# Beyond Borders: Exploring Maximal Extractable Value ("MEV") and Risks in Cross-Domain CEX-DEX Arbitrage

Colin Chan

April 2024

**Abstract**

Price dislocations exist between assets on decentralized exchanges (DEX) and centralized exchanges (CEX). This paper systematically explores the emerging CEX-DEX arbitrage opportunity and presents a long-time scale empirical characterization of the arbitrageurs' profiles, potential revenues, costs and toxic flows incurred by liquidity providers (LPs). Using a panel regression framework, I investigate the significance of factors that influence the profitability of the strategy and highlight the impacts of these arbitrages on the overall ecosystem. These results provide a preliminary evidence into this emerging opportunity, alongside policy recommendations to facilitate healthy arbitrage flows.

**Keywords:** Ethereum, Arbitrage, Cross Domain, AMM

# Contents

# 1 Introduction

Decentralised Finance (DeFi) has revolutionized how financial transactions are executed - decentralized and transparent without financial intermediaries. With a total value locked[1] of US$66.41B[2], these platforms have garnered mainstream adoption to facilitate the exchange of money. However, the space remains plagued by predatory actors (i.e. searchers) which seek to profit from market inefficiencies. They exploit potential opportunities by manipulating the transaction order within blocks on blockchains such as Ethereum to maximize profits (Wahrstätter et al, 2023), a rent seeking behavior known as Maximal Extractable Value ("MEV"). Qin et al (2021) estimated that the total MEV on Ethereum to be between USD 550 - 650 million since 2020.

In fact, price discrepancies also exist between decentralized exchanges (DEX) on Ethereum (on-chain) and centralized exchange (CEX) in an off-chain environment. This dual-faceted approach (Chiplunkar and Gosselin (2023)) involves: DEX trades on-chain with payments for inclusion in the blocks and parallel trades in a CEX environment. Yet, the opacity of CEX transactions renders this strategy largely unexplored.

Against such a backdrop, this paper conducts one of the first systematic studies on CEX-DEX cross domain arbitrages. By leveraging on-chain data through a set of curated heuristics, the paper has identified 642,062 of these arbitrages in the study period from September 15, 2022 to July 31, 2023.

- **Empirical Measurement**: It reveals a relatively concentrated blockspace market dominated by a small group of market participants, with the emergence of searcher builder integrations which co-operate together to arbitrage these opportunities (eg. searcher 0xa69 - beaverbuild).
- **Strategy Analysis**: Pairing the trades with the CEX price on Binance, this research also finds significant latency advantages and the potential revenues and risks of trading different types of assets for each class of arbitrageurs.
- **Welfare Implications**: By quantifying the profits garnered by arbitrageurs, the study maps the value flow within the DeFi ecosystem. This exploration contextualizes the impacts of these arbitrages on market health and stability, with block builders such as Manta Builder exhibiting highly toxic behavior.

These strategies warrant our attention due to the following reasons. First, following Ethereum's transition from Proof of Work (PoW) to Proof of Stake (PoS) (Ethereum, 2022), the blockchain adopted MEV Boost, (a specific implementation of the Proposer Builder Separation (PBS) mechanism) to delineate the block construction from the

[1]Total Value Locked refers to the total US dollar value of digital assets on a particular blockchain network or decentralized application

[2]According to https://defillama.com, as of February 10, 2023

proposal process. This has led to the unintended centralizing effects on the blockchain where entities established infrastructure to compete for order flow and extract MEV opportunities (Wahrstatter et al. (2023) and Gupta et al. (2023)). It includes preferential arrangements where searchers optimize for latency by aligning with builders, and race to the front to secure priority in their order executions. Thus, it is critical to understand the value flow and interactions between market participants within this new paradigm. Second, this form of arbitrage involves moving assets between different ecosystems. Studying these arbitrages will provide a better understanding on this new set of benefits and risks to the arbitrageurs. Third, the study of CEX-DEX arbitrages can also provide valuable insights for policy decisions to improve the value flow within Ethereum, for both users and LPs.

# 2 Background

This section delves into the core principles of Ethereum, the operating models of DEX on the blockchain, and a review of the pertinent literature.

## 2.1 Understanding Ethereum

Ethereum was conceived in 2013 (Buterin, 2014), and introduced a decentralized, immutable ledger system where records, stored in blocks, are cryptographically linked. Central to Ethereum's innovation is the Ethereum Virtual Machine, which hosts programmable smart contracts that autonomously execute transactions. This technology underpins DEX, which enable decentralized peer-to-peer trading.

To transact on the blockchain, users are required to pay ETH (ether, commonly known as gas), the native token of Ethereum. Following the Ethereum Improvement Proposal (EIP)-1559 in 2021 (Thorn, 2021), the Priority Gas Auction (PGA) was established. This system fostered a competitive environment where users vie to have their transactions processed by paying a base fee that is subsequently burned and a priority fee to incentivize block producers. Additionally, users can execute *coinbase transfers* by directly transferring ETH to a block producer, thereby increasing the likelihood of their transaction's inclusion.

**Proposer Builder Separation Design - MEV Boost**

MEV Boost, developed by Flashbots (2022), is a specific PBS implementation. Figure 1 demonstrates the distributed roles among four key actors:

1. Searchers monitor pending transactions in the mempool and create bundles of profitable transactions. They submit these bundles along with a sealed price bid to a network of builders.

2. Block builders gather these transactions and assemble them into blocks. They then compete against each other and submit bids alongside the blocks to a relay.

3. These builder bids are updated throughout the 12 second block time as the block value increases from the priority fees and MEV opportunities.

4. The builder with the highest bid secures their block's inclusion in the blockchain.

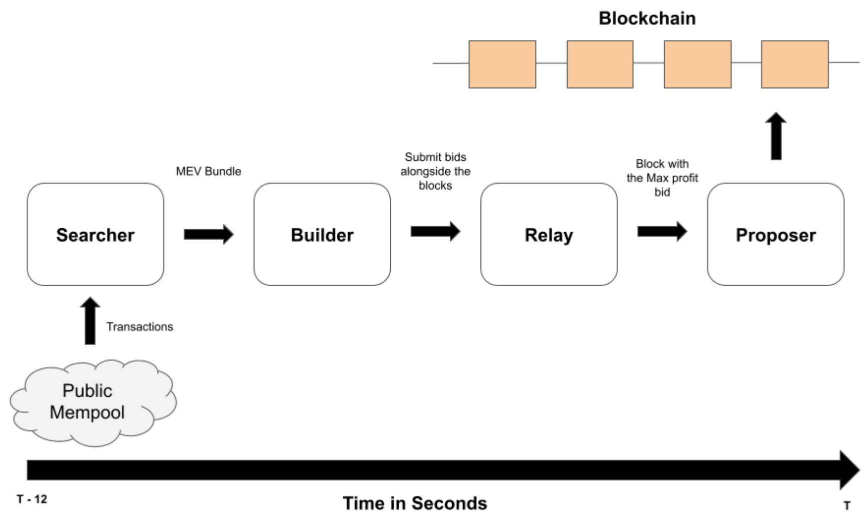



Figure 1: Overview of MEV Boost. Highlights the actors involved to include transactions in a block on Ethereum, over a period of 12 seconds.

## 2.2 Decentralized Exchanges

### Automated Market Making Model

Instead of limit order books, smart contracts are used to facilitate peer-to-peer trading. The Automated Market Making (AMM) model uses liquidity pools (Adams et al, 2020) which holds reserves of 2 or more assets, enabling users to trade against it based on a constant product formula. This determines the market price ($P_{xy} = \frac{y}{x}$) - X and Y are digital assets, while the liquidity pool consists of x units of X and y units of Y.

$$x \cdot y = k \tag{1}$$

where k is the invariant.

Fees are also charged for every transaction in the pool, denoted as f. For each exchange from X to Y, where $\lambda = 1 - f$, the net quantity of X that goes into the swap is $\lambda \Delta x$.

The quantity of Y received from is then mathematically regulated by the above equation. For instance, in a Uniswap liquidity pool containing ETH and USDC, when ETH is purchased, the swapper adds USDC and removes ETH from the pool. This decreases the ETH amount and raises the ETH price due to the constant product k. Each swap generates trading fees, distributed to the LPs of the pool, incentivizing them to add liquidity based on the current exchange rate.

This mechanism provides a fertile ground for arbitrageurs to earn risk-free profits by strategically anticipating the change in pool's reserves as each trade moves the price deterministically. Furthermore, large orders can temporarily dislocate the asset's price from its fair price across exchanges, creating arbitrage opportunities.

**Adverse Selection and Toxic Order Flow**

Order flow is toxic "when it adversely selects market makers, who may be unaware they are providing liquidity at a loss" (Easley et al, 2012). Within AMMs, this phenomenon manifests in a zero-sum scenario where arbitrageurs, armed with more information on the fair value of asset pairs, will out-manoeuvre the LPs. This is mainly due to the stale on the AMM against fluctuating market prices on CEX, giving rise to an information asymmetry that favour arbitrageurs. Milionis et al (2023) described this observation as "loss versus rebalancing" (LVR) which posits that any arbitrage profit—after accounting for swap fees—essentially comes at the LPs' expense, highlighting a fundamental risk associated with providing liquidity in AMMs.

## 2.3 Cross Domain Arbitrage (CEX-DEX )

**Mechanism of CEX-DEX Arbitrages**

- In this identified CEX-DEX arbitrage (0xc43[3]), the arbitrageur swapped 175,070 USDC for 92.70 ETH.
- Based on the timestamp of these transactions, the close price at the 1s interval was queried from leading centralised exchange (Binance). At the time of trade, it can be interpreted that the DEX exchange rate was at 1,888.57 USDC/ETH. On Binance, the approximated rate was at 1,896.68 USDC/ETH.
- As such, this difference in price for the same pair of assets creates an opportunity for arbitrage.

[3]Transaction hash: 0xc4322b4c4f2611dfbb36fa27d7485867d9f1f4f0a26a6988e903a52fd4007db5

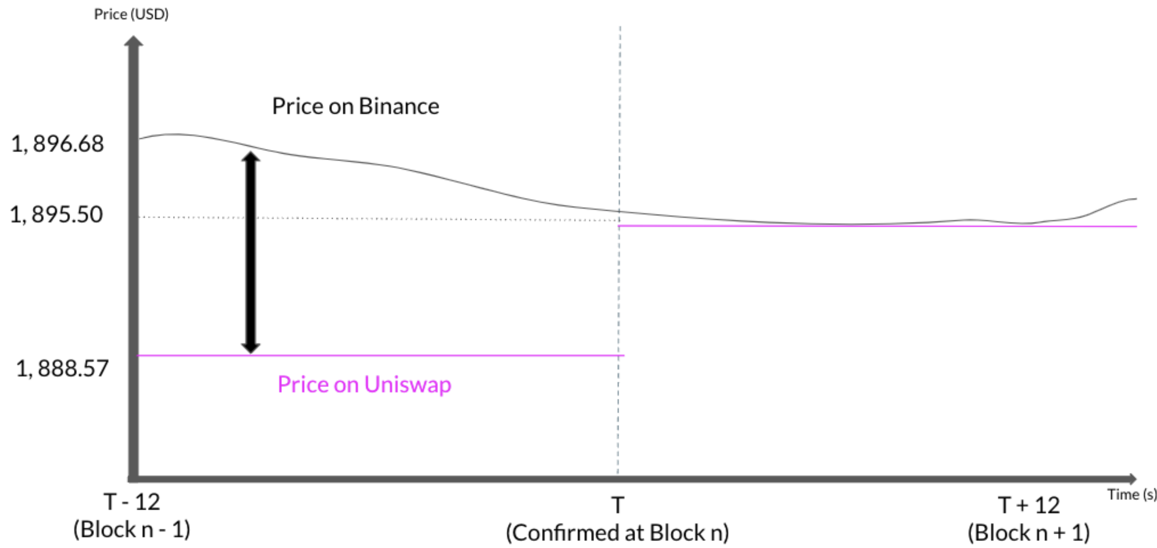



Figure 2: Visualizing Cross Domain Arbitrage, involving ETH and USDC between CEX and DEX over 2 blocks, equivalent of 24 seconds.

