July 2024

Price Discovery on Decentralized Exchanges

Agostino Capponi
Columbia University

Ruizhe Jia
Columbia University

Shihao Yu
Singapore Management University

SMU
SINGAPORE MANAGEMENT UNIVERSITY

SIM KEE BOON INSTITUTE FOR FINANCIAL ECONOMICS

LEE KONG CHIAN SCHOOL OF BUSINESS

Price Discovery on Decentralized Exchanges[1]

Agostino Capponi Ruizhe Jia Shihao Yu

Current version: July 1, 2024

[1]Agostino Capponi, Columbia University, Department of Industrial Engineering and Operations Research, ac3827@columbia.edu. Ruizhe Jia, Columbia University, Department of Industrial Engineering and Operations Research, rj2536@columbia.edu. Shihao Yu, Columbia University, Department of Industrial Engineering and Operations Research, sy3119@columbia.edu. We thank Darrel Duffie, Joel Hasbrouck, Sean Foley, Tālis Putniņš, Bart Zhou Yueshen, Konstantin Sokolov, Sabrina Buti, Carole Comerton-Forde, Andreas Barbon (discussant), Shiyang Huang (discussant), Marius Zoican (discussant), and seminar participants at the Microstructure Exchange (Asia-pacific edition), Hong Kong Baptist University, Luohan Academy, Oxford University DeFi Working Group, Chaincode, The University of Memphis, SGF Conference 2023, European Summer Symposium in Finance Markets 2023, CICF 2023, EFA 2023, for helpful comments. An earlier version of the paper circulated under the title "The Information Content of Blockchain Fees".

SMU Singapore Management University | Sim Kee Boon Institute for Financial Economics

Price Discovery on Decentralized Exchanges

Abstract

Decentralized exchanges (DEXs) allow traders to express their willingness to pay for quick execution through a public priority fee bidding mechanism. This influences the trading strategy of informed traders and creates a distinct price discovery process on DEXs compared to centralized exchanges. We present empirical evidence that high-fee DEX trades contain more private information. Informed traders bid high fees not only to avoid execution risk from blockchain congestion, but also to compete for execution priority. Using a dataset of Ethereum mempool orders, we demonstrate that informed traders employ a "jump bidding" strategy, placing high initial bids to deter potential competitors.

1 Introduction

Decentralized exchanges (DEXs) are trading venues built on public blockchains. As of May 2024, the average trading volume on DEXs amounts to $88.66 billion, which is about 10% of the total traded volume in the crypto market.[1] The trading mechanism on DEXs has distinctive features compared to centralized exchanges. Most DEXs execute orders in discrete time, with the execution sequence of pending transactions determined by the priority fee that traders bid for their orders. The higher the fee, the greater the likelihood that the order will be executed ahead of other pending orders. Additionally, on DEXs, orders submitted through the peer-to-peer blockchain network are publicly visible, disclosing details such as the direction and size of the order, as well as the bid for priority fees.

Despite the significant trading activity on DEXs, the process through which price discovery occurs on these exchanges remains largely unexplored. In particular, the unique trading mechanism and transparency rule of DEXs affect the optimal trading strategies of informed traders and the incorporation of private information into prices, which is a significant component of the price discovery process. Specifically, informed traders have to balance the aggressive execution of trades to capitalize on their information against the risk of revealing that information to the market.[2] On centralized exchanges, informed traders can expedite execution through various strategies such as submitting marketable orders, placing larger orders, or sweeping orders across multiple exchanges. Conversely, they may conceal their intentions by randomizing order flows or timing liquidity to mitigate the risk of information leakage. In contrast, on DEXs, informed traders can explicitly express their willingness to pay for a faster execution through the fee bidding mechanism. Moreover, the transparency of pending orders and their associated fees on DEXs increases the likelihood of

[1] Trading volume data is sourced from https://www.theblock.co/data/decentralized-finance/dex-non-custodial.

[2] The foundational study that established the channel through which private information is incorporated into prices in centralized exchanges is Kyle (1985). In this model, market makers update their beliefs about the fundamental prices of assets from the aggregated order flow. Further studies have analyzed the price informativeness of signals such as block trades (Easley and O'Hara 1987), odd-lot trades (O'Hara, Yao, and Ye 2014), and limit-order-book events (Brogaard, Hendershott, and Riordan 2019).

information leakage for informed traders. These unique characteristics make established theories and empirical designs from traditional CEXs not readily applicable to the analysis of the price discovery process on DEXs. Consequently, a novel analysis is necessary.

Our study takes advantage of the priority fee bidding mechanism and transparency of DEXs to address the following fundamental question: do informed traders on DEXs tend to bid high priority fees for quick execution, or do they lower their bids to prevent information leakage? To address this question, we analyze the most actively traded token pairs on Uniswap, the largest decentralized exchange (DEX), based on their daily average trading volume during our sample period from November 18, 2020, to August 4, 2021. Our analysis begins by demonstrating that high-fee DEX trades contain more private information than low-fee trades, using a reduced-form price impact measure defined as the average post-trade midquote change. We then estimate the permanent price impacts of DEX trade flows within a structural vector-autoregressive (VAR) model, allowing us to capture the private information content of these trades. Our findings reveal that high-fee DEX trade flow is considerably more informative than low-fee DEX trade flow. Specifically, across our four sample token pairs, the average permanent price impact of high-fee DEX trade flow ranges from 4.86 to 5.27 basis points, compared to just 0.5 to 0.62 basis points for low-fee DEX trade flow. The magnitude of high-fee DEX trade flow's price impact is economically significant as the return standard deviation is about 15 basis points.

We further demonstrate that the greater informativeness of high-fee DEX trades is driven by informed traders actively bidding high fees, rather than by uninformed traders naturally clustering at low fees. Specifically, we directly identify informed traders (wallet addresses) based on their volume-weighted average price impact. Then we examine the relative block positions of the trades executed by the informed traders and uninformed ones respectively. The results confirm that informed traders bid high fees to prioritize their trades at the top of the blocks. In contrast, uninformed traders tend to bid lower fees, resulting in their trades being more uniformly distributed throughout the block.

We provide evidence that it is trading on private information, rather than arbitrage on public information, that drives the observed pattern that high-fee DEX trades are more informed. We first identify the two most common arbitrage trades based on public information, namely, sandwich-attack trades and CEX-DEX arbitrage trades. We then exclude them when estimating the structural VAR model. that, with public-information arbitrage trades excluded, high-fee trade flow still has a significantly larger permanent price impact than low-fee trade flow.

Our findings cannot be attributed solely to trade size. Previous literature has shown that larger trades tend to be more informative (Easley and O'Hara 1987). Additionally, traders on DEXs typically bid higher fees for larger trades since the fee is fixed regardless of trade size, making trade size a potential confounding factor. However, we demonstrate that this is not the case. By categorizing trades into distinct size buckets, we observe that even within the same size bucket, high-fee DEX trade flow remains more informative than low-fee trade flow. This indicates that the fees bid by informed traders convey additional private information beyond what is captured by trade size.

Informed traders bid high fees even when they have the option to hide their orders. To empirically validate this claim, we analyze the introduction of private pools on the Ethereum blockchain. This innovation allows traders to send their orders directly to validators, bypassing public mempool exposure. Surprisingly, when we expand our sample to include the period after the introduction of private pools, we find that our results remain qualitatively the same: high-fee DEX trade flows in the public mempool remain more informative than low-fee flows. In other words, informed traders prefer to stay in the public mempool and bid high fees to execute their orders, despite the availability of private pools as an alternative venue.

What drives informed traders to bid high fees for their order executions on DEXs? Could it simply be attributed to increased blockchain congestion resulting from unrelated trading activities? Our findings suggest otherwise. Notably, we observe that informed traders on DEXs are willing to pay fees significantly higher than what is necessary for their trades to be included in a congested block. When we exclude these trades from consideration, the permanent price impact of the high-fee DEX trade flow diminishes by approximately one-third. This underscores that these high-fee trades are indeed initiated by informed traders.

Informed traders on DEXs, when choosing to bid excessively high fees, expose themselves to the risk of leaking information. To explain the rationale behind such a bidding strategy, we utilize a unique tick-by-tick mempool order dataset. We discover that a predominant portion of DEX trades with excessively high fees, fluctuating between 73.29% and 87.95% across token pairs, likely arises from a specific bidding strategy referred to as jump bidding (Daniel and Hirshleifer 1998; Avery 1998). Instead of beginning with a lower bid and then competitively increasing it, informed traders tend to make a large initial fee bid. This jump bidding strategy serves as a signaling mechanism to potential competitors, discouraging them from entering into competition. It is particularly effective on the blockchain, given that fee bidding operates as an all-pay auction, making competitive bidding costly.

Our paper is mostly related to the branch of literature analyzing the trading strategies of informed traders and their implications on price discovery under different market and information structures. Facing no competition, a monopolistic informed trader reveals her information linearly in time through trades (Kyle 1985). Price discovery is fast when there is competition among multiple privately informed traders (Holden and Subrahmanyam 1992; Foster and Viswanathan 1996; Back, Cao, and Willard 2000) or impatience of informed traders due to uncertain timing of the public announcement of private information (Caldentey and Stacchetti 2010) or short information horizon (Kaniel and Liu 2006). In contrast, price discovery is slow when informed traders can time their trades (Collin-Dufresne and Fos 2015). We contribute to the literature by analyzing informed trading in markets like DEXs, where trading intentions are communicated through public auctions featuring fee bidding. Our findings indicate that priority fees serve as a public signal of private information, and emphasize a unique competitive mechanism among informed traders through jump bidding strategies.[3]

[3]Past studies in market microstructure have linked the private information contained in trades to their public characteristics, e.g., block trades versus non-block trades (Easley and O'Hara 1987), odd-lot trades versus round-lot trades (O'Hara, Yao, and Ye 2014), trades executed on ECNs versus the NASDAQ exchange (Barclay, Hendershott, and McCormick 2003).

Our study adds to the emerging and fast-growing body of literature on decentralized exchanges, emphasizing the role of priority fees in incentivizing trading and liquidity provision. Park (2023) focuses on the unintended consequence of public blockchain order processing, which exposes all pending DEX transactions to the risk of a "sandwich attack". Capponi and Jia (2021) investigate the effect of DEX pricing rules on welfare and liquidity provision incentives. Barbon and Ranaldo (2022) study the transaction costs on DEXs. Lehar and Parlour (2021) contrast DEXs running an AMM with a limit-order-book market and focus on the different trade-offs faced by liquidity providers. Aoyagi and Ito (2021) model the coexistence of an AMM-based DEX and a limit-order-book market and study the resulting equilibrium in liquidity provision. Lehar, Parlour, and Zoican (2022) show that the priority fee drives the fragmentation of liquidity supply across DEX liquidity pools. Hasbrouck, Rivera, and Saleh (2022) demonstrate that increasing DEX trading fees can increase DEX trading volume. Han, Huang, and Zhong (2022) show that traders respond to trading on DEXs and more so when the user base of the latter increases. Unlike the above mentioned studies, our work uniquely emphasizes the role of priority fees in shaping the bidding behaviors of informed traders, highlighting its implications for price discovery.

The paper proceeds as follows. In Section 2, we introduce institutional details of DEXs and their unique characteristics. In Section 3, we describe our dataset. We present the empirical methodology in Section 4. In Section 5, we provide evidence that high-fee bidding transactions reveal private information. In Section 6, we study the economic motives behind their high-fee bidding strategy. We conclude in Section 7.

2 Institutional Background on DEXs

DEXs are exchanges that operate on a decentralized blockchain network. Trades are executed through automated smart contracts which allow for peer-to-peer trading without the need for centralized intermediaries. To execute on a DEX, a trader must broadcast the order in the peer-to-peer network of the blockchain on which the DEX is deployed and bid a priority fee. Once the order is received by the blockchain validators[4] it becomes a pending order in their mempools. At discrete times, one validator is chosen to append the next block to the chain. As the block space is limited, the validator will execute orders in her mempool in descending order of priority fees. It is worth noting that DEX orders compete for block space with order flow unrelated to decentralized finance (DeFi) transactions (e.g. cryptocurrency payments and initial coin offerings) but pending in the mempool of the same validator.

DEXs exhibit distinct operational differences compared to CEXs. While CEXs process orders in a continuous manner, DEXs execute them in discrete intervals, determined by their underlying blockchain infrastructure.[5] On CEXs, continuous trading favors fast traders, because the order execution priority depends on the order's arrival time. This leads to intense competition among high-frequency traders, potentially undermining market liquidity (Budish, Cramton, and Shim 2015; Aquilina, Budish, and O'Neill 2021).

Traders on DEXs have to bid a priority fee[6] to determine the execution priority of their orders. If traders are willing to pay a higher fee, it is more likely their orders will be executed quickly. This dynamic in DEX execution stands in contrast to the frequent batch auction design put forth by Budish, Cramton, and Shim (2015) as a remedy against the arms race between fast traders. While both systems process orders in batches, a frequent batch auction treats all orders equally and matches them at a single price. In contrast, DEXs process orders based on the fee amount — those with higher fees are prioritized for earlier execution and receive more favorable prices. For example, a buy order executed before other buy orders from competitors can receive a lower price. Therefore, rather than engaging in a speed-based competition like on CEXs, traders on DEXs compete for execution priority through their choice of priority fees.

When orders are submitted to DEXs via the peer-to-peer blockchain network, they first enter

[4]Validators ensure the authenticity of transactions, incorporate them into new blocks, and then add these blocks to the chain, thereby earning the priority fees corresponding to those transactions.

[5]For instance, on the proof-of-work (PoW) Ethereum blockchain, blocks are validated on average every 13 seconds, though this interval can vary. However, after transitioning to proof-of-stake (PoS) on September 6, 2022, Ethereum reduced the block time to 12 seconds.

[6]On the Ethereum blockchain, traders have to bid a "gas fee" in order to execute their orders.

a pending state within the public mempools of validators before actual execution. This implies that market participants monitoring the network can observe every new order and any subsequent revision. For instance, they will be able to infer the size and direction of the trade, and the priority fee bid for a pending market order. In contrast, market orders are not visible ex-ante on CEXs, they can only be inferred from trade executions. In markets with pre-trade transparency, only resting *limit* orders are observed, not the *market* orders.[7]

CEXs operate largely as LOB markets where incoming market orders are executed against resting limit orders submitted by market makers. As a result, LOB prices adjust as market makers revise their quotes responding to either public information such as news or to private information revealed by the trade executions. For instance, Brogaard, Hendershott, and Riordan (2019) show that price discovery occurs predominantly through limit orders in the Canadian stock market. In contrast, most DEXs employ an automated market maker (AMM) model where liquidity providers deposit tokens, and incoming orders execute against a predefined pricing curve, eliminating the concept of traditional quotes. As a result, AMM prices shift solely based on trade executions, which reflect both public and private information. We delve deeper into the nuances of these informational sources in Section 4.

3 Data

In this section, we describe the dataset used in our empirical analysis. DEXs and other DeFi services are, at present, available on various blockchains, including Ethereum, Binance Smart Chain (BSC), Tron, Arbitrum, Polygon, Avalanche, and Optimism. Among these, Ethereum holds, by far, the dominant market share, surpassing 50% in terms of total value locked (TVL).[8] Given this

[7] Since the introduction of Flashbots in early 2021, traders can now use private relays to send their orders directly to validators. As a result, the traditional "public pool" where all traders' orders were visible to anyone on the network is no longer the sole option. Validators now monitor their own private pools alongside a public mempool. Consequently, traders' pending orders are fragmented across various private pools, and orders in these private pools cannot be observed by other traders using the same pool. Additionally, traders must compete in first-price sealed-bid auctions within the private pools. In Section 5.4.3, we examine the impact of private pools on price discovery in DEXs.

[8] For detailed statistics, we refer to https://defillama.com/chains.

overwhelming prominence, our study focuses exclusively on the Ethereum blockchain. Additionally, DEXs account for approximately 10% of cryptocurrency spot trading, with Uniswap standing out as the most prominent DEX, responsible for over half of the entire DEX trading activity. The remaining 90% of cryptocurrency spot transactions occur on centralized exchanges (CEXs), where Binance dominates the landscape, capturing more than 60% of the CEX trading volume.[9] Given the dominant market shares of both Uniswap and Binance, our analysis is based on these two exchanges. We describe executed trade data in Section 3.1, and mempool order data in Section 3.2.

3.1 Executed trade data

Our dataset includes trades involving four highly traded token pairs, covering the period from November 18, 2020, to August 4, 2021.

For our baseline analysis, we focus on the time frame from November 18, 2020, to February 10, 2021. This timeframe follows the conclusion of Uniswap's staking reward program, which ran from September 18 to November 17, 2020. During this program, Uniswap allocated 20 million UNI, its governance token, to liquidity providers in pools such as WBTC-ETH, ETH-USDC, ETH-USDT, and ETH-DAI to enhance liquidity. This selection includes the period before the first block including Flashbots transactions was mined. The remaining period is used to assess the robustness of our results to the introduction of Flashbots.[10]

We employ a systematic approach to select token pairs from the numerous pairs traded on CEXs and DEXs. The selection process involves the following steps:

1. We start by identifying the top 50 crypto tokens by market capitalization as of the beginning of our sample period, and consider pairs that include both tokens from this list.

2. We then filter to include only those pairs traded on both Uniswap and Binance.

[9] We refer to `https://www.theblock.co/data/decentralized-finance/dex-non-custodial` for trading volume estimates on DEXs, and to `https://www.coingecko.com/en/exchanges` for trading volume estimates of the top crypto exchanges.

[10] Flashbots provide off-chain private channels through which traders can send their orders directly to validators, thereby avoiding the risk of being front-run in the public mempool. The first block mined with Flashbots transactions was on February 10, 2021.

3. Next, we retain token pairs with at least 500 trades per day on average during our sample period. Figure 1 below shows the average daily number of trades and average daily trading volume in USD on Uniswap for all token pairs meeting the first two criteria. As clearly seen from the figure, the trading activities of a few pairs, such as ETH-USDT and ETH-USDC, significantly surpass the rest.

