STAFF REPORTS

NO. 1102
MAY 2024

Information and Market Power in DeFi Intermediation

Pablo Azar | Adrian Casillas | Maryam Farboodi

FEDERAL RESERVE BANK of NEW YORK

Information and Market Power in DeFi Intermediation
Pablo Azar, Adrian Casillas, and Maryam Farboodi
Federal Reserve Bank of New York Staff Reports, no. 1102
May 2024
https://doi.org/10.59576/sr.1102

Abstract

The decentralized nature of blockchain markets has given rise to a complex and highly heterogeneous market structure, gaining increasing importance as traditional and decentralized (DeFi) finance become more interconnected. This paper introduces the DeFi intermediation chain and provides theoretical and empirical evidence for private information as a key determinant of intermediation rents. We propose a repeated bargaining model that predicts that profit share of Ethereum market participants is positively correlated with their private information, and employ a novel instrumental variable approach to show that a 1 percent increase in the value of intermediaries' private information leads to a 1.4 percent increase in their profit share.

JEL classification: G23, D82, L14, L22, G14, D43
Key words: financial intermediation, oligopoly, blockchain, decentralized finance, cybersecurity

---

Azar: Federal Reserve Bank of New York (email: pablo.azar@ny.frb.org). Casillas: MIT Sloan (email: agcasil@mit.edu). Farboodi: MIT Sloan, CEPR, NBER (email: farboodi@mit.edu).

1 Introduction

Financial intermediation matches investors with entrepreneurs, allowing capital to be allocated efficiently. Along intermediation chains, intermediaries help this matching process by monitory, making markets, providing liquidity, improving risk sharing and managing inventory (Leland and Pyle, 1977; Diamond and Dybvig, 1983; Diamond, 1984; Allen and Gale, 1997; Boot and Thakor, 1997; Diamond and Rajan, 2001). As capital moves along the intermediation chain, each intermediary takes a share of the transaction's value as a fee for their service. Understanding how these fees are determined—and how they change depending on market conditions—is crucial for understanding financial markets. If the fees are too high, this creates a friction that prevents the entrepreneur from matching with investors. Conversely, if the fees are too low, the intermediaries may withdraw from the market.

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, 2019; Farboodi et al., 2019). 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.

In this paper, we use novel data from Decentralized Finance (DeFi) in the Ethereum blockchain to provide a more precise empirical picture of how these intermediation rents are driven by the intermediaries' private information. Ethereum is the largest platform for DeFi, and all transactions become public once they are appended to the blockchain. The large amount of DeFi protocols generates a constant stream of arbitrage opportunities, which are very quickly exploited by automated arbitrage bots deployed by arbitrageurs (entrepreneurs). Because the blockchain is public, we can see these transactions once they are added to the blockchain, as well as the associated transaction fees paid to intermediaries.

The study of DeFi intermediation in the Ethereum blockchain is not only interesting because of the transparency of the underlying data, but also because of the growing interconnectedness between DeFi and the traditional financial system. The approval of Bitcoin ETFs has already led to large asset managers holding tens of billions of dollars worth of Bitcoin. If a similar ETF is approved for Ethereum's native cryptocurrency Ether (ETH), it could result in these traditional financial institutions becoming the dominant Ethereum block proposers. This development would position them as crucial participants in the DeFi intermediation chain that we examine in this paper. As the lines between traditional finance and DeFi continue to blur, understanding the dynamics of intermediation in decentralized markets becomes increasingly relevant for both academics and practitioners seeking to navigate this evolving landscape.

1.1 The Origin of DeFi Intermediation: The Need for Privacy

The key friction in DeFi intermediation is that *arbitrageurs* who find arbitrage opportunities need to keep their transactions private before the transactions are appended to the blockchain. That is, they need a way to have their transactions approved without first broadcasting them to the entire network and facing the risk of being frontrun or having the arbitrage opportunity stolen by an adversary. The risk of having a transaction stolen is not only real, but ubiquitous due to the commodification of AI-driven frontrunning bots (Robinson and Konstantopoulos, 2020).

This need for privacy 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 ETH—to select a single block to be appended to the blockchain in a given round.[1] Block builders compete with each other to create the most valuable block they can and then submit a (block, bid) pair to the block proposer, who chooses one winning block.[2] In this intermediation chain, 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.

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

[1]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.

[2]This is a simplification which preserves the economic content. In practice, the builder submits a bid (denominated in ETH), as well as a cryptographic commitment to the block. To prevent the proposer from front-running transactions, the block is only revealed once the proposer has accepted the bid. We provide more details in the Appendix.

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.

This intermediation chain connects arbitrageurs and depositors through multiple layers that preserve the privacy of the arbitrageurs' transactions. If an arbitrageur sends their transaction directly to a builder, only the arbitrageur and the builder will know of that transaction until it gets appended to the blockchain. 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. 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 167 known builders, 3 capture more than 50% of all the builder revenue and blocks proposed. Similarly, even though there are 156,150 block proposers, five large staking pools capture more than 50% of all the proposer revenue and blocks proposed.

