Price Discovery on Decentralized Exchanges[1]

Agostino Capponi Ruizhe Jia Shihao Yu

Current version: February 1, 2023

[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, Bart Zhou Yueshen, Sean Foley, Tālis Putniņš, Marius Zoican, and seminar participants at the Microstructure Exchange (Asia-pacific edition), Hong Kong Baptist University, Luohan Academy, Oxford University DeFi Working Group, for helpful comments.

Price Discovery on Decentralized Exchanges

Abstract

In contrast to centralized exchanges (CEXs) which match orders continuously following a price-time priority rule, decentralized exchanges (DEXs) process orders in discrete time and require traders to bid a blockchain priority fee to determine the execution priority of their orders. We employ a structural vector-autoregressive (structural VAR) model to provide evidence that blockchain fees attached to DEX trades reveal their private information, contributing to price discovery. We show that informed traders bid higher fees not only to avoid execution risk resulting from blockchain congestion but also to compete with each other. Using a unique dataset of Ethereum mempool orders, we further demonstrate that informed traders mostly compete on DEXs through a jump bidding strategy.

1 Introduction

Price discovery, the process in which market participants reach a consensus about the fundamental value of an asset, is a key function of financial markets. How such a process realizes has been a central topic in market microstructure, and it largely depends on various aspects of asset trading including market structure (e.g., dark pools, centralized market versus OTC market as in Zhu (2014); Hagströmer and Menkveld (2019)), transparency rule (e.g., pre-trade transparency versus post-trade transparency, as in Bloomfield and O'Hara (1999); Boehmer, Saar, and Yu (2005)), and trading constraints (e.g., short sell ban as in Boehmer and Wu (2013)).

Decentralized exchanges (DEXs) are trading venues built on public blockchains. They enable trading of digital assets without the need for centralized intermediaries and have gained a sizable trading volume and market share since their inception.[1] In contrast to centralized exchanges (CEXs), which typically execute orders continuously based on a price-time priority rule, DEXs execute orders in discrete time and traders bid a blockchain fee to determine the execution priority of their orders.[2] Given the unique trading mechanism, price discovery on DEXs might have distinct features. For example, does blockchain fee, a new trade characteristic, convey any information? How does blockchain fee bidding affect informed trading, and what does it imply for price discovery? The objective of our paper is to provide answers to the above questions.

To do so, we construct a data set consisting of trades from Uniswap (the largest DEX) and Binance (the largest CEX) and mempool orders data on the Ethereum blockchain. We focus on the six most traded token pairs during our sample period between November 18, 2020, and February 10, 2021. With the trade data, we use a structural vector-autoregressive (structural VAR) model to investigate the information content of DEX trade flows with different levels of fee (Hasbrouck, 1991). The tick-by-tick mempool data tracks all orders submitted to the Ethereum network, as

[1]For the spot trading of cryptocurrencies, the aggregated trading volume on DEXs hovers between 50 billion USD and 200 billion USD, which correspond to a market share between 10% and 20%. See `https://www.theblock.co/data/decentralized-finance/dex-non-custodial`.

[2]For example, for DEXs on the Ethereum blockchain, orders are executed by blockchain validators in block time or every 12 seconds, and their execution priority is determined by what's called gas price bid by traders.

well as the blockchain fee (gas price) bids by traders. Thus, it allows us to investigate the bidding strategy of competing traders on DEXs.

Our main findings are summarized below. We find that DEX trades with high fees reveal private information, and more so compared to DEX trades with low fees: for token pairs involving a non-stablecoin (e.g., Ethereum and Bitcoin), a shock to a high-fee DEX trade flow leads to a much larger permanent impact on market price than that to a low-fee trade flow. For example, a positive shock of one standard deviation to the high-fee DEX trade flow results in a permanent increase of about 8.16 basis points in market price. In contrast, a shock of the same size to low-fee DEX trade flow permanently moves the market price by only about 0.83 basis points.

High-fee DEX trade flow is more informative, indicating that informed traders bid high fees to execute their orders on DEXs. What can be the economic channels? One possible channel is that they bid high fees to avoid execution risk due to blockchain congestion. Such a motive naturally arises as informed traders on DEXs have to compete with other blockchain users for limited block space. As the blockchain becomes congested and the marginal blockchain fee increases, informed traders will increase their bids in order to execute their orders in time. However, we show that it is not the only channel. We find that informed traders can bid excessively high fees for their trades, which is less likely to be driven by them merely avoiding execution risk. Rather, it is more likely to result from them competing with each other.

How do informed traders compete on DEXs? Our analysis of tick-by-tick mempool order data shows that even for DEX trades with excessively high fees, only a small fraction of them are likely to result from priority gas auctions (PGAs) where traders competitively bid up fees (Daian et al., 2020). Instead, they start by bidding a rather high fee, discouraging competition from other traders. Such a pattern fits the jump bidding strategy documented in the auction theory (Avery, 1998), which can be rationalized by a high bidding cost on DEXs and the winner's curse problem.

Our paper relates to several streams of literature. Past studies in market microstructure have linked the private information contained in trades to their public characteristics[3], e.g., block trades

[3] Brogaard, Hendershott, and Riordan (2014) use proprietary data to investigate the information content of private

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).[4] We contribute to the literature by studying the information content of blockchain fees, a featuring characteristic of trades executed on DEXs besides price and trade size. As informed traders on DEXs have to bid fees to get their orders executed, blockchain fees can potentially serve as a new public signal revealing the private information contained in DEX trades.

Our paper also contributes to the nascent yet rapidly growing literature on decentralized exchanges, and the role of blockchain fees in the provision of trading and liquidity incentives. Park (2021) focuses on the unintended consequence of public blockchain order processing, which exposes all pending DEX transactions to the risk of a "sandwich attack". He argues that, in theory, liquidity demanders are able to prevent frontrunning by choosing a very high blockchain fee. Capponi and Jia (2021) investigate how the choice of DEX pricing rules affects welfare and liquidity provision incentives. They show that arbitrageurs can always outbid liquidity providers, in blockchain fee auctions, to exploit the price discrepancy between CEXs and DEXs, which in turn reduces incentives for liquidity provision. Barbon and Ranaldo (2021) compare the price efficiency of CEX and DEX. They argue that the low price efficiency of Uniswap can be partially attributed to high fees which are fixed costs for traders. Parlour (2021) contrast DEXs running an automated market maker (AMM) with CEXs running a central limit order book (CLOB) and focus on the different trade-offs faced by liquidity providers. Aoyagi and Ito (2021) instead model the coexistence of an AMM-based DEX and a CLOB-based CEX and study the resulting equilibrium in liquidity provision. The main contribution of our work relative to this literature is highlighting how blockchain fees convey both private and public information. On DEXs where the market price can only be revised through trading, we show that trades with high fees make DEX prices

trade characteristics e.g., HFT trades versus non-HFT trades.

[4] A related literature is on the trading strategy of the informed trader(s) in various settings, e.g., a monopolistic informed trader (Kyle, 1985) or 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 the private information (Caldentey and Stacchetti, 2010) or short information horizon (Kaniel and Liu, 2006).

informationally efficient.

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, and discuss the results in Section 5. We conclude in Section 6.

2 Institutional background

We introduce institutional features of DEXs in Section 2.1, and characteristics of trade execution on DEXs in Section 2.2. We focus our discussion primarily on the blockchain fee-setting mechanism.

2.1 DEXs

DEXs are peer-to-peer marketplaces, which facilitate trading through automated smart contracts. They operate on blockchain and thus execute trades without an intermediary. As of July 2022, about 15% of crypto spot trading occurred on DEXs. Uniswap is currently the largest DEX by trading volume and accounts for more than half of the total DEX trading volume. The remaining 85% of crypto spot trading is executed on CEXs. The largest one by daily trading volume is Binance, which accounts for more than 75% of the CEX market share. Different from CEXs which utilize limit order books, most DEXs are in the form of Automated Market Makers (AMM). In AMMs, liquidity providers deposit token pairs to the pool, and the pricing schedule is determined by an exogenously specified bonding curve, which is pre-coded in the smart contract. We refer to Capponi and Jia (2021) for additional details about DEXs.

2.2 Trade execution on DEXs and blockchain fees

DEXs rely on blockchain networks (e.g., Ethereum) to receive, process, and execute orders. To execute a trade on a DEX, a trader has to first broadcast the transaction details in the network and bid a fee for her order. The transaction details reveal trade information even before the trade is executed, such as the address of the DEX and the intended trade size and price. Once a transaction is received by a validator, it will be pending in its mempool. If that validator is chosen to append the next block to the chain, then she will execute transactions in her mempool in decreasing order of fees. Since blocks are produced at discrete times, DEX orders are also processed discretely in batches. Because each block has a maximum capacity, transactions with too low fees will not be included in the block, or they need to wait for a long time before being executed.

Every transaction broadcast by a trader is associated with a number called “nonce”. Each nonce can only be used once and in increasing order. In other words, 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”. If a trader wants to modify her pending order or increase the fee bid, she has to broadcast a new transaction with the same nonce as the pending one and increase the bid fee. A validator who receives the new transaction will not execute the old one because she prioritizes transactions with higher fees. For the same reason, a transaction with the same nonce as a previously submitted transaction but with a lower fee will not be executed, because the validator would assign a lower priority to it.

In the Ethereum blockchain, the blockchain fees are referred to as “gas fees”. The execution of each transaction requires a fixed amount of computational resources, measured by the “gas used”. More complicated transactions need more computational work, so they require a higher amount of gas than simple transactions such as payment transfers. For example, a standard ETH transfer requires a gas amount of 21,000 units. Upon bidding blockchain fees, Ethereum users specify the “gas price”, i.e., how much they are willing to pay per unit of gas. The total gas fee paid by users is equal to the gas used multiplied by the gas price bid. Note that Ethereum validators sort and execute transactions in mempools in decreasing order of gas price.

3 Data

We describe the dataset used for our empirical analysis. We introduce the executed trade data in Section 3.1, and the tick-by-tick mempool order data in Section 3.2.