In this case, only 1 side of the arbitrage can be observed on Ethereum and it will be presumed that the CEX leg gets executed. In particular, arbitrageurs face predetermined latency risks since transactions across two domains are not atomic. As seen in Figure 2, the searcher who has made a transaction on DEX will have to wait for their transaction to be finalised during this period. This is different from CEX where prices exhibit greater certainty. Thus, searchers strategize submit their bundles to block builders and ensure that their transaction is successfully executed on-chain.

This research thus intersects with execution risks from latency. The 12-second block proposal interval on Ethereum has catalyzed a 'latency arms race', as arbitrageurs vie for advantageous positioning within the block to execute transactions (Crapis, 2023). Thus, the quest for arbitrage profits extends beyond merely identifying price discrepancies but also precise timing to optimally capture profits within the trade window.

Therefore, this paper seeks to complement the studies by highlighting the prevalence of this issue from a quantitative angle and gain a profound understanding of the economic behaviours of these informed searchers and block builders.

# 3 Methodology

To measure the prevalence of CEX-DEX arbitrages, data for both legs of this opportunity were retrieved from September 15, 2022 to July 31, 2023. A set of heuristics was then applied to identify successful CEX-DEX arbitrages based on the on-chain transactions from the AMM trades. Appendix 10.1 outlines the set of data used and explains the logic to query for blockchain data.

**Overview of Heuristics**

- **Scenario 1- CEX DEX Bundle 1**: All legs of this trade are wholly contained within a single transaction (uniquely identified by the "builder_and_dex_hash" ). The searcher made a *coinbase transfer* to the builder and traded on the DEX within the same transaction hash.

  Arbitrage Example 0xc28
  Coinbase Transfer: 0.00064 ETH from MEV Bot `0x98` to Flashbots:Builder
  DEX Transaction: 7,433.90 ID → 6,633.85 USDC on Uniswap V3

- **Scenario 2- CEX DEX Bundle 2**: The trade is broken into separate transactions that occur within the same block (uniquely identified by the 'builder_hash" and "dex_hash"). The searcher made a *coinbase transfer* to the builder in 1 transaction, and made a trade on the DEX in a separate transaction.

  Arbitrage Example Builder Hash 0x0fc, DEX Hash 0xa88
  Coinbase Transfer: 0.0292 ETH from MEV Bot `0x0b3` to Boba Builder
  DEX Transaction: 28.23 ETH → 52,453.13 USDC on Uniswap V3

Thus, an arbitrage cycle typically consists of 3 transactions - 1 swap (2 transactions) and 1 *coinbase transfer*. Transactions containing symbols which were not supported on Binance were removed. This allowed me to retrieve the prices of the relevant assets to calculate the theoretical profits and risks. The final dataset included 642,062 transactions and 402,739 blocks. For a thorough data description and the metrics used, see Appendix 10.1.

**Assumptions**

- **Hedging and Timing**. Identified arbitrageurs mitigate their exposure to price fluctuations by concurrently engaging in trades across both CEX and DEX platforms within the pre-defined window of t - 12 to t + 12.

- **Liquidity**. It is assumed that liquidity on CEX platforms is infinite where transactions are not subjected to slippage. The prevailing price on Binance reflects the fair price of the asset pair as it represents over 50% of global volume (Lee, 2024).

- **Price Calculations**. While Binance supports various cross pairs, their associated order books may exhibit limited depth and lower trading volumes. Consequently, this analysis excludes direct cross pair prices, opting instead to employ USDT as an intermediary for price triangulation, as illustrated in Figure 3.

## 3.1 Data Segmentation

### Types of pairs

- **Market capitalizations**. BTC and ETH are leading cryptocurrencies with significantly higher market capitalizations and liquidity profiles relative to the other digital assets and were thus, classified as the majors.
- **Nature of Asset**. The inherent characteristics of asset, based on its stability or volatility influences its price dynamics within the trading window. This can be segmented into stablecoins, altcoins and memecoins. [4]

Table 1: Classification of Digital Assets Traded

| Types of Assets | Features |
|---|---|
| Stable | Stablecoins - USDC, USDT, BUSD, DAI |
| Major | BTC, ETH |
| Altcoins | All Remaining Altcoins |
| Memecoins | SHIB, PEPE, ELON, MUSK |

### Groups of Searchers

Table 2: Groups of Searchers

| | Searcher Profile | Classification Method | Number |
|---|---|---|---|
| SBE | Searcher Builder Entities | Based on bi-directional relationship in Section 4 | 5 |
| ATE | Active Trading Entities | Executed over 1000 trades, not classified as SBE | 22 |
| MTE | Medium Trading Entities | Executed 101 - 1000 trades | 24 |
| STE | Small Trading Entities | Executed less than 100 trades | 120 |

The searchers were stratified into different arbitrageur profiles (Table 2) based on a specific selection criteria. This aims to illuminate the strategies and preferences across different categories of arbitrageurs. The descriptive statistics of their trading patterns, including distribution, daily transaction counts, and volumes can be found in Appendix 10.5.

## 3.2 Framework

**Builder/Searcher Coverage**. Defined as $SC_{\text{Bi}}$, this metric is formulated as $SC_{\text{Bi}} = \frac{S_{\text{Bi}}}{S_{\text{T}}}$, where $S_{\text{Bi}}$ represents the number of searchers engaging with the ith builder, and

[4] FTT and Luna were removed due to the abormal volatility experienced from the collapse of FTX in November 2022 and the Terra blockchain in May 2022.

$S_T$ is the aggregate number of searchers interacting with that builder (Titan & Frontier Research, 2023).

**Strategy Analysis**

- **Derivation of Expected Dollar Revenue**. Leveraging the price data from Binance, sampled at one-second intervals, allows for the approximation of theoretical arbitrage revenue. The operational framework in Figure 3 illustrates the sequence of transactions aimed at exploiting price discrepancies between CEX and DEX. These transactions, ideally executed in a synchronized manner, commence with the DEX leg and proceed eastward through a series of three steps.

$$\text{Expected Dollar Revenue per arbitrage (ER)} = \left[\frac{q_y \cdot P_{CEX\left(\frac{y}{USDT}\right)}}{P_{CEX\left(\frac{x}{USDT}\right)}} - q_x\right] \cdot P_{CEX\left(\frac{x}{USDT}\right)} \tag{2}$$

*X: token symbol sold on-chain, Y: token symbol bought on-chain*

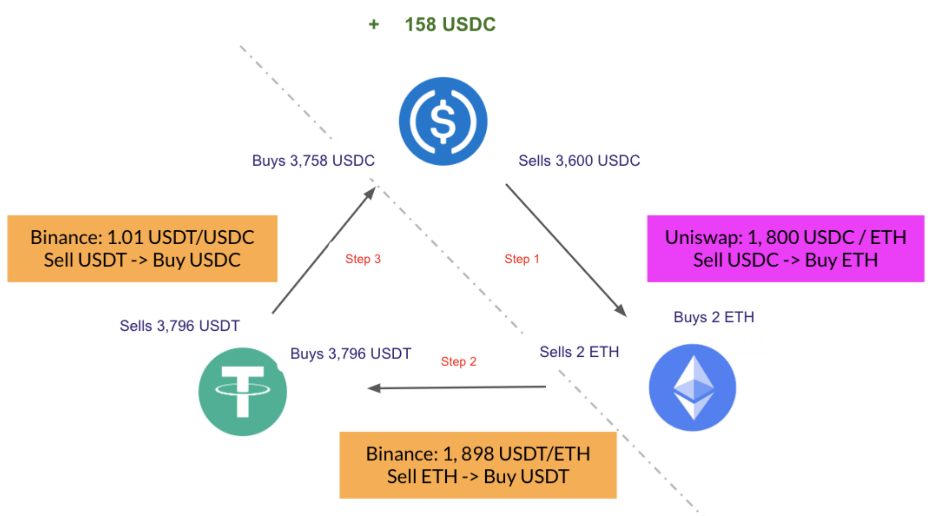



Figure 3: Expected Dollar Revenue, using the ETH - USDC pair with USDT as the intermediate currency.

- **Costs**. This includes variable costs on the network and opportunity costs from latency, which potentially influences the profitability of these strategies.
  - **Gas and *coinbase transfers***. Engaging in trades on DEXs incurs specific costs for arbitrageurs - gas fees, which fluctuate based on Ethereum's network congestion, and *coinbase transfers* paid to block builders.

– **Latency and Opportunity Costs**. Ethereum's 12-second block interval introduces a period of market exposure to price volatility for the arbitrageur. The risks from price fluctuations during the unconfirmed state of the DEX leg include the potential revenue lost due to adverse movements within this trading window. The method compares the realized revenue of each transaction to the hypothetical maximum expected revenue within the window.

$$\mathrm{MR}_i = \max\left(\mathrm{R}_{i,t}\right), \mathrm{LC}_{i,t} = \frac{\mathrm{MR}_i - \mathrm{R}_{i,t}}{\mathrm{R}_{i,t}} \tag{3}$$

where i is the type of asset pair, t is the point in time [0s, 12s]

# 4 Empirical Findings

## 4.1 Prevalence of CEX-DEX Arbitrage

During the measurement period, a predominant portion of CEX-DEX arbitrages was conducted on Uniswap, with 73.61% on v3 and 13.85% on v2. There were also 5 periods of significant increases in the volume traded and number of arbitrages recorded (Figure 4), coinciding with key events. Summary statistics are available in Appendix 10.5.

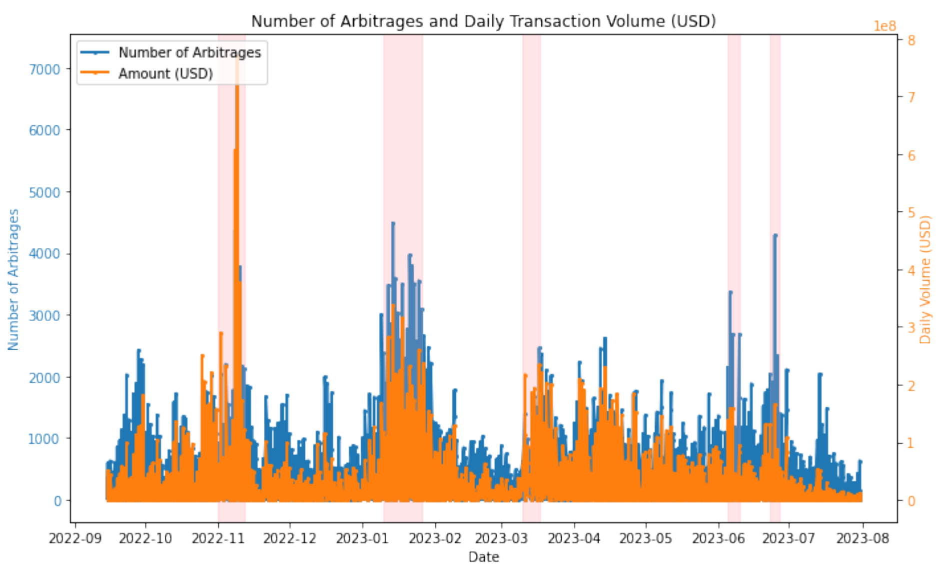


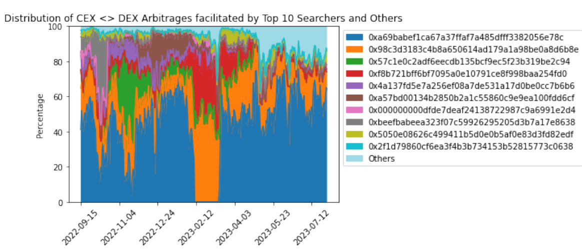

Figure 4: Daily Number of Arbitrages and Volume Recorded. Notably larger volumes and arbitrages were observed during five major events in the study period - (a) FTX Collapse in Nov 2022, (b) Market Recovery in Jan 2023, (c) USDC Depeg in March 2023, (d) SEC lawsuits in June 2023, (e) BTC ETF applications in June 2023. A summary can be found in the Appendix 10.4.