The value of public and private information When arbitrageurs want to submit a transaction, they can do it in one of two ways: broadcast it to the entire network of builders, or submit it privately to a single builder. The first approach is valuable if a arbitrageur wants their transaction to be appended to the blockchain quickly (i.e. in the next few blocks). However, publicly broadcasting even a single leg of a valuable sequence of trades would allow an adversary to swoop in and front-run that transaction, therefore reducing the arbitrageur's profit. For this reason, many transactions, including those related to valuable arbitrage opportunities, are sent directly to specific block builders.

All intermediaries are aware of transactions broadcast to the network, and store them in a special data structure called the *mempool*. However, a given block builder will also know all the transactions sent directly to them, which include the most valuable arbitrage opportunities. This allows us to assign a value to private (public) information—the sum of all fees in private (public) transactions, together with any direct payments to the block builder. Even though these transactions are private before they are added to the blockchain, once they are added we can observe both the transaction and the fee. Since we also observe

Number of Builders	**Number of Proposers**	
167	221,534	
Builder Address	**Share of Total Builder Revenue**	**Share of Blocks**
beaverbuild.org	42.32%	25.05%
Titan Builder	14.63%	15.41%
builder0x69	13.69%	14.94%
Proposer Entity	**Share of Total Proposer Revenue**	**Share of Blocks**
Lido	31.38%	31.75%
Coinbase	10.31%	10.20%
Kraken	5.53%	5.56%
Binance	5.20%	5.71%
Stakefish	4.61%	4.90%

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.

all the transactions broadcast to the network,[3], we can distinguish the value of public and private transactions.

The level of transparency and granularity in the data allows us to empirically test the predictions of a bargaining model with private information. In a nutshell, we model the determination of the profit shares of proposers and block builders in the context of a repeated bargaining model with asymmetric exogenous bargaining powers but endogenous outside options. In each period, the outside options of the block builder and proposer are determined by their future opportunities if they do not agree on adding the block create by the block builder to the blockchain in the current period.

We show that access to valuable *private* arbitrage transactions by builders—i.e., those arbitrage transactions that are privately submitted to them by an arbitrageur— improves their outside option as they are effectively the gatekeeper for the private arbitrage opportunity. As such, private arbitrage transactions increase the profit share of the block builders while

[3]We are grateful to Professor Fan Zhang and the Secure Decentralized Systems Lab's Mempool Guru project for providing us with data about which transactions were and were not broadcast to the network prior to being incorporated into a block.

decreasing that of the proposers. 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.

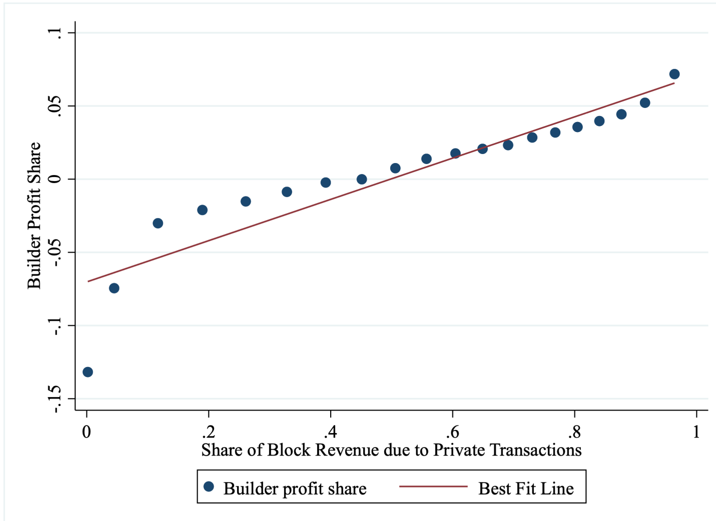
The main prediction of the model is that the block builder's profit share from a block is increasing in his private information. Figure 1 illustrates how this prediction is supported by our data, with a strong positive relationship between the share of a block's revenue that's due to private transactions, and the profit share of the builder.

However, a simple OLS regression to support this finding would introduce biases both from simultaneity and omitted variables. Since the builder decides both the bid made to the proposer—essentially determining their share of the revenue, as well as the transactions that are included in the block, the private revenue share and the builder's profit share are determined simultaneously. Furthermore, there may be other properties of the block, such as the total revenue, which affect the builder's profit share.

To address biases from simultaneity and omitted variables, we introduce a novel instrumental variable approach that uses hacks of crypto institutions as an instrument for the value of private information in a block. When an exchange or decentralized finance protocol's assets get stolen through a hack, the attacker has incentives to be as inconspicuous as possible: they do not want to reveal their IP address, the existence of the hack, or the source of the funds. Furthermore, they may want to sell their stolen assets quickly, paying large transaction fees to dispose of the assets. Therefore, hacks are likely to create valuable private transactions, which increase the value of private information in a block, but not necessarily the value of public information.