3.1 Executed trade data

Our dataset covers trades executed on Binance, the largest CEX, and Uniswap, the largest DEX, for six actively traded token pairs during our sample period, November 18, 2020, through February 10, 2021. Note that our sample period starts after Uniswap's staking reward program was terminated so that we avoid including a structural break in token liquidity as the termination resulted in large token outflows and smaller pool sizes. In addition, our sample period ends before the first block including Flashbots trades was mined. Flashbots allow traders to directly send their orders to miners through private channels to avoid exposing them in the mempool. Our six token pairs are USDC-USDT, DAI-USDT, ETH-USDT, WBTC-ETH, LINK-ETH, and AAVE-ETH, and they fall into two types: "Stable" and "NonStable". "Stable" pairs include two stablecoins pegged to one US Dollar (USDC-USDT, ETH-USDT). "NonStable" pairs include at least one non-stable token, i.e., which is not pegged to any fiat currency (ETH-USDT, WBTC-ETH, LINK-ETH, and AAVE-ETH). Binance trades are publicly available and collected from the Binance website[5], while Uniswap trades are collected through a proprietary node. Below. we provide a detailed description for each of the two datasets.

Uniswap trades Each Uniswap trade contains the hash, the address of the trader, 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, the execution position of the trade in that block, gas price, gas used, trade direction indicating whether it is a buy trade or sell trade in terms of the base token[6], the amount of tokens in the liquidity pool before and after the trades, and the amount of tokens that the trader deposits in and takes out from the liquidity pool.

Using the amount of tokens in the liquidity pool before a trade, we can compute the prevailing "midquote" just before the trade. For example, if there is an amount y of tokens Y and an amount

[5]We refer to https://data.binance.vision/?prefix=data/spot/monthly/ for details.

[6]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, for the token pair ETH-USDT, ETH is the base token and USDT is the quote token.

x of tokens X before the trade, the prevailing midquote is simply the ratio of the amount of two tokens in the pool, x/y. Note that on AMMs like Uniswap, there are no quotes. Thus, we define "midquote" as the hypothetical price for an infinitesimal trade. In addition, based on the amount of tokens that the trader deposits in and takes out from the liquidity pool, we can compute the transaction price of the trade. For example, if Δy amount of token Y is swapped for x amount of token X, then the transaction price is simply the ratio of the amount of two tokens swapped, i.e., $\Delta x/\Delta y$. Last, we use the amount of the base token swapped as the transaction size of the trade, that is, $|\Delta y|$.

Binance trades Each Binance trade record includes a unique identifier for the trade, the timestamp (to the precision of 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 a 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 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

In addition to executed trade data, we obtain tick-by-tick Ethereum mempool order data from amberdata[7]. The dataset includes every new order submission received in the mempool of nodes maintained by amberdata, which either ends up with being 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 millisecond), the address of the trader, nonce, gas price, and gas limit (i.e. the maximum gas allowed to be used).

[7]amberdata is a US data company specializing in market data in decentralized finance.

Our mempool data covers the same sample period of November 18, 2020, through February 10, 2021. With the mempool data, we can track the complete history of order revisions, if they occur, before the final order is executed and recorded as a trade. Hence, we are able to see whether the trader increases the gas price attached to her order to get it executed.

3.3 Summary statistics of executed trades

To provide an overview of trading in our sample token pairs, in Table 1, we report summary statistics of their daily trading volume and daily number of trades on Uniswap and Binance respectively. Several notable observations are in order. First, trading in all six token pairs is fairly active. For instance, the average daily number of trades (daily trading volume) on Uniswap is 997 (≈ 3.4 million USDT), 8,560 (73,489 ETH ≈ 66 million USD) and 1,371 (31,644 ETH ≈ 28 million USD) for USDC-USDT, ETH-USDT and WBTC-ETH respectively.

Second, trading activity on Uniswap and Binance differs significantly across token pairs. For the two Stable token pairs, USDC-USDT and DAI-USDT, trading is much more active on Binance than Uniswap. For example, the average daily trading volume on Binance is about 96 million USDT for USDC-USDT, more than an order of magnitude larger than that on Uniswap. It is because trading is cheaper on Binance than Uniswap, as the latter imposes a larger price impact due to the convexity of the bonding curve, and requires an additional blockchain fee. Importantly, the transaction cost is a relatively large factor when trading Stable token pairs as transactions in them are not information but liquidity driven. In contrast, for NonStable token pairs, trading is in general more active on Uniswap than Binance. Take WBTC-ETH as an example. Its average daily trading volume is about 31644 ETH on Uniswap, much larger than 2023 ETH on Binance.

In Table 2, we further report summary statistics of the execution price, gas price and trade size of Uniswap trades for our six sample token pairs. First, the average trade size of a Uniswap trade is fairly large and is about 4,360 USDT (≈ 4,360 USD), 8.59 ETH (≈ 7,661 USD), and 23.07 ETH (≈ 20,577 USD) for USDC-USDT, ETH-USDT, and WBTC-ETH respectively. Second,

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.

(a) Stable token pairs. Trading volume is denominated in thousand USDT.

Pair		N	Mean	SD	Min	Med	Max
USDC-USDT	TradingVolume-Uniswap	85	3426	1747	646	3456	7577
	TradeCount-Uniswap	85	997	397	504	884	3085
	TradingVolume-Binance	85	96681	59806	22936	82593	276362
	TradeCount-Binance	85	51403	21583	15724	47647	116379
DAI-USDT	TradingVolume-Uniswap	85	1494	1361	56	1155	5830
	TradeCount-Uniswap	85	658	403	174	570	2068
	TradingVolume-Binance	85	11575	10451	2224	9210	77831
	TradeCount-Binance	85	9174	7925	1525	7341	58558

(b) NonStable token pairs. Trading volume is denominated in ETH.

Pair		N	Mean	SD	Min	Med	Max
ETH-USDT	TradingVolume-Uniswap	85	73489	37752	36923	62131	263356
	TradeCount-Uniswap	85	8560	1700	6311	8155	16419
	TradingVolume-Binance	85	1444426	709203	493012	1281734	4245010
	TradeCount-Binance	85	994231	524099	272746	915584	2577496
WBTC-ETH	TradingVolume-Uniswap	85	31644	17748	9014	27141	87965
	TradeCount-Uniswap	85	1371	592	646	1127	3338
	TradingVolume-Binance	85	2023	1993	135	1258	9984
	TradeCount-Binance	85	7886	7529	289	5332	35191
LINK-ETH	TradingVolume-Uniswap	85	10779	6295	3437	9406	42520
	TradeCount-Uniswap	85	1054	380	574	961	2682
	TradingVolume-Binance	85	4387	2687	1071	3856	13598
	TradeCount-Binance	85	10459	6793	2223	9391	29514
AAVE-ETH	TradingVolume-Uniswap	85	7368	4177	1766	6366	29936
	TradeCount-Uniswap	85	609	253	261	551	1514
	TradingVolume-Binance	85	2135	1510	408	1627	10143
	TradeCount-Binance	85	6829	5410	1131	5511	36964

the gas price attached to Uniswap trades varies considerably across 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 change in the overall congestion of the Ethereum network, or from traders' bidding high fees to trade on the information.

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.

(a) Stable token pairs. Transaction size is denominated in thousand USDT.

TokenPair	Variable	N	Mean	SD	1%	10%	Median	90%	99%
USDC-USDT	TxPrice	84779	1.00	0.00	0.99	1.00	1.00	1.00	1.01
	GasPrice	84779	90.91	81.52	16.00	33.00	71.00	164.00	400.00
	TxSize	84779	3.43	8.51	0.01	0.11	1.04	7.66	40.05
DAI-USDT	TxPrice	55919	1.00	0.00	0.99	1.00	1.00	1.01	1.01
	GasPrice	55919	91.08	121.44	18.00	35.00	73.00	155.00	397.82
	TxSize	55919	2.27	5.15	0.01	0.08	0.76	5.05	26.02

(b) NonStable token pairs. 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

4 Methodology

To examine whether DEX trades with a higher fee are more informative, we follow Hasbrouck (1991) and estimate a structural vector-autoregressive (structural VAR) model. In the structural VAR model, we include CEX return and DEX trade flows with different fee levels as endogenous variables. Hence, we can compute the cumulative impulse response of return to a trade flow variable, i.e., its permanent price impact, which is regarded as a measure of its private information content. Barclay, Hendershott, and McCormick (2003) and O'Hara, Yao, and Ye (2014) apply the same approach to examine the informativeness of odd-lot trades versus round-lot trades and ECN trades versus market-maker trades respectively.

4.1 Baseline structural VAR specification

A general structural VAR model can be specified as follows:

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

where $\Phi_1 \ldots \Phi_p$ are standard system matrices of the VAR model. 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 $s \neq t$. Observe that y_t is the endogenous variable vector, and A is the structural matrix capturing the contemporaneous correlations between the endogenous variables. In our specifications below, we specify the endogenous variables included in y_t and the kind of contemporaneous correlations assumed between the endogenous variables.

Our baseline specification for the structural VAR model is as follows:

$$y_t = \begin{pmatrix} r_t^{\text{CEX}} & x_t^{\text{LowFee-DEX}} & x_t^{\text{MidFee-DEX}} & x_t^{\text{HighFee-DEX}} \end{pmatrix}', \qquad A = \begin{pmatrix} 1 & a_{12} & a_{13} & a_{14} \\ 0 & 1 & 0 & 0 \\ 0 & a_{32} & 1 & 0 \\ 0 & a_{42} & a_{43} & 1 \end{pmatrix} \tag{2}$$

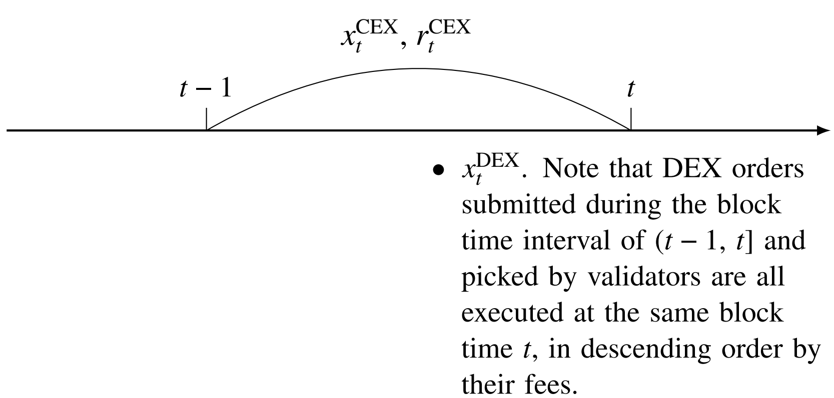
where t indexes block time, and r_t^{CEX} is the Binance midquote return from block time $t-1$ to t. We use $x_t^{\text{LowFee-DEX}}$, $x_t^{\text{MidFee-DEX}}$ and $x_t^{\text{HighFee-DEX}}$ to denote Uniswap trade flows in block t, respectively with low, mid and high blockchain fee levels, as specified in blockchain fee level classification below. Note that we define the trade flow to be the sum of signed trades. For example, the trade flow at a given fee level i in a block t is computed as:

$$x_t^i = \sum_k d_k^i s_k^i. \tag{3}$$

where k indexes trades with fee level i in block t, d_k is the trade direction indicator (+1 for buys and -1 for sells), and s_k is the trade size.

