Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Paying for speed, hedging the wait: Execution risk and optimal priority fees in arbitrage between centralised and decentralised crypto exchanges

Kirill Proskurin

koproskurin@edu.hse.ru

Southwestern University of Finance and Economics

Research Article

Keywords: Decentralised finance, Arbitrage, Execution risk, Priority fee, Hedging, Stochastic control

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11358315/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Paying for speed, hedging the wait: Execution risk and optimal priority fees in arbitrage between centralised and decentralised crypto exchanges

Kirill Proskurin[1,2]

[1] *HSE University, 11 Pokrovsky Boulevard, Moscow 109028, Russia*

[2] *Southwestern University of Finance and Economics, 555 Liutai Avenue, Wenjiang District, Chengdu, Sichuan 611130, China*

Corresponding author: *Kirill Proskurin, e-mail: koproskurin@edu.hse.ru*

Acknowledgments

The author thanks Ivan Petrenko (HSE University) for supervision and helpful comments.

Abstract

In arbitrage between a centralised crypto exchange and a decentralised exchange, the trade on the centralised exchange is filled at once, while the transaction on the decentralised exchange waits in a public queue until it is included in a block. Prices move during this random delay, so the arbitrage carries execution risk. This paper studies the arbitrageur's decision problem with two controls: a priority fee that shortens the expected wait and a futures hedge on the centralised exchange. The delay is modelled as the first jump of a counting process whose intensity increases with the fee. A mean–variance version of the problem gives closed-form rules for both controls, and a dynamic stochastic control version, solved numerically through its Hamilton–Jacobi–Bellman equation, confirms them as a first-order approximation. The minimum-variance hedge does not depend on the fee, while the optimal fee rises with trade size, risk aversion, volatility and the share of risk that cannot be hedged. I calibrate the model on Binance, Uniswap v3 and Ethereum data. The hedge removes 93.9% of the price variance during the wait. In the calm market of 2026, the optimal fee is at its lower bound, so an optimised rule and a fixed-fee rule perform the same. Across nine historical gas regimes, the optimised rule has a higher Sharpe ratio than the fixed-fee rule in eight. On 182,562 on-chain arbitrage swaps, it earns USD 2.59 more per trade than the fixed-fee rule.

Keywords Decentralised finance · Arbitrage · Execution risk · Priority fee · Hedging · Stochastic control

JEL Classification C61 · G11 · G14

1 Introduction

Crypto-assets are traded on two very different kinds of venues. On a centralised exchange (CEX), such as Binance, orders are matched in a limit order book that a company runs. On a decentralised exchange (DEX), such as Uniswap, a smart contract executes the trade against a pool of tokens. This pool is called an automated market maker (AMM), and the trade is recorded on a public blockchain (Angeris et al. 2021; Lehar and Parlour 2025). The same assets trade on both venues. Their prices stay close because arbitrageurs buy where the asset is cheaper and sell where it is more expensive.

This activity is large. In the Uniswap v3 pool of ether and USD Coin that I study, about one in four swaps in early 2026 meets my definition of an arbitrage trade against the CEX price (Section 4.7). Lehar and Parlour (2025) find no long-lived arbitrage opportunities between Uniswap and Binance, so this mechanism seems to work. However, it works only as

long as arbitrageurs are able and willing to trade. Makarov and Schoar (2020) show that when arbitrage capital faces frictions, price gaps across crypto exchanges can become large and persistent.

In this paper I study one friction that is specific to DEXs. The arbitrageur cannot execute both parts of the trade at the same time. The CEX part is filled almost at once. The DEX part is sent to a public waiting room, the mempool, and it is executed only when a block producer includes it in a block. This can take from seconds to minutes, and the exact time is random. Under the current Ethereum fee rules, introduced by Ethereum Improvement Proposal 1559 (EIP-1559), a trader can shorten the expected wait by paying a higher priority fee (Buterin et al. 2019; Roughgarden 2020). While the transaction waits, prices on both venues move. So the spread that the arbitrageur sees when deciding to trade is not the spread that is finally earned. In other words, CEX–DEX arbitrage is not riskless. It carries execution risk. Capponi et al. (2026) find that traders on DEXs pay high priority fees partly to avoid this kind of risk during congestion. For this reason, I think of the priority fee as a price of speed, and of block space as a market where traders buy a reduction in risk.

This problem matters not only for arbitrageurs. First, arbitrage profits are paid by the liquidity providers of DEX pools, who sell to arbitrageurs at stale prices (Milionis et al. 2024). The costs and risks of arbitrageurs affect how far pool prices can move away from the market before someone corrects them. Second, the fees that arbitrageurs pay for priority are an important part of the demand for block space, and these fees rise sharply when the network is congested (Liu et al. 2022). Third, ordinary DEX users get prices that are only as good as arbitrage keeps them. Finally, the problem is a simple example of a general risk-management question that appears in any market with delayed settlement: how much should one pay to shorten the time of exposure, and how much of the exposure can be hedged instead?

I ask three questions. First, how much of the risk during the wait can be hedged on the CEX, and does the best hedge depend on the fee? Second, when is it worth paying for speed, and how should the fee change with trade size, risk aversion, volatility and network congestion? Third, do optimised fee rules work better than the simple fixed-fee rules that many trading bots use, and in which market conditions?