4. We exclude token pairs consisting solely of stablecoins, such as USDC-USDT and DAI-USDT.

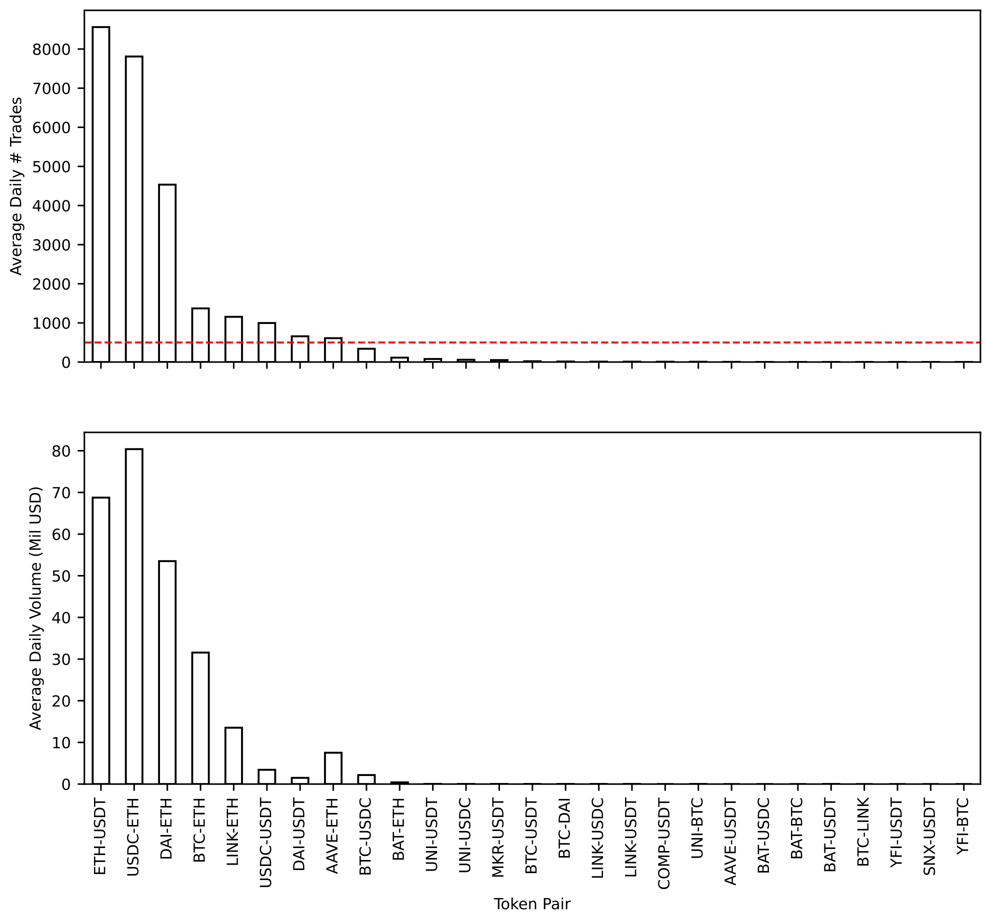
5. We remove token pairs that essentially track the same price. For instance, ETH-USDT, ETH-USDC, and ETH-DAI all effectively track the price of Ethereum. We keep only ETH-USDT, as it has the highest daily average number of trades among the three. As a robustness check, we replicate all main analyses for the other two pairs, ETH-USDC and ETH-DAI, and find similar results. These results are included in the online appendix for brevity.

Based on the criteria mentioned above, we identify four token pairs: ETH-USDT, WBTC-ETH, LINK-ETH, and AAVE-ETH. Binance trades are publicly available and can be accessed from the Binance website[11]. Uniswap trades, on the other hand, are collected through Amberdata[12]. Note that Uniswap v3 was launched on May 5, 2021, causing liquidity to be fragmented between the Uniswap v2 and Uniswap v3 pricing protocols. To provide a comprehensive view of trading activities on Uniswap, we collect executed trades from both Uniswap v2 and Uniswap v3 pools, covering all swap fee levels. When classifying Uniswap trades into high-fee, mid-fee, and low-fee categories, we base the classification on gas prices, without differentiating the pools from which the trades originate. In Section G of the appendix, we construct the trade flows for v2 and v3 separately and estimate their corresponding permanent price impacts. Table A5 reports the detailed results. In general, we find the results are mixed. For example, when comparing Uniswap v2 and the Uniswap v3 pool with a 5bp swap fee (“Uniswap v3-5bp”), the results depend on the token pairs. For ETH-USDT and WBTC-ETH, the Uniswap v2 pool seems to have a larger permanent price impact than

[11] Details can be found at https://data.binance.vision/?prefix=data/spot/monthly/.

[12] Amberdata is a US-based company specializing in market data for decentralized finance.

Figure 1. Average daily trading volume and number of trades. This figure plots the average daily trading volume and number of trades for a sample of token pairs. The trading volume is in million dollars. The sample period is between November 18, 2020 and February 10, 2021. The red dotted line indicates the threshold on the average daily trades below which we discard token pairs.



the Uniswap v3-5bp pool. In contrast, for the other two tokens, the Uniswap v3-5bp pool has a larger permanent price impact. When comparing the price impact of trade flows from the Uniswap v3-5bp pool and the Uniswap v3-30bp pool, the results remain mixed.

Below, we provide a detailed description of each trade dataset.

Uniswap trades. Each Uniswap trade contains the following information types:

- Timestamp: the timestamp of the block in which the trade is included (to the precision of a second), the number of the block in which the trade is included, and the execution position of the trade in that block.

- Identifiers: hash, submission address, and nonce.

  - Hash: the hash is a unique identifier for each new order submitted to the network. Using the hash, we can match an executed trade with its original order. Note that when a trader modifies a pending order, the modified order will be assigned a new hash.

  - Submission address and nonce: nonce is used to track the orders sent from a given submission address. Specifically, the first order of a trader is assigned nonce "0", her second order has a nonce "1", and her N^{th} order has a nonce "N". Importantly, a new order will not be executed if there are pending orders with smaller nonces from the same submission address. It means that if a trader wants to modify or cancel her pending order, she has to broadcast a new order with the same nonce and a higher priority fee. A validator will only execute the new order as she prioritizes orders with higher fees.[13] Through the submission address and the nonce number, we can link an executed or canceled order with its history of revisions.

- Trade characteristics: gas price, gas used, trade direction, and the amount of tokens that the trader deposits in and takes out from the liquidity pool.

  - Gas price and gas used: on the Ethereum blockchain, the priority fees are referred to as "gas fees".[14] The "gas used" of a transaction measures the fixed amount of computational resources needed for its execution. More complicated transactions require more computational work and thus consume a higher amount of gas. Upon bidding, Ethereum users choose the "gas price", that is, the unit price of gas they are willing to pay. Hence, the total gas fee paid by users is equal to the gas used multiplied by the

[13] A trader can cancel a pending order by submitting a new order with the same nonce but a higher priority fee, where she transfers zero amount of the native token of the blockchain (e.g., ETH on the Ethereum blockchain) to her own wallet. The old order is effectively "canceled" as only the new order with a higher priority fee will be executed by validators.

[14] The London Upgrade of the Ethereum blockchain on August 5, 2021, implemented the EIP-1559, and decomposed the gas fee into two parts: base fee and priority fee (tips). The base fee is a reserve price every trader needs to pay, and adjusts to the congestion level of the network. The base fee gets burnt and thus is not earned by validators. The priority fee or tip instead is bid by traders to incentivize validators to include their transactions in the next block. We use the terminology priority fee to stress its role in determining the execution priority of pending orders.

gas price. Note that Ethereum validators sort and execute transactions in mempools in decreasing order of gas price.

- Trade direction: it indicates whether it is a buy trade or sell trade in terms of the base token. We follow the convention used for currency pairs in the foreign exchange market and label the first token appearing in a pair as the base token and the second token as the quote token. For example, the token pair ETH-USDT has ETH as the base token and USDT as the quote token.
- The amount of tokens that the trader deposits in and takes out from the liquidity pool: we use the amount of the base token swapped as the transaction size of the trade.

Binance trades. Each Binance trade record includes a unique identifier for the trade, the timestamp (to the precision of a millisecond), the transaction price, the transaction size in terms of the base token, and an indicator for whether the buyer uses a limit order or a market order, which tells us the direction of the trade: if the buyer uses a market order, then it is classified as a buy trade; otherwise, it is a sell trade.

In addition to executed trades, we obtain event updates of Binance's limit order book (to the precision of a second). With order book event updates, we are able to reconstruct the order book states and calculate the best bid, best ask, and the midquote on Binance, which we use to calculate token pair returns.

3.2 Mempool order data

We obtain tick-by-tick Ethereum mempool order data from Amberdata. Our mempool data covers the same sample period of November 18, 2020, through August 4, 2021. The dataset comprehensively covers every new order submission received within the mempool of nodes maintained by Amberdata. These orders may ultimately be either executed or left unexecuted. Each order comes with the following information: the hash, the timestamp when the order is received by the node (to the precision of a millisecond), the address of the trader, nonce, gas price, and gas limit (i.e. the

Table 1. Summary statistics of daily trading statistics on Uniswap and Binance. This table reports, for each token pair, summary statistics of daily trading volume (TradingVolume) and number of trades (TradeCount) on Uniswap and Binance respectively. N refers to the number of days in our sample period. Trading volume is denominated in ETH.

Pair		N	Mean	SD	Min	Med	Max
ETH-USDT	TradingVolume-Uniswap	85	73,489	37,752	36,923	62,131	263,356
	TradeCount-Uniswap	85	8,560	1,700	6,311	8,155	16,419
	TradingVolume-Binance	85	1,444,426	709,203	493,012	1,281,734	4,245,010
	TradeCount-Binance	85	994,231	524,099	272,746	915,584	2,577,496
WBTC-ETH	TradingVolume-Uniswap	85	31,644	17,748	9,014	27,141	87,965
	TradeCount-Uniswap	85	1,371	592	646	1,127	3,338
	TradingVolume-Binance	85	2,023	1,993	135	1,258	9,984
	TradeCount-Binance	85	7,886	7,529	289	5,332	35,191
LINK-ETH	TradingVolume-Uniswap	85	10,779	6,295	3,437	9,406	42,520
	TradeCount-Uniswap	85	1,054	380	574	961	2,682
	TradingVolume-Binance	85	4,387	2,687	1,071	3,856	13,598
	TradeCount-Binance	85	10,459	6,793	2,223	9,391	29,514
AAVE-ETH	TradingVolume-Uniswap	85	7,368	4,177	1,766	6,366	29,936
	TradeCount-Uniswap	85	609	253	261	551	1,514
	TradingVolume-Binance	85	2,135	1,510	408	1,627	10,143
	TradeCount-Binance	85	6829	5,410	1,131	5,511	36,964

maximum gas allowed to be used).

The mempool data is used for two purposes. First, we can identify whether a transaction comes from public pools or private pools. For trades from private pools, there is no matching pending order in the public pool; Second, for a public transaction, we can track its complete history of order revisions, if they occur, before the final order is executed and recorded as a trade. Consequently, we can observe whether the trader increased the gas price attached to her order in order to ensure its execution.

3.3 Summary statistics of executed trades

In Table 1, we provide an overview of trading characteristics for our sample token pairs. We report summary statistics of their daily trading volume and their daily number of trades on Uniswap and Binance during our sample period between November 18, 2020, and February 10, 2021. Several observations are in order. First, trading in all four token pairs is fairly active. For instance, the average daily number of trades (daily trading volume) on Uniswap is 8,560 (73,489 ETH $\approx$ 66 million

Table 2. Summary statistics of Uniswap trades. This table reports, for each token pair, summary statistics of the transaction price (TxPrice), transaction size (TxSize), and gas price (GasPrice). Gas price is denominated in Gwei, which equals to 10^{-9} ETH. N refers to the number of trades for each token pair during our sample period. Transaction size is denominated in ETH.

TokenPair	Variable	N	Mean	SD	1%	10%	Median	90%	99%
ETH-USDT	TxPrice	727600	891.94	379.26	474.95	546.93	653.14	1397.87	1751.67
	GasPrice	727600	99.77	189.92	15.30	30.00	70.00	181.00	530.00
	TxSize	727600	8.59	34.08	0.01	0.13	1.37	15.34	124.89
WBTC-ETH	TxPrice	116520	30.48	5.44	22.56	23.88	31.44	38.28	42.33
	GasPrice	116520	126.13	220.12	17.00	37.00	88.00	240.00	652.00
	TxSize	116520	23.07	59.40	0.02	0.23	3.99	64.76	235.89
LINK-ETH	TxPrice	89630	0.02	0.00	0.01	0.01	0.02	0.02	0.03
	GasPrice	89630	114.48	242.18	16.00	34.00	78.89	205.70	669.82
	TxSize	89630	10.22	24.36	0.02	0.19	2.82	27.20	86.39
AAVE-ETH	TxPrice	51811	0.16	0.05	0.09	0.11	0.15	0.27	0.31
	GasPrice	51811	110.91	177.61	15.56	30.72	80.00	203.00	565.62
	TxSize	51811	12.08	19.93	0.03	0.20	4.85	29.95	88.13

USD) and 1,371 (31,644 ETH ≈ 28 million USD) for ETH-USDT and WBTC-ETH respectively.

Second, trading activity on Uniswap and Binance differs significantly across token pairs. For ETH-USDT, trading is much more active on Binance than on Uniswap. For example, the average daily trading volume on Binance is about 144 million ETH, more than an order of magnitude larger than that on Uniswap. In contrast, for other token pairs, trading is in general more active on Uniswap than Binance.

In Table 2, we report summary statistics of the execution price, gas price and trade size of Uniswap trades for our four token pairs. The average size of an Uniswap trade is fairly large and amounts to about 8.59 ETH (≈ 7,661 USD) and 23.07 ETH (≈ 20,577 USD) for ETH-USDT and WBTC-ETH respectively. There is significant heterogeneity in the gas price paid for Uniswap trades. Take WBTC-ETH as an example. While a Uniswap trade in WBTC-ETH has an average gas price of 126.13 Gwei (1Gwei = 10^{-9}ETH), its standard deviation is 220.12, which is about twice the size of the mean. Such a large variation can result from either a change in the overall congestion of the Ethereum network or from traders' bidding high fees to trade on the information.

4 Empirical Methodology

Although there is no cash flow news for crypto assets, there is value generated by the enterprises issuing these assets. For example, protocol-level tokens, such as Aave, offer Web3 services such as decentralized borrowing and lending, and thus the value of the token is linked to the future adoption of their protocols. Oracle-related tokens, such as LINK, are associated with Oracle services provided by ChainLink, which feed price information from off-chain sources to on-chain smart contracts. Its value is linked to the usage of Oracle services in the DeFi ecosystem. Blockchain-related tokens, such as ETH and WBTC, have value linked to the adoption of the Ethereum and Bitcoin blockchains, respectively.

Some market participants can acquire private information about the value of these tokens from several sources. First, it can come from predicting news events. For example, the announcement by AAVE DAO about their new stablecoin GHO led to a rapid increase in AAVE's price. Similarly, the announcement of Uniswap voting on a fee switch and Ethereum ETF S-1 filings have caused short-term price movements in ETH. Second, there are some short-horizon proprietary trading signals: sophisticated proprietary traders and market makers, such as Wintermute and Symbolic Capital, often operate with short-term strategies based on algorithmic signals. These signals might utilize indicators such as short-term open interest changes, liquidation events, and funding payments in derivative markets. This information has often not yet been incorporated into spot market prices.

The de facto standard approach for estimating the private information contained in trades was proposed by Hasbrouck (1991a) and Hasbrouck (1991b). The security return and trades are modeled as a structural vector-autoregressive (VAR) system that characterizes the dynamic interactions between them. With the structural VAR model, one can estimate the persistent impact of trades on security prices, a proxy for private information, by computing the cumulative impulse responses of security return to trade innovations over a substantially long period. In addition, one is able to decompose the total efficient price variance, a measure of the total amount of information, into a component correlated with trades and a component uncorrelated with trades. The former component reflects the amount of private information conveyed through trades while the latter captures the amount of public information such as news.[15]

To examine whether informed traders on DEX bid high fees, we follow the above mentioned approach and estimate a structural VAR model with CEX return and DEX trade flows of different fee levels. Below, we first introduce our fee-level classification method. Then we detail our structural VAR specification.

4.1 Constructing DEX trade flows with different fee levels

We adopt a quantile-based, rolling-window approach to construct DEX trade flows with three different priority fee levels. Specifically, to classify trades in the current block t, we first sort them along with all trades from the previous 20 non-empty blocks, i.e., block $t-20$ to block $t-1$ based on their priority fee in descending order. We then label trades in the block t as follows: trades in the top quartile (above the 75th percentile) are classified as high-fee trades, those in the bottom quartile (below the 25th percentile) as low-fee trades, and all other trades (between the 25th and 75th percentiles) as mid-fee trades.

There are two important points to note. Firstly, we use quantiles rather than standard deviations for trade classification because quantiles are more robust to outliers. Secondly, we employ rolling windows instead of the full sample when calculating the quantiles. This method allows us to account for the impact of time-varying fees due to blockchain congestion. For example, during periods of heightened activity, high fees are consistently prevalent across all blockchain transactions, including those occurring on DEXs. Therefore, use of full-sample quantiles may result in classifying all DEX trades during this period as high-fee trades, while categorizing all DEX trades outside that period as medium-fee or low-fee trades. In practice, when informed traders decide on

[15] Hasbrouck (1991a) and Hasbrouck (1991b) consider security return and a single aggregated signed trade flow. While different functional forms of the aggregated signed trade flow, such as the trade direction indicator and the squared trade flow, can be included in the structural VAR, all trade variables essentially originate from the same aggregated signed trade flow. Recent studies extend the approach by incorporating multiple trade flows, enabling a comparative analysis of the informativeness of different trade flows, for example, ECN trades versus NASDAQ trades (Barclay, Hendershott, and McCormick 2003), odd-lot trades versus round-lot trades (O'Hara, Yao, and Ye 2014), and orders versus trades (Fleming, Mizrach, and Nguyen 2018; Brogaard, Hendershott, and Riordan 2019).

a fee to bid for their trades, they are more concerned about the fee levels *relative* to others' bids in the current block or in the recent past blocks, instead of the *absolute* fee levels. Importantly, it is only the ranking of fees that decides the execution priority of the orders. In addition, by only using fee levels in past rolling windows, we avoid the look-ahead bias.

