: Layer 2. 10, 11, 15, 16
LP: Liquidity Provider. 16, 35

PnL: Profit and Loss. 25
PoS: Proof-of-Stake. 8, 9
PoW: Proof-of-Work. 8, 9

RPC: Remote Procedure Call. 2, 24

SQL: Structured Query Language. 2

USDC: Circle USDC Token. 16

CHAPTER 1

Introduction

We find ourselves in a digital era where data has become the new oil, especially in the financial sector, with the increasing popularity of quantitative strategies and an abundance of data regarding all different types of securities. In investment decision-making, there can never be too much information, professionals in this field must deal with the concept of big data to uncover valuable insights derived from those datasets. Data that can be leveraged to gain relevant information for generating alpha, mitigating risk, or enhancing decision-making processes, is among the most crucial resources in the financial industry, as well-informed investors are more likely to achieve success.

In financial markets, data is primarily utilised to comprehend past events and, consequently, to predict the probability of future outcomes (Hajizadeh et al., 2010). The expertise and tools required to handle data effectively are one of the main differentiating factors that distinguish institutional investors from retail investors. The first ones are typically more adept and quicker at processing new information and making informed investment decisions, while retail investors often lack the same level of data and information access, leading to comparatively lower success rates. Studies have shown that these retail investors tend to position right in the short term, as the imbalance in their open interest tends to be correlated with market returns in the short term, despite this, they have difficulty being profitable in their trades (Barber et al., 2023). Current literature attributes this to a different set of behavioural biases, such as over-trading and difficulty in processing new information, affecting the decision-making process of these traders.

Even though these traders tend not to be profitable, data regarding order-flow and tick data has been proven to be rather interesting in providing insights into their behaviour and the market. This work aims to explore the behaviour of retail traders in the digital assets market, focusing on the two prominent traded cryptocurrencies Bitcoin (BTC) and Ethereum (ETH) while leveraging blockchain technology to access and extract valuable data regarding their trades. The objective is to provide insights into how retail traders operate in this relatively new and rapidly evolving market, and how their behaviour contrasts with their actions in more mature markets. This research offers valuable information for both academic research and practical applications for investment firms, as the empirical analysis will examine aspects such as position size, leverage, directional alignment, and trading frequency using order-flow and tick data from an on-chain decentralised exchange where perpetual futures contracts are traded.

Blockchain technology, which emerges from the broader concept of distributed ledger technology, offers the advantages of decentralization and immutability, providing more transparency and traceability in transactions. These characteristics make blockchain data more accessible and cost-effective compared to conventional financial market data, as these second ones require one investor to have more capital to make sense to pay for this data. For this study, data from the Decentralised Perpetuals Exchange (GMX) will be used, as this is one of the largest decentralised perpetual futures trading platforms on the blockchain. GMX allows for both spot and leverage trading as it enables users to open positions with leverage up to 100 times the value of their collateral. With a significant trading volume exceeding $135 billion and an extensive user base of over 300,000 users, GMX is recognized as one of the leading decentralised trading platforms for trading perpetual futures with leverage. Perpetual futures contracts, unlike traditional futures contracts, do not expire, enhancing their liquidity and simplifying trading by eliminating the need to manage multiple contracts with different maturities.

A first process to efficiently extract all this data from the blockchain was built. This made it possible to extract all Bitcoin and Ethereum trades of perpetual futures contracts taken on GMX from January 1st 2022 until December 31st 2023. To be able to analyse returns by position, a second process was built to aggregate all trade events data corresponding to each position. The dataset comprised 674,917 positions, allowing to analyse of the data for 6 different hypotheses and reaching results in light of those.

In summary, this thesis leverages the transparency and accessibility of blockchain data to conduct a comprehensive analysis of retail trader behaviour in the cryptocurrency market. This work seeks to drive innovation in the digital assets market, helping its maturation and recognition as a distinct financial asset class, and offering practical implications for future academic studies and applications within investment firms, as it provides valuable insights into retail traders' behaviour in the cryptocurrency market. The development of this work is intended to contribute to the development of research in this field while leveraging on past research from the equities, fixed income, and foreign exchange markets.

Traditional financial market data acquisition is often costly, however, this research leverages blockchain technology to obtain data directly from the source, where trades are settled. By retrieving data from the blockchain using an Arbitrum Remote Procedure Call (RPC) Node, provided by Infura, and storing it in an Structured Query Language (SQL) database for analysis, the methodology of this work employs a cost-effective and transparent approach, with all data processing conducted using Python and open source properly licensed libraries. The sourced data will regard all information regarding Bitcoin (BTC) and Ethereum (ETH) trades on the platform, their size, leverage utilisation, position return and other components. This methodology not only reduces the costs associated with data acquisition but also enhances the transparency and reliability of the analysis.

Before delving deeper into the hypotheses and analyses to be performed, the next chapter will provide a literature review on retail traders and their characteristics, as well as the main concepts of blockchain, tokenized securities, decentralised protocols, and perpetual futures trading. This will offer a comprehensive understanding of the current landscape of this emergent market and the foundation upon which the future of this technology is being built. In the third chapter, we will explore the hypotheses to be tested, examining whether the performance of retail traders can be explained by any of the biases previously referenced, or if other factors influence their performance. The fourth chapter will focus on testing the hypotheses and conducting the empirical analysis, with the ultimate goal of establishing a delta-neutral, low-risk liquidity provision approach, enabling an investor to capture the fees paid by retail traders to trade on these platforms.

In our investigation into the factors influencing retail trader behaviour, we were able to identify trading fees as the most significant determinant. Our analysis revealed that the majority of traders and trading volume exhibited an average cumulative return between -1% and 1%. The substantial impact of fees on traders' lack of success is evident, as these fees significantly reduce their returns. Additionally, traders who fall into the tails of the distribution tend to engage in fewer trades, averaging around 1.5 trades, but these trades generally have longer duration and higher leverage, making them pay higher borrowing fees on their trades. Ultimately, we proposed a liquidity provision strategy where liquidity providers can capture a portion of the fees paid by retail traders. This strategy, designed to be delta-neutral, with a volatility of 1.044% and an average yearly return of 12.88%, offers a systematic approach to benefit from the trading activities of retail participants.

CHAPTER 2

Literature Review

Retail traders are typically investors who tend to trade with their own money and have less capital than institutional investors. These market players tend to be less professionalized, having less access to advanced technological and analytical tools to help investment decision-making, compared to more proficient institutional ones. Even though these are considered to be less proficient market players, in recent years, institutional investors have been paying more attention to their trading and investing patterns.

Back in January of 2021, a surge in Gamestop (GME) trading volume made hedge funds concerned about an imminent short squeeze triggered by retail traders and what is called "Dumb Money". This unprecedented battle of David vs Goliath was getting attention from Wall Street, even before Robinhood, the main platform responsible for the order-flow routing of Gamestop, blocked traders from buying the stock and its derivatives, while also being accused of routing the order-flow to Citadel (Malz, 2021). There has been research pointing out the forecasting properties of order-flow and tick data, as it can be used to predict exchange rate movements (Sager & Taylor, 2007), or for stock market prediction with neural networks using tick data (Selvamuthu et al., 2019). After the Gamestop episode, hedge funds went public saying they are now paying more attention to this type of market movement by "Dumb Money", given the huge bill the short hedge funds paid afterwards (McCabe, 2022).

2.1. Retail Traders

According to a Nasdaq article (Mackintosh, 2020), retail investors accounted approximately for 20% of the United States equities market trading volume, a figure that has seen substantial growth in recent years, since the COVID-19 pandemic. The article also highlights the report from the Federal Reserve regarding the significant ownership of U.S. stocks from American households and non-profit organizations, as these own around 77% of the U.S. equities market capitalization, including stocks they own directly, as well as those held through mutual funds and pension funds. The most recent report from the United States Federal Reserve (2024) shows that, as of January 2024, retail investors and non-profit organizations together hold nearly $123 trillion in assets.

Institutional investors are generally more skilled and quicker at processing new information and making informed investment decisions, whereas retail traders often do not have the same level of access to data and information, resulting in comparatively lower success rates. Behavioural finance literature suggests that retail traders exhibit both emotional and cognitive biases, which contribute to their under-performance in the market

(Talhartit et al., 2022). Cognitive biases include conservatism, information availability, domestic market bias, and decision-framing bias. Emotional biases encompass regret and loss aversion, overconfidence, and status quo bias, significantly influencing traders' decisions. This thesis will also explore additional behavioural biases affecting traders' performance.

2.2. Behavioural Biases

One differentiating factor in retail traders' profitability is market volatility. In research about traders' profitability in Options markets (Jarnecic & Liu, 2006), it was concluded that a good part of both institutional and retail traders' profitable positions was related to liquidity provision. While analysing the key differentiating factors between institutional and retail traders, it was found that in periods of high market volatility, institutional traders were more likely to make position-taking profits, while the retail traders in these scenarios would be even more likely to incur losses. The capability of institutional investors to process information and take actions based on it was the main explanation for why retail traders are likely to do the opposite of institutional investors in periods of high volatility.

As suggested by Baig et al. (2022) retail traders have a good connection with market volatility in periods of crisis, as evidenced by the 2021 COVID-19 crisis. The boom of retail trading activity results in more market instability and price volatility in the assets that these were trading. This leads to a snowball effect, where market volatility increases and decisions from retail trades lead to bigger losses, since these traders tend to accrue losses when volatility is higher and they have more of a hard time making decisions, having their biases come into play when they decide to open or close positions.

As mentioned by Jarnecic and Liu (2006), the results achieved in their research seemed to be consistent with previous research by Davis and Steil (2001), where it is suggested retail traders are less informed because of their ability to absorb and process information. According to their research, a big portion of returns is correlated with how quickly data can be ingested and insights can be extracted from it. This explains why big financial firms have invested hefty amounts of money into developing systems that can process data and output information to help decision-making processes, or even have automatic processes make the decision themselves. In the research by Barber and Odean (2013) where they analyse the behaviour of retail traders, it is also mentioned that asymmetric information/informational disadvantages are one of the reasons why individual investors underperform. Overconfidence is also one of the other reasons, in the sense that retail traders think that they know more or have a better analysis of data than the other players in the market. Barber and Odean also present evidence about the known fault of retail traders selling winners and holding losers, as they attribute this to several past documented reasons, between prospect theory, regret aversion, mental accounting, and self-control issues. Retail traders are also more likely to buy a stock that they previously returned a profit, than somewhere they returned a loss, even if the financial conditions are alike.

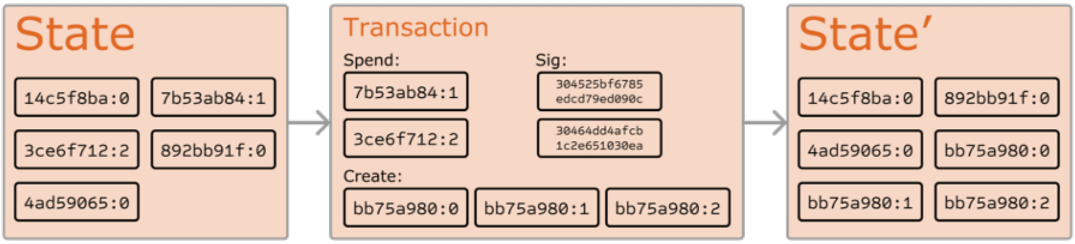
Glaser and Weber (2007) provided insights about the topic of overconfidence and over-trading, where retail investors who think they have above-average trading skills in past trading performance, did not present evidence to support those claims and were also the ones more likely to be over-trading. This ties directly to the feeling of sensation-seeking by traders. Research suggests that the feeling when trading is similar to what people seek when they gamble, as this helps to explain the excessive trading by some individuals. Barber and Odean (2013) also suggest that when there is a lack of activities that satisfy this sensation-seeking, these individuals are likely to trade more, recalling past research (Dorn et al., 2012) where evidence was presented that a one-standard-deviation increase in multi-state lottery games (i.e. Powerball and Mega Millions) is associated with a 1% reduction in small trader market participation. One good example of this is the retail trading boom during the 2021 COVID lockdowns, where people started, not only to get more into trading but also into other activities like online gambling (Gurrola-Pérez et al., 2022).