## 4.2 MEV Boost Insights

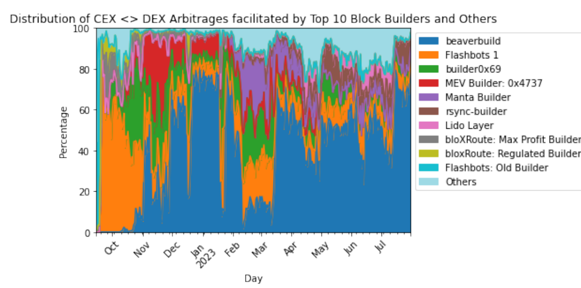
### Distribution of Arbitrages amongst Searchers and Builders

**Searcher Dynamics**: These arbitrage opportunities were highly concentrated around a few key players as the top five searchers collectively contributed to 82.29% of the identified arbitrages. Searcher 0xa69 led this group, participating in 46.81% of all transactions, with searcher 0x98c trailing at 19.45%. A period of inactivity was recorded for searcher 0xa69 between February 10 and March 13, 2023.

**Block Builder Dynamics**: The block building segment was initially dominated by Flashbots, a non-profit organization known for introducing the PBS mechanism. However, October 2022 bucked this trend with the emergence of new builders such as beaverbuild, builder0x69, Manta Builder, and Rsync, rumored to be associated with High Frequency Trading Firms (HFTs) (Gupta et al., 2023). beaverbuild gradually increased its presence, ultimately accounting for 36.31% of blocks built for arbitrages, overtaking Flashbots 1 (18.25%) and builder0x69 (15.53%). Manta-builder, despite an initial surge in market share, discontinued its operations as of May 12th, 2023, coinciding with the scaling back of crypto trading operations by firms like Jane Street and Jump Trading (Doherty and Yang, 2023).




(a) Distribution of Arbitrages amongst Searchers




(b) Distribution of Arbitrages amongst Block Builders

Figure 5: Distribution of Arbitrages Amongst Searchers and Block Builders

### Identifying Searcher-Builder Entities

In Figure 13, the intensity of the blue color represents the proportion of CEX-DEX arbitrages facilitated by each block builder, corresponding to specific searchers.

- **Searchers**: Among the top 10 searchers, a group of them appear to solely interact with certain block builders. For example, searcher 0xf8b directed 100.0% of its CEX-DEX arbitrages to Manta Builder; searcher 0xa69 predominantly engaged with beaverbuild (80.53%), ; searcher 0x57c conducted all of its arbitrages through

MEV Builder: 0x4737[5].

- **Block Builders**: The top 10 block builders, categorized by the number of arbitrages facilitated, showed a distinct pattern of interactions with specific searchers. Notably, MEV Builder: 0x4737 and Manta Builder exclusively facilitated arbitrages for searchers 0x57 and 0xf8b, respectively.

Potential 'searcher-builder' entities were determined using the Searcher Coverage and Builder Coverage (see Section 3.2), based on a bi-directional relationship of 30%. These will be isolated to determine the influence of vertical integration on the profitability.

Table 3: Searcher Builder Entities

| **Searcher** | **block_builder** | **SC (%)** | **BC (%)** |
|---|---|---|---|
| 0xa69babef1ca67a37ffaf7a485dfff3382056e78c | beaverbuild | 71.97 | 80.53 |
| 0x57c1e0c2adf6eecdb135bcf9ec5f23b319be2c94 | MEV Builder: 0x4737 | 85.17 | 100.0 |
| 0xf8b721bff6bf7095a0e10791ce8f998baa254fd0 | Manta Builder | 85.86 | 99.99 |
| 0xbeefbabeea323f07c59926295205d3b7a17e8638 | MEV Builder: 0xb646 | 39.61 | 100.0 |
| 0x1264f83b093abbf840ea80a361988d19c7f5a686 | Boba Builder | 100.0 | 99.12 |

## 4.3 The Economics of Strategies

This section delves into the economics of these opportunities by approximating the theoretical revenue, costs and profits by each class of arbitrageurs.

### 4.3.1 Expected Risk Rewards

To understand the payoff structure for the different classes of asset pairs, Figure ?? displays the expected revenues[6] they received and compared it with the risk [7] throughout the trading window.

- **Meme-Alt Outliers**: The unique performance of meme-alt pairs yielded the highest returns but yet recorded a lower risk. This could be possibly attributed to the relatively small dataset which only consisted of 176 SHIB/MATIC pairs.

- **Market Confidence and Stability**: Stable-stable pairs exhibited the greatest stability while meme-stable pairs recorded the highest risks due to the volatile nature of memecoins. On the contrary, major-stable pairs yielded the lowest expected revenues, which could be possibly explained by the relatively mature market for the majors (BTC, ETH) and thus, smaller relative price differentials exist. In the meanwhile, meme-alts had the highest risk given the large swings in prices from SHIB.

[5]Details on the addresses of the top 10 searchers and block builders are provided in Appendix 10.11
[6]Expected Revenue = Average revenue throughout the trading window of 24 seconds
[7]Risk = Standard Deviation of Revenue throughout the trading window of 24 seconds

Figure 9 shows that MTE and STE, characterized by smaller trading sizes and volumes, often exhibit higher expected revenues. However, this is juxtaposed with occasional losses and more pronounced deviations in spreads. It suggests that these entities might concentrate their trading decisions on specific market opportunities, albeit with fewer overall trades. In contrast, the active arbitrageurs, SBE and ATE, consistently demonstrate a stable level of expected revenues across various asset pairs Figure 13). The narrower profit margins observed in this group can be attributed to the larger number of trades they execute. Their strategy appears to revolve around exploiting market inefficiencies, enabling them to maintain a steady flow of revenue despite the thinner profit margins per trade.

### 4.3.2 Transaction Cost Analysis

The total cost of the transaction includes the variable costs of gas and *coinbase transfer* made by the arbitrageur. Figure 10.8 shows the daily average of the total cost incurred for these arbitrages, with visible spikes in the months of November 2022 (FTX Collapse), March (USDC Depeg) and May 2023 (Memecoin Frenzy with PEPE), coinciding with notable periods of volatility. In general, the overall percentage of the total cost per trade has risen throughout the period, averaging from 50% in September 2022 to 67% in July 2023. In fact, approximately 65.92% of these costs can be attributed to the *coinbase transfers*, which has remained relatively consistent.

- **Total Transaction Costs**: While the SBEs had one of the lowest average costs, it is important to recognise the large percentage of arbitrages captured by the 4 identified pairs of searcher builders. Upon closer examination into the number of daily transactions, these arbitrages were made under very different environments. The top days where the SBE recorded the most number of arbitrages were during the collapse of FTX between November 8 and 11, with the broader recovery observed in mid to end January. In contrast, ATE's arbitrages were focussed on the periods of 23 to 30 September 2022 (start of post-Merge with Flashbots as the preferred avenue) and 21 to 23 January 2023. These were noticeably periods of reduced volatility, with fewer on-chain transactions (Etherscan, n.d).

- **Transaction Costs by Notional Size of Trade**: Transaction costs varied significantly with the notional size of the trade. As the order size becomes larger, *coinbase transfers* contribute a larger percentage of the total costs incurred - from 13.82% for the 1st percentile to as high as 93.76% for the 99th percentile of trade sizes (Figure 15). It means that arbitrageurs are willing to pay a larger percentage of the total costs on *coinbase transfers* to confirm their transactions. This strategic choice is pronounced in scenarios involving larger swaps where the risk of failure to execute on-chain becomes more substantial.

- **Opportunity Costs from Latency**. To highlight the differentials on a daily basis, Figure 16 highlights the costs of latency at 1s, 6s and the maximum of

12s of the trading horizon. As t increases, there exists greater uncertainty between Binance and on-chain prices, resulting in greater uncertainties. It becomes evident that arbitrageurs with lower latency can capture a larger percentage of profits compared to their competitors , enabling them to time their trades accurately and swiftly exploit these discrepancies within the trading horizon. This suggests a nascent marketplace and potentially a concentrated one with builder-searcher entities dominating the majority of all transactions. Thereby, they enjoy significant competitive advantages without significant changes to the overall costs.

It can be seen that there is relatively high and competitive transaction costs as the trade size increases. In particular, *coinbase transfers* become crucial for arbitrageurs to directly interface with block builders and include their transactions on-chain.

### 4.3.3 Profits per Trade

Upon examination of arbitrage transactions, 75.57% yielded positive returns, whereas the remaining 24.43% were loss-making. The extent of profitability exhibited considerable diversity across distinct trading activities. According to Figure 17, a monthly aggregation of the financial performance of each trader category revealed that SBE and ATE consistently accrued substantial profits. In contrast, MTE and STE demonstrated more pronounced fluctuations in their earnings. Specifically, SBE and ATE achieved an average profit per transaction of $7.67 and $11.19 respectively, coupled with a reduced standard deviation, suggesting a higher level of market acumen that facilitates regular arbitrage success.

# 5 Welfare Implications - Toxic Order Flow for LPs

In the context of AMMs, the positive profit realized by each arbitrageur translates into a corresponding loss borne by the LPs. This phenomenon is exacerbated by volatile price movements on Binance since stale prices in liquidity pools are not updated and get instantly arbitraged against. In total, an estimated $27.48 million of losses was incurred by LPs. In fact, Figure 18 elucidates an overarching trend: transactions of greater magnitude lead to heftier losses for LPs. This graphical depiction of trade profitability from the arbitrageurs' viewpoint, against their notional sizes' percentiles, resonates with prior empirical findings by 0xfbifemboy (2023).

**Toxic Flow Analysis by Pair Type**

The analysis unveils a nuanced insight into the profitability of LPs when segmented by pair types. Particularly for pairs involving stablecoins (stable-stable pairs), the data reveals a compelling turning point where larger transactions begin to tilt in favor of LP profitability. This phenomenon can be attributed to the fee structure inherent in these swaps, which, in scenarios of substantial trades, outweighs the nominal price discrepancies typically associated with such low-volatility assets. Consequently, LPs in stable-stable markets enjoy a buffered stance against drastic price movements during volatility, thereby mitigating the risk of toxic flow and adverse selection. This dynamic also facilitates a conducive environment for arbitrageurs to seamlessly arbitrage mispricings in stablecoins across various platforms, further minimizing the potential negative repercussions on LPs.

### Toxic Flow Analysis by Arbitrageur Profile

The trades were segmented according to the profile of the arbitrageurs to visualize their average profits against the logarithmic average notional size. Memecoins and altcoins, known for their volatile risk profiles were filtered out to provide a more accurate reflection of trading behaviors aligned with conventional risk tolerance thresholds. As depicted in Figure 18, distinct clusters can be seen which highlight varying degrees of toxic flow associated with different arbitrageur segments. This segmentation corroborates findings from a seminal analysis conducted on Uniswap V3, which focused on discerning toxic flows within the ETH/USDC pools (0xfbifemboy, 2023), thereby reinforcing the consistency of these observed patterns.

- **Least Toxic Arbitrageurs**. STE which have less than 100 trades, typically engage in transactions of minimal notional size. These entities are more likely to incur losses given that they are the least sophisticated.
- **Toxic Arbitrageurs**. Contrarily, arbitrageurs which made at least 100 trades, particularly MTE and ATE exhibit pronounced toxicity. This group is distinguished by larger average trade sizes, which correlate with increased losses for LPs. The heightened volatility associated with their preference for altcoin pairs exacerbates this effect, introducing greater uncertainty and price volatility. Thereby, this imposes additional financial losses on LPs due to their delayed responses to market fluctuations.