After labeling the trades, we compute the (signed) trade flow at a given fee level i in a block t as:

$$x_t^i = \sum_k d_{t,k}^i s_{t,k}^i, \tag{1}$$

where k indexes trades with fee level i in block t, d_k is the trade direction indicator (+1 for buyer-initiated trades and -1 for seller-initiated trades), and s_k is the trade size. Hence, for each block, we construct three DEX trade flows: $x_t^{\text{LowFee-DEX}}$, $x_t^{\text{MidFee-DEX}}$, and $x_t^{\text{HighFee-DEX}}$. If we have no observation for one type of trade in a block, the corresponding trade flow simply takes a zero value.

It is worth highlighting two choices made in the above classification method. First, we choose three different fee levels, high-fee (above 75% quantile), mid-fee (between 25% and 75% quantile), and low fee (below 25% quantile), instead of two, high-fee (above 50%) and low-fee (above 50%), to guarantee a clear distinction between the high-fee and low-fee group.

Second, we set the length of the rolling window to 20 non-empty blocks to balance two competing forces. On the one hand, a too-short window makes our quantile estimates noisy due to a small number of trades. For example, if we only use the current block and if that contains only a few trades, then two trades with very similar gas prices will fall into different categories. On the other hand, a too-long window might include trades with priority fees bid too long ago to reflect the current congestion level of the blockchain. Admittedly, the choice of 20 non-empty blocks remains arbitrary. As a robustness check, we repeat the fee level classification with window lengths of 10 and 40 blocks, and the structural VAR results stay qualitatively the same. We report the detailed results in Appendix H.3.

4.2 Measuring the informativeness of DEX trade flows

To estimate the informativeness of DEX trades, we follow Hasbrouck (1991b) and adopt a structural VAR model as follows:

$$Ay_t = \alpha + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \varepsilon_t \tag{2}$$

where $\Phi_1 \ldots \Phi_p$ are system matrices of the lagged terms of the structural VAR model. The term ε_t is the vector of structural innovations and satisfies the following conditions: $\mathrm{E}(\varepsilon_t) = 0$; $\mathrm{E}(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$; $\mathrm{E}(\varepsilon_t \varepsilon_s') = 0$ for $s \neq t$. Note that the covariance matrix Σ_ε is diagonal as the contemporaneous relations among the endogenous variables are directly modeled in A. The term $y_t = (r_t, \boldsymbol{x_t})'$ is the endogenous variable vector with return r_t and trade variables $\boldsymbol{x_t}$. The structural matrix A captures the contemporaneous correlations between the endogenous variables.

Based on the structural VAR model, we can calculate the permanent price impact of a trade flow. We provide a brief discussion next and refer to Appendix A for additional details. The permanent price impact (PPI) of a trade flow variable k is defined as the cumulative impulse response of the midquote return to a unit shock in the trade flow, that is,

$$\mathrm{PPI}_k = \frac{\sum_{j=0}^{\infty} \partial r_{t+j}}{\partial \varepsilon_{k,t}} = [\Theta(1)]_{1,k}, \quad k > 1 \tag{3}$$

where $[\Theta(1)]_{1,k}$ denotes the $(1,k)$-*th* entry of $\Theta(1) = \Theta_0 + \Theta_1 + \Theta_2 + \ldots$, the vector moving average parameters. PPI measures the impact of trades on security prices over a substantially long period. Thus it captures private information contained in the trade as it will induce permanent price changes as opposed to transitory price changes due to inventory control effects.

4.3 Designing the Structural VAR for DEXs

When applying the structural VAR to DEXs, one needs to adapt it to account for its unique trading mechanism. We adopt the following three implementation details:

Choice of the endogenous variable vector For our baseline specification, we include the midquote return on Binance and three trade flows on Uniswap with different priority fee levels in the endogenous variable vector y_t, that is,

$$y_t = \begin{pmatrix} r_t^{\text{CEX}} & x_t^{\text{CEX}} & x_t^{\text{LowFee-DEX}} & x_t^{\text{MidFee-DEX}} & x_t^{\text{HighFee-DEX}} \end{pmatrix}', \tag{4}$$

where t indexes block time, and r_t^{CEX} is the Binance midquote return from block time $t-1$ to t. x_t^{CEX} denote the Binance trade flow aggregated from block time $t-1$ to t. The variables $x_t^{\text{LowFee-DEX}}$, $x_t^{\text{MidFee-DEX}}$ and $x_t^{\text{HighFee-DEX}}$ denote Uniswap trade flows in block t, respectively with low, mid and high priority fee levels, as specified in the above priority fee level classification. Note that informed traders trading on their private signals are likely to split their trades between CEXs and DEXs in order to reduce price impacts. As a result, CEX and DEX trade flow can be positively correlated. Therefore, we include CEX trade flow in the structural VAR specification as a control. We use the midquote return on Binance in the baseline specification as Binance operates a CLOB market and market makers can easily revise the quotes without trades. In Appendix F, we include the return calculated from Uniswap prices in the endogenous vector and re-estimate the permanent price impacts of DEX trade flows of different levels of fees. Table A4 reports the detailed estimation results and shows that our key finding holds: high-fee DEX trade flow has a much larger permanent price impact than low-fee flow.

Timestamp convention To effectively integrate the return on Binance, which operates on a continuous clock for order execution, and the trade flows on Uniswap, which operate on a discrete clock, into a unified structural VAR model, it is important to establish a timestamp convention that encompasses the characteristics of both clocks. To do so, we define r_t^{CEX} as the log difference between the Binance midquote at block time $t-1$ and t respectively. All three Uniswap trade flows, $x_t^{\text{LowFee-DEX}}$, $x_t^{\text{MidFee-DEX}}$ and $x_t^{\text{HighFee-DEX}}$, are computed based on trades executed at block time t as in Equation 1. In contrast, the trade flow on Binance, x_t^{CEX}, is computed by considering trades executed on Binance within the time period between the block time $t-1$ and t.[16]

Resolution of the contemporaneous correlations To pin down the permanent price impacts of the Uniswap trade flows, one needs to specify the A matrix in the structural VAR model, as it influences the reduced-form system matrices (see Equation 2). This matrix dictates the contemporaneous causality between the endogenous variables.

We resolve the contemporaneous correlations as follows. We allow CEX returns to contemporaneously cause DEX trade flows but not vice versa. This restriction ensures that the price impact of DEX trade flows captures private information. Specifically, by not allowing DEX trade flows to affect CEX returns contemporaneously or within the same block interval, any price impact from DEX trade flows can only occur starting from the next block interval. Since public information is incorporated into Binance prices during the contemporaneous block interval, our measure of price impact, which captures the effect of trades in the current block interval on Binance returns in the subsequent block interval, is more likely to reflect private information contained in the trade flows rather than public information.

To further illustrate the difference between price impacts due to public and private information, consider the following example. Suppose the price on Binance increases due to positive public news during block time $t-1$ and t. Arbitrageurs will race to trade against the now stale and lower price on Uniswap at block time t, re-aligning the prices on Binance and Uniswap. Consequently, we observe a contemporaneous impact of DEX trade flows on Binance prices from block time $t-1$ and t. However, this price impact reflects only the public news already known to market participants, not any new private information. By excluding the contemporaneous price impact of DEX trade flows, we avoid mistakenly attributing it to private information. Instead, we focus on

[16] Because r_t^{CEX} is defined over block time $t-1$ and t and Uniswap trade flows, $x_t^{\text{LowFee-DEX}}$, $x_t^{\text{MidFee-DEX}}$ and $x_t^{\text{HighFee-DEX}}$, are defined at block time t, it might appear that the return on Binance occurs prior to the Uniswap trade flows. In reality, these two factors can contemporaneously influence each other. Although Uniswap trades are all executed at the same block time t, market makers on Binance can monitor pending orders in the mempool and predict which orders will be executed based on their priority fees. They can then adjust their quotes on Binance accordingly. For instance, if there are many pending buy orders in the mempool, market makers might might revise up their quotes on Binance.

the impact of DEX trade flows executed at block time t on Binance returns from t to $t+1$, which is more likely to capture private information rather than public information.

Additionally, we do not impose a specific contemporaneous causality among the trade flows. Instead, we try all possible contemporaneous causalities based on a generic recursive structure to obtain the lower and upper bound of the permanent price impacts and information shares of the trade flows (Hasbrouck 1995; Barclay, Hendershott, and McCormick 2003; O'Hara, Yao, and Ye 2014). In econometric terms, we use Cholesky factorization to decompose the reduced-form covariance matrix, which is equivalent to imposing a recursive structure on the A matrix. Without loss of generality, we let the A matrix be a lower-triangular matrix. Thus we assume that the first variable in the endogenous variable vector, y_t, contemporaneously causes the second variable onwards and the second variable contemporaneously causes the third variable onwards, and so on. For example, if we place the low-fee Uniswap trade flow before the high-fee flow, we assume that the former contemporaneously causes the latter. Thus, we are likely to obtain an upper bound for the permanent price impact and information share of the low-fee Uniswap trade flow.[17]

Other implementation details include: (1) the model is estimated at block-by-block frequency, although the priority fee level classification is based on a 20-block rolling window; (2) we set the number of lags in the structural VAR model to 5. In Appendix H.4, we change the number of lags included in the structural VAR model to 10 and 20, and show that estimation results remain qualitatively the same; Note that by choosing lags of 5, 10, and 20 blocks, we focus on an information horizon that is neither too brief nor too long, which thus makes informed traders' choice of priority fees nontrivial; (3) As the base currency varies across token pairs, to ease comparison and aggregation across token pairs, we standardize all trade flow variables such that they have zero

[17] In some earlier applications of the structural VAR approach (see, e.g., Hasbrouck 1991a; Hasbrouck 1991b; Fleming, Mizrach, and Nguyen 2018; Brogaard, Hendershott, and Riordan 2019), trades are assumed to contemporaneously cause quote revisions, not vice versa. Adopting the same timing convention is problematic in our application. First, our estimation frequency is block-by-block, which on average is about 12 seconds. Given such a relatively long interval, the contemporaneous correlations between the Binance return and Uniswap trade flows can be quite large. Second, even though we could impose a specific contemporaneous causality between the Binance return and Uniswap trade flows, it is not straightforward to impose one among the Uniswap trade flows themselves. Although Uniswap trades are executed in descending order of priority fees, meaning the high-fee trade flow is executed before the mid-fee and low-fee trade flows, all are executed at the same block time t. For example, low-fee trades might be submitted before the high-fee trades and thus enter the mempool earlier, they can be executed at the same block time t.

mean and unit variance. Hence, the impulse responses reported below should be interpreted as permanent price impacts in basis points per standard deviation increase in the trade flow.

5 Do High-fee DEX Trades Reveal Private Information?

We provide evidence that high-fee DEX trades contain private information. We begin by presenting results based on a reduced-form price impact measure in Section 5.1. In Section 5.2, we consider the permanent price impact measure estimated from the structural VAR model. In Section 5.3, we show that high-fee DEX trades being more informed is the result of informed traders bidding high fees as opposed to uninformed traders clustering at low fees. In Section 5.4, we show that the results are robust with the respect to trade sizes and to the introduction of private pools.

5.1 A reduced-form price impact measure

While a structural VAR model is useful for distinguishing between permanent and transitory price impacts, it imposes specific functional forms and assumptions on the contemporaneous relationships between return and trade variables. Therefore, before estimating the permanent price impact using the structural VAR model, we first employ a straightforward, reduced-form price impact measure commonly used in the microstructure literature. Specifically, for a single trade i, the price impact measure is calculated as:

$$\mathrm{RPI}_i = \frac{d_i\left(\mathrm{Mid}_{t+\Delta_t} - \mathrm{Mid}_t\right)}{\mathrm{Mid}_t},$$

where i indicates trades, Mid_t represents the Binance midquote *just before* trade t, and $\mathrm{Mid}_{t+\Delta_t}$ represents the prevailing Binance midquote Δ_t after the trade. Selecting an appropriate time interval Δ_t is important: too short an interval may capture only transitory price impacts, while too long an interval might incorporate other informational events, such as news. We therefore experiment with different window lengths, specifically 20 blocks, 60 blocks, and 120 blocks as window lengths, which roughly correspond to 5 minutes, 15 minutes, and 30 minutes, respectively. The variable d_i denotes the trade direction indicator. In CEXs with a limit order book, the midquote is calculated as the average of the best bid and ask prices. After calculating the price impact for each individual trade, we compute the size-weighted average relative price impact for trades with low, mid, and high fees.

Table 3 presents the average relative price impact measures over different time horizons for low-fee, mid-fee, and high-fee DEX trades respectively. It shows that across all four token pairs, the average RPI increases in the fee level. For example, when measuring the price impact over a 20-block horizon, the average RPI of the high-fee trades in ETH-USDT is 9.27 basis points, while that of the low-fee trades is only 1.41 basis points.

Table 3. Reduced-form price impacts of DEX trades with different priority fee levels. This table reports the reduced-form price impacts for DEX trades with different priority fee levels. The reduced-form price impact is calculated as the midquote change during a specific horizon after a trade occurs. The information in the parentheses, 20 blocks, 60 blocks, and 120 blocks, refer to the horizons over which we calculate the price impact measures. The specific formulas are given by Equation 5.1. The reduced-form price impact is in basis points.

Token Pair	Fee Level Measure	Low-Fee DEX Trades	Mid-Fee DEX Trades	High-Fee DEX Trades
ETH-USDT	RPI (20 Blocks)	1.41	2.74	9.27
	RPI (60 Blocks)	2.35	4.04	12.42
	RPI (120 Blocks)	2.82	3.6	11.91
LINK-ETH	RPI (20 Blocks)	0.72	3.36	9.35
	RPI (60 Blocks)	2.32	5.44	10.77
	RPI (120 Blocks)	1.95	5.49	12.3
WBTC-ETH	RPI (20 Blocks)	1.84	4.91	9.82
	RPI (60 Blocks)	2.72	6.98	13.95
	RPI (120 Blocks)	1.74	9.09	15.98
AAVE-ETH	RPI (20 Blocks)	2.14	8.85	11.72
	RPI (60 Blocks)	3.55	12.05	14.43
	RPI (120 Blocks)	5.35	13.29	14.55

5.2 Permanent price impact from the structural VAR model

Prior to discussing the estimation results of the structural VAR model, we present summary statistics for the return and trade flow variables of each token pair in Table 4. Two observations are in order. First, consistent with the liquidity summary statistics in Table 1, the magnitude of trade

Table 4. Summary statistics of CEX return, CEX trade flow, and DEX trade flow variables. This table reports, for each token pair, summary statistics of the return and trade flow variables used in the structural VAR estimation. r_t^{CEX} is Binance return from block time $t-1$ to t. x_t^{CEX} is Binance trade flow. x_t^{DEX} is Uniswap trade flows. $x_t^{\text{LowFee-DEX}}$, $x_t^{\text{MidFee-DEX}}$ and $x_t^{\text{HighFee-DEX}}$ are Uniswap trade flows consisting of trades from the low-, mid- and high-fee category in block t. Both r_t^{CEX} and r_t^{DEX} are in basis points. N refers to the number of blocks for each token pair during our sample period. All trade flow variables are denominated in ETH.

		N	Mean	SD	Min	50%	Max
ETH-USDT	r_t^{CEX}	370291	0.03	10.27	-476.61	0.00	368.22
	x_t^{CEX}	370291	-0.32	221.19	-7370.94	0.11	10152.33
	x_t^{DEX}	370291	0.15	40.76	-3111.34	0.04	2154.22
	$x_t^{\text{LowFee-DEX}}$	370291	-0.03	10.29	-2345.49	0.00	1241.70
	$x_t^{\text{MidFee-DEX}}$	370291	-0.06	21.37	-1897.53	0.00	2147.57
	$x_t^{\text{HighFee-DEX}}$	370291	0.23	33.18	-3498.28	0.00	2217.48
WBTC-ETH	r_t^{CEX}	81892	-0.05	9.12	-269.32	0.00	245.93
	x_t^{CEX}	81892	-0.02	9.93	-395.21	0.00	1991.97
	x_t^{DEX}	81892	-0.25	56.17	-2750.21	0.22	2331.24
	$x_t^{\text{LowFee-DEX}}$	81892	0.07	15.87	-475.92	0.00	698.13
	$x_t^{\text{MidFee-DEX}}$	81892	0.07	36.64	-2750.21	0.00	726.66
	$x_t^{\text{HighFee-DEX}}$	81892	-0.40	39.13	-771.15	0.00	2331.24
LINK-ETH	r_t^{CEX}	72951	-0.07	16.10	-494.76	0.00	467.55
	x_t^{CEX}	72951	-0.47	16.73	-2047.56	0.00	432.04
	x_t^{DEX}	72951	-0.08	22.57	-1187.08	0.00	652.36
	$x_t^{\text{LowFee-DEX}}$	72951	-0.04	5.32	-202.07	0.00	161.16
	$x_t^{\text{MidFee-DEX}}$	72951	-0.10	14.47	-1187.08	0.00	652.36
	$x_t^{\text{HighFee-DEX}}$	72951	0.06	16.11	-432.35	0.00	541.94
AAVE-ETH	r_t^{CEX}	42975	0.14	29.89	-509.77	0.00	582.37
	x_t^{CEX}	42975	-0.31	10.83	-676.27	0.00	239.78
	x_t^{DEX}	42975	0.14	19.59	-417.79	0.10	374.95
	$x_t^{\text{LowFee-DEX}}$	42975	0.07	5.51	-150.28	0.00	225.81
	$x_t^{\text{MidFee-DEX}}$	42975	0.02	12.78	-417.79	0.00	192.39
	$x_t^{\text{HighFee-DEX}}$	42975	0.05	13.75	-221.06	0.00	374.95

flows on Uniswap versus Binance varies significantly across token pairs. For ETH-USDT, the trade flow magnitude is much larger on Binance than on Uniswap. For instance, the standard deviation of the per-block-time trade flow of ETH-USDT on Binance is approximately 220 ETH, which is more than five times larger than the standard deviation of about 40 ETH on Uniswap.