Another topic that presented influence on the traders' performance, especially the ones who over-trade, is the cost associated with trading fees. Research regarding the trades in the Taiwan Stock Exchange (Barber et al., 2014), concluded that although the best-performing traders of one year end up with positive returns in the year afterwards, their pre-fees and after-fees results are very different. This difference is even bigger in the worst-performing traders, who end up with even worse returns due to the fees paid to trade. On top of these fees, most of the bad traders also tend to carry unnecessary taxable events while taking profits (selling) in winners (usually to buying losers). The topic of trading fees is well connected with sensation seeking, as one of the reasons for more trading fees and unnecessary taxable events is over-trading.

Although retail traders tend to be less likely to return a profit, there has been a growing interest in data regarding retail traders' positions and it is reflected in the cost of this data, ranging from a few tens of thousands of dollars from 1 year of data regarding the stock in the SP500 index, or hundreds of thousands if someone is looking to have a larger and more complete database with derivatives data (TickData, 2023). This has made data hardly obtainable by the common retail trader that does not have the resources or capital to invest in data, or the capabilities to extract information from it. Financial market data acquisition is often costly, however, this research will use data retrieved from the blockchain, where perpetual futures trades are settled. The applied methodology to fetch this data will demonstrate that is possible to, not only reduce the costs associated with data acquisition but also enhance the transparency and reliability of the analysis and its results.

2.3. Blockchain: A Decentralised Ledger

The concept of a blockchain emerged from the work of researchers Haber and Stornetta (1991), as they were working on a practical solution to keep the backup of digital documents and aimed to make the timestamps of those documents more secure. They published their first paper in 1991, which expressed the use of a chain to cryptographically secure blocks to protect the integrity of past information. The first conceptual model of a blockchain was then presented in the Bitcoin whitepaper by Nakamoto (2008). As proposed, the Bitcoin Network would allow for peer-to-peer transfers of a digital currency called Bitcoin, and the ones actively validating the network would all have a copy of it, to ensure both the transparency and verifiability of all transactions made, making it much harder to change the past state of the blockchain without taking control over the entire network of validators that would be spread all over the globe.



Ethereum Whitepaper: https://ethereum.org/content/whitepaper

FIGURE 1. Illustration of state changes with new transactions on the Bitcoin Network

Blockchain is a distributed ledger technology where every transaction is executed and recorded, without the need for a third party as an intermediary, while its records can be publicly accessible, or in the case of permissioned ledger systems where the records can only accessible by a pre-determined entity. Transactions are validated, executed, and chronologically recorded on the blockchain. A record of a transaction on the blockchain is immutable, in the sense that a past transaction cannot be deleted or altered (Swan, 2015). The consensus/validation system in a blockchain network refers to the process by which the network participants (nodes) agree on the validity of transactions and the state of the blockchain (Sriman et al., 2020). This is a crucial component of blockchain technology, as it ensures its integrity and security. The two main consensus mechanisms that are used in the blockchain space are Proof-of-Work (PoW) and Proof-of-Stake (PoS).

As proposed in the Bitcoin whitepaper (Nakamoto, 2008), in the PoW mechanism, miners compete to solve a mathematical equation to validate transactions and add new blocks to the blockchain. The miner who successfully solves the puzzle first is rewarded with cryptocurrencies for their effort and contribution to the network, as this reward is used as the financial incentive for more miners to validate the network. The PoW mechanism is designed to be computationally intensive, making it difficult for any single entity attempting to control the network. This mechanism is also energy-intensive, as it requires significant computational power to solve the cryptographic problems.

A few years after the release of the Bitcoin whitepaper Nakamoto (2008), the Ethereum Protocol developed a new concept for the blockchain (the Ethereum blockchain), where an increase in block space would allow more complex transactions to be executed, validated and packed in the blocks (Buterin, 2014). Although this concept was first presented 25 years ago by a computer scientist and cryptographer named Nick Szabo (Levi & Lipton, 2022) smart contracts were popularised by the Ethereum developers community. These allow for more complex tasks to be fully executed on the blockchain with a certain set of rules imposed by every contract. Smart contracts are self-executing sets of rules that are encoded into lines of code, stored on the blockchain, and automate the enforcement and execution of terms, without the need for intermediaries. These were a form of taking the security and immutability of the blockchain technology and using it for something more than peer-to-peer transactions of digital currency, like the Bitcoin network. Smart contracts were the catalyst for the development of decentralised settlement of more complex transactions/agreements. Nowadays, smart contracts are developed to execute several types of tasks in a decentralised way, while every transaction is recorded on the blockchain. Although the Ethereum blockchain started with a similar Proof-of-Work mechanism as Bitcoin, they evolved into a Proof-of-Work mechanism.

In the PoS mechanism, validators are selected to validate transactions and add new blocks to the blockchain based on the amount of cryptocurrency they stake. Validators are required to lock a portion of their cryptocurrency as a security deposit (their “stake”), which they can lose if they attempt to validate invalid transactions or a false state of the blockchain. This mechanism is generally more energy-efficient than PoW, as it does not require the same level of computational power to solve the puzzles to validate transactions. However, the PoS mechanism may be more susceptible to centralization, as typically validators with larger stakes have a higher chance of being selected to validate transactions.

In addition to introducing the concept of smart contracts, the Ethereum blockchain also modified the system for transaction execution and validation. Unlike the Bitcoin blockchain, where transactions are placed in a queue and pay a fee, known as a ”gas fee,” to be executed and reward the network, Ethereum's system is more complex due to the inclusion of smart contracts. On the Bitcoin network, gas fees are paid in Bitcoin, the native token, but the Ethereum blockchain requires gas fees to be paid in its native token (ETH), with fees proportional to the computational effort needed for transaction validation. This change creates an economic model where the cost of executing transactions on the Ethereum network reflects their complexity, incentivizing efficient use of network resources, while also creating demand for the token.

The central premise of blockchain technology is simple: instead of relying on a trusted intermediary to keep a record of transactions among network members, all members maintain their copies of the same record and ensure that all copies are synchronized. This synchronization eliminates the need for an intermediary that might want to take advantage of a system and act in its interest (Sunyaev et al., 2021). Bitcoin emerged as a response from tech engineers to address trust issues in the global financial system, especially following the subprime crisis at the end of the 2000s. Initially, blockchain technology aimed to facilitate peer-to-peer payments without the need for financial institutions. In recent years, with the development of smart contracts, blockchain technology has advanced to execute more complex tasks, enabling a wide range of applications and real-life use cases to be fully decentralised. As highlighted by the Eigen Labs Team (2024), existing digital platforms cannot be fully qualified as digital commons, as they are controlled without democratic governance, leading to significant frictions, such as leading to issues like centralized control, subjective policies, innovation velocity constraints, and unilateral policy changes. Blockchain technology offers an alternative paradigm for building the global digital commons where it is possible to have permissionless user access, user governance, and share value across the participants. All these properties arise due to the permissionless and verifiability characteristics of the blockchain. Making this technology a promising framework for building open and verifiable digital commons.

Blockchain technology is transitioning from its early stages of development to a more mature phase, evidenced by significant growth in both development and utilization in recent years (Swan, 2017). According to the "Global Cryptocurrency User Base" (2024), it is estimated that currently, the number of users of cryptocurrencies and blockchain technology is approaching 600 million. Institutional investors and commercial enterprises are beginning to enter this new asset class with more conviction, while a cryptocurrency derivatives market is also taking shape as part of the broader development and expansion of the digital assets market.

Decentralised systems, such as blockchain, face limitations, one being the restricted number of transactions each block can handle. This constraint is particularly evident on the Ethereum blockchain, which, despite technological advancements over the years, can only process about 12 transactions per second. The rise of smart contracts and on-chain protocols has led to increased transaction volumes on the Ethereum blockchain, subsequently driving up transaction costs. This occurs because the validation system requires each transaction to adjust the gas fee (the fee paid to validators for processing the transaction) to increase the likelihood of being selected from the queue. This is what is commonly known as the blockchain trilemma. The blockchain trilemma regards the difficulty in the fact that there need to be compromises in some sense to make up for one of the three aspects of the blockchain: security, decentralisation, and scalability (Sregantan, 2023). One cannot scale the blockchain in the number of transactions that are processed, without the sacrifice of decentralisation or security.

Given the increased price of gas fees to execute transactions, developers created separate mechanisms with their validation systems that "live inside" the Ethereum mainnet blockchain. Layer 2 (L2) operate on the base layer blockchain, providing a mechanism to process transactions off-chain, while ensuring the security and finality guaranteed by the main blockchain, like the Ethereum blockchain. These systems significantly increase the number of transactions per second and reduce costs associated with transactions, at the cost of the security risk of a separate mechanism that processes the transactions off-chain.

Ethereum employs various L2 solutions, such as Optimistic Rollups and zk-Rollups. Rollups aggregate multiple transactions into a single transaction, compressing the data and submitting it to the main chain (Sguanci et al., 2021). As an example, if two entities transacted 20 times between them in a short period, instead of having 20 separate transactions, the Layer 2 mechanism would only send the result of these transactions to the main blockchain. This means that all transactions inside Layer 2 carry a lower cost compared to an Ethereum mainnet transaction, making every transaction on L2 a fraction of the cost of one in the Ethereum mainnet blockchain. This made the gas fees for the execution of the transactions cheaper, allowing for several new services and protocols to exist where once it was almost impossible.

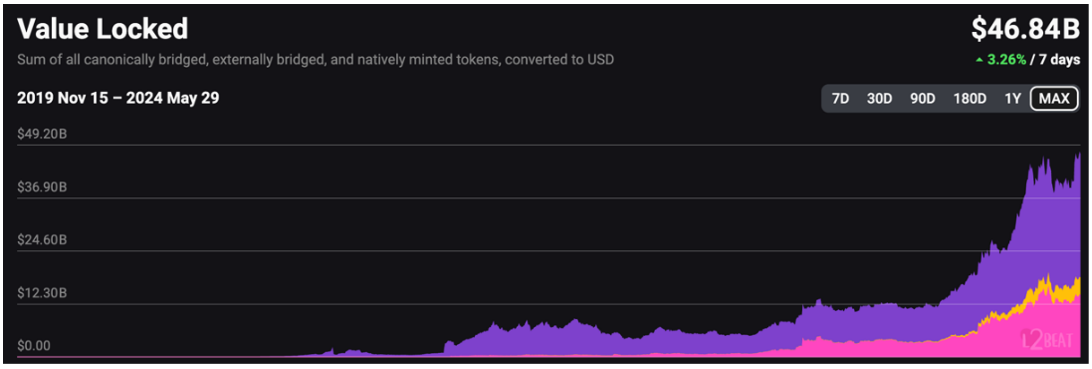


Figure 2. Total dollar amount in Layer 2 of Ethereum. Purple represents canonically bridge tokens; Pink represents natively minted tokens; Yellow represents externally bridged tokens. Source: L2Beat.com

Nowadays, there are interesting approaches that aim to develop a better-decentralised system, such as the one presented Taraxa Foundation (2022), that can overcome this trilemma without compromising the security with an implementation of a Directed Acyclic Graphs (DAG) a mathematical concept used in graph theory and computer science of a directed graph with no directed cycles (Thulasiraman & Swamy, 1992) instead of a linear block increment for the state of the blockchain, while still using a Proof-of-State validation system.

2.4. Cryptocurrencies and Tokenized Securities

Although the concept of a distributed ledger is not recent, a digital currency like Bitcoin was the first application of blockchain technology, back in 2010 (Swan, 2017). Cryptocurrencies are just one of the many real-world scenarios where blockchain technology can be used. Besides transactions and payments, there are other types of applications being developed, including property records, contractual agreements, and identity confirmations (Rejeb et al., 2023).