Timestamp convention Because Binance runs a continuous central limit order book while Uniswap executes trades in batches based on block time, it is necessary to specify a timestamp convention that encompasses both Binance returns and Uniswap trade flows. In Figure 1, we provide a visual illustration of the chosen convention. Specifically, r_t^{CEX} is the log difference between the price of the last Binance trade before block time $t-1$ and that of the last Binance trade before block time t.[8] Then all 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 in batch at block time t. Next, we detail our timestamp convention and our blockchain fee-level classification scheme.

Figure 1. Timestamp convention. This figure illustrates our time convention. t is block time. r_t^{CEX} is the log return from Binance defined over the time interval between $t-1$ and t. Note that we do not have quote updates from Binance. The return is calculated based on trade prices, not midquotes. x_t^{CEX} is the trade flow on Binance obtained by summing the trades executed between block time $t-1$ and t (See Equation 3). x_t^{DEX} is the signed trade flow on Uniswap at block time t.



Blockchain fee level classification We specify the classification scheme used in the structural VAR specification above, where we include DEX trade flows with three different levels of gas price, the blockchain fee traders bid on the Ethereum network. We adopt a rolling-window approach to calculate the benchmark blockchain fee. Specifically, to classify trades in the current block t, we

[8] Note that we do not have quote updates, but only trades data, from Binance. Hence, the return is calculated based on trade prices, and not midquotes.

first sort them together with all trades within the last 20 non-empty blocks[9], i.e., block $t-20$ to block $t-1$ based on their blockchain fee in descending order. Then trades in block t which fall in the top quartile (i.e., above 75% quantile) are labeled as high-fee trades, $x_t^{\text{HighFee-DEX}}$; Trades which fall in the bottom quartile (i.e., below 25% quantile) are labeled as low-fee trades, $x_t^{\text{LowFee-DEX}}$; All other trades are labeled as mid-fee trades (i.e., between 25% and 75% quantile), $x_t^{\text{MidFee-DEX}}$. Hence, for each block, we construct three DEX trade flows. For some blocks, we might only have observations for one or two types of trades, in which case trade flows for the remaining type(s) are set to zero.

Resolution of the contemporaneous correlations Last, the structural matrix A is specified to only allow for the following contemporaneous relations: (1) all three DEX trade flow variables cause CEX return but not vice versa[10]; (2) low-fee DEX trade flow causes mid-fee and high-fee DEX trade flows; (3) mid-fee DEX trade flow causes high-fee DEX trade flow. We impose such a recursive structure so to obtain a lower bound of the price impact of high-fee DEX trade flow.[11]

Permanent price impact of trade flows After estimating the structural VAR model, we can easily obtain the vector moving average (VMA) representation to compute the impulse responses of return and trade variables to shocks in the structural innovations:

$$y_t = \Theta(L)\varepsilon_t = \Theta_0\varepsilon_t + \Theta_1\varepsilon_{t-1} + \Theta_2\varepsilon_{t-2} + \cdots \tag{4}$$

[9]As a robustness check, we have repeated the classification based on trades within the last 10 or 40 blocks instead, and all results are qualitatively the same. In Appendix A.1, we report the structural VAR results based on alternative window lengths.

[10]Such restriction is normally made in empirical market microstructure literature: trades/trade flows are assumed to affect return contemporaneously but not vice versa. The economic intuition is that market makers will only revise their quotes after seeing the incoming trades and updating their beliefs about the new equilibrium price. We refer to Hasbrouck (1991) for detailed explanations.

[11]The economic intuition is as follows. Including contemporaneous low-mid and mid-mid DEX trade flows in the equation of high-fee DEX trade flow equation means that there will possibly be some positive correlation between the trade flows. A similar approach is used in, for example, O'Hara, Yao, and Ye (2014) where they assume odd-lot trades are caused by round-lot and mixed-lot trades in one of their specifications to obtain a lower bound of the price impact of odd-lot trades.

where $\Theta(L)$ is the polynomial of the lag operator $\Theta(L) = \Theta_0 + \Theta_1 L + \Theta_2 L^2 + \cdots$. 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,

$$\mathrm{PPI}_k = \frac{\sum_{j=0}^{\infty} \partial r_{t+j}^{\mathrm{CEX}}}{\partial \varepsilon_{k,t}} = [\Theta(1)]_{1,k}, \quad k > 1 \tag{5}$$

where $[\Theta(1)]_{1,k}$ denotes the $(1,k)$-*th* entry of $\Theta(1)$, the impulse response of return to trade flow variable k.

Information share of trade flows In addition to permanent price impact, we can compute the information shares of trade flow variables via the approach of random walk decomposition (See Hasbrouck, 1991, for detailed proofs). The information share measure weighs the permanent price impact of the CEX return or a trade flow variable $[\Theta(1)]_{1,k}$ by its own structural innovation variance, $\sigma_{\varepsilon_k}^2$. Hence, if two trade flow variables have the same permanent price impact, the one with a larger innovation variance will have a larger information share. Formally, the information share (IS) of the CEX return or a trade flow variable k to price discovery is computed as:

$$\mathrm{IS}_k = \frac{[\Theta(1)]_{1,k}^2 \sigma_{\varepsilon_k}^2}{\sum_k [\Theta(1)]_{1,k}^2 \sigma_{\varepsilon_k}^2} \tag{6}$$

Implementation details We implement the structural VAR estimation in the following ways: (1) the model is estimated at block-by-block frequency, although the blockchain 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[12]; (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 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.

[12]In Appendix A.2, 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.

5 Empirical results

In what follows, we present the main results from our empirical analysis. First, in Section 5.1 we report results from our structural VAR analysis and show our key finding: high-fee DEX trade flow contains more private information than low-fee DEX trade flow. Then, in Section 5.2, we conduct several robustness checks for our key finding. Last, in Section 5.3, we provide plausible economic channels behind our key finding.

5.1 Priority fees and price discovery

Below we examine whether blockchain fees play an important role in the price discovery process through DEX trade flows. To do so, we estimate a structural VAR model as in Equation 2 where we include CEX return and DEX trade flows with different fee levels. In Section 5.1.1, we provide summary statistics of the CEX return, DEX trade flows, and CEX trade flow, variables used in the structural VAR model. We then analyze the impulse response analysis and report the permanent price impact and information shares of DEX trade flows in Section 5.1.2 and Section 5.1.3 respectively. In Section 5.1.4, we examine the speed of price discovery through DEX trade flows.

5.1.1 Summary statistics of CEX return and DEX trade flows

Before discussing the estimation results from the structural VAR model, for each token pair, we report summary statistics of the return and trade flow variables in Table 3. Several observations are in order. First, as expected, returns of NonStable token pairs are much more volatile. For instance, per-block-time ($\approx$12 seconds) standard deviation of Binance return, r_t^{CEX}, is about 0.79, 10.27, and 9.12 basis points for USDC-USDT, ETH-USDT and WBTC-ETH respectively. These results are expected because NonStable pairs consist of risky tokens such as Bitcoin and Ethereum and thus their prices respond to both short-term liquidity shocks and long-term information shocks. In contrast, Stable token pairs are only affected by short-term liquidity shocks as both of their tokens

Table 3. 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 point. N refers to the number of blocks for each token pair during our sample period.

(a) Stable token pairs. All trade flow variables are denominated in thousand USD.

		N	Mean	SD	Min	50%	Max
USDC-USDT	r_t^{CEX}	66949	-0.00	0.79	-52.68	0.00	37.12
	x_t^{CEX}	66949	1.58	112.29	-3892.72	0.00	3277.79
	x_t^{DEX}	66949	0.01	7.95	-268.57	-0.02	227.86
	$x_t^{\text{LowFee-DEX}}$	66949	0.02	2.62	-118.09	0.00	85.00
	$x_t^{\text{MidFee-DEX}}$	66949	0.03	4.82	-145.88	0.00	150.00
	$x_t^{\text{HighFee-DEX}}$	66949	-0.04	5.75	-268.57	0.00	227.86
DAI-USDT	r_t^{CEX}	45868	-0.00	1.45	-46.36	0.00	40.56
	x_t^{CEX}	45868	-1.20	33.79	-1162.58	0.00	778.95
	x_t^{DEX}	45868	0.01	5.14	-142.20	-0.00	141.16
	$x_t^{\text{LowFee-DEX}}$	45868	0.02	1.90	-60.01	0.00	50.65
	$x_t^{\text{MidFee-DEX}}$	45868	-0.00	3.12	-81.14	0.00	64.00
	$x_t^{\text{HighFee-DEX}}$	45868	-0.01	3.55	-142.20	0.00	94.93

are pegged to the US Dollar.

Second, consistent with the liquidity summary statistics in Table 1, the magnitude of trade flows on Uniswap versus Binance differs significantly across token pairs. For the two Stable token pairs, USDC-USDT and DAI-USDT, the magnitude of trade flow is much larger on Binance than on Uniswap. For example, the standard deviation of per-block-time trade flow of USDC-USDT on Binance is about 112 thousand USD, more than an order of magnitude larger than that of about eight thousand USD on Uniswap. In contrast, for the rest of the token pairs except for ETH-USDT, 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.