### Toxic Flow by Block Builders

The CEX-DEX arbitrages facilitated by the integrated block builders (of SBE) were isolated to gauge the impact on LPs. The cumulative profits and notional sizes were amalgamated for each block, providing a comprehensive evaluation of the overall toxicity associated with arbitrages within these blocks. The empirical data, as illustrated in Figure 19, revealed a pronounced bimodal distribution, marked by two distinct peaks. Among these, Boba Builder emerged as a notably balanced entity, with its order flow distribution almost symmetrically straddling the toxic and non-toxic thresholds. Conversely, the analysis identified a tilt towards high toxicity in the blocks associated with beaverbuild and Manta builder. Manta builder, in particular, has been previously spotlighted for its adept exploitation of statistical alphas within Uniswap V3 (0xfbifemboy , 2023), underscoring the varying degrees of toxicity facilitated by builders for CEX-DEX arbitrages.

# 6 Factors Influencing Profitability

This section investigates the determinants of the arbitrage's profitability within the CEX-DEX ecosystem. Equation 4 tests the profits earned against the volatility and event periods, with control variables (*Arbitrages, Volume*) used.

$$\begin{aligned}\log(\text{Profit } (\%)_i) =& \beta_1 \cdot \log(\text{Volatility}_i) + \beta_2 \cdot \mathit{Event} + \beta_3 \cdot \log(\text{Arbitrages}_{daily}) \\ &+ \beta_4 \cdot \log(\text{Volume}_{daily}) + \text{error}_{i,t}\end{aligned} \tag{4}$$

where i is the transaction

*Profit (percentage terms)* is the total profits eared by arbitrageurs after accounting for gas costs and *coinbase transfers*, *Volatility* refers to the standard deviation in the price of the token pair during a 24 second trading window (2 blocks), *Event* is an indicator variable equal to 1 if the transaction occurred within the interval for the 5 events identified, *Arbitrages* refers to the number of daily arbitrages, and *Volume* is the total volume of arbitrages per day in USD.

Logarithmic transformation was applied to the profits, daily number and volumes of arbitrages, and volatility of trades. This was used to standardize the distribution of data points and create a uniform scale for analysis to mitigate the impact of outliers.

The coefficients derived from the model quantify the impact of a 1-unit increase in volatility and presence of an event on the logarithmically transformed profits of each arbitrageur profile. The results in Table 6 suggest that the model exhibits a notably high adjusted r-squared value and low standard errors for the variables, indicating a strong fit that effectively explains the majority of the observed variations. A positive coefficient indicates a beneficial effect on profitability, while a negative coefficient suggests a detrimental impact.

For the assets' volatility, the coefficients show a slightly positive relationship with profits for SBE and ATE, a very strong negative relationship for MTE, and a moderately negative relationship for STE. This confirms the intuition that while higher token volatility generally increases profitability for arbitrageurs, the less sophisticated ones (i.e MTE and STE) are the most negatively affected and might incur greater losses. The 'event' variable similarly exerts a stronger direct influence on the SBE and ATE profitability but yet yield highly mixed results for the MTE and STE. In particular, the p-value for the event variable in the latter group exceeds 0.05, suggesting that it might not have a statistically significant impact on their profits. This could be plausibly explained by their higher variance in the success of their trades as seen in their distribution of profits

(Figure 17), despite a lower number of trades executed.

From the perspective of LPs, this also means that they are more likely to face toxic flows and get rapidly arbitraged due to significant price dislocations that present skewed arbitrage opportunities.

**Event vs Token Volatility Risks**

Based on the regression, the difference in absolute values of the variables (event vs volatility) indicate that certain arbitrageurs are able to profit more significantly during event driven uncertainties as compared to others. This could be due to the nature of the arbitrage which entails different risks. The former (event) measures jump risk which typically refers to sudden and significant movements in the price of the assets over a very short time frame. Examples include black swan events such as the collapse of FTX and USDC depeg which introduces unpredictable market reactions and systemic risks. This typically poses unique challenges for arbitrageurs as it can quickly erode profits and induce significant losses if the risks are not managed effectively. On the contrary, normal price risks are the typical volatilities and idiosyncratic fluctuations that can arise from market sentiments. These can typically be well managed by statistical arbitrage strategies to profit from the short term price movements.

Therefore, the abilities of the SBE and ATE to navigate both types of risks and even enjoy a greater profit during sudden price changes, points to their abilities to exploit jump risks and sophisticated risk management systems. The fact that they were more likely to step in to provide liqudiity and trade during these event periods (Figure 9), when other arbitrageurs are likely to withdraw due to fragmented or scarce liquidity, further adds credence to their expertise.

# 7 Policy Recommendations

**Distribution Mechanism**

While MEV boost provided a mechanism to separate builders from proposers, MEV-Share (Flashbots, n.d.) was a new design to return some of the MEV revenue from searchers back to users. In this case, as seen in Section 5 , CEX-DEX arbitrages remain a source of toxic MEV facilitated by integrated searcher-builders, causing LPs to become unprofitable. Therefore, a similar redistribution mechanism can be implemented, by encouraging arbitrageurs to return value to LPs.

- **Protocol Level:** MEV Share introduces a new entity called Matchmaker between the searcher and builder. It acts in the best interests of the user and seek to distribute any additional profits back to the user who made the DEX transaction.

However, this adds another entity to the MEV supply chain and introduces a risk of centralisation.

- **AMM Level:** CoW DAO's new AMM creates 'surplus-capturing batch auctions' with the introduction of 'solvers' (CoW Protocol, 2024). They bid to rebalance AMM pools whenever there is an arbitrage opportunities. The solver which returns the most surplus to the pool wins the right to rebalance the pool and reduces the LVR losses incurred by LPs.

### Reduce Block Time

The average block time of 12 seconds on Ethereum enables a significant amount of revenue earned by arbitrageurs. Further simulations can be conducted to determine the influence of block time on an arbitrageurs' response. For instance, in scenarios with shorter block times, the observed price differentials between CEX and DEX tend to diminish. This reduction in price gaps translates to lower levels of toxic flows—undesirable arbitrage trading activities that can adversely affect LPs.

The opportunity costs associated with latency further highlight this relationship, as depicted in Figure 16. The figure illustrates how varying block times impact arbitrage opportunities and, consequently, the resulting revenue for arbitrageurs. In addition, to achieve the same expected gains, these arbitrageurs may potentially trade more, increasing volume and thus the base fee earned by LPs. However, this requires a more significant change at the consensus level to amend the block time.

### Top of Block vs Rest of Block Opportunities

This novel solution has been suggested previously in academic literature (Gupta et al, 2023) to segment the space within each block into 2 distinct categories - top of block, rest of block. Since every transaction modifies the AMM 's invariant which influences the exchange rate, being earlier in the block minimizes the market risks and exposure to potentially less favourable rates. This is supported by Figure 6 where majority of these transactions tend to be the first few to occur in a block. The separation of blockspace enables a more efficient allocation based on the needs of users - searchers against market participants who are seeking to trade on a DEX.

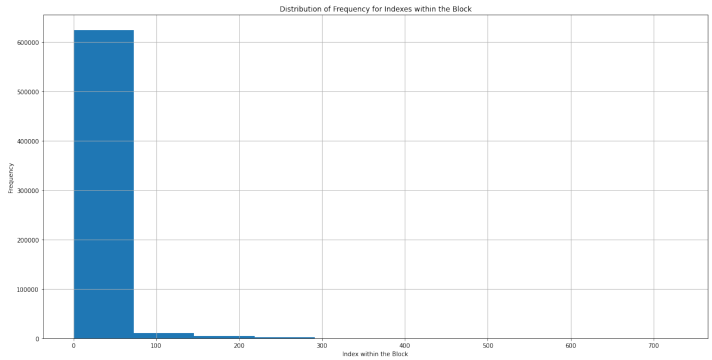



Figure 6: Distribution of Index for DEX Transactions within the block

**Figure 6. Distribution of DEX transaction Index Within the Block**. The histogram shows the position of the transaction made by arbitrageurs within the block.

### Order Size Fee Mechanism

The solution addresses the observed phenomenon where larger orders executed by arbitrageurs, particularly those facilitated by integrated searcher-builder entities, tend to introduce greater toxicity into the liquidity pool. To counteract the potential negative impact of these sizeable trades on LPs and the overall health of the pool, a proposed solution involves imposing an additional tax on arbitrageurs executing orders beyond a certain size threshold based on the pool's microstructures. As it is possible that these large arbitrageurs would simply split their trades into smaller orders instead, this can be fortified with an additional condition by the protocol to aggregate the trading volume per block by the arbitrageur in the pool.

Earlier observations support the notion that excessively large orders contribute significantly to pool toxicity. By implementing this fee mechanism, the DEX aims to discourage arbitrageurs from executing disproportionately large trades, while generating additional revenue that can be directed back into the protocol for LPs.

### Volatility Fee Mechanism

A volatility index can be integrated to provide a dynamic fee mechanism to reflect the degree of price fluctuations within the market. As seen from the model, increased levels of arbitrage activity and toxic flows are associated with periods of elevated volatility. Since this is associated with higher risks for LPs, the dynamic adjustment of transaction fees based on the volatility index aims to provide compensation and incentivize LPs to continue supplying liquidity.

$$\text{Fee} = \text{Base Fee} + (\text{Volatility Factor} \times \text{Volatility Index})$$

- **Base Fee** represents the standard fee that users would pay under normal market conditions.
- **Volatility Factor** determines how much the fee will increase in response to changes in the Volatility Index. A higher Volatility Factor leads to more significant fee adjustments during volatile periods. Therefore, LPs of memecoins and long tail assets can be adequately compensated for the increased risk from market fluctuations.
- **Volatility Index** serves as the dynamic input, reflecting the current level of market volatility. This can be calculated using price oracles from data providers such as Chainlink and Pyth network.

**New AMM Designs: Emergence of RFQ-based exchanges**

RFQ-based exchanges function similarly to OTC trading where arbitrageurs can submit orders to the DEX. Market makers then execute these trades on-chain by behaving as both the LP and searcher as they aggregate liquidity across venues and provide competitive pricing. For instance, Uniswap has recently launched Uniswap X (Adams, 2023) - permissionless Dutch auction that outsources routing to third party fillers which compete to offer the best quotes. As such, RFQ models enable market participants to request and negotiate prices, leading to a reduction in slippage, better alignment with market prices, and promotes tighter spreads. This reduces the overall value leakage from the ecosystem.

# 8 Conclusion

In conclusion, this paper stands at the forefront of exploring the CEX-DEX arbitrage landscape post-Merge. The extensive analysis has unveiled the changing market dynamics of this new domain. For instance, the blockspace ecosystem reveals a narrative of consolidation with key players such as beaverbuild, Manta Builder and Boba Builder actively engaged in CEX-DEX arbitrages. Through a set of metrics, the paper has also dissected the anatomy of profitability, unveiling the fleeting nature of gains and the critical timing for optimal returns. Furthermore, specific searcher-builder entities will consistently contribute to order flow toxicity, providing insights into potential focal points for regulatory interventions in the CEX-DEX arbitrage landscape. The final contribution would be the impacts of volatility and key events on the likelihood of profitability which have shown to be of significance in influencing a strategy's performance.

Future research should aim to broaden the dataset to capture a wider spectrum of arbitrage opportunities. The analysis can be extended to encompass private order flows with multiple private fragmented mempools - searchers send transactions directly to builders and interact through private endpoints, enabling potential arbitrages to go unnoticed. This has emerged as a growing alternative as nearly 10% of all transactions on Ethereum are private orderflows (Wahrstatter, n.d.). Next, this paper assumes these arbitrageurs trade within the 12 second trading horizon defined by the Ethereum block time. However, majority of them are HFTs and are likely trading over a longer window. For instance, Frontier Research (2023) highlighted a markout analysis with periods between 5 mins and as far as 21 days. They estimated the EV for the Uniswap V3 ETH-USDC pool to be approximately $100M in 2022 based on a 5-minute markout price. Thus, the data can be extended to a longer time horizon and determine the variation in expected profits.

