Natural Centralization in Decentralized Finance*

Pablo Azar† Adrian Casillas‡ Maryam Farboodi§

October 7, 2025

Abstract

We ask whether concentrated market power can emerge in environments designed for perfect competition. Studying the Ethereum blockchain—a setting with permissionless entry and transparent rules—we show that centralization arises endogenously when information asymmetry interacts with risk-sharing. Using novel data distinguishing private from public order flow, a 1 percent increase in private information value causally increases intermediary profit shares by 0.57 percent. A dynamic bargaining model explains this as an equilibrium response: informed intermediaries gain rents by threatening to delay trade. Our results demonstrate how informational frictions alone can generate persistent oligopoly in frictionless digital markets.

*This paper subsumes the earlier draft "Information and Market Power in DeFi Intermediation." We are grateful to Vasco Carvalho, Douglas Diamond, Lars Hansen, Philipp Kircher, Michael J Lee, Ye Li, Yueran Ma, Jonathan Parker, Harald Uhlig, Keeyoung Rhee, Antoinette Shoar and Jeremy Stein for helpful comments and suggestions. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

†Federal Reserve Bank of New York; pablo.azar@ny.frb.org

‡NYU Stern; ac12001@stern.nyu.edu

§MIT Sloan, NBER and CEPR; farboodi@mit.edu

1 Introduction

A central question in economics and finance concerns the determinants of market structure. While canonical models suggest that environments with low barriers to entry and homogeneous products should lead to competitive outcomes, many markets exhibit persistent concentration. Understanding the sources of market power that generate and sustain such concentration is a foundational problem in industrial organization and finance. Is market power primarily the result of explicit frictions—such as barriers to entry, technological standards, capital requirements, or regulation, or does it arise endogenously from more primitive economic forces? It has been particularly challenging to answer this question empirically, as it is difficult to find settings where confounding institutional factors are verifiably absent.

This paper documents the emergence of an endogenous market structure, dominated by a few intermediaries who earn persistent information rents, in a novel setting that closely approximates a frictionless ideal: the Ethereum blockchain. This environment provides a unique real-world laboratory for studying the sources of market power as its core institutional design is defined by three features that, in theory, should foster intense competition: (1) *permissionless entry*, allowing any agent to become an intermediary without license; (2)*standardized protocols*, which commoditize the process of transaction aggregation; and (3) a *transparent public ledger*, where all transactions and resulting payoffs are universally observable.

Our central hypothesis is that this concentrated market structure is driven by two forces: information asymmetries and economies of scale. We argue that privileged access to private, unobservable order flow constitutes a durable source of market power, allowing a small set of intermediaries—known as block builders—to extract significant rents. Even in a system with a public ledger, the ability to see and sequence transactions before they are executed creates a valuable, excludable asset. This private information allows intermediaries to identify and capture arbitrage opportunities that are invisible to competitors, conferring a decisive advantage that is self-reinforcing: superior information attracts more order flow, which in turn solidifies the intermediary's informational advantage and market position. At the same time, risk-sharing forces and technological sophistication allow a second set of intermediaries—known as block proposers—to concentrate Ethereum deposits. This concentration is the second source of intermediation rents in this market.

We test this hypothesis using a comprehensive dataset of all transactions on the proof-of-stake Ethereum blockchain, which allows us to precisely distinguish between revenue derived from public versus private information. To establish causality, we develop an instrumental variables strategy that exploits two sources of plausibly exogenous variation in the value of

private information: major crypto-market crises and hacks of specific protocols. These events create sharp, systemic shifts in the demand for intermediation and the composition of order flow that are uncorrelated with individual intermediaries' characteristics. Our two-stage least squares estimates are unambiguous: a 1% increase in the value of a block builder's private information increases its profit share by 0.57%. In stark contrast, revenue from public information—which is contestable by all block builders—is fully competed away and captured by block proposers.

To formalize the economic mechanism, we develop a dynamic bargaining model that captures how informed intermediaries can strategically leverage private information over time to generate market power. In the model, a block builder holding a valuable private arbitrage opportunity can threaten to withhold it from the assigned block proposer and wait for a more favorable match in the future. This threat to delay constitutes a valuable outside option, enabling the builder to extract rents in each bargaining round. The model's predictions align with our empirical findings, showing why only private, non-contestable information can sustain rents for block builders, while public information is fully competed away.

This paper's primary contribution is to provide some of the cleanest empirical evidence to date on the role of information as a fundamental determinant of market structure and intermediation rents. Studying a setting where most traditional frictions are absent allows us to isolate an economic force capable of producing a "natural oligopoly." Our results show that even in technologically advanced, ostensibly decentralized markets, information asymmetry can drive significant and durable concentration—shaping financial intermediation.

Our analysis contributes directly to several foundational questions in market microstructure. The mechanism we identify provides direct evidence on the value of mitigating information leakage, a central challenge for informed institutions (Kyle, 1985; Han et al., 2020). Furthermore, our setting offers a transparent view of market segmentation. The public order flow is analogous to a 'lit' exchange, while private order flow resembles opaque off-exchange venues such as dark pools or payment-for-order-flow arrangements (Zhu, 2013; Battalio et al., 2016; Barbon et al., 2019). These environments raise fundamental questions about rent allocation when informed order flow is separated from public liquidity—questions that are notoriously difficult to study empirically in fragmented traditional markets.

Our empirical and theoretical analysis of blockchain market structure is increasingly relevant beyond academia. Blockchain-based markets are being woven into the core infrastructure of global finance, drawing the attention of policymakers and major firms. The SEC's approval of spot Bitcoin and Ethereum ETFs marked the first wave of institutional adop-

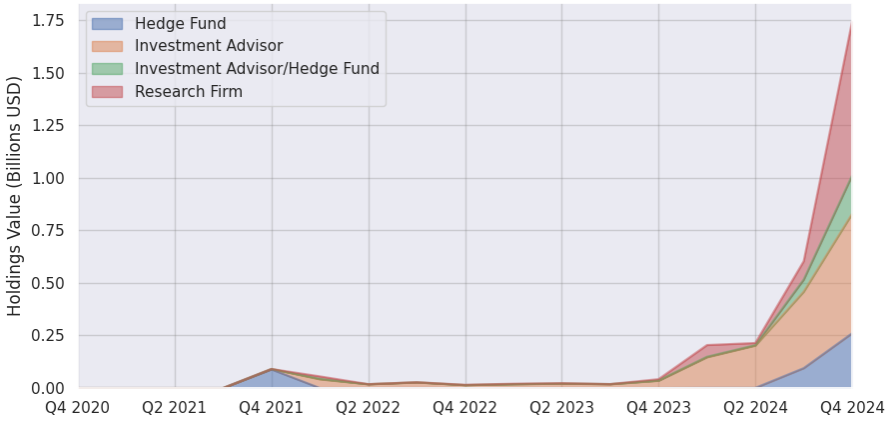


Figure 1: Institutional Ethereum holdings (in Billions USD) from Q4 2020 to Q4 2024. The charts show the dramatic increase in institutional investment across different types of financial entities, with Investment Advisors and Research Firms dominating holdings by the end of 2024.

tion, followed by tokenized money-market funds from firms such as BlackRock (Figure 1). In 2025, the Guiding and Establishing National Innovation for U.S. Stablecoins (GENIUS) Act created the first comprehensive federal framework for stablecoins, prompting traditional financial institutions to explore issuing their own.[1] As the boundary between decentralized and traditional finance narrows, the concentration dynamics we study matter not only for theory, but for those designing and regulating the next generation of financial systems.

1.1 The DeFi Intermediation Chain

The decentralized structure of Ethereum allows for numerous competing Decentralized Finance (DeFi) protocols—including decentralized exchanges, crypto-collateralized stablecoins, and lending platforms. This multiplicity of protocols creates a constant stream of violations of the law of one price, together with ensuing arbitrage opportunities. Importantly, Ethereum allows arbitrageurs who find price discrepancies to obtain a guaranteed non-negative profit by submitting groups of transactions *atomically*—either all transactions are executed, or none of them are. These transactions therefore constitute risk-free arbitrage opportunities.

However, the Ethereum ecosystem also introduces a novel limit to arbitrage in the spirit of Shleifer and Vishny (1997): arbitrageurs with private information need a way to have their trades appended to the public blockchain without alerting potential rivals. This ten-

[1]Hannah Lang, 2025, "Companies plan stablecoins under new law, but experts say hurdles remain," Reuters, www.reuters.com (accessed August 18, 2025).

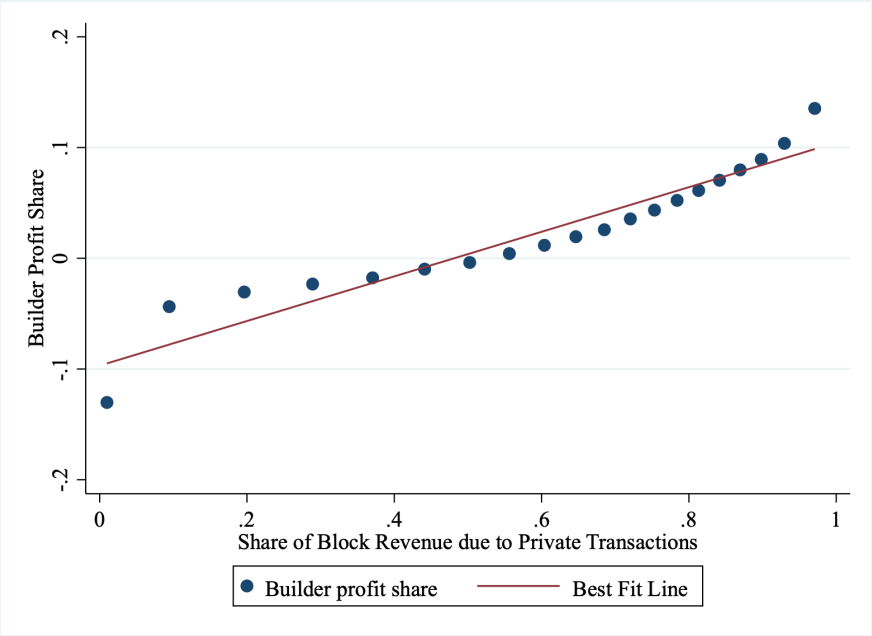
sion between the public nature of blockchain trades and the private needs of arbitrageurs shapes the DeFi intermediation chain, giving rise to a multi-layered structure: 1) arbitrageurs identify mispricing across different DeFi protocols or between centralized and decentralized exchanges; 2) block builders aggregate transactions into blocks, acting as gatekeepers of private information and potentially extracting rents in the process; 3) block proposers, selected randomly via the proof-of-stake mechanism, choose one winning block among bids from multiple builders; and 4) ETH depositors, including individual holders and centralized exchanges, participate in this process through delegating their stake to proposers, in a process we call *delegated staking*. This introduces an additional layer of intermediation, mirroring traditional financial structures in a decentralized context.

1.2 Impact of Private Information on Profit Sharing

The level of transparency and granularity in the data allows us to causally identify the impact of private information on profit sharing between block builders and block proposers. Figure 2 depicts our main finding. It illustrates that access to valuable *private* arbitrage transactions by builders—i.e., those arbitrage transactions that are privately submitted to them by an arbitrageur— increase their profit share since they are effectively the gatekeeper for the private arbitrage opportunity. On the other hand, the block revenue that is associated with public arbitrage opportunities is widely accessible to the proposer through other block builders. As such, none of the block builders can gain from the publicly available arbitrage transactions and the proposers capture most of the corresponding revenue.

To address biases from simultaneity and omitted variables, we employ very stringent fixed effects and introduce two novel instrumental variables: a dummy for major crypto market crises and a dummy for hacks of exchanges and decentralized protocols. These instruments are designed to capture variations in both total block revenue and the value of private information. We identify crises as the FTX bankruptcy (November 8-12, 2022) and the SVB run (March 9-12, 2023), which led to a large number of blockchain transactions. The dummy corresponding to crypto hacks is set to 1 on days on which an Ethereum protocol or exchange is hacked, according to data from DefiLlama. As shown in Section 4, both crises and hacks affect overall block value and private information value, but to different degrees. Crises tend to generate more public transactions, while hacks primarily increase private information value. This differential impact allows these two instrumental variables to effectively span our set of two explanatory variables.

Using this instrumental variable approach, we show in Section 4 that a 1% increase in the value of private information leads to 0.57% increase in the builder profit share. After



Source: Dune Analytics and Mempool Guru Project

Figure 2: Block builder's profit share as a function of the share of the block's revenue that is due private information. There is a strong positive relationship between the two variables. This figure is generated with binscatter, using 20 bins.

controlling for the value of private information, the effect of larger block revenue on the builder profit share is negative. This indicates that the builder's market power is driven exclusively from the value of their private information, and not from the total value of the block they produce.

1.3 A Bargaining Model of DeFi Intermediation

We complement our empirical analysis with a dynamic bargaining model to shed light on how information asymmetries in decentralized finance contribute to the distribution of profits along the DeFi intermediation chain. The core economic mechanism is the interaction between the source of information asymmetry between block builders and proposers, the market structure of the block builder segment of the market, and the randomness inherent in the proof-of-stake technology.