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

(b) NonStable token pairs. 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

5.1.2 Permanent price impacts of DEX trade flows

If DEX trade flows with high fees contain more private information than those with low fees, we should expect the former to have a larger permanent price impact. 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 5. In Table 4, we report the cumulative impulse responses of CEX return to DEX trade flows with different fee levels. The results show that high-fee DEX trade flow has a larger permanent price impact and thus contains more private information than low-fee DEX trade flow.

We begin by discussing the results for pairs of the NonStable token pairs. These token pairs consist of at least one non-stable coin, and should thus experience frequent private information shocks. The results show that the cumulative impulse response of CEX return, r_t^{CEX}, to high-fee DEX trade flow, $x_t^{\text{HighFee-DEX}}$, is statistically significant. In addition, the cumulative impulse responses of CEX return to high-fee DEX trade flow are much larger in magnitude compared with the responses to mid- and low-fee DEX trade flows. A positive shock to high-fee DEX trade flow equal to one standard deviation of that flow leads to an 8.16 basis points increase in CEX return while a similar size shock applied to the low-fee trade flow leads to a much smaller increase of 0.83 basis points.

Next, we discuss the estimation results for pairs of Stable tokens, which are consistent with our expectations. Stable token pairs carry little private or public information because both tokens of the pair are stable coins pegged to the US Dollar. Hence, without short-term liquidity shocks, token pairs should always be priced at one. As a result, traders of Stable pairs are either liquidity traders who would like to exchange one stablecoin for the other or arbitrageurs who respond to public information such as transitory price discrepancy of the token pairs between CEXs and DEXs. Both types of trades can only impose a transitory impact on the prices, but not a permanent one. These results are consistent with intuition: the cumulative impulse responses of DEX trade flows are statistically insignificant, regardless of fee levels.

In addition, the results show that, for NonStable token pairs, the cumulative impulse responses of DEX trade flows to CEX return shocks are statistically significant, and their magnitudes increase with fees. For instance, a one-basis point shock to CEX returns of NonStable token pairs leads to an increase of high-fee DEX trade flow of about 0.03 standard deviations, which is much larger than the response of medium-fee and low-fee DEX trade flow. While CEX price can quickly adjust through quote revisions upon the release of new public information. In most DEXs prices are determined by a pre-coded pricing function and cannot be immediately adjusted. Rather, they can only be updated through trades executed in subsequent blocks. The large response of high-fee DEX trade flow to the CEX price shocks help update the DEX price towards the efficient price.

Table 4. Cumulative impulse responses between CEX return and DEX trade flows with different blockchain fee levels. This table reports the impulse responses between the CEX return and DEX trade flows with different blockchain fee levels, cumulative over 20 blocks. Impulse responses are obtained by estimating the structural VAR model specified in Equation 2. The estimation is done for each pair-day and statistical inference is based on pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis point. DEX trade flows are standardized and thus in their standard deviations. *, ** and *** indicate significance levels at 1%, 5% and 10% respectively.

PairType	Variable	r^{CEX}	$x^{\text{LowFee-DEX}}$	$x^{\text{MidFee-DEX}}$	$x^{\text{HighFee-DEX}}$
Stable	r^{CEX}	0.8***	0.01	0.0	0.0
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{LowFee-DEX}}$	0.01	0.96***	−0.03***	−0.03***
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.01	−0.08***	0.89***	−0.08***
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	−0.01	−0.08***	−0.2***	0.81***
		(0.01)	(0.01)	(0.01)	(0.01)
NonStable	r^{CEX}	1.06***	0.83***	3.8***	8.16***
		(0.01)	(0.12)	(0.27)	(0.37)
	$x^{\text{LowFee-DEX}}$	0.0***	1.01***	−0.02***	−0.02***
		(0.0)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.01***	−0.05***	1.02***	0.05***
		(0.0)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	0.03***	−0.09***	−0.18***	1.15***
		(0.0)	(0.01)	(0.01)	(0.01)

5.1.3 Information shares of DEX trade flows

Table 5. Information shares of DEX trade flows with different gas price levels. This table reports the information shares of the CEX return and DEX trade flows with different blockchain fee levels. Information shares are computed using the formula in Equation 6. The estimation is done for each pair-day and statistical inference is based on pair-day estimates. Numbers in brackets are standard errors.

PairType Variable	Stable	NonStable
r^{CEX}	97.28	77.09
	(0.24)	(0.79)
$x^{\text{LowFee-DEX}}$	1.05	0.75
	(0.14)	(0.06)
$x^{\text{MidFee-DEX}}$	1.0	4.09
	(0.14)	(0.29)
$x^{\text{HighFee-DEX}}$	0.66	18.06
	(0.1)	(0.67)

In Table 5, we compute the information share of DEX trade flows associated with different fee levels. The information share approach considers both the permanent price impact of a trade flow variable and its own (unexpected) variance (See Equation 6). For example, if two trade flow variables have the same permanent price impact, the one with a larger (unexpected) variance will have a larger information share.

The results show that, for NonStable token pairs, CEX return contributes the largest share to price discovery, which reflects public information, and high-fee DEX trade flow has a much larger information share than low-fee DEX trade flow: the high-fee DEX trade flow contributes about 18.06% to price discovery, which is much larger than 4.09% of the mid-fee trade flow, and 0.75% of the low-fee DEX trade flow. In contrast, for Stable token pairs, CEX return itself contributes to almost all (97.28%) price discovery, and DEX trade flows contain barely any private information.

5.1.4 DEX trade flows and speed of price discovery

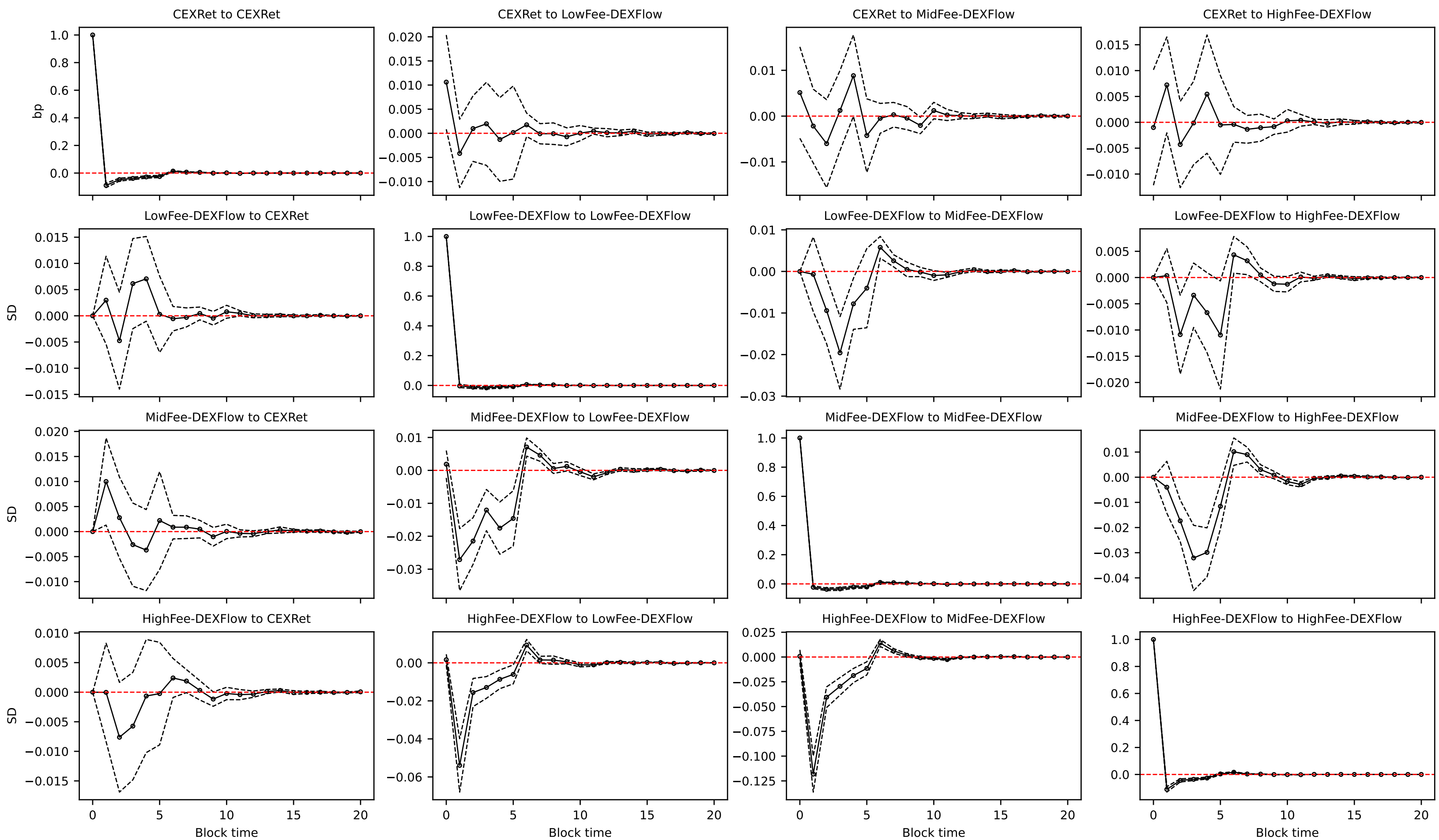
In previous sections, we have shown that high-fee DEX trade flow has a much larger permanent price impact than low-fee DEX trade flow, contributing more to price discovery. However, as the permanent price impact is defined as the cumulative impulse responses of CEX return, it can not speak to the speed of price discovery. How quickly does the CEX price adjust to the private information revealed through DEX trade flows? To examine it, we turn to the dynamics of impulse responses of CEX return to DEX trade flows.

Panel (a) of Figure 2 indicates that the impulse responses of CEX return to DEX trades for Stable token pairs are statistically insignificant over all periods, regardless of the bid fees. These results are consistent with the cumulative return impulse responses tabulated in Table 4. Moreover, the impulse responses of DEX trade flows to CEX returns, are statistically insignificant for all periods and all fee groups.