In contrast, for the rest of the token pairs, the absolute trade flow is larger on Uniswap than on Binance. For example, the standard deviation of per-block-time trade flow of WBTC-ETH is about 56 ETH on Uniswap compared with 10 ETH on Binance. Second, for all token pairs on Uniswap, trade flows with high fees are larger in magnitude than flows with middle and low gas fees. For example, the standard deviation of ETH-USDT high-fee trade flow is 33.18 ETH, which is more

Table 5. Permanent price impact of DEX trade flows with different priority fee levels. This table reports the permanent price impacts of the CEX trade flow and DEX trade flows with high, medium, and low priority fee levels. Permanent price impacts are defined as the cumulative impulse responses of the CEX return to trade flows in the structural VAR model (see Equation 3). Upper bounds (UB) and lower bounds (LB) are obtained by considering all possible sequences of the recursive contemporaneous causality among the trade flows. The last column reports the difference between the lower bound of the permanent price impact of the high-fee DEX trade flow and the upper bound of the permanent price impact of the low-fee DEX trade flow. The estimation of the structural VAR is done for each pair-day and statistical inference is based on variations across the pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis points. DEX trade flows are standardized and thus in their standard deviations. *, ** and *** indicate significance levels at 1%, 5% and 10% respectively.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDT	0.81***	1.06***	−0.06	−0.04	0.25***	0.29***	2.45***	2.66***	2.49***
	(0.16)	(0.18)	(0.07)	(0.08)	(0.08)	(0.09)	(0.32)	(0.34)	(0.34)
LINK-ETH	1.47***	1.57***	0.45	0.51	1.48***	1.73***	4.58***	4.97***	4.07***
	(0.41)	(0.45)	(0.37)	(0.36)	(0.48)	(0.52)	(0.64)	(0.7)	(0.65)
WBTC-ETH	3.59***	4.58***	0.03	0.23	1.22***	1.88***	5.7***	6.87***	5.47***
	(0.61)	(0.79)	(0.29)	(0.39)	(0.51)	(0.73)	(0.63)	(0.84)	(0.78)
AAVE-ETH	1.62	2.06	1.57	2.06**	3.89***	4.57***	7.34***	8.49***	5.28***
	(1.1)	(1.3)	(1.14)	(1.02)	(0.88)	(0.88)	(0.91)	(1.06)	(1.31)
Pooled	1.82***	2.02***	0.5*	0.62**	1.71***	2.03***	4.86***	5.27***	4.23***
	(0.38)	(0.33)	(0.31)	(0.28)	(0.28)	(0.31)	(0.34)	(0.38)	(0.41)

than three times larger than that of low-fee trade flow.

In a structural VAR framework, the permanent price impact of a particular trade flow is estimated by the cumulative impulse responses of return to its unexpected component, as specified in Equation 3. In Table 5, we report the cumulative impulse responses of CEX return to DEX trade flows with different fee levels. In addition, to account for contemporaneous relations between the endogenous variables, we report the upper and lower bound of each variable's permanent price impact based on estimates across all possible orderings of the Cholesky decomposition.

We find that DEX trade flows with medium and high fees have positive and statistically significant permanent price impacts, hence being informed. There are several reasons why informed traders prefer executing on DEXs. One reason is related to transaction costs, which, as shown in Barbon and Ranaldo (2022), are comparable between centralized and decentralized exchanges. Another reason is the exposure to counterparty risk in CEXs, as evidenced by the recent example of FTX, where traders give custody of their tokens. Additionally, on CEXs, fast traders must invest ex-ante in speed technologies such as real-time market data and co-location services to gain priority. In contrast, DEXs allow informed traders to jump the queue by simply bidding a higher fee, effectively making an ex-post, order-by-order speed investment (Brolley and Zoican 2023).[18]

Our findings indicate that high-fee DEX trade flows have a much higher permanent price impact compared to low-fee transactions. As outlined in the implementation details of the structural VAR model, we analyze time spans of 5, 10, and 20 blocks. This choice of information horizon provides informed traders with the option to opt for lower fees in order to hide their information. However, our results suggest otherwise. Averaging across token pairs, the permanent price impact of the high-fee DEX trade flow, $x_t^{\text{HighFee-DEX}}$, has a lower and upper bound of 4.86 and 5.27 basis points respectively, meaning that a one standard deviation positive shock to the high-fee DEX trade flow leads to a permanent increase of CEX prices between 4.86 and 5.27 basis points. In contrast, the permanent price impact of low-fee DEX trade flow, $x_t^{\text{LowFee-DEX}}$, ranges between 0.5 and 0.62 basis points. Thus, a one standard deviation positive shock to the low-fee DEX trade flow only leads to a permanent increase in CEX prices between 0.5 and 0.62 basis points. To make a conservative comparison between the permanent price impacts of high-fee and low-fee DEX trade flows, we calculate the difference between the lower bound of the former and the upper bound of the latter. This results in a value of 4.23 basis points (see the last column, $\Delta^{\text{HighFee}-\text{LowFee}}$).

A sample *t*-test shows that the positive difference is statistically significant. Notably, this difference is also economically significant, being about eight times larger than the permanent price impact of the low-fee DEX trade flow. In Appendix C, we analyze the speed at which the CEX price adjusts to shocks in high-fee DEX trade flows. The impulse response function plots in Figure A1 demonstrate that the CEX price adjusts quickly to these shocks, with most of the cumulative price impact generated by DEX trade flows being realized within the subsequent block time.

[18] Observe that our findings do not imply that informed traders prefer DEXs over CEXs. The estimates in Table 5 indicate that the permanent price impacts of CEX flows are also statistically and economically significant. Depending on token pairs and market conditions, Binance may offer deeper liquidity, better transaction costs, and greater anonymity.

5.3 Do informed traders bid high fees or uninformed traders cluster at low fees?

In previous sections, we demonstrated that high-fee DEX trades contain more private information than low-fee trades, as evidenced by both the reduced-form and permanent price impact measures derived from a structural VAR model. This suggests that informed traders, relative to uninformed ones, are more likely to pay higher fees to execute their orders. However, it remains unclear whether informed traders consistently bid high fees on an absolute basis. Specifically, the empirical pattern of more private information and larger permanent price impacts in high-fee trades could stem from two potential scenarios:

- **Scenario 1**: Informed traders bid high fees while uninformed traders are uniformly distributed across fees.
- **Scenario 2**: Informed traders distribute their fee bids evenly across all levels, while uninformed traders tend to concentrate their bids at lower fees.

To evaluate the two scenarios, we identify informed traders by analyzing the distribution of bid fees. We compile all trades associated with each submission address and calculate the volume-weighted reduced-form price impact measure, as detailed in Section 5.1. An address is classified as informed if it satisfies the following criteria:

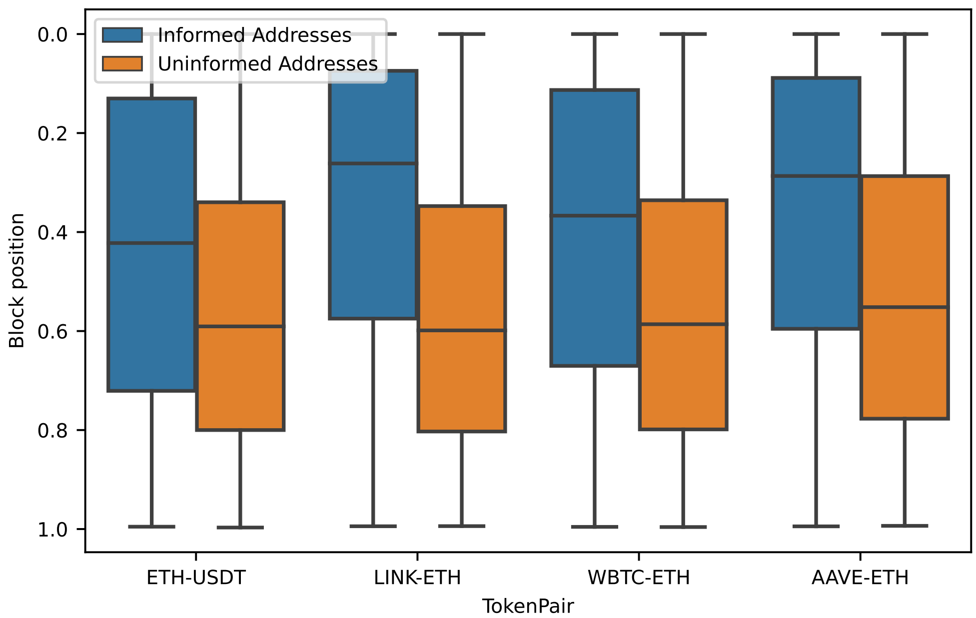
1. The number of trades is within the top 25% quantile of all addresses.
2. The volume-weighted price impact is within the top 25% quantile of all addresses.

These criteria ensure that informed addresses are relatively active in trading and have a significant, positive price impact.

To investigate whether informed and uninformed traders bid high or low fees, we analyze the relative block position of their trades. For example, if a trade occupies the fifth spot in a block containing 100 transactions, its relative block position is calculated as $5/100 = 0.05$. We opt for relative block position over absolute fees because it accounts for variations in blockchain congestion. Ultimately, it is the trade's position within the block that determines its execution priority.

Figure 2 displays the block positions of trades from informed and uninformed addresses. The data indicates that trades from informed addresses are primarily located at the front of the block, suggesting higher fee bids to expedite order execution. In contrast, trades from uninformed addresses are more evenly spread across the block. This pattern supports the first scenario that informed traders are more likely to bid higher fees to secure faster transaction processing.

Figure 2. Relative block position of trades from informed addresses and uninformed addresses. This figure plots the relative block positions of trades submitted by informed and uninformed traders. For instance, a trade at the fifth position in a block of 100 transactions would have a relative block position of $5/100 = 0.05$. An address qualifies as belonging to an informed trader if it falls within the top 25% quantile for both the number of trades and its volume-weighted price impact compared to all addresses.



5.4 Robustness to public information, trade size and private pools

We assess the robustness of our key findings to confirm that they are not determined solely by (i) arbitrage trades based on public information, (ii) the size of the trades, and (iii) they remain valid with the inclusion of private trading pools.

5.4.1 Public information arbitrage trades

Recall from Section 4 that we do not allow DEX trade flows to cause price changes during the contemporaneous block. This helps avoid wrongly attributing price changes caused by public information, for example, through market makers adjusting their quotes on CEXs in response to news, to DEX trade flows and thus overestimating the private information content of the DEX trade flows. Nevertheless, there might be some edge cases when market makers did not update their quotes in time within the contemporaneous block period, but only managed to do so in the next blocks. This delay can be attributed to imperfect monitoring technology employed by market makers on CEXs. For instance, their systems may not promptly process publicly disclosed news or new orders submitted by informed traders to the public mempool of the blockchain. Under these circumstances, price changes caused by public information might only realize in subsequent periods.

To rule out the possibility that high-fee bidding order flows are only driven by public information arbitrage competition, we identify two most popular types of public information arbitrage trades, namely CEX-DEX arbitrage (e.g. Capponi and Jia 2021 and Milionis, Moallemi, Roughgarden, and Zhang 2022), and sandwich attacks (Park 2023). Then we exclude them and re-estimate the permanent price impacts of the trade flows from the structural VAR model.

We identify sandwich-attack trades using the the heuristics outlined by Wang, Zuest, Yao, Lu, and Wattenhofer (2022), see also Appendix D.1 for more details. We classify trades as CEX-DEX arbitrage trades if they pass the criteria described in Appendix D.2. The estimates reported in Table 6 show that the aggregate price impact estimates obtained by excluding public information arbitrage trades are comparable to those in Table 5 when we include all trades. In particular, the differential price impact between high and low fee order flows is statistically and economically significant.

In Appendix E, we use the variation in blockchain congestion to further support our proposed

Table 6. Permanent price impact of DEX trade flows: public information arbitrage trades excluded. We remove trades from two popular arbitrage trades: sandwich-attach trades and CEX-DEX arbitrage trades. The heuristics used to identify these two types of trades can be found in Section D.1 and Section D.2 respectively.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDT	0.96***	1.17***	−0.05	−0.05	0.17**	0.2***	2.44***	2.52***	2.49***
	(0.13)	(0.15)	(0.05)	(0.05)	(0.08)	(0.08)	(0.29)	(0.3)	(0.3)
LINK-ETH	0.63**	0.76***	0.51***	0.56***	1.9***	1.94***	5.21***	5.23***	4.65***
	(0.31)	(0.31)	(0.22)	(0.22)	(0.32)	(0.32)	(0.53)	(0.53)	(0.55)
WBTC-ETH	3.28***	3.62***	0.31	0.39*	1.69***	1.87***	4.89***	5.07***	4.5***
	(0.46)	(0.52)	(0.22)	(0.22)	(0.3)	(0.32)	(0.52)	(0.52)	(0.59)
AAVE-ETH	1.21	1.56**	1.88***	2.0***	5.5***	5.87***	8.3***	8.49***	6.3***
	(0.77)	(0.75)	(0.7)	(0.7)	(0.76)	(0.75)	(0.72)	(0.73)	(0.94)
Pooled	1.49***	1.75***	0.67***	0.73***	2.33***	2.45***	5.26***	5.27***	4.53***
	(0.24)	(0.25)	(0.2)	(0.2)	(0.24)	(0.25)	(0.29)	(0.29)	(0.33)

channel of private information trading against the channel of public information arbitrage.

In scenarios of public information arbitrage, where two traders have symmetric information, they engage in Bertrand competition, bidding up to the arbitrage profit. This bidding behavior remains unaffected by blockchain congestion levels, as the profit margin significantly exceeds the cost of increased fees (marginal gas).

Conversely, in private information trading scenarios, each trader bids to secure a trading advantage based on their unique information. If competing against an uninformed trader, the informed trader needs only to outbid the marginal gas fee, which varies with blockchain congestion. Thus, the equilibrium fee that each trader bids tends to be a weighted average of the winning fee and the marginal fee, increasing with higher congestion.

Empirical analysis as reported by Table A3 supports this theory; on days with higher median gas fees (classified as high-gas days), the price impact difference between high-fee and low-fee DEX trades is more pronounced. This suggests that fees bid by informed traders rise with blockchain congestion, aligning with the private information competition channel rather than the public information arbitrage channel. A toy model to formally illustrate the two channels described and the detailed empirical results can be found in Section E

5.4.2 Trade size

Priority fee is a fixed cost regardless of the trade size. Consequently, traders are more willing to pay a higher priority fee for larger trades, as it is relatively cheaper on a per-unit basis. Therefore, trade size and priority fee are positively correlated. In addition, it is a well-known fact that large trades tend to have a larger price impact than small trades (Easley and O'Hara 1987). Thus, trade size has a potential confounding effect on the price impact of fees.

To account for the potential confounding effect of trade size, we fit a step-wise function between the gas price and trade size and then use the relative gas price as the classification variable. Specifically, we first, for each token pair, sort all trades into ten equally-sized buckets based on their size. Then for each size bucket, we calculate the relative gas price—gas price deviation relative to the bucket median. Last, we classify trades into "high-fee", "mid-fee", and "low-fee" based on the relative gas price, instead of the original gas price. The relative gas price essentially measures whether, and to what extent, a trader over- or under-bid for her transaction of a particular size. In such way, we control the effect of the trade size.

Table 7 presents the results of the permanent price impact of trade flows at different relative gas price levels. The findings indicate that, even after controlling for trade size, high-fee DEX trade flows have a larger permanent price impact than low-fee flows.

Table 7. Permanent price impact of DEX trade flow: fee level classification based on relative gas controlling for trade size.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDT	0.93***	1.14***	0.58***	0.64***	1.07***	1.15***	2.14***	2.25***	1.5***
	(0.13)	(0.15)	(0.09)	(0.09)	(0.15)	(0.15)	(0.25)	(0.26)	(0.2)
LINK-ETH	0.56**	0.7***	1.86***	2.0***	3.2***	3.34***	4.14***	4.14***	2.13***
	(0.27)	(0.27)	(0.27)	(0.28)	(0.41)	(0.4)	(0.45)	(0.45)	(0.42)
WBTC-ETH	3.39***	3.75***	1.5***	1.63***	3.08***	3.34***	3.5***	3.67***	1.86***
	(0.43)	(0.47)	(0.27)	(0.28)	(0.43)	(0.43)	(0.43)	(0.43)	(0.46)
AAVE-ETH	1.36**	1.64***	4.96***	5.73***	4.8***	5.95***	6.81***	7.33***	1.08
	(0.67)	(0.67)	(0.73)	(0.74)	(0.73)	(0.73)	(0.61)	(0.62)	(1.0)
Pooled	1.53***	1.76***	2.25***	2.48***	3.11***	3.36***	4.22***	4.28***	1.73***
	(0.22)	(0.22)	(0.23)	(0.23)	(0.25)	(0.25)	(0.24)	(0.25)	(0.3)

5.4.3 Private pool transactions

Trading on DEXs has undergone several infrastructural changes since its inception. One of the most significant changes was the introduction of Flashbots in early 2021. Flashbots enable traders to send orders directly to validators through a private pool, thus bypassing the public mempools.[19] In these private pools, order characteristics such as direction, size, and priority fee bid are not observable, leading traders to compete in a first-price sealed bid auction mechanism. In contrast, before the introduction of Flashbots, traders could only broadcast orders to the public peer-to-peer network, making the characteristics of pending orders, including direction, size, and priority fee bid, publicly observable.[20]

The introduction of private pools might lead to a segmentation of informed and uninformed order flows. Before private pools, informed traders with short-lived information traded in a fully transparent public mempool, where their trading intentions, expressed through trade size and gas fees, were observable by back-runners on DEXs. This visibility often led to competition and high fee bids to prioritize execution. With the advent of private pools, these informed traders might shift their activity there. However, competition still exists in private pools, particularly among informed traders with similar information. Therefore, it is ex-ante unclear whether the introduction of private pools fundamentally changes the trading dynamics in public pools.