Consider the bargaining process between the block builders and the proposer when adding each block to the blockchain. Observe that the block builders are always better informed than the proposer about the content and the value of the block that they built. However, if the information advantage is common among all block builders, through public arbitrage

opportunities, the competition among the block builders prevents them from exploiting it, even if it is very valuable. On the other hand, if the information asymmetry between a block builder and the proposer stems from transactions that are private to that specific builder, through private arbitrage opportunities, it constitutes a "private information advantage." Unlike a public information advantage, the block builder is able to monetize their private information advantage through the threat of withholding the information in this period and selling it next period to the proposer randomly selected by the proof-of-stake mechanism —who is not necessarily the current proposer.

In order to capture the above intuition, we propose a repeated bargaining model where the outside option of the agents links the different time periods. This simple economic mechanism enables us to simultaneously explain both of our main empirical findings. First, a higher value of private transactions in a block increases the profit share of the block builder. However, controlling for the value of private transactions, higher block revenue increases the profit share of the proposer while decreasing that of the block builder.

1.4 Related Literature

There is an extensive literature spanning different aspects of financial intermediation (see Leland and Pyle (1977), Campbell and Kracaw (1980), Diamond and Dybvig (1983), Diamond (1984), Allen (1990), Allen and Gale (1997), Boot and Thakor (1997), Diamond and Rajan (2001), as well as Gorton and Winton (2003) and references therein). The prevalence of intermediation rents in financial markets have been widely documented in the empirical literature (Green et al., 2007; Di Maggio et al., 2017; Hollifield et al., 2017; Li and Schürhoff, 2018; Farboodi et al., 2025). However, identifying the source of these rents empirically has proven challenging as the balance sheet of large financial intermediaries is opaque and it is hard to acquire data about their comparative advantage. We contribute to this literature by first distinguishing the role of financial intermediaries in blockchain systems, cryptocurrency and DeFi, which represent the most recent developments in financial technology. Second, we identify private information as a source of intermediation rents in this market.

The paper also speaks to foundational theories in the industrial organization of financial markets. Guerrieri et al. (2010) and Guerrieri and Shimer (2014) consider competitive search models with adverse selection and analyze the impact of private information on bilateral matching outcomes, probability and terms of trade, as well as market liquidity. Our setting provides a transparent environment to test these predictions, cleanly isolating the value of private information from other confounding frictions. In this respect, our work also contributes to the literature on intangible capital as a source of firm value and market power

(Crouzet et al., 2022; Terry et al., 2022). We show that privileged access to order flow functions as a measurable, excludable intangible asset that generates persistent rents, providing a specific microfoundation for how such advantages arise and are sustained.

Our paper contributes to a fast-growing literature on blockchain technology. Raskin and Yermack (2018) provide a preliminary overview of financial systems built on blockchains. Cong and He (2019), Abadi and Brunnermeier (2018) and Biais et al. (2019) expand on consensus mechanisms, focusing on proof-of-work. There is a small but growing body of work that studies the economics of Bitcoin, both theoretically and empirically. Athey et al. (2016), Cong et al. (2021a), Pagnotta and Buraschi (2018) and Sockin and Xiong (2023) develop alternative theoretical frameworks to study the decentralized Bitcoin network. Prat and Walter (2021) provide an estimate of the computing power of Bitcoin network and Cong et al. (2021b) study the effect of mining pools on energy consumption, which is a significant input to proof-of-work consensus. Budish (2025) suggests lack of security facing external attacks as a limit to economic viability of Bitcoin.

A number of papers consider the degree of decentralization in blockchain, focusing on proof-of-work consensus protocols (Cong et al., 2023; Huberman et al., 2021; Ferreira et al., 2023; Makarov and Schoar, 2021; Capponi et al., 2025; Cong et al., 2021b; Lehar and Parlour, 2023; Budish, 2018). In addition, some previous work studies staking systems, including the game-theoretic properties of proof-of-stake consensus mechanisms (Saleh, 2021; Schwarz-Schilling et al., 2023), the valuation of native tokens such as ETH (Fanti et al., 2021), and the valuation of non-native tokens that can be staked in DeFi protocols (Cong et al., 2025). In contrast to these papers, we focus on the emergence of intermediation in a financial sector built on proof-of-stake technology and study the concentration of this intermediated market. We emphasize the influence of arbitrage opportunities, as documented in Makarov and Schoar (2020), on the degree of market concentration, and show that the combination of proof-of-stake consensus and smart contracts can lead to a high degree of concentration in the Ethereum crypto intermediation market.

The nature and type of arbitrage opportunities in a proof-of-stake blockchain is at the core of our analysis. Daian et al. (2020); Gupta et al. (2023); Heimbach et al. (2024) provide early empirical evidence and classification of MEV and private and public arbitrage opportunities in a blockchain network. Alternatively, Milionis et al. (2023) theoretically models the relationship between public and private transactions and market volatility. Capponi et al. (2024) provide a game-theoretic model of proposer-builder separation. We use a simplified definition of private transactions, where a transaction is private if it is not broadcast to the network before it appears on the blockchain, a simple and intuitive definition that can

be measured precisely in the data, in contrast with heuristic based definitions proposed by Gupta et al. (2023) and Heimbach et al. (2024). In Section 6, we use their classification to show the robustness of our empirical findings. Our instrumental variable approach relies on the observation that crypto crises can be an exogenous shock that creates arbitrage opportunities (Liu et al., 2023). To the best of our knowledge, we are the first to use these crises as instrumental variables in empirical analyses of DeFi.

The rest of the paper is organized as follows. Section 2 describes the institutional details of the DeFi intermediation chain. Section 3 provides details of the block level data from the Ethereum blockchain that we use for our empirical analysis. Section 4 describes the instrumental variable approach and presents the main empirical results. Section 5 proposes a model to provide the economic mechanism that underlies the empirical findings. Section 6 provides a number of robustness exercises. Section 7 concludes.

2 Origins of DeFi Intermediation: The Need for Privacy

The Ethereum blockchain handles two main types of transactions: simple payment transactions and smart contract interactions. Simple transactions transfer ETH or tokens between addresses. Smart contracts are blockchain-stored programs that execute when users send transactions to their addresses, triggering predefined functions. For example, users can interact with decentralized exchange contracts to swap tokens or with lending protocols to deposit collateral. These contracts automatically execute complex financial operations, such as updating inventories or calculating prices and interest rates, without a central operator.

Risk-free arbitrage. Ethereum's diverse ecosystem of competing DeFi protocols creates a large number of violations of the law of one price.[2] Since Ethereum allows for transaction batching, arbitrageurs can submit multiple transactions which are executed atomically—either all together or not at all—ensuring a guaranteed non-negative profit for the arbitrageur. The unique aspect of blockchain-based arbitrage, particularly in Ethereum's DeFi ecosystem, is the ability to execute truly risk-free arbitrage through atomic transactions that have no execution risk. This feature, combined with the transparency of the blockchain, creates a novel environment for arbitrage that is not directly paralleled in traditional financial systems.

The risk-free nature of these arbitrage opportunities makes them very attractive. Yet, the nature of PoS blockchain technology introduces a novel *limit to arbitrage*: the possibility

[2]Most of these violations occur through differences in pricing across exchanges. However, some additional arbitrages are due to automatic liquidations of collateral at fire sale prices, and its resale at market prices.

of front running. The key friction stems from arbitrageurs' need to keep their transactions private before they are appended to the blockchain if they want to appropriate the corresponding profits. There are constant opportunities for arbitrageurs to identify mispricings across different DeFi protocols or between centralized and decentralized exchanges. However, to capitalize on these opportunities, arbitrageurs require a mechanism to have their transactions approved without broadcasting them to the entire network, thus avoiding the risk of being front-run or having their arbitrage opportunity stolen by competitors. [3]

This privacy requirement is the limit to arbitrage faced by each individual arbitrageur. It introduces unique challenges in exploiting arbitrage opportunities on the Ethereum blockchain and shapes the structure of DeFi intermediation. Moreover, it gives rise to a specific form of information rent captured by some intermediaries in the process of creating the Ethereum blockchain. In this section, we focus on the emergence of the DeFi intermediation chain as a consequence of the need for privacy. Sections 4 and 5 explain the information rents in the DeFi intermediation chain, both empirically and theoretically.

DeFi intermediation chain. We define the "DeFi intermediation chain" to be the market structure that underlies the creation and distribution of ETH, the native cryptocurrency of Ethereum. It consists of four groups of agents who interact with each other through an intermediation chain. *Arbitrageurs* form the initial segment of the chain. These are typically high-frequency trading algorithms or bots that continuously scan for profitable transactions, including both arbitrage and non-arbitrage opportunities, to be submitted to the Ethereum blockchain.

The need for privacy by arbitrageurs, as explained above, leads to the rise of *block builders* as intermediaries. Arbitrageurs who find an arbitrage can send their transactions directly to a block builder, who incorporates them into an aggregate block. If these arbitrage transactions are valuable, arbitrageurs usually pay an additional fee or direct payment to the builder in order to make sure the builders incorporate their transactions into blocks. The total value that is generated by adding a block to the blockchain—colloquially known as the block's Maximum Extractable Value (MEV)—is the sum of arbitrageur profits, transaction fees paid to the block builders, and any direct payments sent by arbitrageurs to builders in order to incentivize them to add their transactions to the block.

The next layer of intermediaries are *block proposers*, who are selected at random via the proof-of-stake (PoS) consensus mechanism—with probability proportional to their stake of

[3]The risk of having a transaction stolen is not only real, but ubiquitous due to the commodification of AI-driven front-running bots (Robinson and Konstantopoulos, 2020).

ETH—to select a single block to be appended to the blockchain in a given round.[4] Block builders compete with each other to create the most valuable block they can. They then submit a cryptographic commitment to the block along with a bid to the block proposer, who chooses one winning (block, bid) pair without being able to view the block's contents. To prevent the proposer from front-running arbitrageurs' transactions, the encrypted block is only decrypted and revealed once the proposer has accepted the bid.[5]

Because block proposers are selected at random with probability proportional to their stake, proposers who have larger amounts of ETH obtain a much more consistent stream of revenue than proposers with a very small amount of ETH. This leads to the final layer of the intermediation chain, where *ETH depositors* pool their assets together into large staking pools, and get a share of the profits that these pools obtain from proposing blocks.

Figure 3 illustrates the market structure of the Ethereum blockchain's production network, highlighting the DeFi intermediation chain. The diagram shows four key participant groups: arbitrageurs, block builders, proposers, and depositors. Arbitrageurs and depositors are represented by multiple icons, indicating their larger numbers and diverse nature. Block builders and proposers have fewer icons, reflecting their more concentrated roles. The arrows between groups represent the flow of transactions within the DeFi intermediation chain. Importantly, the arrow from Depositors to Proposers reflects that depositors choose proposers to stake their ETH with, influencing the distribution of staking power.

The solid red arrows indicate the selected paths during the matching process in each time slot, to build a single block. Furthermore, in each time slot the Proof-of-Stake consensus mechanism selects the proposer for each single block (dashed red arrow pointing to the red proposer), who in turn adds a block to the Ethereum blockchain (dashed red arrow pointing from the proposer to the blockchain).

Delegated Staking Many ETH holders do not participate in the above protocols directly, but rather through intermediaries such as centralized exchanges or liquid staking smart contracts. A centralized exchange is a company such as Coinbase, Binance, or Kraken, which takes depositors' ETH and uses it to participate in the proof-of-stake Ethereum consensus protocol. By staking their depositors' ETH, these centralized exchanges earn returns, which

[4]As a brief note on terminology, we note here that the term block proposer is related to the technical architecture of the Ethereum blockchain. Once proposers select a block, they *propose* it to a random small group of *attesters*, who verify that all transactions in the block are valid and there is no double spending in a block. From an economic point of view, the attesters do not receive any revenue related to DeFi intermediation, and therefore we do not study them in this paper. We also highlight that the case where a proposer's selected block is not accepted by the attesters is extremely rare, and the attester mechanism exists solely to ensure honest behavior by the proposer.

[5]We provide more details of this procedure in the Appendix.