To instrument for the total revenue of the block, we use a dummy variable that is one during days of major crises in the crypto markets, and zero otherwise. We identify two such crises: the FTX bankruptcy unfolding from November 8, 2022 to November 12, 2022, and the SVB run occurring between March 9th, 2023 and March 12th, 2023. These crypto crises led to a large number of blockchain transactions, generating a large amount of revenue. As we show in Section 3, both crypto hacks and crypto crises increase the value of the entire block as well as the value of private information. However, hacks of crypto institutions increase the value of private information by a much larger amount than they increase revenue, while crypto crises increase revenue by a larger amount than the value of private information, allowing the two instrumental variables to span our set of two explanatory variables.

Using this instrumental variable approach, we show in Section 3 that a 1% increase in the value of private information leads to a 1.4% increase in the builder profit share. After



Source: Dune Analytics and Mempool Guru Project

Figure 1: 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.

1.2 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). There is also a wealth of papers that empirically document the prevalence of intermediation rents in financial markets (Green et al., 2007; Di Maggio et al., 2017; Hollifield et al., 2017; Li and Schürhoff, 2019; Farboodi et al., 2019). We contribute to this literature by first distinguishing the role of financial intermediaries in the most recent segment of the financial sector, the cryptocurrency market. Second, we identify private information as a source of intermediation rents in this market.

Furthermore, 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. Abadi and Brunnermeier (2018) and Biais et al. (2019) expand on consensus mechanisms, focusing on proof-of-work.

Economics of Blockchain Technology 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 (2020) 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.

A number of papers consider the degree of decentralization in cryptocurrency markets. Huberman et al. (2021) argue that the decentralized nature of Bitcoin network prevents monopoly pricing. Alternatively, Ferreira et al. (2023) argue that the joint behavior of miners, mining pools, and firms producing specialized mining equipment can lead to few large firm capturing the governance of proof-of-work blockchains. Makarov and Schoar (2021) provide a detailed analysis of the Bitcoin network and show that the Bitcoin ecosystem is dominated by large and concentrated players. On the other hand, Cong et al. (2021c) argue that, since individual miners can join multiple mining pools, the system can still be decentralized. The authors apply the same argument to proof-of-stake systems. Capponi et al. (2023) empirically study the effect of private information on aggregate welfare in a proof-of-work environment and argue that private information enhances welfare.

In contrast with these papers, we focus on 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.

MEV in the Ethereum Blockchain The earliest paper to formalize MEV was Daian et al. (2019), which began hinting at the threats to blockchain security and market efficiency it posed. Gupta et al. (2023) show that integrated searcher-builders are more likely to capitalize private arbitrage transactions at the head of a block in times of high centralized exchange volatility, showing that the first transactions in a block are all but reserved for arbitrages between centralized exchanges and decentralized exchanges. Heimbach et al. (2024) expand on the heuristics proposed in Gupta et al. (2023) to classify arbitrages as either *atomic*, with all transactions occurring on the blockchain, and *non-atomic*, with arbitrage transactions exploiting mispricing between centralized and decentralized exchanges. They show that more than a fourth of the volume on the 5 most popular decentralized exchanges is due to these non-atomic arbitrage transactions. Since centralized exchanges' ledgers are not on the blockchain, these kind of transactions rely on private information held by arbitrageurs. They further show that only 11 such arbitrageurs are responsible for 80% of these non-atomic arbitrage transactions.

In our main analysis, 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. This is a very 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 our robustness results, we use the more complex heuristic of Heimbach et al. (2024) to show that our results also hold under their definition of private transactions.

There is also a body of theoretical literature which models the relationship between public and private MEV and market volatility. Milionis et al. (2023b) model MEV as the adverse selection to an automated market maker employed by DEXs and derives a closed form solution to the theoretical amount of available MEV. Moreover, this solution shows that the volume of private arbitrages is a direct function of token price volatility. Milionis et al. (2023a) introduce trading fees into this framework and shows that fees reduces the amount of available arbitrage profits based on the amount of time it takes to find a profitable arbitrage.

Crypto Crises and Arbitrage Liu et al. (2023) show that the Terra crash led to large arbitrage oppportunities. While the Terra crash occurred before Ethereum's switch to proof-of-stake consensus, we use a similar observation to construct our instrumental variables. We show that crypto crises—such as the FTX bankruptcy, the SVB collapse, and cyber-attacks on crypto institutions and protocols—can be an exogenous shock that creates private arbitrage opportunities. 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 provides details of the block level data from Ethereum blockchain that we use for our empirical analysis. Section 3 describes the instrumental variable approach and presents the main empirical results. Section 4 proposes a model to provide the economic mechanism that underlies the empirical findings. Section 5 provides a number of robustness exercises. Section 6 concludes.

2 Data

We use Dune Analytics[4] 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., 2023) 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 January 31, 2024—75.9% of the blocks satisfy both of these conditions, and are considered MEV blocks.[5]

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,[6] minus the payment to the proposer at the end of the block. The net profit for the proposer, $\Pi_{P,t}$, is

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

[5] The total number of blocks in our sample is 3,588,414, and the number of blocks satisfying the MEV conditions is 2,723,653. Of these, we remove 138 blocks that have zero revenue, to end up with 2,723,585 positive-revenue MEV blocks.

