

Blockchain Analysis of the Bitcoin Market

Igor Makarov¹ and Antoinette Schoar^{*2}

¹London School of Economics

²MIT Sloan, NBER and CEPR

May 10, 2026

Abstract

This paper presents a comprehensive analysis of the Bitcoin ecosystem using blockchain data from 2009-2025, examining transaction patterns, mining concentration, and ownership distribution. Our findings show that the Bitcoin network has become increasingly dominated by large, concentrated players, contradicting its original decentralized vision. The top 50 miners control nearly 50% of network power, while the top 0.01% of individual investors hold approximately 25% of circulating supply. Exchanges and trading-related entities account for 70% of all blockchain activity, suggesting that Bitcoin functions primarily as a traded financial asset rather than a medium of exchange.

JEL classification: G12, G15, F38

*Igor Makarov: Houghton Street, London WC2A 2AE, UK. Email: i.makarov@lse.ac.uk. Antoinette Schoar: 62-638, 100 Main Street, Cambridge MA 02138, USA. Email: aschoar@mit.edu. We thank Lucian Bebchuk, Anna Cieslak, Daniel Ferreira, Allen Ferrell, Campbell Harvey, Anil Kashyap, Christine Parlour, Jon Steinsson, Harald Uhlig as well as participants at AQR seminar, the CB&DC seminar at the Bank of Canada, CFTC seminar, Chicago Money and Banking seminar, Crypto Currency Panel at Princeton University, Harvard Law and Economics seminar, LSE finance lunch, McGill finance seminar, MIT finance lunch, NBER Monetary Economics Fall Meeting 2021, Neapolis University, Q Group Spring seminar Sussex University Finance and Stochastics Seminar, UBC finance seminar, UNC Decentralized Finance (DeFi) Conference and especially our discussant Hyun Song Shin. We thank Jiageng Liu and Ziyi Zeng for excellent research assistance. We also thank Kyrlyo Chykhrazdze for very helpful comments and help with the Crystal Blockchain data. We thank Flipside, Nansen, and Arkham Intelligence for providing access to their data.

Cryptocurrencies have experienced explosive growth in value and public attention since Bitcoin’s inception in 2009, sparking intense debate about their role in the global economy. Bitcoin, as the first and still the largest cryptocurrency by market capitalization, has become a barometer for the entire digital asset ecosystem. Yet despite widespread adoption and regulatory attention, many fundamental questions about Bitcoin’s usage, network structure and largest participants remain unanswered. Understanding these dynamics is essential for policymakers, regulators, and market participants as they seek to integrate cryptocurrencies into the mainstream financial sector.

To address these questions, we develop new methodologies that leverage the pseudonymous nature of the Bitcoin data. Specifically, we conduct three major pieces of analysis that focus on the core functions of the Bitcoin architecture: We document the transaction volume and network structure of the main participants, the concentration of miners which ensure the integrity of the blockchain ledger, and finally, the ownership concentration and trading behavior of the largest individual holders of Bitcoin, the so called “whales”. Our approach involves several methodological contributions. First, we create algorithms to filter out the transitory volume associated with peeling chains to obtain a more accurate estimate of bitcoin flows between distinct economic entities on the blockchain. Second, we develop techniques to trace the distribution of mining rewards from pools to individual miners, which provides the first comprehensive analysis of mining capacity of individual miners. Finally, we design methods to identify wallets of large individual investors and track their ownership over time, which enables us to measure ownership concentration among private holders and the dynamics of their entry and exit behavior.

Our analysis shows that the Bitcoin ecosystem has become increasingly dominated by large and concentrated players. Thus, the reality of the Bitcoin network and the composition of the central players in the ecosystem differ substantially from the idealized, decentralized model that had been envisioned by its creator, Satoshi Nakamoto, and early crypto enthusiasts.

Transaction Volume and Network Structure. We show that 90% of Bitcoin transaction volume consists of within-entity transfers for wallet management and obfuscation purposes rather than genuine economic activity. We develop algorithms to filter out this spurious volume and trace economically meaningful payments between entities on the Bitcoin network. Among the remaining 10% of real volume, trading venues (such as exchanges, ETFs, investment funds, and probable trading venues) account for approximately 70% of all activity, while illegal transactions, gambling, and merchant payments collectively account for less than 3% of total volume. These results highlight the central role of trading venues in the Bitcoin ecosystem and underscore the

predominant use of Bitcoin as a traded financial asset rather than a payment medium.

Despite Bitcoin’s dramatic price appreciation over the past decade, on-chain volume denominated in BTC has remained remarkably stable, and the composition of transaction types has been relatively consistent throughout the last decade.

Trading venues not only generate the most volume, but they are also the most connected nodes on the blockchain. The high degree of interconnectedness has important implications for transparency of the system and implementation of Know-Your-Customer (KYC) and Anti-Money Laundering (AML) regulations. We show that on the one hand, such connectivity could enhance the transparency of the Bitcoin network if all major entities comply with regulations like the European Union’s Markets in Crypto-Assets (MiCA) regulation, which introduced comprehensive reporting requirements for crypto-asset service provider.

On the other hand, this same interconnectedness could facilitate the movement of tainted bitcoins (those associated with illicit activities) if even a few entities operate with lax KYC standards. To highlight the challenges of enforcing KYC rules, we study the network related to Hydra Market, which was one of the largest darknet marketplaces prior to its shutdown by German authorities in coordination with U.S. law enforcement in 2022. We show that the highest volume entities interacting *directly* with Hydra Market were exchanges known at the time for having relatively weak KYC standards. Once funds reached these exchanges, they were commingled with other flows, rendering them untraceable and enabling onward transfers to a wide range of destinations across the Bitcoin ecosystem. While exchanges known for enforcing stricter KYC standards, such as Coinbase or Gemini, had limited *direct interactions* with Hydra Market, they had significantly larger transaction volumes with clusters that received large inflows from Hydra Market and were adjacent to Hydra Market in the network.

These results show that enforcing regulations on a select set of exchanges or within single jurisdictions is insufficient due to regulatory arbitrage opportunities, and the fragmented global regulatory landscape enables tainted funds to re-enter mainstream networks through sophisticated obfuscation strategies; thus enforcing KYC and AML regulations on a limited set of exchange is typically insufficient.

The current fragmentation of the global regulatory landscape creates arbitrage opportunities for malicious actors. While the tracing of flows on the Bitcoin ledger might be easier than tracking cash and bills across jurisdictions, the physical nature of cash makes it more difficult for bad actors to transport and use at scale. Consequently, as Bitcoin adoption continues to grow, the utilization of Bitcoin in the grey economy is also likely to expand—though it may face increasing competition from other cryptocurrencies that are more difficult to trace.

Composition of Bitcoin Miners. In a second major piece of analysis, we study the concentration of Bitcoin miners—the entities responsible for processing transactions, verifying blocks, and securing the blockchain. In return for this service, miners receive newly created bitcoins and transaction fees. In Bitcoin’s early years, mining was carried out by enthusiasts using personal computers. Today, it is dominated by specialized firms, some of which are publicly listed companies.

Bitcoin’s proof-of-work (PoW) protocol requires that a majority of mining power be controlled by honest participants. The original design of Satoshi Nakamoto envisioned fully decentralized mining that would not depend on a few large players. In such a world, collusion would be difficult and the failure of any single miner would have negligible consequences for network security. If, however, a single entity or a group of colluding miners were to command a majority of mining power, they could execute a so-called 51% attack—altering previously verified records on the ledger.

This decentralized ideal, however, clashes with the economics of PoW mining. By design, the probability of mining a block is proportional to the hash rate a miner devotes to the network, which creates strong incentives to pool resources and share rewards as a form of mutual insurance against the volatility of individual mining outcomes. As a result, Bitcoin mining is dominated by large mining pools, a feature that has attracted considerable public attention and concern.

Some observers downplay the risk posed by pool concentration, arguing that although pools exert substantial influence over the protocol, they do not directly control the miners who contribute hash rate. If a pool were observed engaging in malicious behavior, its constituent miners could exit and join competing pools. This argument shifts attention from pool-level concentration to the concentration of the underlying mining entities themselves.

Understanding miner-level concentration is important for several reasons. First, the power a pool operator has over individual miners depends on how easily miners can shift capacity across pools, which in turn depends on the underlying size distribution of the miners themselves. Second, if mining capacity is concentrated among a small number of large entities, a coalition of such entities could, in principle, coordinate a 51% attack. Third, concentration affects Bitcoin’s systemic risk: the more concentrated mining capacity is, the easier it becomes for an external party to influence the protocol by targeting or pressuring a small number of large miners. Finally, even absent a 51% attack, a small group of dominant miners may coordinate to protect economic rents or resist protocol upgrades that threaten their profitability.

We note that collusion at the entity level faces its own obstacles. Coordinating dozens of miners, potentially spread across multiple jurisdictions, may be more difficult

than coordinating a handful of pools. Our goal is not to argue that entity-level collusion is necessarily easier than pool-level collusion, but rather to provide the first systematic measurement of miner-level concentration, which is a prerequisite for any rigorous assessment of these risks.

Unlike information about mining pools, which is commonly available, information about individual miners is not readily accessible. We identify individual miners by tracking the distribution of mining rewards from the largest 23 mining pools to the miners that work for them. Since each pool uses its own algorithm to distribute rewards, we build separate algorithms for each pool.

We show that the top 0.1% or roughly 50 miners control about 50% of the mining capacity. Prior to 2022, mining concentration showed counter-cyclical patterns, declining during Bitcoin price run-ups as smaller miners entered, then increasing during downturns and halving events. However, since 2022, following the explosive growth of large-scale mining farms in the U.S., mining concentration has remained persistently high, regardless of price fluctuations.

Ownership concentration. Finally, we examine the ownership structure, concentration of Bitcoin holdings, and the trading behavior of whales. Since its inception, there has been persistent interest in identifying the largest Bitcoin holders and determining the extent of their ownership. Ownership concentration is not just a question of curiosity but carries significant economic implications. It influences the governance of the Bitcoin ecosystem, the potential for market manipulation and broader systemic risk. Moreover, understanding ownership patterns also sheds light on who stands to benefit most from Bitcoin’s price appreciation—whether it is a narrow group of early adopters or a broader segment of the public. It is particularly interesting in the context of Bitcoin since one of the original promises of this new architecture was to democratize financial services.

Determining the concentration of ownership is more complicated than just identifying addresses with the largest holdings of bitcoins, since many of these addresses belong to institutions that hold Bitcoin on behalf of many other investors. Consequently, to get insights into the concentration of Bitcoin ownership, one must focus on individual balances held outside of intermediary-controlled wallets. We develop a suite of algorithms based on graph analysis to classify clusters into those belonging to individual investors or those belonging to intermediaries and to trace ownership of Bitcoin over time, when rich clusters rebalance their holdings.

We show that ownership concentration among individual Bitcoin holders significantly exceeds wealth inequality in traditional assets. The top 10,000 individual addresses, which represent about 0.01% of Bitcoin holders, control approximately 5

million bitcoins (about 25% of the circulating supply). This concentration level is substantially higher than U.S. stock market ownership, where the top 1% of shareholders own 50% of the market. It is also much higher than the wealth concentration of the US population. We also show that these results are not driven by abandoned bitcoin accounts nor the holdings of Satoshi Nakamoto.

We conclude the analysis of whales by studying how whales established their positions and exit over time. Contrary to some prevailing beliefs, only a small fraction of whales entered early after the inception of Bitcoin. Instead, there is sustained entry over the entire period, and many of the large positions were established even when the price of Bitcoin exceeded \$10,000. The data also reveal pronounced cyclicity in the entry and exit behavior of whales: Entry accelerates when there are sharp run-ups in the Bitcoin price, while the peaks are slightly more muted post 2018, these patterns are stable over time. Interestingly, even among the whales that end up holding their Bitcoin position till the end of the sample, the so called “hodlers”, we see that the timing of their entry is very sensitive to the Bitcoin price. One might speculate that fear of missing out (FOMO) and financial motives are important drivers of many, even the early whales.

When looking at the exit decisions of whales (those which sell more than 95% of their stake) we again see strong sensitivity to the price of Bitcoin. A sizable fraction of whales tends to act counter-cyclical and reduce their Bitcoin holdings during periods of large run ups in the price of Bitcoin. However, we see a change in the size of the stakes that whales are selling over time. Pre-2014 many whales either exited their positions completely or sold large stakes in response to sharp increases in the price of Bitcoin. After 2014 we instead see an increase in partial exits, i.e. whales selling between 5% to 25% of their positions, during a price run-up. These results might suggest that early on a significant fraction of whales might have been worried that the Bitcoin phenomenon would be short lived and tried to realize their profits as soon as there was a price run up.

Overall, while absolute returns fluctuate with Bitcoin’s price cycles, the average and median realized returns at which whales choose to exit remain consistently similar across different time periods and market peaks and troughs. This finding suggests that whales have relatively stable targets for the profits they want to realize from holding Bitcoin, independent of when they entered the market or the overall price environment.

Related Literature. Our paper contributes to a fast-growing literature on cryptocurrencies and blockchains. [Härdle et al. \(2020\)](#) and [Halaburda et al. \(2022\)](#) provide general overviews of the economics of cryptocurrency and blockchain technology. [Budish \(2025\)](#), [Abadi and Brunnermeier \(2018\)](#), [Biais et al. \(2019\)](#) or [Lewis-Pye and](#)

[Roughgarden \(2023\)](#) analyze the consensus mechanisms and limitations of the proof-of-work protocol.

We also build on a small number of earlier papers that study the composition of Bitcoin transactions in the early period (2009–2015). [Athey et al. \(2016\)](#) use machine learning techniques to identify clusters and classify them into different types of industries and regions. [Tasca et al. \(2018\)](#) document the growing role of exchanges, broadly consistent with our findings. [Foley et al. \(2019\)](#) highlight the prevalence of illegal transactions on the Bitcoin blockchain. Their methodology leads to substantially higher estimates of illegal transactions in 2015–2017 than our baseline. This divergence reflects three methodological differences, which we document carefully in our paper: [Foley et al. \(2019\)](#) exclude exchange transactions from their denominator, recursively impute illegal status rather than relying on identified entities, and do not filter out short-term pass-through clusters that can inflate volume counts. Wallet-level blockchain data are also used by [Griffin and Shams \(2020\)](#) to study whether tether issuance affects bitcoin prices. [Graf von Luckner et al. \(2023\)](#) study Bitcoin transactions between selected P2P exchanges during 2017–2022 and estimate that about 13% of transactions can be classified as remittances to emerging markets.

In addition, we relate to the growing literature on forensic blockchain analysis, which applies computational methods to investigate cryptocurrency transactions for regulatory and security purposes. Notable contributions include [Cong et al. \(2025\)](#), [Khalilov and Levi \(2018\)](#), [Wu et al. \(2020\)](#), [Liu et al. \(2021\)](#), and [Weber et al. \(2019\)](#).

Finally, several papers study the economics of Bitcoin mining. [Prat and Walter \(2021\)](#) examines the relationship between the price of Bitcoin and investments in hashing capacity. [Easley et al. \(2019\)](#) and [Huberman et al. \(2021\)](#) develop equilibrium models of Bitcoin mining fees. [Cong et al. \(2020\)](#) proposes a theory of mining pools that suggests mining pools escalate the arms race between miners and significantly increase the energy consumption of proof-of-work blockchains. [Ferreira et al. \(2023\)](#) model the joint behavior of miners, mining pools, and firms producing specialized mining equipment. Earlier work in computer science has also pointed to concentration within mining pools. [Romiti et al. \(2019\)](#) trace payout transactions within BTC.com, AntPool, and ViaBTC and find that in each pool, fewer than 20 members receive over 50% of all BTC payouts. We contribute to this literature by developing a suite of algorithms to systematically identify individual miners and measure individual mining concentration across the entire Bitcoin network.

The remainder of this paper proceeds as follows. In [Section 1](#), we outline basic fact about the Bitcoin blockchain. [Section 2](#) describes our data sources and data construction. [Section 3](#) analyzes transaction volumes and network structure. [Section 4](#)

examines mining concentration. Section 5 we document the ownership concentration of Bitcoin participants and the wealth dynamics of whales. Section 6 concludes.

1. Bitcoin Blockchain

In this section, we outline the key features of the Bitcoin blockchain that are most relevant to our analysis. For additional details see, for example, [Antonopoulos and Harding \(2023\)](#).

All Bitcoin transactions are recorded on the Bitcoin blockchain, which is a distributed public ledger maintained by a network of participants, the so-called miners. The network organizes transactions into blocks that are added to the blockchain approximately every 10 minutes by the miners. Each block typically contains between 1,500 and 3,000 transactions, depending on transaction complexity and network congestion. A typical Bitcoin transaction includes a list of senders and recipients represented by pseudonymous addresses, the amount of bitcoins sent and received, and a time-stamp of the transaction.

While the pseudonymous nature of the Bitcoin blockchain enables tracing the movement of bitcoin between addresses, analyzing Bitcoin data presents two main challenges. First, the identity of the entity controlling a specific address is generally unknown. Second, the same entity can control multiple addresses, as users can freely create an unlimited number of addresses to receive and send bitcoins. These challenges are not unique to Bitcoin but are common across most cryptocurrencies.

The first task of identifying entities on the blockchain typically involves two approaches: (1) classifying an entity’s type based on its function within the network, which is accomplished by analyzing utilization patterns of unknown and known addresses; (2) associating entities on the blockchain with their real-world counterparts. This step relies on learning entity’s addresses either through direct interaction with them or through matching address balances with the reported holdings of publicly known entities.

The second task of grouping addresses together that belong to the same entity relies on address clustering heuristics, which leverage specific aspects of the blockchain design. The two most popular Bitcoin heuristics that have previously been developed are the union-find and change address methods.¹ The union-find heuristic is the most conservative clustering method. It assumes that all input addresses in a single transaction are controlled by the same entity, since signing a transaction requires control of all

¹See [Ron and Shamir \(2012\)](#) and [Meiklejohn et al. \(2013\)](#) for early results, and [Möser and Narayanan \(2022\)](#) for the latest developments.

inputs.² In practice, users specify only the destination addresses and transfer amounts. Their digital wallet software then automatically selects which of the user’s addresses to use to fulfill the transaction. When the same input addresses are repeatedly used across transactions, clustering algorithms can infer and group them as belonging to a single entity.

In comparison, the change address heuristic relies on the unspent transaction output (UTXO) model used by Bitcoin. In this model, an address’s balance is not stored directly but is computed from its transaction history. For efficiency, the Bitcoin protocol allows a user to spend only the exact outputs they have previously received.

For example, suppose an address received 5, 7, and 10 bitcoins, totaling 22 BTC. To send 8 BTC, the user must select a combination of prior outputs that equals or exceeds 8 BTC—such as 10 BTC, or 5 + 7, etc. The difference, or “change,” is returned to an address controlled by the sender—either a known address or a newly generated one. The heuristic aims to identify this change address and associate it with the sender’s cluster.

As the term “heuristic” indicates, all of these clustering methods are inherently imperfect. The union-find heuristic may miss addresses that are never co-spent, while the change address heuristic risks over-clustering due to inaccurate identification of change outputs. Additionally, since these heuristics are rule-based, users can deliberately manipulate them to conceal connections between different addresses. For example, when a hacker demands a ransom payment to a specific Bitcoin address, this address becomes publicly known. If the hacker later sends bitcoin from this address to an exchange or merchant, these funds could be flagged as coming from illegal activity. To prevent this, hackers often obfuscate the transaction trail by creating multiple addresses and splitting the initial payment among them. This process is typically repeated many times, resulting in “peeling chains”—sequences of transactions where small amounts are peeled off for use while the majority moves through a series of new addresses, making the funds difficult to track across the blockchain.

2. Data

We now describe how we obtain and process the Bitcoin data and outline the main building blocks of the blockchain database used in our analysis. The step by step details can be found in the Internet Appendix. The initial step requires parsing the

²See [Ron and Shamir \(2012\)](#) or [Meiklejohn et al. \(2013\)](#). Bitcoin mixing services, such as CoinJoin, allow users to combine their coins with others’ to obfuscate ownership and defeat this heuristic. The BlockSci software used in our analysis accounts for such cases by excluding CoinJoin transactions from its clustering algorithm.

blocks from the blockchain into a structured database and applying address clustering.