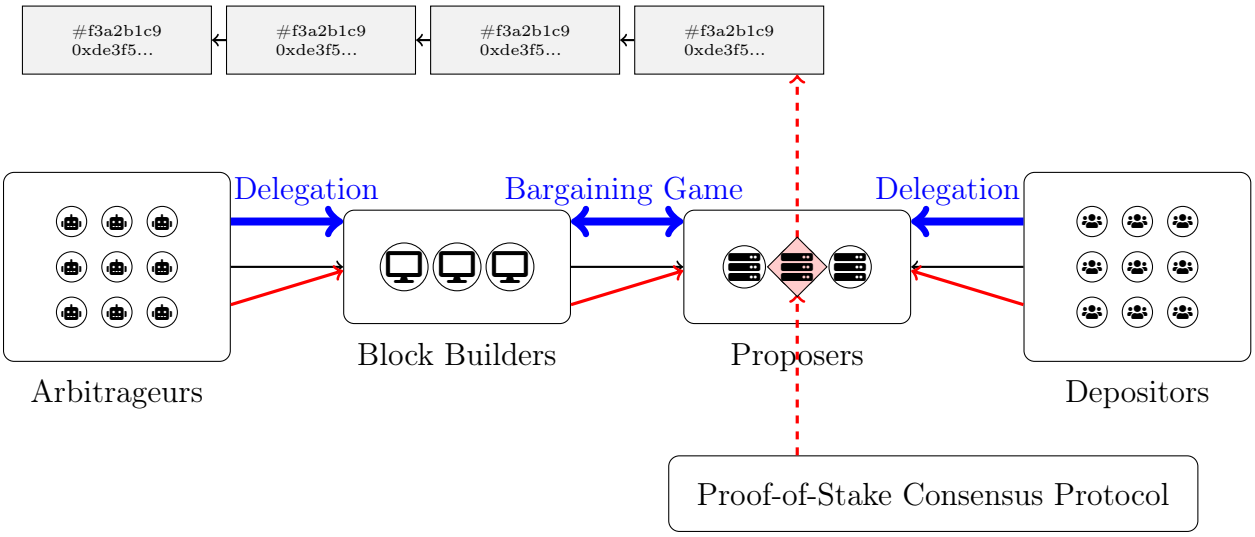


Figure 3: Market structure of the DeFi intermediation chain, Ethereum blockchain's production network.

Notes: The figure depicts the proof-of-stake consensus protocol (bottom row), the key participants of the DeFi intermediation chain: arbitrageurs, block builders, proposers and depositors (middle row),and the produced Ethereum blockchain (top row). The interactions at each time period t (time slot of 12 second) are depicted in red. The red arrows among the market participants indicate the selected paths in the DeFi intermediation chain required to produce each block, while the black signify alternative paths that were not selected. The dashed red arrow from the consensus protocol to the proposers indicates that at each time t, the random outcome of the proof-of-stake consensus protocol is a proposer, with the red diamond displaying the randomly selected proposer. Finally, the dashed red arrow from proposer to the blockchain at the top represents the block proposal process. The blockchain is the outcome of these repeated interactions.

they then pass on to their customers after taking a spread.[6] These intermediaries arise because they have the technological sophistication to participate in the proof-of-stake protocol.[7] ETH holders who do not have this level of sophistication may still earn returns by buying ETH from a centralized exchange, and asking the exchange to stake their ETH for them. As such, proposers are *delegated stakers* in the DeFi intermediation chain.

Figure 4 shows these delegated stakers' market shares over time. We can see that the overall shares do not change significantly, indicating that the relationship between depositors and their delegated proposers remains stable over time.

[6]Recently, the SEC has reached a settlement with Kraken to prevent it from acting as such an intermediary for American customers.

[7]In particular, participants need to continuously run a server which listens for transactions, engages with block builders, proposes valid blocks when called upon to do so, and verifies other proposers' blocks. Any deviation from the protocol, due for example to a software bug, a hardware failure or a network outage, is penalized by debiting ETH from the participants' account—a process known as slashing. In equilibrium, sophisticated agents can easily satisfy these requirements, so slashing is extremely rare, with only 0.04% of participants having been slashed according to CoinTelegraph `https://cointelegraph.com/news/only-0-04-of-ethereum-proposers-have-been-slashed-since-2020-says-core-dev`.

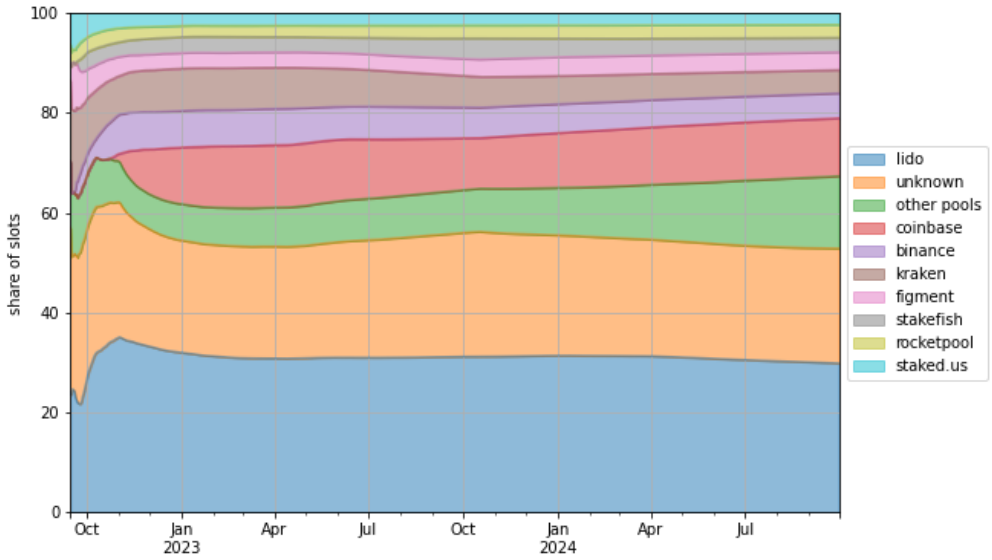


Figure 4: This figure shows how validator shares have evolved over time. The shares are relatively stable, showing that the overall market structure of delegated staking has not significantly shifted over time.

Throughout the chain, each node receives a payoff for their service. Arbitrageurs keep a large amount of their arbitrage profits, but pay transaction fees to the builders. The builders pay the block proposers to ensure their blocks are added to the blockchain. As such, the block proposer's net revenue is equal to the bid of the winning bidder, while the winning block builder's net revenue consists of all transaction fees and direct payments to them, minus the bid that they pay to the block proposer. Finally, block proposers who represent staking pools pay a large share of their revenues to the individual investors who pooled their ETH with them.

Even though the Ethereum blockchain is permissionless and there is near free-entry into intermediation, informational frictions and risk-sharing in DeFi lead to concentration among intermediaries as demonstrated in Table 1. Of the 108 known builders, 3 capture more than 50% of all the builder revenue and blocks proposed. Similarly, even though there are more than a hundred thousand block proposers, 5 large staking pools capture more than 50% of all the proposer revenue and blocks proposed.

Number of Builders	**Number of Proposers**	
108	189,126	
Builder	Share of Total Revenue	Share of Total Blocks
beaverbuild.org	34.25	32.3
Titan Builder	33.41	15.11
rsync-builder.xyz	7.93	16.28
Proposer	Share of Total Revenue	Share of Total Blocks
Lido	31.14	29.81
Coinbase	11.56	11.59
Binance	4.74	5.01
Kraken	4.68	4.64
Figment	3.32	3.59

Source: Dune Analytics

Table 1: The first panel reports the total number of block builders and proposers. The second panel shows that the largest 3 block builders account for more than 50% of aggregate builder revenue, and more than 50% of the number of blocks added to the Ethereum blockchain. The third panel shows that the largest 5 block proposers account for more than 50% of aggregate proposer revenue, and more than 50% of the blocks added to the Ethereum blockchain.

3 Data

We use Dune Analytics[8] to obtain block-level data, the identity of the builder and proposer, the MEV revenue for the block, and the revenue split between the builder and the proposer. We use data from the Secure Decentralized Systems Lab's Mempool Guru project (Yang et al., 2024) to keep track of which transactions were broadcast to the network before being appended to the blockchain, and which were not broadcast to the network. We classify transactions broadcast to the network as public, and transactions not broadcast as private.

Let B_t denote the block added to the blockchain at time slot t. We consider the block to be an MEV block if two conditions hold. First, the block builder is different than the block proposer. Second, the last transaction of the block is issued from the block builder to the block proposer. In our sample—which spans from the switch to proof-of-stake in September 15, 2022, to September 30, 2024—77.6% of the blocks satisfy both of these conditions, and are considered MEV blocks.[9]

[8] `https://www.dune.com`

[9] The total number of blocks in our sample is 5,326,069, and the number of blocks satisfying the MEV

The key independent variable in our analysis is the value of private transactions in the block generated at time t. We define private and public transactions as follows.

Definition 1.

Public Transaction *A transaction in the block added at time t is public if is broadcast to the network before time t, and is not a direct payment to the builder who built the block generated at time t.*

Private Transaction *A transaction in the block added at time t is private if either it is not broadcast to the network before time t or it is a direct payment addressed to the builder who built the block generated at time t.*

The key concept behind this definition is that public transactions are *non-exclusive*: any builder can collect their value if they are chosen as the block builder at time t. Private transactions, however, are *exclusive*: only the builders that know about them, or the builder whom the payment is addressed to, can collect the value of these transactions.

Let Rev_t denote the total revenue from block t. Moreover, let $\Pi_{B,t}$ and $\Pi_{P,t}$ denote the net profit of the block builder and proposer, respectively. The net profit for the builder, $\Pi_{B,t}$, consists of the sum of direct payments they receive and priority gas fees,[10] minus the payment to the proposer at the end of the block. The net profit for the proposer, $\Pi_{P,t}$, is the value of the block's final transaction. The key dependent variables in our analysis are the profit shares of the builder and proposer are denoted as $\theta_{B,t} = \frac{\Pi_{B,t}}{Rev_t}$ and $\theta_{P,t} = \frac{\Pi_{P,t}}{Rev_t}$, respectively.

Table 2 presents the summary statistics of the key variables. It shows that all profits are highly skewed to the right, with the majority of blocks generating minimal revenue. On average, a block generates 0.14 ETH in revenue, more than 85% of which is captured by the proposer.

There are many blocks where the builder makes negative profits. This behavior is likely to ensure that the builder's block is chosen and is adopted as strategy to build market share: by subsidizing proposers during regular periods, builders aim to dominate the market share of proposed blocks, attracting arbitrageurs with lucrative arbitrage opportunities when they arise, thereby securing blocks that yield positive profits.[11] Our main analysis considers only blocks where the builder profit share is greater than -10%, which represent 94.9% of the

conditions is 4,132,184.

[10] Any Ethereum transaction must pay a base gas fee Fee_{Base} to be included in the block. This fee is always "burnt" and removed from the system, and is not part of the builder's revenue. However, if the transaction is valuable or important, the user who submits the transaction may choose to pay the builder an excess gas fee Fee_{Excess}, which is part of the builder's revenue.

[11] Primarily empirical evidence support the outcome of this strategy. Results are available upon request.

	Mean	Std. Dev.	Min	5th	Median	95th	99th	Max	Skewness	Kurtosis
Rev_t	0.14	1.49	0.00	0.02	0.05	0.35	1.27	802.79	228.41	83247.25
$\Pi_{B,t}$	0.02	0.66	-0.30	-0.00	0.00	0.03	0.17	791.74	616.15	615521.41
$\Pi_{P,t}$	0.12	1.22	0.00	0.02	0.05	0.32	1.09	691.96	260.66	108839.84
$\theta_{B,t}$	0.05	0.12	-0.10	-0.04	0.01	0.29	0.58	1.00	3.23	15.94
$\theta_{P,t}$	0.95	0.12	0.00	0.71	0.99	1.04	1.08	1.10	-3.23	15.94
$\log Private_t$	0.08	0.17	0.00	0.00	0.03	0.27	0.78	6.69	8.82	128.29
$\log Public_t$	0.02	0.05	0.00	0.01	0.02	0.06	0.14	5.20	26.73	1295.27
Hack Dummy	0.07	0.25	0.00	0.00	0.00	1.00	1.00	1.00	3.42	12.71
Crisis Dummy	0.01	0.11	0.00	0.00	0.00	0.00	1.00	1.00	8.58	74.63
Observations	3924921									

Source: Dune Analytics and Mempool Guru Project

Table 2: Summary Statistics

MEV blocks in our sample.[12]

4 Information Rents in Ethereum Intermediation Chain

In this section, we estimate how a block's share of private revenue affects the builder's profit share. The simplest specification would be a regression of the form

$$\theta_{B,t} = \alpha + \beta \log Private_t + \epsilon_t.$$

However, estimating this regression with OLS would introduce biases in three ways. First, there is an omitted variable bias because there are many characteristics of the block which can affect $\theta_{B,t}$ which are not captured in the specification. The most important variable that is missing from the specification is the other transactions included in the block, $Public_t$. Second, the specification above does not capture any relationships between builders and proposers which may lead the builders to treat some proposers more or less favorably.

To address the potential for omitted variable bias and pre-existing relationships between builders and proposers, we estimate the slightly more complicated regression (1):

$$\theta_{B,t} = \beta \log Private_t + \gamma \log Public_t + \psi_{i(t)} + \eta_{j(t)} + \phi_{i(t),j(t)} + \epsilon_t. \quad (1)$$

[12]Section 6 and Appendix C report the results for the full sample and show that they are even stronger than the restricted sample. We limit ourselves to blocks with small subsidies in the main text to ensure that the results are not driven by subsidies.