[6] 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

	Mean	Std. Dev.	Min	5th	Median	95th	Max	Skewness	Kurtosis
Rev_t	0.14	1.52	0.00	0.02	0.06	0.37	691.96	225.07	76506.49
$\Pi_{B,t}$	0.01	0.40	-0.30	-0.00	0.00	0.02	386.27	474.65	366718.03
$\Pi_{P,t}$	0.14	1.39	0.00	0.02	0.05	0.35	691.96	254.00	95968.37
$\theta_{B,t}$	0.03	0.08	-0.10	-0.02	0.01	0.16	1.00	4.66	32.36
$\theta_{P,t}$	0.97	0.08	0.00	0.84	0.99	1.02	1.10	-4.66	32.36
$\log Private_t$	0.07	0.17	0.00	0.00	0.03	0.27	6.54	8.44	119.52
$\log Public_t$	0.03	0.05	0.00	0.01	0.02	0.07	5.20	24.48	1115.99
Hack Dummy	0.07	0.26	0.00	0.00	0.00	1.00	1.00	3.23	11.46
Crisis Dummy	0.02	0.14	0.00	0.00	0.00	0.00	1.00	6.95	49.32
Observations	2627618								

Source: Dune Analytics and Mempool Guru Project

Table 2: Summary Statistics

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 90% 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.[7] Our main analysis considers only blocks where the builder profit share is greater than -10%, which represent 96.5% of the blocks in our sample.[8]

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.

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

[8]In Section 5, we show that our results also hold for the full sample, as well as a further restricted sample that contains only blocks where the builder makes non-negative profits.

3 Market Power in the 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 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. Second, 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 revenue Rev_t of the block. Finally, 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):

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

Including the revenue term 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.

Finally, to address simultaneity, we use an instrumental variables approach with two instrumental variables. Both IVs 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.[9] The second dummy, $Crisis_t$ is equal to 1 if block t is appended to the

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

blockchain during either the FTX or SVB crises.[10] The first-stage specification is given by

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

Table 3 shows how these instrumental variables affect the value of private information and revenue. We can see that private information increases more—-relative to revenue—during hacks than during crises. Furthermore, the instruments are exogenous, since both hacks and crypto crises are unexpected and not caused by the bargaining process between builders and proposers.

	(1) $\log Private_t$	(2) $\log Private_t$	(3) $\log Rev_t$	(4) $\log Rev_t$
Hack Dummy	0.0072***	0.0059***	0.0039***	0.0043***
	(0.0007)	(0.0006)	(0.0007)	(0.0007)
Crisis Dummy	0.1208***	0.1236***	0.1300***	0.1289***
	(0.0092)	(0.0099)	(0.0093)	(0.0100)
Constant	0.0715***		0.0985***	
	(0.0022)		(0.0020)	
Observations	2607730	2277574	2607730	2277574

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 these instruments, with Column (1) having no fixed effects, and Column (2) having builder, proposer and builder × proposer fixed effects. Columns (3) and (4) show analogous results for $\log Public$, the log of the value of Public arbitrages in a given block. Columns (5) and (6) show analogous results using $\log Rev$, the log revenue of the block. All standard errors are clustered at the builder × proposer level.

Table 3: First Stage Regression Results

Results Table 4 shows the results of OLS and 2SLS regressions where $\theta_{B,t}$ is the dependent variable, and $\log Private_t$, $\log Rev_t$ are the independent variables. Columns (1) and (2) show the OLS results without and with builder × proposer fixed effects. Columns (3) and (4) show the results from the instrumental variable regressions using the Hacks and Crises dummies

[10] The FTX crises occured 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.

	Builder Profit Share $\theta_{B,t}$			
	(1) OLS No FE	(2) OLS FE	(3) IV No FE	(4) IV FE
$\log Private_t$	0.143***	0.111***	1.367***	1.484***
	(0.0138)	(0.0149)	(0.177)	(0.235)
$\log Rev_t$	-0.0713***	-0.0511***	-1.240***	-1.360***
	(0.00906)	(0.00950)	(0.175)	(0.223)
Constant	0.0237***		0.0511***	
	(0.00244)		(0.00535)	
N	2607730	2277574	2607730	2277574
F Statistic			583.25	127.22
Robust F Statistic			220.937	26.100

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 fixed effects, respectively. Columns (3) and (4) show 2SLS results, without and with 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 4: OLS and Two-Stage Least Squares Results

as instruments—again without and with builder × proposer fixed effects.[11] 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 1.4%. 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.

4 Model: Information-Driven Market Power

In this section we provide a stylized model to illustrate the determination of profit shares of the proposers and block builders in the DeFi interemediation chain. In order to clarify the mechanism, we will abstract away from the rest of the DeFi intermediation chain and

[11] 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.

restrict attention to the bargaining game between the proposers and block builders in the process of building the blockchain.