Cryptocurrencies, also known as digital currencies, are digital assets that can be used as a medium of exchange. In most cryptocurrencies, all records and details of their transactions are stored in a public ledger, based on blockchain technology (Y. Chen & Bellavitis, 2020). The ability to conduct a transaction without needing an intermediary, such as a financial institution, is the concept presented in the first sentence of the manifesto that created Bitcoin, the first and most famous cryptocurrency, established in 2009. Signed by the mysterious Satoshi Nakamoto, the Bitcoin whitepaper was the first document to combine the idea of a digital asset (cryptocurrency) with the concepts of distributed ledger technology (Nakamoto, 2008).

The recent spark of interest from the major investment firms in the cryptocurrency market, especially in Bitcoin and Ethereum, is a consequence of the growing desire from investors to have exposure to the asset class (Lamothe-López et al., 2024). Given the small dimension of this emerging market compared to traditional markets, this move from financial institutions allows for easier exposure from the overall investors in the market, in the form of Exchange Traded Fund (ETF)s. The cryptocurrencies market is still relatively small with a total capitalization of around $2.5 trillion (Forbes, 2024a), compared to the hundreds of trillions of dollars of the market capitalization of traditional and more mature markets.

Cryptocurrencies are just an example of tokenization, as the growing interest from financial institutions as them looking into the tokenization of real-world assets and securities for several reasons. According to the report from McKinsey and Company (2023), the main benefits of asset and security tokenization are capital efficiency, the democratization of access, savings in operational costs with cheaper infrastructures, and facilitation and transparency in compliance, audit and reporting activities. Most financial institutions look into making their operational costs cheaper, and especially when it comes to financial markets, issuance and settlement costs are a big hit on the balance sheets (Berentsen et al., 2020). Trade processing savings can go up to a few tens of billions of dollars every year. On the transparency side, it becomes easier for regulators to follow all the activity if the records are kept on a blockchain of their own. As of now, different protocols are tokenizing securities, between those securities there are treasury bills, exchange-traded funds, and even tokenization of loans.

With the evolution of smart contracts, the number of applications developed using decentralised protocols, also known as onchain protocols, has seen considerable growth. These protocols are decentralised systems and applications that operate directly on a blockchain, leveraging the underlying blockchain's architecture to provide a variety of services without requiring a centralized authority or intermediary. Prominent examples include Decentralised Finance (DeFi) protocols, which facilitate financial activities such as lending, borrowing, and trading directly on the blockchain without the need for intermediaries (Shah et al., 2023). The operation of onchain protocols relies heavily on blockchain consensus mechanisms, ensuring that all nodes in the network agree on the validity of transactions. This includes protocols that enable peer-to-peer lending and borrowing, automated market-making activities that allow token swaps without intermediaries, and other financial services that function autonomously on the blockchain. As the underlying value of digital assets grows, these decentralised activities are expanding, reflecting the increasing adoption and utility of blockchain technology.

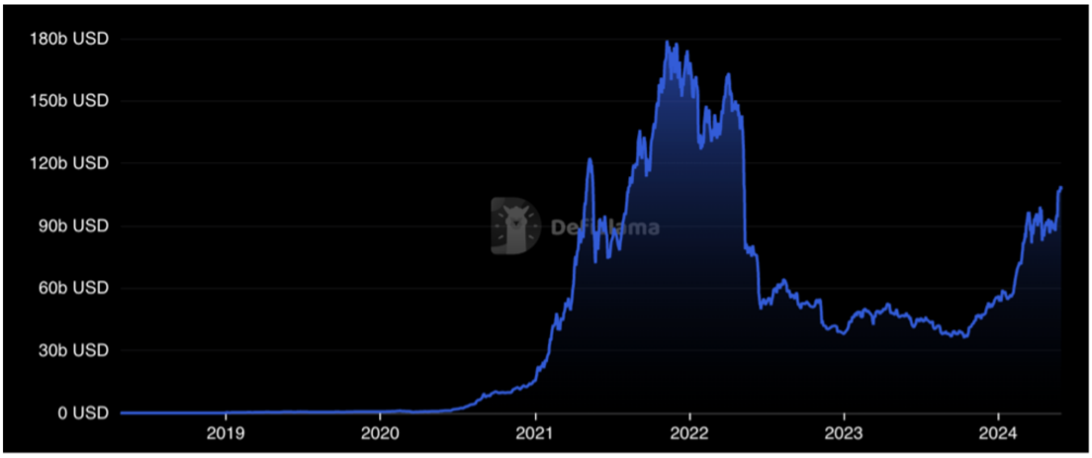


FIGURE 3. The total dollar value of DeFi Protocols across the major blockchains. Source: DeFiLama.com

Decentralised Finance represents a pivotal advancement in blockchain technology, aligning with Satoshi Nakamoto's vision of a permissionless system to democratize financial services and diminish dependence on traditional intermediaries. Leveraging smart contracts on platforms such as Ethereum, DeFi facilitates Decentralised Applications (dApps) that offer financial services including lending, borrowing, and trading without centralized oversight (Zetzsche et al., 2020). This innovation significantly enhances transparency, security, cost-efficiency, and transaction speed. By removing barriers to entry and fostering continuous innovation through its open-source framework, DeFi extends financial inclusivity to underserved populations. Furthermore, DeFi introduces new income opportunities for the "common investor" such as market-making and liquidity provision, which were previously mostly done by well-capitalized financial institutions that had resources to support such operations. The interoperability among DeFi protocols establishes a cohesive financial ecosystem, broadening the array of financial tools accessible to users, while these maintain custody over their assets, instead of a broker being responsible for it.

In times when financial institutions are looking into a future of tokenized securities, if these are issued, traded, and settled on the blockchain, there is a need for mechanisms from decentralised exchanges to take the role of broker/market maker in those transactions. Decentralised platforms allow for trading without the need of a middleman to play the role of market maker, due to their Automated Market-Maker (AMM). These are a type of decentralised exchange mechanism used in the digital assets space. Unlike traditional exchanges that use order books to match buyers and sellers, AMMs use liquidity pools and algorithms written onto the smart contract to facilitate and execute trades. Users provide liquidity to these pools and, in return, earn fees from trades.

The total liquidity in a pool is given by a simple equation, provided by the Uniswap Labs Team (Adams et al., 2020), that tracks the product of the amount of asset x and y:

$$k = x \cdot y, \tag{2.1}$$

where:

· x represents the quantities of asset x.

· y represents the quantities of asset y.

· k is a constant that represents the amount of liquidity provided by both assets.

This rather simple formula, allowed for continuous and automated trading without the need for a central intermediary. When a trade is executed, the token balances are adjusted to keep this product of quantities accurate, as this happens, the quantity of the two assets in the pool, sets the price/exchange rate automatically, without the need for a traditional order book or market makers that take this role.

The exchange rate at time p_t of two assets x, y in a liquidity pool (not accounting for fees) is then determined by the following formula:

$$p_t = \quad (r_t^x)/(r_t^y\ ), \tag{2.2}$$

where:

· r_t^x represents the pool reserves of asset x at time t.

· r_t^y represents the pool reserves of asset y at time t.

This mechanism created an automatic price adjustment based purely on supply and demand, providing a more stable, reliable, and transparent pricing model compared to traditional centralized exchanges. A predictable pricing formula also reduces slippage and makes it easier for traders to correct price differences between the liquidity pool and the broader market (arbitrage opportunities), contributing to overall market efficiency.

2.5. Perpetual Futures Trading on Decentralised Exchanges

Perpetual futures contracts are the most popular traded derivative in the digital assets market and make over $100b in volume each day (He et al., 2022). Although these are different from the traditional concept of a futures contract, the standard pricing method for these derivatives still applies.

A futures contract is a financial agreement to buy or sell an asset at a predetermined price on a future date. The terms of this contract are the size, expiration date, and settlement method for the operation. Futures contracts are used primarily for hedging against price fluctuations or for speculative purposes by investors and traders. As provided by Hull (2011), the no-arbitrage pricing formula of a futures contract for a stock with no dividend yield is following:

$$F_t = S_t \cdot e^{(r \cdot (T-t))}, \tag{2.3}$$

where:

· S_t is the spot price of the underlying asset at time t.

· r is the risk-free interest rate.

· T is the time to maturity of the futures contract.

· $T - t$ is the time remaining until the contract's expiration.

Perpetual futures contracts, unlike traditional futures contracts, do not expire, enhancing their liquidity and simplifying trading by eliminating the need to manage multiple contracts with different maturities. At $t = 0$, a long and short counterpart agrees on an initial futures price without exchanging funds, except for posting margin with the exchange and profits or losses continuously calculated and allocated to their margin accounts based on prevailing market prices.

Due to the lack of a fixed expiration date, perpetual futures are not guaranteed to converge to the spot price of their underlying asset. To address this, a funding rate mechanism is employed where long position holders periodically (usually every 8 hours) pay short position holders a funding rate that aims to shorten the gap between the perpetual futures price and the spot price, incentivizing trades that minimize this discrepancy and ensure price alignment.

For this contract to happen, the buyer pays a funding rate (p_t) to the short position holder, which is given by the following formula:

$$p_t = S_t \cdot \int_t^{t+\Delta t} FR \, dt, \tag{2.4}$$

where:

· S_t is the spot price of the underlying asset at time t.

· FR is the funding rate applied over the interval Δt.

Before L2s it was not feasible for someone to pay up to \$40 to execute a trade. However, with transaction prices under a few cents, this made it much easier for these protocols to exist on Layer 2s. One of the surging applications of smart contracts and DeFi was perpetual futures trading platforms on the blockchain. As of today, the major decentralised trading platforms account for around 10% of the total trading value in the cryptocurrency market ("DEX to CEX Spot Trade Volume", 2024). As stated in research regarding the impact of perpetual futures contracts on the volatility of Bitcoin prices (E. Chen et al., 2024), there are 3 types of perpetual futures. The ones described above are usually available on almost every Centralised Exchange (CEX), but after the collapse of the FTX exchange, several decentralised protocols emerged offering these products. The first type is AMM, similar to CEXs that work through an order book, the trading of the assets sets the price. The second ones are Oracle Pricing protocols. These second ones take the price from an onchain Oracle system that provides a secure price for the asset using a decentralised validation system like the one provided by Chainlink.

There are multiple Decentralised Exchange (DEX) which one can get data from, however, for this paper the focus will be GMX protocol, as this is one of the major decentralised exchanges with perpetual futures trading, that allows users to use leverage up to 100x the amount of collateral, as this is the value someone has to put upfront as a guarantee for their trade. This platform on the Arbitrum Blockchain, a Layer 2 of Ethereum, has accumulated over 125 billion dollars in trading volume over the past 2 years.

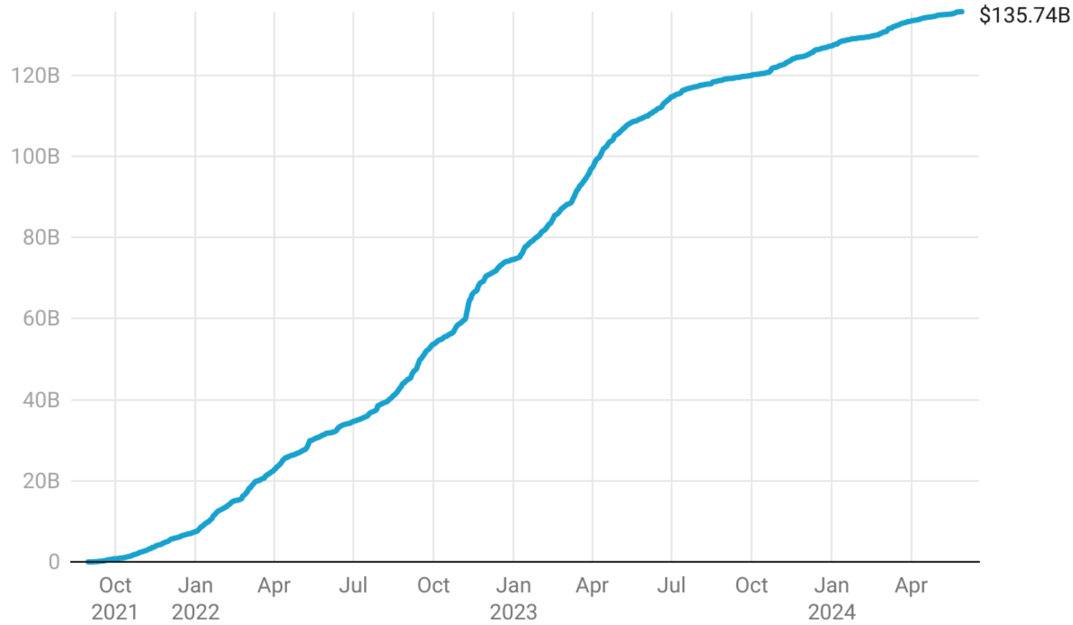


Figure 4. Total transacted volume on GMX Platform over time

GMX utilizes a unique liquidity pool model, where liquidity providers (LP) deposit assets into a pool, that consists of an index of digital assets including mostly Bitcoin, Ethereum, and USDC the know U.S. Dollar pegged stablecoin issued from Circle, that earns fees from traders borrowing leverage. The pool supports trading with minimal slippage and small price impact for large trades, making it an attractive option for traders.