We download the Bitcoin blockchain data using the open-source software Bitcoin Core and parse these data into a database using the BlockSci platform.³ Our dataset covers the complete Bitcoin blockchain from its inception in January 2009 through June 13, 2025, comprising 901,128 blocks, 1.2 billion transactions, and 1.78 billion addresses.⁴ Our choice of BlockSci as parsing software is motivated by the fact that it is free, publicly available, and specifically designed for academic research with blockchain data. However, our analysis methodology can be easily adapted to any other parser that enables the construction of a table of Bitcoin flows between address clusters.

As explained in Section 1 above, one of the main challenges in working with Bitcoin data is identifying which addresses belong to the same entity. As a starting point, we use the union-find clustering algorithm and address its tendency to under-cluster (failing to group related addresses together) in subsequent analyses. We select this heuristic for two main reasons. First, unlike the change address heuristic, it is defined unambiguously with clear implementation rules. Second, as the most conservative clustering approach, it primarily errs by under-clustering addresses. Under-clustering mistakes are significantly easier to correct than over-clustering mistakes, as they merely involve merging additional clusters. Applying the union-find clustering algorithm to the data yields 959 million nonempty clusters—a significant reduction compared to the original 1.78 billion addresses.

2.1. Bitcoin flow database

Next, to improve the efficiency of our analysis of bitcoin flows, we build a database of Bitcoin flows between the clusters produced by the union-find algorithm. Each transaction is represented by a list of records, each consisting of the block ID, transaction ID, the sending cluster ID, the receiving cluster ID, and the amount of bitcoin received. The coinbase transactions (newly mined bitcoin entering circulation) are recorded as flows from a special cluster with an ID of -1 . Since the union-find algorithm assigns all sending addresses to the same cluster, one transaction can be represented by $n + 1$ records, where n is the number of recipient addresses.

³Bitcoin Core and BlockSci are available at <https://bitcoin.org/en/bitcoin-core/> and <https://github.com/citp/BlockSci>, respectively. Since the first version of this paper was written, BlockSci has ceased active development. As of this writing, it can fully process all transactions except those involving Taproot. While Taproot transactions are recognized and correctly processed by all key functions, including clustering algorithms, the alphanumeric string representation of Taproot addresses is not supported. This limitation does not impact our analysis, as we work with internal identifiers rather than address strings.

⁴The number of addresses depends on the definition of equivalent addresses. See https://citp.github.io/BlockSci/reference/addresses/equiv_address.html for further details.

For CoinJoin transactions where multiple users combine their transactions to enhance privacy, we apply a proportional splitting rule. Each transaction record is duplicated m times, where m is the number of distinct input clusters. The received amount is divided among these m clusters in proportion to their contributions to the total input value.

2.1.1. Pass-through Clusters

Many Bitcoin clusters have a very short lifespan and minimal activity, making them unlikely to represent stand-alone or economically independent entities. We refer to these as *pass-through clusters*. These clusters are frequently part of peeling chains, where funds move through a series of addresses to obscure their origin. Since addresses in peeling chains are typically never co-spent with other addresses, the union-find algorithm classifies each as a separate entity, creating a significant bias in the number of independent entities. If not corrected, this bias could dramatically overestimate the level of participation and activities on the blockchain.

To address this problem, we build two additional versions of the Bitcoin flow database for our analysis of Bitcoin volume and network structure.

There are two approaches one could devise for handling pass-through clusters. One would be to modify the clustering algorithm to assign addresses in pass-through clusters to existing clusters based on behavioral patterns. However, since different entities use their own specific rules to generate new addresses, this method increases the incidence of clustering mistakes. Therefore, we adopt an alternative approach in this paper, where we trace volume flowing through these clusters back to original sources and forward to final destinations, effectively removing them from the analysis. This methodology represents a novel contribution of our study.

We would like to note that our objective is not to identify all instances of obfuscating addresses—a task that would be extremely challenging given their variety and complexity, and one that lies beyond the scope of this paper—but rather to eliminate the most prevalent and impactful cases in order to ensure that the analysis of blockchain volume is meaningful.

We differentiate between two types of pass-through clusters. We classify the first type of pass-through cluster if the following four conditions are satisfied:

1. The cluster has only one incoming transaction and one outgoing transaction;
2. The cluster has a balance less than 546 Satoshi (approximately 0.00000546 bitcoin, the minimum economically viable transaction amount) after the two transactions;

3. The two transactions are fewer than 1068 blocks apart on the blockchain (approximately one week, given Bitcoin’s 10-minute average block time);
4. The incoming transaction is not a CoinJoin transaction.

For a non-CoinJoin transaction, the first condition ensures (with the union-find clustering algorithm) that the pass-through cluster receives its flows from a single cluster. This makes it straightforward to eliminate this cluster and reassign its volume: we simply assign flows from the cluster that sends to the pass-through cluster directly to the clusters that receive flows from the pass-through cluster, effectively skipping the intermediary cluster. There are 510 million pass-through clusters of the first type in our data, making them the most common type of clusters on the Bitcoin blockchain. Thus, after removing these clusters, the Bitcoin flow database is left with about 449 million clusters, which also brings significant gains in computational performance.

We define a second type of pass-through clusters by relaxing the first condition to allow for multiple incoming and outgoing transactions. This accommodates more complex obfuscation patterns while still capturing their transient nature. As with the first type, we require that the first and the last transaction are fewer than 1068 blocks apart and that a balance less than 546 Satoshi remains after the last transaction. This second type accounts for an additional 53 million clusters in our dataset and captures entities that serve primarily as temporary holders rather than final destinations for bitcoin.

Of the total 53 million type-2 pass-through clusters, 8 million clusters receive flows from a single cluster, making them very similar to the type-1 clusters. Refactoring the volume associated with type-two pass-through clusters that receive flows from more than one cluster cannot be performed unambiguously in the general case. To illustrate this, consider a scenario in which clusters A_1 and A_2 each send one bitcoin to a short-term intermediary cluster, which subsequently sends one bitcoin to clusters B_1 and B_2 . While it is possible that B_1 receives its bitcoin from A_1 , it is equally plausible that it originated from A_2 , or from a linear combination of both.

In our analysis, we assume that the volume flowing to downstream clusters is allocated in proportion to the volume received from upstream clusters. Applying this rule to the example above results in an equal distribution: each of clusters B_1 and B_2 receives 0.5 bitcoin from both A_1 and A_2 .

While this proportional allocation rule may introduce errors in individual cases, we expect such discrepancies to average out across a large number of transactions. Moreover, in practice, clusters that send funds to the same pass-through cluster are likely controlled by the same entity, which mitigates the impact of any specific volume-splitting rule. Further details on volume refactoring for type-two clusters are provided

in the Internet Appendix.

We show in Section 3.2 that accounting for volume associated with pass-through clusters is important for accurately interpreting Bitcoin transaction activity. Our analysis reveals that only approximately 10% of the raw transaction volume recorded on the blockchain reflects genuine transfers of value between distinct economic entities.

2.2. Identifying Bitcoin Entities

We focus on the following categories of Bitcoin entities that play an important role in the Bitcoin economy:

1. *Trading venues*: This category includes cryptocurrency exchanges (e.g. Coinbase or Binance), over-the-counter (OTC) trading desks, ETFs, funds and online wallets. These entities serve as the primary on- and off-ramps between the traditional financial system and the cryptocurrency ecosystem.
2. *Mining pools and miners*: These entities validate transactions and secure the Bitcoin network. Mining pools coordinate the efforts of multiple miners to increase the probability of receiving block rewards, while individual miners are entities that run mining rigs and supply the computing power to mining pools or operate independently.
3. *Whales*: In this category we identify individual, large investors known as “whales” in crypto-parlance. These entities typically exhibit distinctive holding patterns with limited outflows relative to their total holdings.
4. *Payment processors*: These entities facilitate the use of bitcoin in commercial transactions and allow merchants to accept cryptocurrency payments while receiving fiat currency, for example, BitPay and Coinbase Commerce.
5. *Gambling sites*: This category is comprised of online casinos, betting platforms, and gaming services that use bitcoin as their primary currency, for example, BC.GAME and FreeBitco.in.
6. *Illegal services*: These include darknet marketplaces (such as the now-defunct Silk Road and Hydra Market), ransomware payment addresses, stolen coins, scam operations, addresses linked to terrorism financing and child exploitation, sanctioned entities, and mixing services designed to obfuscate transactions by attempting to disguise the source and destination of funds.

2.2.1. Entity Identification Methodology

To identify and classify Bitcoin entities we employ two complementary approaches. First, we collect publicly disclosed Bitcoin addresses from multiple sources. Follow-

ing the collapse of FTX in 2022, several major exchanges—including BitMEX, OKX, and Binance—began publishing their wallet addresses as part of proof-of-reserves audits to enhance transparency.⁵ We systematically search cryptocurrency forums (bitcointalk.org) and blockchain explorers (OXT) to identify addresses that users report through direct interaction with these entities. For institutional investors, we cross-reference disclosed bitcoin holdings in public financial statements with on-chain addresses that match these reported amounts.

We also integrate data from several analytic platforms such as Bitfury Crystal Blockchain, Nansen, Flipside, and Arkham Intelligence. Bitfury Crystal Blockchain is a leading provider of anti-money laundering tools and cryptocurrency analytics, with particularly strong coverage of illicit services and gambling-related entities. Nansen, Flipside, and Arkham Intelligence offer more comprehensive coverage of institutional investors.

When conflicts arise between different identification sources, we prioritize officially disclosed addresses, followed by data from analytics platforms, and then user-reported addresses, applying manual verification for high-volume entities.

We believe that we have compiled the most comprehensive dataset of labeled Bitcoin addresses available to researchers. Table 1 presents summary statistics on the number of identified entities, clusters, and addresses across six entity categories. In total, our data cover 13,751 entities, including 1,578 exchanges, 104 investment funds, 332 payment processors and online marketplaces, and 594 gambling sites. For comparison, WalletExplorer—one of the most widely used public sources for wallet labeling—provides information on fewer than 280 entities and either omits or offers only partial coverage of major exchanges such as Binance, Coinbase, OKX, Bitfinex, investment funds, etc.⁶

Second, for entities lacking reliable labels from the aforementioned data sources, we develop custom identification algorithms that leverage network topology and transaction patterns. This methodology is a novel contribution of our paper and can be grouped into three main categories.

The first group, detailed in Section 4, consists of algorithms designed to identify individual miners by tracing the distribution of mining pool rewards to miners. This analysis requires only limited external data. It relies on information on the blocks that were mined by a specific mining pools, which we obtain from coinbase messages provided by Flipside. The approach also incorporates exchange labels to stop the tracing process when a miner collects rewards through an exchange-hosted account. Since mining pools have different distribution processes, we build individual tracing algorithms

⁵See, for example, <https://public.bitmex.com/?prefix=data/por1/> and <https://www.okx.com/proof-of-reserves/download> for publicly available proof-of-reserves data.

⁶See <https://www.walletexplorer.com/>.

for the 23 largest mining pools. Using this methodology, we identify approximately 403,500 miners who have received at least 25 rewards, with a total value of \$1,000 or more.

The second group, detailed in Section 5, consists of algorithms to identify large private Bitcoin investors and track their ownership over time. The analysis addresses two primary challenges: (1) distinguishing between addresses controlled by intermediaries and those privately held by individual investors, and (2) determining the ownership when a large cluster rebalances its holdings. The analysis leverages the knowledge of entity labels and the set of traced miners, as many miners—particularly those active early in the sample were able to accumulate substantial Bitcoin positions. Using this methodology, we identify approximately 42,000 whales, including 7,000 hodlers, aka whales that have sold less than 10% of their holdings after establishing a position.

Finally, in Section 3.2 we examine the top 10,000 clusters with the highest Bitcoin volume in each year and analyze their Bitcoin flows and utilization patterns to identify previously unlabeled clusters that are likely to correspond to unidentified trading venues and investment funds.

Our identification methodology has several limitations worth noting. First, while we achieve high coverage of major entities, many smaller entities remain unidentified. Second, our approach may not detect sophisticated entities that deliberately avoid established behavioral patterns.

3. Blockchain Volume and Network Structure

Since Bitcoin’s inception, there has been considerable debate about its primary use cases and economic function. While early proponents envisioned it as a peer-to-peer electronic payment system or an alternative form of money, critics have argued that Bitcoin operates primarily as a speculative asset or a vehicle for illicit transactions. To assess how Bitcoin is actually used in practice, we analyze the flow of bitcoins on the Bitcoin blockchain.

Our analysis shows that only about 10% of the raw transaction volume recorded on the blockchain corresponds to genuine transfers of value between distinct economic entities. The remaining 90% consists of within-entity activity, including wallet management, change-of-address flows, and peeling chain volume. Among inter-entity transactions, we find that trading venues account for approximately 70% of the volume, highlighting Bitcoin’s dominant role as a traded financial asset rather than a medium of exchange for goods and services.

We also examine the network structure of the Bitcoin network. We document a

high degree of concentration and interconnectedness among major entities, particularly exchanges. The high degree of cross-border flow and interconnection between exchanges has important implications for the enforcement of Know-Your-Customer (KYC) and Anti-Money-Laundering (AML) regulations, as transactions can easily flow between jurisdictions with different regulatory frameworks.

3.1. Within and Between-entity Volume

We begin our analysis of Bitcoin volume by decomposing total transaction volume into two components: within-entity volume and between-entity volume. The within-entity volume is generated when an entity sends bitcoins to itself, similar to moving cash from the left to the right pocket, and therefore, is not economically meaningful. This volume is the sum of two parts. The first is the within-cluster volume based on the clusters produced by the union-find clustering algorithm described in Section 2, which we further refine using the results from our analysis of miners and large Bitcoin investors. This volume includes internal wallet management activity, such as rebalancing and change-of-address flows. The second is the transitory volume associated with peeling chains, as discussed in Section 2.1.1. The precise allocation between these two parts depends on the clustering method used and is therefore not economically meaningful on its own.

The between-entity volume is the remaining volume, which represents transfers between different entities. While within-entity transactions do not correspond to genuine transfers of value, they are nonetheless recorded on the blockchain and may contribute to network congestion, particularly during periods of high activity. Figure 1 shows that the share of the between-entity volume has remained stable over time, consistently accounting for approximately 10% of total transaction volume on the Bitcoin blockchain.

3.2. Bitcoin Blockchain Volume

We next analyze the composition of between-entity volume to understand what types of entities drive Bitcoin usage on the Bitcoin network and how these patterns have evolved over time. For each month, and for every pair of entities i and j , we compute the volume of bitcoins sent from i to j and received by i from j . In a second step, we aggregate these bilateral flows by the entity types described in Section 2.2.

To address the issue of incomplete entity identification and ensure that we are not missing major players on the blockchain, we further analyze the 10,000 clusters with the highest Bitcoin volume in each year. Across the 2015–2025 sample, the union of

these annual top sets comprises 94,723 distinct clusters: 15,638 known clusters linked to 1,313 identified entities, 924 unknown miner clusters identified by tracing reward distributions from mining pools, and 78,161 other unknown clusters. For each cluster, we compute a set of diagnostic metrics, including the share of volume received from and sent to trading venues, the number of exchanges with which the cluster interacts, the number of addresses it contains, the total number of transactions, and aggregate volume.

We then compare the above cluster statistics across known entities to identify which metrics are most effective at distinguishing known trading venues and investment funds from other types of entities. Since the majority of large known clusters are associated with exchanges and investment funds, we aim to be conservative and mitigate potential bias arising from oversampling these entities. To this end, we prioritize minimizing the misclassification of non-trading venues as trading venues.

We find that a simple rule that requires the share of volume received from trading venues to exceed 60% performs well in practice. This heuristic results in only 2.48% (0.86% volume weighted) false discovery rate, with most misclassifications involving clusters associated with Tesla, SpaceX, payment processors, miners, and exchange stolen coins. At the same time, the rule successfully captures a large number of known trading venues. We apply this rule to the set of unknown clusters and identify 21,083 clusters that satisfy the criterion. In what follows, we refer to these as probable trading venues (PTVs).

Figure 2 presents the average monthly transaction volume generated by different types of entities on the Bitcoin blockchain from the beginning of 2015 through December 2024. Panel A reports volume in bitcoins (BTC), while Panel B shows the fraction of the average monthly volume associated with each entity type. As shown in Figure IA.1 in the Appendix, the composition of transaction volume remains stable over our period, indicating that Panel B is representative of composition of volume over the full sample period.

Panel A reveals a striking fact. Even though the Bitcoin price has increased dramatically over the past decade, on-chain volume denominated in BTC has remained remarkably stable. Panel B shows that the majority of between-entity volume—around 70%—is generated by transactions involving exchanges, investment funds, and PTV clusters. The classification of these and other entities is described in Section 2.2. The purple bin in the figure decomposes this exchange-related volume into three components: volume flowing to known exchanges (44.9% of total volume), volume flowing to PTVs (8.5%), and volume flowing from exchanges and PTVs to Other addresses (12%).

Other known entities account for only a minor share of total volume throughout the sample. For example, flows to addresses associated with illegal activity, scams, and gambling (orange bin) together constitute about 4% of volume in 2015; by the end of our sample in 2024, this share has fallen below 0.5%. Similarly, miners (light green bin) account for about 2% of volume by the end of 2024.

Our findings highlight the dominant role of trading venues on the Bitcoin network. These results are broadly consistent with those of [Tasca et al. \(2018\)](#), who study the early Bitcoin period (2009–2015), but on casual reading appear at odds with the estimates of [Foley et al. \(2019\)](#), who report \$76 billion in illegal activity on the Bitcoin blockchain, attributing 46% of all Bitcoin transactions in 2015 to illicit use, declining to 20% by 2017.

There are three important differences in methodology that explain this divergence. First, [Foley et al. \(2019\)](#) exclude exchange-related transactions. Our analysis covers the entire blockchain transaction flow, which substantially expands the denominator. Given the importance of trading and speculation in the use of Bitcoin, we believe it is important to include these transactions. Second, [Foley et al. \(2019\)](#) do not exclude short-term clusters when calculating volume. We show in Section 3.3 that short-term clusters are commonly used to obfuscate the origin of funds in illegal networks. Consequently, many such clusters are likely classified as separate illegal entities by [Foley et al. \(2019\)](#), which can significantly inflate their illegal-volume estimates. Finally, our baseline estimates rely on volume linked to identified illegal entities. Although our database covers more than 10,000 such entities, some illegal actors are undoubtedly missing. In contrast, [Foley et al. \(2019\)](#) impute illegal status. A cluster is classified as illegal if at least 50% of its transaction volume involves clusters already deemed illegal, and this classification is applied recursively: starting from an initial seed set, the illegal clusters in iteration $n + 1$ are those whose transaction volume with illegal clusters identified in iteration n exceeds the threshold.

To estimate the potential impact of missing illegal clusters, we follow an approach similar to that of [Foley et al. \(2019\)](#). For each unknown cluster, we compute three quantities: (a) volume to and from clusters identified as darknet marketplaces, darknet services, terrorism financing, illegal services, sanctioned entities, child exploitation, or mixing services; (b) volume from clusters identified as ransom-payment addresses, stolen coins, or scams; and (c) volume to seized-asset addresses. We define illegal volume as $V_{il} = (a) + (b) + (c)$ and classify a cluster as illegal if

$$\frac{V_{il}}{V_{in} + V_{out}} \geq 0.1.$$

We denote the resulting set of clusters by $I(1)$. This initial set contains approximately 6 million clusters. The choice of the initial 10% threshold is somewhat arbitrary. Since our goal is to obtain an upper bound on illegal volume, we set the initial threshold at a relatively low value. We also vary the threshold to 2% and 20% and obtain similar results.⁷

To iterate from $I(n)$ to $I(n+1)$, we define $V_{I(n)}$ as the volume to and from clusters in $I(n)$ and classify a cluster as illegal if

$$\frac{V_{il} + V_{I(n)}}{V_{in} + V_{out}} \geq k,$$

where k is a pre-specified threshold. By construction, the set of illegal clusters is non-decreasing in n . We denote by $I(\infty)$ the limiting set: $I(\infty) = \lim_{n \rightarrow \infty} I(n)$.