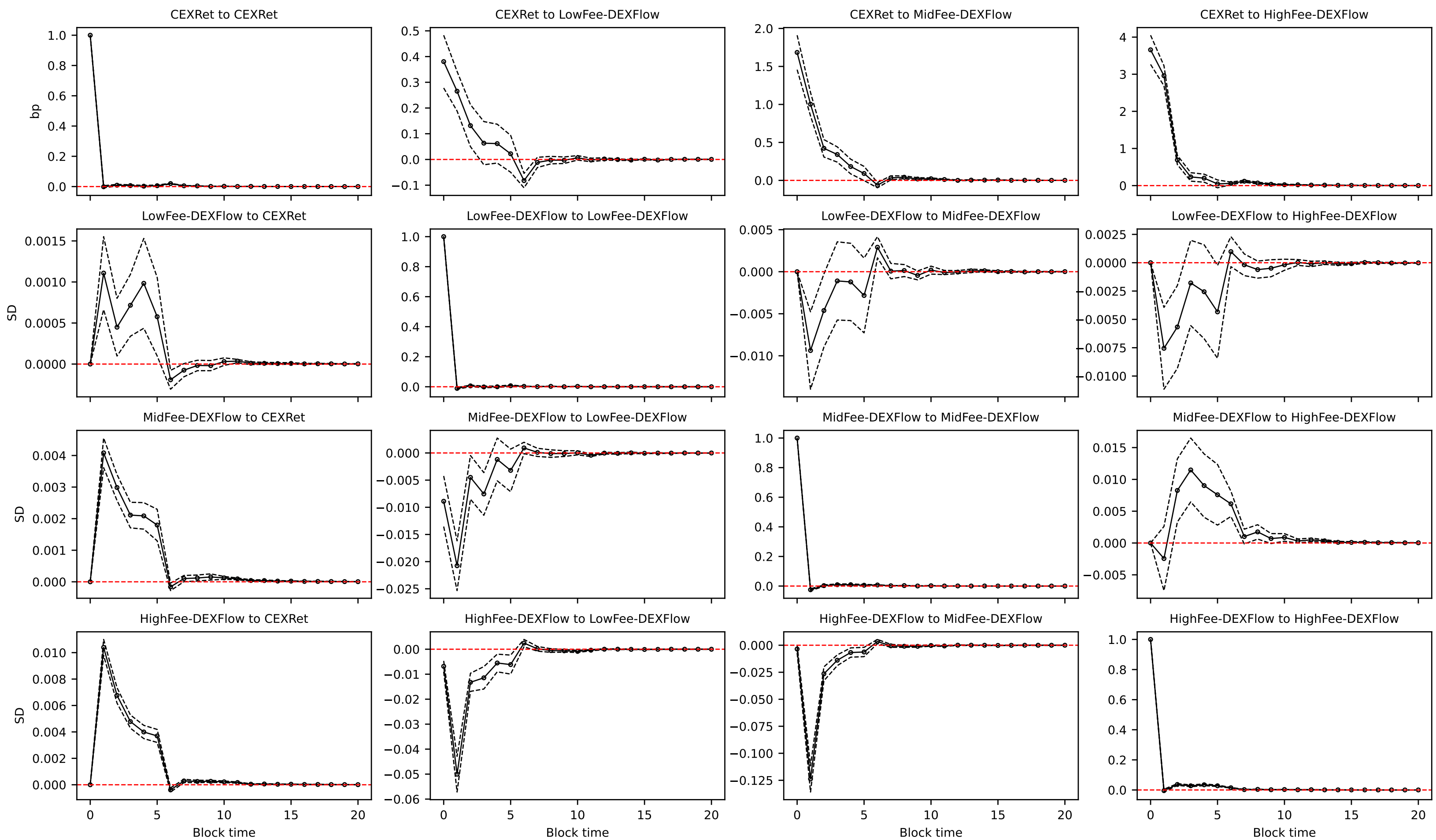
In Panel (b) of Figure 2, we plot the impulse responses for NonStable token pairs. While the return impulse response is significant and large in the contemporaneous ($t = 0$) and the next block ($t = 1$), it becomes much smaller in magnitude from the second block ($t = 2$) on. It indicates that CEX return responds significantly and quickly to DEX trade flows and most price discovery through DEX trade flow is realized within two blocks. Hence, traders are able to learn the private information contained in the high-fee DEX trade flow quickly and update their beliefs on the new price. In contrast, the impulse responses of high-fee and mid-fee trade flows to CEX returns are significant and large for about five blocks ($t = 1$ to $t = 5$). This suggests that the response of DEX trade flows to public information is more sticky and takes several blocks of time.

Figure 2. Impulse response functions between CEX return and DEX trade flows with different fee levels. This figure plots the impulse responses between the CEX return and DEX trade flows with different fee levels over a horizon of 20 blocks. Impulse responses are obtained by estimating the structural VAR model specified in Equation 2. CEX return is measured in basis points and DEX trade flows are standardized and thus measured in standard deviation units. We perform the estimation for each pair-day, and the statistical inference is based on pair-day estimates. Dashed black lines represent 95% confidence bands.

(a) Stable token pairs.



(b) NonStable token pairs.



5.2 Robustness checks

5.2.1 Robustness: Accounting for CEX trade flow

Traders execute their trades both on centralized and decentralized exchanges. To control for cross-venue arbitrage trades between Uniswap and Binance, we include Binance trade flows in the endogenous variable vector and consider the robustness of the results against the following alternative specification:

$$y_t = \begin{pmatrix} r_t^{\text{CEX}} & x_t^{\text{CEX}} & x_t^{\text{LowFee-DEX}} & x_t^{\text{MidFee-DEX}} & x_t^{\text{HighFee-DEX}} \end{pmatrix}', \quad A = \begin{pmatrix} 1 & a_{12} & a_{13} & a_{14} & a_{15} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & a_{32} & 1 & 0 & 0 \\ 0 & a_{42} & a_{43} & 1 & 0 \\ 0 & a_{52} & a_{53} & a_{54} & 1 \end{pmatrix} \tag{7}$$

x_t^{CEX} is the signed trade flow on Binance aggregated between block time $t-1$ and t. The variables r_t^{CEX}, $x_t^{\text{LowFee-DEX}}$, $x_t^{\text{MidFee-DEX}}$ and $x_t^{\text{HighFee-DEX}}$ have all been introduced above, and we recall that they represent CEX return and Uniswap trade flows with low, mid and high gas fee levels respectively.

Similarly, we specify the structural matrix A in such a way that we only allow CEX trade flow to contemporaneously affect DEX trade flows, but not vice versa. In terms of economics, we impose such structural restrictions in order to control informed traders splitting their trades on both CEXs and DEXs. The idea is intuitive: assuming traders trade their private information on both CEXs and DEXs, we should expect CEX and DEX trade flows to be highly correlated. Hence, after the CEX trade flow is controlled, if CEX prices still respond to DEX trade flows, it must be the case that the DEX trade flows contain other private information. In addition, as in the first specification, we maintain the following assumptions on the contemporaneous relations: (1) low-fee DEX trade flow causes mid-fee and high-fee DEX trade flows; (3) mid-fee DEX trade flow causes high-fee DEX trade flow.

Table 6 and 7 indicate that the results remain qualitatively the same as the baseline model: high-fee DEX trade flow has a much larger permanent price impact and information share than mid- and low-fee DEX trade flows. Adding CEX trade flow only slightly reduces the economic magnitude of the permanent price impacts of high-fee DEX trade flow. For example, for NonStable token pairs, the cumulative CEX return impulse response to one positive standard deviation shock in high-fee DEX trade flow is 7.76 basis points when CEX trade flow is controlled for, which is marginally smaller than 8.16 basis points when CEX trade flow is not controlled for. In addition, the permanent price impact of high-fee DEX trade flow is larger than that of CEX trade flow: for NonStable token pairs, one standard deviation of a positive shock to the high-fee DEX trade flow leads to an increase of about 4.38 basis points in the CEX price, compared to one standard deviation of a positive shock to the high-fee DEX trade flow which yields an increase of 7.76 basis points.

In summary, high-fee DEX trade flow continues to have a large permanent impact on CEX prices even after controlling for CEX trade flow, which strengthens our claim that DEX trade flow captures private information not contained in CEX trade flow.

5.2.2 Robustness: Resolution of the contemporaneous correlations between CEX return and DEX trade flows

In our baseline specification for the structural VAR model, we assume DEX trade flows contemporaneously affect CEX return, but not vice versa. Accordingly, we specify the causality matrix A

Table 6. Cumulative impulse responses between CEX return, CEX trade flow, and DEX trade flows with different gas prices. This table reports the impulse responses between the CEX return, CEX trade flow, and DEX trade flows with different gas price levels, cumulative over 20 blocks. Impulse responses are obtained by estimating the structural VAR model specified in Equation 9. The estimation is done for each pair-day and statistical inference is based on pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis point. CEX trade flow and DEX trade flows are standardized and thus in their standard deviations. *, ** and *** indicate significance levels at 1%, 5% and 10% respectively.