Ultimately, the exploration of Ethereum's blockspace markets and HFT behavior opens new avenues for understanding the dynamics of profitability in digital asset markets. Moving forward, the continued investigation into these domains provides a necessary step towards fostering a more efficient, transparent, and equitable financial ecosystem on the blockchain.

# 9 References

## References


[1] 0xfbifemboy. (2022, November 21). Usage of Markout to Calculate LP Profitability in Uniswap V3. Medium. https://crocswap.medium.com/usage-of-markout-to-calculate-lp-profitability-in-uniswap-v3-e32773b1a88e

[2] 0xfbifemboy. (2023, March 10). Discrimination of Toxic Flow in Uniswap V3: Part 4. Medium. https://crocswap.medium.com/discrimination-of-toxic-flow-in-uniswap-v3-part-4-c09656ec016e

[3] Adams, H.(2023, July 17). Introducing the Uniswap X protocol. Uniswap Protocol. https://blog.uniswap.org/uniswapx-protocol

[4] Adams, H., Zinsmeister, N., & Robinson, D. (2020, March). Uniswap v2 Core. https://uniswap.org/whitepaper.pdf

[5] Allison, I. (2022, November 2). Divisions in Sam Bankman-Fried's crypto empire Blur on his trading Titan Alameda's balance sheet. https://www.coindesk.com/business/2022/11/02/divisions-in-sam-bankman-frieds-crypto-empire-blur-on-his-trading-titan-alamedas-b

[6] Adams, H.(2023, July 17). Introducing the Uniswap X protocol. Uniswap Protocol. https://blog.uniswap.org/uniswapx-protocol

[7] Berwick, A. (2023, June 5). Highlights from SEC complaint against Crypto Exchange Binance. Reuters. urlhttps://www.reuters.com/markets/us/highlights-sec-complaint-against-crypto-exchange-binance-2023-06-05/

[8] Bovaird, C. (2023, January 30). Why are crypto markets having such a good January? Why Are Crypto Markets Having Such A Good January? https://www.forbes.com/sites/cbovaird/2023/01/30/why-are-crypto-markets-having-such-a-good-january/?sh=4bd954b83f9a

[9] Buterin, V. (2014). Ethereum: A next-generation smart contract and decentralized application platform. https://ethereum.org/669c9e2e2027310b6b3cdce6e1c52962/Ethereum_Whitepaper_-_Buterin_2014.pdf

[10] Chiplunkar, A., & Gosselin, S. (2023, February 14). A new game in town. https://frontier.tech/a-new-game-in-town

[11] CoW Protocol (2024, February 14). Cow DAO launches the first MEV-capturing AMM. https://blog.cow.fi/cow-dao-launches-the-first-mev-capturing-amm-bc7199e217a3

[12] Crapis, D. (2023, March 4). Latency arms race concerns in blockspace markets. Ethereum Research. https://ethresear.ch/t/latency-arms-race-concerns-in-blockspace-markets/14957

[13] Doherty, K., & Yang, Y. (2023, May 9). Jane Street, jump slow crypto trading as US Regulators crack down. Bloomberg.com. https://www.bloomberg.com/news/articles/2023-05-09/jane-street-jump-pull-back-crypto-trading-amid-us-crackdown#xj4y7vzkg

[14] Easley, D., Lopez, M., & Hara M. (2012, February). Flow Toxicity and Liquidity in a High Frequency World. https://www.stern.nyu.edu/sites/default/files/assets/documents/con_035928.pdf

[15] Etherscan. (n.d.). Etherscan. https://etherscan.io/accounts/label/mev-bot

[16] Ethereum (ETH) Blockchain Explorer. (n.d.). https://etherscan.io/chart/tx

[17] Ethereum. (2022). The Merge. https://ethereum.org/en/roadmap/merge/

[18] Flashbots. (2022, September 23). MEV-Boost in a Nutshell. https://boost.flashbots.net/

[19] Flashbots. (n.d.). MEV-Share. https://docs.flashbots.net/flashbots-protect/mev-share

[20] Goswami, R. (2023, June 8). SEC sues coinbase over exchange and staking programs, stock drops 12%. CNBC. https://www.cnbc.com/2023/06/06/sec-sues-coinbase-over-exchange-and-staking-programs-stock-drops-14percent.html

[21] Gupta, T., Pai, M. M., & Resnick, M. (2023, May 30). The centralizing effects of private order flow on proposer-builder separation. https://arxiv.org/abs/2305.19150

[22] Kharif, O., & Greifeld, K. (2023, June 21).Bitcoin Jumps on Speculation BlackRock ‘May Know Something’ https://www.bloomberg.com/news/articles/2023-06-20/wisdomtree-files-for-us-spot-bitcoin-etf-on-heels-of-application-by-blackrock#xj4y7vzkg

[23] Lee, S. P. (2024, January 30). Market share of centralized crypto exchanges, by trading volume. https://www.coingecko.com/research/publications/centralized-crypto-exchanges-market-share

[24] Milionis, J., Moallemi, C., Roughgarden, T., & Lee, A. (2023, November 27). Automated Market Making and Loss-Versus-Rebalancing. https://anthonyleezhang.github.io/pdfs/lvr.pdf

[25] Qin, K., Zhou, L., & Gervais, A. (2021, December 10). Quantifying blockchain extractable value: How dark is the forest? https://arxiv.org/abs/2101.05511

[26] Sandor, K. (2022, March 11). Circle confirms $3.3B of USDC's cash reserves stuck at failed Silicon Valley Bank. https://www.coindesk.com/business/2023/03/11/circle-confirms-33b-of-usdcs-cash-reserves-stuck-at-failed-silicon-valley-bank

[27] Thorn, A. (2021, April 1). EIP-1559: A major upgrade for Ethereum. Galaxy Digital. https://www.galaxy.com/insights/research/eip-1559-major-ethereum-upgrade/

[28] Titan & Frontier Research. (2023, June 13). Builder Dominance and Searcher Dependence. https://frontier.tech/builder-dominance-and-searcher-dependence

[29] Wahrstätter, A., Zhou, L., Qin, K., Svetinovic, D., & Gervais, A. (2023, May 25). Time to bribe: Measuring block construction market. https://arxiv.org/abs/2305.16468

[30] Wahrstätter. (n.d.). Ethereum Mempool Dashboard. https://mempool.pics

# 10 Appendix

## 10.1 Data Sources

- Blockchain analytics platform Dune Analytics was used to query for on-chain data in the Ethereum ecosystem. This would return the list of transactions that corresponded to the DEX leg of the arbitrage and include key information such as token pair, amounts bought and sold, searcher and builders involved.

- Based on the timestamp of these transactions, the close price at the 1s interval was queried from leading centralised exchange (Binance). The prices for the traded pair, alongside the asset and USDT are obtained. USDT will be used as the primary base asset in these calculations given its resilience and historical 1:1 parity. The period retrieved per transaction consists of prices from t - 12 to t + 12, where t = block time when the block was created on-chain.

- To obtain the amount of gas and *coinbase transfers* involved in the transaction, I accessed Etherscan's API for its endpoints (eth_getTransactionReceipt).

**On-Chain Dataset Retrieved**

- **Block Information**: For every block, the obtained information includes block number, timestamp, miner (validator) address, block hash, gas used, base fee and burned fee.

- **Transaction Information**: For each transaction, I first collect the basic information, including hash, block number, success status (0 means failure and 1 means success), sender address, receiver address, ETH value, input, gasused, gasprice, and gasfee. I have also collected the following information:

  - **Traces**: the information of internal transactions that are the smart contract calls within the transaction. Specifically, every internal transaction should record the sender address, receiver address, ETH value, and the call graph information.
  - **Swaps**: token amount, token address, token sender, token receiver, token symbols, DEX, version type where the swap occurred.
  - **Searcher and Builder metadata**: The individual entities involved in the trade are recorded and their Ethereum addresses are mapped to known entities (Appendix 1.1). The Etherscan team has labelled some contracts as "MEV Bot" and "Block Builder".

These on-chain transactions and its relevant off-chain price data were then combined into a common dataframe to determine the theoretical profits and risks involved.

## 10.2 Identification Algorithm

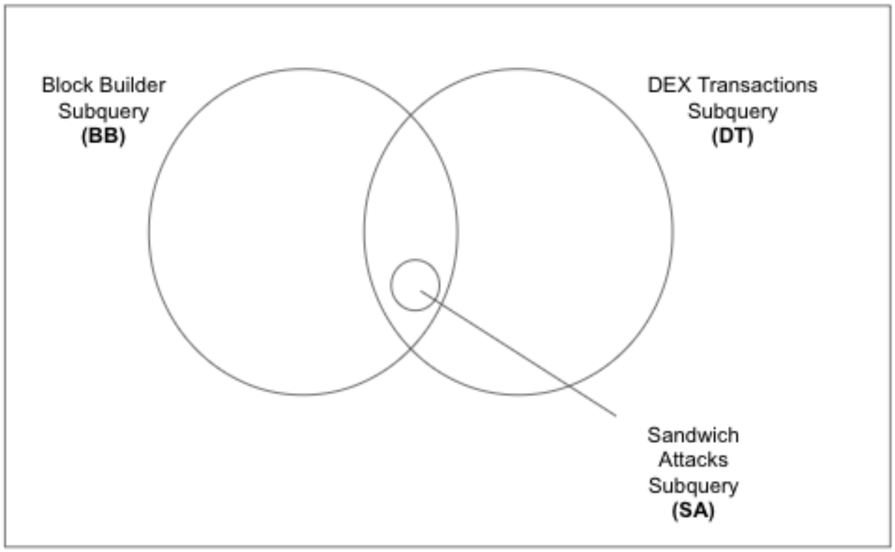



Figure 7: Overview of Identification Algorithm

*Illustration is not drawn to scale and the size does not signify the magnitude and/or number of transactions identified*

**Figure 7. Overview of Identification Algorithm** Figure 5 presents the logic of the identification algorithm through the construction of subqueries to fulfil each of the aforementioned heuristics.

- **Block Builder Subquery:** All transactions which consisted of an internal payment to a block builder were extracted in a subquery. The transaction hash, sender, block builder, and amount of ETH transferred were recorded.
- **DEX Transactions Subquery:** The relevant DEX transactions were retrieved in another subquery, consisting of the transaction hashes, name of the DEX, sender, symbols, and addresses for the token, the amounts transacted, and the USD equivalent of the transaction. This will consist of 2 transactions within 1 swap. For example, an arbitrageur transfers WETH to Uniswap DEX and receives USDC from it.
- **Sandwich Attacks Subquery:** Transactions which were identified as sandwich attacks by organizations such as Flashbots, were extracted.

$$\text{Scenario } 1 = \text{BB} \cap (\text{DT} \cap \text{SA}') \tag{5}$$

$$\text{Scenario 2} = (\text{BB} \cap \text{Scenario 1}') \cap (\text{DT} \cap \text{SA'}) \tag{6}$$

## 10.3 Data Cleaning

There exists large outliers from scenarios such as the emergence of memecoins or new listings on Binance, where these arbitrageurs seek to rapidly arbitrage the small windows of inefficiencies. As a result, winsorization was applied to the dataset, at the 5th and 95th percentile providing a more holistic overview of the landscape.

## 10.4 Summary of Events

**FTX Collapse: 1 to 12 November**: On November 2, 2022, CoinDesk published an article raising concerns about the solvency of FTX's affiliated trading arm, Alameda Research (Allison, 2022). Subsequently, Binance announced its decision to liquidate FTX's native token, FTT, leading to a decline in investor confidence and a wave of withdrawals from FTX. The resulting liquidity crisis forced FTX and its subsidiaries to file for bankruptcy on November 11, significantly impacting the cryptocurrency market as its total valuation fell below US$1 trillion.