Figure IA.2 in the Internet Appendix shows the number of clusters in $I(n)$ as a function of n for $k = 0.5$ and $k = 0.25$. The limiting set $I(\infty)$ is very sensitive to the threshold. For $k = 0.5$, $I(\infty)$ contains about 10 million illegal clusters; for $k = 0.25$, this number rises to 43 million. The reason for this sensitivity, as we discuss in Section 3.3, is the high degree of interconnectedness of the Bitcoin network. Setting $k = 0.1$ effectively classifies all unknown clusters as illegal.

Figure IA.3 in the Internet Appendix shows the share of illegal volume based on identified and imputed illegal actors for $k = 0.5$ and $k = 0.25$. For $k = 0.5$ —the threshold used by Foley et al. (2019)—the imputation has only a limited impact on the baseline estimate, resulting in the share of illegal volume being below 1% after 2020. Even for $k = 0.25$, the imputed share of total volume remains relatively low: approximately 10% in 2015–2017, declining to about 4% for the remainder of the sample. On net, this sensitivity analysis shows that our estimate of illegal volume is robust to plausible classes of omitted illegal clusters.

While our estimates indicate that a relatively small share of total Bitcoin volume on the blockchain is tied to illegal use, this should not be interpreted as dismissing the role of Bitcoin in facilitating illegal or “gray” activities. As we show in Section 3.3, the pseudonymous nature of Bitcoin, the high interconnectedness of its entities, and the limited traceability of transaction flows make it an appealing instrument for the gray economy and pose serious challenges for AML regulation.

⁷A nonzero threshold is necessary because, as we show in Section 3.3.3, even exchanges with high KYC standards, such as Coinbase and Kraken, have some direct interactions with illegal services in the data.

3.2.1. On-chain and off-chain exchange volume

We conclude this section by comparing on-chain and off-chain volumes associated with exchanges. Nearly all Bitcoin exchanges are centralized. When market participants deposit their bitcoins with these exchanges, they de facto transfer their coins to the exchange. Exchanges typically aggregate all customer deposits and store them in cold wallets—Bitcoin addresses held on devices that are not connected to the internet. As a result, the trading activity of individual investors on exchanges cannot be directly observed using blockchain data. De jure, however, customers retain legal ownership of their bitcoins and can trade them for other assets. Such trades are typically executed through centralized limit order books maintained by the exchanges and reported externally via application programming interfaces (APIs).

To compute off-chain volume, we use daily trading volume data from the 15 largest exchanges identified in [Makarov and Schoar \(2020\)](#). These data are sourced from Kaiko, a private data provider that has been collecting comprehensive cryptocurrency trading information since 2014. Table 2 reports the monthly time-series correlation between off-chain and on-chain volumes for each of the 15 exchanges, along with the average ratio of off-chain to on-chain volume.

The table reveals strong correlation between off-chain and on-chain volumes across exchanges. The average monthly correlation is about 70%, with the lowest correlation of 31% observed for Coinbase and the highest, at 96%, for the now-defunct FTX. The table also indicates that while off-chain volume generally exceeds on-chain volume, the two are of comparable magnitude.

These findings suggest that on-chain volume may serve as a novel and informative metric for assessing the prominence of cryptocurrency exchanges. Traditionally, data aggregators such as CoinMarketCap have relied on measures including off-chain trading volume, website traffic, and the number of Twitter followers. A key requirement for any such ranking metric is resilience to manipulation. However, existing indicators are not fully manipulation-proof, see for example, [Aloosh and Li \(2024\)](#) and [Cong et al. \(2023\)](#). While some exchanges may inflate their reported off-chain volumes through practices such as wash trading, manipulating on-chain volume may be more costly and thus less susceptible to such distortions.

The strong time-series correlation between on-chain and off-chain volume also suggests that the act of transferring bitcoins to an exchange could be an informative indicator of impending trading activity. This is consistent with the common practice of monitoring the on-chain activity of Bitcoin whales as a source of valuable trading signals.

Finally, the high correlation may also reflect the movement of arbitrage capital. As

shown in Makarov and Schoar (2020), periods of elevated trading activity on exchanges often coincide with increased price discrepancies across platforms. To exploit these arbitrage opportunities, traders must frequently transfer bitcoins from lower-price to higher-price exchanges. This reallocation of arbitrage capital generates on-chain flows that are positively linked to off-chain trading activity, thereby reinforcing the observed correlation between on-chain and off-chain volume.

3.3. Network Structure

We conclude our analysis of Bitcoin flows with an examination of the network structure of the Bitcoin ecosystem. The ability of parties to transact and send funds pseudonymously across geographic or regulatory boundaries is often seen as one of the benefits of cryptocurrencies. But it also presents challenges for the enforcement of KYC and AML regulations.

We provide an analysis of the role that exchanges and large regulated entities can play in shaping how the legitimate (“white”) and illicit (“gray”) parts of the Bitcoin economy interact. We first document the interconnectedness and dominant position that the largest and most active entities on the network have, and the role they can play in improving the transparency of the Bitcoin network if they were to comply with regulations such as the European Union’s MiCA regulation or similar proposals. As a second step, we then provide the example of Hydra Market, one of the largest darknet marketplaces prior to its shutdown in 2022, to illustrate the challenge of system-wide AML regulation if some exchanges fail to comply.

3.3.1. Top Entities

Given the rapidly evolving nature of the Bitcoin ecosystem, we focus our analysis on the most recent period—from the beginning to the end of 2024. For each cluster, we calculate the total transaction volume over the year and select the top 25 clusters with the highest cumulative volume.

To understand how these entities interact, we construct a directed, weighted network graph based on their bilateral transaction flows. Each node in the graph represents one of the top 25 clusters, and a directed edge from node i to node j corresponds to the total bitcoin volume received by entity i from entity j .⁸

Figure 3 displays the resulting network. Node sizes are proportional to the total volume received by each entity, while edge thickness reflects the larger of the two

⁸In a directed network, an edge from node i to node j and an edge from node j to node i are treated as distinct edges.

bilateral flow volumes between a pair of entities. In cases where two clusters exchange flows in both directions, the edge is drawn in the direction of the dominant flow and is marked with a red segment. The figure also displays entity names where available. Of the top 25 clusters, 23 correspond to known entities in our dataset. These include major exchanges such as Coinbase and Binance; custodians like Fidelity and Coinbase Prime; market makers such as Wintermute and Cumberland; ETF providers like BlackRock; and trading firms like Jane Street. The unknown two clusters likely correspond to unidentified clusters of over-the-counter desks.

Figure 3 illustrates that on-chain activity is dominated by institutional entities, with the two largest exchanges, Coinbase and Binance, playing particularly prominent roles. The figure also reveals a high degree of interconnectedness among the top entities, despite their operations spanning different jurisdictions and geographic regions. For instance, Bybit is headquartered in Dubai, while Binance currently operates without an official central headquarters.⁹

The high degree of interconnectedness among entities has important implications for KYC and AML regulation. On the one hand, such connectivity could enhance the transparency of the Bitcoin network if all major entities comply with KYC and AML regulations and report information about their counterparties. On the other hand, this same interconnectedness could facilitate the movement of tainted bitcoins (those associated with illicit activities) if even a few entities operate with lax KYC standards.

To illustrate these points, we now provide two examples. First, we demonstrate the potential impact that strict enforcement of KYC and AML regulations could have on improving the transparency of counterparties on the blockchain. Second, we highlight the limitations of these regulations in the presence of heterogeneity in adherence to, or coverage of, such rules across different jurisdictions.

3.3.2. Travel Rule

Panel B also reveals that a substantial portion—approximately 30%—of the transaction volume on the Bitcoin blockchain consists of flows between unknown clusters.¹⁰ This as well poses potential concerns from an AML perspective, particularly as many governments, including the United States, explore the broader adoption of Bitcoin into the mainstream financial system.

In response to such challenges, the European Union’s MiCA regulation introduced

⁹This pattern becomes even more pronounced when smaller entities are included in the analysis.

¹⁰A significant portion of this volume is associated with the top volume clusters, which we conservatively do not classify as PTV. However, if one were to apply the probability distribution implied by the labeled clusters, a substantial share of this “other” volume would be attributed to the PTV category.

comprehensive reporting requirements for crypto-asset service providers (CASPs). Under MiCA, CASPs must collect and provide information about both the originator and the beneficiary of crypto-asset transfers, irrespective of the transaction amount. This includes adherence to the so-called “Travel Rule,” which mandates that all crypto-asset transfers carry identifying information on both parties involved.

Figure 4 presents a hypothetical decomposition of transaction volume under the assumption that MiCA regulations were applied to all identified trading venues in our sample. To account for the potential obfuscation effects of peeling chains, this exercise restricts the analysis to factoring out only the volume associated with type-1 pass-through clusters. As detailed in Sections 2.1.1 and 3.1, these clusters exhibit a simple and transparent structure, contribute significantly to the raw transaction volume on the blockchain, and allow for unambiguous volume refactoring. We therefore assume that these clusters to be reported as part of the Travel rule.

The blue area in Figure 4 shows the volume to the known trading venues. It accounts for a slightly smaller share than in Panel B of Figure 2 because the total volume now contains the volume in type-2 pass-through clusters. Nevertheless, it remains substantial, hovering around 40%.

The red area depicts the volume associated with what we define as level-1 clusters—clusters that directly send or receive bitcoins from known trading venues. When combined with known trading venues, these clusters account for approximately 85% of total blockchain volume. While this significantly expands the proportion of volume that can be attributed to identifiable entities, about 15% still remains unaccounted for.

Lastly, the green area represents the volume associated with level-2, which are clusters that transact directly with level-1 clusters. These entities could potentially be identified if level-1 cluster owners were required to disclose information about their counterparties. This would be consistent with provisions in the MiCA regulation, which stipulate that owners of stand-alone wallets exhibiting complex transaction histories may be subject to enhanced disclosure requirements. It is also possible that some level-2 clusters are, in fact, misclassified level-1 clusters, particularly if certain level-1 clusters belong to yet-unidentified CASPs.

Figure 4 illustrates that if both level-1 and level-2 clusters could be reliably identified, up to 95–98% of the transaction volume on the blockchain would become attributable. This suggests that the effective implementation of MiCA could substantially enhance transparency regarding the parties transacting on the Bitcoin blockchain.

That said, these results likely represent an upper bound on the share of volume that can be identified. First, some trading venues may fall outside the scope of MiCA. Second, market participants may alter their on-chain behavior in response to the reg-

ulation, potentially reducing the effectiveness of identification efforts.

To illustrate the challenges of enforcing KYC regulations and how tainted bitcoins can reenter the “white” segment of the Bitcoin ecosystem through less rigorously regulated exchanges and complex transaction patterns designed to obscure the origin of funds, we analyze the network structure surrounding Hydra Market in the following section.

3.3.3. Hydra Market

Hydra Market operated from 2015 until its shutdown in April 2022. For our analysis, we focus on the period from 2020 to June 2021. During this timeframe, Hydra Market received 179,904 bitcoins from 2,292,310 distinct clusters and sent bitcoins to 461,100 clusters. The 2,292,310 sending clusters themselves received flows from 4,555,871 clusters, while the 461,100 receiving clusters subsequently transferred bitcoins to 1,139,225 additional clusters.

Figure 5 depicts the resulting network. For clarity of presentation, we retain only nodes that transferred at least 10,000 bitcoins within this network. In calculating volume for this analysis, we deliberately exclude flows involving clusters associated with known or PTV clusters studied in Sections 3.2.

Node size reflects the total amount of bitcoin received from Hydra Market, while edge thickness is proportional to the volume exchanged between two distinct entities. As in the earlier figure, when two clusters send flows to each other, the edge direction reflects the dominant flow and is highlighted with a red segment. Nodes are color-coded as follows: green nodes are identified clusters, purple nodes are unknown clusters, and orange are short-lived clusters with a life-span below one week.

Figure 5 reveals that the highest-volume entities interacting directly with Hydra Market were exchanges known at the time for having relatively weak KYC standards, including LocalBitcoins, Bitzlato, Binance, and HTX(Huobi). Once funds reached these exchanges, they were commingled with other flows, rendering them untraceable and enabling onward transfers to a wide range of destinations across the Bitcoin ecosystem.

The figure also indicates that direct interactions between Hydra Market and exchanges known for enforcing stricter KYC standards, such as Coinbase and Gemini, were limited. However, these exchanges exhibited significantly larger transaction volumes with clusters that were adjacent to Hydra Market in the network. For instance, Coinbase directly sent 345 bitcoins to, and received 86 bitcoins from, Hydra Market. In contrast, it sent approximately 247,000 bitcoins to, and received 89,000 bitcoins from, neighboring clusters, which also received significantly larger flows from Hydra

Market. This pattern shows that tainted flows may reach regulated entities indirectly, highlighting the complexity of tracing illicit activity across multiple transactions.

As shown in Figure 5, the majority of flows to and from Coinbase occurred via pass-through clusters, entities that were typically created for the sole purpose of obfuscating the origin of funds. A typical “whitening” transaction involves mixing tainted bitcoins (those traceable to Hydra Market in this specific example) with “clean” funds that have no links on the blockchain to illicit activity. Each round of mixing reduces the identifiable share of tainted volume. This process is typically repeated multiple times until the resulting flows are sufficiently cleaned to be sent to exchanges with KYC requirements, such as Coinbase.

The above analysis demonstrates that enforcing KYC and AML regulations on a limited set of exchanges—or even on all entities within a single jurisdiction—is typically insufficient. Although the traceability of Bitcoin on the blockchain offers tools to counteract certain obfuscation strategies, tainted flows can still reenter the mainstream network through a sequence of “whitening” transactions, when the entities serving as entry points do not uniformly enforce KYC and AML standards.

The current fragmentation of the global regulatory landscape creates arbitrage opportunities for malicious actors. To effectively prevent the inflow of tainted funds, regulators in stricter jurisdictions may need to impose restrictions on the types of counterparties that regulated entities can engage with, or require that all incoming funds be verified as “clean.” If rigorously enforced, such measures could lead to a segmentation of the Bitcoin ecosystem into distinct “white” and “grey” sectors.

While the tracing of flows on the Bitcoin ledger might be easier than tracking cash and bills across jurisdictions, the physical nature of cash makes it more difficult for bad actors to transport and use at scale. Consequently, as Bitcoin adoption continues to grow, the utilization of Bitcoin in the grey economy is also likely to expand—though it may face increasing competition from other cryptocurrencies that are more difficult to trace.

4. Miners

Miners serve as the backbone of the Bitcoin blockchain’s verification process. By processing and validating transactions, they ensure the integrity of the distributed ledger. In return for their computational work, miners are compensated with newly minted bitcoins and transaction fees collected from confirmed transactions. This is the so-called “trustless trust” architecture. Bitcoin’s proof-of-work (PoW) protocol assumes that the majority of mining power is held by honest participants. If a single

miner—or a coordinated group—were to control a majority of the network’s mining power, the system would be vulnerable to a so-called 51% attack. In such a scenario, the malicious entity could reverse or reorder previously validated transactions, thereby compromising the blockchain’s trustless consensus.

But collusion between miners does not need to involve overt attacks such as a 51% takeover. Instead, coordination may serve to protect economic rents or resist upgrades to the protocol that threaten incumbents’ profitability. For instance, large miners may oppose the adoption of more efficient consensus mechanisms or rule changes if these would reduce their expected returns or render their existing infrastructure obsolete.

Moreover, the centralization of hashing power increases systemic risk. As mining becomes concentrated among a small number of actors, the network becomes more vulnerable to targeted disruptions. A malicious party could compromise Bitcoin’s consensus protocol not by dominating global hash rate directly, but by targeting a handful of large miners.

Satoshi Nakamoto, Bitcoin’s pseudonymous creator, originally envisioned a decentralized mining environment made up of numerous, small and independent miners. In such a setting, collusion would be difficult, and the failure of any individual miner would have negligible impact on network security.

In practice, however, the economics of PoW mining runs counter to this decentralized ideal. The production function of individual miners displays significant scale economics due to fixed costs such as co-locating with cheap and abundant energy providers, reliance on specialized and sophisticated mining equipment among others. In addition, the likelihood of successfully mining a block—and earning the corresponding reward—is proportional to a miner’s share of the total hashing power. This creates strong incentives for miners to pool resources and share rewards, thereby insuring one another against the volatility of individual mining success. As a result, mining activity is now dominated by large mining pools.

The concentration of mining pools has drawn significant public and academic attention ([Cong et al., 2020](#)). Some observers argue that such concentration does not necessarily threaten the protocol, noting that pool operators do not directly control their contributing miners. Therefore, if any pool is noticed to engage in rogue behavior its miners can leave it and join other pools, thereby disciplining the operator. However, this argument becomes weaker when a substantial portion of a pool’s capacity is provided by a few large, and often affiliated, mining operations.

These considerations underscore the importance of measuring not only the concentration of mining pools but also the concentration of mining power among individual miners, which we address in the following section.

4.1. Identifying Individual Miners

We build a new methodology for identifying individual miners which consists of three main steps: (1) mapping blocks to mining pools, (2) tracing reward distribution patterns to identify individual miner addresses, and (3) applying filtering criteria to focus on economically significant participants. This section details each component of our identification process.

4.1.1. Pool Identification and Block Mapping

We begin by establishing which blocks were mined by each mining pool using two complementary data sources. First, we exploit the common industry practice that mining pools embed their name or identifier in the coinbase message of any blocks they mine. We obtain coinbase message data from Flipside and systematically parse these messages to establish block-to-pool correspondences. This approach captures the majority of recent mining activity, as most established pools consistently include identifiable information in coinbase messages.

Second, we supplement this dataset with pool-block mappings obtained from a blockchain explorer BTC.com that maintains comprehensive pool attribution data. This additional source proves particularly valuable for the earlier portion of our sample period (2015-2017), when coinbase message practices were less standardized and pool identifiers were often ambiguous or absent. The combination of these sources allows us to construct a comprehensive mapping of coinbase reward addresses associated with each identifiable mining pool.

4.1.2. Individual Miner Address Identification

Once pool-level reward addresses are identified, we trace the subsequent distribution of rewards to individual miners. Each mining pool employs its own internal reward distribution mechanism and we therefore have to build pool specific tracing algorithms. We recover pools' distribution mechanisms by analyzing transaction chains originating from pool coinbase addresses. The typical distribution process follows a hierarchical structure: rewards flow from the initial coinbase address to one or more primary distribution addresses, which then send payouts to a group of miners, retain the remainder (change), and initiate a subsequent transaction from new intermediate addresses. This process continues until the rewards are fully distributed.

To distinguish legitimate miner payouts from internal pool operations and smaller pools that are part of larger pools, we apply several identification criteria. We classify an output address as representing an individual miner if: (1) it receives funds through

the pool’s distribution chain and (2) it does not subsequently redistribute these funds to other addresses in a pattern consistent with further pool operations (such as sending its rewards to other miners’ addresses, or having rewards that are unstable over time or come in integer numbers). Since pools may also use distribution addresses for unrelated operational activities (e.g., purchasing equipment), we filter out any addresses identified as internal to the pool.

4.1.3. Filtering and Validation

To focus our analysis on economically meaningful mining operations, we apply two filtering criteria to our candidate miner addresses. First, we exclude addresses that receive total rewards equivalent to less than \$1,000 (valued at time of receipt) over our sample period. This threshold eliminates small-scale hobby miners while retaining addresses representing substantial mining investments. Second, we exclude addresses with fewer than 25 separate reward transactions, ensuring that our sample captures miners with sustained participation rather than sporadic activity. After applying these filters, our methodology identifies 403,510 individual miners across all traced pools.