As $Rev_t = Private_t + Public_t$, including the $Public_t$ allows us to capture the effect of both private information as well as the total revenue of the block. The fixed effects terms $\psi_{i(t)}, \eta_{j(t)}, \phi_{i(t),j(t)}$ capture relations between the builders and proposers. Because the proposer shares of ETH staked are relatively stable over time, these fixed effects capture the relative market power of different proposers.

Finally, there is simultaneity bias because the block builder simultaneously decides their payment to the proposer (determining $\theta_{B,t}$) and the transactions that they want to insert into the block. That is, the builder has to decide how much of their private information they want to capitalize in during this block, and how much of that value they want to share with the proposer.

To address simultaneity, we use an instrumental variables approach with two instruments, both of which are dummy variables. The first dummy, $Hacked_t$, is equal to 1 if block t is appended to the blockchain on a day where there is a crypto protocol hack, and 0 otherwise.[13] The second dummy, $Crisis_t$ is equal to 1 if block t is appended to the blockchain during either the FTX or SVB crises.[14] The first-stage specification is given by

$$\begin{aligned}
\log Private_t &= \widehat{\beta}_1 Hacked_t + \widehat{\gamma}_1 Crisis_t + \widehat{\psi_{1,i(t)}} + \widehat{\eta_{1,j(t)}} + \widehat{\phi_{1,i(t),j(t)}} + \widehat{\epsilon_{1,t}}; \\
\log Public_t &= \widehat{\beta}_2 Hacked_t + \widehat{\gamma}_2 Crisis_t + \widehat{\psi_{2,i(t)}} + \widehat{\eta_{2,j(t)}} + \widehat{\phi_{2,i(t),j(t)}} + \widehat{\epsilon_{2,t}}.
\end{aligned}$$

The exclusion restriction for our two instruments — crypto hacks and major market crises — is that these events influence the builder share exclusively through their impact on the composition of blocks, particularly the balance between private and public transactions, rather than through any direct effect on the builder-proposer negotiation process. Furthermore, the instruments are exogenous, since both hacks and crypto crises are unexpected and not caused by the bargaining process between builders and proposers.

These two instruments span our two explanatory variables, due to the distinct ways in which hacks and crises affect the DeFi ecosystem. Crypto crises, such as the FTX collapse or the SVB crisis, impact the entire crypto market in a broad, systemic manner. This is reflected in our first-stage results shown in Table 3. During crises, both private and public revenue increase, with coefficients of approximately 0.12 for log private revenue and 0.01 for log public revenue. In contrast, hacks typically affect individual protocols or platforms within the Ethereum ecosystem. These events are more likely to generate private revenue

[13]The list of hacks is obtained from DefiLlama (`https://defillama.com`), and we keep only hacks which affected only the Ethereum chain.

[14]The FTX crises occurred between November 8 2022, and November 12 2022. The SVB crisis unfolded between March 9 2023, and March 12 2023. Other crypto crises, such as the Terra crash, occurred before the transition to proof-of-stake and are therefore not in our sample.

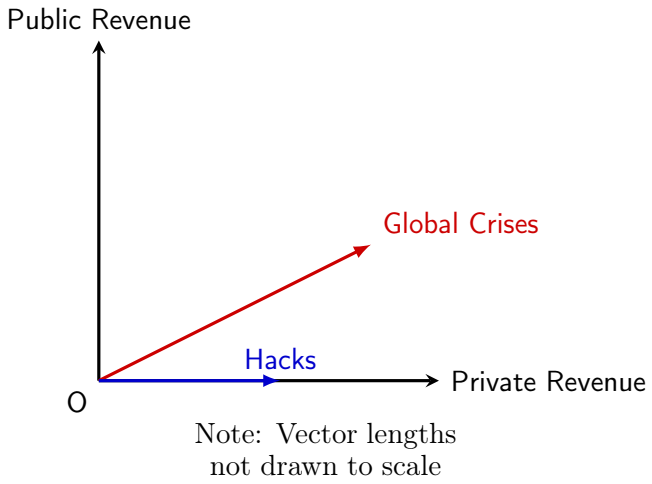


Figure 5: Effect of the instrumental variables on an average block's private and total revenue. Crypto crises increase private and total revenue by similar amounts, while Hacks tend to increase private revenue disproportionately—spanning our space of explanatory variables.

opportunities for insiders or those with early knowledge of the hack. This is evidenced by the differential impact of hacks in our first-stage results: the coefficient on log private revenue is positive, while the coefficient on log public revenue is a precise 0. Figure 5 illustrates the differential impact of the two instruments, and shows how they span the space of explanatory variables.

Results Table 4 shows the results of OLS and 2SLS regressions where $\theta_{B,t}$ is the dependent variable, and $\log Private_t$, $\log Public_t$ are the independent variables. Columns (1) and (2) show the OLS results without and with builder $\times$ proposer fixed effects. Columns (3) and (4) show the results from the instrumental variable regressions using the Hacks and Crises dummies as instruments—again without and with builder $\times$ proposer fixed effects.[15] The results using instrumental variables and fixed effects are very strong, showing that a 1% increase in the value of private arbitrages increases the builder's revenue share by 0.57%. We highlight that the coefficient on the revenue control is negative. This follows from a simple economic intuition: since a block can contain multiple sources of revenue, many of which are public, a larger revenue after accounting for private arbitrages will shift market power to the proposer, and away from the builder.

[15]We use the commands *reg*, *reghdfe*, *ivreg2* and *ivreghdfe* to compute each of these four columns. Note that when using reghdfe or using ivreghdfe the constant is not reported because it is a normalization factor chosen algorithmically to ensure all fixed effects have zero mean.

	(1) $\log Private_t$	(2) $\log Private_t$	(3) $\log Public_t$	(4) $\log Public_t$
Hack Dummy	0.0107***	0.0094***	-0.0011***	-0.0001
	(0.0006)	(0.0007)	(0.0003)	(0.0002)
Crisis Dummy	0.1187***	0.1230***	0.0219***	0.0139***
	(0.0095)	(0.0098)	(0.0015)	(0.0014)
Constant	0.0732***		0.0243***	
	(0.0015)		(0.0008)	
Observations	3924921	3419294	3924921	3419294

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table shows the first stage estimation results for our different 2SLS specifications. Columns (1) and (2) show how $\log Private$ is affected by our instrumental variables, without and with builder, proposer, and builder × proposer fixed effects. Columns (3) and (4) show analogous results for $\log Public$. All standard errors are clustered at the builder × proposer level.

Table 3: First Stage Regression Results

5 Model: Information-Driven Market Power

In this section, we develop a stylized model to formalize the central economic mechanism driving our empirical results: the process by which private information confers market power. The DeFi intermediation chain involves three key relationships: between arbitrageurs and block builders, builders and proposers, and proposers and ETH depositors. Our formal model abstracts from the first and third of these relationships to focus squarely on the core bargaining game between builders and proposers, where the rents from private information are ultimately divided. The market structures of the other two relationships, while important for the overall ecosystem, can be understood through the lens of more standard economic forces—namely vertical integration, coordination externalities, risk-sharing, and delegation—which we discuss at the end of this section.

We model the construction of the blockchain as the stationary steady state of an infinitely repeated game between two types of agents— proposers and block builders. Time is indexed by $t \in \{0, 1, 2, ..., \}$.There are N proposers, indexed by $n \in \{1, ..., N\}$, each with stake w_n. Motivated by the empirical observation that proposers' market shares of staked ETH are very stable over time, we assume that proposers' stake is constant. There are M block builders, indexed by $m \in \{1, ..., M\}$. Agents are profit maximizers and they do not discount the future.

	(1) OLS No FE	(2) OLS FE	(3) IV No FE	(4) IV FE
$\log Private_t$	0.155***	0.141***	0.422***	0.575***
	(0.0149)	(0.0187)	(0.0709)	(0.138)
$\log Public_t$	-0.126***	-0.0448***	-3.048***	-4.787***
	(0.0118)	(0.0113)	(0.363)	(1.106)
Constant	0.0390***		0.0906***	
	(0.00318)		(0.00588)	
N	3924921	3419294	3924921	3419294
F Statistic			417.05	65.16
Robust F Statistic			106.549	12.886

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table shows our multivariate estimation results when the builder profit share is the dependent variable. Columns (1) and (2) show OLS results, without and with builder, proposer, and builder × proposer fixed effects. Columns (3) and (4) show 2SLS results, without and with builder, proposer, and builder × proposer fixed effects. All standard errors are clustered at the builder × proposer level. The instrumental variables are $Hacked_t$ and $Crisis_t$.

Table 4: OLS and Two-Stage Least Squares Results

In each period t, a proposer n is chosen among the N proposers via the proof-of-stake consensus mechanism with probability $\psi_n = \frac{w_n}{\sum_{i=1}^{N} w_i}$ to add the next block to the blockchain. Let $B_{m,t}$ denote the block built by builder m at time t with total value $R_{m,t}$.[16] We denote the set of all the blocks created at time t by $\mathcal{B}_t = \{B_{m,t}\}_{m\in\{1,\dots,M\}}$.

At each time t, blocks are made up of three type of transactions: regular transactions, which have a low payoff and arise every period, and two types of arbitrage transactions that are high payoff but arise rarely. Thus, the majority of time periods are regular periods with no arbitrage opportunities. An arbitrage opportunity arises at small Poisson rate $\rho \ll 1$ per period. The arbitrage opportunity is public with i.i.d. probability π_u and private with complementary probability $\pi_r = 1 - \pi_u$. Public arbitrage transactions are known by all block builders while each private arbitrage transaction is known by a single block builder. A private arbitrage opportunity becomes publicly known at Poisson rate $\rho_d < 1$ per period, where $\rho_d \gg \rho$. As all block builders know each public arbitrage transaction, they all include it in their respective block. As such, a arbitrage transaction in period t is included in every $B_{m,t} \in \mathcal{B}_t$.

[16] One can assume that there is fixed cost $c > 0$ associated with building a block. It does not change any of the results and does not add any intuition, thus we set $c = 0$.

For expositional purposes, we assume a very simple process for transaction payoffs. We assume that the payoff of all arbitrage transactions, public or private, comes from the same distribution. The payoff of regular transactions come from of a different distribution that has a significantly lower support. More precisely, assume all blocks that include the arbitrage transaction in period t have the same high revenue, $\bar{R}_t$ and all the other blocks at time t have the same low revenue, $\underline{\mathrm{R}}_t$, and $\underline{\mathrm{R}}_t = \underline{\mathrm{R}}$ and $\bar{R}_t = \bar{R}$, $\forall t$.[17] Table 2 documents that the distribution of block revenue is strongly skewed to the right. Motivated by this empirical evidence, we assume $\bar{R} \gg \underline{\mathrm{R}}$.

Let "public blocks" denote all the blocks in a regular period or in a period with public arbitrages, with the same value $\widehat{R_t}$. The value $\widehat{R_t}$ varies across periods— low in regular periods, $\underline{\mathrm{R}}$, and high in periods with publicly known arbitrages, $\bar{R}$. Thus, $\widehat{R}_t \in \{\underline{\mathrm{R}}, \bar{R}\}$.

Alternatively, a "private block" includes a private arbitrage transaction. Recall that a private arbitrage that arrives in period t is picked up by a single builder, $\tilde{m}$, who incorporates it in his block $B_{\tilde{m},t}$.[18] In that sense, the private arbitrage opportunity is the "private information" of builder $\tilde{m}$, thus, we call $B_{\tilde{m},t}$ the "private block" at time t. The low Poisson arrival rate of arbitrage transactions imply that every other block at time t is a public block with low total revenue, i.e., $\forall\, B_{m,t} \in \mathcal{B}_t, m \neq \tilde{m}$, we have $R_{m,t} = \underline{\mathrm{R}}$.

With a slight abuse of notation, let p_t denote the proposer selected, b_t the block builder, B_t the block added to the blockchain, and R_t denote the block's revenue, in period t. Proposer p_t and block builder b_t *trade* and divide the *net* block value with exogenous bargaining powers $(\xi_P, \xi_B) = (1-\delta, \delta)$,and *endogenous* outside options $\Upsilon_{P,t}$ and $\Upsilon_{B,t}$ for the proposer and the block builder, respectively. The asymmetric bargaining powers can be simply derived from a variation of an alternating-offer bargaining game a la Rubinstein (1982) played by the proposer and block builder in virtual time within each period.[19]

[17] In general, $\underline{\mathrm{R}}_t$ and $\bar{R}_t$ can be random variables themselves with different supports, as long as the upper bound of support of $\underline{\mathrm{R}}_t$ is sufficiently smaller than the lower bound of support of $\bar{R}_t$. The extreme skewness of block revenue distribution documented in Table 2 supports this assumption.

[18] The assumption that an arbitrageur with a private arbitrage offers his block to a single block builder is consistent with the empirical pattern that block builders try to build market share in order to capture arbitrageurs. It is also the optimal strategy for the arbitrageur, as it ensures a high profit for him while almost certainly being added to the chain by safeguarding his information advantage.