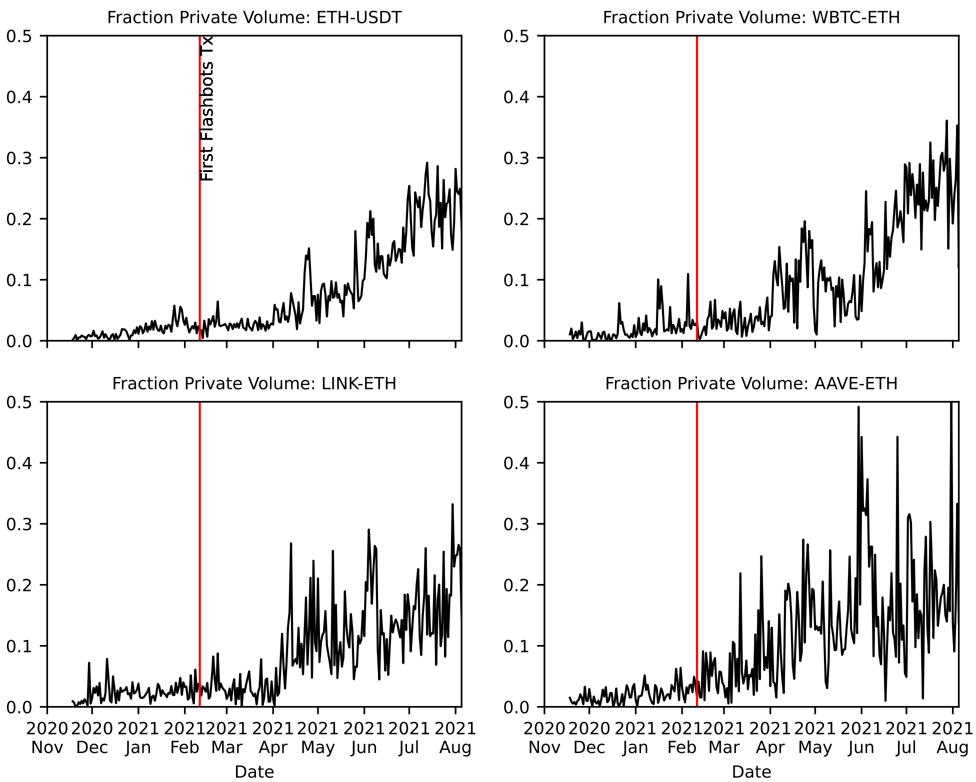
To identify private pool transactions, we use the following criterion: when traders use the private relay to send their orders directly to validators, their orders do not appear in the public mempools. Therefore, we analyze public mempool data to identify transactions that were likely executed through private pools. Specifically, for an executed Uniswap trade, we attempt to find a

[19]After the Proof-of-Stake (PoS) transition of Ethereum, the orders are sent to block builders instead of proposers directly.

[20]Note that these private submission pools are different from dark pools in the equities market. The latter refers to marketplaces where orders are not publicly displayed. In other words, there is no pre-trade transparency. However, private pools instead refer to private mempools of miners to which traders can submit their orders. Liquidity on DEXs is always publicly visible and there is pre-trade transparency. A closer example in traditional markets is the information leakage risk faced by brokers' clients. For example, Maggio, Franzoni, Kermani, and Sommavilla (2019) provide evidence that brokers leak informed traders' information to their best clients. Orders submitted to the public mempool potentially suffer from such information leakage or outright front-running problem. So traders might choose to submit to private pools instead.

matching pending mempool order with the same hash. If no such match is found, it indicates that the order never appeared in the public mempool and was likely submitted through the private relay. Figure 3 plots the fraction of private trades following the identification method above. It shows that the fraction of private trade volume picked up gradually after the Flashbots introduction. Towards the end of 2021, the fraction stays around 20%.

Figure 3. Fraction of private pool trade volume. This figure plots the fraction of private pool trades during the sample period between November 18, 2020, and August 4, 2021, for each of the four sample token pairs.



To determine whether the finding that DEX trades with higher priority fees are more informative remains robust under the new market structure with private pools, we re-estimate the SVAR model for the period from February 10, 2021, to August 4, 2021, when trading volume via private pools increased significantly. Table 8 presents the permanent price impacts of private trade flow and public trade flows of different fee levels. There are two main observations. First, while private trade flows have a positive permanent price impact, with the upper bound being statistically significant, their magnitude is much smaller than that of public, high-fee trade flows. This suggests that although some informed trades have moved to private pools, the majority remain in the public mempool. Second, even with a significant volume migrating to private pools, our key finding from before the introduction of private pools still holds: public, high-fee DEX trade flows are much more informative than public, low-fee trade flows. This demonstrates that the public mempool remains an important venue for price discovery, driven by the mechanism of priority fee bidding.

Table 8. Price impacts of DEX public trade flows with different priority fee levels and private trade flow: with private pools. This table reports the permanent price impacts of public DEX trade flows with high, medium, and low priority fee levels, and private DEX trade flow. As before, we include CEX trade flow as a control but it is omitted for space considerations. The sample period is between February 10, 2020 and August 4, 2021. Permanent price impacts are defined as the cumulative impulse responses of the CEX return to DEX trade flow in the structural VAR model (see Equation 3). Upper bounds (UB) and lower bounds (LB) are obtained by considering all possible sequences of the recursive contemporaneous causality among the endogenous variables. The last column reports the difference between the lower bound of the permanent price impact of the public, high-fee DEX trade flow and the upper bound of the permanent price impact of the public, low-fee DEX trade flow. The estimation of the structural VAR is done for each pair-day and statistical inference is based on variations across the pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis points. DEX trade flows are standardized and thus in their standard deviations. *, ** and *** indicate significance levels at 1%, 5% and 10% respectively.

Variable	$x^{\text{Lit-LF-DEX}}$		$x^{\text{Lit-MF-DEX}}$		$x^{\text{Lit-HF-DEX}}$		$x^{\text{Dark-DEX}}$		$\Delta^{\text{Lit-HF - Lit-LF}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDT	0.0	0.04	0.31***	0.36***	1.63***	1.79***	0.52***	0.64***	1.59***
	(0.05)	(0.04)	(0.07)	(0.08)	(0.11)	(0.13)	(0.07)	(0.08)	(0.12)
LINK-ETH	0.35	0.54*	0.69***	0.92***	1.71***	2.02***	0.33	0.64**	1.17***
	(0.32)	(0.33)	(0.28)	(0.27)	(0.33)	(0.37)	(0.32)	(0.3)	(0.46)
WBTC-ETH	0.65***	1.08***	1.25***	2.36***	2.41***	3.42***	0.47*	0.9***	1.33**
	(0.25)	(0.39)	(0.24)	(0.79)	(0.37)	(0.83)	(0.24)	(0.36)	(0.62)
AAVE-ETH	0.31	2.29	4.17***	8.58**	4.32***	5.94***	0.32	3.69***	2.03
	(0.94)	(1.93)	(1.15)	(4.32)	(1.35)	(1.18)	(1.76)	(0.71)	(2.87)
Pooled	0.35*	0.84**	1.66***	2.52***	2.43***	2.92***	0.49	1.16***	1.58***
	(0.21)	(0.41)	(0.31)	(0.89)	(0.31)	(0.37)	(0.39)	(0.18)	(0.61)

6 Why Do Informed Traders Bid High Fees?

In this section, we investigate the motives behind the high-fee bidding strategy. Informed traders in the public mempool face the trade-off between faster execution speed and higher information leakage risk. Our empirical results suggest that the benefit of faster execution speed outweighs the cost of information leakage.

6.1 Can high fees be explained by blockchain congestion?

Trading on DEXs is not the only activity on a blockchain. Other activities, such as payment transfers, borrowing and lending, non-fungible token (NFT) auctions, and initial coin offerings (ICOs), also occupy block space. Notably, if there is a surge in non-DEX activities leading to block congestion, the marginal priority fee required to execute a transaction increases, thereby driving up transaction costs for DEX traders.

During such times, in contrast to a patient and uninformed trader, a trader who possesses short-lived private information, e.g, over the next several blocks, might bid a high fee to avoid execution risk if the gain from her trade is large.[21] Ideally, she would like to set her bid to the marginal fee to guarantee execution in the next block. However, the marginal fee of the next block is not perfectly predictable. For example, even if the informed trader actively monitors all pending orders received by its mempool, due to network latency, pending orders seen by her can be different from the ones seen by the validators. As a result, she will bid a fee higher than the expected marginal fee to reduce her execution risk.

This implies that if an informed trader only faces execution risk, she will choose a high, but not excessively high, blockchain priority fee for her trades compared to other transactions in the same block. Consequently, her trades will likely be positioned around the middle of the block, rather than at the very top. By bidding a medium fee, it becomes difficult for other informed traders to detect her presence, thereby keeping the information leakage risk low.

An informed trader will bid a high fee if the blockchain network is congested. However, congestion is not the only factor; she may also bid a high fee when facing competition from other traders. This competition arises when private information is shared among multiple informed traders who receive either the same or highly correlated private signals (See, e.g., Holden and Subrahmanyam 1992; Foster and Viswanathan 1996; Back, Cao, and Willard 2000). When competing

[21] We note that impatient and uninformed traders (e.g., liquidity traders who receive marginal calls and have to liquidate their positions) can bid high priority fees to avoid execution risk as well. However, their trades contain no private information and thus can not drive our findings in the above section that high-fee trades are more informative.

with other traders holding similar information, an informed trader might have to bid a significantly higher priority fee than other non-DEX transactions in the same block, especially when the potential profit from the information is substantial. In such cases, we might observe DEX trades with excessively high fees positioned at the very top of the block.

Identify "excessively-high-fee trades" As explained above, competition among informed traders can lead to excessively high priority fees for DEX trades compared with other non-DEX transactions executed in the same block. How high a fee needs to be in order to be regarded as "excessive"? To choose the right threshold for the priority fee, we use the inter-quartile range (IQR) method, a commonly used outlier detection approach in statistics.[22] Specifically, for each block, we first calculate the 25% quantile (Q25) and 75% quantile (Q75) of the priority fees of all executed transactions in the block[23], including both DEX trades and non-DEX transactions. Then we calculate the IQR, defined as the difference between the 75% quantile and 25% quantile, that is, $\text{IQR} = \text{Q75} - \text{Q25}$. Finally, we obtain the threshold $\text{Q75} + 1.5 \times \text{IQR}$ and label DEX trades with a priority fee higher than the threshold as "excessively-high-fee trades".[24]

Information content of "excessively-high-fee trades" Note that DEX trades with excessively high fees or located at the very top of the block can include three different types of trades: (1) trades driven by competition among privately informed traders; (2) trades driven by competition among arbitrageurs on public information (e.g., price discrepancies between CEXs and DEXs); (3) trades by impatient and uninformed traders (e.g., liquidation trades triggered by marginal calls). However, only the first type of trades, which are driven by competition among privately informed traders, contain private information and thus can have permanent price impacts on the CEX returns.

[22]We prefer the IQR method, a quantile-based approach, over other outlier detection methods based on standard deviations as the priority fee distribution is not normal but right-skewed.

[23]We obtain the executed transactions data on the Ethereum blockchain from Blockchair (https://gz.blockchair.com/ethereum/transactions/).

[24]Alternatively, one can identify such trades based on their block position. As transactions executed in the same block are ranked based on their priority fees in descending order. Thus, transactions with higher priority fees will be placed more at the front of the block. Specifically, one can choose a threshold for the block position, say top 10%, and then label DEX trades located more before the threshold. We tested the alternative approach and the results are qualitatively the same.

Table 9. Permanent price impact (PPI) of DEX trade flows with different priority fee levels: Excluding "excessively-high-fee trades". This table reports the permanent price impacts of the DEX trade flows with high, medium, and low priority fee levels. Permanent price impacts are defined as the cumulative impulse responses of the CEX return to DEX trade flow in the structural VAR model (see Equation 3). Upper bounds (UB) and lower bounds (LB) are obtained by considering all possible sequences of the recursive contemporaneous causality among the endogenous variables. The last column reports the difference between the lower bound of the permanent price impact of the high-fee DEX trade flow and the lower bound of the permanent price impact of the low-fee DEX trade flow. The estimation of the structural VAR is done for each pair-day and statistical inference is based on variations in the pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis points. DEX trade flows are standardized and thus in their standard deviations. *, ** and *** indicate significance levels at 1%, 5% and 10% respectively.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDT	1.04^{***}	1.15^{***}	-0.06	-0.05	0.34^{***}	0.38^{***}	1.13^{***}	1.24^{***}	1.19^{***}
	(0.14)	(0.15)	(0.06)	(0.07)	(0.08)	(0.09)	(0.15)	(0.16)	(0.17)
LINK-ETH	0.41	0.48^{*}	0.52^{**}	0.57^{**}	2.03^{***}	2.17^{***}	2.74^{***}	2.85^{***}	2.17^{***}
	(0.27)	(0.28)	(0.25)	(0.25)	(0.3)	(0.31)	(0.33)	(0.35)	(0.37)
WBTC-ETH	2.93^{***}	3.84^{***}	0.48^{*}	0.51^{*}	1.77^{***}	2.26^{***}	3.39^{***}	4.31^{***}	2.88^{***}
	(0.39)	(0.53)	(0.24)	(0.26)	(0.31)	(0.42)	(0.43)	(0.52)	(0.56)
AAVE-ETH	1.62^{**}	1.87^{**}	1.36^{*}	1.61^{***}	6.31^{***}	6.84^{***}	5.93^{***}	6.31^{***}	4.32^{***}
	(0.81)	(0.84)	(0.7)	(0.69)	(0.69)	(0.71)	(0.7)	(0.71)	(0.94)
Pooled	1.47^{***}	1.74^{***}	0.56^{***}	0.64^{***}	2.58^{***}	2.83^{***}	3.29^{***}	3.57^{***}	2.65^{***}
	(0.24)	(0.26)	(0.19)	(0.19)	(0.23)	(0.25)	(0.24)	(0.26)	(0.29)

To determine if our identified trades include those driven by private information, we reconstruct DEX trade flows with different priority fee levels excluding all "excessively-high-fee" trades. We then re-implement the structural VAR analysis. The rationale is that if a significant portion of high-fee trades results from competition among privately informed traders, we should observe a notable reduction in their permanent price impact after excluding the "excessively-high-fee" trades.

Table 9 reports the permanent price impacts of DEX trade flows when "excessively-high-fee trades" are excluded. Compared to the baseline results in Table 5, where all trades are included, the lower and upper bounds of the permanent price impact of high-fee DEX trade flows drop significantly from 4.86 and 5.27 basis points to 3.29 and 3.57 basis points, respectively. These results indicate that our key finding—high-fee DEX trade flow is more privately informed—is at least partially driven by competition among privately informed traders, in addition to avoiding execution risk.

6.2 Explaining high fee through jump bidding

The analyses above show that competition among privately informed traders on DEXs leads them to bid excessively high priority fees. However, high fees increase the risk of information leakage, particularly given the public nature of mempools. For instance, "back-runners" can exploit this information (Yang and Zhu 2020) or "predators" (Brunnermeier 2005), despite lacking private signals, can infer them from public signals such as fees. What is the rationale for informed traders to bid high fees? Below, we dive deeper into their bidding strategy to answer this question.

Identify trades from priority gas auctions (PGAs) As pending orders in the mempools are publicly visible to all traders who actively monitor them, one natural bidding strategy for informed traders is to competitively bid up their priority fees, a process known as the priority gas auction (PGA) in the literature (Daian et al. 2020). But is it the dominant bidding strategy?

For an executed trade to quality as a PGA trade, we require the following criteria:

1. **The executed trade has at least one matched mempool order with the same submission address and nonce.** Recall that a trader on DEX needs to attach a number called “nonce” to each of her orders. The most important property of a nonce is that each number can only be used once and it must be used in a consecutively increasing order. For example, a new order broadcast by a trader needs to have a new nonce increased by 1 compared with the previous order. More importantly, a trader’s order with a larger nonce cannot be executed before one with a smaller nonce. This implies that if a trader wants to modify her pending order, e.g., increase the fee, she needs to broadcast a new order with the same nonce as the pending one. Hence, the first criterion on submission address and nonce guarantees that the matched mempool orders are previous revisions of the final executed order.

2. **The gas price of the executed trade must be higher than that of its matched order(s).** We observe the gas price attached to both mempool orders and the executed trade. The second criterion requires that the executed trade must have a higher gas fee than its matched

order(s) (i.e., those with the same submission address and nonce) so that we capture trades associated with fee competition.

3. **All matched orders of the executed trade must arrive at the mempool within five blocks.** Specifically, to be matched with a trade executed at block time t, orders must arrive in the mempool during the block time interval of $(t-5, t]$. We believe gas bidding due to competition should happen within a fairly short time window. If the window is too long, the bid update is more likely to result from patient liquidity traders revising their fees to reduce the waiting time.

Fraction of PGA trades We implement the foregoing identification strategy above and Table 10 reports, for each token pair, the fraction of PGA trades for both trades with excessively high fees ("excessively-high-fee trades") and other trades ("other trades"). The overall fraction of executed trades identified as PGA trades is very small. For instance, for the group of "other trades", less than 5% are identified as PGA trades across the six token pairs.

Table 10. Percentages of priority gas auction (PGA) trades. This table shows the fraction of trades identified as priority gas auction (PGA) trades, for "excessively-high-fee trades" and other trades.