PairType	Variable	r^{CEX}	x^{CEX}	$x^{\text{LowFee-DEX}}$	$x^{\text{MidFee-DEX}}$	$x^{\text{HighFee-DEX}}$
Stable	r^{CEX}	0.72***	0.37***	0.01	0.0	0.0
		(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
	x^{CEX}	−0.23***	1.21***	0.0	0.01	0.01
		(0.03)	(0.02)	(0.01)	(0.01)	(0.01)
	$x^{\text{LowFee-DEX}}$	0.02*	0.0	0.96***	−0.04***	−0.03***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.0	0.0	−0.08***	0.89***	−0.08***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	−0.02**	0.0	−0.08***	−0.2***	0.81***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
NonStable	r^{CEX}	1.03***	4.38***	0.78***	3.64***	7.76***
		(0.01)	(0.23)	(0.12)	(0.26)	(0.37)
	x^{CEX}	0.01***	1.24***	−0.01	0.03***	0.17***
		(0.0)	(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{LowFee-DEX}}$	0.0***	0.02***	1.01***	−0.02***	−0.02***
		(0.0)	(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.01***	0.09***	−0.05***	1.02***	0.04***
		(0.0)	(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	0.03***	0.22***	−0.09***	−0.19***	1.14***
		(0.0)	(0.01)	(0.01)	(0.01)	(0.01)

as below:

$$y_t = \left(r_t^{\text{CEX}} \quad x_t^{\text{LowFee-DEX}} \quad x_t^{\text{MidFee-DEX}} \quad x_t^{\text{HighFee-DEX}} \right)', \qquad A = \begin{pmatrix} 1 & a_{12} & a_{13} & a_{14} \\ 0 & 1 & 0 & 0 \\ 0 & a_{32} & 1 & 0 \\ 0 & a_{42} & a_{43} & 1 \end{pmatrix}. \tag{8}$$

However, the causality can go in the other direction. For example, market makers for derivatives need to constantly execute hedging trades in the underlying security in response to price

Table 7. Robustness: Information shares of CEX trade flow and DEX trade flows with different blockchain fee levels. This table reports the information shares of the CEX return, CEX trade flow, and DEX trade flows with different gas price levels. Information shares are computed based on Equation 6. The estimation is done for each pair-day and statistical inference is based on pair-day estimates. Numbers in brackets are standard errors.

PairType Variable	Stable	NonStable
r^{CEX}	70.5	67.34
	(1.33)	(0.94)
x^{CEX}	26.96	12.51
	(1.35)	(0.83)
$x^{\text{LowFee-DEX}}$	0.97	0.73
	(0.13)	(0.06)
$x^{\text{MidFee-DEX}}$	0.93	3.77
	(0.14)	(0.28)
$x^{\text{HighFee-DEX}}$	0.63	15.65
	(0.09)	(0.64)

shocks. Besides, CEX return and DEX trade flows are both defined at block-time frequency, which on average lasts 12 seconds on the Ethereum blockchain. Thus, the timestamp is too coarse to decide the right ordering between the CEX return and DEX trade flows, which might result in a large contemporaneous correlation between the two. In Table 8, we report the contemporaneous correlations between CEX return and DEX trade flows from the estimation results of the reduced-form VAR model. It shows that while the correlations between CEX return and low-fee and medium-fee DEX trade flow are rather small, that between CEX return and high-fee DEX trade flow are moderate at about 0.18 for NonStable and token pairs.

Given that the magnitude of the correlation between CEX return and high-fee DEX trade flow is not negligible, changing the ordering between the two can affect the estimation results of the structural VAR model. To assess the sensitivity of our results to the ordering, we flip the causality assumption and allow CEX return to contemporaneously affect DEX trade flows but not vice versa. Specifically, we impose the causality matrix A as follows and re-implement the structural VAR

Table 8. Contemporaneous correlations between CEX return and DEX trade flows. This table reports the contemporaneous correlations between the CEX return and DEX trade flows. We obtain the contemporaneous correlations by estimating the correlation matrix of the innovation terms from the reduced-form VAR. We estimate the reduce-form VAR for each token pair and day pair and report the average correlations by token pair type.

Pair	Variable	r^{CEX}	$x^{\text{LowFee-DEX}}$	$x^{\text{MidFee-DEX}}$	$x^{\text{HighFee-DEX}}$
Stable	r^{CEX}	1.00	0.01	0.00	-0.00
	$x^{\text{LowFee-DEX}}$	0.01	1.00	0.00	0.00
	$x^{\text{MidFee-DEX}}$	0.00	0.00	1.00	0.00
	$x^{\text{HighFee-DEX}}$	-0.00	0.00	0.00	1.00
NonStable	r^{CEX}	1.00	0.01	0.07	0.18
	$x^{\text{LowFee-DEX}}$	0.01	1.00	-0.01	-0.01
	$x^{\text{MidFee-DEX}}$	0.07	-0.01	1.00	-0.00
	$x^{\text{HighFee-DEX}}$	0.18	-0.01	-0.00	1.00

estimation.

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 \\ a_{31} & a_{32} & 1 & 0 \\ a_{41} & a_{42} & a_{43} & 1 \end{pmatrix} \tag{9}$$

Table 4 reports the cumulative impulse responses between CEX return and DEX trade flows under the alternative causality matrix A. There are two key observations. First, the magnitude of the cumulative impulse responses of CEX return to DEX trade flows drops for all three fee levels. Such a drop is expected as we now impose the restriction that the contemporaneous price impact of DEX trade flows is zero. In reality, traders can monitor pending orders in mempool and adjust their quotes on CEXs in real time. So what we obtain here are the lower bounds of the price impact of DEX trade flows. Second, although the overall magnitude of DEX trade flows' price impact drops, our key results hold. For NonStable token pairs, the price impact of high-fee DEX trade flow remains larger than that of medium-fee and low-fee DEX trade flows.

Table 9. Cumulative impulse responses between CEX return and DEX trade flows with different blockchain fee levels. Alternative causality matrix A. This table reports the impulse responses between the CEX return and DEX trade flows with different blockchain fee levels, cumulative over 20 blocks. Impulse responses are obtained by estimating the structural VAR model specified in Equation 2. The estimation is done for each pair-day and statistical inference is based on pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis point. DEX trade flows are standardized and thus in their standard deviations. *, ** and *** indicate significance levels at 1%, 5% and 10% respectively.

PairType	Variable	r^{CEX}	$x^{\text{LowFee-DEX}}$	$x^{\text{MidFee-DEX}}$	$x^{\text{HighFee-DEX}}$
Stable	r^{CEX}	0.8***	0.0	0.0	0.01
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{LowFee-DEX}}$	0.02	0.96***	−0.03***	−0.03***
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.01	−0.08***	0.89***	−0.08***
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	−0.02*	−0.08***	−0.2***	0.81***
		(0.01)	(0.01)	(0.01)	(0.01)
NonStable	r^{CEX}	1.13***	0.41***	1.94***	4.41***
		(0.01)	(0.1)	(0.17)	(0.22)
	$x^{\text{LowFee-DEX}}$	0.0***	1.01***	−0.02***	−0.03***
		(0.0)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.02***	−0.05***	1.0***	0.01
		(0.0)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	0.05***	−0.1***	−0.23***	1.08***
		(0.0)	(0.01)	(0.01)	(0.01)

5.2.3 Robustness: Controlling for the confounding effect of trade size

Blockchain fee is a fixed cost regardless of the trade size. Thus, traders are willing to pay a higher blockchain fee for large trades as it is relatively cheaper. So, trade size and 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 alleviate the concern, we further partition DEX trades based on their size in addition to fees and examine whether, within the same size group, trades with higher fees have a larger price impact. Specifically, we classify DEX trades into two size groups: a large-size group consisting of trades with a size above its 90% quantile and a small-size group consisting of trades with a size below its 90% quantile. Thus, our large-size group captures very large trades on the right tail of the size distribution. We choose 90% quantile as the cutoff point so that for the large-size group, the average trade size is similar across our three fee levels as shown in Table 10. For some token pairs such as WBTC-ETH, LINK-ETH, and AAVE-ETH, within the large-trade group, the average trade size of the mid-fee group is even larger than the high-fee group.

Table 10. Average trade size by trade size group and blockchain fee level. N refers to the number of trades in our sample.

	GasPriceLevel	LowFee	MidFee	HighFee
TokenPair	TxSizeLevel			
USDC-USDT	Below Q90(TxSize)	1.23	1.48	2.01
	Above Q90(TxSize)	17.55	19.93	21.72
DAI-USDT	Below Q90(TxSize)	0.87	1.02	1.31
	Above Q90(TxSize)	12.19	12.87	13.90
ETH-USDT	Below Q90(TxSize)	1.65	2.12	3.16
	Above Q90(TxSize)	55.44	63.57	68.60
WBTC-ETH	Below Q90(TxSize)	5.79	9.27	17.04
	Above Q90(TxSize)	148.94	151.65	126.60
LINK-ETH	Below Q90(TxSize)	2.86	4.63	9.31
	Above Q90(TxSize)	50.90	62.22	49.73
AAVE-ETH	Below Q90(TxSize)	3.74	6.86	12.67
	Above Q90(TxSize)	55.44	58.84	48.02

Based on our size and fee grouping above, we construct six DEX trade flows: small-size and low-fee DEX trade flow ($x^{\text{S-L-DEX}}$), small-size and medium-fee DEX trade flow ($x^{\text{L-M-DEX}}$), small-size and high-fee DEX trade flow ($x^{\text{L-H-DEX}}$), large-size and low-fee DEX trade flow ($x^{\text{L-L-DEX}}$), large-size and medium-fee DEX trade flow ($x^{\text{L-M-DEX}}$), and large-size and high-fee DEX trade flow ($x^{\text{L-H-DEX}}$). Then we estimate a structural VAR model based on the six DEX trade flows.

Table 11 reports the cumulative impulse responses between the CEX return and the six DEX trade flows by trade size and fee level. Focusing on the NonStable token pairs, the results show that, consistent with the literature, large trades contain more private information. We see that DEX trade flows in the large-trade group have larger price impacts than flows in the small-trade group. More importantly, the results further show that, within the same trade size group, high-fee DEX trade flow has a larger price impact than medium-fee and low-fee flows. It is worth noting that the difference between the price impact of high-fee and low-fee DEX trade flows is more pronounced for the large-trade group, which reflects the positive interaction effect between blockchain fees and trade size. In summary, the above results show that blockchain fee contains additional information content not captured by the trade size.

Table 11. Cumulative impulse responses between CEX return and DEX trade flows by trade size and fee level. This table reports the cumulative impulse responses between CEX return and DEX trade flows by trade size and fee level. Impulse responses are obtained by estimating the structural VAR model. The estimation is done for each pair-day and statistical inference is based on pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis point. CEX trade flow and DEX trade flows are standardized and thus in their standard deviations. *, ** and *** indicate significance levels at 1%, 5% and 10% respectively.