When liquidity providers deposit their assets to the pool, they are given a token that represents their position, called GLP. This token acts as an index for the assets pool, which consists of 50% in stablecoins and the rest is distributed with 45% between BTC and ETH, while the remaining 5% are smaller cryptocurrencies. The GLP token can be minted using any index asset and redeemed to any index asset.

As GLP holders provide liquidity for leverage trading, they will make a profit when leverage traders make a loss and vice versa, while also earning 70% of all fees paid by the traders to borrow the leverage, a mechanism that makes GMX attractive for both traders and liquidity providers.

As in perpetual futures contracts the funding rate is adjusted to bring the spot price closer to the futures price, the borrowing fee paid per hour by retail traders on GMX is a function of the pool usage by traders. If more traders are borrowing leverage, the interest paid by these to the liquidity providers is also higher. The borrowing fee at time t (BF_t) is given by the following equation:

$$BF_t = \left(\frac{\sum_{i=1}^{n} B_{i,t}}{\sum_{i=1}^{n} A_{i,t}} \right) \cdot 0.01\%, \tag{2.5}$$

where:

· $B_{i,t}$ is the total assets borrowed from the pool at time t.

· $A_{i,t}$ is the total pool assets value at time t.

This equation gives flexibility to the borrowing rate to be higher when the liquidity pool utilization rate is also higher.

Even though all the records for all transactions are being recorded and stored on the blockchain, accessing this data is only possible through a blockchain archive node. Nowadays multiple services offer free node data requests for standard use, that allow someone to query that from the blockchain, however, if someone needs a high volume of data, with high frequency, and low latency requests, one should either pay for the services or have themselves an archive node. Currently, an archive node for the Arbitrum Blockchain holds around 9Tb of data.

In the next chapter, we will outline the methodology for this analysis and the processes established to efficiently store and analyze the necessary data for our research. In the fourth chapter, we will analyze if there are any relations between the behavioural biases set by previous research that affect traders' performance and the data that we have, or if the consensus of traders acts randomly at opening trades.

Due to the high number of positions and trades data that the GMX platform can provide, we will be able to understand if variables like the amount of leverage or trading frequency are correlated with traders' returns, or if the fees that these platforms charge takes the best part of the returns of these traders, as stated in previous research. In the case this second scenario is proven, we will delve into the liquidity provision approach that the GMX platform allows, as a way to benefit from the high fees paid by traders to liquidity providers.

CHAPTER 3

Methodology and Data

As highlighted in previous chapters, the opportunity to develop innovative research in this field is one of the main motivations behind this work. Given how relatively recent blockchain technology and the cryptocurrency/tokenized securities market are, research on these topics is starting to emerge more often.

3.1. Methodology

In the first approach to this topic, an analysis of the main differences between the best, worst and average traders will be performed, as highlighting the key differences between these two is a key factor moving forward in this work. Afterwards, the empirical analysis will be able to dig deeper into their performances over time and if there is any correlation with the market, as well as the factors that influence their trading returns, between the use of leverage, their trading frequency and trading/borrowing fees.

These analyses will help us establish if any patterns lead to the increase/decrease of the success rate of traders. Although research that was referenced in this paper points to possible relations between behavioural biases and traders' returns, there has been research related to retail traders in the futures market (Ferko et al., 2023) that points to a somehow randomness for the criteria of retail traders opening or closing trades. Some of these less experienced market participants tend to start trading as a way to learn about the topic and assess if they have the skills to do it.

In the end, data will serve as the analysis of the 6 hypotheses being raised for the analysis. After understanding more about the behaviour of the traders on the platform and which of these views is our data aligned with, we will look into developing a systematic liquidity provision approach to capture the fees paid by retail traders to the liquidity providers on this platform. Before reaching the systematic approach, the goal thought our hypothesis is to get a better understanding of how these traders behave to support the reasoning behind the implementation of the liquidity provision strategy.

3.1.1. How are the best and worst-performing traders distributed?

The goal is to identify how these traders are distributed to help differentiate the most successful traders from the least successful ones. It is expected that this provides insights on which group of traders has more influence in the overall returns and which ones should be analyzed in further analysis. To achieve this, traders were categorized into quintiles based on their cumulative returns over the observed period. The middle quintile represents traders whose cumulative returns are proximate to the mean return. The remaining four quintiles are delineated by one and two standard deviations below and above the mean, respectively. This distribution will facilitate the analysis of the differences between traders in both extremities of the distribution (first and fifth quintiles) and allow for comparisons with those exhibiting average performance (third quintile).

The cumulative return for a trader over some time t to T is given by:

$$CR_a = \frac{\sum_{j=1}^{M_a} P_{a,j}}{\sum_{j=1}^{M_a} C_{a,j}}, \tag{3.1}$$

where:

· CR_a is the cumulative return for a trader a.

· $P_{a,j}$ is the return of a position j taken by trader a between t and T.

· $C_{a,j}$ is the collateral of a position j taken by trader a between t and T.

· M_a is the total number of positions taken by trader a between t and T.

The objective is to draw a picture that will help further down in the analysis to look for specific differences in the factors that influence traders through the middle quintile, in comparison with the first and last quintiles that will correspond to both extremities of the distribution. A set of variables and measurements will be considered in comparing these groups, such as the average time that a position lasts, what is the average leverage amount, the position size, and whether the position is long or short which indicates a possible bias of traders. Other relevant factors will also be examined to provide a comprehensive understanding of the performance disparities among traders.

3.1.2. Is there a correlation between the imbalance in open interest and market returns? How do these traders perform over time in comparison with the market?

This analysis seeks to determine whether the imbalance in the daily open interest is correlated with market returns. Past research suggests a short-term correlation between market returns and retail traders' order imbalance. The order imbalance in open orders in a day can be expressed as follows:

Let:

· L denotes the total volume of long positions opened over a day.

· S denotes the total volume of short positions opened over a day.

The order imbalance (I) can be calculated as:

$$I = \frac{(L - S)}{(L + S)}, \tag{3.2}$$

where:

· $(L - S)$ represents the difference between the total volume of long positions and the total volume of short positions opened.

· $(L+S)$ represents the total volume of both long and short positions opened combined.

This study will investigate whether similar trends can be identified in the provided data sample, using the market daily returns and the daily longs/shorts order imbalance. To accomplish this, it will be conducted a Pearson Correlation test to assess if this relation exists and how strong it is. Pearson's correlation coefficient is a measure of the linear relationship between two variables. It assesses whether there is a linear association between two continuous variables X and Y (Curto, 2021).

Given two variables X and Y, with n pairs of observations for $i = 1, 2, ...n$, the Pearson's correlation coefficient r is calculated as:

$$r = \frac{\sum_{i=1}^{n}(x_i, \bar{x})(y_i, \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i, \bar{x})^2 \sum_{i=1}^{n}(y_i, \bar{y})^2}}, \tag{3.3}$$

where:

· $\bar{x}$ and $\bar{y}$ and are the means of X and Y, respectively.

· x_i and y_i are individual data points.

The hypothesis for testing the significance of Pearson's correlation coefficient (r) typically follows:

1. Null Hypothesis (H_0): The null hypothesis states that there is no linear relationship between the two variables X and Y.

$$H_0 : r = 0 \tag{3.4}$$

2. Alternative Hypothesis (H_1): The alternative hypothesis proposes that there is a linear relationship between the two variables.

$$H_1 : r \neq 0 \tag{3.5}$$

Following the first analysis, the objective is to provide more detail on how the performance of traders evolved over the 2 years. Past research tells us that traders tend to have a long bias on their positioning, and since data is respective to two years of trading (2022 and 2023), it is relevant that the analysis delves into their returns by year. During the 2-year time, BTC returned -62% in the first year and over 150% in the second year. We will conduct a data exploratory analysis to see if traders' returns are related to the market returns, as research tells us that even if the traders present a long bias when the market goes up, their returns tend to lag.

In this analysis, we will also assess whether these traders generally achieve positive or negative returns and if there is a relation between their returns and the annual market returns for Bitcoin and Ethereum. The analysis will include all trades within the specified time frame and will separately evaluate performance in BTC and ETH markets.

3.1.3. Is there a correlation between traders' performance and the use of leverage?

This analysis aims to explore the relationship between traders' performance and the use of leverage. Past research indicates that leveraged trading can lead to more volatile returns (Baig et al., 2022), given that this platform allows up to 100x leverage, it is crucial to investigate how increased leverage impacts trading performance. These are also more likely to open positions when the underlying stock volatility is high, and the use of leverage can be a factor to consider in their overall performance (Barber et al., 2023). However, when the underlying stock volatility is high, we assume that this would be bad for traders, but we would like to test the hypothesis of traders being able to both return a profit or a loss from the use of leverage, as big swings in high volatility moments can lead to returns going both ways.

For this analysis, we will test the hypothesis of the leverage amount being correlated with traders' returns using a Pearson correlation test, as shown above.

3.1.4. Is there a correlation between trading frequency and traders' cumulative returns?

This analysis will investigate the relationship between trading frequency and trader returns over time. Specifically, it will examine whether there is a correlation between the total number of trades executed by each trader across all markets and their overall returns. By analyzing this data, the goal is to determine if a higher trading frequency correlates with better or worse returns, thereby addressing the hypothesis proposed in past research that trading less frequently can lead to better returns.

Although past research tells us that trading frequency can have a negative correlation with traders' returns (Barber et al., 2014), we will try to assess if this is true for our data, or if more factors are hidden in the trading frequency, like the fees being paid and taxable events, that damage trader's returns.

To accomplish this, a Pearson Correlation test will be conducted between the total number of trades of each trader and their cumulative returns, to assess if there is a correlation between how frequently they trade and their cumulative returns.

3.1.5. How big is the impact of trading fees on the overall performance of retail traders?

As a continuation of the previous analysis, this research will assess if trading fees have an impact on the overall performance of traders. Given that the GMX platform charges high fees for borrowing leverage to trade, the analysis will assess if there is an impact on the overall trader's performance when trades take more time in their trades, as in this platform fees are paid in function on the time that a trader borrows the leverage. To evaluate this, the analysis will focus on the average span of blocks of each trader along the main quintiles (middle, and tails).

The bigger focus will be on the difference between the best and worst trades (opposite quintiles), as we will try to investigate if the overall returns of these traders are positive or negative, even with the market growing 156% during the year 2023, we will benchmark those results against 2022 and see if there is difficulty in traders to return a profit due to trading fees, that take a lot from their performance, even if they manage to be right about the directionality of the market.

3.1.6. Is it possible to develop a consistent liquidity provision approach?

As suggested in past research, most of the profitable trades by institutional investors come from liquidity provision positions, as these institutions are rewarded with fees to be the middleman making sure there is enough supply to buy and sell. In a decentralised system, there is no need for financial institutions to play this role, and the liquidity provision approach on the GMX platform is proof of that. In this work, we will delve into the profitability of developing a delta hedge liquidity provision approach that aims to reward retail traders with low volatility. The volatility (σ) of returns that will be tracked is given by the square root of the variance of its returns as given in by Markowitz (1952):

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(r^t - \mu)^2} \tag{3.6}$$

where:

· r^t is the return of the asset at time t.