TokenPair	ExplicitCompetition ExcessiveGas	Non-PGA trades	PGA trades
ETH-USDT	Other trades	97.68	2.32
	Excessively-high-fee trades	87.95	12.05
LINK-ETH	Other trades	95.06	4.94
	Excessively-high-fee trades	73.29	26.71
WBTC-ETH	Other trades	96.61	3.39
	Excessively-high-fee trades	84.79	15.21
AAVE-ETH	Other trades	95.39	4.61
	Excessively-high-fee trades	81.72	18.28

Surprisingly, when we examine the "excessively-high-fee trades," which likely include trades influenced by competition, we find that only a small portion of these trades are classified as PGA trades. The percentage of PGA trades within the category of "excessively-high-fee trades" varies across the four token pairs, ranging from 12.05% for ETH-USDT to 26.71% for LINK-ETH. These findings indicate that the PGA bidding strategy is not the prevailing choice among informed traders.

Instead of engaging in competitive fee bidding, they opt for initiating trades with a significantly high fee, reminiscent of the jump bidding strategy observed in auction theory (Daniel and Hirshleifer 1998; Avery 1998).

The motivation for adopting such a bidding strategy is twofold. First, by bidding a high fee initially, an informed trader can signal that her valuation of the information is high. If bidding is costly, this signal can deter potential competitors from participating. Second, even if all traders value the information equally and there is no bidding cost, it remains optimal for others to drop out, as winning over an aggressive bid from the jump bidder subjects them to a greater Winner's Curse.

7 Conclusion

In this paper, we investigate the trading strategies of informed traders and their implications for price discovery on decentralized exchanges (DEXs). DEXs execute orders in batches, with execution priority determined by the priority fee bid by traders. Our evidence indicates that this result is driven by informed traders bidding high fees to execute their orders, rather than uninformed traders naturally bidding low fees. Additionally, we offer a plausible explanation for this high-fee bidding behavior using a unique dataset of Ethereum mempool orders. Our findings suggest that informed traders bid high fees as part of a jump bidding strategy: by placing a high initial bid, they signal their high valuation of the information, thereby discouraging competition from other potential bidders.

It is important to note that, while we have demonstrated the benefits of the priority fee bidding mechanism in DEXs for enhancing price discovery, we cannot unequivocally assert its positive impact on overall market quality. Faster information revelation may increase the adverse selection risk faced by liquidity providers, potentially leading to reduced liquidity provision. Future research should consider both price discovery and market liquidity to comprehensively assess the impact of the priority fee bidding mechanism on the overall market quality of DEXs.

A Detailed Derivation of Permanent Price Impact Measure

Below, we provide more details about the derivation of the two trade informativeness measures from the structural VAR model. After estimating the structural VAR model, we can easily obtain the vector moving average (VMA) representation with structural innovations as follows:

$$y_t = \Theta(L)\varepsilon_t = \Theta_0\varepsilon_t + \Theta_1\varepsilon_{t-1} + \Theta_2\varepsilon_{t-2} + \cdots, \tag{5}$$

where $\Theta(L)$ is the polynomial of the lag operator $\Theta(L) = \Theta_0 + \Theta_1 L + \Theta_2 L^2 + \cdots$ and $\Theta_0, \Theta_1, \ldots$ are the VMA system matrices. Then the permanent price impact (PPI) of a trade flow variable k is defined as the cumulative impulse response of the midquote return to a unit shock in the trade flow, that is,

$$\mathrm{PPI}_k = \frac{\sum_{j=0}^{\infty} \partial r_{t+j}^{\mathrm{CEX}}}{\partial \varepsilon_{k,t}} = [\Theta(1)]_{1,k}, \quad k > 1 \tag{6}$$

where $[\Theta(1)]_{1,k}$ denotes the $(1,k)$-*th* entry of $\Theta(1) = \Theta_0 + \Theta_1 + \Theta_2 + \ldots$.

One challenge of measuring the private information content of trades through price impact measures is that trades can move prices through two confounding effects: inventory control and asymmetric information effects. For example, upon executing buy trades, market makers can revise their quotes upwards either because they would like to induce future sell trades to revert their inventory back to the original level or because they learn positive private information from the buy trades. However, as pointed out in Hasbrouck (1991a), we can partially resolve these two effects if we measure the impact of trades on the security prices over a substantially long period. This allows to distinguish the transitory nature of inventory control effects, as market makers adjust their quotes back to pre-trade levels when their inventories revert back from private information. The latter is conveyed through trades, driven by asymmetric information, and permanently embedded in security prices.

B Summary Statistics of CEX and DEX Trading Volumes

Table A1 reports the summary statistics of the DEX trading volume.

Table A1. Summary statistics of the DEX trading volume by block time.

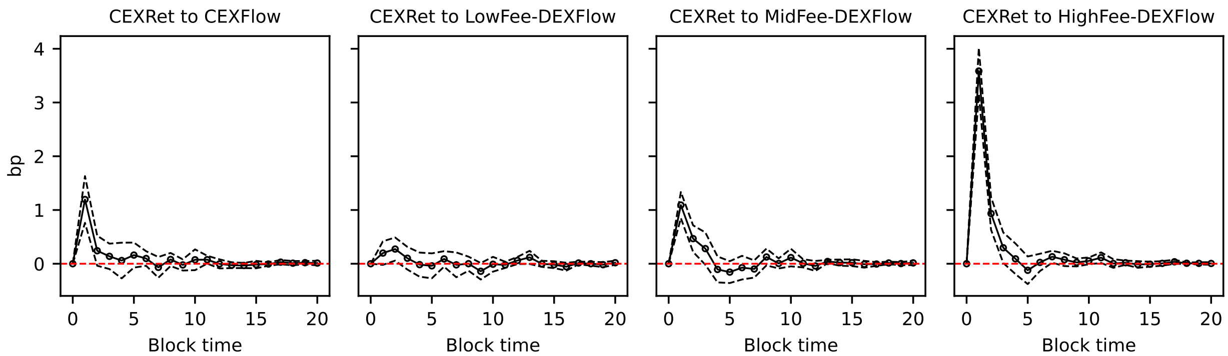
		N	Mean	SD	Min	50%	Max
ETH-USDT	$\|x_t\|^{CEX}$	358936	337.09	609.51	0.0	141.67	18755.65
	$\|x_t\|^{DEX}$	358936	16.91	65.67	0.0	3.30	5665.07
	$\|x_t\|^{HighGas\text{-}DEX}$	358936	8.90	42.62	0.0	0.00	4396.11
	$\|x_t\|^{MidGas\text{-}DEX}$	358936	6.50	45.45	0.0	0.62	5665.07
	$\|x_t\|^{LowGas\text{-}DEX}$	358936	1.51	17.29	0.0	0.00	3609.29
WBTC-ETH	$\|x_t\|^{CEX}$	41303	3.79	17.74	0.0	1.30	2278.09
	$\|x_t\|^{DEX}$	41303	32.32	89.35	0.0	6.00	4508.88
	$\|x_t\|^{HighGas\text{-}DEX}$	41303	14.65	58.31	0.0	0.00	4508.88
	$\|x_t\|^{MidGas\text{-}DEX}$	41303	14.36	63.28	0.0	0.17	2574.74
	$\|x_t\|^{LowGas\text{-}DEX}$	41303	3.30	29.49	0.0	0.00	3157.51
LINK-ETH	$\|x_t\|^{CEX}$	50844	6.77	22.20	0.0	1.50	2078.45
	$\|x_t\|^{DEX}$	50844	12.93	39.44	0.0	3.00	3156.18
	$\|x_t\|^{HighGas\text{-}DEX}$	50844	5.81	23.15	0.0	0.00	1739.54
	$\|x_t\|^{MidGas\text{-}DEX}$	50844	5.87	31.36	0.0	0.10	3052.18
	$\|x_t\|^{LowGas\text{-}DEX}$	50844	1.24	6.83	0.0	0.00	332.52
AAVE-ETH	$\|x_t\|^{CEX}$	33179	5.18	14.28	0.0	1.28	685.40
	$\|x_t\|^{DEX}$	33179	14.41	30.24	0.0	5.00	1033.87
	$\|x_t\|^{HighGas\text{-}DEX}$	33179	5.84	17.23	0.0	0.00	1033.87
	$\|x_t\|^{MidGas\text{-}DEX}$	33179	6.99	24.36	0.0	0.07	990.18
	$\|x_t\|^{LowGas\text{-}DEX}$	33179	1.58	10.28	0.0	0.00	980.44

C Speed of Price Adjustments

In Section 5.2, we have provided evidence that the high-fee DEX trade flow has a much larger permanent price impact than the low-fee DEX trade flow. However, as the permanent price impact is defined as the cumulative impulse responses of CEX return, it can not speak to the speed of price discovery. How quickly does the CEX price adjust to high-fee bidding order flow? To examine it, we turn to the dynamics of the impulse responses of the CEX return to a one-standard-deviation shock to the DEX trade flows in Figure A1.[25]

[25]To obtain conservative estimates for the return impulse responses to the DEX trade flows, we use the following sequence for the endogenous variables: CEX return, low-fee DEX trade flow, mid-fee DEX trade flow, and high-fee DEX trade flow. In other words, we impose the restriction that the CEX return contemporaneously causes the DEX trade flows and low-fee flow causes mid-fee and high-fee flows.

Figure A1. Impulse response functions of CEX return to DEX trade flows with different priority fee levels. This figure plots the impulse responses of the CEX return to a one-standard-deviation shock to CEX trade flow and high-fee, mid-fee, and low-fee DEX trade flows over the horizon of 20 blocks. Impulse responses are obtained by converting the estimated structural VAR model to its VMA form. To have conservative estimates of the price impacts of the high-fee DEX trade flow, we use the following sequence of variables: CEX return, CEX trade flow, low-fee DEX trade flow, mid-fee DEX trade flow, and high-fee DEX trade flow (see Equation 4) for the Cholesky decomposition. CEX return is measured in basis points and trade flows are standardized and thus measured in standard deviation units. We estimate the structural VAR model for each pair-day, and the statistical inference is based on variations across pair-day estimates. Dashed black lines represent 95% confidence bands.



It shows that the return impulse responses are significant and peak at the first subsequent period (block time $t = 1$). Note that given the specific recursive structure that we impose, the return impulse responses to all three DEX trade flows at the contemporaneous period ($t = 0$) are zero. In addition, the return impulse responses drop significantly from the second period (block time $t = 2$), especially to a shock in the high-fee DEX trade flow. It indicates that CEX return responds quickly to DEX trade flows and most of the cumulative price impact generated by DEX trade flows is realized within the subsequent block time.

Note that across all flows, there is no return response at block time $t = 0$. This is because we do not allow trade flows to contemporaneously impact returns in order to capture private information, as emphasized in Section 4. Nevertheless, we do observe short-term transitory price impacts followed by price reversals. For block time $t = 1$ and $t = 2$, there is positive price impact even for the low-fee flow, just its magnitude is smaller compared to that of middle fee and high-fee flow.

D Selection Criteria for Public Information Arbitrage Trades

D.1 Sandwich attack trades identification

We provide a brief review of procedure developed in Wang, Zuest, Yao, Lu, and Wattenhofer (2022) to identify sandwich attack trades: T_{A1} and T_{A2}.

1. T_{A1} and T_{A2} are included in the same block and in this order.

2. T_{A1} and T_{A2} have different transaction hashes

3. T_{A1} and T_{A2} swap assets in the same liquidity pool, but in opposite directions. The input amount for the swap in T_{A2} is equal to the output amount of the swap in T_{A1}.

4. Every transaction T_{A2} is mapped to exactly one transaction T_{A1}.

D.2 CEX-DEX arbitrage trades identification

We identify trades as CEX-DEX arbitrage trades if they pass the following four criteria following Heimbach, Pahari, and Schertenleib (2024):

1. The trade is a simple swap transaction and does not use more than 400,000 gas. In addition, the trade is not labeled as a sandwich attack (including the front-running trade, victim trade, and the back-running trade).

2. The swap executed by the transaction is the first swap of all swaps, if there are any, of the same direction within the same block. For example, if there are two buy trades, we only count the first one as a potential CEX-DEX arbitrage trade.

3. The direction of the swap is the opposite of the CEX-DEX price deviation.

4. The swap transaction does not fall within the bottom 25% of the block.

Table A2 provides the summary statistics of CEX-DEX arbitrage trades for the four token pairs considered in this paper.

Table A2. Occurrence of public-information arbitrage trades. This table reports, for each token pair, the average number of all trades, CEX-DEX arbitrage trades, and sandwich-attack trades per block. In addition, it reports the proportions of CEX-DEX arbitrage trades and sandwich-attack trades relative to the total number of trades. The sample period is between November 18, 2020, and February 15, 2021.

	All (#)	CEX-DEX arb (#)	Sandwich attack (#)	CEX-DEX arb (%)	Sandwich attack (%)
ETH-USDT	8560.58	860.58	95.8	10.05	1.12
WBTC-ETH	1371.25	118.49	5.47	8.64	0.4
LINK-ETH	1054.76	93.88	6.05	8.9	0.57
AAVE-ETH	610.93	62.86	2.2	10.29	0.36

E High-fee bidding and blockchain congestion

In Section 5.4, we have provided direct empirical evidence which rules out that high-fee bidding transactions are due to public information arbitrage competition. In this appendix section, we corroborate this result by providing indirect evidence in support of this claim.

We begin by constructing a theoretical gas fee bidding model, where bids account for the level of blockchain congestion and for the type of arbitrage competition, i.e., public or private arbitrage competition. Assume the existence of two potentially informed arbitrageurs, each of whom has a probability p of being informed. Denote by a the value of the arbitrage opportunity for the traders. Further, let $\kappa < a$ be the marginal bid from other traders whose objective is to bid just enough to secure a position in the block. The bid κ would grow with the congestion level of the blockchain. We distinguish two separate settings:

- **Public information arbitrage**. Under a setting of full transparency, the two arbitrageurs have symmetric information. Hence, they engage in a Bertrand competition and bid their true valuation a. As a result, the bid is independent of the blockchain congestion level.
- **Private information trading**. Under this setting, each informed arbitrageur chooses a gas

fee bid g to maximize his expected revenue given by

$$p(a-g)F(g)+(1-p)(a-g), \tag{7}$$

where $F(x)$ is the probability that the competing arbitrageur bids a fee smaller or equal than x. In the above expression, the first term is the product of the probability p that the other arbitrageur is informed and the expected arbitrageur's revenue conditioned on the event that the other arbitrageur is informed. The latter is given by the arbitrage value a net of the paid fee g multiplied by the probability that the competitor loses by bidding less than g. The second term is the product of the probability $1-p$ that the competing trader is uninformed and the expected arbitrageur's revenue is conditioned on the event that the other arbitrageur is uninformed. In such a case, the other trader makes the marginal bid g.

Observe that $F(\kappa)=0$ because if the arbitrageur offers less than the marginal bid, he would surely lose in the auction. Moreover, the other trader would never bid more than $pa+(1-p)\kappa$ because that would mean bidding more than the arbitrage value a if the other trade is informed, which never results in a profit. Hence, $F(pa+(1-p)\kappa=1$. Therefore, the optimal bid g is distributed in the interval $[\kappa, pa+(1-p)\kappa]$. To recover the profit-maximizing bidding distribution F, we take the first order condition in (7), leading to

$$-pF(g)+p(a-g)F'(g)-(1-p)=0.$$

The above expression is a first-order differential equation, whose solution can easily be verified to be

$$F(g)=\frac{(g-\kappa)(1-p)}{p(a-g)}$$

in the interval $[\kappa, pa+(1-p)\kappa]$. It can then be easily verified that the expected fee bid is increasing in the κ.

We next empirically verify whether bids vary with the gas fee. We classify days into high-gas, mid-gas, and low-gas days based on their median gas price. High-gas days are those with a median gas price above the 75%. Table A3 reports average price impacts of DEX trade flows of different fee levels for each day group. The table shows that the price impact difference between high-fee DEX trade and low-fee is higher in high-gas days. Take ETH-USDT as an example. The price impact of high-fee DEX trade is more than 10 times more than the low-fee for high-gas days (1.35/0.12) while it is only about 5 times for low-gas days (2.13/0.57). Using the above derived theoretical model, we conclude that these orders must come from informed traders who compete to exploit private information arbitrage opportunities.

Table A3. Permanent price impacts: congested days versus non-congested days. This table presents the permanent price impacts of DEX trade flows with different fee levels for days with different congestion levels. We classify days into high-gas, mid-gas, and low-gas days based on their median gas price. High-gas days are those with a median gas price above the 75%.

Pair	GasLevel Variable	HighGasDay	LowGasDay	MidGasDay
ETH-USDT	x^{CEX}	0.29	0.12	0.73
	r^{CEX}	1.0	1.01	1.01
	$x^{\text{HighFee-DEX}}$	1.35	2.13	2.28
	$x^{\text{LowFee-DEX}}$	0.12	0.57	0.11
	$x^{\text{MidFee-DEX}}$	0.15	0.54	0.29
LINK-ETH	x^{CEX}	0.91	-0.37	0.97
	r^{CEX}	1.01	0.95	1.06
	$x^{\text{HighFee-DEX}}$	2.94	0.74	4.41
	$x^{\text{LowFee-DEX}}$	0.89	0.5	0.44
	$x^{\text{MidFee-DEX}}$	1.74	0.18	1.73
WBTC-ETH	x^{CEX}	1.73	0.35	2.73
	r^{CEX}	1.13	1.18	1.2
	$x^{\text{HighFee-DEX}}$	2.73	2.03	3.87
	$x^{\text{LowFee-DEX}}$	0.52	0.56	0.45
	$x^{\text{MidFee-DEX}}$	1.5	0.94	1.49
AAVE-ETH	x^{CEX}	2.59	-0.01	0.7
	r^{CEX}	0.9	0.99	1.02
	$x^{\text{HighFee-DEX}}$	7.65	5.4	7.09
	$x^{\text{LowFee-DEX}}$	2.3	0.74	1.59
	$x^{\text{MidFee-DEX}}$	3.48	4.35	5.02

F Using DEX Return in the Structural VAR

In the baseline results, we calculate returns based on the Binance midquotes. As a robustness check, we use the returns based on the Uniswap "midquote". As we do not have quotes from Uniswap, an automated market maker, we calculate the midquote as the ratio of the amount of two tokens in the pool, y/x, i.e., the hypothetical price for an infinitesimal trade. Table A4 reports the estimation results with the DEX return. The results are largely consistent with what we have when using the Binance return: The permanent price impact of the high-fee DEX trade flow is much higher than that of the low-fee DEX flow.