We model the construction of the blockchain as an infinitely repeated game between two types of agents— proposers and block builders. Time is indexed by $t = 0, 1, \ldots$.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 coins 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.

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 $R_{m,t}$.[12] We denote the set of all the blocks created at time t by $\mathcal{B}_t = \{B_{m,t}\}_{m \in \{1,...,M\}}$.

At each time t, blocks are made up of three type of transactions: Normal 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 normal periods with no arbitrage opportunities. An arbitrage opportunity arises at small Poisson rate $\rho \ll 1$ per period, and disappears at Poisson rate $\rho_d < 1$ per period, where $\rho_d \gg \rho$. 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 identified by many different arbitrageurs and thus are known to all block builders who all include it in their respective block. As such, a public arbitrage transaction in period t is included in every $B_{m,t} \in \mathcal{B}_t$.

Let "public blocks" denote all the blocks in a normal period or a in a period with public arbitrages. For simplicity, assume all public blocks in period t have the same value $\widehat{R_t}$. The value $\widehat{R_t}$ varies across periods— low in normal periods, $\underline{\mathrm{R}}_t$, and high in periods with public arbitrages, $\bar{R}_t$. Thus, $\widehat{R}_t \in \{\underline{\mathrm{R}}_t, \bar{R}_t\}$.[13] Table 2 documents that the distribution of block revenue is strongly skewed to the right. Motivated by this empirical evidence, we assume $\bar{R}_t \gg \underline{\mathrm{R}}_t$.

Alternatively, a private arbitrage that arrives in period t is identified by a single arbitrageur and picked up by a single builder, $\tilde{m}$, who incorporates it in his block $B_{\tilde{m},t}$.[14] In

[12]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 $c = 0$.

[13]$\underline{\mathrm{R}}_t$ and $\bar{R}_t$ can be random variables themselves, with different supports.

[14]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.

As long as few arbitrageurs identify each arbitrage opportunity and few block builders incorporate each arbitrage opportunity in their block, the same argument goes through.

that sense, the private arbitrage opportunity is the "private information" of builder $\tilde{m}$. We call $B_{\tilde{m},t}$ the "private block" at time t. 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 $R_{m,t} = \underline{\mathrm{R}}_t$. For simplicity, assume the value of a private arbitrage block is governed by the same random variable as a public arbitrage block, and $R_{\tilde{m},t} = \bar{R}_t$.

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.[15]

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 2, 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 $\delta \to 0$, in which case $\Pi_{B,t}$ and $\Pi_{P,t}$ simplify to

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

[15]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 $(\Upsilon_{P,t}, \Upsilon_{B,t})$.

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 3.

Proposition 1 (Information-Driven Market Power). *Existence of private arbitrage transactions in a given block increase the profit share of the block builder and decreases that of the proposer. Alternatively, higher total block revenue has the opposite impact on the profit share of the block builder and the proposer, controlling for the value of private arbitrages.*

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

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

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

Next, consider a period t when there is a builder with a private block, $B_{\tilde{m},t}$ and let n denote the proposer who is chosen in period t, $p_t = n$. Furthermore, let X denote the maximum profit that proposer n can obtain from existence of the private block $B_{\tilde{m},t}$ in the future if he does not choose $B_t = B_{\tilde{m},t}$. X is also the maximum that proposer n can extract from the total revenue of the private arbitrage block $B_{\tilde{m},t}$ at time t if he decides to choose $B_t = B_{\tilde{m},t}$.

If proposer n does not choose the private block $B_{\tilde{m},t}$ in period t to add to the blockchain, it remains in the pool of available blocks for future proposers, until the private arbitrage opportunity disappears (at rate ρ_d). Meanwhile, as this private arbitrage transaction is not included in any other bock, its corresponding value remains unexploited. Proposer n has an i.i.d. probability $\psi_n = \frac{w_n}{\sum_{i=1}^N w_i}$ to be chosen each period after t. In each future period $\tau > t$ that $n = p_\tau$, if block $B_{\tilde{m},t}$ is still available, n can obtain at most $\bar{R}_t$ from choosing $B_\tau = B_{\tilde{m},t}$. Finally, when n does not choose $B_t = B_{\tilde{m},t}$, he chooses a public block with normal transactions only in period t, with total value $\underline{\mathrm{R}}_t$. Thus, Equation (7) implies that he receives $\underline{\mathrm{R}}_t$ today. As such, an upper bound for X is given by

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

Since $\underline{\mathrm{R}}_t \ll \bar{R}_t$, $X < \bar{R}_t$. In other words, proposer n is willing to leave profit $\frac{\rho_d}{\psi_n+\rho_d(1-\psi_n)}\bar{R}_t - \underline{\mathrm{R}}_t$ 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, for a private block, using Equations (2) and (3) we have

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

$$\Pi^{\text{private}}_{P,t} \leq \left(1 - \frac{\rho_d}{\psi_n + \rho_d(1-\psi_n)}\right)\bar{R}_t + \underline{\mathrm{R}}_t \;\Rightarrow\; \theta^{\text{private}}_{P,t} < 1 \tag{9}$$

Comparing Equation (6) with (8) and (7) with (9) clearly illustrate 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 shorter lived private arbitrage opportunities lead to a higher profit share for the block builder. This is intuitive as it corresponds to a more limited availability of the private arbitrage opportunity, which in turn makes the private information of the block builder more valuable and improves his bargaining position.[16]

[16]It is worth mentioning that we have abstracted away from any relationship building between proposers

5 Robustness

In this section we show the robustness of our empirical results in two dimension.