**General Market Recovery: 10 to 26 January** : Following the tumultuous events of 2022, capital began flowing back into cryptocurrency markets. January 2023 witnessed an overall market upsurge of over 35%, indicating a recovery phase (Bovaird, 2023).

**USDC Depeg: 19 to 17 March**: Due to macroeconomic uncertainties in the traditional banking sector, Silicon Valley Bank ("SVB") fell victim as a result of poor risk management practices. Reports that Circle, the issuer of the USDC stablecoin, had deposited 8% of its funds in SVB caused market panic (Sandor, 2023). This led to a frenzied redemption of USDC, culminating in the stablecoin's devaluation to as low as 86 cents, and a consequential collapse in trading pairs involving USDC.

**SEC files several high profile lawsuits against CEXs - Binance and Coinbase, along with the foundations of several altcoins: 5 to 10 June**: On June 5, 2023, the US SEC initiated a lawsuit against the largest crypto exchange, Binance (Berwick, 2023), followed by a similar action against Coinbase on June 6 (Goswani, 2023). The allegations focused on regulatory non-compliance and improper operational mingling. Additionally, the SEC classified 13 cryptocurrencies, including Solana (SOL), Cardano (ADA), and Filecoin (FIL), as securities, leading to a sharp decline in their values.

**Institutionalisation of Bitcoin: 23 to 27 June**: June 15 marked a significant turn with Blackrock's unexpected application for a Bitcoin ETF, followed by similar moves from traditional financial giants like VanEck, ARK Invest, Fidelity Investments, Wisdom Tree, and Invesco. This series of events instilled confidence in the market, driving Bitcoin's price up by 12% and breaching the $30,000 threshold (Kharif, 2023).

## 10.5 Tables and Figures

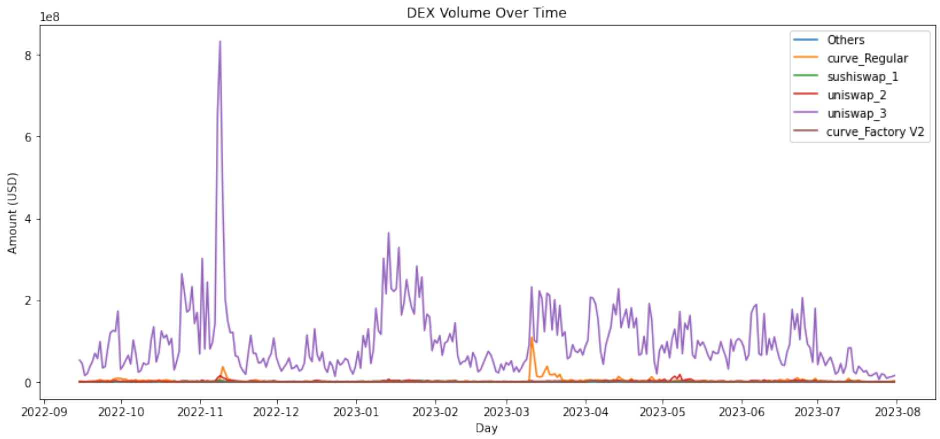
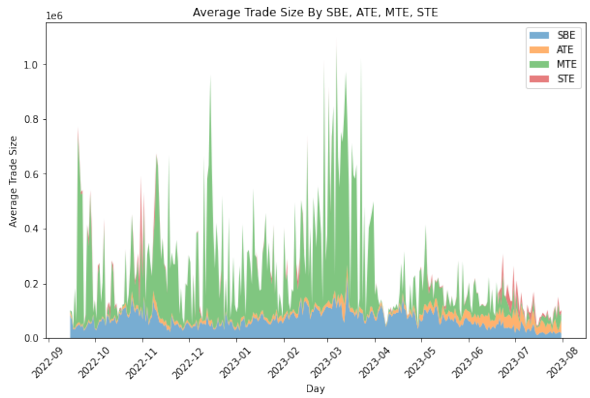
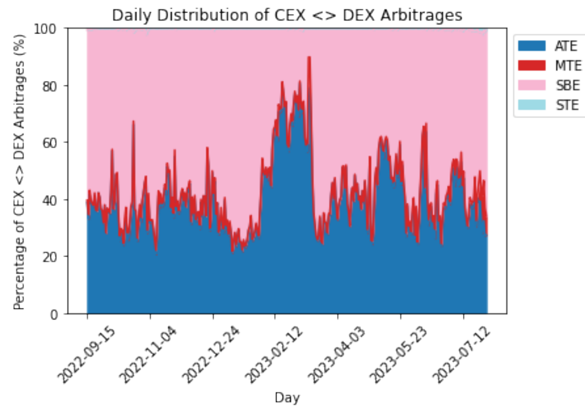



Figure 8: Daily Volume Traded on Each DEX, Top 5 + Others

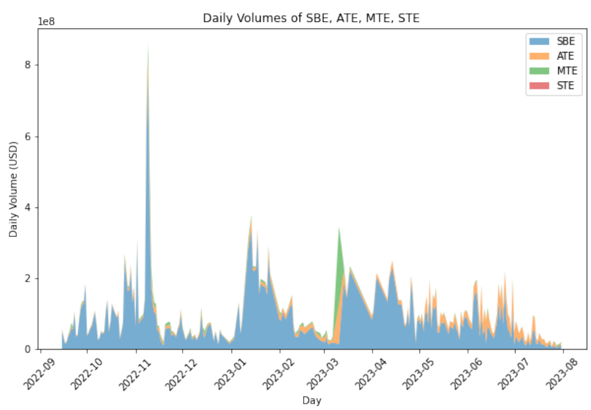
**Figure 8. Summary Statistics**. Figure 6 highlights the daily volume of CEX-DEX arbitrages recorded in the sample. The top 5 DEXs by volume were labelled, with the remaining classified as "Others".



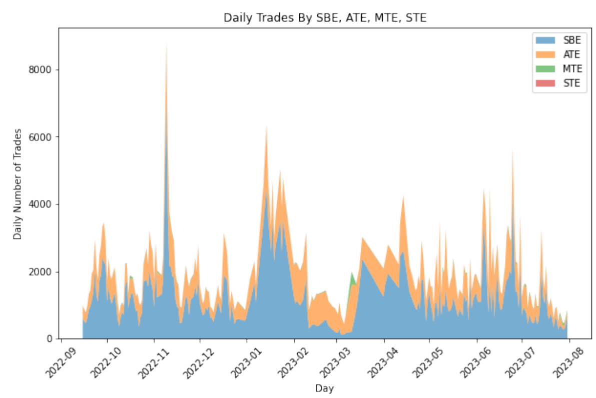

Daily Average Trade Size




Distribution of Trades




Daily Volume Traded




Daily Number of Arbitrages

Figure 9: Trading Statistics by Each Arbitrageur Profile

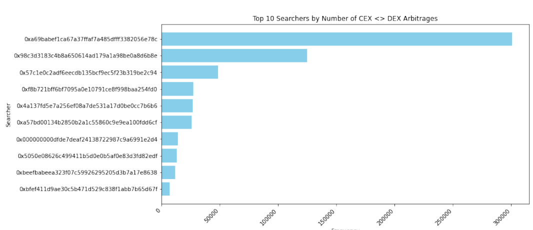
**Figure 9. Trading Statistics by Each Arbitrageur Profile.** The figures report the sum of the daily average volume, daily number of trades, average trade size and distribution of arbitrages amongst the arbitrageurs.

**Table 4. Number of Trades Executed by Asset Pair and Arbitrageurs**. The table reports the breakdown in the total number of trades by arbitrageur profile, for each pair type.

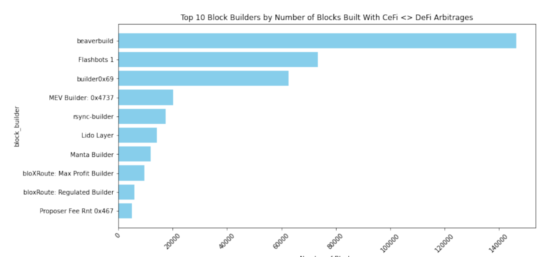
Table 4: Number of Trades Executed by Asset Pair and Arbitrageurs

| | SBE | ATE | MTE | STE |
|---|---|---|---|---|
| major-alt | 145,646 | 153,427 | 3,891 | 609 |
| major-stable | 182,503 | 26,918 | 2,307 | 381 |
| major-major | 41,807 | 5,737 | 161 | 19 |
| alt-stable | 12,601 | 27,789 | 430 | 66 |
| major-meme | 14,771 | 14,826 | 175 | 33 |
| stable-stable | 2,929 | 1,171 | 458 | 30 |
| alt-alt | 1,588 | 1,151 | 10 | 3 |
| meme-stable | 75 | 374 | 1 | 0 |
| meme-alt | 0 | 175 | 0 | 0 |
| Total | 401,920 | 231,568 | 7,433 | 1,141 |

## 10.6 Block Builders and Searchers




(a) Top 10 Searchers by Number of Arbitrages




(b) Top 10 Block Builders by Number of Blocks

Figure 10: Distribution of Arbitrages

**Figure 10. Distribution of Arbitrages Amongst Searchers and Block Builders**. Figure 10 highlights the distribution of arbitrages facilitated by each searcher and block builder. Figure 10a shows the number of CEX-DEX arbitrages executed by the top 10 searchers; 10b displays the number of arbitrages by the top 10 block builders (based on number of blocks built).

**Table 5. Searcher and Block Builder Distribution**. The tables report the percentage of CEX-DEX arbitrages facilitated by the top 10 searchers and percentage of blocks facilitated by block builders which contain CEX-DEX arbitrages.

Table 5: Searcher and Block Builder Distribution

| Searcher | % of Arbitrages |
|---|---|
| 0xa69babef1ca67a37ffaf7a485dfff3382056e78c | 46.81 |
| 0x98c3d3183c4b8a650614ad179a1a98be0a8d6b8e | 19.45 |
| 0x57c1e0c2adf6eecdb135bcf9ec5f23b319be2c94 | 7.61 |
| 0xf8b721bff6bf7095a0e10791ce8f998baa254fd0 | 4.25 |
| 0x4a137fd5e7a256ef08a7de531a17d0be0cc7b6b6 | 4.17 |
| 0xa57bd00134b2850b2a1c55860c9e9ea100fdd6cf | 4.08 |
| 0x000000000dfde7deaf24138722987c9a6991e2d4 | 2.22 |
| 0x5050e08626c499411b5d0e0b5af0e83d3fd82edf | 2.04 |
| 0xbeefbabeea323f07c59926295205d3b7a17e8638 | 1.84 |
| 0xbfef411d9ae30c5b471d529c838f1abb7b65d67f | 1.10 |

| Block Builder | % of Blocks Built |
|---|---|
| beaverbuild | 36.31 |
| Flashbots 1 | 18.25 |
| builder0x69 | 15.53 |
| MEV Builder: 0x4737 | 5.00 |
| rsync-builder | 4.32 |
| Lido Layer | 3.54 |
| Manta Builder | 2.99 |
| bloXRoute: Max Profit Builder | 2.37 |
| bloxRoute: Regulated Builder | 1.52 |
| Proposer Fee Rnt 0x467 | 1.25 |