For the lower bounds, the permanent price impact of the low-fee DEX flows are negative. One possible explanation is that the high-fee paying informed traders might time the low-fee flow so that they can trade *after* them to reduce the price impact and save transaction costs. So the low-fee flow, which is not directional in the first place, might appear to be on the wrong side all the time.

Table A4. Permanent price impact of DEX trade flows on DEX return.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDT	2.02^{***}	2.28^{***}	-0.51^{***}	0.57^{***}	-0.69^{***}	1.73^{***}	4.51^{***}	5.71^{***}	3.94^{***}
	(0.14)	(0.16)	(0.13)	(0.07)	(0.16)	(0.1)	(0.35)	(0.36)	(0.38)
LINK-ETH	1.01^{***}	1.21^{***}	-1.3^{***}	-0.13	-2.7^{***}	0.3	6.38^{***}	6.42^{***}	6.52^{***}
	(0.33)	(0.34)	(0.36)	(0.32)	(0.52)	(0.47)	(0.57)	(0.63)	(0.7)
WBTC-ETH	4.46^{***}	4.87^{***}	-1.57^{***}	-0.92^{***}	-1.07^{***}	0.26	3.11^{***}	3.52^{***}	4.02^{***}
	(0.46)	(0.48)	(0.32)	(0.34)	(0.42)	(0.45)	(0.5)	(0.5)	(0.63)
AAVE-ETH	1.82^{***}	2.31^{***}	-2.81^{***}	-1.07	-3.77^{***}	1.28	8.26^{***}	8.74^{***}	9.33^{***}
	(0.75)	(0.75)	(0.87)	(0.92)	(0.95)	(1.0)	(1.34)	(0.97)	(1.16)
Pooled	2.32^{***}	2.65^{***}	-1.52^{***}	-0.39	-2.06^{***}	0.85^{***}	5.8^{***}	5.89^{***}	6.19^{***}
	(0.25)	(0.25)	(0.26)	(0.26)	(0.3)	(0.3)	(0.34)	(0.41)	(0.42)

G Permanent Price Impact of Uniswap v3 Trade Flows

In our baseline results, we merged the trades from Uniswap v2 with those from Uniswap v3 pools. Given Uniswap v3 has a different liquidity provision mechanism, i.e., the concentrated liquidity, it might be interesting to investigate whether Uniswap v3 flow is more informative or not. In addition, Uniswap v3 allows for multiple pools with different fees for the same token pairs. Thus it would be interesting see for example whether the trade flow from low-fee v3 pool is more informative compare to high-fee pool.

Table A5 reports the estimation results. In general, we find the results are mixed. For example, when comparing Uniswap v2 and Uniswap v3 pool with 5bp swap fee ("Uniswap v3-5bp"), the results depend on the token pairs. For ETH-USDT and WBTC-ETH, Uniswap v2 pool seems to have a larger permanent price impact than Uniswap v3-5bp pool. In contrast, for the other two tokens, Uniswap v3-5bp pool has a larger permanent price impact.

When comparing the price impact of trade flows from Uniswap v3-5bp pool and Uniswap v3-30bp pool, the results remain mixed. It might result from our sample period, which stops before the EIP-1559 on August 15, 2022. We might have captured the transitional period where volume started to move from Uniswap v2 to Uniswap v3 pools which started in May 2022. For a follow-up study, examining the informativeness of the Uniswap v3 pools of different swap fees with an expanded sample period would be interesting in itself. For example, the theory model of Lehar, Parlour, and Zoican (2022) predicts that informed traders should prefer low-fee pools due to the lower transaction cost. However, our key focus is on whether the priority fee reveals private information of the informed traders, not the different price impacts across pools.

Table A5. Permanent price impact of Uniswap v2 and v3 trade flows.

Variable	x^{CEX}		x^{v2}		$x^{\text{v3-5bp}}$		$x^{\text{v3-30bp}}$		$\Delta^{\text{v2 - v3-5bp}}$	$\Delta^{\text{v2 - v3-30bp}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB	LB - UB
ETH-USDT	−0.02	0.07	0.99***	1.86***	0.9***	1.42***	1.24***	2.0***	−0.43***	−0.18
	(0.06)	(0.06)	(0.09)	(0.12)	(0.09)	(0.13)	(0.08)	(0.12)	(0.14)	(0.14)
LINK-ETH	−0.23	−0.18	1.56***	1.68***	0.09	0.13	0.28	0.77**	1.43***	0.16
	(0.22)	(0.22)	(0.26)	(0.28)	(0.19)	(0.19)	(0.32)	(0.35)	(0.32)	(0.37)
WBTC-ETH	0.32	0.36	1.57***	2.11***	1.83***	2.22***	0.61***	1.42***	−0.65***	−1.6***
	(0.25)	(0.26)	(0.23)	(0.24)	(0.19)	(0.21)	(0.15)	(0.18)	(0.27)	(0.24)
AAVE-ETH	−0.41	−0.2	4.4***	5.25***	0.49	0.59	3.5***	4.39***	3.81***	2.91***
	(0.64)	(0.65)	(0.75)	(0.76)	(0.85)	(0.85)	(0.78)	(0.78)	(1.0)	(1.0)
Pooled	−0.07	0.0	2.14***	2.68***	0.85***	1.07***	1.39***	2.12***	1.07***	0.32
	(0.18)	(0.18)	(0.22)	(0.23)	(0.22)	(0.23)	(0.22)	(0.23)	(0.29)	(0.29)

H Additional Robustness Checks

In the appendix, we conduct several other robustness checks.

H.1 SVAR results with information shares

As a robustness check, we compute another trade informativeness measure, the information shares of the trade flow variables (Hasbrouck 1991b). The basic idea is that we can decompose the logarithm of the midquote q_t into an efficient price component m_t and a microstructure noise term s_t:

$$q_t = m_t + s_t. \tag{8}$$

The efficient price component m_t is a random walk with innovation w_t: $m_t = m_{t-1} + w_t$ where $Ew = 0$, $Ew_t^2 = \sigma_w$, and $Ew_t w_\tau = 0$ for $\tau \neq t$. The microstructure noise term s_t is a zero-mean process that is jointly covariance stationary with w_t. Thus, the variance of the innovation σ_w has a natural interpretation as the total new information incorporated into the efficient price, which can be attributed to both public news and private information conveyed through trades.

As the above structural random walk decomposition is unobservable, to estimate the efficient price variance and decompose it to a trade-correlated, private information component and a trade-uncorrelated, public information, component, we need to resort to its reduced-form representation. First, we can rewrite the VMA representation above (see Equation 5) as follows:

$$\begin{pmatrix} r_t^{\text{CEX}} \\ \boldsymbol{x_t} \end{pmatrix} = \begin{pmatrix} \Theta^a(L) & \boldsymbol{\Theta^b(L)} \\ \boldsymbol{\Theta^c(L)} & \boldsymbol{\Theta^d(L)} \end{pmatrix} \begin{pmatrix} \varepsilon_{r,t} \\ \boldsymbol{\varepsilon_{x,t}} \end{pmatrix} \tag{9}$$

where $\boldsymbol{x_t}$[26] is the vector containing the Uniswap DEX trade flows. $\Theta^a(L)$ and $\boldsymbol{\Theta^b(L)}$, $\boldsymbol{\Theta^c}(L)$ and $\boldsymbol{\Theta^d}(L)$ are the polynomial of the lag operators with Θ^a, $\boldsymbol{\Theta^b}$, $\boldsymbol{\Theta^c}$ and $\boldsymbol{\Theta^d}$ being the conformable VMA system matrices in the return and trade equations. $\text{Var}(\varepsilon_{r,t}) = \sigma^2_{\varepsilon_r}$ and $\text{Var}(\boldsymbol{\varepsilon_{x,t}}) = \Sigma_{\varepsilon_x}$.

[26] We use normal font for scalar variables and bold font for vectors/matrices.

Skipping detailed proofs (See Hasbrouck 1991b), we have:

$$\sigma_w^2 = \boldsymbol{\Theta}^{\boldsymbol{b}}(\mathbf{1})\Sigma_{\varepsilon_x}\boldsymbol{\Theta}^{\boldsymbol{b}}(\mathbf{1})' + [\Theta^a(1)]^2\sigma_{\varepsilon_r}^2. \tag{10}$$

where the first component represents the private information conveyed through trades and the second component represents the public information through the news.

For a given A matrix, the structural innovations $\boldsymbol{\varepsilon_t}$ have zero contemporaneous correlations by construction, which implies that the covariance matrix Σ_{ε_x} is diagonal. Thus, we can further attribute the private information component to the contribution of each trade flow uniquely: $\boldsymbol{\Theta}^{\boldsymbol{b}}(\mathbf{1})\Sigma_{\varepsilon_x}\boldsymbol{\Theta}^{\boldsymbol{b}}(\mathbf{1})' = \sum_k[\Theta_k^b(1)]^2\sigma_{\varepsilon_k}^2$ where k indexes trade flow. $[\Theta_k^b(1)]^2$ is the k-th element of the vector of $\boldsymbol{\Theta}^{\boldsymbol{b}}(\mathbf{1})$ and $\sigma_{\varepsilon_k}^2$ is the k-th element of the diagonal of Σ_{ε_x}. Finally, to normalize each variable's information contribution, an absolute measure, to its information share, a relative measure bounded between 0 and 1, we divide it by the total efficient price variance. Formally, the information share (IS) of the trade flow variable k is computed as:

$$\mathrm{IS}_k = \frac{[\Theta_k^b(1)]^2\sigma_{\varepsilon_k}^2}{\sigma_w^2} \tag{11}$$

Compared with permanent price impact, the information share measure is a more comprehensive measure of trade informativeness as it weighs the permanent price impact of the trade flow variable by its own structural innovation variance. So, if two trade flow variables have the same permanent price impact, the one with a larger innovation (unexpected) variance will have a larger information share. Table A6 reports the results.

Two principal insights emerge from our analysis. Firstly, combined trade flows from both CEX and DEX contribute to roughly 15% of the variance in the efficient price innovation. The remainder, 85%, is contributed by the innovation in CEX returns (which we have excluded from Table A6 for succinctness). This dominant influence of CEX return innovation captures the significant role of public information in shaping the efficient price. Specifically, upon encountering public news, market makers on the CEX swiftly adjust their stable quotes, even before DEX trades occur. This

Table A6. Information shares of DEX trade flows with different priority fee levels. This table reports the information shares of the CEX trade flow and DEX trade flows with different priority fee levels. Information shares are computed using the formula in Equation 11. Upper bounds (UB) and lower bounds (LB) are obtained by considering all possible sequences of the recursive contemporaneous causality among the trade flows. The last column reports the difference between the lower bound of the information share of the high-fee DEX trade flow and the upper bound of the information share of the low-fee DEX trade flow. The estimation of the structural VAR model is done for each pair-day and statistical inference is based on variations across pair-day estimates. Numbers in brackets are standard errors.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDT	1.57	1.97	0.29	0.32	0.43	0.44	5.33	5.71	5.01***
	(0.3)	(0.33)	(0.06)	(0.08)	(0.06)	(0.07)	(0.7)	(0.75)	(0.71)
LINK-ETH	2.46	2.7	2.41	2.49	2.95	3.05	8.47	8.69	5.98***
	(0.45)	(0.47)	(0.69)	(0.65)	(0.54)	(0.56)	(1.21)	(1.23)	(1.38)
WBTC-ETH	6.81	8.44	1.86	1.95	3.4	3.73	12.49	13.94	10.54***
	(1.11)	(1.28)	(0.3)	(0.31)	(0.78)	(1.01)	(1.48)	(1.63)	(1.6)
AAVE-ETH	4.66	4.9	4.09	4.37	3.7	4.01	6.26	6.69	1.89
	(0.87)	(0.88)	(1.01)	(1.07)	(0.61)	(0.6)	(0.95)	(1.08)	(1.4)
Pooled	3.65	4.15	2.12	2.17	2.53	2.62	7.93	8.29	5.76***
	(0.36)	(0.39)	(0.33)	(0.33)	(0.27)	(0.29)	(0.56)	(0.58)	(0.66)

implies that the informational content of public news is reflected in the quote changes directly, not predicted by trades.[27]

Secondly, in alignment with our findings on the permanent price impact, high-fee DEX trade flow contributes a much larger share to price discovery than low-fee DEX trade flow. Specifically, the informational contribution of high-fee DEX trade flows ranges between 7.93% and 8.29%. In contrast, for low-fee DEX trade flows, this figure spans from 2.12% to 2.17%. When we measure the gap between the lower bound of the high-fee DEX trade flow's information share and the upper limit of the low-fee DEX trade flow, a noteworthy difference of 5.76% emerges. This difference is both statistically and economically meaningful. Given that trades are regarded as conveyors of private information, our data analysis highlights the pivotal role of high-fee DEX trade flows in integrating such private insights into the prevailing market prices.

[27]It is worth observing that similar patterns have been noted in past studies: 65% for a selection of 177 NYSE stocks in 1989's first quarter (Hasbrouck 1991b), and values ranging from 72% to 88% for NASDAQ stocks sampled in June 2000 (Barclay, Hendershott, and McCormick 2003).

H.2 Unstandardized DEX trade flows

In our baseline results, we standardize all DEX trade flows to have zero mean and unit variance. So permanent price impacts of the DEX trade flows mean cumulative return impulse responses to a one standard deviation shock in the trade flows. The benefit is that we can have a fair comparison and compute average measures across token pairs with different trading volumes and liquidity. For example, an one-ETH order flow might be small in liquid tokens like ETH-USDT but large in less liquid ones such as AAVE-ETH. As a robustness check, we use the unstandardized the DEX trade flows instead so that the permanent price impacts are measured in the units of dollars. It facilitates the comparison between the trade flows of different fee levels within the same token pair.

Table A7 reports the estimation results. They show that in terms of per-ETH permanent price impact, the high-DEX trade flow remains more informative than the low-fee DEX trade flow.

Table A7. Robustness: Permanent price impact of DEX trade flows with different priority fee levels: Unstandardized trade flows. This table reports the permanent price impacts of the CEX trade flow and DEX trade flows with high, medium, and low priority fee levels. Permanent price impacts are defined as the cumulative impulse responses of the CEX return to DEX trade flow in the structural VAR model (see Equation 3). Upper bounds (UB) and lower bounds (LB) are obtained by considering all possible sequences of the recursive contemporaneous causality among the endogenous variables. The last column reports the difference between the lower bound of the permanent price impact of the high-fee DEX trade flow and the lower bound of the permanent price impact of the low-fee DEX trade flow. The estimation of the structural VAR is done for each pair-day and statistical inference is based on variations in the pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis points. DEX trade flows standardized. For Stable pairs, the DEX trade flows are in thousand USDT. The DEX trade flows are in ETH. *, ** and *** indicate significance levels at 1%, 5% and 10% respectively.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDT	0.0***	0.0***	−0.01*	0.0	0.01***	0.01***	0.08***	0.08***	0.08***
	(0.0)	(0.0)	(0.01)	(0.01)	(0.0)	(0.0)	(0.01)	(0.01)	(0.01)
LINK-ETH	0.05**	0.06***	0.08*	0.09**	0.15***	0.16***	0.32***	0.33***	0.23***
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.05)
WBTC-ETH	0.6***	0.67***	0.04	0.05	0.05***	0.05***	0.12***	0.12***	0.07*
	(0.09)	(0.1)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)
AAVE-ETH	0.21**	0.25***	0.36***	0.41***	0.48***	0.53***	0.62***	0.65***	0.21
	(0.1)	(0.1)	(0.14)	(0.14)	(0.05)	(0.06)	(0.05)	(0.05)	(0.15)
Pooled	0.21***	0.24***	0.12***	0.13***	0.18***	0.19***	0.29***	0.3***	0.15***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)

H.3 Priority fee level classification

In the baseline estimation, we use 20 blocks as the length of the rolling window in fee-level classification. As a robustness check, we try two different window lengths, 5 blocks, and 10 blocks, to classify DEX trades and then redo the structural VAR estimation. Table A8 reports the estimation results of the cumulative return impulse responses based on DEX trade flows from the two alternative gas level classifications. It shows that the results are largely unchanged compared with the baseline results in Table 5.

H.4 Lag order choice

In our baseline specification for the structural VAR model, we include lagged return and trade flow variables of the last five blocks. As a robustness check, we vary the number of lags in the structural VAR specification. Table A9 report the permanent price impacts when the number of lags is set to 10 and 20 respectively. It shows that the results are qualitatively the same as the baseline results.

Table A8. Permanent price impact of DEX trade flows with different priority fee levels: Fee level classification based on a rolling window of alternative lengths. This table reports the permanent price impacts of the DEX trade flows with high, medium, and low priority fee levels. Permanent price impacts are defined as the cumulative impulse responses of the CEX return to CEX trade flow and DEX trade flow in the structural VAR model (see Equation 3). Upper bounds (UB) and lower bounds (LB) are obtained by considering all possible sequences of the recursive contemporaneous causality among the endogenous variables. The last column reports the difference between the lower bound of the permanent price impact of the high-fee DEX trade flow and the upper bound of the permanent price impact of the low-fee DEX trade flow. The estimation of the structural VAR is done for each pair-day and statistical inference is based on variations across the pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis points. DEX trade flows are standardized and thus in their standard deviations. *, ** and *** indicate significance levels at 1%, 5% and 10% respectively.