[19] To be precise, assume the proposer and the block builder play an alternating-offer bargaining game a la Rubinstein (1982) in virtual time in each period t, and the block builder always makes the first offer. The proposer and the block builder face probabilities of within-period trade-breakdown $1-\delta_1$ and $1-\delta_2$, respectively. This implies $\delta = \frac{\delta_2(1-\delta_1)}{1-\delta_1\delta_2}$. Thus, in equilibrium, the initial block builder's offer, i.e. the observed bid of the block builder for each block, is set to achieve the profits implied by the bargaining game with bargaining powers $(\xi_P, \xi_B) = (1-\delta, \delta)$ and endogenous outside options $(\Upsilon_{P,t}, \Upsilon_{B,t})$.

The alternating-offer game of Rubinstein (1982) is a game of public information between the two bargaining parties. Note that the block builders and proposers play this game infinitely many times and after each block is added its information become public. Furthermore, within each period, the proposer has the monopoly to

The net value from adding block B_t to the blockchain is given by R_t, the total value generated by block B_t, minus the sum of outside options of the block proposer and the block builder, $S_t = R_t - \Upsilon_{P,t} - \Upsilon_{B,t}$. Consistent with the notation in Section 3, let $\Pi_{P,t}$ ($\theta_{P,t}$) and $\Pi_{B,t}$ ($\theta_{B,t}$) denote the profit (profit share) of the proposer and builder in period t, respectively. They are given by:

$$\Pi_{B,t} = \delta S_t + \Upsilon_{B,t},$$
$$\Pi_{P,t} = (1-\delta) S_t + \Upsilon_{P,t}.$$

Table 2 shows that the mean block builder's profit is very low, close to zero. As such, we are particularly interested in the case where $\delta \to 0$,[20] in which case $\Pi_{B,t}$ and $\Pi_{P,t}$ simplify to

$$\Pi_{B,t} = \Upsilon_{B,t}, \tag{2}$$
$$\Pi_{P,t} = R_t - \Upsilon_{B,t} \tag{3}$$

which in turn imply

$$\theta_{B,t} = \frac{\Upsilon_{B,t}}{R_t}, \tag{4}$$
$$\theta_{P,t} = \frac{R_t - \Upsilon_{B,t}}{R_t} = 1 - \frac{\Upsilon_{B,t}}{R_t} \tag{5}$$

Recall that block builder and proposer outside options, $\Upsilon_{B,t}$ and $\Upsilon_{P,t}$, are equilibrium outcomes that are determined endogenously. In turn, they determine the profit levels and profit shares of the block builder and the proposer. Equations (2), (3), (4) and (5) highlight the main intuition of the model— that the source of block builders' revenue is their outside option.

Proposition 1 summarizes the main theoretical results of the model that rely on this intuition and tightly connect to the empirical patterns documented in Section 4.

Proposition 1 (Information-Driven Market Power). *If arbitrage transactions are sufficiently more profitable than regular transactions, the following two statements hold*

1. *Higher value of private transactions in a given block increases the profit share of the block builder and decreases that of the proposer.*

choose the block of his choice and can punish any block builder who has not bid truthfully in a prior period. As such, we assume that the proposer can infer the true value of block B_t perfectly from the bid he observes at period t and thus we can use the alternating-offer game of Rubinstein (1982) within each period.

[20] This assumption is made for expositional purposes only. Proposition 1 holds for general $\delta < 1$. The proof is available upon request.

2. *Higher value of public transactions in a given block, controlling for the value of private transactions, has the opposite impact. It decreases the profit share of the block builder and increases that of the proposer.*

The proof of proposition 1 shows that block builders' outside option is governed by their private information. We call this the information-driven market power of block builders. In order to get some intuition for this result, it is most insightful to consider the determination of the outside options.

First, assume period t is a period with only public blocks, i.e., with no private arbitrage transactions. In this case, the selected proposer p_t can choose any $B_{m,t} \in \mathcal{B}_t$ and none of them has an advantage over the others. On the other hand, block builders cannot do anything other than offering their block to proposer p_t only in period t. In particular, any block that is chosen as B_t and is added to the blockchain at time t includes the same set of transactions. As such, all the other blocks in $\mathcal{B}_t$ lose their value as soon as B_t is added to the blockchain. This implies that the block builder b_t's outside option is zero. Thus, for the public blocks added to the blockchain Equations (2) and (3) reduce to

$$\Pi_{B,t}^{\text{public}} = 0 \tag{6}$$

$$\Pi_{P,t}^{\text{public}} = \widehat{R}_t \tag{7}$$

Next, consider a period t when there is a builder $\tilde{m}$ with a private block, $B_{\tilde{m},t}$. Let n denote the index of the proposer who is chosen in period t, and let $\psi_n = \frac{w_n}{\sum_{i=1}^{N} w_i}$ denote their share of the stake. The proposer p_t has the choice to pick the block $B_{\tilde{m},t}$ and add it to the blockchain at time t to have $B_t = B_{\tilde{m},t}$, or not.

Let X denote the expected profit that proposer p_t obtains from the private block $B_{\tilde{m},t}$ if he does not choose to add $B_{\tilde{m},t}$ at time t, i.e., $B_t \neq B_{\tilde{m},t}$. The proposer p_t can extract X from the total revenue of the private arbitrage block $B_{\tilde{m},t}$ at time t if he decides to choose it at time t and have $B_t = B_{\tilde{m},t}$.

If proposer p_t does not choose the private block $B_{\tilde{m},t}$ in period t to add to the blockchain, the private arbitrage transaction will not be exploited at time t. In each following period, it will turn public with probability ρ_d, in which case it will be added to the blockchain right away and exploited. Otherwise it will stay part of a private block that builder $\tilde{m}$ creates in the future periods until another proposer accepts the bid and adds the private block to the blockchain.

Proposer $p_t = n$ has an i.i.d. probability ψ_n to be chosen by the proof-of-stake consensus mechanism in each period t. In each future period $\tau > t$ that he is the proposer, $p_\tau = p_t$,

there are three contingencies. First, the private arbitrage transaction is still private and unexploited, in which case it is included in block $B_{\tilde{m},\tau}$ and proposer p_t can obtain X from choosing $B_\tau = B_{\tilde{m},t}$. Second, the private arbitrage just turned public, in which case it is included in every block at period τ and proposer p_t obtains $\bar{R}$ from it. Third, the private arbitrage has already turned public in period τ', $t < \tau' < \tau$ and is already exploited by the proposer in that period, in which case proposer p_t gains nothing from it in period τ.[21]

As such, an upper bound for X is:

$$\bar{X} = \frac{1-(1-\psi_n)(1-\rho_d)}{\rho_d}\underline{\mathrm{R}} + \psi_n \bar{R}$$

which is increasing in ψ_n and decreasing in ρ_d. These comparative statistics are intuitive: A higher stake share ψ_n implies that it is more likely for p_n to be the proposer next period and add this block. Alternatively, an increase in ρ_d implies that it is more likely that this private arbitrage transaction turns to a public one before proposer p_t is chosen again, in which case it is exploited by some other proposer. Put differently, $\frac{1}{\psi_n}$ and ρ_d act as the effective discount rate for proposer p_t.

For any pair (ψ_n, ρ_d) such that $\psi_n < 1$ and $\rho_d > 0$, $\underline{R} < \frac{\rho_d(1-\psi_n)}{\rho_d(1-\psi_n)+\psi_n}\bar{R}$ guarantees $\bar{X} < \bar{R}$. In other words, when $\underline{\mathrm{R}} \ll \bar{R}$ proposer n is willing to leave profit $\bar{R} - \bar{X}$ for block builder $\tilde{m}$ in order to be able to add the block $B_{\tilde{m},t}$ in period t to the blockchain. Thus, using Equations (2) and (3), for a private block we have

$$\Pi_{B,t}^{\text{private}} \geq (1-\psi_n)\bar{R}_t - \frac{1-(1-\psi_n)(1-\rho_d)}{\rho_d}\underline{\mathrm{R}} > 0 \;\Rightarrow\; \theta_{B,t}^{\text{private}} > 0 \tag{8}$$

$$\Pi_{P,t}^{\text{private}} \leq \frac{1-(1-\psi_n)(1-\rho_d)}{\rho_d}\underline{\mathrm{R}} + \psi_n\bar{R} < \bar{R} \;\Rightarrow\; \theta_{P,t}^{\text{private}} < 1 \tag{9}$$

Comparing Equation (6) with (8) and (7) with (9) clearly illustrates the opposite impact of the private arbitrage opportunities on profit shares of block builders and proposers. It also shows that higher block revenue leads to a higher share for the block builder only if the block revenue comes from private arbitrages. All the rest of the block revenue is captured by the proposer and thus increases his profit share, while decreasing the profit share of the block builder. As such, Proposition 1 provides a consistent mechanism for the empirical findings of Table 4.

Equation (8) also implies that a private arbitrage transaction which remains private

[21] To be precise, this calculation gives an upper bound on X as it assumes that if the private arbitrage transaction is still private at time τ, in the interim periods τ', $t < \tau' < \tau$, none of the proposers have added the block $B_{\tilde{m},\tau'}$ to the blockchain, so the private arbitrage opportunity remains unexploited until time τ. This only strengthens the result as we need $X < \bar{R}$, which is true if an upper bound in X is less than $\bar{R}$.

longer leads to a higher profit share for the block builder. This is intuitive as it corresponds to a longer duration for the limited availability of the arbitrage transaction, which in turn makes the private information of the block builder more valuable and improves his bargaining position.[22]

5.1 Market Structure Beyond the Core Bargaining Game

Our model restricts attention to the economic mechanism that leads to the rent sharing patterns between block builders and block proposers, as documented in Table 4. In what follows, we explain how standard economic forces give rise to the remaining relationships that our core model above abstracts away from—those between arbitrageurs and block builders and between block proposers and ETH depositors.

The Arbitrageur-Builder Relationship. The concentration observed in the block builder market is sustained by two distinct economic forces that segment arbitrageurs. First, for large, sophisticated trading firms, the optimal strategy is vertical integration. Firms such as Symbolic Capital Partners and Wintermute, which generate substantial high-value, private order flow, operate their own exclusive builders (BeaverBuild and RSync, respectively). This structure allows them to internalize the full value of their information, avoid leakage to competitors, and ensure their transactions are prioritized. By directing their proprietary flow to an in-house builder, they solve a fundamental commitment problem and maximize the probability of their arbitrage's execution.

Second, for smaller, independent arbitrageurs who lack the scale to operate their own builder, the market resembles a coordination game with network externalities. These arbitrageurs face a strong incentive to route their transactions to a single, dominant public builder (e.g., Titan). A builder's probability of winning the block auction is increasing in the total value of the transactions it aggregates. Consequently, by coordinating on a single builder, small arbitrageurs collectively increase the value of that builder's blocks, enhancing their own execution probability and expected profits. In equilibrium, it is not rational for an individual small arbitrageur to deviate to a less popular builder, as doing so would place

[22]It is worth mentioning that we have abstracted away from any relationship building between proposers and block builders, which is what gives rise to the negative builder profit shares. Furthermore, this simplified model does not address the determination of stakes of the proposers and the detailed interaction between block builders and arbitrageurs. These simplifications are crucial to highlight the main mechanism that gives rise to information-driven market power for block builders. We plan to incorporate the stylized model in a full model of the DeFi Intermediation chain that includes arbitrageurs, block builders, proposers and depositors and features inter-period dependencies. The full model includes optimal strategy of arbitrageurs as well as depositors and determines M, N and $\{w_1, \cdots, w_n\}$ endogenously.

their transaction in a block with a lower probability of being selected, stranding their capital and arbitrage opportunity.

The Proposer-Depositor Relationship. The proposer market exhibits a similar pattern of concentration, driven by the classic mechanisms of risk-sharing and delegation. The proof-of-stake consensus algorithm allocates the right to propose a block probabilistically, based on the amount of staked capital. For an individual depositor, this creates a high-variance return profile: long periods of no income punctuated by rare, large rewards. Risk-averse depositors, who prefer a smooth consumption path, have a clear incentive to pool their capital with large, professional proposers. These staking pools can absorb the idiosyncratic risk of the selection lottery and provide depositors with a stable, predictable stream of income in exchange for a fee. These incentives are strongly amplified by the fat-tailed distribution of block revenue.

Furthermore, participating as a proposer requires significant technical expertise and operational security to run and maintain the necessary software and hardware without interruption. Deviations from the protocol, even if accidental, can be penalized through 'slashing,' resulting in a loss of staked capital. This creates a strong motive for delegation. Depositors can avoid these operational burdens and risks by entrusting their assets to specialized firms that benefit from economies of scale in technology, security, and expertise. The resulting market structure is a natural oligopoly of large, trusted proposers who act as delegated asset managers for a diffuse base of capital holders.

In addition—as shown in Figure 4—we observe that the empirical market shares of proposers are relatively stable over time, so that the proposer fixed effects in our regressions account for the relative market power that different proposers may have due to their large concentrations of ETH deposits.

6 Robustness

In our main analysis, we removed blocks where the builder profit share was less than -10%, consisting of about 5% of the sample, to prevent outliers from skewing the results. In this section, we show our results hold—in fact they are stronger—when we don't remove these blocks. We also show that the results hold when we condition on the builder profit share being non-negative, and when we use alternative definitions of private transactions.