## Searcher - Builder Relationships

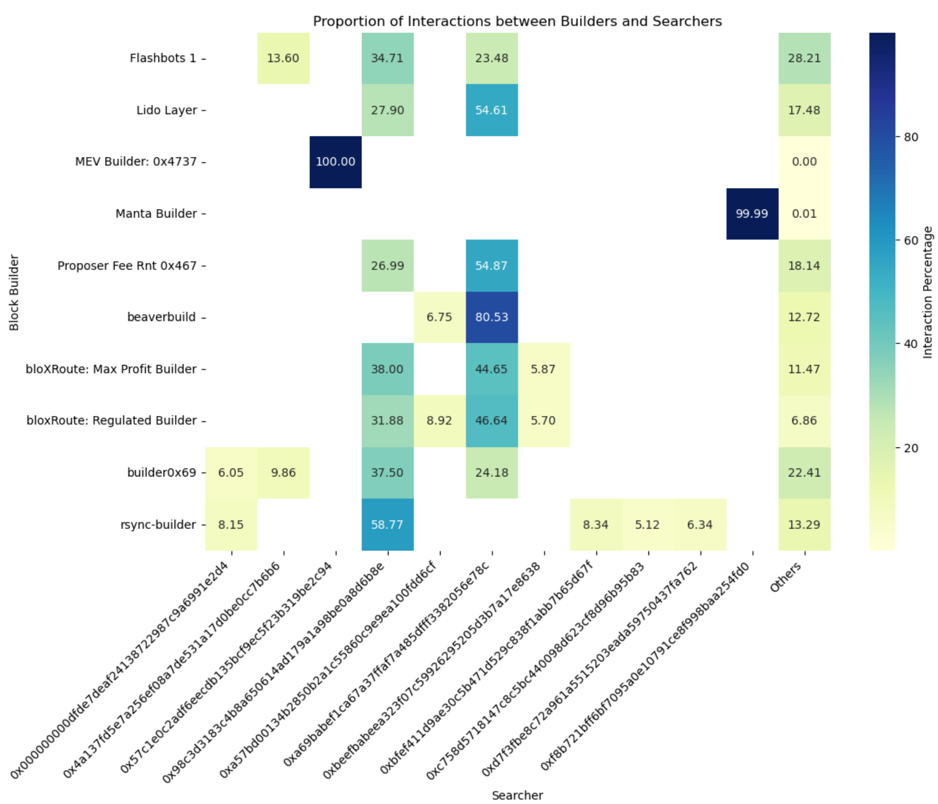



Figure 11: Searcher and Builder Coverage

**Figure 11. Searcher and Builder Coverage**. The heatmaps reveal the percentage of interactions by the top 10 searchers/block builders with their counterparty (block builder/searcher respectively). In this case, since the top 10 searchers interact with the same group of block builders and vice versa, 1 heatmap will suffice to represent the relationships.

## 10.7 Risk-Reward Analysis

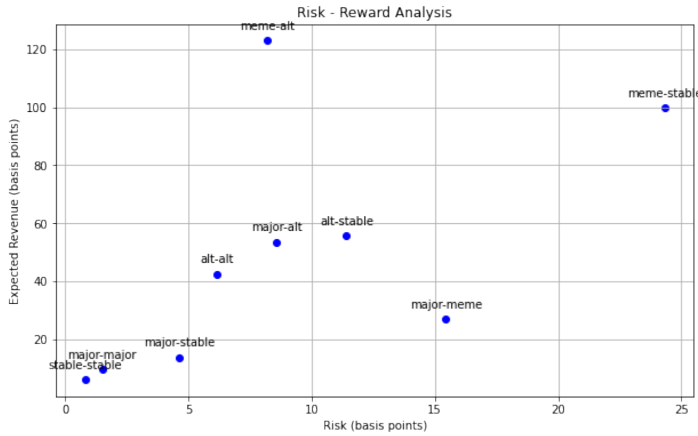
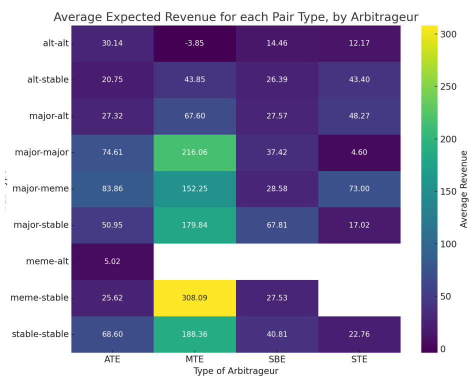
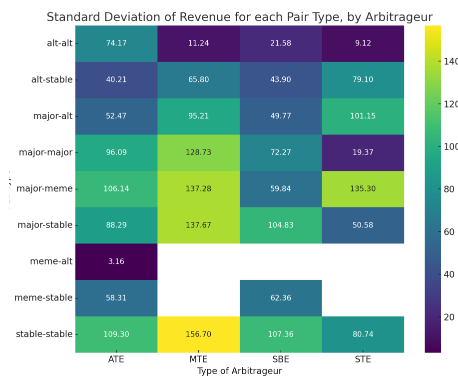



Figure 12: Risk Reward Analysis by Asset Pair

**Figure 12. Risk Reward Payoff Structure By Asset Pair** Figure 12 presents the risk reward payoff structure by asset pair. Expected Revenue refers to the average revenue throughout the trading window of 24 seconds. Risk refers to the standard deviation of revenue throughout the trading window of 24 seconds.




(a) Average Revenue By Arbitrageur

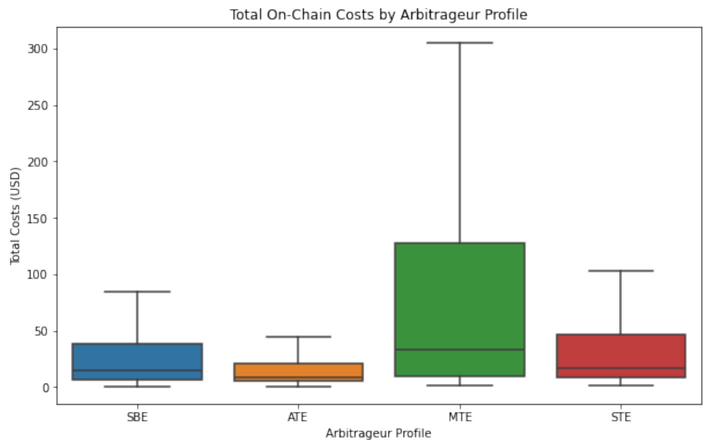


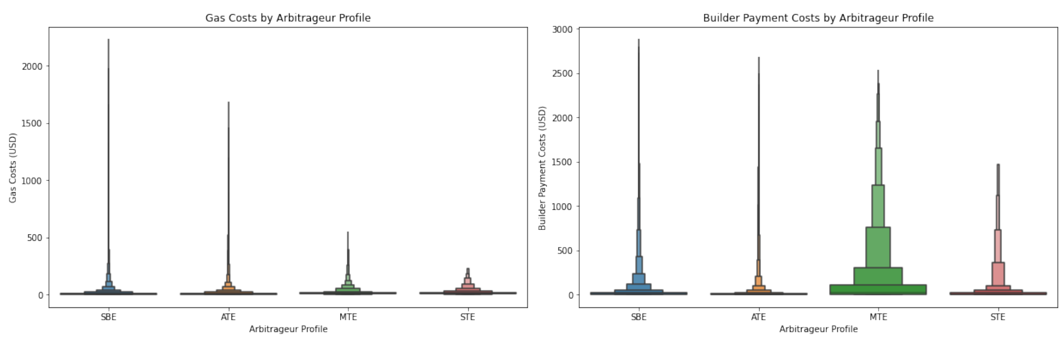

(b) Spread of Revenue By Arbitrageur

Figure 13: Overview of Average and Spread of Revenue Earned by Arbitrageur

**Figure 13. Descriptive Statistics on Revenue**. The figures display the descriptive statistics for the revenue earned by each type of arbitrageur, by pair-type. This includes the average and standard deviation in dollar terms. Figure 13a highlights the average revenue for each type of arbitrageur profile; 13b features the overview of the spread revenue for the arbitrageurs, excluding major-meme by STE due to the anomalies.

## 10.8 Transaction Cost Analysis



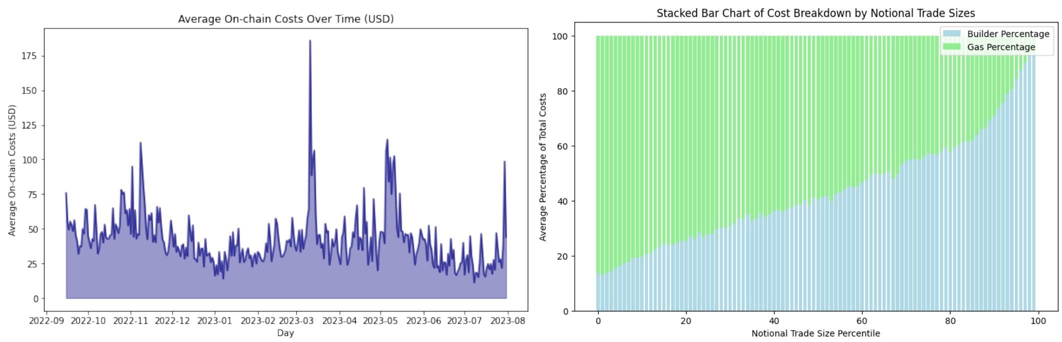




(a) Distribution of Gas Costs

(b) Distribution of coinbase transfers

Figure 14: Breakdown of Transaction Costs

**Figure 14. Total Cost of Transaction by Arbitrageur Profile**. The figure presents the boxplot distribution of the total costs incurred by each arbitrageur profile in USD It highlights the 25th percentile, mean and 75th percentiles for all the transactions. Figure 14a shows a boxplot distribution of the gas costs while Figure 14b features the distribution of *coinbase transfers* by arbitrageur profiles.




(a) Total Costs Incurred ($USD) (b) Distribution of Total Costs by Order Size Percentiles

Figure 15: Distribution of Transaction Costs

**Figure 15. Total Cost of Transaction - Gas, *coinbase transfers*.** The figures present an analysis of the total costs of the CEX-DEX arbitrages in USD. Figure 15a shows a time series visualisation of the total costs during the study period; Figure 15b features the distribution of *coinbase transfers* and gas out of the total costs, based on the percentiles of the order sizes.

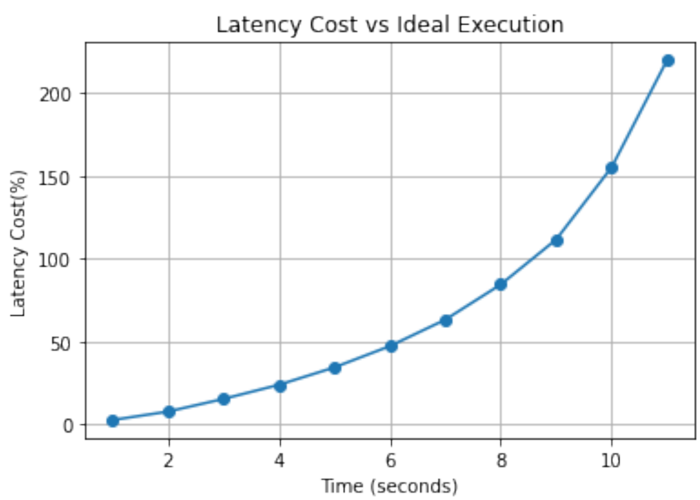
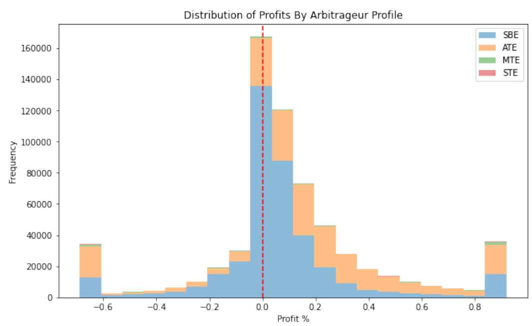
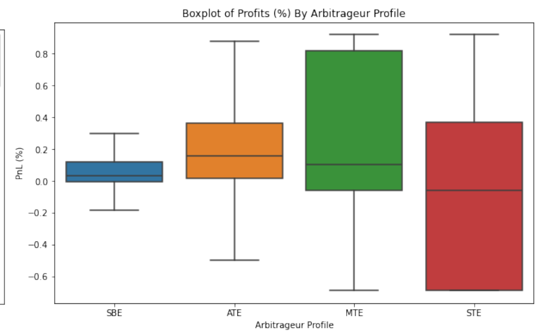



Figure 16: Opportunity Costs from Latency

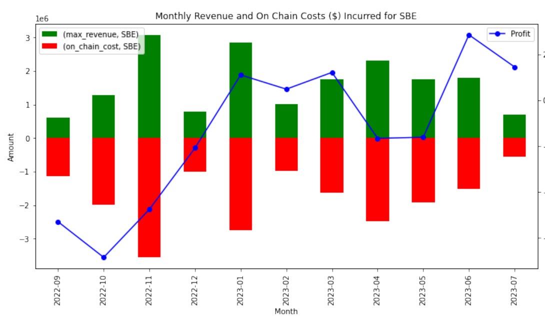
**Figure 16. Opportunity Costs from Latency** The figure presents a visual of the costs of latency prior to transaction confirmation on-chain. Table 10 represents the average costs for all transactions recorded at the intervals from 1s to 12s, with the former being 1s away from transaction being confirmed, and the latter being 12s away.