4.1.4. Measurement and Coverage

Because the rewards of a miner are proportional to its contribution of hashing power, we measure the capacity of each miner by the volume of bitcoins they receive from the pools through distribution transactions. While mining pools employ varying fee structures and payout schemes (Cong et al., 2020), we expect these variations to have minimal impact on our capacity measure because of the competitive nature of the mining pool market.

Table IA.1 in the Internet Appendix reports the total number of blocks mined and the corresponding volume of bitcoins for the mining pools used in our analysis. Pools for which we successfully trace individual miners are indicated in bold font, while private pools that do not distribute rewards externally are shown in italics.¹¹

Figure IA.5 in the Internet Appendix shows the aggregate mining capacity of traced pools and their corresponding individual miners from January 2015 to December 2024. Our tracing coverage increases over time: in early 2015, we successfully trace approximately 60% of total network mining rewards, while by late 2024, this figure reaches nearly 90%. Throughout the sample period, we maintain approximately 90% coverage

¹¹Our analysis focuses on pools active during our sample period. Some exchange-affiliated pools such as Huobi Pool intermingle mining reward distributions with exchange-related transactions, making reliable isolation of individual miner payouts challenging. These pools are excluded from our individual miner analysis but retained in our pool-level statistics.

of rewards distributed by the pools we identify.

4.2. Concentration of Mining Capacity

We now analyze the concentration of mining capacity across individual miners. For each month, we rank active miners by their mining capacity and compute the share of total capacity held by different quantiles. The results for the top 50%, 10%, 5%, 0.5%, and 0.1% of miners are shown in the left panel of Figure 6. The figure reveals that Bitcoin mining is highly concentrated. The top 50% of miners consistently control nearly all mining capacity; the top 10% account for approximately 90%, and the top 0.1% control close to 50%.

To document the magnitude of the concentration we compute the number of miners required to collectively control 10%, 20%, 30%, 40%, and 50% of total mining capacity, as shown in the right panel of Figure 6. The 50% threshold is of particular interest, as it directly relates to the potential for a 51% attack (blue line). Prior to 2022 we see a strong counter-cyclical pattern in mining concentration. Concentration declined sharply following run ups in the price of Bitcoin, but to rise during downturns—such as in 2018 and 2022. Concentration also increases after Bitcoin halving events which occurred in July 2016 and May 2020. For example, between 2015 and 2016, fewer than 50 miners were typically sufficient to exceed this threshold. In early 2018, the number peaked at approximately 700 miners. This pattern suggests that smaller miners tended to enter and exit the market in response to changes in the economic attractiveness of Bitcoin.

Starting in 2022, however, the number of miners needed to control 51% dropped below 50 and has stayed at this level with minor variations until the end of the sample period. Even during the sharp run up of the Bitcoin price in 2024, we do not see the same spike in the entry of smaller miners anymore. One of the main factors behind this increase in concentration might be the entrance of large U.S. public Bitcoin mining companies such as Marathon Digital Holdings and Riot Blockchain.

5. Ownership of Bitcoin

Since Bitcoin’s launch in 2009, considerable fascination and attention has focused on the question of who owns the largest share of coins and how concentrated these holdings are. Websites devoted to tracking large Bitcoin addresses—the so-called “rich lists”—are widely followed in the crypto community. While often framed as a matter of curiosity, the issue of ownership concentration has important economic implications; it

affects, for example, the governance of the bitcoin ecosystem, the potential for market manipulation, and systemic risks. It is also important to understand who stands to benefit most from price appreciation: is it a select group of early adopters or the wider public?

Answering these questions requires more than simply identifying “rich addresses”, i.e. those with the largest balances. Many of these addresses belong to institutions that hold Bitcoin on behalf of many other investors. In particular, centralized exchanges, ETFs or custodial wallet services manage substantial quantities of Bitcoin deposited by users. These entities typically pool coins owned by many different users and store them in secure but co-mingled wallets, making it impossible to infer the holdings of individual investors from the aggregate balances of intermediaries. Consequently, to get insights into the concentration of Bitcoin ownership, one must focus on individual balances held outside of intermediary-controlled wallets.

5.1. Known Wallets

We begin our analysis of Bitcoin ownership by examining the holdings of labeled wallets. These are categorized into exchanges, investment funds, miners, governments, and wallets associated with illicit activity. A significant portion of the bitcoins attributed to government-controlled wallets originates from assets seized through enforcement actions against illegal operations. A smaller fraction belongs to governments like El Salvador who are holding bitcoins for investment purposes.

Figure 7 illustrates the evolution of Bitcoin balances across the five labeled entity groups over time. Among these, exchanges consistently hold the largest share of known balances. Their aggregate holdings stood at approximately 2.4 million bitcoins at the beginning of 2015 and grew steadily, peaking at around 4.1 million bitcoins in early 2020. This trend reflects the dominant role of centralized exchanges as the primary gateway for retail investors to enter the cryptocurrency market during this period.

The next largest group is investment funds, which includes Bitcoin ETFs, hedge funds, and custodians that hold bitcoin on behalf of other institutional investors. The holdings of these entities have grown rapidly—from about 500,000 BTC at the beginning of 2020 to about 2.5 million BTC by the end of 2024—making this the fastest-growing segment of the Bitcoin market.

As a result, the combined balance held by intermediaries such as exchanges and funds has reached almost 6 million by the end of 2024. The remainder of the balances held by governments, miners and illegal wallets is around 700,000 BTC and has not changed significantly over the last five years.

5.2. Unknown Wallets

We now turn to the analysis of bitcoin holdings held outside known wallets. As previously mentioned, any meaningful analysis of ownership concentration must distinguish between addresses controlled by intermediaries and those held directly by investors. Moreover, to understand the wealth dynamics of whales, one also needs to solve another important challenge: How to trace the ownership of Bitcoin when rich clusters rebalance their holdings. In most cases, clusters use a new set of addresses that are never co-spent with the existing cluster addresses. As a result, we need to devise an algorithm to decide when a rebalancing represents a change in ownership and when the ownership remains unchanged. To address these challenges, we develop a multi-step approach that combines blockchain analysis with ownership tracing algorithms, which we describe in the next section.

5.2.1. Identifying Large Bitcoin Investors

We start with identifying candidate rich clusters that are likely controlled by individuals. First, we exclude pass-through clusters and all clusters linked to known institutional entities. Second, we select clusters that ever had a maximum balance of at least 100 BTC. The threshold of 100 BTC is somewhat arbitrary, but we set the threshold purposefully low to start from a wide set of potential whales. The 100 BTC threshold corresponds to an investment position of one million dollars at the Bitcoin price of \$10,000, which was reached for the first time in November 2017. It excludes many of the early addresses (about 20,000) with a balance of 50 bitcoins mined by Bitcoin’s creator, Satoshi Nakamoto, which are widely believed to be inactive.¹²

We then apply three additional filters to exclude large clusters that are likely controlled by institutional investors or are very short-term traders:

1. The time when the cluster first holds 1 BTC to the time when it last holds 1 BTC should be more than 30 days.
2. A cluster should have fewer than 100 addresses.
3. A cluster should receive and send flows to fewer than 50 unknown clusters.

The particular thresholds are set to separate individual investors from institutional wallets based on the behaviors we observe for known institutions: Institutional investors tend to rebalance their portfolios significantly more often than individual investors, interact with more entities, and have more transactions, leading to a higher number of

¹²We do not assign the ownership of these addresses to one person. The reason for this choice is that there is ongoing uncertainty surrounding the identity of Satoshi Nakamoto, including the possibility that the name represents a group rather than a single person.

used addresses. Applying these criteria yields approximately 162 thousand candidate clusters.

Whales often redistribute their bitcoin holdings across different addresses to enhance privacy or improve security. As a result, when a cluster transfers its balance to another cluster, it becomes necessary to determine whether ownership has changed hands or remains with the same entity.

We implement the following tracing algorithm to assess ownership continuity: We trace outflows from each of the candidate clusters forward in time starting from the earliest clusters and evaluate the recipient clusters that the bitcoins are sent to. If a recipient cluster receives all its inflows from the original cluster, we assume it is part of the same ownership entity. In contrast, if the holdings are sent to a recipient cluster that receives some of its inflows from other addresses that cannot be traced to the original cluster, we assume that the holdings now belong to a different owner.

There might be cases in which whales may control multiple addresses that never co-spend, causing bitcoin flows to appear as transfers to new owners when, in reality, they are movements between addresses owned by the same entity. This would lead us to break the continuity of ownership too quickly. We believe the likelihood of this type 2 error is small relative to the large type 1 error that would arise, if we were to include such ownership changes.

The tracing process is computationally intensive. We describe further methodological details in the Internet Appendix. After running our tracing algorithm, we consolidate the original set of 162 thousand clusters into approximately 114 thousand merged clusters.

To reduce the risk of misclassifying unidentified custodial or exchange cold wallets as individual investors, we implement two further filters. First, for each merged cluster, we identify the counterparties that make up its largest inflows and outflows. If more than 90% of both inflows and outflows are in the same cluster, we exclude it, interpreting this as a potential cold wallet. Second, to reduce false positives in identifying institutional entities, we impose an inactivity constraint. We require that any cluster has a balance above 50 BTC for more than 180 days and any cluster with a balance exceeding 50 BTC has a period of inactivity of at least 90 days. Institutions typically adjust the balance of their cold wallets on a more frequent basis than large individual investors. Therefore, whales are more likely to meet this inactivity constraint.¹³

¹³We also vary the threshold for the inactivity constraint between 90 days to 180 days. When we expand the inactivity constraint, we might be able to filter out more unclassified wallets of institutions, but naturally we might also lose some legitimate whales who trade more frequently. We repeat all our analysis, using these different cut offs, and we do not see any material difference in the results reported below.

After applying these additional filters, we obtain a sample of 42,000 whales, of which 7,000 have never sold a significant portion of their Bitcoin holdings. With these data at hand, we are now ready to look at the concentration of individual Bitcoin ownership.

5.2.2. Ownership Concentration

By construction, our definition of a whale requires a cluster’s maximum balance to exceed 100 bitcoins. As a result, our sample includes all individual clusters holding more than 100 BTC, but it may omit a substantial number of smaller clusters with balances below this threshold.

This omission is innocuous when the focus is on the upper end of the ownership distribution but may introduce bias when analyzing the lower end. Because it is impractical to study and trace the ownership of all wallets, we adopt a simple decision rule to classify low-balance wallets as belonging to individual investors. Specifically, we require that a wallet exhibit no transaction activity throughout the entire year of 2024 in order to be included in this category. This inactivity constraint is more stringent than the one we applied to large-balance wallets to account for the fact that we do not impose any other constraints in this segment of the distribution.

Some of these clusters may correspond to old or even forgotten addresses, while others are likely held by long-term investors. By the end of 2024, there are approximately 646,000 such clusters, collectively controlling 4.35 million bitcoins.

To compute the concentration of Bitcoin ownership, we sort individual clusters, including both the previously identified whales and low-balance clusters that remained dormant throughout 2024, according to their balance at the end of 2024 and then plot their cumulative balance against the number of individual clusters that are holding these bitcoins.

Figure 8 presents the results. It shows that despite the growing adoption of Bitcoin among retail investors, ownership remains highly concentrated among a small number of large holders, even at the end of 2024, as illustrated by the blue line. The top 1,000 clusters control 2.6 million bitcoins, while the top 10,000 clusters collectively hold approximately 5.1 million bitcoins, which is a bit more than a quarter of all outstanding bitcoins.

A concern frequently raised within the crypto community is that some early investors may have lost access to their wallets by misplacing private keys or passwords. If these investors acquired bitcoins at very low prices in the early years, their addresses may appear to hold substantial wealth, even though the ownership is no longer accessible or economically meaningful. The blockchain does not contain explicit indicators for addresses with lost keys. However, as a proxy, we can identify addresses that have

remained inactive for many years, which may suggest that the private key has been lost.¹⁴

To ensure that our findings on ownership concentration are not driven by stale or abandoned balances, we analyze the contribution of addresses to overall concentration based on the year of their last transaction. We define a transaction as either an outflow greater than zero or an inflow whose dollar value exceeds the square root of the contemporaneous Bitcoin price (about \$10 when Bitcoin trades at \$100 and about \$300 at \$100,000). This price-adaptive threshold is intended to filter out small transactions that can be received passively, even by abandoned addresses. We exclude these small flows when determining the last time an address was actively used.

We then re-plot the ownership concentration in Figure 8, this time excluding any wallets that have not engaged in a transaction since the beginning of 2014 (red line). We choose 2014 as the cutoff year because it marks the first time Bitcoin’s price reached \$1,000. At that valuation, even an investor holding 100 bitcoins would have had an investment worth over \$100,000—an amount significant enough to provide strong incentives to safeguard private keys and make sure not to lose them.

We see the concentration among the top 1,000 whales drops from 2.6 million BTC to 2.35 million, and for the largest 10,000 whales the combined ownership drops from 5.1 million BTC to about 4.6 million. We see a greater divergence in the measured ownership concentration (blue and red lines) beyond the top 25,000 wallets, which is primarily driven by the exclusion of wallets mined by Bitcoin’s creator, Satoshi Nakamoto. Overall, the existence of stale balances does not significantly change our concentration results.

5.2.3. Benchmarking Concentration

To benchmark the ownership concentration of Bitcoin we can use as a baseline comparison the wealth distribution of the US population. [Saez and Zucman \(2020\)](#) show that as of the end of 2020, the wealth share of the top 10% households in the US is almost 80% of wealth, the top 1% hold more than 35% of wealth, and the top 0.1% hold about 16%. Wealth here is defined as the market value of all non-financial and financial assets net of all debts, excluding consumer durables and unfunded pensions. Similarly, we can look at the concentration of stock market wealth in the US. Like Bitcoin this is a risky asset and might therefore be a better comparison. According to data from the Federal Reserve Bank’s FRED database, as of the end of 2024, the top 1% of U.S. households held approximately 50% of the U.S. stock market, while the top 0.1% held around 23%. In comparison, crypto.com estimates that there were about

¹⁴Of course, some people might just hold for the long run even if they have the password.

106 million holders of Bitcoin as of December 2024. So the top 10,000 holders make about 0.01% of Bitcoin holders. As of December 20, 2024, these top 10,000 addresses held 5.1 million of the 19.8 million bitcoins that were in circulation at that point, which is about 26% of all the wealth of Bitcoin. Thus, the concentration of Bitcoin wealth is significantly higher than the wealth distribution or the concentration of stock ownership of the US population.

Finally, we want to note that our measurement of concentration represents a lower bound, since we cannot rule out that some of the largest addresses are controlled by the same entity, or that certain investors hold substantial amounts of Bitcoin on exchanges.

5.3. Whales

Given that whales hold more than a quarter of the total Bitcoin supply, their behavior may have a disproportionate impact on Bitcoin’s long-term price stability and overall price dynamics. In this section, we present what is, to the best of our knowledge, the first comprehensive study of how whales trade, establish, and exit their positions.

A common perception is that most whales are true believers and early adopters of Bitcoin who are not driven by financial motives, but by the libertarian ideal behind cryptocurrencies. If this narrative were true for most whales, we would expect that the entry and exit decisions of whales would not be sensitive to the price of Bitcoin. We document significant cross-sectional heterogeneity in the behavior of whales: A large fraction, about 50%, sells their stakes within two years of establishing their positions and another 25% within five years. Only about 20% of whales end up holding their position in the long run, the so-called hodlers. In addition, both entry and exit rates of whales are pro-cyclical and increase with the price of Bitcoin. This behavior is quite stable over time and over multiple price cycles, suggesting that fear of missing out (FOMO) or market sentiment plays an important factor in the decision of whales to participate but also to liquidate their positions. If the average whales were predominantly driven by non-financial motivations, we would not expect buying and selling behaviors to move closely with the price of Bitcoin. In particular, we might not expect them to sell when the price of Bitcoin is going up, since this might be an indication of wider Bitcoin adoption. In fact, we find that the very earliest whales exited the market much faster in response to price increases than later cohorts, which again runs counter to the view that the early adopters were not driven by financial considerations.¹⁵

The sample consists of 42,000 whales we identified in Section 5.2.1, of which 7,000

¹⁵Of course, exit decisions for the latest entrants, post 2020, are increasingly right censored.

never sold a significant fraction of their Bitcoin holdings.¹⁶ The latter are the so-called hodlers, which we define as whales that sold less than 10% throughout our sample period. We set the threshold at 10% to allow for noise generated by small amounts of rebalancing or sales that whales might undertake.¹⁷ The remainder of the whales sell at least a fraction of their position and about 67% exited their position completely at some point in our sample period. The descriptive statistics in Table 3 show that the average whale holds a balance of 425 BTC over the time they are in the sample, but there is significant skewness among whales – the median whale balance is 165 BTC, while whales at the 95th percentile hold 1396 BTC. We also break out the statistics for hodlers versus non-hodlers. Since hodlers and non-hodlers hold their bitcoin over different time periods, we report the maximum balance that a given whale ever held over the time they are in the sample. On average, the balances of hodlers are somewhat larger than non-hodlers: the median average balance of hodlers versus non-hodlers is 197 BTC versus 158 BTC. In terms of maximum balances, however, the largest whales are non-hodlers – the median max balance is 200 BTC for hodlers versus 233 BTC for non-hodlers, and at the 95th percentile of the distribution the largest non-hodlers are almost twice as large as hodlers. Finally, the average whale is in the sample for 2.5 years, conditioning on whales that exit.

5.3.1. Entry Dynamics of Whales

To shed light on when whales entered the market, in Figure 9 we plot the logarithm of the number of whales that establish their position each month, from 2009 to 2024 (blue line). Contrary to prevailing beliefs, only a small fraction of whales entered shortly after the inception of Bitcoin. Instead, we see sustained entry throughout the entire period, even in the second half of the sample when the price of Bitcoin exceeded \$10,000. The data also reveal a pronounced cyclicity in the entry behavior of whales: Entry accelerates sharply during times of run-ups in the Bitcoin price, such as mid-2011, when Bitcoin first reached \$30, 2014, 2017, and 2021. Although after 2018 periods of steep price hikes still drive increased whale entry, the peaks become more muted. It appears that many investors who want to hold Bitcoin might have already established a position by then or the price had become prohibitively high. The gray line in Figure 9 plots the deviations from the smoothed log bitcoin price to highlight

¹⁶We drop from this sample any whale addresses where we have a gap in the traceability of balances, which leaves us with 38,900 whales. We confirmed that the results in the paper hold if we include the whales that have missing data.

¹⁷The results are unchanged if we vary the threshold from 0 to 10%.

the peaks in the price.¹⁸

Hodlers (the red line in Figure 9) exhibit entry patterns that move in parallel with non-hodler whales. Like other whales, hodlers entered throughout the entire sample period and show an increased entry probability during periods of sharp price increases.¹⁹

To quantify the relationship between whale entry and the bitcoin price, we now run regressions of the log number of whales that enter each week as a function of the deviations from the smoothed price. Here again the idea is that if a majority of whales were entering this market for idiosyncratic or non-financial reasons, we would not expect entry rates to vary with the price of bitcoin. We use entry rates at the weekly level in the regression analysis to pick up finer variation in whale behavior over time, but the results are unchanged if we repeat the analysis at the monthly level.

$$\ln(\text{NumEntries}_t) = \beta_0 + \beta_1 \text{LogPriceDeviation}_t + \sum_{i=2}^3 \gamma_i \text{Cohort}_i \times \text{LogPriceDeviation}_t + \text{CohortFE} + \text{YearFE} + \varepsilon_t, \quad (1)$$

where $\text{LogPriceDeviation}_t$ is the log price of Bitcoin minus the exponentially smoothed price. Cohort Dummies are defined over three sub periods that roughly divide the data into three periods: Cohort 1 runs from 2009 to 2013-12-04, when there was a first run up in Bitcoin prices. Cohort 2 runs from 2013-12-04 to 2017-12-16, the second steep run up. And Cohort 3 starts from 2017-12-16 until the end. The idea is to be able to differentiate the behavior of early versus later entrants. We also redid all the analysis with annual cohorts and the results are qualitatively very similar.