· μ is the mean return over the period.

· N is the total number of returns in the period.

The goal is to demonstrate that there are interesting investment opportunities in the realm of decentralised finance that should be further explored.

3.2. Data Extraction

To be able to perform the intended analysis, a data extraction process using the principals of event sourcing architecture was developed to extract onchain data to a SQL database using Python. Event Sourcing is an architectural pattern where the state of an application is derived by recording all changes to its state as a sequence of events, in this case, similar to the architecture of a blockchain, where events (transactions) change the state of the blockchain. In this case, Python fetches all relevant events from the GMX position manager smart contract: Increase Position, Decrease Position, Update Position, Close Position, and Liquidate Position. These events capture each state transition and are persisted in an event log to a database on the SQL local server. By replaying these events, the blockchain state can be reconstructed at any block number, enabling traceability, and the ability to derive different views of the data. This approach facilitates scalability, and the ability to be efficient in retrieving all relevant blockchain events and transactions for further analysis while avoiding complex logic implementations that would require superior infrastructures.

To access blockchain data, the most cost-efficient way (as it is free) is to use an RPC node endpoint that is provided by web3 data companies like Infura and Alchemy. These RPC nodes can respond to requests for blockchain data regarding blocks, transactions, smart contracts executions and others. This work will deal with a good amount of data, for it to be cost-effective, we will need to fetch blockchain data and properly store it for it to be used for the data analysis part, otherwise, performing each of those analyses using the RPC node requests would be both inefficient and slow. The event sourcing architecture implemented helped to make the data extraction process take around 20 minutes to retrieve all necessary data and store it on the database while using a free RPC node from Infura.

The data extraction process only extracted events from the smart contract transactions on the blockchain, a further process was developed to aggregate all transactions as positions. This meant that this process would help with the analysis since most of it is based on the positions that retail traders took and their underlying performances. The result from this second process was also stored on a SQL server.

3.3. Data Exploratory Analysis

The data extracted for this work comes from all the trades that took place between the 1st of January 2022 and the 31st of December 2023, on the GMX platform. The dataset encompasses all transactions from block number 4,221,290 to block number 165,788,864 on the Arbitrum blockchain. After developing a process to efficiently fetch this data from the blockchain, the analysis will be conducted on data regarding 5 types of events: Increase Positions, Decrease Positions, Update Positions, Close Positions, and Liquidate Positions.

The distribution of events during this time frame is the following:

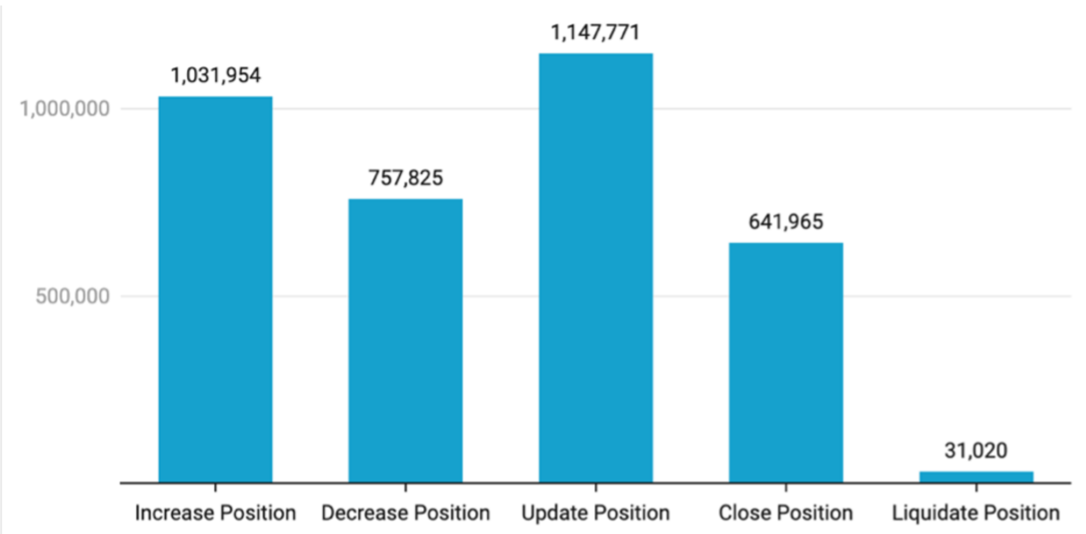


FIGURE 5. Number of database records distributed by event type

(1) **Increase Position Events**: these events relate to actions/trades of increasing the size of a position. These events can also be interpreted as opening a trade or increasing the size of a position that is already open. As an example, this can be opening a 20 dollars trade, or increasing the size of our current position with 10 more dollars.
(2) **Decrease Position Events**: events are the opposite of the above, as they relate to actions of decreasing the position size. These events can also be interpreted as decreasing a part of the total amount of a position.
(3) **Update Position Events**: events related to updates of the position like increasing or decreasing the trade at a new price. Increase and Decrease events keep track of the deltas to the position, as update events maintain the state of the position.
(4) **Close Position Events**: these events track closed position values. These events carry the Profit and Loss (PnL) of a position, as well as the final price at which it was closed.
(5) **Liquidate Position Events**: events appear in the case of liquidations of some positions and like close position events, keep track of the final values of the position once liquidated.

Although these trades that are fetched from the Arbitrum blockchain appear as individual events, the second process was developed to provide an aggregated view for each position. This would help as the intended analysis would be to look into the positions taken by traders, and not their single trades. The final dataset is comprised of 127,582 different wallets, for a total of 674,917 positions. Similar to a normal wallet, a cryptocurrency wallet holds the tokens and executes transactions on a blockchain. Each user may possess multiple wallets within each blockchain, however, each wallet address is unique and cannot be replicated by another individual transacting assets within their wallet with another private key that allows access to the specific wallet.

The aggregated information from all these events regarding a position will provide the full data about the open and close timestamps, return of the position, total collateral deposited in the position, amount of leverage, and other types of information like liquidations. Following the results from this process, the distribution of positions by asset shows that traders have a bigger tendency, not only to trade Ethereum more often than Bitcoin but also have more total volume.

A long position means someone bought the asset as they believe the price will go up and they will return a profit out of that price movement. The short position is the opposite, however, a trader must borrow assets in order to short those same assets and earn a profit if the price of the underlying asset trades lower than the price the short position was opened. In the GMX platform, a Bitcoin long and a Bitcoin short are considered to be two different markets, as these markets depend on the type of position and the underlying asset being traded.

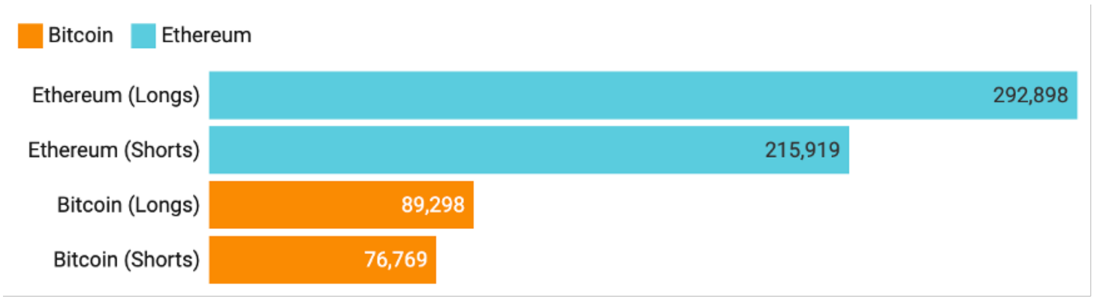


FIGURE 6. Number of positions by token and market type Long/Short

The total Bitcoin traded volume in this period was 15.91 billion dollars, while Ethereum's trading volume was almost over double at 37.67 billion dollars, for a total of 53.59 billion dollars of traded volume between the two cryptocurrencies.

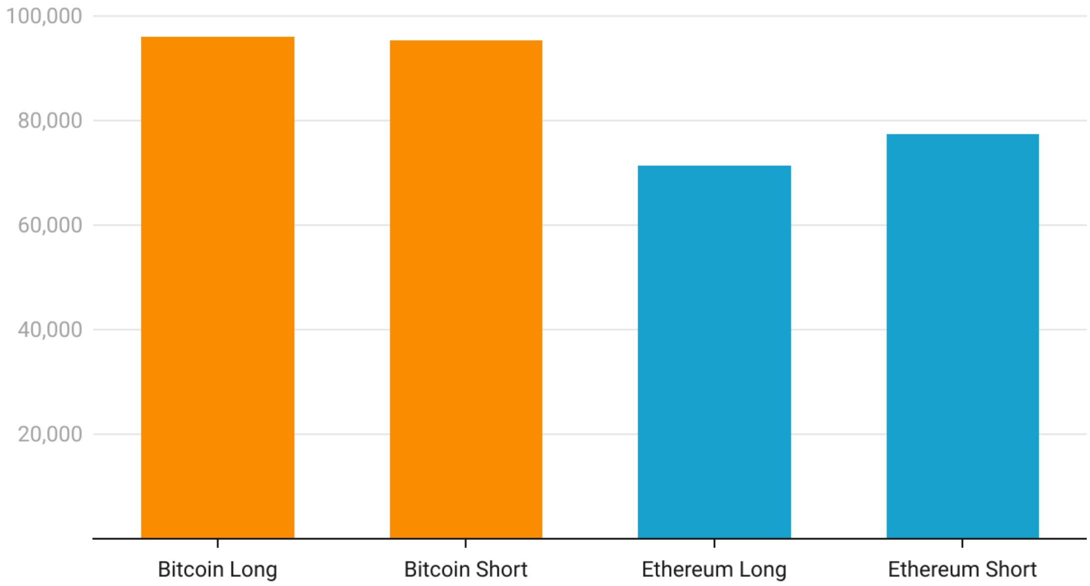


FIGURE 7. Average Position Size by Market

Although a bigger traded volume, the average position size on Ethereum trades was around 18% lower than Bitcoin's. Given the distribution of returns by token and by the position directionality (long or short), the total return of traders during this time frame was 8,805,382.48 dollars. This represents a return of around 0.0164% of the total traded volume.

Taking a deeper look into the numbers, it is easy to spot that Bitcoin Long positions were the only ones that returned a profit when we account for the return of all the positions taken in Bitcoin and Ethereum.

As Table 1 shows, Bitcoin Longs were the only ones profitable in this time-frame. Looking into the data, it is also possible to see that this was also the market with the smaller biggest nominal losing trade at -$995,134.07. The market with the biggest nominal

Market	Total P&L	Total Positions Value	Total Positions Return
Bitcoin Shorts	-$1,707,815.02	$7,325,127,875.90	-0.02%
Bitcoin Longs	$28,817,201.71	$8,585,671,483.73	0.34%
Ethereum Shorts	-$1,592,653.45	$16,730,846,841.86	-0.01%
Ethereum Longs	-$16,711,350.77	$20,936,460,654.11	-0.08%

TABLE 1. Performance by market.

return trade was Ethereum Longs with a return of $4,238,612.33. Analyzing the total liquidations by market, it becomes evident that the worst-performing markets also have the biggest number of liquidation events.

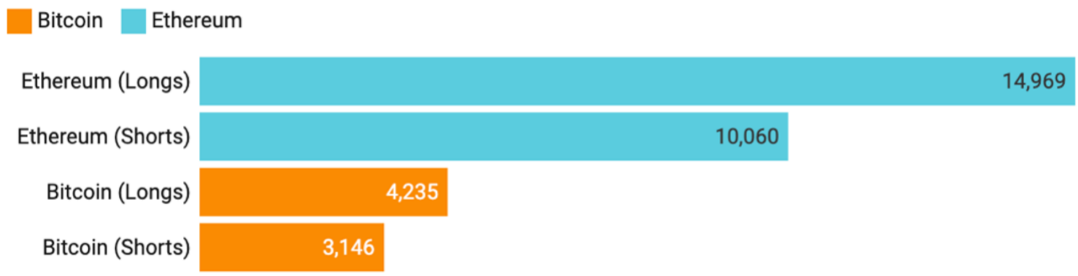


FIGURE 8. Number of liquidation events by market

In percentage of total trades, Ethereum Longs had the biggest percentage of liquidated positions at around 5.11% of the total number of positions in this market. Bitcoin Shorts had the best ratio of liquidated positions per total of positions at around 2.7%.