In Section 5.1 we change the sample of blocks in two different ways. First, we extend the sample to include the blocks in which the block builder subsidizes the proposer at very high rates. Second, we limit the sample only to blocks in which block builders make weakly positive profits.

In Section 5.2 we use a more involved algorithm to measure builder's private information. In particular, we employ the heuristic introduced by Heimbach et al. (2024). This heuristic classifies a transaction as a private arbitrage only if it is based on information that is not in the blockchain.

Appendix B presents the results of these robustness estimation exercises. Importantly, in every exercise all the instrumental variable regression coefficients have the same sign as the baseline estimation and remain statistically significant.

5.1 Blocks With And Without Subsidies

In our main analysis, we removed blocks where the builder profit share was less than -10%, consisting of about 3.5% of the sample, to prevent outliers from skewing the results. In this section, we show our results hold 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. 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.

Table 5 in Appendix B shows that there exist blocks where the builder gives a very large subsidy to the proposer, with the largest subsidy being 56.13 ETH. This heavily skews the builder profit share $\theta_{P,t}$, with the share being highly negative for blocks with large subsidies. We show in this subsection that our results apply even when we restrict ourselves to blocks with a builder profit share $\theta_{B,t} \geq -0.1$, and when we restrict ourselves only to blocks with non-negative builder profit share $\theta_{B,t} \geq 0$.

---

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.

The results for these specifications are shown in Tables 6 through 9 in Appendix B. Tables 6 and 7 show the first and second stage estimation results for the full sample. Alternatively, Tables 8 and 9 show the results conditioned on $\theta_{B,t} \geq 0$. While the effect of private information is slightly attenuated from 5.5 to around 1.5, all coefficients are still significant and the instruments are still strong.

5.2 Alternative Definition of Private Information

In our main analysis, we use the simplest possible 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. 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.

As an alternative measure of private information, we use the heuristic introduced by Heimbach et al. (2024), who consider only arbitrage transactions which are based on information that is not in the blockchain. Under this heuristic, a group of transactions is private if and only if one of the transactions involves a direct transfer to the builder, the transactions do not appear in the public mempool, one of the transactions is a DEX swap as classified by the ZeroMEV API,[17], and the swap involves a token that is traded in a centralized exchange.

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.

[17] A DEX swap is an exchange of one token for another in a decentralized exchange. We use the ZeroMEV API for a classification of swap transactions[18]

Tables 10 and 11 show the first-stage and 2SLS regression results using the alternative definition of private information. We can see that the instruments are still strong, the results are significant, and there's a positive effect for the value of private information. The only qualitative difference is that all of our other regression results have a negative coefficient for $\log Rev_t$ in all columns, while in Table 11 the OLS coefficient for $\log Rev_t$ is positive, and it becomes negative only when using instrumental variables. The intuitive explanation for this is that our definition of private arbitrage is much more restrictive, and many transactions which would have been classified as private in our main analysis are no longer classified as such. 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 $\log Private_t$.

6 Conclusion

We examine 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 need for privacy—driven by the commodification of AI frontrunners—leads to the emergence of block builders as intermediaries, who capture a share of the block revenue by incorporating private transactions into aggregate blocks.

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 1.4% increase in the block builder profit share. We propose a repeated bargaining model to provide an economics 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 more 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," Technical Report, National Bureau of Economic Research 2018.

Allen, Franklin, "The market for information and the origin of financial intermediation," *Journal of financial intermediation*, 1990, *1* (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," Technical Report 2016.

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

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

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 Ye Wang, "Maximal Extractable Value and Allocative Inefficiencies in Public Blockchains," *Available at SSRN 3997796*, 2023.

Cong, Lin William, 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.

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

Daian, Philip, Steven Goldfeder, Tyler Kell, Yunqi Li, Xueyuan Zhao, Iddo Bentov, Lorenz Breidenbach, and Ari Juels, "Flash boys 2.0: frontrunning, transaction reordering, and consensus instability in decentralized exchanges," 2019.

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.

Farboodi, Maryam, Gregor Jarosch, and Guido Menzio, "Intermediation as rent extraction," Technical Report, National Bureau of Economic Research 2019.

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.

Gupta, Tivas, Mallesh M Pai, and Max Resnick, "The centralizing effects of private order flow on proposer-builder separation," 2023.

Heimbach, Lioba, Vabuk Pahari, and Eric Schertenleib, "Non-atomic arbitrage in decentralized finance," 2024.