PairType	Variable	r^{CEX}	$x^{S\text{-}L\text{-}DEX}$	$x^{S\text{-}M\text{-}DEX}$	$x^{S\text{-}H\text{-}DEX}$	$x^{L\text{-}L\text{-}DEX}$	$x^{L\text{-}M\text{-}DEX}$	$x^{L\text{-}H\text{-}DEX}$
Stable	r^{CEX}	0.79***	−0.01	0.0	0.01	0.01	0.0	0.01
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	$x^{S\text{-}L\text{-}DEX}$	0.01	1.07***	0.06***	0.02***	−0.09***	−0.06***	−0.04***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	$x^{S\text{-}M\text{-}DEX}$	0.01	0.0	1.08***	0.09***	−0.07***	−0.13***	−0.11***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	$x^{S\text{-}H\text{-}DEX}$	0.0	−0.01	0.02***	1.04***	−0.06***	−0.08***	−0.13***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	$x^{L\text{-}L\text{-}DEX}$	0.01	0.01	0.02**	0.0	0.97***	−0.04***	−0.03***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	$x^{L\text{-}M\text{-}DEX}$	0.0	0.01	0.0	0.02***	−0.07***	0.9***	−0.07***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	$x^{L\text{-}H\text{-}DEX}$	−0.01*	−0.01**	−0.02***	−0.01	−0.08***	−0.21***	0.83***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
NonStable	r^{CEX}	1.06***	0.1	0.59***	1.51***	0.81***	3.7***	7.82***
		(0.01)	(0.12)	(0.11)	(0.16)	(0.12)	(0.27)	(0.38)
	$x^{S\text{-}L\text{-}DEX}$	0.0***	1.04***	0.03***	0.01***	−0.03***	−0.04***	−0.03***
		(0.0)	(0.01)	(0.0)	(0.0)	(0.01)	(0.01)	(0.0)
	$x^{S\text{-}M\text{-}DEX}$	0.01***	0.02***	1.09***	0.06***	−0.03***	−0.06***	−0.05***
		(0.0)	(0.01)	(0.01)	(0.01)	(0.0)	(0.01)	(0.01)
	$x^{S\text{-}H\text{-}DEX}$	0.01***	0.01*	0.03***	1.1***	−0.03***	−0.07***	−0.03***
		(0.0)	(0.0)	(0.01)	(0.01)	(0.0)	(0.01)	(0.01)
	$x^{L\text{-}L\text{-}DEX}$	0.0***	−0.03***	0.02***	0.0	1.02***	−0.01***	−0.02***
		(0.0)	(0.01)	(0.0)	(0.0)	(0.01)	(0.01)	(0.0)
	$x^{L\text{-}M\text{-}DEX}$	0.01***	−0.01***	−0.03***	0.03***	−0.04***	1.02***	0.04***
		(0.0)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	$x^{L\text{-}H\text{-}DEX}$	0.03***	0.01	0.02***	0.08***	−0.09***	−0.18***	1.12***
		(0.0)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

5.3 Blockchain fees and information: Economic channels

In the above section, we have shown that high-fee DEX trade flow contains more private information than low-fee DEX trade flow, suggesting that privately informed traders bid high fees to execute their orders on DEXs. Next, we provide plausible economic channels to explain the results and use mempool order data to test them.

5.3.1 Two potential economic channels

Channel #1: Execution risk due to blockchain congestion Trading on DEXs is not the only activity on a blockchain. Other non-DEX activities such as payment transfer, borrowing and lending, non-fungible token (NFTs) auctions, and initial coin offerings (ICOs) take up limited block space as well. In particular, if there is a surge of non-DEX activities which make blocks congested, the marginal blockchain fee needed to execute a transaction increases, driving up the transaction cost for traders on DEXs.

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.[13] 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.

What this implies is that, if an informed trader only faces execution risk, she will choose a high, but not too high blockchain priority fee for her trades, compared with other transactions in the same block. In terms of the block position, her trades will likely be located around the middle

[13]We note that impatient and uninformed traders (e.g., liquidity traders who receive marginal calls and have to liquidate their positions) can bid high blockchain 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.

of the block, but not at the very top.

Channel #2: Competition among informed traders An informed trader will bid a high fee if the blockchain network is congested. However, this may not be the only channel; she might bid a high fee if she faces competition from other traders.

It is unclear ex-ante whether such a channel exists as theoretical literature has mixed predictions about informed trading and its implications on price discovery. Competition arises when private information is not only possessed by one informed trader; instead, there are multiple 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).

Another possibility is that there are "back-runners" (Yang and Zhu, 2020) or "predators" (Brunnermeier, 2005) who are not endowed with private signals but infer them from public signals such as order imbalance or blockchain fees in the context of DEXs. However, informed traders can select the timing of their trades. For example, they might trade when the liquidity of the target token pairs is high such that their trades lead to a lower price impact and can not be easily detected (Collin-Dufresne and Fos, 2015).

When facing competition from other traders with the same or similar information, an informed trader might have to bid a blockchain fee much higher than the rest of non-DEX transactions in the same block, especially when the potential profit from the information is high. In such cases, we might observe DEX trades with excessively high fees and located at the very top of the block.

5.3.2 Do privately informed traders compete on DEXs?

Identify "excessively-high-fee trades" As explained above, competition among informed traders can lead to excessively high blockchain 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 blockchain fee, we use the inter-quartile range (IQR) method, a commonly used outlier detection approach in statistics.[14] Specifically, for each block, we first calculate the 25% quantile (Q25) and 75% quantile (Q75) of the blockchain fees of all executed transactions in the block[15], 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, IQR = Q75 − Q25. Finally, we obtain the threshold Q75 + 1.5 × IQR and label DEX trades with a blockchain fee higher than the threshold as "excessively-high-fee trades".[16]

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.

To examine whether our identified trades include the first type of trades with private information, we reconstruct DEX trade flows with different blockchain fee levels excluding all "excessively-high-fee" trades and then re-implement the structural VAR analysis. The idea is that if a significant share of high-fee trades result from competition among privately informed traders, we should see their permanent price impact become significantly smaller in magnitude after we exclude the "excessively-high-fee trades".

Table 12 reports the cumulative impulse responses between the CEX return and DEX trade flows when "excessively-high-fee trades" are excluded. It shows that, compared with the baseline

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

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

[16]Alternatively, one can identify such trades based on their block position. As transactions executed in the same block are ranked based on their blockchain fees in descending order. Thus, transactions with higher blockchain 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.

results where all trades are included in Table 4, the cumulative impulse response of CEX return to a unit standard deviation shock to high-fee DEX trade flow drops significantly in magnitude from 8.16 to 5.36 basis points for NonStable token pairs. The results illustrate that our key results—high-fee DEX trade flow is more privately informed —are driven by competition among privately informed traders in addition to them avoiding execution risk.

Table 12. Cumulative impulse responses between CEX return and DEX trade flows: Excluding "excessively-high-fee trades". This table reports the impulse responses between the return and trade flow variables, cumulative over 20 blocks. Impulse responses are obtained by estimating the structural VAR model specified in Equation 2. The estimation is done for each pair-day and statistical inference is based on pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis point. DEX trade flows are standardized and thus in their standard deviations. *, ** and *** indicate significance levels at 1%, 5% and 10% respectively.

PairType	Variable	r^{CEX}	$x^{\text{LowFee-DEX}}$	$x^{\text{MidFee-DEX}}$	$x^{\text{HighFee-DEX}}$
Stable	r^{CEX}	0.79***	0.01	0.0	0.0
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{LowFee-DEX}}$	0.01	0.96***	−0.04***	−0.03***
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.01	−0.08***	0.88***	−0.06***
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	−0.02**	−0.05***	−0.15***	0.88***
		(0.01)	(0.01)	(0.01)	(0.01)
NonStable	r^{CEX}	1.07***	0.95***	3.94***	5.36***
		(0.01)	(0.14)	(0.28)	(0.31)
	$x^{\text{LowFee-DEX}}$	0.0***	1.01***	−0.02***	−0.01***
		(0.0)	(0.01)	(0.01)	(0.0)
	$x^{\text{MidFee-DEX}}$	0.01***	−0.04***	1.02***	0.03***
		(0.0)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	0.02***	−0.05***	−0.07***	1.06***
		(0.0)	(0.01)	(0.01)	(0.01)

5.3.3 How do informed traders compete on DEXs?

In the above section, we have shown that competition among privately informed traders on DEXs is a significant driving force of our key results: high-fee DEX trade flow contains more private information. Next, we investigate what fee bidding strategy informed traders use to compete with each other on DEXs.

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 is that informed traders competitively bid up their blockchain 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 two matched mempool orders 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 guarantee 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 gas fee competition.

3. *All matched orders of the executed trade must arrive at the mempool between block time $t-1$ and t.* Specifically, to be matched with a trade executed at block time t, orders must arrive in the mempool during the block time interval of $(t-1, t]$. 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 13 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"). There are two notable observations. First, the overall fraction of executed trades identified as PGA trades is very small. For example, for the group of "Other trades", less than 1% of them are identified as PGA trades across the six token pairs.

Table 13. 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
USDC-USDT	Other trades	99.95	0.05
	Excessively-high-fee trades	99.40	0.60
DAI-USDT	Other trades	99.95	0.05
	Excessively-high-fee trades	99.45	0.55
ETH-USDT	Other trades	99.87	0.13
	Excessively-high-fee trades	96.05	3.95
LINK-ETH	Other trades	99.35	0.65
	Excessively-high-fee trades	90.24	9.76
WBTC-ETH	Other trades	99.74	0.26
	Excessively-high-fee trades	95.15	4.85
AAVE-ETH	Other trades	99.46	0.54
	Excessively-high-fee trades	93.55	6.45

Surprisingly, even if we zoom in on the "excessively-high-fee trades" which include trades likely driven by competition, only a minority of them are identified as PGA trades. Across the six token pairs, the fraction of PGA trades out of "excessively-high-fee trades" varies between 0.55% for DAI-USDT and 9.76% for LINK-ETH. The result suggests that PGA type of bidding strategy is not the dominant one used by informed traders. Instead of competitively bidding up the fee, they start with bidding a very high fee, which resembles the jump bidding strategy in auction theory (Daniel and Hirshleifer, 1998; Avery, 1998).