Table 4 presents the results. Column (1) shows that the entrance of whales is positively related to the deviation from the smoothed log Bitcoin price. In fact, the price deviation explains a large fraction of the variation in entry behavior, the adjusted R-squared of the regression is 0.36. The estimated coefficient on the log price deviation suggests that entry is strongly pro-cyclical, and weeks with a 10% increase in the price deviation see a more than 4% increase in new whales entering. In Column (2), we

¹⁸Figure IA.6 in the appendix shows the weekly log BTC price, the smoothed log price and the deviation from the log price. We use a simple-exponential smoothing of the weekly log price using $\alpha = 0.01$. The implied effective look-back window is about 100 weeks, and the half-life, the point at which a past observation's weight has decayed to 50% of the current weight, is 69 weeks. This slow-moving trend captures the long-run price level, so the deviation reflects the cumulative appreciation of Bitcoin over roughly the preceding one to two years.

¹⁹The latter portion of our sample is subject to censoring, as recent hodlers may not have had sufficient time to sell their positions.

include year fixed effects as a robustness check, and the coefficient on the log price deviation does not change materially, but the R-squared increases to 0.59.

To test for differential sensitivity of entry to the Bitcoin price across cohorts, in Column (3) we add interactions of cohort dummies with the log price deviation. The omitted category is the first cohort dummy, from 2009 to 2013. The estimated coefficients on the interaction terms with cohorts 2 and 3 are negative, but only the interaction with cohort 2 is statistically significant. This pattern persists when restricting the analysis to the largest whales (those holding more than 600 BTC) in Column (4). The only difference to Column (3) is that the coefficient on the interaction term with cohort 3 is now positive but only borderline significant. Thus, whales that entered in the earliest cohort demonstrate similar or even higher sensitivity to price changes as later entrants, running counter to the hypothesis that most early adopters were not driven by financial considerations.

Columns (5)-(7) repeat this analysis for hodlers. Hodlers show similar entry dynamics and sensitivity to price deviations as other whales, and again early hodler cohorts demonstrate slightly higher price sensitivity compared to later cohorts. In Column (8) we also report the results for the 20% largest hodlers, and the results are qualitatively unchanged. Since the number of these largest hodlers that enter each week is zero in almost 40% of cohort-weeks, due to the smaller number of whales, we run this analysis as a logit specification, where the dependent variable is one if any whale enters in a given week, and zero otherwise.²⁰

These results run counter to the popular narrative that most early adopters were predominantly true believers in the technology or the libertarian ideal of Bitcoin and were not influenced by sharp changes in the price of Bitcoin. This suggests that fear of missing out (FOMO) seems to have been a significant factor in the entry of whales, similar to many other participants in the Bitcoin market, see [Kogan et al. \(2024\)](#).

5.3.2. Whales Exits

To further shed light on the role of Bitcoin prices on the behavior of whales, we now analyze how whales exit or reduce their positions as a function of the Bitcoin price. In Figure 10, we plot the fraction of whales each month that sell between 5% to 25% of their balance in a month (blue line) and those that sell more than 95% of their balance (red line). As before the gray line is the log deviation from the smoothed Bitcoin price. We see that in each month about 2% of whales sell between 5% to 25% of their Bitcoin on average, and a slightly lower fraction exits completely. But sales

²⁰We have also rerun this specification at the monthly level or only conditional on weeks with positive exits and again the results are qualitatively unchanged.

spike significantly in months when the Bitcoin price went up sharply, this is true for both partial sales, and full exits. Overall, a sizable fraction of whales tend to behave counter-cyclically and reduce their Bitcoin holdings during periods of large run ups in the price of Bitcoin.

Interestingly, we see a change in the composition of selling behavior over time. Pre-2014 many whales either exited their positions completely in response to large increases in the price of Bitcoin, or sold large stakes of more than 95%. After 2014 there is marked reduction in the fraction of whales that exit their positions completely or sell a large share after a run up in BTC prices. Instead, we see an increase in partial exits, i.e. whales selling between 5% to 25% of their positions, during a price run up. These results might suggest that early on some whales might have been worried that the Bitcoin phenomenon would be short-lived and tried to realize their profits as soon as there was a price run up. In later periods, whales seem to have been more confident in the survival of Bitcoin and did not liquidate their entire position but only sold parts of their holdings in response to a price spike.

We now confirm the intuition from Figure 10 with a regression analysis of the fraction of whales selling above 95% of their holdings in each month on the deviation from the bitcoin price. The data is constructed at the cohort x week level, which allows us to analyze any differences in the exit behavior across cohorts. This is naturally an unbalanced panel since later cohorts only started later and thus can only exit at later points.

The results in Table 5 Columns (1) and (2) confirm a strong positive correlation between the exit rate of whales and the deviation from the log bitcoin price, whether we control for year fixed effects or not. In Column (3) we then interact the weekly price deviation with our cohort fixed effects, which are defined as before. The omitted category is cohort 1. We see that the estimated coefficients on the interaction terms of the price deviation with cohort 2 as well as cohort 3 are negative and significant. But the coefficient is much larger for cohort 3 suggesting that these whales have essentially zero residual price sensitivity for full exits. In Columns (4) and (5) we then repeat the analysis for the largest 20% of the whales and the results are qualitatively similar. Overall, this analysis confirms that a significant fraction of whales exits when the price of Bitcoin is going up quickly, thus are trying to be contrarian. But the sensitivity to price deviations is slightly lower for whales that entered post 2018. Whales seem to try and time the market to maximize their returns, which again suggests that financial and trading motives play a large role, even for the earliest cohorts.

Could part of this reduced sensitivity to the price of Bitcoin in the post-2018 cohorts be explained by the fact that a larger fraction of these whales are right-censored and

thus are still waiting to exit? To rule out this concern, we measure for each month the fraction of entering whales that sell their position within two years from entry, 2 to 5 years, 5 to 8 and more than 8 years from entering. Figure [IA.8](#) in the Internet Appendix shows that the fraction of whales that sell within two years of entry varies with the price of Bitcoin, i.e. there is a much larger fraction of quick exits during times of a price run ups. But the overall composition of whales that are willing to exit quickly versus hold for longer does not change markedly over the entire period that we can observe.

Since the price of Bitcoin has seen a large appreciation as well as significant volatility over time, whales entering at different times see very different price paths, and potentially different returns on their holdings. One might wonder if this has changed the cumulative returns at which whales on average are willing to sell their holdings. To provide a simple metric for this behavior we calculate for each whale that exits more than 25% of their holdings in a given month the realized return, as the price at exit divided by the cost at which the whale entered the position. For most whales the denominator is simple, since they established their position once or with a few purchases in the same month. We then plot the mean and median log realized return for the whales that exit in each month in Figure [11](#).²¹ Here again we see significant stability in the realized returns at which whales exit over each pricing cycle. While the monthly returns vary with the changes in the price of Bitcoin the average and median realized returns at each peak and trough of the Bitcoin price remain similar over time.

6. Conclusions

Our analysis shows that the Bitcoin ecosystem has become dominated by a small number of large participants across mining, ownership, and transaction activity. This reality contrasts with the decentralized vision that initially inspired its creation. We document that just 50 miners control roughly half of the network’s mining capacity, while 0.01% of investors hold over a quarter of the circulating supply. Furthermore, the largest trading venues not only generate approximately 70% of transaction volume, but they are also the most connected nodes on the blockchain.

We show that a high degree of concentration and interconnectedness can create important challenges for the transparency of the system and the implementation of KYC and AML regulations. Similarly, the concentration of miners and whales increases

²¹We also repeat this analysis only focusing on whales that exit more than 50% or 95% of their holdings and the results are qualitatively the same, but the level of realized cumulative returns is slightly larger for higher exits.

the dependence of the system on a small set of decision makers and with it the potential for systemic risk.

References

- Abadi, J. and Brunnermeier, M. (2018). Blockchain economics. Working Paper 25407, National Bureau of Economic Research.
- Aloosh, A. and Li, J. (2024). Direct evidence of Bitcoin wash trading. *Management Science*, 70(12):8875–8921.
- Antonopoulos, A. M. and Harding, D. A. (2023). *Mastering Bitcoin*. O’Reilly Media, 3rd edition.
- Athey, S., Parashkevov, I., Sarukkai, V., and Xia, J. (2016). Bitcoin Pricing, Adoption, and Usage: Theory and Evidence. Research Papers 3469, Stanford University, Graduate School of Business.
- Biais, B., Bisiere, C., Bouvard, M., and Casamatta, C. (2019). The blockchain folk theorem. *The Review of Financial Studies*, 32(5):1662–1715.
- Budish, E. (2025). Trust at scale: The economic limits of cryptocurrencies and blockchains. *The Quarterly Journal of Economics*, 140(1):1–62.
- Cong, L. W., Harvey, C. R., Rabetti, D., and Wu, Z.-Y. (2025). An anatomy of crypto-enabled cybercrimes. *Management Science*, 71(4):3622–3633.
- Cong, L. W., He, Z., and Li, J. (2020). Decentralized Mining in Centralized Pools. *The Review of Financial Studies*, 34(3):1191–1235.
- Cong, L. W., Li, X., Tang, K., and Yang, Y. (2023). Crypto wash trading. *Management Science*, 69(11):6427–6454.
- Easley, D., O’Hara, M., and Basu, S. (2019). From mining to markets: The evolution of bitcoin transaction fees. *Journal of Financial Economics*, 134(1):91–109.
- Ferreira, D., Li, J., and Nikolowa, R. (2023). Corporate capture of blockchain governance. Technical Report 4.
- Foley, S., Karlsen, J. R., and Putniņš, T. J. (2019). Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies? *The Review of Financial Studies*, 32(5):1798–1853.
- Graf von Luckner, C., Reinhart, C. M., and Rogoff, K. (2023). Decrypting new age international capital flows. *Journal of Monetary Economics*, 138:104–122.

- Griffin, J. M. and Shams, A. (2020). Is bitcoin really untethered? *The Journal of Finance*, 75(4):1913–1964.
- Halaburda, H., Haeringer, G., Gans, J., and Gandal, N. (2022). The microeconomics of cryptocurrencies. *Journal of Economic Literature*, 60(3):971–1013.
- Huberman, G., Leshno, J. D., and Moallemi, C. (2021). Monopoly without a monopolist: An economic analysis of the Bitcoin payment system. *The Review of Economic Studies*, 88(6):3011–3040.
- Härdle, W. K., Harvey, C. R., and Reule, R. C. G. (2020). Understanding cryptocurrencies. *Journal of Financial Econometrics*, 18(2):181–208.
- Khalilov, M. C. K. and Levi, A. (2018). A survey on anonymity and privacy in bitcoin-like digital cash systems. *IEEE Communications Surveys & Tutorials*, 20(4):2543–2585.
- Kogan, S., Makarov, I., Niessner, M., and Schoar, A. (2024). Are cryptos different? evidence from retail trading. *Journal of Financial Economics*, 159:103897.
- Lewis-Pye, A. and Roughgarden, T. (2023). Permissionless consensus. *arXiv preprint arXiv:2304.14701*. arXiv:2304.14701 [cs.DC].
- Liu, X. F., Jiang, X.-J., Liu, S.-H., and Tse, C. K. (2021). Knowledge discovery in cryptocurrency transactions: A survey. *IEEE Access*, 9:37229–37254.
- Makarov, I. and Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2):293–319.
- Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G. M., and Savage, S. (2013). A fistful of bitcoins: Characterizing payments among men with no names. In *Proceedings of the 2013 Conference on Internet Measurement Conference*, IMC ’13, page 127–140, New York, NY, USA. Association for Computing Machinery.
- Möser, M. and Narayanan, A. (2022). Resurrecting address clustering in bitcoin. In *Financial Cryptography and Data Security*, pages 386–403, Cham. Springer International Publishing.
- Prat, J. and Walter, B. (2021). An equilibrium model of the market for bitcoin mining. *Journal of Political Economy*, 129(8):2415–2452.

- Romiti, M., Judmayer, A., Zamyatin, A., and Haslhofer, B. (2019). A deep dive into Bitcoin mining pools: An empirical analysis of mining shares. *arXiv preprint arXiv:1905.05999*. Presented at WEIS 2019, arXiv:1905.05999 [cs.CR].
- Ron, D. and Shamir, A. (2012). Quantitative analysis of the full bitcoin transaction graph. *IACR Cryptology ePrint Archive*, page 584.
- Saez, E. and Zucman, G. (2020). The rise of income and wealth inequality in america: Evidence from distributional macroeconomic accounts. *Journal of Economic Perspectives*, 34(4):3–26.
- Tasca, P., Hayes, A., and Liu, S. (2018). The evolution of the bitcoin economy: Extracting and analyzing the network of payment relationships. *The journal of risk finance*, 19(2):94–126.
- Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., and Leiserson, C. E. (2019). Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. *arXiv preprint arXiv:1908.02591*.
- Wu, J., Liu, J., Zhao, Y., and Zheng, Z. (2020). Analysis of cryptocurrency transactions from a network perspective: An overview. *arXiv preprint arXiv:2011.09318*.

Figures

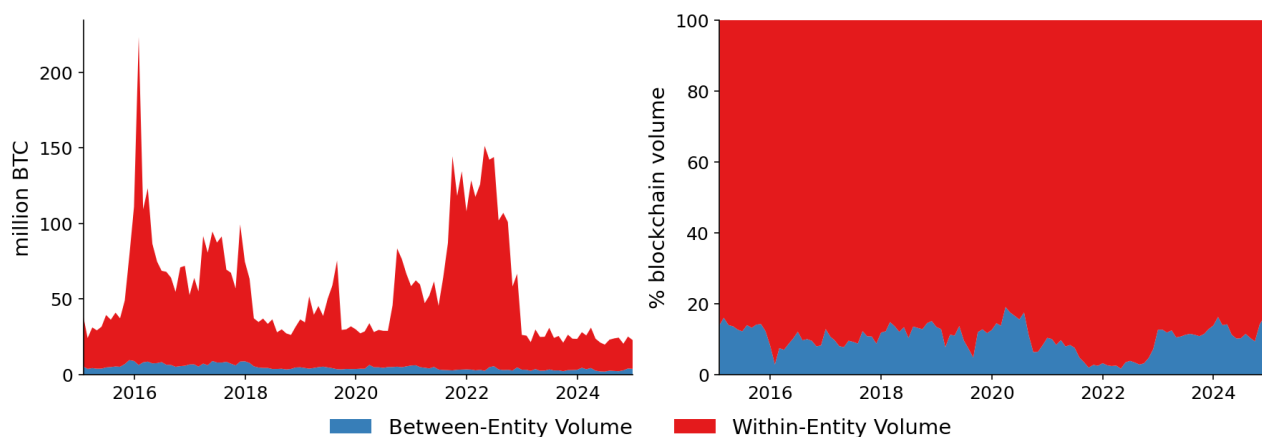


Figure 1: Decomposition of volume: Between- vs. within-entity volume.

The figure shows the decomposition of total Bitcoin blockchain volume at the monthly level into two components. The top (red) part shows within-entity volume, which is the volume that is generated when an entity sends bitcoins to itself. The bottom part (blue) is between-entity, which represents transfers between clusters controlled by different users.

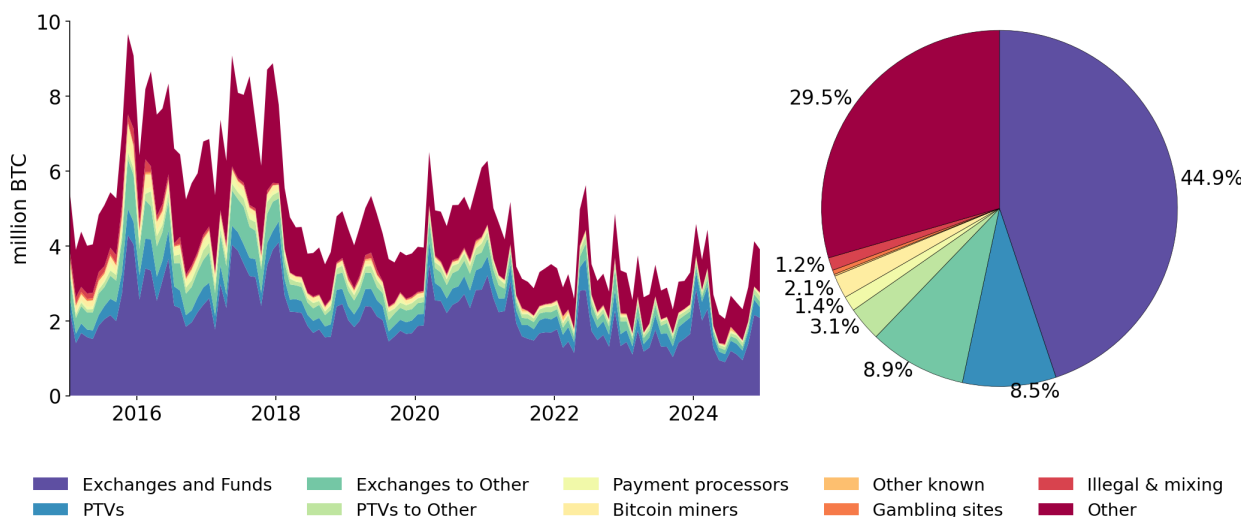


Figure 2: Decomposition of between-entity volume.

The figure shows the monthly volume generated by different entities on the blockchain, measured as the amount of bitcoins that are sent to different types of entities in a given month, from January 2015 to December of 2024. The panel on the left shows the volume in BTC. The panel on the right shows the percentage of average volume of each entity over the entire sample period. PTVs are probable trading venues. Other represents addresses that are unclassified. We break out volume to Other if it is generated by exchanges or PTVs.

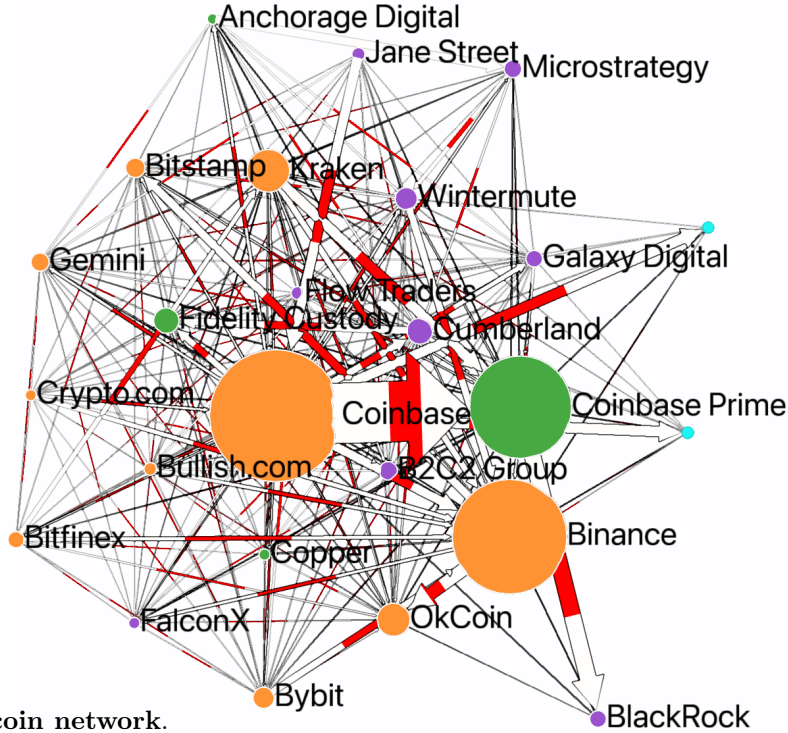


Figure 3: Bitcoin network.

The figure shows the bitcoin network for the top 25 entities with the largest volume in 2024. This is a directed, weighted network graph, where a node i corresponds to cluster i and a link from i to j corresponds to the total Bitcoin flows over the year 2024 from cluster i to j . Node size is proportional to total volume received by the entity; link width is proportional to the volume carried by that link. When two clusters send flows to each other, the link is drawn in the direction of the larger flow and marked with a red segment. Orange nodes signify exchanges, purple nodes are institutional investors, green nodes are custodians, and turquoise are unknown entities.