The relation between leverage and returns has an interesting expression in the data, which is seen as an inverse relation between the amount of leverage and the average performance of the positions.

Market	Average Position Leverage
Bitcoin Shorts	22.88x
Bitcoin Longs	24.17
Ethereum Shorts	19.23x
Ethereum Longs	17.7x

TABLE 2. Average position leverage to the amount of collateral provided in positions

Previous analysis showed that Ethereum Longs are the ones having the worst performance in this time frame, however, these are also the ones having the lowest average leverage on their positions at 17.46x. Opposed to them, Bitcoin Longs, the ones who got the best performance, are the ones with the biggest average at 24.17x. Although this data points to an inverse relation between leverage use and traders' returns, a hypothesis will be further tested on this topic. With the current data, it would be wrong to assume a possible relation between the amount of leverage used by traders and their returns.

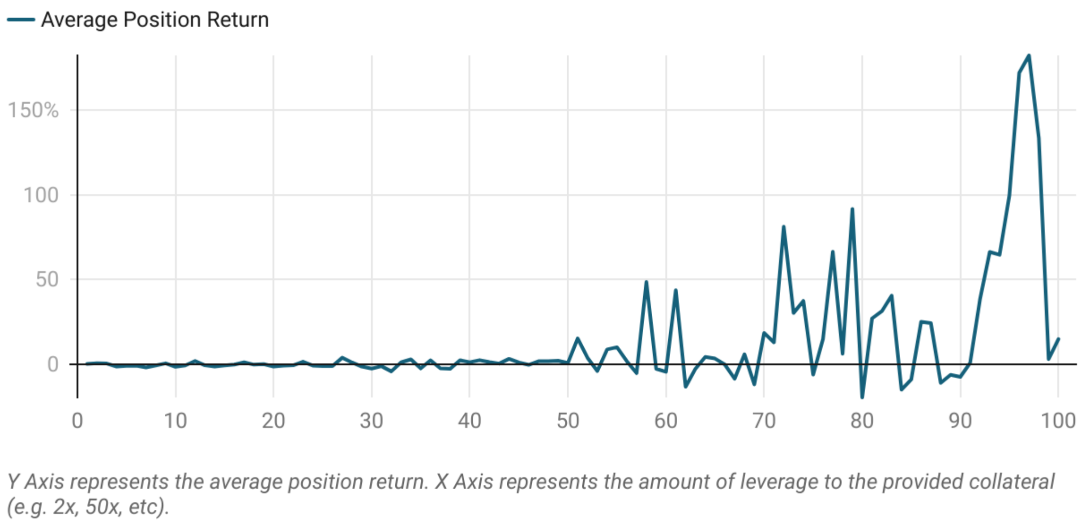


FIGURE 9. Average Position Return in Function of Leverage

As seen in the figure above, when the position leverage stays below 50x, the average return of the positions looks relatively more stable when compared to the positions with leverage usage above 50x. A relation can't be drawn this easily, because the number of positions below 50x leverage is far greater at 653,800, compared to the 21,089 positions that had higher leverage than 50x. This happens because the GMX interface allows users to have a leverage scroll button to use up to 50x leverage. To use a higher leverage multiple, users need to do it algorithmically through a smart contract. This makes it harder for retail traders to have access to higher leverage.

In the next chapter, we will test the prior proposed set of analyses and hypotheses, as well as look for relations between these traders' performances, to see if it is possible to explain traders' returns at some level by those variables, or if their lack of knowledge makes it difficult to trace patterns. At the end of the chapter, this work will look deeper into a liquidity provision approach on the GMX protocol, where it will use the information gathered in the hypothesis testing and empirical analysis to propose an investment strategy with low volatility and interesting risk-adjusted returns.

CHAPTER 4

Empirical Analysis and Hypothesis Testing

To test the possibility of profiting from retail traders, there is the need to first understand how they are distributed, how they behave and what influences their returns.

4.1. How are the best and worst-performing traders distributed?

First, an approach to select the best and worst-performing traders was put in place for further analysis. This approach is similar to the one used by (Barber et al., 2023). In this work, traders were divided into quintiles where the middle one corresponds to traders who had a cumulative return over the period close to the average and the other four quintiles correspond to one and two standard deviations under and above the average.

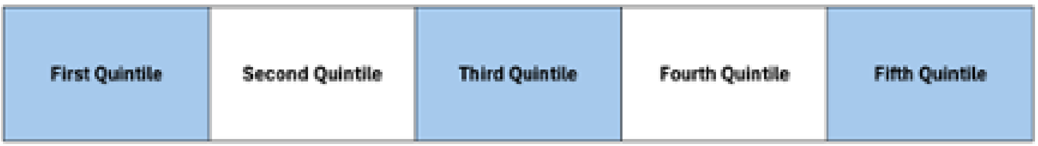


Figure 10. Illustration of Quintile Partitioning

Taking into account the above representation, if the cumulative returns of each trader are adjusted by their weight in the total traded volume, the following average trading returns from each quintile are achieved:

Quintile	Cumulative Return	Number of Traders
Q1	-97.19%	7,769
Q2	-57.13%	4,485
Q3	-0.64%	108,038
Q4	46.72%	3,482
Q5	142.03%	2,045

Table 3. Distribution of traders for the 5 quintiles

Compared to other quintiles, the middle one represents the average of the majority of the traders, with 85.87% of all traders tracked in this dataset. The average return of all of these traders sits around -0.64% for a total of 108,038 traders.

As represented in the figure above, the distribution of traders who had cumulative returns between -1% and 1% represents 57.42% of all traders in the dataset and accounts for more than 97.5% of the traded volume in this quintile. This indicates that most of

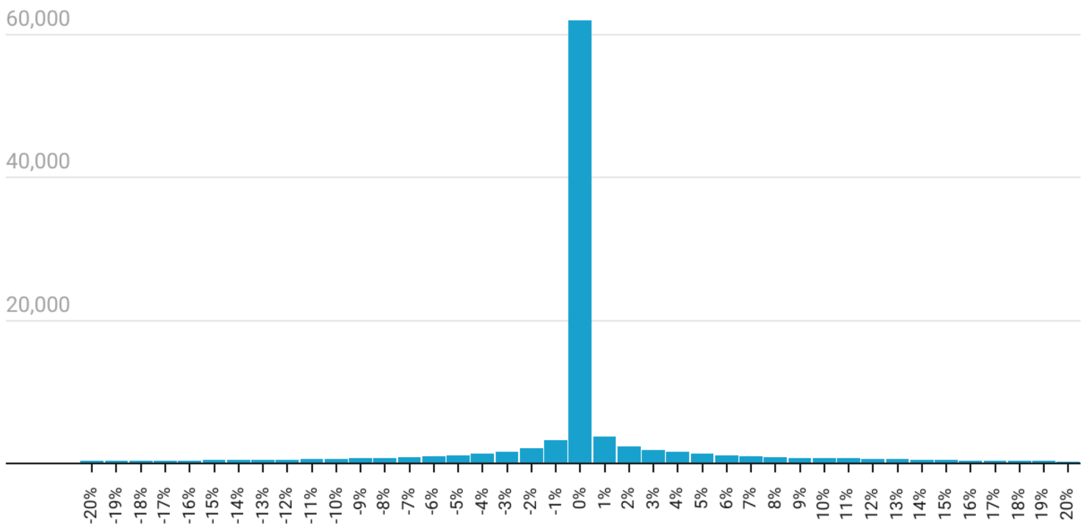


FIGURE 11. Distribution of cumulative returns by the traders in the centre quintile with returns between -20% and 20%

the traders on this platform end up with returns close to 0% over the 2 years' time.

Quintile	Total Traded Volume
Q1	\$170M
Q2	\$538M
Q3	\$55,000M
Q4	\$400M
Q5	\$289M

TABLE 4. Traded Volume by quintile

As expected, there is evidence of a difference in returns and trading volumes between the traders that fall in the tails of the distribution and the ones closer to the average quintile. Data shows that, although traders in one of the tails of the distribution present big differences when compared to the traders that fall in the middle quintile, the characteristics of the traders in opposite tails of the distribution seem more alike than expected, as further analysis will demonstrate.

4.2. Is there a correlation between the imbalance in open interest and market returns?

Past research suggests a short-term correlation between market returns and order imbalance. To investigate whether similar trends can be identified in the dataset, an analysis was conducted to determine whether the imbalance in the daily open interest is correlated with daily market returns. To accomplish this, all increase position events were aggregated by day.

These values and the daily market returns, first for Bitcoin and secondly for Ethereum, were used for the Pearson Correlation test. The Pearson coefficient for all Bitcoin trades was -0.0777, and the p-value was 0.03575. These results indicated that there is a small negative linear correlation between the daily open interest imbalance and market daily returns for Bitcoin. For the Ethereum traders, the Pearson correlation coefficient was -0.03077, and the p-value was 0.40655. The p-value was higher than the significance level of 0.05. This result led us to fail to reject the null hypothesis.

After performing the statistical analysis, the results indicated that in the Bitcoin daily open orders imbalance there is a small inverse correlation with market returns. The fact that this correlation is weaker than the presented in past research, might be due to various factors, such as differences in market conditions, data sample characteristics, or other external influences. One possible factor that can cause the different results is the fact that cryptocurrencies behave differently for other more mature markets, as their underlying volatility is considerably higher, making movements in the price happen faster than in traditional markets.

4.3. How do these traders perform over time in comparison with the market?

As shown in Table 3, traders that fall in the middle quintile, have an average cumulative return very close to 0 at -0.64%. If we adjust this number to the weight each trader had on the total traded volume of this quintile, the weighted average return comes to 0.68%.

If all traders in one of the tails of the distribution are aggregated, their average cumulative returns are -41.87% and if adjusted for the total traded volume, the weighted average return is -7.32%. Even though the traders in the fifth quintile have traded more volume, if the cumulative returns of both tails' traders are adjusted for volume, the total return of these two quintiles is still negative. It is relevant to state that, in this time frame, Bitcoin returned -9.06% and Ethereum returned -40.41%.

Asset	2022	2023	Total
Bitcoin	-62.02%	156%	-9.06%
Ethereum	-67.7%	84.44%	-40.41%
Tail Traders	-63.66%	13.4%	-7.32%
Mid Traders	-0.38%	2.48%	0.68%

Table 5. Comparison table with returns by year of market and traders (The returns for tail and mid traders are adjusted for volume weighted average returns)

As highlighted in the table above, the aggregate of the traders in the tails of the distribution show that these are more likely to underperform the market when compared to traders in the middle quintile. Having worse returns both in 2022 and even in 2023 when the market rallied upwards of 80% for Ethereum and 150% for Bitcoin due to the announcements of spot ETFs approvals and Bitcoins' halving (Forbes, 2024b).

4.4. Is there a correlation between traders' performance and the use of leverage?

Data shows that the average collateral of positions taken by traders on the first and fifth quintiles is much smaller than the traders on the middle quintile. The average position opened by retail traders in the tails of the distribution is 14 times smaller at $2,898.79, compared to traders in the middle quintile had average position collaterals of $40,243.5.

One of the testing hypotheses for traders' bad performance was the use of leverage. It was tested for all trades taken on the Decentralised Perpetuals Exchange between 2022 and 2023, there is a weak correlation between the leverage amount used and the trade return. The estimate for the Pearson coefficient was 0.0034 with a p-value of 0.01725. This means that there is no linear relation between the use of leverage and trade profitability. This correlation test was also done for two subsets of trades: Bitcoin trades and Ethereum trades. Both presented the same results of weak linear relation between the amount of leverage and the profitability of the trade.