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.

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*, 2019, *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," Technical Report, National Bureau of Economic Research 2021.

Milionis, Jason, Ciamac C. Moallemi, and Tim Roughgarden, "Automated market making and arbitrage profits in the presence of fees," 2023.

_ , _ , _ , and Anthony Lee Zhang, "Automated market making and loss-versus-rebalancing," 2023.

Nakamoto, Satoshi, "Bitcoin: A peer-to-peer electronic cash system," *Decentralized business review*, 2008, p. 21260.

Pagnotta, Emiliano and Andrea Buraschi, "An equilibrium valuation of bitcoin and decentralized network assets," Technical Report 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: Journal of the Econometric Society*, 1982, pp. 97–109.

Sockin, Michael and Wei Xiong, "A model of cryptocurrencies," Technical Report, National Bureau of Economic Research 2020.

Yang, Sen, Fan Zhang, Ken Huang, Xi Chen, Youwei Yang, and Feng Zhu, "SoK: MEV Countermeasures: Theory and Practice," 2023.

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.[19]

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)).[20] 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,[21] leading to high transaction fees and very volatile congestion charges.

[19]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).

[20]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.

[21]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.[22] 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.

[22] 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.[23] 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 frontruns 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.

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

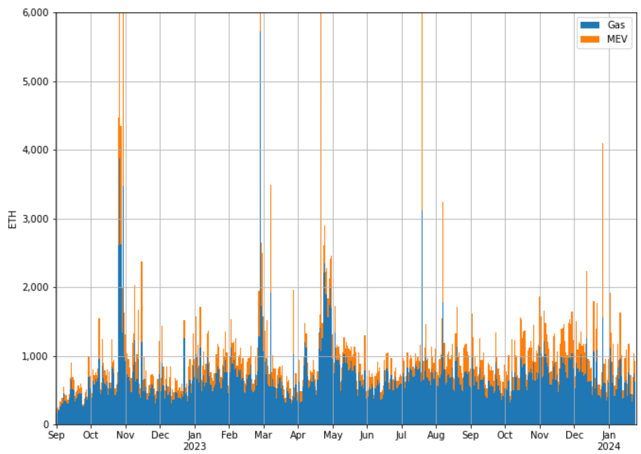
Since the probability of being a block proposer is proportion the amount of ETH staked, we can interpret this as a yield

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

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 2 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

[23] 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 2: 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 frontrunning. 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 frontrunning, 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 frontrunning, 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—[24] the block proposer at this time learns nothing which would allow her to frontrun 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.[25]

[24] 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^*.

[25] 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

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 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. [26] 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 Robustness Regression Tables

In this Appendix, we show the robustness regression Tables described in Section 5. Table 5 shows summary statistics for the full sample. Tables 6 through 9 show our regressions using different subsamples of our dataset. Tables 10 and 11 show the regressions using an alternative definition of private information.

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 frontrun them, there is an informational advantage for blockbuilders to operate their own relaying software.

[26] 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.

	Mean	Std. Dev.	Min	5th	Median	95th	Max	Skewness	Kurtosis
Rev_t	0.14	1.49	0.00	0.02	0.05	0.36	691.96	229.03	79243.77
$\Pi_{B,t}$	0.01	0.40	-56.13	-0.00	0.00	0.02	386.27	472.18	369899.17
$\Pi_{P,t}$	0.13	1.36	0.00	0.02	0.05	0.35	691.96	257.98	99142.42
$\theta_{B,t}$	0.01	0.90	-947.07	-0.06	0.01	0.15	1.00	-555.01	495297.49
$\theta_{P,t}$	0.99	0.90	0.00	0.85	0.99	1.06	948.07	555.01	495297.57
$\log Private_t$	0.07	0.17	0.00	0.00	0.03	0.27	6.54	8.55	122.62
$\log Public_t$	0.03	0.05	0.00	0.01	0.02	0.07	5.20	24.77	1146.61
Hack Dummy	0.07	0.26	0.00	0.00	0.00	1.00	1.00	3.24	11.50
Crisis Dummy	0.02	0.14	0.00	0.00	0.00	0.00	1.00	6.88	48.39
Observations	2723585								

Source: Dune Analytics and Mempool Guru Project

Table 5: Summary Statistics for the Full Sample

	(1) $\log Private_t$	(2) $\log Private_t$	(3) $\log Rev_t$	(4) $\log Rev_t$
Hack Dummy	0.0076***	0.0063***	0.0045***	0.0048***
	(0.0007)	(0.0006)	(0.0007)	(0.0007)
Crisis Dummy	0.1141***	0.1162***	0.1228***	0.1212***
	(0.0094)	(0.0101)	(0.0094)	(0.0101)
Constant	0.0699***		0.0968***	
	(0.0022)		(0.0019)	
Observations	2679416	2341828	2679416	2341828

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 these instruments, with Column (1) having no fixed effects, and Column (2) having builder, proposer and builder $\times$ proposer fixed effects. Columns (3) and (4) show analogous results for $\log Public$, the log of the value of Public arbitrages in a given block. Columns (5) and (6) show analogous results using $\log Rev$, the log revenue of the block. All standard errors are clustered at the builder $\times$ proposer level.