Appendix C presents the results of these robustness exercises. Importantly, in each of these robustness checks, we obtain a significant positive effect for the value of the builder's private information on their profit share. Furthermore, our instrumental variables based on

crypto crises and cyber-attacks remain strong throughout all our alternative specifications.

6.1 Blocks With And Without Subsidies

Table 5 in Appendix C displays the summary statistics of the full sample. It shows that there are blocks where the builder gives a very large subsidy to the proposer, with the largest subsidy being 56.13 ETH. This subsidy skews the builder profit share $\theta_{P,t}$ heavily, with the share being highly negative for blocks with large subsidies. In the main text, we restrict ourselves to blocks with a builder profit share $\theta_{B,t} \geq -0.1$. In this section, we show that our findings are robust to including all the blocks in the sample.

Tables 6 and 7 in Appendix C present the first and second stage estimation results for the full sample, including blocks that contain transactions with large proposer subsidies. The instruments stay strongly significant and the effect of private information is amplified around three folds.

6.2 Alternative Definition of Private Information

In the main analysis, we use a simple definition of private information, where a transaction is private if and only if it is not broadcast to the network before being appended to the blockchain. We use this definition in order to be able to utilize the classification of private versus public transactions provided by a third party, Secure Decentralized Systems Lab's Mempool Guru project (Yang et al., 2024).

However, there may be many arbitrage opportunities trades—while sent privately to a builder—are observed by many entrepreneurs. For example, any price discrepancies between decentralized exchanges (DEXs) on the blockchain may be observed by multiple entrepreneurs who have algorithms scanning the blockchain for such trades. Even if we observe an arbitrageur sending such transactions privately to a builder, it is possible that other arbitrageurs have sent them to other builders, making the arbitrage essentially public.

Because we know the details of these arbitrages, we can apply a number of heuristics which identifies them as essentially public albeit having been sent privately. Such an arbitrage meets all of the following criteria. First, these arbitrages consist of at least two transactions: one which is intended to buy low and another to sell high. Second, the gross position of this strategy must be greater than or equal to zero. Finally, the net position of all but the first and last transaction must net to zero, so that the strategy forms a loop. These heuristics are akin to those used by research teams such as EigenPhi to detect atomic arbitrages. These transactions are then labeled as public and excluded from the pool of private transactions.

The intuition for this heuristic is that an arbitrage is private if it can only be discovered using some private, off-chain signal. The vast majority of private arbitrages are arbitrages between centralized exchanges (CEXs) and decentralized exchanges. A CEX-DEX arbitrage takes advantage of mispricings quoted between two or more exchanges, but one is a centralized exchange whose prices and orders are off-chain. In contrast to DEX-DEX arbitrages, its legs are not executed simultaneously, so there is inventory risk as the arbitrageur holds the off-chain position. For this reason, an arbitrageur wants their public position to execute as soon as possible and exclusively, so they will certainly include a direct payment to a block builder and do so privately. Moreover, as only the on-chain leg is observable on the blockchain, the strategy looks like a swap between two tokens, one of which is traded on a centralized exchange.

Appendix C present the first and second stage estimation results using the alternative definition of private information. The instruments stay strongly significant and the effect of private information is amplified more than six folds.

It is worth noting that an interesting qualitative difference surfaces with the alternative definition of private information. While in all other specification, OLS and IV, value of public transactions negatively impact the profit share of the block builder, in Table 9 the OLS coefficient for $\log Public_t$ is positive and becomes negative only when using instrumental variables. We believe the reason is that as the definition of private arbitrage is much more restrictive, many transactions which would have been classified as private in our main analysis are no longer classified as such in this specification. Therefore, a block with a large number of transactions has many other sources of revenue for the builder that are not covered in the alternate definition of $\log Private_t$.

7 Conclusion

We examine the emergence of centralization in markets that are specifically designed to be decentralized by considering the impact of private information on the profit shares of financial intermediaries in Decentralized Finance (DeFi). Using novel DeFi transaction data, we find that the combination of depositor risk sharing and arbitrageurs' need for privacy leads to the emergence of block proposers and block builders as respective intermediaries in the Ethereum blockchain. Furthermore, we quantify the intermediation profits generated by block builders' access to valuable private information.

Employing an instrumental variable approach using crypto crises and thefts of funds from crypto institutions and protocols as instruments, we find that a 1% increase in the value of

private information leads to a 0.57% increase in the block builder profit share. We propose a repeated bargaining model to provide an economic mechanism for our empirical findings. This evidence highlights the crucial role of private information in determining the revenue shares of intermediaries in decentralized financial markets. As traditional and decentralized finance become increasingly interconnected, understanding the dynamics of intermediation in decentralized markets becomes increasingly relevant for both academics and practitioners.

References

Abadi, Joseph and Markus Brunnermeier, "Blockchain economics," Working Paper 25407, National Bureau of Economic Research December 2018.

Allen, Franklin, "The market for information and the origin of financial intermediation," *Journal of Financial Intermediation*, 1990, *1*, 3–30.

_ **and Douglas Gale**, "Financial markets, intermediaries, and intertemporal smoothing," *Journal of Political Economy*, 1997, *105* (3), 523–546.

Athey, Susan, Ivo Parashkevov, Vishnu Sarukkai, and Jing Xia, "Bitcoin pricing, adoption, and usage: Theory and evidence," *Stanford University Graduate School of Business Research Paper*, 2016, (16).

Barbon, Andrea, Marco Di Maggio, Francesco A Franzoni, and Augustin Landier, "Brokers and order flow leakage: Evidence from fire sales," *The Journal of Finance*, 2019, *74* (6), 2707–2749.

Battalio, Robert, Andriy Shkilko, and Robert Van Ness, "To pay or be paid? The impact of taker fees and order flow inducements on trading costs in U.S. options markets," *The Journal of Financial and Quantitative Analysis*, 2016, *51* (5), 1637–1662.

Biais, Bruno, Christophe Bisière, Matthieu Bouvard, and Catherine Casamatta, "The blockchain folk theorem," *The Review of Financial Studies*, 2019, *32* (5), 1662–1715.

Boot, Arnoud W.A. and Anjan V. Thakor, "Financial system architecture," *The Review of Financial Studies*, 1997, *10* (3), 693–733.

Budish, Eric, "The economic limits of bitcoin and the blockchain," Working Paper 24717, National Bureau of Economic Research June 2018.

_ , "Trust at scale: The economic limits of cryptocurrencies and blockchains," *The Quarterly Journal of Economics*, 2025, *140* (1), 1–62.

Buterin, Vitalik, "A next-generation smart contract and decentralized application platform," *Whitepaper*, 2014.

Campbell, Tim S and William A Kracaw, "Information production, market signalling, and the theory of financial intermediation," *The Journal of Finance*, 1980, *35* (4), 863–882.

Capponi, Agostino, Ruizhe Jia, and Kanye Ye Wang, "Maximal extractable value and allocative inefficiencies in public blockchains," *Journal of Financial Economics*, 2025, *172*, 104–132.

_ , _ , and Sveinn Olafsson, "Proposer-builder separation, payment for order flows, and centralization in blockchain," Technical Report 4723674, Social Science Research Network 2024.

Cong, Lin William and Zhiguo He, "Blockchain disruption and smart contracts," *The Review of Financial Studies*, 2019, *32* (5), 1754–1797.

_ , Ke Tang, Yanxin Wang, and Xi Zhao, "Inclusion and democratization through web3 and defi? Initial evidence from the ethereum ecosystem," Working Paper 30949, National Bureau of Economic Research February 2023.

_ , Ye Li, and Neng Wang, "Tokenomics: Dynamic adoption and valuation," *The Review of Financial Studies*, 2021, *34* (3), 1105–1155.

_ , Zhiguo He, and Jiasun Li, "Decentralized mining in centralized pools," *The Review of Financial Studies*, 2021, *34* (3), 1191–1235.

_ , Zhiheng He, and Ke Tang, "The tokenomics of staking," Working Paper 33640, National Bureau of Economic Research April 2025.

Crouzet, Nicolas, Janice C. Eberly, Andrea L. Eisfeldt, and Dimitris Papanikolaou, "The Economics of Intangible Capital," *Journal of Economic Perspectives*, August 2022, *36* (3), 29–52.

Daian, Philip, Steven Goldfeder, Tyler Kell, Yunqi Li, Xueyuan Zhao, Iddo Bentov, Lorenz Breidenbach, and Ari Juels, "Flash boys 2.0: Frontrunning in decentralized exchanges, miner extractable value, and consensus instability," in "2020 IEEE symposium on security and privacy (SP)" IEEE 2020, pp. 910–927.

Diamond, Douglas W, "Financial intermediation and delegated monitoring," *The Review of Economic Studies*, 1984, *51* (3), 393–414.

_ and Philip H Dybvig, "Bank runs, deposit insurance, and liquidity," *Journal of Political Economy*, 1983, *91* (3), 401–419.

_ and Raghuram G Rajan, "Liquidity risk, liquidity creation, and financial fragility: A theory of banking," *Journal of Political Economy*, 2001, *109* (2), 287–327.

Fanti, Giulia, Leonid Kogan, and Pramod Viswanath, "Economics of proof-of-stake payment systems," *MIT Sloan Working Paper*, 2021, (5845).

Farboodi, Maryam, Gregor Jarosch, Guido Menzio, and Ursula Wiriadinata, "Intermediation as rent extraction," *Journal of Economic Theory*, 2025, p. 106029.

Ferreira, Daniel, Jin Li, and Radoslawa Nikolowa, "Corporate capture of blockchain governance," *The Review of Financial Studies*, 2023, *36* (4), 1364–1407.

Gorton, Gary and Andrew Winton, "Financial intermediation," in "Handbook of the Economics of Finance," Vol. 1, Elsevier, 2003, pp. 431–552.

Green, Richard C, Burton Hollifield, and Norman Schürhoff, "Dealer intermediation and price behavior in the aftermarket for new bond issues," *Journal of Financial Economics*, 2007, *86* (3), 643–682.

Guerrieri, Veronica and Robert Shimer, "Dynamic adverse selection: A theory of illiquidity, fire sales, and flight to quality," *American Economic Review*, 2014, *104* (7), 1875–1908.

_ , _ , and Randall Wright, "Adverse selection in competitive search equilibrium," *Econometrica*, 2010, *78* (6), 1823–1862.

Gupta, Tivas, Mallesh M Pai, and Max Resnick, "The centralizing effects of private order flow on proposer-builder separation," *arXiv: Preprint*, 2023, *2305* (19150).

Han, Munhee, Sanghyun Kim, and Vikram K. Nanda, "Splitting and shuffling: Institutional trading motives and order submissions across brokers," Technical Report 3429452, Social Science Research Network 2020.

Heimbach, Lioba, Vabuk Pahari, and Eric Schertenleib, "Non-atomic arbitrage in decentralized finance," in "2024 IEEE Symposium on Security and Privacy (SP)" 2024, pp. 3866–3884.

Hollifield, Burton, Artem Neklyudov, and Chester Spatt, "Bid-ask spreads, trading networks, and the pricing of securitizations," *The Review of Financial Studies*, 2017, *30* (9), 3048–3085.

Huberman, Gur, Jacob D Leshno, and Ciamac Moallemi, "Monopoly without a monopolist: An economic analysis of the bitcoin payment system," *The Review of Economic Studies*, 2021, *88* (6), 3011–3040.

Kyle, Albert S, "Continuous auctions and insider trading," *Econometrica*, 1985, *53* (6), 1315–1335.

Lehar, Alfred and Christine A Parlour, "Battle of the bots: Flash loans, miner extractable value and efficient settlement," Technical Report 4382292, Social Science Research Network 2023.

Leland, Hayne E and David H Pyle, "Informational asymmetries, financial structure, and financial intermediation," *The Journal of Finance*, 1977, *32* (2), 371–387.

Li, Dan and Norman Schürhoff, "Dealer networks," *The Journal of Finance*, 2018, *74* (1), 91–144.

Liu, Jiageng, Igor Makarov, and Antoinette Schoar, "Anatomy of a run: The terra luna crash," Working Paper 31160, National Bureau of Economic Research April 2023.

Maggio, Marco Di, Amir Kermani, and Zhaogang Song, "The value of trading relations in turbulent times," *Journal of Financial Economics*, 2017, *124* (2), 266–284.

Makarov, Igor and Antoinette Schoar, "Trading and arbitrage in cryptocurrency markets," *Journal of Financial Economics*, 2020, *135* (2), 293–319.

_ and _, "Blockchain analysis of the bitcoin market," Working Paper 29396, National Bureau of Economic Research October 2021.

Milionis, Jason, Ciamac C. Moallemi, Tim Roughgarden, and Anthony Lee Zhang, "Automated market making and loss-versus-rebalancing," *arXiv: Preprint*, 2023, *2208* (06046).

Nakamoto, Satoshi, "Bitcoin: A peer-to-peer electronic cash system," *Whitepaper*, 2008.

Pagnotta, Emiliano and Andrea Buraschi, "An equilibrium valuation of bitcoin and decentralized network assets," Technical Report 3142022, Social Science Research Network 2018.