Looking deeper into the quintiles, data shows that although there is a weak correlation between leverage and returns, the average leverage amount for tail traders was 28% higher at 13.87x the amount of collateral, in comparison with the traders in the third quintile who had an average leverage of 10.91x the amount of collateral. Even though these values might hint at a possible relation between leverage and volatility of the returns and similarities between traders in both extremities of the distribution, further analysis of the subsequent hypothesis will help better understand these two.

4.5. Is there a correlation between trading frequency and traders' cumulative returns?

As stated by past research by Barber et al. (2014), retail traders are likely to open and close trades more often than professional investors. To test this hypothesis of trades who trade more, tend to have worse returns, data regarding the total amount of trades by each trader was computed to learn more about a possible correlation. The Pearson Correlation test indicated a coefficient of 0.0146 and a p-value of 0.00000023. This means that there is a weak linear relation between the use of leverage and trade profitability. This correlation test was also done for two subsets of trades: Bitcoin trades and Ethereum trades. Both presented the same results as prior of a weak linear relation between the number of trades and the profitability of the trader.

Looking into the distribution by quintiles, data also shows that tail quintiles don't trade as often. The average number of trades by a trader in the first or fifth quintile is around 1.25, while traders in the third quintile have an average number of trades at 5.92.

Although the relation between the trading frequency and traders' returns seems to be weak and does not fully explain why the cumulative returns of traders in the tails lag the market, one possible explanation for this is trading fees. As the next analysis will show, these can have a greater impact than what might have been anticipated.

4.6. How big is the impact of trading fees on the overall performance of retail traders?

Past research attributes the behaviour of over-trading to sensation seeking, similar to the one people get when they gamble. For tail traders, something similar might be occurring: these traders don't trade a lot, mostly because, like in a game of blackjack at a casino, they went all-in into one of the sides, picking Long or Short, and sticking to the trade until they got something out of it, even if it meant losing all the money (liquidation). The ride-or-die analysis comes from the data regarding the average span of blocks (a measure of time in the blockchain) that each trader averages in their trades. When it comes to tails traders, this average tends to be 10x bigger than the traders on the other quintiles of the distribution.

Quintile	Average Number of Blocks
Q1	6,620,802
Q2	3,075,327
Q3	610,131
Q4	7,353,445
Q5	15,673,024

Table 6. Average number of blocks that a trade span

On the GMX platform trading fees are especially high, as borrowing from the liquidity providers means that these second ones need to be compensated for lending the assets and not capturing the potential returns.

As tail traders tend to stick longer to their trades with a median time of around 5 days per trade when compared to 12 minutes by the middle quintile traders, the costs of trading fees tend to pile up and take a lot from their returns. Not only do these tail traders tend to stick longer to their trades, but their average amount of leverage is 28% higher than traders in the middle quintile, making the borrowing fees also higher.

In past research, there was evidence that, although retail traders are not that bad at picking the direction of the mark (even if a long bias is presented by these traders), even if they manage to pick the winning side, trading fees will take a lot from their returns. Picking all the tail traders by year, for all trades they did during 2022, when the cryptocurrency market took a big hit, with Bitcoin returning -62.02% and Ethereum returning -67.7%, retail traders on the tails lost a staggering -63.66% when adjusted for the volume of each trade. During 2023, while the cryptocurrency market went on another rally, with Bitcoin returning 156% and Ethereum returning 84.44%, the cumulative returns of traders on the tails were only 13.4%.

During the two years of data, even though around 60% of all trades and 56.5% of all trading volume were Longs, the average leverage amount on Shorts was slightly higher at 20.82x, while Longs had an average leverage of 19.55x. Even though there is evidence of a long bias from the traders, meaning that in theory, they should had a better performance. If we take the following assumptions for a given trade:

· The average leverage multiple of 13.8x from tails traders,

· The median amount of time for a trade from the tail traders,

· An average pool utilization rate of 50%.

An estimation can be reached that, for one trade, 8.28% of the return gets “eaten” by trading fees of the platform (not accounting for other fees like gas fees to execute the operations on the blockchain).

When it comes to the rest of the traders that take the other 99% of the volume, these are not so “interesting”, with well-distributed returns that also tend to be skewed relative to the market, that is why we see the small difference from the performance in 2022 and 2023. These are relatively better at picking the “correct” side and do not pay as much fees as the tails traders who stick to their traders for longer. However, the majority of their traders will still end up with returns close to 0%.

4.7. Is it possible to develop a consistent liquidity provision approach?

As seen in the previous chapters, retail traders have a hard time making a consistent profit on their trades. Either because they over-trade or pay obscene fees to be able to maintain their margin for the high-leverage amounts that they use. Although not very interesting on the traders' side, like a casino, the house benefits from this type of environment, especially when trading assets in not-so-mature or developed markets like cryptocurrencies. In traditional markets, being the house (or the market maker) is not so easy as one needs to be a financial institution with enough capital to deploy to cover the cost of a market-making operation, however, in an environment that aims for decentralisation, getting to be the seat is easier, more affordable, and incentivised.

To save the trouble of having liquidity to provide their traders, decentralised platforms like GMX tend to have liquidity pools where liquidity providers deposit their cryptocurrencies and receive a big slice from the fees paid by the traders to borrow leverage. On this platform, liquidity providers are lending their digital assets to traders in exchange for a fee that they believe to compensate for the impermanent loss that they have. This impermanent loss is the effect of not getting the asset return because they are lending both the asset and their underlying return to the traders. The concept of impermanent loss is familiar to market makers across all markets since this is the risk they take to earn the fees, as they believe this is a good risk/rewards ratio to have.

Although the digital assets market is considered to be highly volatile, Decentralised Perpetuals Exchange compensates the liquidity providers by charging high borrowing fees to the traders. As explained before, the slice of the liquidity pool where liquidity providers deposit their tokens is distributed in the form of a token called GLP. This token not only accrues the return of the assets on the pool but also the fees paid by traders to lend those assets, which afterwards are distributed *pro-rata* to the liquidity providers. The rise in the value of the assets that compose the liquidity pool is reflected in the price of the GLP token, as this represents the shares that the liquidity providers own of that pool, and the borrowing fees are distributed *pro-rata* to each LP.

The question begs how good liquidity provisioning in this decentralised market is, and if it is a worthwhile approach to capture the high fees to make a profit out of the retail traders. To answer this question, the first step is to analyze the past returns of these liquidity providers.

The following equation gives the price of the GLP token (P_{GLP}):

$$P_{GLP} = \frac{V_{pool}}{N_{GLP}}, \tag{4.1}$$

where:

· V_{pool} · is the dollar value of all assets deposited in the pool.

· N_{GLP} is the number of outstanding GLP tokens.

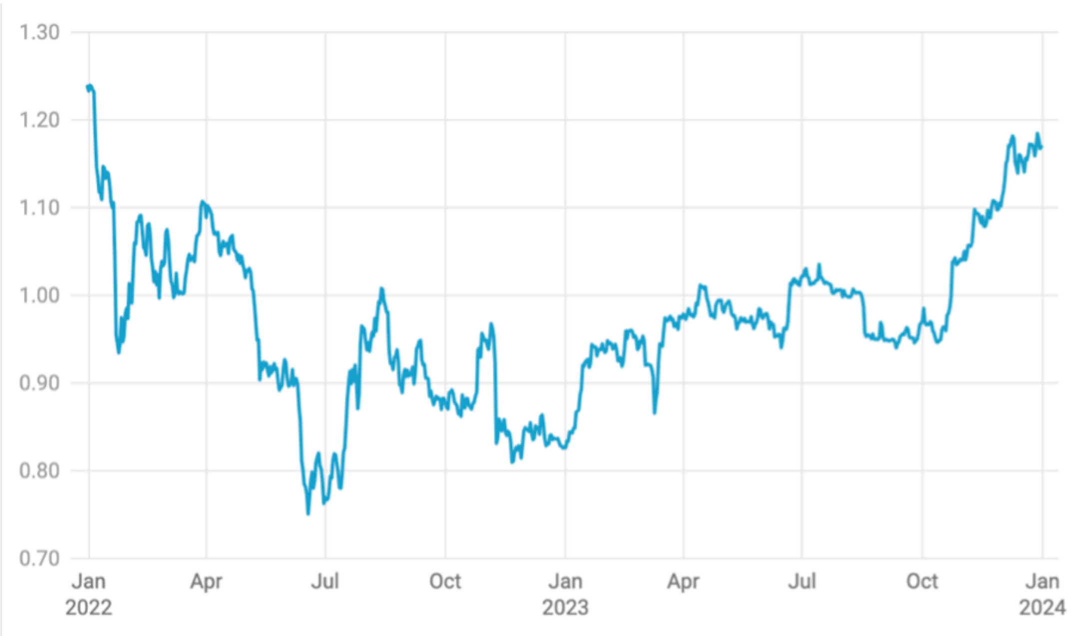


Figure 12. GLP Price from Jan 2022 to Dec 2023

The GLP token has returned – 5.55% during the 2 years with a volatility of 25.35%, which is given by the standard deviation of the daily returns. During 2022, while both Bitcoin and Ethereum went down around 60%, the GLP token returned -23.41%, while in 2023, with both reference assets returning 156% and 64%, respectively, the GLP token went up 41.39%, however, the fees that were distributed to LPs are not being accounted for the return of the GLP token. If a liquidity provider had kept all fees, it would have earned 38.66% over the 2 years, on top of the GLP return. This breaks down into:

· 23.28% in 2022 · 15.37% in 2023

Let´s say one investor does not like the idea of being exposed to the delta of the underlying that would compose the GLP token, but he still wants to capture the fees paid by liquidity providers. As explained before, at all times, around 50% of the GLP pool composition remains in stablecoins. This means that the liquidity provider would need to hedge the other 50% of the exposer. Having a delta hedge approach of half of the exposer means the liquidity provider would only get two-thirds of the fees on his capital since one-third of the capital needs to be used to hedge the other third that is exposed to the delta of non-stable assets that compose the GLP pool.

In conclusion, this would mean that a liquidity provider could manage to get 15.5% (two-thirds of the 23.28%) in 2022 and 10.25% (two-thirds of the 15.37%) in 2023, while being delta-hedged. A compounded return of 27.34% over the 2 years would have a very low volatility of 1.044% for the liquidity provider while managing to capture on his portfolio, the brilliancy of all retail traders using this platform over the 2 years.

CHAPTER 5

Conclusion

The ease of access to financial markets via online trading platforms, which offer high leverage, has made retail trading a phenomenon in recent years. From meme stocks to cryptocurrencies, retail traders, using complex derivatives products, often struggle to make a profit.

In his book, Housel (2020) states that one cannot cheat the markets and only capture the returns on the right side of the distribution, without "paying" the risk of getting the returns from the left side, otherwise, investing like this would be to capture arbitrage opportunities. This serves as a warning to the average retail trader who often views financial markets through biased lenses, seeking opportunities to earn money with minimal to no effort.

According to the data analyzed in this work, in the emergent market of cryptocurrencies, it seems that the overall behaviour of traders when it comes to the outcomes of their trades, does not differ much, although there were some differences in what influences their returns the most. During a 2-year time frame, around 60% of the traders presented returns very close to 0. While trying to understand if there was some linear correlation between trade variables that could explain the returns of the traders, outcomes showed weak correlations, as indicated by low Pearson coefficients.

One of the hypothesis tests was on the correlation between the daily open interest and the daily market returns, where the main goal was to see if there was any relation between those two that could explain the "why?" or "when?" retail traders are opening trades. The results of this test indicated a weak Person coefficient, meaning that market returns have a very small linear relation with the imbalance in retail traders' long or short trades.