## 10.9 Profitability Analysis

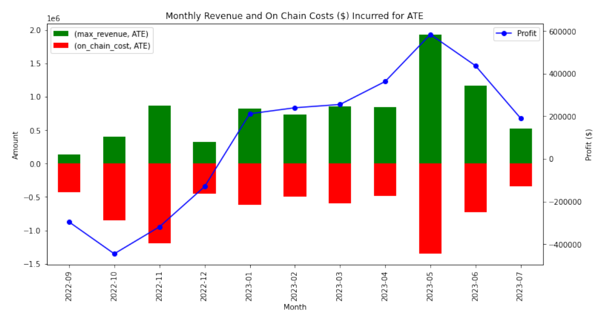


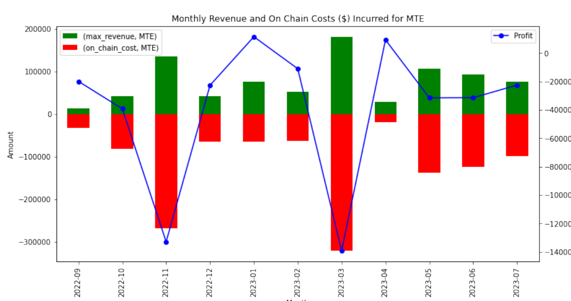

(a) Profit Distribution by Arbitrageur Profile (b) Boxplot of Profits by Arbitrageur Profile



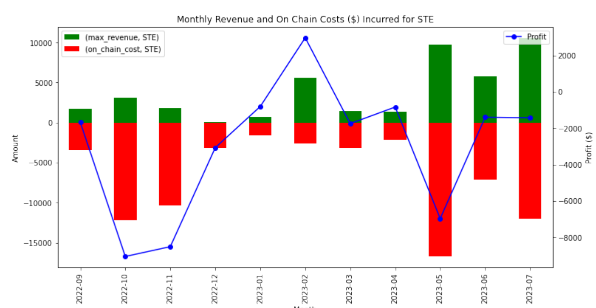

SBE




ATE




MTE




STE

Figure 17: Total Revenue and Total Costs Incurred by Arbitrageur Profile

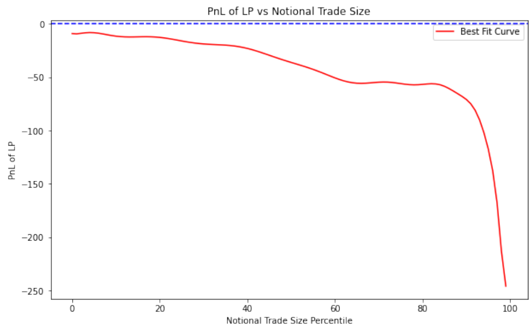
**Figure 17. Total Revenue and Total Costs incurred by Arbitrageur Profile** The figure presents the estimated profits earned by each arbitrageur profile, by accounting for the total revenue and costs incurred. Figure 17a presents a histogram distribution of the profitability; 17b highlights the boxplot distribution of the profitability by arbitrageur profile. The 4 bar graphs represent the monthly maximum theoretical revenue and total costs incurred by arbitrageur profile.

Table 6: Regression Results for Volatility and Event Days

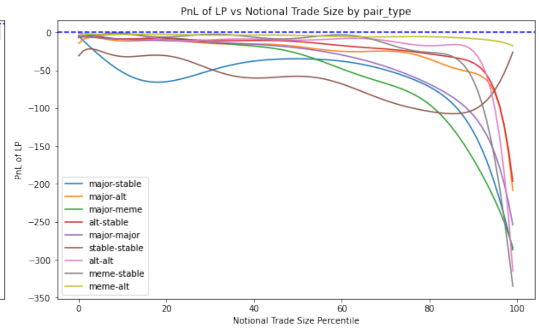
| | *Dependent variable: Profit (%)* | | | |
|---|---|---|---|---|
| | SBE | ATE | MTE | STE |
| arbitrages | 0.486*** | 0.007 | 0.707*** | -1.071*** |
| | (0.012) | (0.011) | (0.073) | (0.379) |
| event | 0.039*** | 0.097*** | -0.405*** | -0.915*** |
| | (0.008) | (0.008) | (0.055) | (0.272) |
| volatility | 0.019*** | 0.016*** | -0.209*** | -0.114*** |
| | (0.002) | (0.002) | (0.008) | (0.025) |
| volume | -0.357*** | -0.095*** | -0.330*** | 0.407*** |
| | (0.005) | (0.005) | (0.030) | (0.156) |
| Observations | 284354 | 195372 | 4960 | 510 |
| $R^2$ | 0.763 | 0.596 | 0.397 | 0.233 |
| Adjusted $R^2$ | 0.763 | 0.596 | 0.396 | 0.227 |
| Residual Std. Error | 1.561 (df=284350) | 1.367 (df=195368) | 1.584 (df=4956) | 2.221 (df=506) |
| F Statistic | 228956.390*** (df=4; 284350) | 72098.166*** (df=4; 195368) | 814.452*** (df=4; 4956) | 38.494*** (df=4; 506) |
| *Note:* | | | | *p<0.1; **p<0.05; ***p<0.01 |

**Table 6. Regression Results for Volatility and Event Days**. The table reports the regression results on the influence of the volatility of the token pair and events, on profit (percentage terms) by arbitrageur profile, normalized against the daily volume and number of arbitrages.

## 10.10 Toxic Flow Incurred by LPs



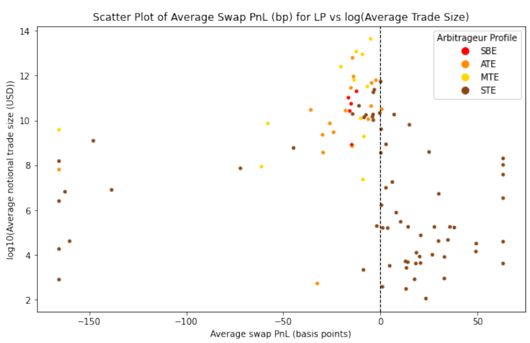

(a) Overall Profits of LP



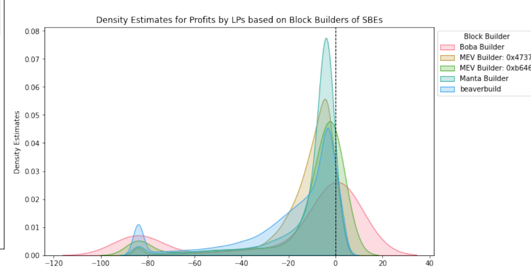

(b) Profits of LPs by Pair Type

Figure 18: Distribution of Profits for LPs by Notional Trade Size (Percentiles)

**Figure 18. Profits by LPs** The figure presents the estimated profits earned by LPs based on percentiles of the notional trade size. Figure 18a presents the best fit curve for an aggregated overview of profits by LPs; 18b highlights the best fit curves, segmented by the pair types.




(a) By Arbitrageur Profile




(b) By Integrated Block Builders

Figure 19: Distribution of Profits for LPs by Arbitrageur Profile

**Figure 19. Profits by LPs based on the Arbitrageur Profile**. The figure presents the estimated profits earned by the LPs when interacting with each arbitrageur profile. Figure 19a presents the scatterplot of the average profit earned against the average notional size of the trade; 19b highlights the distribution of estimated profits earned by LPs when these arbitrages are facilitated by the integrated block builders.

## 10.11 Attribution of Known Block Builders

| No. | Address | Block Builder |
|---|---|---|
| 0 | 0x95222290dd7278aa3ddd389cc1e1d165cc4bafe5 | beaverbuild |
| 1 | 0x1f9090aae28b8a3dceadf281b0f12828e676c326 | rsync-builder |
| 2 | 0x690b9a9e9aa1c9db991c7721a92d351db4fac990 | builder0x69 |
| 3 | 0xdafea492d9c6733ae3d56b7ed1adb60692c98bc5 | Flashbots 1 |
| 4 | 0x4675c7e5baafbffbca748158becba61ef3b0a263 | Proposer Fee Recipient 0x467 |
| 5 | 0x388c818ca8b9251b393131c08a736a67ccb19297 | Lido Execution Layer |
| 6 | 0xbaf6dc2e647aeb6f510f9e318856a1bcd66c5e19 | MEV Builder : 0xbaf |
| 7 | 0xfeebabe6b0418ec13b30aadf129f5dcdd4f70cea | eth-builder |
| 8 | 0xb4c9e4617a16be36b92689b9e07e9f64757c1792 | MEV Builder: 0xb4c |
| 9 | 0x6d2e03b7effeae98bd302a9f836d0d6ab0002766 | Propose Fee Recipient 0x6d2 |
| 10 | 0xe688b84b23f322a994a53dbf8e15fa82cdb71127 | Proposer Fee Recipient: 0xe69 |
| 11 | 0x333333f332a06ecb5d20d35da44ba07986d6e203 | MEV Builder: 0x33 |
| 12 | 0x4838b106fce9647bdf1e7877bf73ce8b0bad5f97 | Titan Builder |
| 13 | 0xebec795c9c8bbd61ffc14a6662944748f299cacf | Proposer Fee Recipient: 0xebe |
| 14 | 0xbd3afb0bb76683ecb4225f9dbc91f998713c3b01 | MEV Builder: 0xbd |
| 15 | 0x5124fcc2b3f99f571ad67d075643c743f38f1c34 | MEV Builder: 0x968 |
| 16 | 0xe94f1fa4f27d9d288ffea234bb62e1fbc086ca0c | Propose Fee Recipient: 0xe94 |
| 17 | 0xffee087852cb4898e6c3532e776e68bc68b1143b | StakeFish Recipient |
| 18 | 0x3b64216ad1a58f61538b4fa1b27327675ab7ed67 | Boba Builder |
| 19 | 0xf2f5c73fa04406b1995e397b55c24ab1f3ea726c | bloXRoute: Max Profit Builder |
| 20 | 0x199d5ed7f45f4ee35960cf22eade2076e95b253f | bloxRoute: Regulated Builder |
| 21 | 0x473780deaf4a2ac070bbba936b0cdefe7f267dfc | MEV Builder: 0x4737 |
| 22 | 0xb646d87963da1fb9d192ddba775f24f33e857128 | MEV Builder: 0xb646 |
| 23 | 0xb64a30399f7f6b0c154c2e7af0a3ec7b0a5b131a | Flashbots: Old Builder |
| 24 | 0xaab27b150451726ec7738aa1d0a94505c8729bd1 | Eden Network: Builder |
| 25 | 0x5F927395213ee6b95dE97bDdCb1b2B1C0F16844F | Manta Builder |