Prat, Julien and Benjamin Walter, "An equilibrium model of the market for bitcoin mining," *Journal of Political Economy*, 2021, *129* (8), 2415–2452.

Raskin, Max and David Yermack, "Digital currencies, decentralized ledgers and the future of central banking," in "Research handbook on central banking," Edward Elgar Publishing, 2018, pp. 474–486.

Robinson, Dan and Georgios Konstantopoulos, "Ethereum is a dark forest," August 2020. Accessed: 2024-04-04.

Rubinstein, Ariel, "Perfect equilibrium in a bargaining model," *Econometrica*, 1982, *50* (1), 97–109.

Saleh, Fahad, "Blockchain without waste: Proof-of-stake," *The Review of Financial Studies*, 2021, *34* (3), 1156–1190.

Schwarz-Schilling, Caspar, Fahad Saleh, Thomas Thiery, Jennifer Pan, Nihar Shah, and Barnabé Monnot, "Time is money: Strategic timing games in proof-of-stake protocols," in "5th Conference on Advances in Financial Technologies," Vol. 282 Schloss Dagstuhl-Leibniz-Zentrum für Informatik 2023, pp. 1–17.

Shleifer, Andrei and Robert W Vishny, "The limits of arbitrage," *The Journal of Finance*, 1997, *52* (1), 35–55.

Sockin, Michael and Wei Xiong, "A model of cryptocurrencies," *Management Science*, 2023, *69* (11), 6684–6707.

Terry, Stephen J, Toni M Whited, and Anastasia A Zakolyukina, "Information versus Investment*," *The Review of Financial Studies*, 08 2022, *36* (3), 1148–1191.

Yang, Sen, Fan Zhang, Ken Huang, Xi Chen, Youwei Yang, and Feng Zhu, "SoK: MEV countermeasures," in "DeFi '24: Proceedings of the Workshop on Decentralized Finance and Security" Association for Computing Machinery 2024, pp. 21–30.

Zhu, Haoxiang, "Do dark pools harm price discovery?," *The Review of Financial Studies*, 2013, *27* (3), 747–789.

Appendix

A Institutional Details

The Bitcoin Blockchain and Proof-of-Work The bitcoin blockchain (Nakamoto (2008)) is both the first blockchain ever created, and the largest by market capitalization. The goal of the blockchain is to achieve consensus about who owns how many units of a digital asset, the bitcoin cryptocurrency (BTC). This consensus is established by a proof-of-work protocol, where (approximately) every 10 minutes a new block of transactions is "mined" and appended to the blockchain. As a payment for their service, the miner receives both a mining reward—reflected by the minting of new bitcoin which are credited to the miner's balance—and transaction fees paid by users who want their transactions included in the block. In every one of these 10 minute intervals, there is competition among users to be the miner and collect the rewards. In the most simple terms, the miner is the first user who can solve a cryptographic puzzle—the solution of which can be verified by all other participants.[23]

Because there is a competition to mine the next block, the bitcoin blockchain essentially has an all-pay auction every 10 minutes, where prospective miners perform trillions of computations attempting to be the first to solve the cryptographic puzzle. This competition is very wasteful and does not allow for high throughput of transactions. In addition, the bitcoin blockchain has a drawback in that it only keeps track of bitcoin balances, but does not have provisions for generating consensus on the balances of other assets.

The Ethereum Blockchain and Smart Contracts The Ethereum blockchain is the second largest blockchain by market capitalization, and the largest blockchain that allows the execution of general smart contracts (Buterin (2014)).[24] The native cryptocurrency of the Ethereum blockchain is Ether, or ETH for short.

Up until September 2022, Ethereum achieved consensus through a proof-of-work algorithm, which immediately led to challenges for operating smart contracts at scale. Since proof-of-work algorithms have very low throughput, the demand for smart contract operations was much larger than the available computing power of the Ethereum Virtual Machine,[25] leading to high transaction fees and very volatile congestion charges.

[23]The consensus algorithm of bitcoin is more complex than described in this short paragraph, with incentives designed to prevent participants from re-mining a block. Interested readers are directed to the original bitcoin whitepaper in Nakamoto (2008).

[24]The Bitcoin blockchain allows the execution of a restricted set of smart contracts through bitcoin script, but the feasible operations are very limited compared to the Turing-Complete Ethereum Virtual Machine.

[25]The bottleneck here is not the raw computing power of an individual node, but rather the amount of

Proof-Of-Stake Consensus One way to address these challenges is a proof-of-stake algorithm, where block proposers get chosen randomly with probability proportional to their stake.[26] To prevent a rogue block proposer from appending an invalid block to the blockchain (e.g. one that has double spending), a small group of verifiers is also chosen at random. The verifiers attest to the block's validity. As long as the stake is sufficiently distributed, the block proposer and verifiers will be independent with very high probability, and a valid block will be added to the chain. Since only a very small fraction of participants needs to be sampled to ensure the correctness of each new block, the amount of computation needed to obtain consensus is vastly reduced.

Consensus Layer Yield In Ethereum, all participants who stake their ETH receive some yield for accurately executing the consensus protocol. For every block, both the block proposer and verifiers receive some reward (in ETH) for accurately participating in the protocol. This reward varies depending on how much aggregate ETH has been staked, the time it takes for the verifiers to produce their attestation of correctness, and of course, the correctness of the proposed block. We define the Consensus Layer Reward as the expected payment (in ETH) to the block proposer and verifiers for correctly participating in the consensus protocol.

$$R_{consensus} = \mathbb{E}[\text{Reward from Participating in Consensus}]$$

Since the probability of being chosen as a proposer or verifier is proportional to the amount of ETH staked, the expected reward that one obtains from participating in the protocol is also proportional to the amount of ETH one has staked. Therefore, this reward can be interpreted as a yield on the staked Ether.

$$y_{consensus} = \frac{\mathbb{E}[R_{consensus}]}{\text{Amount of ETH staked}}.$$

Execution Layer Yield The most popular smart contracts on Ethereum are decentralized finance applications, including Crypto-Collateralized stablecoins such as MakerDAO, and decentralized exchanges such as Uniswap and Curve. Any user of the Ethereum blockchain can attempt to find arbitrage opportunities arising from these protocols. For example, an underwater MakerDAO loan can be liquidated at fire sale prices, and the collateral can be

computing power needed to agree on the state of the blockchain at any given time, including the state of all the smart contracts being executed.

[26] In practice, the participants in proof-of-stake algorithms use a decentralized pseudorandom number generator, which is implemented with cryptographic tools to prevent any coalition below a given size from biasing the random number generation process.

immediately sold in a decentralized exchange at a higher price, yielding an instant arbitrage in two transactions.[27] Similarly, price discrepancies among the hundreds of different decentralized exchanges can lead to arbitrage opportunities.

Since the data on these applications is public, there can be many arbitrageurs competing to exploit all of these arbitrage opportunities. This gives Ethereum block proposers some power to decide who obtains these arbitrage profits. When determining the order of transactions in a block, the block proposer can prioritize some arbitrageurs over others. The extreme case of this is when the block proposer observes the incoming transactions and front-runs arbitrage opportunities that are suggested to them by potential arbitrageurs. In the long run, this would dissuade the arbitrageurs from operating, or would lead to vertical integration between arbitrageurs and block proposers. In the data we don't observe this vertical integration. Instead, we see that there are specialized arbitrageurs—called *block builders*, who share some of their surplus with the block proposers. This sharing of the surplus of the execution layer reward gives the block proposers some expected income per block from arbitrage opportunities. We define the Execution Layer Reward as the expected payment to the block proposer (in ETH) from arbitrageurs for incorporating their transactions into a block.

$$R_{execution} = \mathbb{E}[\text{Block Proposer Reward from Arbitrage Opportunities}]$$

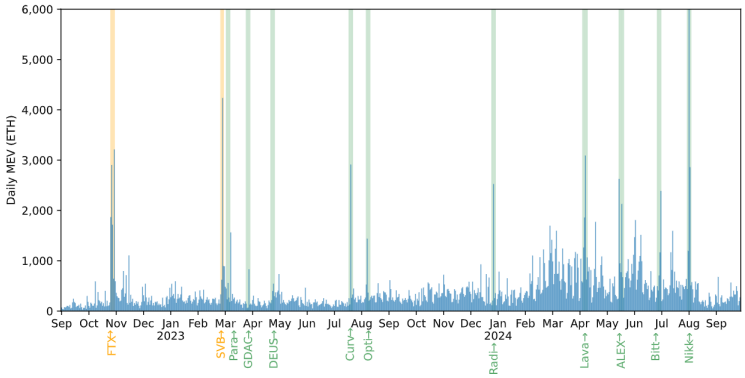
Since the probability of being a block proposer is proportion the amount of ETH staked, we can interpret this as a yield

$$y_{execution} = \frac{\mathbb{E}[R_{execution}]}{\text{Amount of ETH staked}}.$$

Maximal Extractable Value (MEV) The Maximal Extractable Value of a block represents the revenue that can be extracted from the ordering of transactions in a block, which is in excess of revenue from the value of transactions alone. Figure 6 shows the time series of MEV since the merge. We observe that the execution layer component is more volatile than the consensus layer component.

MEV Searchers MEV searchers are automated arbitrageurs who identify mispricings and the potential for near-riskless profit and bundle together transactions that, upon being incorporated into the blockchain, execute their arbitrage strategy. The success of these

[27] The arbitrage is instant because the transaction that buys the collateral at fire sale prices and the one that sells the purchased collateral in decentralized exchanges occur in the same block.



Source: Dune Analytics

Figure 6: Daily Gas and MEV Revenue

strategies is contingent on immediately capitalizing on public and private information, so their transactions carry high priority fees and even direct payments to builders in order to guarantee their incorporation towards the front of the next block. Because the immediacy and position of these transactions matter, their value is considered MEV.

The Market Inefficiencies of MEV The arbitrage opportunities in decentralized finance, combined with the decentralized consensus protocol of the Ethereum blockchain, create economic inefficiencies. First, there may be competition among arbitrageurs to get their transactions incorporated into blocks—and to prevent competitors from placing their transactions. For example, an arbitrageur may pay transaction fees high enough to buy the entire space in a block, preventing anybody else from interfering in their trades. Additionally, there is a problem with front-running. If an arbitrageur finds a profitable trade and submits it directly to a block proposer, there is no inherent reason besides reputation for why the proposer can't just clone the transaction and submit it themselves to collect the profit. In the long run, this discourages arbitrageurs from participating in the market.

Proposer-Builder Separation To prevent front-running, members of the Ethereum community advocated for the principle of *Proposer-Builder Separation* (PBS). Under this principle, the *builder* who collects all the transactions in a block, including the profitable arbitrage opportunities, is not the same as the *block proposer* who is chosen by the consensus protocol to propose the next block. Instead, there are multiple builders—in essence arbitrageurs—who compete to build the most profitable block of transactions. The builders will collect all the MEV of the block, and split this revenue with the proposer through a *proposer fee*—essentially a bid that incentivizes the proposer to choose the builder's block over all others.

In order to prevent front-running, the process through which these blocks are proposed is as follows

- **Block Builder's Action** Block builder i creates a block B_i. She submits a pair (B_i, p_i) to a relay, where p_i is the proposed payment to the proposer.
- **Relayer's Action** The relay j receives multiple pairs $(B_{i_1}, p_{i_1}), ..., (B_{i_n}, p_{i_n})$. The relay verifies that the blocks are valid (and potentially, that they don't have transactions from sanctioned Ethereum accounts), and chooses the highest bid (B_j^*, p_j^*) among the valid block proposals.

  Each relay j communicates (H_j^*, p_j^*) to the block proposer, where $H_j^* = H(B_j^*)$ is a hash function of the block B_j^*. Since the block proposer only observes a hash of the block—and hash functions are essentially random—[28] the block proposer at this time learns nothing which would allow her to front-run the arbitrage opportunities collected in the block.
- **Block Proposer's Action** The block proposer may either
  1. choose a relay j^* who "wins" the round—in which case the relay j^* reveals the block B_j^* to the block proposer; or
  2. the block proposer rejects all bids and proposes some "outside-option" block B_{out} that they construct themselves.
- **Payoffs** The payoffs of the game are as follows
  1. If the block proposer accepts the bid (H_j^*, p_j^*), she will receive a payoff of p_j^*. The block builder will receive both the consensus layer and execution layer reward. We assume the relay is competitive, and receives zero payoff.[29]
  2. If the block proposer rejects all bids, she receives both the consensus layer and execution layer rewards associated with B_{out}. In this case, the payoff to the block

[28] The technical term here is that a hash function is computationally hiding. Under widely accepted computational assumptions such as the existence of one-way functions, the receiver of a message H_j^* would have to do an astronomical amount of computation to recover an input B_j^* such that $H(B_j^*) = H_j^*$. This is also true if the received wanted to partially recover some bits from the input B_j^*.