The other two tests were evaluating hypotheses raised in previous research. The first one looked for a possible linear relation between leverage and traders' returns, while the second one looked for correlations with trading frequency. Both the results showed a weak linear relation between trader variables and traders' returns. In the first test, a more interesting result was found, where most of the traders who end up in one of the tails of the distribution tend to trade with more leverage than traders in the middle quintile. In the second test, while digging deeper into the dataset, it was found that traders in the middle quintile are the ones trading more often than traders in the tails of the distribution. These weak relations occur mostly because 60% of all trades on the Decentralised Perpetuals Exchange returned between -1% and 1%, resulting in an average trader return close to zero.

The conclusions of these analyses allowed us to understand that the behaviour of retail traders is in line with some of the past research. When it comes to the behavioural biases, although the correlations were weaker, the overall distribution of traders' cumulative returns and the similarities between the first and fifth quintiles are in line with the research from Ferko et al. (2023). When it comes to the trading frequency from traders in the tails of the distribution, the main difference emerges. As our results point out traders that are liquidated, or have an outstanding return, do not trade so often after that, as the number of trades per trader in these two quintiles was comparatively lower than the middle quintile.

Despite the weak linear relation in some of the behavioural biases that previous research highlighted as influencing profitability, one clear trend emerged: traders who hold bigger positions for a longer period were more likely to end up in one of the extremes of the distribution. Those in the tails of the distribution tend to hold trades 10x longer than those in the middle quintiles. This is a big insight that helped to explain traders' behaviour.

Decentralised trading platforms are no different in allowing retail traders to use high leverage. Additionally, in the Decentralised Perpetuals Exchange, users pay high fees to maintain the borrowed leverage in their positions. This significantly influences the performance of traders over time, especially those who stick to longer positions, as this is the case for traders in the tails of the distribution, corresponding to the first and fifth quintiles.

The use of leverage helped explain why the consensus of these traders, in contrast with traders who end up closer to the average of the distribution, have a hard time picking up with the market. These traders seem to have similar returns to the market when it was going down, as in 2022, but lack this similarity and lag against the market when it is going up, like in 2023. During the market downturn of 2022, traders' returns mirrored the market at -63.66%, while Bitcoin returned -62.02% and Ethereum -67.7%. In contrast, these tail traders had a cumulative return of 13% in 2023, where they lagged the market, with Bitcoin returning over 150% and Ethereum over 80%. This disparity is mostly due to the high fees paid by leveraged traders, which even when the market is going in their favour, makes their life much more difficult when it comes to returning a profit over time. For an average trader in one of the tails, the borrowing fees paid on one trade over 5 days can cost over 8%.

Even though these decentralised trading platforms charge high fees, they offer an interesting opportunity for liquidity providers, who are fairly rewarded for lending their assets to more risk-averse investors. Liquidity providers could have outperformed the market by getting 38.66% in fees on top of the -5.55% from the GLP token return. Even with a delta hedge approach, liquidity providers would have earned an average of 12.88% per year, over two years, with an annualized volatility of 1.044%.

Further research about this topic should focus on incorporating more variables, such as the underlying asset returns before a trader opens a position, to test the hypothesis that past returns influence trade openings and sizes. Including more variables related to price performance could enhance models predicting the likelihood of a trader falling into each quintile. The use of Principal Component Analysis could provide better insights into how these new variables contribute to predicting trade outcomes and the probability of a trader landing in each quintile. Developing a model to predict quintile probabilities could inform trading strategies that replicate or counter traders, potentially creating derivative products with better performance than the GLP token.

This work showed the focus on data management in financial markets as an increasingly important skill to have, especially when working with blockchain data, as the trend seems to be that more securities will transition to be issued and settle on the blockchain. As the market matures, market-making activities will become more attractive to large financial firms seeking indirect exposure to digital assets. With the approval of Bitcoin and Ethereum ETFs, investors will soon look for better risk-adjusted exposure to this asset class. Quantitative approaches to market-making on decentralised protocols could democratize access to these activities, traditionally requiring significant capital and expertise in infrastructure development. Liquidity provision approaches, like the one presented in this work, will gather interest from more risk-averse investors and help both expand and mature the digital assets market.

Bibliography

Adams, H., Zinsmeister, N., & Robinson, D. (2020). Uniswap v2 core. https://uniswap.org/whitepaper.pdf

Baig, A. S., Blau, B. M., Butt, H. A., & Yasin, A. (2022). Do retail traders destabilize financial markets? an investigation surrounding the covid-19 pandemic. *Journal of Banking and Finance*, *144*. https://doi.org/10.1016/j.jbankfin.2022.106627

Barber, B. M., Lee, Y. T., Liu, Y. J., & Odean, T. (2014). The cross-section of speculator skill: Evidence from day trading. *Journal of Financial Markets*, *18*, 1–24. https://doi.org/10.1016/j.finmar.2013.05.006

Barber, B. M., Lin, S., & Odean, T. (2023). Resolving a paradox: Retail trades positively predict returns but are not profitable. https://doi.org/10.1017/S0022109023000601

Barber, B. M., & Odean, T. (2013). Chapter 22 - the behavior of individual investors. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-44-459406-8.00022-6

Berentsen, A., Borio, C., Claessens, S., Fornaro, D., Holden, H., Shin, H. S., Shirakami, T., Wooldridge, P., Bech, M., Hancock, J., Rice, T., & Wadsworth, A. (2020). On the future of securities settlement 1 we thank.

Buterin, V. (2014). Ethereum: A next-generation smart contract and decentralized application platform. https://ethereum.org/en/whitepaper/

Chen, E., Ma, M., & Nie, Z. (2024). Exploring the impact: How decentralized exchange designs shape traders' behavior on perpetual future contracts. https://arxiv.org/abs/2402.03953

Chen, Y., & Bellavitis, C. (2020). Blockchain disruption and decentralized finance: The rise of decentralized business models. *Journal of Business Venturing Insights*, *13*, e00151. https://doi.org/https://doi.org/10.1016/j.jbvi.2019.e00151

Curto, J. D. (2021). *Econometrics and statistics - over 100 problems (with solutions): Applications in 'r/rstudio' and 'excel'*. Independently published.

Davis, E. P., & Steil, B. (2001). *Institutional investors*.

DEX to CEX spot trade volume. (2024). https://www.theblock.co/data/decentralized-finance/dex-non-custodial/dex-to-cex-spot-trade-volume

Dorn, A. J., Dorn, D., & Sengmueller, P. (2012). Trading as gambling. http://ssrn.com/abstract=2160996

Eigen Labs Team. (2024, April). Eigenlabs github. https://github.com/Layr-Labs/whitepaper/blob/master/EIGEN_Token_Whitepaper.pdf

Ferko, A., Mixon, S., & Onur, E. (2023, October 1). Retail traders in futures markets.

Forbes. (2024a). Digial assets - crypto prices. https://www.forbes.com/digital-assets/crypto-prices/

Forbes. (2024b). How this halving will impact bitcoin. https://www.forbes.com/sites/digital-assets/2024/04/07/how-this-halving-will-impact-bitcoin/

Glaser, M., & Weber, M. (2007). Overconfidence and trading volume. *Geneva Risk and Insurance Review*, *32*, 1–36. https://doi.org/10.1007/s10713-007-0003-3

Global cryptocurrency user base. (2024). https://www.statista.com/statistics/1202503/global-cryptocurrency-user-base/

Gurrola-Pérez, P., Lin, K., Speth, B., & Gurrola-Perez, P. (2022). Wfe research retail trading: An analysis of current trends and drivers retail trading: An analysis of global trends and drivers. www.world-exchanges.orgElectroniccopyavailableat: https://ssrn.com/abstract=4562259

Haber, S., & Stornetta, S. (1991). How to time-stamp a digital document. *Journal of Cryptology*, 99–111.

Hajizadeh, E., Ardakani, H., & Shahrabi, J. (2010). Application of data mining techniques in stock markets: A survey. *Journal of Economics and International Finance*, *109-118*.

He, S., Manela, A., Ross, O., & von Wachter, V. (2022, December). Fundamentals of Perpetual Futures. https://ideas.repec.org/p/arx/papers/2212.06888.html

Housel, M. (2020). *The psychology of money: Timeless lessons on wealth, greed, and happiness*. Harriman House Publishing.

Hull, J. (2011). *Options, futures, and other derivatives*. Pearson College Div.

Jarnecic, E., & Liu, K. (2006). Trader profits by institutional and retail participants in options markets.

Lamothe-López, P., Lamothe-Fernández, P., & Rodríguez-Valencia, L. (2024, April). The inclusion of bitcoin and other cryptocurrencies in investors' portfolios. IntechOpen. https://doi.org/10.5772/intechopen.1003990

Levi, S., & Lipton, A. (2022). An introduction to smart contracts and their potential and inherent limitations. *Journal of Information Security*.

Mackintosh, P. (2020). Who counts as a retail investor? nasdaq.

Malz, A. (2021). The gamestop episode: What happened and what does it mean? *Cato Journal*.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, Vol. 7, No. 1., pp. 77-91.*

McCabe, C. (2022, January 16). Day traders as 'dumb money'? the pros are now paying attention. https://www.wsj.com/articles/fund-managers-pay-attention-to-retail-day-traders-11642132135

McKinsey & Company. (2023). *Tokenization: A digital-asset déjà vu*. McKinsey & Company.

Nakamoto, S. (2008, October 31). Bitcoin whitepaper. https://bitcoin.org/en/bitcoin-paper

Rejeb, A., Appolloni, A., Rejeb, K., Treiblmaier, H., Iranmanesh, M., & Keogh, J. G. (2023, May). The role of blockchain technology in the transition toward the circular economy: Findings from a systematic literature review. https://doi.org/10.1016/j.rcradv.2022.200126

Sager, M., & Taylor, M. (2007). Commercially available order flow data and exchange rate movements: Caveat emptor. *Journal of Money, Credit and Banking*.

Selvamuthu, D., Kumar, V., & Mishra, A. (2019). *Indian stock market prediction using artificial neural networks on tick data*. SpringerOpen.

Sguanci, C., Spatafora, R., & Vergani, A. M. (2021). Layer 2 blockchain scaling: A survey. *CoRR*, *abs/2107.10881*. https://arxiv.org/abs/2107.10881

Shah, K., Lathiya, D., Lukhi, N., Parmar, K., & Sanghvi, H. (2023). A systematic review of decentralized finance protocols. *International Journal of Intelligent Networks*, *4*, 171–181. https://doi.org/https://doi.org/10.1016/j.ijin.2023.07.002

Sregantan, N. (2023, November 28). What is the blockchain trilemma? https://www.dbs.com.sg/personal/articles/nav/investing/what-is-the-blockchain-trilemma#

Sriman, B., Kumar, S., & Shamili, P. (2020). Blockchain technology: Consensus protocol proof of work and proof of stake. *Intelligent Computing and Applications*, 395–406.

Sunyaev, A., Kannengießer, N., Beck, R., Treiblmaier, H., Lacity, M., Kranz, J., & Luckow, A. (2021). Token economy. *Business & Information Systems Engineering*, 457–478.

Swan, M. (2015). Blockchain: Blueprint for a new economy. *O'Reilly*.

Swan, M. (2017). Anticipating the economic benefits of blockchain. *Technology innovation management review*, *7*(10), 6–13.

Talhartit, I., Jillali, S. A., Kabbouri, M. E., & Kabbouri, M. E. (2022). Behavioral finance: How are traders' financial decisions and performance impacted by behavioral biases under uncertainty? *International Journal of Sciences Academic Research*, *2022*, 4588–4597. http://www.scienceijsar.com

Taraxa Foundation. (2022). Taraxa architecture. https://docs.taraxa.io/tech-whitepaper/taraxa-architecture

Thulasiraman, K., & Swamy, M. (1992). 5.7 acyclic directed graphs. *Graphs: theory and algorithms*, *118*.

TickData. (2023). Tickdata website. https://www.tickdata.com/

United States Federal Reserve. (2024, June 7). Balance sheet of households and nonprofit organizations. https://www.federalreserve.gov/releases/z1/dataviz/z1/balance_sheet/chart/#series:financial-assets;units:usd

Zetzsche, D., Arner, D., & Buckley, R. (2020). Decentralized finance. *Journal of Financial Regulation*, 172–203.