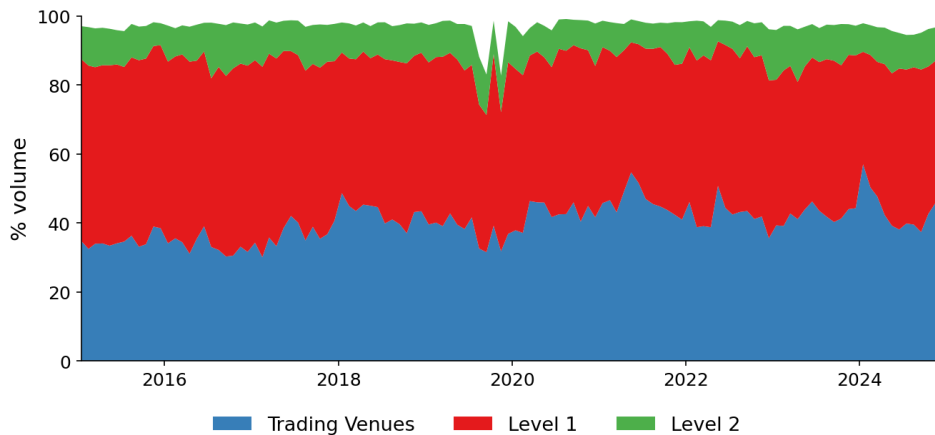


Figure 4: MiCA regulation.

The figure presents a hypothetical decomposition of transaction volume under the assumption that Markets in Crypto-Assets (MiCA) regulations were applied to all identified trading venues in our sample. The blue area shows the volume to the known trading venues. Level 1 refers to volume of clusters that directly send or receive bitcoins from known trading venues. Level 2 refers to volume associated with clusters that transact directly with the Level 1 clusters.

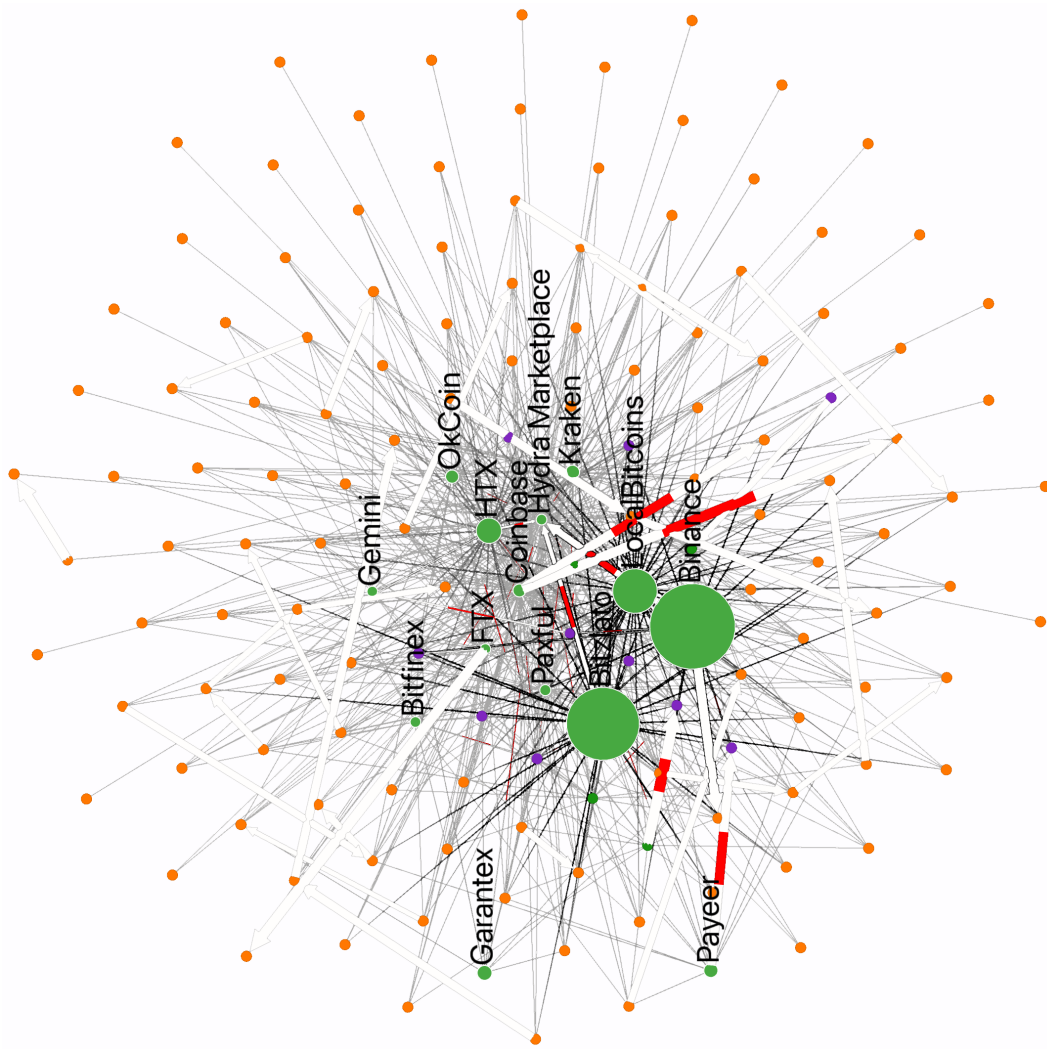


Figure 5: Hydra market network.

The figure shows the clusters in the Bitcoin network that are connected to Hydra Market, one of the largest darkweb sites. We include only entities that send at least 10,000 BTC within the network over the 2020-2021 period with transfers to and from exchanges and PTV excluded. The node size reflects the sum of the total amount of bitcoins sent to and received from Hydra Market. The link size is proportional to the volume between two different entities. In cases where two clusters send flows to each other, the direction of the link between clusters is based on the largest flow, and the link is depicted with a red segment. Green nodes signify identified clusters, purple nodes are unknown clusters, and orange depict short-lived clusters with a life-span below one week.

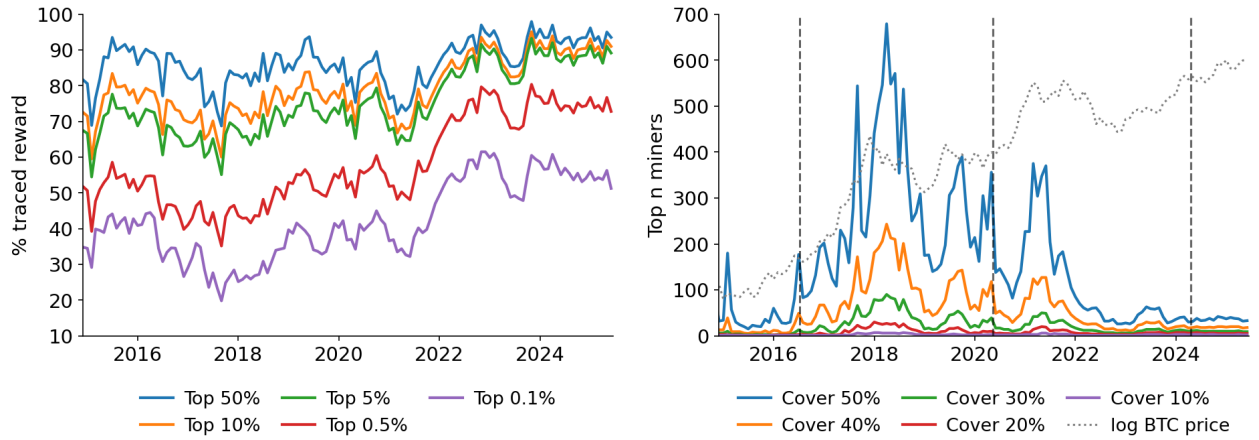


Figure 6: Concentration of mining capacity.

This figure documents the concentration of mining capacity based on Coinbase rewards that miners receive from pools. Each month, we sort active miners by the amount of Coinbase rewards they receive and calculate the percentage of total mining capacity controlled by different quantiles of the miner distribution. The left panel shows the results for the top 50%, 10%, 5%, 0.5%, and 0.1% miners. The right panel shows the number of miners that are necessary to cover 10%, 20%, 30%, etc. of total mining capacity. The dashed lines indicate Bitcoin halving dates. The dotted line shows the log-price of Bitcoin in USD, scaled to fit the plot.

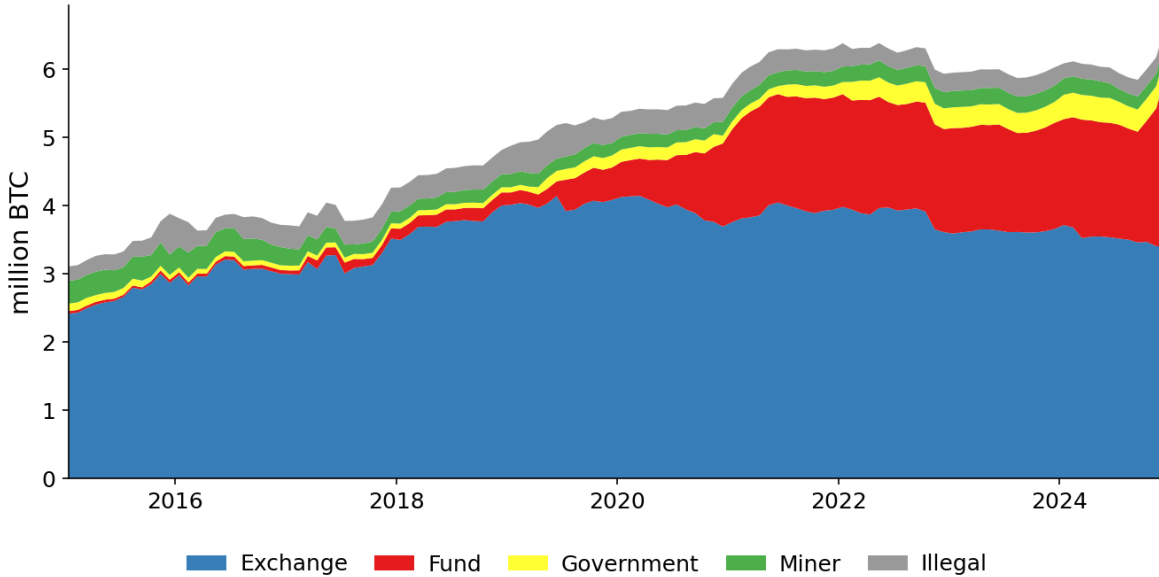


Figure 7: Bitcoin ownership.

The figure shows the amount of Bitcoin held in the wallets of identified entities, from Jan 2015 until Dec 2024. The blue area depicts Bitcoin held by wallets of exchanges, red by funds such as ETFs, yellow by government, mostly a result of seized coins, green by miners and grey by illegal entities.

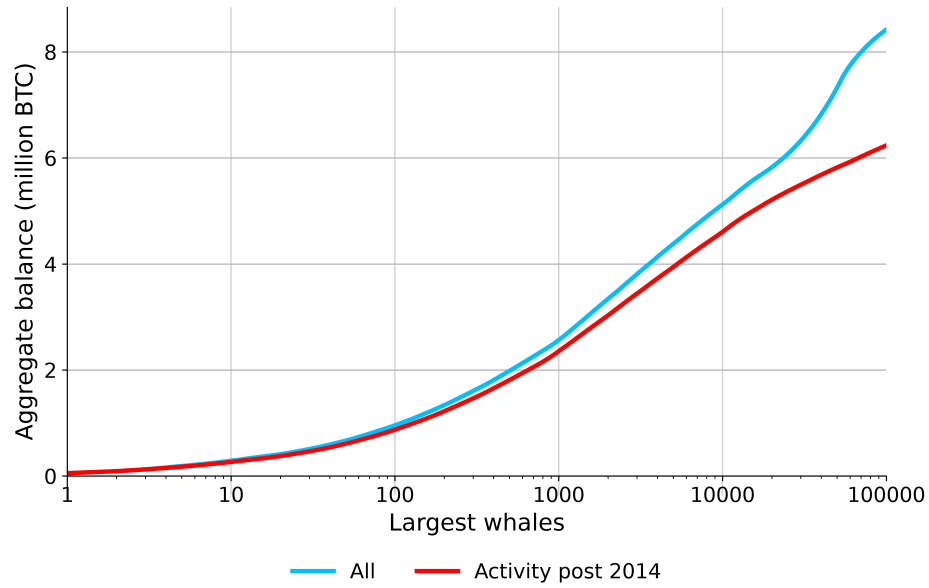


Figure 8: Ownership concentration of individual addresses.

The blue line shows the concentration of Bitcoin held by largest individual holders, “whale”. We sort individual wallets based on their balance at the end of 2024, and plot their cumulative balance against the number of individual wallets holding these bitcoins. The red line excludes any wallets that have not engaged in a transaction since the beginning of 2014 to filter out potentially stale wallets.

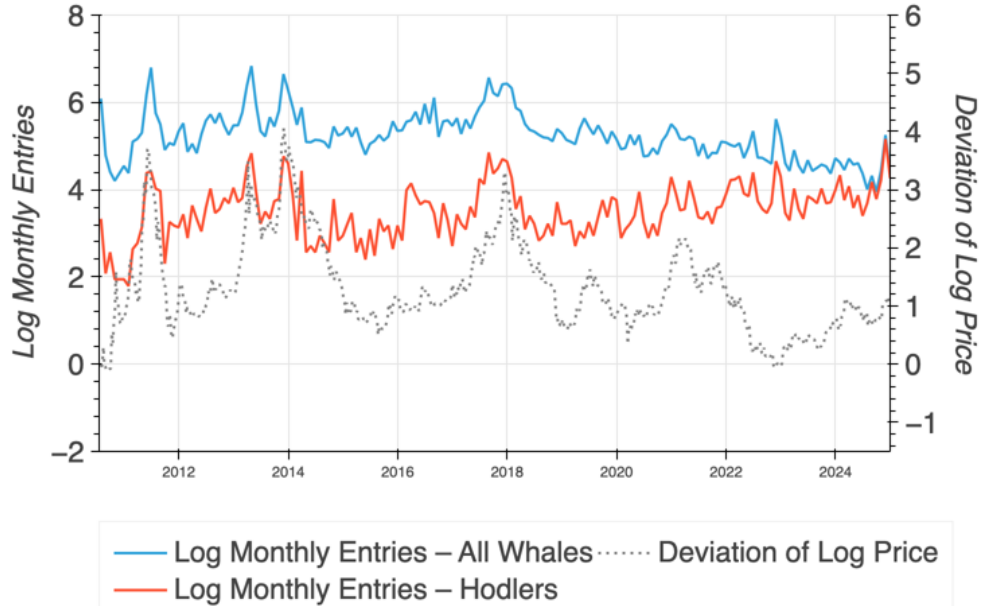


Figure 9: Whales Entries.

The figure shows the log of the number of whales entering the sample in a given month, between 2010–2025. The blue line counts all whales, while the orange line restricts to those that still hold Bitcoin at the end of the sample (hodlers). The grey dotted series shows the deviation of the log BTC price from its exponential-smoothed trend (computed on weekly prices with $\alpha = 0.01$; see Figure [IA.6](#)) to allow visual comparison of entry and Bitcoin price cycles.

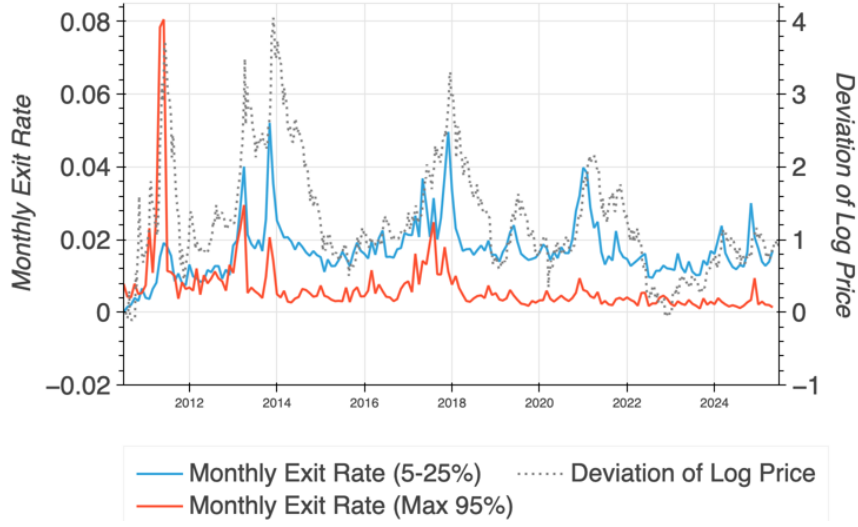


Figure 10: Whales Monthly Exits Large vs. Small.

This figure documents the monthly exit behavior of whales. The red line plots the fraction of whales in a month that reduce their holdings by more than 95% relative to their maximum balance to indicate full exits. The blue line plots the share of whales each month that sell between 5-25% of their balance relative to the end-of-previous-month balance (partial exits). The grey dotted series (right axis) reproduces the log-price deviation from the weekly smoothed trend (see Figure IA.6).

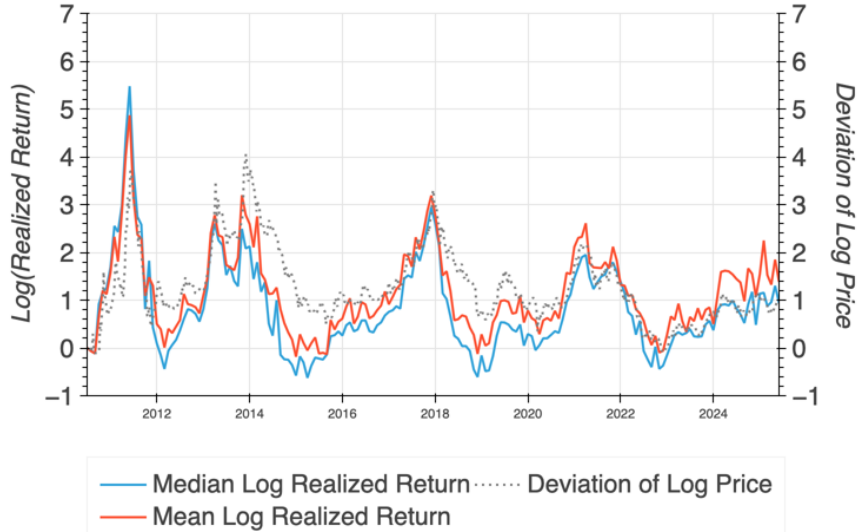


Figure 11: Whales Monthly Realized Return.

This figure shows the monthly median and mean of the logarithm of realized returns of all whales that sell their holdings in a given month. We include any sale in which a whale sells more than 25% of its holding compared to its peak balance. We compute the realized return as $\log(\text{sale price}/\text{volume-weighted average acquisition price})$. The grey dotted line represents the log-price deviation from the weekly smoothed trend used elsewhere (see Figure IA.6).

Tables

Group	Entity count	Cluster count	Address count
Exchanges	1,578	435,585	183,997,673
Gambling	594	8,640	23,291,568
Government	7	1,255	1,880
Illegal	6,171	38,802	26,170,995
Investment	104	10,326	1,470,717
Miner	161	342,692	1,005,428
Other	676	4,681	2,040,697
Payment	332	6,836	52,349,400
Scams	4,128	16,026	4,618,468

Table 1. Address types and counts.

This table reports summary statistics of the number of identified entities, clusters, and addresses across six prominent entity categories.

Exchange	Corr	Obs	Vol ratio
Bitstamp	0.92	96	1.48
Bittrex	0.86	85	0.43
Coinbase	0.31	96	1.60
Gemini	0.69	96	1.03
Kraken	0.56	96	1.33
Poloniex	0.85	96	1.41
Zaif	0.75	83	14.61
bitFlyer	0.88	96	7.64
Binance	0.43	89	7.28
Bithumb	0.60	89	5.08
Huobi	0.83	84	5.15
Coinone	0.65	82	7.24
Upbit	0.49	70	5.88
Kucoin	0.71	69	13.30
FTX	0.96	35	4.03

Table 2. On- and off-chain exchange volume.