The motivation for adopting such a bidding strategy is that, by bidding a high fee in the first place, an informed trader can discourage competition from other traders. First, by bidding a high fee, she can signal that her valuation of the information is high, and if bidding is costly, it is optimal for potential competitors to drop out. Second, even if all traders value the information the same 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 one to a greater Winner's Curse.

6 Conclusion

Decentralized exchanges (DEXs) have gained a significant market share in crypto trading since their inception. Unlike centralized exchanges (CEXs) which continuously execute incoming transactions based on their arrival time, DEXs process transactions in batches and prioritize their executions based on fees bid by users. Thus, the blockchain fee is an important choice variable for informed traders on DEXs.

In this paper, we study the price discovery process on DEXs. Using a structural VAR model, we show that, compared with low-fee trades, high-fee trades reveal more private information. We further test possible economic channels using a unique data set of Ethereum mempool orders. We find that informed traders not only bid high fees to avoid execution risk arising from blockchain congestion, but also to compete with each other. In addition, we show that informed traders compete with each other by following a jump bid strategy.

A Other robustness checks

In this appendix, we conduct two other robustness checks.

A.1 Blockchain fee level classification

In our baseline structural VAR specification, we include (signed) DEX trade flows with high, mid, and low-fee levels respectively. Specifically, high-fee (low-fee) DEX trade flow is computed based on trades with a gas price above 75% (below 25%) quantile of the gas prices of all trades in the past 20 blocks on a rolling window basis.[17] On the one hand, a too-short window makes our quantile estimates noisy. For example, if we only use trades in the current block to implement the classification, two trades with very similar gas prices will fall into different categories. On the other hand, a too-long window might include trades with gas prices too distant to reflect the current congestion level of the blockchain. Thus, we set the window length to 20 in the baseline results to strike a balance.

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 A1 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 4.

A.2 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 included in the structural VAR specification. Table A2 report the return impulse responses when the number of lags is set to 10 and 20 respectively. It shows that the results are qualitatively the same as the

[17] See Section 4 for details of our classification scheme.

Table A1. Cumulative impulse responses between CEX return and DEX trade flows: Gas price level classification based on a rolling window of alternative numbers of blocks. This table reports the impulse responses between the CEX return and DEX trade flows with different gas price levels based on the alternative classification rule, cumulative over 20 blocks. Impulse responses are obtained by estimating the structural VAR model specified in Equation 2. The estimation is done for each pair-day and statistical inference is based on pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis point. DEX trade flows are standardized and thus in their standard deviations. *, ** and *** indicate significance levels at 1%, 5% and 10% respectively.

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

PairType	Variable	r^{CEX}	$x^{\text{LowFee-DEX}}$	$x^{\text{MidFee-DEX}}$	$x^{\text{HighFee-DEX}}$
Stable	r^{CEX}	0.8***	0.01	0.0	0.01
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{LowFee-DEX}}$	0.01	0.98***	−0.03***	−0.02***
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.01	−0.08***	0.89***	−0.1***
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	−0.01	−0.11***	−0.2***	0.82***
		(0.01)	(0.01)	(0.01)	(0.01)
NonStable	r^{CEX}	1.06***	0.96***	4.0***	8.04***
		(0.01)	(0.14)	(0.26)	(0.37)
	$x^{\text{LowFee-DEX}}$	0.0***	1.0***	−0.01**	−0.02***
		(0.0)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.02***	−0.05***	1.01***	0.08***
		(0.0)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	0.03***	−0.09***	−0.17***	1.13***
		(0.0)	(0.01)	(0.01)	(0.01)

baseline results.

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

PairType	Variable	r^{CEX}	$x^{\text{LowFee-DEX}}$	$x^{\text{MidFee-DEX}}$	$x^{\text{HighFee-DEX}}$
Stable	r^{CEX}	0.79***	0.01	−0.01*	0.01**
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{LowFee-DEX}}$	0.01	0.95***	−0.05***	−0.02***
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.02*	−0.08***	0.89***	−0.06***
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	−0.01	−0.07***	−0.19***	0.79***
		(0.01)	(0.01)	(0.01)	(0.01)
NonStable	r^{CEX}	1.06***	0.97***	3.53***	8.19***
		(0.01)	(0.13)	(0.25)	(0.38)
	$x^{\text{LowFee-DEX}}$	0.0***	1.01***	−0.02***	−0.02***
		(0.0)	(0.01)	(0.01)	(0.0)
	$x^{\text{MidFee-DEX}}$	0.01***	−0.05***	1.02***	0.03***
		(0.0)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	0.03***	−0.08***	−0.18***	1.16***
		(0.0)	(0.01)	(0.01)	(0.01)

Table A2. Cumulative impulse responses of CEX return and DEX trade flows with different gas price levels: Alternative number of lags in the structural VAR specification. This table reports the impulse responses between the CEX return and DEX trade flow variables based on alternative numbers of lags included in the structural VAR estimation, cumulative over 20 blocks. Impulse responses are obtained by estimating the structural VAR model specified in Equation 2. The estimation is done for each pair-day and statistical inference is based on pair-day estimates. Row variables are response variables and column variables are shock variables. CEX return is in basis point. 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.

PairType	Variable	r^{CEX}	$x^{\text{LowFee-DEX}}$	$x^{\text{MidFee-DEX}}$	$x^{\text{HighFee-DEX}}$
Stable	r^{CEX}	0.69***	0.02***	0.01	−0.01
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{LowFee-DEX}}$	0.0	0.95***	−0.08***	−0.08***
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.01	−0.13***	0.83***	−0.16***
		(0.01)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	−0.02*	−0.11***	−0.27***	0.75***
		(0.01)	(0.01)	(0.01)	(0.01)
NonStable	r^{CEX}	1.03***	1.04***	4.15***	8.69***
		(0.01)	(0.17)	(0.3)	(0.41)
	$x^{\text{LowFee-DEX}}$	0.0***	1.02***	−0.02**	−0.05***
		(0.0)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.02***	−0.06***	1.02***	0.06***
		(0.0)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	0.04***	−0.11***	−0.23***	1.17***
		(0.0)	(0.01)	(0.01)	(0.01)

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

PairType	Variable	r^{CEX}	$x^{\text{LowFee-DEX}}$	$x^{\text{MidFee-DEX}}$	$x^{\text{HighFee-DEX}}$
Stable	r^{CEX}	0.54***	0.04***	0.01	−0.01
		(0.02)	(0.01)	(0.01)	(0.01)
	$x^{\text{LowFee-DEX}}$	0.0	0.92***	−0.14***	−0.17***
		(0.02)	(0.02)	(0.02)	(0.02)
	$x^{\text{MidFee-DEX}}$	0.01	−0.18***	0.75***	−0.24***
		(0.02)	(0.02)	(0.01)	(0.02)
	$x^{\text{HighFee-DEX}}$	0.0	−0.14***	−0.33***	0.68***
		(0.02)	(0.01)	(0.02)	(0.02)
NonStable	r^{CEX}	0.98***	1.24***	4.15***	8.7***
		(0.01)	(0.23)	(0.33)	(0.45)
	$x^{\text{LowFee-DEX}}$	0.0***	1.02***	−0.02*	−0.06***
		(0.0)	(0.01)	(0.01)	(0.01)
	$x^{\text{MidFee-DEX}}$	0.03***	−0.05***	1.0***	0.06***
		(0.0)	(0.01)	(0.01)	(0.01)
	$x^{\text{HighFee-DEX}}$	0.06***	−0.14***	−0.29***	1.13***
		(0.0)	(0.01)	(0.02)	(0.01)

References

Aoyagi, Jun and Yuki Ito (Mar. 2021). *Coexisting Exchange Platforms: Limit Order Books and Automated Market Makers*. SSRN Scholarly Paper. Rochester, NY.

Avery, Christopher (Apr. 1998). "Strategic Jump Bidding in English Auctions". In: *Rev Econ Studies* 65.2, pp. 185–210.

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

Barbon, Andrea and Angelo Ranaldo (2021). "On The Quality Of Cryptocurrency Markets: Centralized Versus Decentralized Exchanges". In: *SSRN Journal*.

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

Bloomfield, Robert and Maureen O'Hara (Jan. 1999). "Market Transparency: Who Wins and Who Loses?" In: *Rev. Financ. Stud.* 12.1, pp. 5–35.

Boehmer, Ekkehart, Gideon Saar, and Lei Yu (2005). "Lifting the Veil: An Analysis of Pre-trade Transparency at the NYSE". In: *The Journal of Finance* 60.2, pp. 783–815.

Boehmer, Ekkehart and Juan (Julie) Wu (Feb. 2013). "Short Selling and the Price Discovery Process". In: *Rev. Financ. Stud.* 26.2, pp. 287–322.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (Aug. 2014). "High-Frequency Trading and Price Discovery". In: *Rev. Financ. Stud.* 27.8, pp. 2267–2306.

Brunnermeier, Markus K. (2005). "Information Leakage and Market Efficiency". In: *Rev. Financ. Stud.* 18.2, pp. 417–457.

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

Capponi, Agostino and Ruizhe Jia (Mar. 2021). *The Adoption of Blockchain-based Decentralized Exchanges*. SSRN Scholarly Paper. Rochester, NY.

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

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

Daniel, Kent D. and David A. Hirshleifer (July 1998). *A Theory of Costly Sequential Bidding*. SSRN Scholarly Paper. Rochester, NY.

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

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

Hagströmer, Björn and Albert J. Menkveld (Dec. 2019). “Information Revelation in Decentralized Markets”. In: *The Journal of Finance* 74.6, pp. 2751–2787.

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

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

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

Kyle, Albert S. (Nov. 1985). “Continuous Auctions and Insider Trading”. In: *Econometrica* 53.6, p. 1315.

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

Park, Andreas (2021). “The Conceptual Flaws of Constant Product Automated Market Making”. In: *SSRN Journal*.

Parlour, Christine A. (2021). “Decentralized Exchanges”. In: *SSRN Journal*.

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

Zhu, Haoxiang (Mar. 2014). “Do Dark Pools Harm Price Discovery?” In: *Rev. Financ. Stud.* 27.3, pp. 747–789.