Table 6: Full Sample First Stage Regression Results

	(1) OLS No FE	(2) OLS FE	(3) IV No FE	(4) IV FE
$\log Private_t$	0.154***	0.138***	3.004***	5.512***
	(0.0245)	(0.0358)	(0.771)	(1.759)
$\log Rev_t$	-0.0449***	-0.0401**	-3.396***	-5.886***
	(0.0137)	(0.0193)	(0.928)	(1.908)
Constant	0.00176		0.128***	
	(0.00333)		(0.0359)	
N	2679416	2341828	2679416	2341828
F Statistic			592.33	128.77
Robust F Statistic			230.179	24.570

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 fixed effects, respectively. Columns (3) and (4) show 2SLS results, without and with 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 7: Full Sample OLS and Two-Stage Least Squares Results

Builder Profit Share $\theta_{B,t}$				
	(1) $\log Private_t$	(2) $\log Private_t$	(3) $\log Rev_t$	(4) $\log Rev_t$
Hack Dummy	0.0086***	0.0069***	0.0052***	0.0052***
	(0.0008)	(0.0007)	(0.0009)	(0.0008)
Crisis Dummy	0.1406***	0.1440***	0.1493***	0.1494***
	(0.0161)	(0.0182)	(0.0163)	(0.0184)
Constant	0.0801***		0.1081***	
	(0.0030)		(0.0028)	
Observations	2134770	1862430	2134770	1862430

Standard errors in parentheses

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

Note: This table shows the first stage estimation results when $\theta_{B,t} \geq 0$. Columns (1) and (2) show how $\log Private$ is affected by these instruments, with Column (1) having no fixed effects, and Column (2) having builder, proposer and builder × proposer fixed effects. Columns (3) and (4) show analogous results for $\log Public$, the log of the value of Public arbitrages in a given block. Columns (5) and (6) show analogous results using $\log Rev$, the log revenue of the block. All standard errors are clustered at the builder × proposer level.

Table 8: First Stage Regression Results when $\theta_{B,t} \geq 0$

	Builder Profit Share $\theta_{B,t}$			
	(1) OLS No FE	(2) OLS FE	(3) IV No FE	(4) IV FE
$\log Private_t$	0.160***	0.128***	1.586***	1.596***
	(0.0142)	(0.0163)	(0.164)	(0.215)
$\log Rev_t$	-0.101***	-0.0786***	-1.455***	-1.468***
	(0.00900)	(0.00966)	(0.162)	(0.198)
Constant	0.0351***		0.0670***	
	(0.00277)		(0.00461)	
N	2134770	1862430	2134770	1862430
F Statistic			441.75	113.99
Robust F Statistic			164.056	34.042

Standard errors in parentheses

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

Note: This table shows our multivariate estimation results when $\theta_{B,t} \geq 0$. Columns (1) and (2) show OLS results, without and with builder × proposer fixed effects, respectively. Columns (3) and (4) show 2SLS results, without and with 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 where $\theta_{B,t} \geq 0$

	(1) $\log Private_t$	(2) $\log Private_t$	(3) $\log Rev_t$	(4) $\log Rev_t$
Hack Dummy	0.0055***	0.0051***	0.0035***	0.0046***
	(0.0004)	(0.0005)	(0.0007)	(0.0008)
Crisis Dummy	0.0582***	0.0580***	0.1486***	0.1451***
	(0.0059)	(0.0069)	(0.0108)	(0.0119)
Constant	0.0174***		0.1003***	
	(0.0013)		(0.0020)	
Observations	2242059	1933070	2242059	1933070

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 these instruments, with Column (1) having no fixed effects, and Column (2) having builder, proposer and builder × proposer fixed effects. Columns (3) and (4) show analogous results for $\log Public$, the log of the value of Public arbitrages in a given block. Columns (5) and (6) show analogous results using $\log Rev$, the log revenue of the block. All standard errors are clustered at the builder × proposer level.

Table 10: First Stage Regression Robustness Results Using Private Arbitrages

	Builder Profit Share $\theta_{B,t}$			
	(1) OLS No FE	(2) OLS FE	(3) IV No FE	(4) IV FE
$\log Private_t$	0.106***	0.0847***	1.283***	0.964***
	(0.00937)	(0.0110)	(0.186)	(0.226)
$\log Rev_t$	0.0210***	0.0222***	-0.464***	-0.317***
	(0.00317)	(0.00430)	(0.0862)	(0.102)
Constant	0.0234***		0.0513***	
	(0.00251)		(0.00660)	
N	2242059	1933070	2242059	1933070
F Statistic			177.95	93.52
Robust F Statistic			63.502	25.051

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 fixed effects, respectively. Columns (3) and (4) show 2SLS results, without and with 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 11: OLS and Two-Stage Least Squares Using Private Arbitrages