This table reports the monthly time-series correlation between off-chain and on-chain volumes for the largest 15 Bitcoin exchanges, along with the average ratio of off-chain to on-chain volume. Obs is the number of monthly observations for each exchange; Corr the monthly time-series correlation between off-chain and on-chain volume, and Vol ratio the average ratio of off-chain to on-chain volume.

Statistic	Count	Mean	SD	5%	25%	50%	75%	95%
Average balance (all whales)	38.9k	425.50	1256.12	49.62	101.00	164.71	344.07	1395.70
Average balance (hodlers)	6.9k	438.55	1129.08	99.87	117.13	196.64	400.00	1294.05
Max balance (hodlers)	6.9k	455.07	1157.00	100.00	122.00	200.00	405.38	1370.77
Average balance (non-hodlers)	32.0k	422.70	1281.68	44.67	99.78	158.45	330.99	1422.21
Max balance (non-hodlers)	32.0k	701.28	3123.28	101.59	144.37	233.11	503.01	2288.93
Avg. duration (exited)(months)	25.9k	28.68	24.17	6.66	11.38	20.59	37.97	77.69
Freq Selling >5% Max Bal (non-hodlers)	32.0k	2.67	2.24	1.00	1.00	2.00	4.00	7.00

Table 3. Summary Statistics of Bitcoin Whales.

This table reports the Average Bitcoin balance that whales held on the blockchain, reported for all whales and those that never sold (hodlers) and those who sold. Max balance is the maximum balance a whale held at any point over the sample period. Avg Duration (exited) is the number of months on average that whales were in sample before exiting, conditional on those that exited. Freq. Selling is the number of times that whales sold more than 5% of their balance.

	All	All	All	Top 20 All	Hodler	Hodler	Hodler	Top 20 Hodler
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(Price Deviation)	0.441*** (0.024)	0.413*** (0.036)	0.487*** (0.044)	0.323*** (0.044)	0.166*** (0.036)	0.409*** (0.050)	0.499*** (0.053)	1.047*** (0.217)
Cohort 2			0.543*** (0.097)	-0.079 (0.116)			0.369** (0.157)	0.047 (0.467)
Cohort 3			-0.072 (0.094)	-0.733*** (0.098)			1.084*** (0.125)	1.809*** (0.404)
Log(Price Deviation)xCohort 2			-0.263*** (0.057)	-0.118* (0.069)			-0.250*** (0.091)	-0.517* (0.284)
Log(Price Deviation)xCohort 3			-0.062 (0.060)	0.116* (0.068)			-0.559*** (0.075)	-1.492*** (0.280)
Year Effects	No	Yes	No	No	No	Yes	No	No
Cohort Effects	No	No	Yes	Yes	No	No	Yes	Yes
Reg Type	OLS	OLS	OLS	OLS	OLS	OLS	OLS	Logit
Observations	755	755	755	755	755	755	755	755
Adjusted (Pseudo) R^2	0.360	0.591	0.428	0.336	0.032	0.330	0.147	0.053

*p<0.1; **p<0.05; ***p<0.01

Table 4. Whales Entries and BTC Price.

This table reports results from a regression of the log monthly number of new whales establishing a wallet on the blockchain on the deviations of the Bitcoin price from its exponentially smoothed trend. Columns 1–4 use the full sample of whales. Column 4 repeats the specification for the subsample of the largest 20% of whales by peak holdings (>600 BTC). Columns 5–8 report the same specifications for the subset of whales that hold their Bitcoin till the end of the sample (hodlers). In Column 8 we estimate the relationship for the largest 20% of whales, but we a logit model, where the dependent variable is a dummy equal to 1 if there is a least 1 hodler entering in a month, 0 otherwise. Robust standard errors are reported in parentheses; significance levels are denoted by asterisks.

	All	All	All	Top 20	Top 20
	(1)	(2)	(3)	(4)	(5)
Log(Price Deviation)	0.087*** (0.016)	0.101*** (0.034)	0.140*** (0.044)	0.041** (0.018)	0.056*** (0.018)
Cohort 2			0.142*** (0.029)		0.086*** (0.023)
Cohort 3			0.288*** (0.031)		0.207*** (0.024)
Log(Price Deviation)xCohort 2			-0.060** (0.025)		-0.014 (0.016)
Log(Price Deviation)xCohort 3			-0.140*** (0.029)		-0.080*** (0.028)
Year Effects	No	Yes	Yes	Yes	Yes
Cohort Effects	No	No	Yes	No	Yes
Observations	1773	1773	1773	1773	1773
Adjusted R^2	0.079	0.244	0.312	0.091	0.123

*p<0.1; **p<0.05; ***p<0.01

Table 5. Whales exits and price.

This table estimates the relationship of whale exits on the deviation of the Bitcoin price from its exponentially smoothed trend. The dependent variable is the weekly number of whales that sell more than 95% of their stake. The data are constructed at the cohort $\times$ weekly cell. Columns 1–3 use the full sample of whales and progressively introduce calendar-year effects and fixed effects for entry cohorts (plus cohort-specific interactions with the log-price deviation). Columns 4 and 5 replicate the Columns (2) and (3) for the subsample comprising the top-20% of whales by peak holdings (>600 BTC). Robust standard errors are reported beneath coefficient estimates; significance levels are denoted by asterisks.

Internet Appendix:

In this Appendix, we provide a step-by-step description of the data construction process.

Step 1: Obtaining raw data The blockchain data are downloaded using the open-source software of Bitcoin Core and further parsed into a database amenable for analysis using the BlockSci program.²² As of June 13, 2025, there are 901,128 blocks of 1.2 billion Bitcoin transactions and 1.78 billion addresses.²³

Step 2: Address clustering Address clustering involves several steps. In the first step, we use the built-in union-find clustering algorithm. Applying this algorithm to the data yields 959 million clusters. Being the most conservative clustering heuristics, the multi-input algorithm tends to undercluster addresses. We correct clustering mistakes in later steps by merging additional clusters.

Step 3: Bitcoin flow database To speed up computations, we build a database of Bitcoin flows between the clusters produced by the union-find algorithm. Each transaction is represented by a list of records, each consisting of the block ID, transaction ID, the sending cluster ID, the receiving cluster ID, and the amount of bitcoins received. The coinbase transactions are recorded as the flows from a special cluster with an ID of -1 . Since the multi-input algorithm assigns all sending addresses to the same cluster, one transaction can be represented by $n + 1$ records, where n is the number of recipient addresses. For CoinJoin transactions, we split each record into m copies, where m is the number of distinct clusters on the sending side. The amount of bitcoins received by each cluster is split proportionally to the bitcoins contributed by each sending cluster.

Step 4: Identifying pass-through clusters Many Bitcoin clusters have a very short lifespan and are therefore unlikely to represent stand-alone or economically independent entities. In what follows, we call these clusters pass-through clusters. These types of pass-through addresses are often created by wallet programs or are part of a user’s attempt to either consolidate their Bitcoin addresses or spread their holdings among several addresses.

We differentiate between two types of pass-through clusters. We classify a cluster as a pass-through cluster of the first type if the following four conditions are satisfied.

²²Bitcoin Core and BlockSci are available at <https://bitcoin.org/en/bitcoin-core/> and <https://github.com/citp/BlockSci>, respectively.

²³The number of addresses depends on the definition of equivalent addresses. See https://citp.github.io/BlockSci/reference/addresses/equiv_address.html for more details.

1. The cluster has only one incoming transaction and one outgoing transaction;
2. The cluster has a balance less than 546 Satoshi after the two transactions;
3. The two transactions are fewer than 1068 blocks apart on the blockchain (less than a week apart);
4. The incoming transaction is not a CoinJoin transaction.

There are 510 million pass-through clusters of the first type in the data, making them the most common type of clusters on the Bitcoin blockchain.

We define a second type of pass-through clusters by relaxing the first condition to allow for multiple incoming and outgoing transactions. This accommodates more complex obfuscation patterns while still capturing their transient nature. As with the first type, we require that the first and the last transaction are fewer than 1068 blocks apart and that a balance less than 546 Satoshi remains after the last transaction. This second type accounts for an additional 53 million clusters in our dataset and captures entities that serve primarily as temporary holders rather than final destinations for bitcoin.

Step 5: Factoring out pass-through clusters We build two versions of Bitcoin flow database where we factor out volume associated with pass-through clusters. In the first version, we only factor out type-1 clusters. In the second version, we factor out both type-1 and type-2 clusters.

Factoring the volume of type-1 clusters is straightforward. For a non-CoinJoin transaction, the first condition ensures (with the multi-input clustering algorithm) that the pass-through cluster receives its flows from a single cluster. Therefore, to eliminate this cluster and reassign its volume, we simply assign flows from the cluster that sends to the pass-through cluster to the clusters that receive flows from the pass-through cluster.

Factoring the volume of type-2 clusters that receive flows from a single cluster is also straightforward and follows the same steps as in the case of type-1 clusters.

Refactoring volume through type-two pass-through clusters that receive flows from more than one cluster cannot in general be done unambiguously. In our analysis, we assume that outflows to downstream clusters are allocated in proportion to the cumulative inflows received from each upstream cluster at the time the outflow occurs, processing all transactions in chronological blockchain order. To illustrate, suppose pass-through cluster P first receives w_1 bitcoins from cluster A_1 , then sends v_1 and v_2

to clusters B_1 and B_2 , then receives w_2 from cluster A_2 , and finally sends v_3 and v_4 to clusters B_3 and B_4 (where B_3 and B_4 may or may not coincide with B_1 or B_2).

Because A_1 is the sole funding source when the first two outflows occur, the algorithm assigns v_1 and v_2 entirely as flows from A_1 to B_1 and B_2 , respectively. It then computes the residual balance attributable to A_1 : $b_1 = w_1 - (v_1 + v_2)$. When P next sends v_3 and v_4 , two funding sources are present, so each outflow is split in proportion to the balances from A_1 and A_2 . Specifically, the attributed flow from A_1 to B_j is $v_j \times b_1 / (b_1 + w_2)$, $j = 3, 4$, and the attributed flow from A_2 to B_j is $v_j \times w_2 / (b_1 + w_2)$, $j = 3, 4$.

More generally, consider any outgoing transaction that sends v_i to downstream cluster B_i , $i = 1, \dots, N$. The algorithm first computes the residual balance b_j held on behalf of each upstream cluster A_j , $j = 1, \dots, M$. It then attributes volume from A_j to B_i as

$$\frac{b_j}{\sum_{k=1}^M b_k} \times v_i.$$

The example above corresponds to the special case $M = 2$ and $N = 2$, with $b_1 = w_1 - (v_1 + v_2)$ and $b_2 = w_2$.

Step 6: Cluster labels To link clusters to real entities, we use an external table, *Bitcoin Entities*, to assign name labels to clusters. The table contains a list of addresses associated with identified entities. The information in this tables comes from various sources described in Section 2.2.1. To save on space, this table only stores a set of representative entity addresses which correspond to the clusters produced by the union-find algorithm.²⁴

Step 7: Coinbase signature data To track miner rewards, we start by identifying which blocks were mined by each mining pool. We obtain this information from two sources. First, we leverage the common practice where mining pools include their name in the coinbase message of the blocks they mine. We acquire the coinbase message data for each block from Flipside and process these messages to find the pool-block correspondence.

Second, we supplement this data with pool-block information scraped from BTC.com. This helps improve pool coverage in earlier samples where block messages do not clearly indicate the mining pool responsible for the block. Using this information, we obtain

²⁴The clustering outcome from the multi-input algorithm depends on the blockchain size. Since more data make it more likely to for a given address to be reused with other addresses, clusters obtained from the union-find algorithm tend to increase in size with the size of the Bitcoin blockchain. As a result, some entity addresses in the table can correspond to the same clusters.

each pool’s coinbase reward collection addresses.

Step 8: Finding addresses of individual miners Each pool employs its own distribution algorithm. Most pools distribute their rewards from a set of special distribution addresses. We find these addresses by examining the distribution pattern of each pool. Figure [IA.1](#) shows an example of the distribution tree in the case of AntPool.

Since pools employ many miners it is usually impossible to distribute rewards to all miners in one transaction. Therefore, many pools use long peeling chains to accomplish this task. The distribution of the rewards starts from a designated distribution address. It distributes the rewards to a large number of miners; collects the change in a new one-off address that distributes the reward to the next set of miners, and so on.

In the next step, we take all distribution transactions and collect all output addresses that take part in these transactions. Occasionally, some pools use distribution addresses for other purposes, possibly buying equipment or the like. Therefore, we eliminate from this set of addresses any “internal” addresses that belong to the pool. The remaining addresses are candidates for addresses of individual miners. There are a total of 1.74 million of such addresses. To eliminate “recreational” miners, we filter out addresses that receive rewards with an equivalent value of less than \$1,000 or that have fewer than 25 reward distributions over the entire sample period, yielding 403,510 miners.

We use these tracing results to compile the miner database, *Miners*, which contains individual miners’ addresses, their corresponding clusters, and time-series of their pool mining rewards. We also use the tracing results to update the *Bitcoin Entities* database with mining pool addresses.

Step 9: Large Bitcoin investors Next, we identify and trace Bitcoin ownership of whales. As described in the main text, we exclude pass-through clusters and all clusters linked to known institutional entities. To reduce computational burden, we restrict our analysis when tracing flows to the clusters that have received at least one BTC in one transaction output and study bitcoin flows larger than 0.1 BTC to and from these clusters. There are 18.6 million of such clusters, which are not pass-through clusters and named entities. We denote the set of these clusters by $\mathcal{R}$.

Next, we select clusters that ever had a maximum balance of at least 100 BTC, and then apply three additional filters to exclude large clusters that are likely controlled by institutional investors or are very short-term traders:

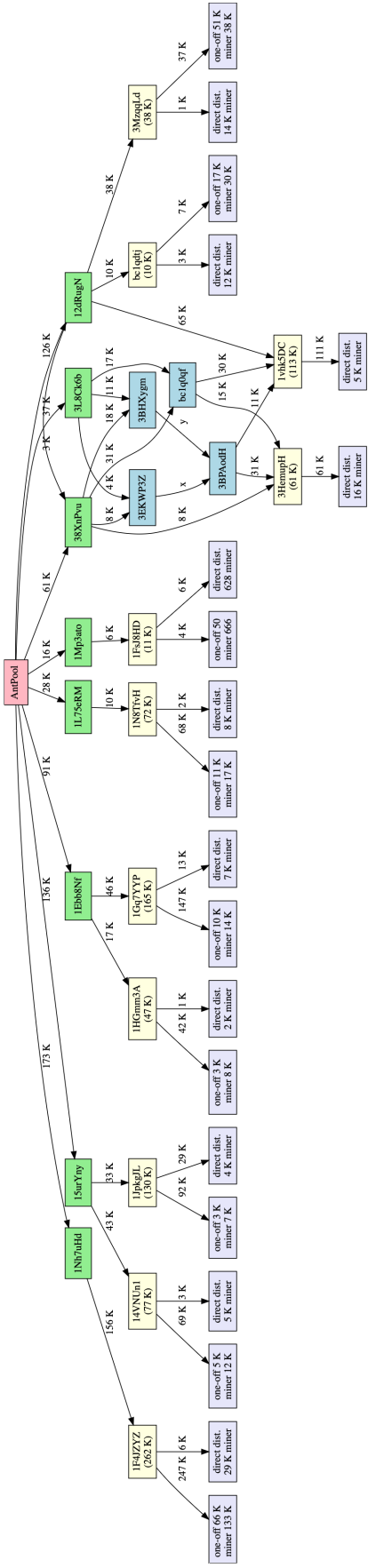


Figure IA.1: Reward distribution tree of AntPool. Bitcoin addresses are represented by the first seven letters in green and yellow boxes. Miners are summarized in blue boxes. Miners are grouped by how they receive reward. Some miners receive a direct distribution from the reward distribution addresses. Some miners receive from one-off distribution addresses that trace back to the frequently-used distribution addresses. Numbers next to arrows represent the Bitcoin volume in BTC. For reward distribution addresses, only their largest source of bitcoins is connected to them with an arrow. Flows of smaller volume are omitted from the plot.

1. The time when the cluster first holds 1 BTC to the time when it last holds 1 BTC should be more than 30 days.
2. A cluster should have fewer than 100 addresses.
3. A cluster should receive and send flows to fewer than 50 unknown clusters.

Applying these criteria yields a candidate subset $\mathcal{P}$ of rich clusters consisting of approximately 162 thousand clusters.

Next, we implement the ownership tracing algorithm, where we trace outflows from each of the candidate clusters forward in time starting from the earliest clusters and evaluate the recipient clusters that the bitcoins are sent to. If a recipient cluster receives all its inflows from the original cluster, we assume it is part of the same ownership entity.

Our goal is twofold: (1) to find a minimal subset of clusters, $\mathcal{P}_{min} \subseteq \mathcal{P}$ such that no cluster in $\mathcal{P}$ receives all its flows from another cluster in $\mathcal{P}_{min}$; and (2) for each cluster $P_i \in \mathcal{P}_{min}$, to identify a set of clusters $\mathcal{C}_i$ such that all flows to any cluster C in $\mathcal{C}_i$ can be traced back to P_i . This is a computationally intensive problem, as bitcoins can travel from P_i to C along multiple paths and through various clusters. Therefore, it is important to restrict a set of clusters that can belong to $\mathcal{C}_i$ for some i . Henceforth, we will refer to any cluster $P_i \in \mathcal{P}_{min}$ as a root cluster and any cluster $C \in \mathcal{C}_i$ as a child cluster.

Let $\mathcal{C} = \cup_i \mathcal{C}_i$ be a set of child clusters. Since the set $\mathcal{P}$ does not have any known clusters, no child cluster can receive flows from known clusters. Furthermore, no child cluster can receive flows from clusters that receive flows from known clusters, and so on.

Formally, denote the set of clusters in $\mathcal{R} - \mathcal{P}$ that receive flows from known clusters by $\mathcal{NC}_1$. Define the set of clusters $\mathcal{NC}_{j+1}$, for $j = 1, 2, \dots$ as the clusters in $\mathcal{R} - \mathcal{P}$ that receives flows from $\mathcal{NC}_j$. Let $\mathcal{NC} = \cup_j \mathcal{NC}_j$. Then $\mathcal{C} \subseteq \mathcal{R} - \mathcal{NC}$.

We sequentially find sets $\mathcal{NC}_j$, $j = 1, 2, \dots$. After 1000 iterations, the set $\mathcal{NC}$ has 17.8 million clusters out of 18.6 million clusters in $\mathcal{R}$. To solve problems (1) and (2) above, we use the following algorithm.

1. Sort all clusters in $\mathcal{P}$ by their creation date.
2. Start with the oldest available cluster from $\mathcal{P}$ and for each such cluster find all its child clusters. To this end, denote the currently processed cluster by P .
 - (a) Sort all P 's outflows in chronological order.

- (b) Start with the earliest P 's outflow and find all clusters that receives this outflow. Denote the set of such clusters by A . For each cluster in A , we need to decide whether it is a child cluster of P .

We first discard any cluster in A that does not belong to $\mathcal{R}$ or belongs to $\mathcal{NC}$. Next, for each remaining cluster $A_i \in A$, we check if it has an inflow that precedes the currently processed outflow. If it has, it cannot be a child cluster of P , so we dropped any such cluster from a set of potential child clusters of P .

Denote the set of remaining potential child clusters by C_P . This is the set of clusters that received flows that can be traced to P and that have not be ruled out.

- (c) Consider the next earliest outflow either from P and from any of the potential child clusters C_P . Process this output following the same steps as in point (b) with the only difference, that when checking if a cluster has earlier inflows we only check earlier inflows that have not been marked as coming from P by that point.
- (d) After tracing all outflows forward, we have a set of clusters that receive outflows that can be tracked back to P and that do not have any other preceding inflows. It is still possible that some of these clusters can receive inflows from other clusters after the last inflow that can be tracked to P . We remove any such clusters. The remaining set is the set of child clusters of P .