(a) Fee level classification based on a rolling window of 10 blocks.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDT	0.91***	1.14***	−0.09	−0.06	0.33***	0.38***	2.47***	2.67***	2.53***
	(0.13)	(0.15)	(0.06)	(0.05)	(0.07)	(0.08)	(0.29)	(0.31)	(0.29)
LINK-ETH	0.56*	0.69***	0.54***	0.58***	2.27***	2.42***	4.69***	4.96***	4.11***
	(0.29)	(0.29)	(0.22)	(0.22)	(0.29)	(0.31)	(0.48)	(0.51)	(0.49)
WBTC-ETH	2.82***	3.56***	0.62***	0.77***	1.45***	1.96***	4.55***	5.44***	3.79***
	(0.38)	(0.47)	(0.23)	(0.25)	(0.31)	(0.41)	(0.47)	(0.59)	(0.56)
AAVE-ETH	1.02	1.31**	1.88***	2.3***	5.81***	6.56***	7.23***	8.08***	4.93***
	(0.68)	(0.62)	(0.59)	(0.6)	(0.65)	(0.66)	(0.58)	(0.63)	(0.81)
Pooled	1.3***	1.64***	0.74***	0.88***	2.51***	2.82***	4.75***	5.1***	3.87***
	(0.22)	(0.22)	(0.17)	(0.18)	(0.23)	(0.24)	(0.25)	(0.27)	(0.28)

(b) Fee level classification based on a rolling window of 40 blocks.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDT	0.9***	1.14***	−0.07*	−0.05	0.18***	0.21***	2.47***	2.67***	2.51***
	(0.13)	(0.15)	(0.04)	(0.03)	(0.06)	(0.07)	(0.29)	(0.31)	(0.29)
LINK-ETH	0.53*	0.67***	0.41*	0.47**	1.74***	1.92***	4.86***	5.23***	4.39***
	(0.29)	(0.27)	(0.22)	(0.22)	(0.28)	(0.31)	(0.52)	(0.55)	(0.54)
WBTC-ETH	2.95***	3.59***	0.31	0.43*	1.12***	1.47***	4.86***	5.59***	4.43***
	(0.45)	(0.5)	(0.21)	(0.24)	(0.26)	(0.32)	(0.47)	(0.56)	(0.54)
AAVE-ETH	0.92	1.13*	2.18***	2.54***	4.68***	5.36***	7.67***	8.54***	5.13***
	(0.71)	(0.65)	(0.61)	(0.65)	(0.63)	(0.64)	(0.63)	(0.66)	(0.94)
Pooled	1.3***	1.58***	0.71***	0.83***	1.96***	2.23***	4.96***	5.32***	4.13***
	(0.23)	(0.23)	(0.18)	(0.19)	(0.21)	(0.22)	(0.27)	(0.28)	(0.31)

Table A9. Permanent price impact of DEX trade flows with different priority fee levels: Alternative number of lags in the structural VAR specification. This table reports the permanent price impacts of the DEX trade flows with high, medium, and low priority fee levels. Permanent price impacts are defined as the cumulative impulse responses of the CEX return to CEX trade flow and DEX trade flow in the structural VAR model (see Equation 3). Upper bounds (UB) and lower bounds (LB) are obtained by considering all possible sequences of the recursive contemporaneous causality among the endogenous variables. The last column reports the difference between the lower bound of the permanent price impact of the high-fee DEX trade flow and the upper bound of the permanent price impact of the low-fee DEX trade flow. The estimation of the structural VAR is done for each pair-day and statistical inference is based on variations across the pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis points. DEX trade flows are standardized and thus in their standard deviations. *, ** and *** indicate significance levels at 1%, 5% and 10% respectively.

(a) 10 lags of CEX return and DEX trade flows included in the structural VAR.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDT	0.92***	1.16***	−0.09	−0.07	0.38***	0.45***	2.63***	2.85***	2.69***
	(0.13)	(0.16)	(0.06)	(0.06)	(0.1)	(0.11)	(0.3)	(0.33)	(0.32)
LINK-ETH	0.42	0.56*	0.28	0.31	2.21***	2.4***	5.14***	5.47***	4.83***
	(0.31)	(0.32)	(0.32)	(0.32)	(0.42)	(0.44)	(0.5)	(0.53)	(0.56)
WBTC-ETH	2.69***	3.68***	0.7**	0.87***	2.1***	2.62***	4.54***	5.33***	3.68***
	(0.65)	(0.65)	(0.35)	(0.37)	(0.45)	(0.51)	(0.5)	(0.66)	(0.69)
AAVE-ETH	1.66*	1.94**	1.34*	1.76**	6.31***	6.97***	8.45***	9.24***	6.7***
	(0.96)	(0.91)	(0.79)	(0.82)	(0.89)	(0.91)	(0.82)	(0.88)	(1.08)
Pooled	1.4***	1.76***	0.56***	0.7***	2.77***	3.11***	5.21***	5.54***	4.51***
	(0.3)	(0.3)	(0.23)	(0.24)	(0.3)	(0.31)	(0.31)	(0.33)	(0.37)

(b) 20 lags of CEX return and DEX trade flows included in the structural VAR.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDT	0.78***	1.01***	−0.19**	−0.15*	0.44***	0.53***	2.63***	2.84***	2.78***
	(0.16)	(0.18)	(0.09)	(0.08)	(0.12)	(0.13)	(0.33)	(0.36)	(0.36)
LINK-ETH	1.04**	1.18***	0.6	0.67	1.99***	2.11***	5.24***	5.55***	4.57***
	(0.47)	(0.46)	(0.52)	(0.52)	(0.52)	(0.49)	(0.58)	(0.6)	(0.77)
WBTC-ETH	3.65***	4.45***	0.75	1.07**	1.32**	2.43***	5.05***	6.59***	3.98***
	(0.6)	(0.91)	(0.5)	(0.51)	(0.59)	(0.8)	(0.71)	(0.99)	(0.86)
AAVE-ETH	1.4	1.78	2.03*	2.52***	6.64***	7.37***	8.23***	8.77***	5.71***
	(1.15)	(1.15)	(1.04)	(1.06)	(1.18)	(1.15)	(1.37)	(1.4)	(1.78)
Pooled	1.63***	1.85***	0.79***	1.0***	2.68***	3.05***	5.35***	5.77***	4.34***
	(0.36)	(0.38)	(0.32)	(0.33)	(0.38)	(0.39)	(0.44)	(0.46)	(0.55)

Online Appendix for "Price Discovery on Decentralized Exchanges"

In this online appendix, we report main results for the other two Ethereum token pairs: ETH-USDC and ETH-DAI. We decide to exclude these two token pairs as they are very similar to ETH-USDT included in the main body of the paper. All three token pairs are essentially tracking the Ethereum price. The results of these two new tokens are largely the same as the ETH-USDT.

Table OA1. Permanent price impact of DEX trade flows. The results are for two new token pairs: ETH-USDC and ETH-DAI.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDC	0.45***	0.53***	0.08	0.09	0.46***	0.46***	2.76***	2.77***	2.67***
	(0.09)	(0.1)	(0.08)	(0.08)	(0.11)	(0.11)	(0.32)	(0.33)	(0.34)
ETH-DAI	2.33***	2.55***	0.24	0.31*	0.94***	1.03***	3.55***	3.71***	3.23***
	(0.35)	(0.38)	(0.17)	(0.18)	(0.23)	(0.24)	(0.41)	(0.42)	(0.46)
NonStable-Pooled	1.52***	2.02***	−0.24	0.5	1.72***	2.25***	4.51***	4.71***	4.0***
	(0.48)	(0.41)	(0.41)	(0.65)	(0.47)	(0.47)	(0.41)	(0.36)	(0.8)

Table OA2. Simple price impacts measure. The results are for two new token pairs: ETH-USDC and ETH-DAI.

Token Pair	Fee Level Measure	Low-Fee DEX Trades	Mid-Fee DEX Trades	High-Fee DEX Trades
USDC-ETH	RPI (20 Blocks)	1.46	2.51	10.57
	RPI (60 Blocks)	3.05	3.13	10.93
	RPI (120 Blocks)	2.92	2.77	9.86
DAI-ETH	RPI (20 Blocks)	1.54	2.95	14.69
	RPI (60 Blocks)	2.45	2.21	16.53
	RPI (120 Blocks)	3.58	1.42	15.22

Table OA3. Permanent price impact of DEX trade flows: Excessive gas trades excluded. The results are for two new token pairs: ETH-USDC and ETH-DAI.

Variable	x^{CEX}		$x^{\text{LowFee-DEX}}$		$x^{\text{MidFee-DEX}}$		$x^{\text{HighFee-DEX}}$		$\Delta^{\text{High - Low}}$
	LB	UB	LB	UB	LB	UB	LB	UB	LB - UB
ETH-USDC	0.51***	0.55***	0.11	0.11	0.49***	0.5***	1.46***	1.46***	1.34***
	(0.1)	(0.1)	(0.09)	(0.09)	(0.11)	(0.11)	(0.16)	(0.16)	(0.2)
ETH-DAI	2.38***	2.52***	0.19	0.27	1.16***	1.23***	2.45***	2.58***	2.19***
	(0.38)	(0.39)	(0.2)	(0.2)	(0.24)	(0.24)	(0.33)	(0.34)	(0.36)
NonStable-Pooled	1.44***	1.54***	0.15	0.19*	0.83***	0.86***	1.95***	2.02***	1.77***
	(0.21)	(0.22)	(0.11)	(0.11)	(0.14)	(0.14)	(0.19)	(0.19)	(0.21)

Figure OA1. Block position of trades from informed addresses and uninformed addresses. The results are for two new token pairs: ETH-USDC and ETH-DAI.

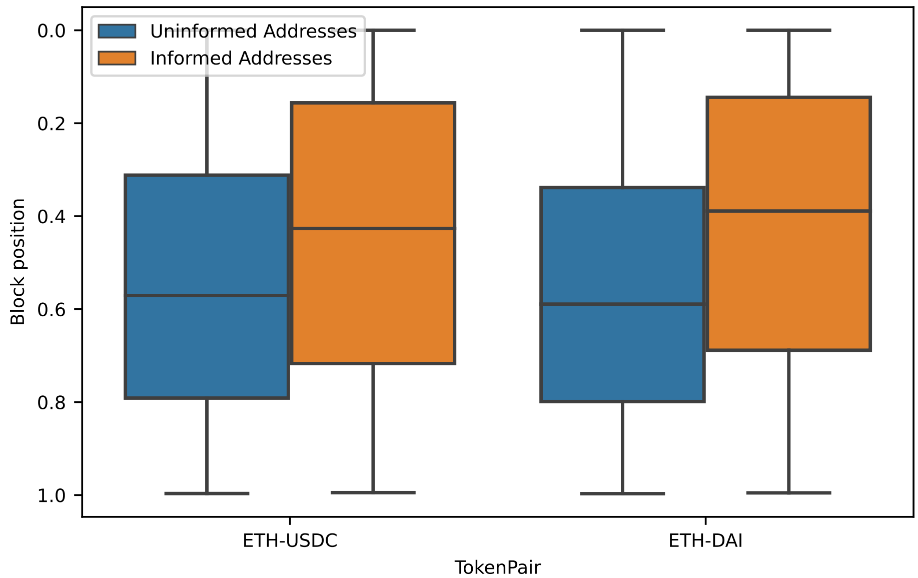


Table OA4. Percentages of priority gas auction (PGA) trades. This table shows the fraction of trades identified as priority gas auction (PGA) trades, for "excessively-high-fee trades" and other trades.

TokenPair	ExplicitCompetition ExcessiveGas	Jump bidding trades	PGA trades	Other
ETH-USDC	Other trades	95.05	2.53	2.43
	Excessively-high-fee trades	84.43	13.20	2.37
ETH-DAI	Other trades	94.66	2.79	2.54
	Excessively-high-fee trades	83.79	14.06	2.14

References

Aoyagi, Jun and Yuki Ito (2021). *Coexisting Exchange Platforms: Limit Order Books and Automated Market Makers*. URL: https://papers.ssrn.com/abstract=3808755. preprint.

Aquilina, Matteo, Eric Budish, and Peter O'Neill (2021). "Quantifying the High-Frequency Trading "Arms Race"". *The Quarterly Journal of Economics* 137.1, pp. 493–564.

Avery, Christopher (1998). "Strategic Jump Bidding in English Auctions". *Review of Economic Studies* 65.2, pp. 185–210.

Back, Kerry, C. Henry Cao, and Gregory A. Willard (2000). "Imperfect Competition among Informed Traders". *The Journal of Finance* 55.5, pp. 2117–2155.

Barbon, Andrea and Angelo Ranaldo (2022). *On The Quality Of Cryptocurrency Markets: Centralized Versus Decentralized Exchanges*. URL: https://papers.ssrn.com/abstract=3984897. preprint.

Barclay, Michael J., Terrence Hendershott, and D. Timothy McCormick (2003). "Competition among Trading Venues: Information and Trading on Electronic Communications Networks". *The Journal of Finance* 58.6, pp. 2637–2665.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2019). "Price Discovery without Trading: Evidence from Limit Orders". *The Journal of Finance* 74.4, pp. 1621–1658.

Brolley, Michael and Marius Zoican (2023). "Liquid Speed: A Micro-Burst Fee for Low-Latency Exchanges". *Journal of Financial Markets* 64, p. 100785.

Brunnermeier, Markus K. (2005). "Information Leakage and Market Efficiency". *Review of Financial Studies* 18.2, pp. 417–457.

Budish, Eric, Peter Cramton, and John Shim (2015). "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response*". *The Quarterly Journal of Economics* 130.4, pp. 1547–1621.

Caldentey, René and Ennio Stacchetti (2010). "Insider Trading With a Random Deadline". *Econometrica* 78.1, pp. 245–283.

Capponi, Agostino and Ruizhe Jia (2021). *The Adoption of Blockchain-based Decentralized Exchanges*. URL: https://papers.ssrn.com/abstract=3805095. preprint.

Collin-Dufresne, Pierre and Vyacheslav Fos (2015). “Do Prices Reveal the Presence of Informed Trading?: Do Prices Reveal the Presence of Informed Trading?” *The Journal of Finance* 70.4, pp. 1555–1582.

Daian, Philip et al. (2020). “Flash Boys 2.0: Frontrunning in Decentralized Exchanges, Miner Extractable Value, and Consensus Instability”. *2020 IEEE Symposium on Security and Privacy (SP)*. 2020 IEEE Symposium on Security and Privacy (SP), pp. 910–927.

Daniel, Kent D. and David A. Hirshleifer (1998). *A Theory of Costly Sequential Bidding*. URL: https://papers.ssrn.com/abstract=161013. preprint.

Easley, David and Maureen O’Hara (1987). “Price, Trade Size, and Information in Securities Markets”. *Journal of Financial Economics* 19.1, pp. 69–90.

Fleming, Michael J., Bruce Mizrach, and Giang Nguyen (2018). “The Microstructure of a U.S. Treasury ECN: The BrokerTec Platform”. *Journal of Financial Markets* 40, pp. 2–22.

Foster, F. Douglas and S. Viswanathan (1996). “Strategic Trading When Agents Forecast the Forecasts of Others”. *The Journal of Finance* 51.4, pp. 1437–1478.

Han, Jianlei, Shiyang Huang, and Zhuo Zhong (2022). *Trust in DeFi: An Empirical Study of the Decentralized Exchange*. URL: https://papers.ssrn.com/abstract=3896461. preprint.

Hasbrouck, Joel (1991a). “Measuring the Information Content of Stock Trades”. *The Journal of Finance* 46.1, pp. 179–207.

— (1991b). “The Summary Informativeness of Stock Trades: An Econometric Analysis”. *Review of Financial Studies* 4.3, pp. 571–595.

— (1995). “One Security, Many Markets: Determining the Contributions to Price Discovery”. *The Journal of Finance* 50.4, pp. 1175–1199.

Hasbrouck, Joel, Thomas J. Rivera, and Fahad Saleh (2022). *The Need for Fees at a DEX: How Increases in Fees Can Increase DEX Trading Volume*. URL: https://papers.ssrn.com/abstract=4192925. preprint.

Heimbach, Lioba, Vabuk Pahari, and Eric Schertenleib (2024). *Non-Atomic Arbitrage in Decentralized Finance*. URL: http://arxiv.org/abs/2401.01622. preprint.

Holden, Craig W. and Avanidhar Subrahmanyam (1992). "Long-Lived Private Information and Imperfect Competition". *The Journal of Finance* 47.1, pp. 247–270.

Kaniel, Ron and Hong Liu (2006). "So What Orders Do Informed Traders Use?" *The Journal of Business* 79.4, pp. 1867–1913.

Kyle, Albert S. (1985). "Continuous Auctions and Insider Trading". *Econometrica* 53.6, p. 1315.

Lehar, Alfred and Christine A. Parlour (2021). *Decentralized Exchanges*. URL: https://papers.ssrn.com/abstract=3905316. preprint.

Lehar, Alfred, Christine A. Parlour, and Marius Zoican (2022). *Liquidity Fragmentation on Decentralized Exchanges*. URL: https://papers.ssrn.com/abstract=4267429. preprint.

Maggio, Marco Di, Francesco Franzoni, Amir Kermani, and Carlo Sommavilla (2019). "The Relevance of Broker Networks for Information Diffusion in the Stock Market". *Journal of Financial Economics* 134.2, pp. 419–446.

Milionis, Jason, Ciamac C. Moallemi, Tim Roughgarden, and Anthony Lee Zhang (2022). *Automated Market Making and Loss-Versus-Rebalancing*. URL: http://arxiv.org/abs/2208.06046. preprint.

O'Hara, Maureen, Chen Yao, and Mao Ye (2014). "What's Not There: Odd Lots and Market Data: What's Not There: Odd Lots and Market Data". *The Journal of Finance* 69.5, pp. 2199–2236.

Park, Andreas (2023). "The Conceptual Flaws of Decentralized Automated Market Making". *Management Science*.

Wang, Ye, Patrick Zuest, Yaxing Yao, Zhicong Lu, and Roger Wattenhofer (2022). "Impact and User Perception of Sandwich Attacks in the DeFi Ecosystem". *CHI Conference on Human Factors in Computing Systems*. CHI '22: CHI Conference on Human Factors in Computing Systems. New Orleans LA USA: ACM, pp. 1–15.

Yang, Liyan and Haoxiang Zhu (2020). "Back-Running: Seeking and Hiding Fundamental Information in Order Flows*". *The Review of Financial Studies* 33.4. Ed. by Andrew Karolyi, pp. 1484–1533.