[29] The assumption that relays are competitive seems to line up with the observed data. There are multiple relays, and both block builders and block proposers can connect to more than one relay. Furthermore, the code for relays is open-source, and therefore non-exclusive and non-rival. In practice, there is some vertical integration between relayers and block builders, with flashbots and Bloxroute operating both block builder bots as well as relays. Since the builders have to trust the relays not to front-run them, there is an informational advantage for block builders to operate their own relaying software.

builder is 0. In addition—as in the previous case—the relayer is assumed to be competitive and has 0 payoff.

MEV-Boost In practice, there is open source software, called *MEV-Boost*, which implements this proposer-builder separation. After the transition to proof-of-stake, MEV-Boost gained widespread adoption, with around 90% of blocks in Ethereum being selected through MEV-Boost, and around 75% of all blocks having different builders and proposers. [30] The most popular relay is the flashbots relay. However, it has recently faced increased competition from other relays. The main difference between flashbots and their competitors is that flashbots will not accept any block that contains transactions with accounts sanctioned by the Treasury's Office of Foreign Asset Control (OFAC). Many other relays, including Bloxroute-Max-Profit, do not take OFAC regulations into account when deciding which blocks to accept.

B Proofs

Proof Proposition 1. It is sufficient to characterize $\Upsilon_{B,t}$ across different periods.

Classify periods into two groups: First, periods where all transactions are public, (whether a public arbitrage opportunity is present or not), and second, periods where a private arbitrage transaction happens. Note that as arbitrage opportunities arise with a small Poisson rate independently, the probability that two arbitrage opportunities of any kind happen in the same period is vanishingly small.

In a public period, every block has the exact same value. As such, all block builders act as perfect competitors and undercut each other. Therefore, all the profits from adding a block to the blockchain in this period are appropriated by the proposer chosen by the proof-of-stake consensus mechanism. As such, $\Upsilon_{B,t} = \theta_{B,t} = 0$, $\theta_{P,t} = 1$.

Next, consider a private period t. We would like to show that $\Upsilon_{B,t} > 0$ in private periods. Let $B_{\tilde{m},t}$ denote the private block and n denote the chosen proposer (i.e., $p_t = n$). When $\delta \to 0$, the builder and proposer profit shares are given by Equations (4) and (5), respectively. Let X be the profit proposer n can obtain from the existence of the private arbitrage transaction which is in block $B_{\tilde{m},t}$, if he doesn't choose $B_t = B_{\tilde{m},t}$ in period t.

In order to show that $\Upsilon_{B,t} > 0$, it is more convenient to characterize X and prove that it is smaller than $\bar{R}$. To characterize X, consider the following observations. First, if n

[30] The data on MEV-Boost, including relay and block builder market shares, is obtained from Toni Wahrstätter's website `https://mevboost.pics/`, and is augmented with data from the individual relays' websites.

doesn't choose $B_{\tilde{m},t}$ now, he chooses a public block with value $\underline{\mathrm{R}}_t$ in period t. Second, in the stationary steady state, in each future period $\tau > t$, he will be the proposer with probability ψ_n. In each such period, there are three contingencies: 1) the private arbitrage transaction is still private and unexploited, in which case it is included in block $B_{\tilde{m},\tau}$ and proposer n adds it and obtains X from choosing $B_\tau = B_{\tilde{m},t}$; 2) the private arbitrage just turned public, in which case it is included in every block at period τ and proposer n obtains $\bar{R}$ from it; or 3) the private arbitrage has already turned public in period τ', $t < \tau' < \tau$ and is already exploited by the proposer n' who was chosen in that period, in which case proposer p_t gains nothing from it in period τ.

Thus, the expected value that n gets from not adding block $B_{\tilde{m},t}$ to the blockchain at time t is given by:

$$\begin{aligned} X \leq & \underline{\mathrm{R}} \\ & + \psi_n \left[\rho_d \bar{R} + (1-\rho_d)X\right] \\ & + (1-\psi_n)\psi_n \left[\rho_d \times 0 + (1-\rho_d)\left[\rho_d \bar{R} + (1-\rho_d)X\right]\right] \\ & + (1-\psi_n)^2 \psi_n \left[\rho_d \times 0 + (1-\rho_d)\rho_d \times 0 + (1-\rho_d)^2\left[\rho_d \bar{R} + (1-\rho_d)X\right]\right] + \cdots \end{aligned}$$

The first term, $\underline{\mathrm{R}}$, is what proposer n gets from adding any other block at time t. Each line j that follows corresponds to proposer n payoff if he is chosen as the proposer at time $\tau = t + j$ for the first time after time t. If in any period $\tau' = t + j'$, $j' \in \{1, \cdots j-1\}$, this private arbitrage transactions has become public, it is exploited already and proposer n gets nothing. If the arbitrage transaction is still private until period $j-1$, proposer n can obtain $\rho_d \bar{R} + (1-\rho_d)X$ from it at time $\tau = t + j$ assuming that no interim proposers have exploited it.[31] Either the arbitrage turns public with probability ρ_d, in which case proposer n gets all the surplus $\bar{R}$, or it remains private in which case proposer n gets R. As such, we have

$$\begin{aligned} X \leq & \underline{\mathrm{R}} + \psi_n \left[\rho_d \bar{R} + (1-\rho_d)X\right] \sum_{\tau=t+1}^{\infty} \left((1-\psi_n)(1-\rho_d)\right)^{\tau-t-1} \\ = & \underline{\mathrm{R}}_t + \psi_n \left[\rho_d \bar{R} + (1-\rho_d)X\right] \frac{1}{1-(1-\psi_n)(1-\rho_d)} \end{aligned}$$

[31] The assumption of no exploitation of the private arbitrage by interim proposers implies the inequality, which provides an upper bound on X as explained in the text.

Thus

$$X\left(1-\frac{\psi_n(1-\rho_d)}{1-(1-\psi_n)(1-\rho_d)}\right) \leq \underline{\mathrm{R}}+\frac{\psi_n\rho_d}{1-(1-\psi_n)(1-\rho_d)}\bar{R}$$
$$X\frac{\rho_d}{1-(1-\psi_n)(1-\rho_d)} \leq \underline{\mathrm{R}}+\frac{\psi_n\rho_d}{1-(1-\psi_n)(1-\rho_d)}\bar{R}$$

Thus, an upper bound for X is:

$$\bar{X}=\frac{1-(1-\psi_n)(1-\rho_d)}{\rho_d}\underline{\mathrm{R}}+\psi_n\bar{R}$$

For any pair (ψ_n, ρ_d) such that $\psi_n < 1$ and $\rho_d > 0$, let

$$\underline{\mathrm{R}}^{\max}=\frac{\rho_d(1-\psi_n)}{\rho_d(1-\psi_n)+\psi_n}\bar{R}.$$

If $\underline{R} < \underline{\mathrm{R}}^{\max}$, then $\bar{X} < \bar{R}$ and thus $X < \bar{R}$. Thus assumption $\underline{\mathrm{R}} \ll \bar{R}$ guarantees that $X < \bar{R}$. As such, proposer n is willing to leave profit $\bar{R}-\bar{X}$ for block builder $\tilde{m}$ to add $B_{\tilde{m},t}$ to the blockchain in period t, $\Upsilon_{B,t} \geq \bar{R}-\bar{X} > 0$.

Substituting the lower bound for $\Upsilon_{B,t}$ in Equations (2) and (3) leads to Equations (8) and (9), where the last inequality in each equation is satisfied if $\underline{R} < \underline{\mathrm{R}}^{\max}$. These inequalities directly imply $\theta_{B,t}^{\text{private}} > 0$ and $\theta_{P,t}^{\text{private}} < 1$. □

C Robustness Regression Tables

In this Appendix, we show the robustness regression Tables described in Section 6. Table 5 shows summary statistics for the full sample. Tables 6 through 7 show our regressions using the full sample. Tables 8 and 9 show the regressions using an alternative definition of private information.

	Mean	Std. Dev.	Min	5th	Median	95th	99th	Max	Skewness	Kurtosis
Rev_t	0.13	1.45	0.00	0.02	0.05	0.34	1.22	802.79	234.22	87570.02
$\Pi_{B,t}$	0.01	0.64	-56.13	-0.00	0.00	0.03	0.16	791.74	628.64	643411.84
$\Pi_{P,t}$	0.12	1.19	0.00	0.02	0.05	0.31	1.05	691.96	266.84	114216.68
$\theta_{B,t}$	0.02	0.77	-947.07	-0.10	0.01	0.28	0.57	1.00	-585.59	591582.47
$\theta_{P,t}$	0.98	0.77	0.00	0.72	0.99	1.10	1.37	948.07	585.59	591582.46
$\log Private_t$	0.07	0.17	0.00	0.00	0.03	0.26	0.75	6.69	8.98	133.29
$\log Public_t$	0.02	0.05	0.00	0.01	0.02	0.06	0.14	5.20	27.20	1346.59
Hack Dummy	0.07	0.25	0.00	0.00	0.00	1.00	1.00	1.00	3.46	12.94
Crisis Dummy	0.01	0.11	0.00	0.00	0.00	0.00	1.00	1.00	8.57	74.44
Observations	4132184									

Source: Dune Analytics and Mempool Guru Project

Table 5: Summary Statistics for the Full Sample

	(1) $\log Private_t$	(2) $\log Private_t$	(3) $\log Public_t$	(4) $\log Public_t$
Hack Dummy	0.0116***	0.0104***	-0.0008***	0.0001
	(0.0006)	(0.0007)	(0.0002)	(0.0002)
Crisis Dummy	0.1136***	0.1158***	0.0213***	0.0134***
	(0.0095)	(0.0099)	(0.0013)	(0.0012)
Constant	0.0700***		0.0237***	
	(0.0015)		(0.0008)	
Observations	4132184	3604011	4132184	3604011

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table shows the first stage estimation results with our full sample. Columns (1) and (2) show how $\log Private$ is affected by our instrumental variables, without and with builder, proposer, and builder $\times$ proposer fixed effects. Columns (3) and (4) show analogous results for $\log Public$. All standard errors are clustered at the builder $\times$ proposer level.

Table 6: First Stage Regression Results For The Full Sample

	(1) OLS No FE	(2) OLS FE	(3) IV No FE	(4) IV FE
$\log Private_t$	0.205***	0.184***	0.562***	1.180***
	(0.0163)	(0.0203)	(0.140)	(0.400)
$\log Public_t$	-0.0722***	-0.0275	-7.161***	-16.28***
	(0.0242)	(0.0278)	(1.469)	(4.062)
Constant	0.0106***		0.155***	
	(0.00256)		(0.0303)	
N	4132184	3604011	4132184	3604011
F Statistic			415.75	61.45
Robust F Statistic			111.961	10.649

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table shows our multivariate estimation results with our full sample. Columns (1) and (2) show OLS results, without and with builder, proposer and builder × proposer fixed effects. Columns (3) and (4) show 2SLS results, without and with builder, proposer, and builder × proposer fixed effects. All standard errors are clustered at the builder × proposer level. The instrumental variables are $Hacked_t$ and $Crisis_t$.

Table 7: OLS and Two-Stage Least Squares Results For The Full Sample

	(1) $logPrivate_t$	(2) $logPrivate_t$	(3) $logPublic_t$	(4) $logPublic_t$
Hack Dummy	0.0105***	0.0093***	0.0012***	0.0020***
	(0.0006)	(0.0007)	(0.0003)	(0.0003)
Crisis Dummy	0.0995***	0.1022***	0.0493***	0.0408***
	(0.0082)	(0.0084)	(0.0041)	(0.0046)
Constant	0.0671***		0.0280***	
	(0.0015)		(0.0009)	
Observations	4132184	3604011	4132184	3604011

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table shows the first stage estimation results for our different 2SLS specifications using an alternative definition of public and private information. Columns (1) and (2) show how $\log Private$ is affected by our instrumental variables, without and with builder, proposer, and builder × proposer fixed effects. Columns (3) and (4) show analogous results for $\log Public$. All standard errors are clustered at the builder × proposer level.

Table 8: First Stage Regression Results Using Alternative Definition of Private and Public Information

	(1) OLS No FE	(2) OLS FE	(3) IV No FE	(4) IV FE
$logPrivate_t$	0.201***	0.180***	1.785***	3.423***
	(0.0166)	(0.0204)	(0.283)	(0.796)
$logPublic_t$	0.0259***	0.0375***	-5.405***	-10.59***
	(0.00822)	(0.00838)	(1.102)	(2.593)
Constant	0.00905***		0.0555***	
	(0.00238)		(0.0168)	
N	4132184	3604011	4132184	3604011
F Statistic			204.59	40.22
Robust F Statistic			86.589	11.137

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: This table shows our multivariate estimation results when the builder profit share is the dependent variable, using an alternative definition of private and public information. Columns (1) and (2) show OLS results, without and with builder, proposer and builder × proposer fixed effects. Columns (3) and (4) show 2SLS results, without and with builder, proposer, and builder × proposer fixed effects, respectively. All standard errors are clustered at the builder × proposer level. The instrumental variables are $Hacked_t$ and $Crisis_t$.

Table 9: OLS and Two-Stage Least-Squares Results Using Alternative Definition of Private and Public Information.