3. We remove any clusters that appear as a child cluster of the currently processed cluster P from the set $\mathcal{P}$.

4. Return to point (2) and process the next earliest cluster from $\mathcal{P}$.

We run the above algorithm and obtain 114K root clusters and 537K their child clusters. We merge each root cluster with its child clusters, and impose two additional constraints on the resulting clusters to classify them as controlled by individual large investors.

1. the cluster has a balance above 50 BTC for more than 180 days.
2. the cluster balance above 50 BTC does not change for at least 90 days.

Finally, for each root-child merged cluster, we trace its flows to top source and target clusters, and drop any clusters that have the same source and target cluster with the proportion of inflows and outflows above 90%.

After applying these additional filters, we obtain a sample of 42,000 whales, of which 7,000 have never sold a significant portion of their Bitcoin holdings. With these data at hand, we are now ready to look at the concentration of individual Bitcoin ownership. When analyzing the trading behavior of whales, we drop from this sample any whale addresses where we have a gap in the traceability of balances, which leaves us with 38,900 whales. We confirmed that the results in the paper hold if we include the whales that have missing data.

Step 10: Probable trading venues (PTV) Finally, we identify high-volume unlabeled cluster that likely belong to trading venues. To this end, we select the top 10,000 clusters with the highest Bitcoin volume in each year. Across the 2015–2025 sample, the union of these annual top sets comprises 94,723 distinct clusters: 15,638 known clusters linked to 1,313 identified entities, 924 unknown miner clusters identified by tracing reward distributions from mining pools, and 78,161 other unknown clusters. For each cluster, we compute a set of diagnostic metrics, including the share of volume received from and sent to trading venues, the number of exchanges with which the cluster interacts, the number of addresses it contains, the total number of transactions, and aggregate volume.

When deciding on the classification rule, we prioritize minimizing the misclassification of non-trading venues as trading venues. We find that a simple rule that requires the share of volume received from trading venues to exceed 60% performs well in practice. This heuristic results in only 2.5% false discovery rate (0.86% volume-weighted). Applying this rule to the set of unknown clusters yields 21,083 clusters. We refer to these as probable trading venues (PTVs).

Internet Appendix Figures

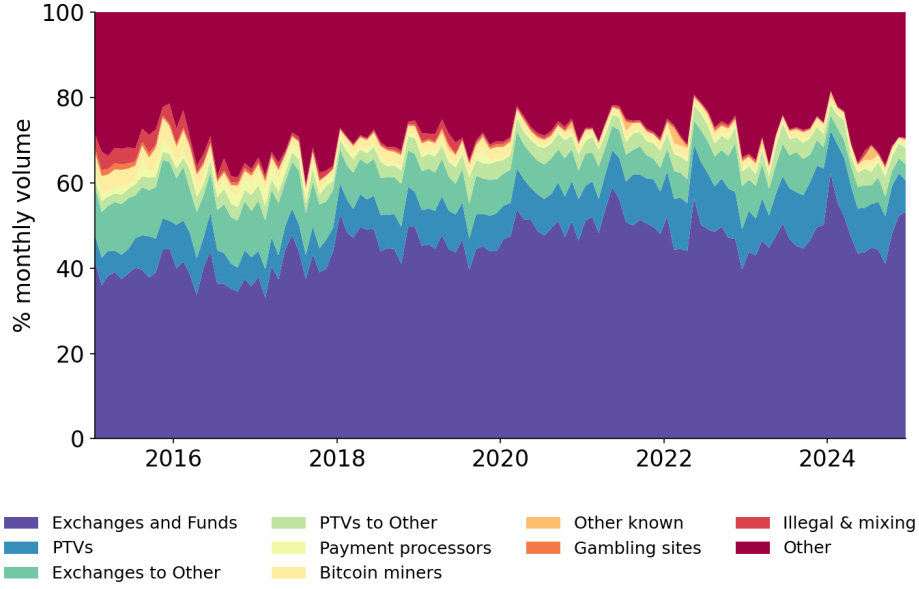


Figure IA.1: Decomposition of real volume share over time.

This figure documents the composition of transaction volume in terms of shares over time from 2015 to the end of 2024. The composition of transaction volume illustrated in Figure 2 right panel remains stable over our period.

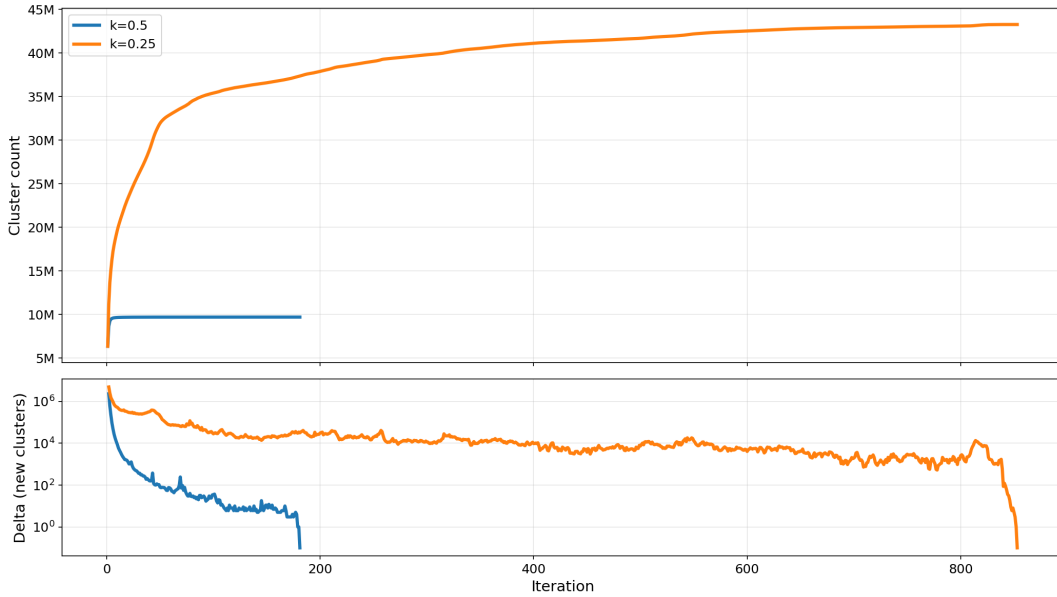


Figure IA.2: The number of clusters in $I(n)$.

This figure documents the number of clusters in the set of imputed clusters $I(n)$ as a function of n . The construction of sets $I(n)$ is described in Section 3.2.

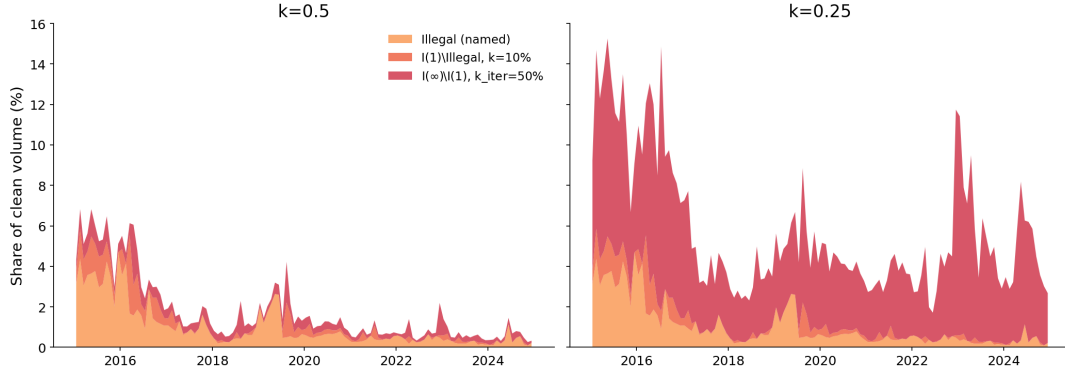


Figure IA.3: Imputed illegal volume share over time.

This figure documents the share of imputed illegal volume over time from 2015 to the end of 2024 for two threshold values: $k = 0.5$ (left panel) and $k = 0.25$ (right panel). The construction of sets $I(1)$ and $I(\infty)$ is described in Section 3.2.

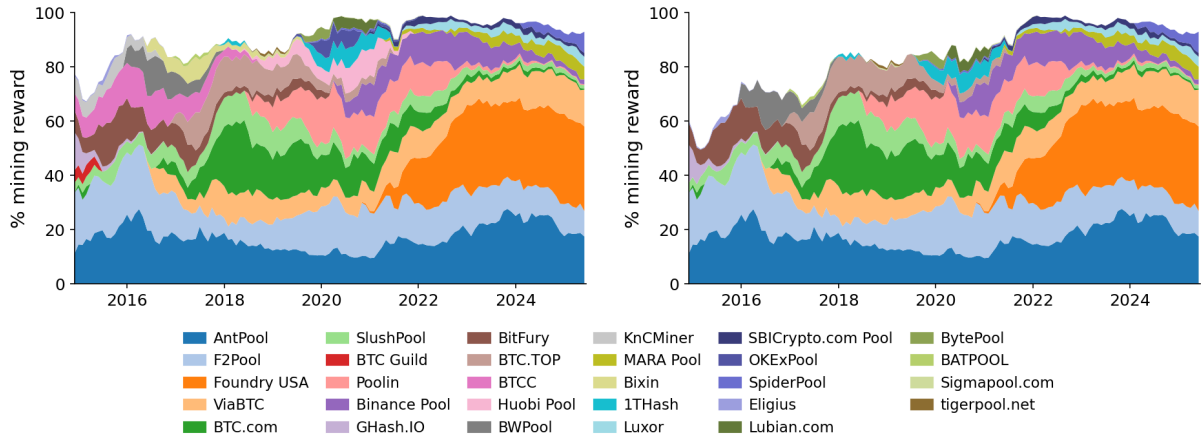


Figure IA.4: Coverage of mining pools.

The figure shows the share of bitcoins mined by each pool in each month from 2015 to December 2024. The left panel shows the shares of all known pools. The right panel shows the subset of pools for which we trace the Coinbase rewards of the pool to individual miners that are working with the pools.

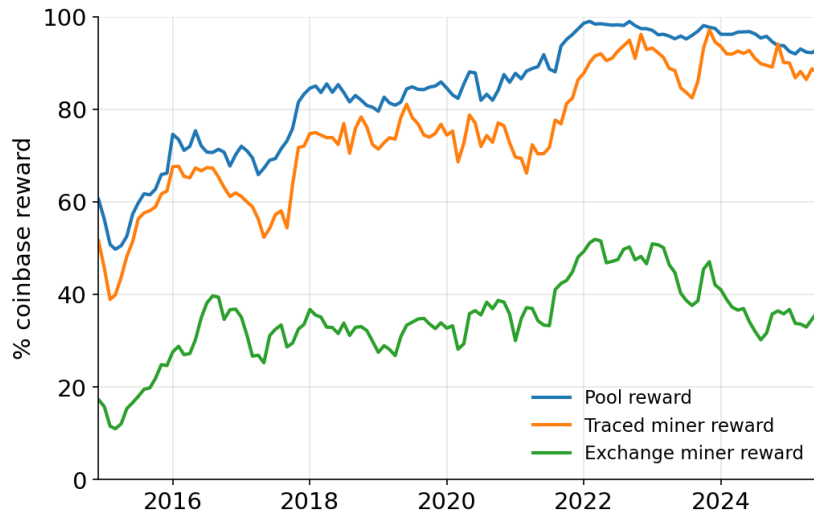


Figure IA.5: Coinbase reward traced to miners.

The figure shows our coverage of the mining capacity on the Bitcoin blockchain on a monthly basis from 2015 until December 2024. The blue line shows the aggregate coinbase rewards reported by the large pools that we can trace at an aggregate level. These are 23 mining pools, including three private pools. The red line shows the distributed mining rewards that we can trace on the blockchain from the pool distribution addresses to the underlying miners. The green line shows the rewards collected by exchange-wallet miners; these are miners who collect their rewards with addresses that are linked to an exchange.

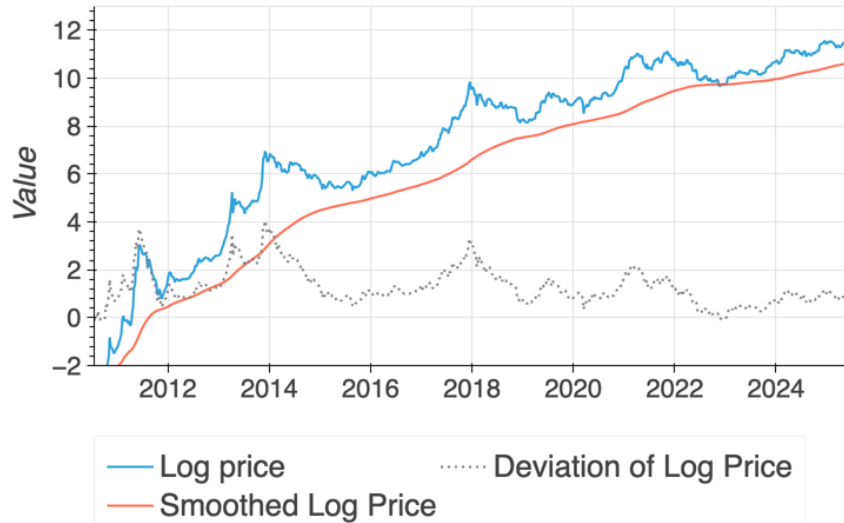


Figure IA.6: BTC Price and deviation from smoothed price.

The figure shows the Price trend of Bitcoin over our sample period. The blue line traces the weekly log USD price of Bitcoin from July 2010 through June 2025. The red line overlays an exponential-smoothing trend fitted to the same series. We smooth the weekly log-price series with simple-exponential smoothing using $\alpha = 0.01$. With weekly observation frequency, this places 1% weight on the most recent price and $(1 - \alpha)^n$ on observations n weeks old. The implied effective look-back window is about $1/\alpha \approx 100$ weeks, and the half-life—the point at which a past observation’s weight has decayed to 50% of the current weight—is $\ln 0.5 / \ln(1 - \alpha) \approx 69$ weeks. Because this slow-moving trend captures the long-run price level, the deviation reflects the cumulative appreciation over roughly the preceding one to two years. The grey dotted line plots the resulting deviation—log price minus smoothed trend.

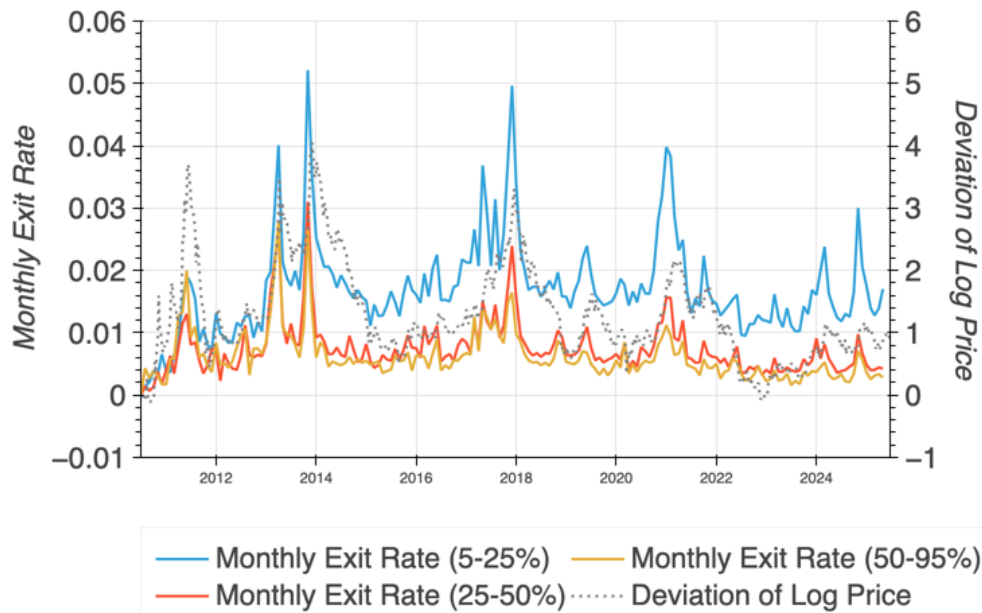


Figure IA.7: Whales Monthly Exits.

This figure documents the monthly fraction of whales that sell different fractions of their holdings. For each calendar month we calculate whether a whale reduces its BTC holdings by (i) 5–25%, (ii) 25–50%, or (iii) 50–95% relative to its end-of-previous-month balance. The figure plots the resulting exit rates, which are defined as the fraction of whales in each bin respectively. The grey dotted series overlays the deviation of the log price from the weekly smoothed trend on the right axis (see Figure IA.6).

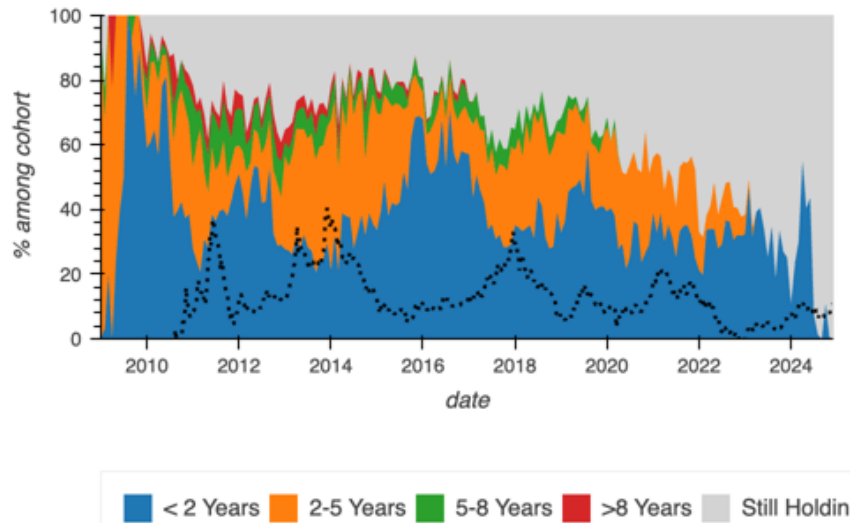


Figure IA.8: Whale Holding Duration.

This figure shows the distribution of the time that whales held on to their stake. For every monthly entry cohort of whales we classify each address by the length of time it remains active before exiting their wallet, where an exit is defined as the first month in which the whale disposes of at least 95% of its balance—or until the sample ends if no such sale occurs. Durations are grouped into four bins (< 2 years, 2–5 years, 5–8 years, > 8 years) with a residual “still holding” category. The stacked area plots, for each cohort on the horizontal axis, the percentage of whales falling into each duration bin; the gray top band represents whales that have not exited by 2025-06. The dotted black line is the log-price deviation series.

Internet Appendix Tables

Pool	Bitcoins mined	Blocks mined
AntPool	1,079,996	97,700
F2Pool	970,625	80,109
BTC.com	453,968	42,757
ViaBTC	369,984	45,915
SlushPool	344,675	28,262
<i>BitFury</i>	341,412	18,175
BTCC	336,625	17,330
Foundry USA	288,153	55,456
Poolin	247,796	28,422
BTC.TOP	217,631	18,250
BWPool	127,100	7,746
Binance Pool	125,637	20,772
KnCMiner	109,025	4,466
Huobi Pool	85,706	9,525
Bixin	73,787	5,753
GHash.IO	47,375	1,912
1THash	45,887	5,382
OKExPool	29,925	3,249
<i>MARA Pool</i>	29,421	6,169
Luxor	25,228	4,819
BTC Guild	24,625	985
SBICrypto.com Pool	19,846	3,594
Eligius	13,425	552
BytePool	12,525	1,002
SpiderPool	12,171	2,859
<i>Lubian.com</i>	12,043	1,783
BATPOOL	5,512	441
tigerpool.net	3,562	285
Sigmapool.com	2,581	294

Table IA.1. Summary statistics of mining pools.

This table reports the number of blocks and the number of bitcoins mined by each pool over the period 2015-2025. Italic indicates pools we have classified as private (single-firm pools); bold indicates pools whose miner-level rewards we trace. Pools that are both italic and bold are private pools whose internal payouts we recover from the blockchain.