To answer these questions, I model the waiting time as a random variable whose mean falls when the fee rises. The arbitrageur chooses the fee and a futures hedge on the CEX at the same time. A mean–variance version of the model gives closed-form rules for both choices. A dynamic version, written as a stochastic control problem with a controlled execution intensity,

shows that these simple rules are a good approximation. I then calibrate the model on public data from Binance, Uniswap v3 and the Ethereum blockchain. I test it in the calm market of 2026, in nine historical periods of network congestion and on 182,562 arbitrage swaps observed on chain.

The paper has three contributions. First, it formulates CEX–DEX arbitrage as a decision problem under a controlled execution intensity, in which the trader chooses the fee and the hedge jointly. In this problem the two decisions separate: execution risk splits into a part that can be hedged on the CEX and a part that cannot, and only the second part is worth paying for with a higher fee. Second, it gives an explicit formula for the price that an arbitrageur should be ready to pay for speed and shows, through the Hamilton–Jacobi–Bellman (HJB) equation of the dynamic problem and its numerical solution, that this formula is a first-order approximation of the dynamic policy. Third, it calibrates the model on public exchange and on-chain data and shows in several historical periods when optimising the fee matters in practice and when a simple rule is enough. The paper therefore contributes a tractable model of decisions under execution delay, rather than an equilibrium model of the fee market.

The paper is organised as follows. Section 2 explains how CEX–DEX arbitrage works and reviews related studies. Section 3 presents the model, the data and the calibration. Section 4 reports the results. Section 5 discusses the economic insights, the implications for different market participants and the limitations. Section 6 concludes. Appendix A gives the details of the dynamic model, Appendix B describes the numerical method, and Appendix C contains extensions to several hedging instruments and several opportunities.

2 Background and related literature

2.1 How CEX–DEX arbitrage works

Fig. 1 shows one arbitrage cycle. Suppose that ether is more expensive on the CEX than in the DEX pool. The arbitrageur sells ether on the CEX, where the order is filled at once. At the same time, the arbitrageur sends a transaction that buys ether from the pool. Ethereum produces a new block about every 12 seconds. Each transaction pays a base fee and a priority fee. The protocol sets the base fee from recent demand for block space. The user chooses the priority fee, which goes to the block producer (Buterin et al. 2019). Block producers prefer transactions with higher priority fees, so these transactions are usually included sooner. Until the transaction is included, the arbitrageur has an open position: the CEX part is done, but the DEX part is not. If the pool price moves in the wrong direction during this time, part or all of

the spread is lost. The trade also loses some value to slippage, because a large swap moves the pool price against the trader (Adams et al. 2021; Angeris et al. 2021).

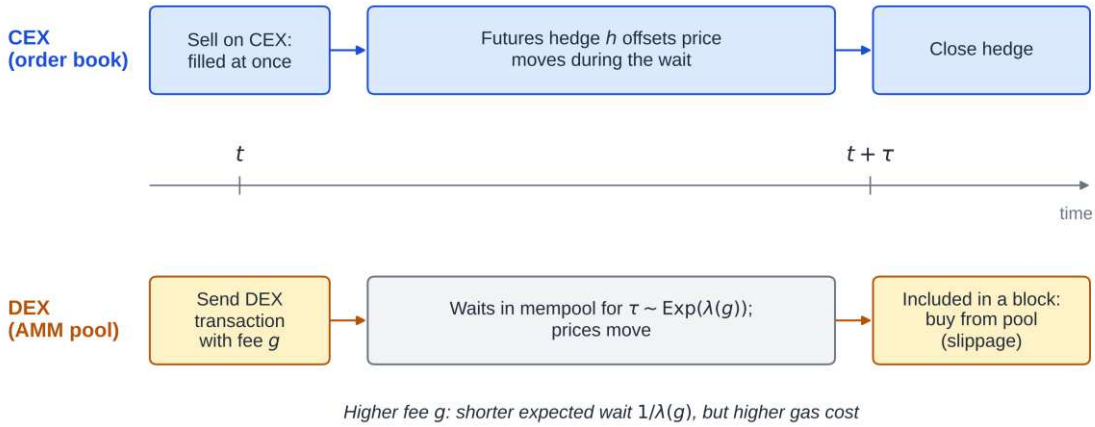


Fig. 1 One cycle of CEX–DEX arbitrage. At time t the arbitrageur sees a spread, fills the CEX leg at once and sends the DEX transaction, which is executed after a random delay τ. The priority fee g shortens the expected delay, and a futures hedge h on the CEX offsets price moves during the wait. The schematic is not drawn to scale

So the arbitrageur has two tools. The first tool is to pay more for speed. This shortens the period of exposure, but it costs more in fees. The second tool is to hedge, for example by opening an opposite position in futures on the CEX. Then price changes during the wait matter less. In short, the spread decides whether to enter the trade, while fees, slippage and the random delay decide how much is actually earned.

2.2 Related literature

My work is related to four groups of studies.

Arbitrage and price alignment across crypto venues. Makarov and Schoar (2020) document large and recurring arbitrage spreads across crypto exchanges. They link these spreads to frictions that limit the movement of arbitrage capital. Lehar and Parlour (2025) study Uniswap pools and find that they are stable and do not leave long-lived arbitrage opportunities against Binance. Hägele (2024) reviews studies that compare CEX and DEX markets, and Kirste et al. (2024) study market quality under conventional and automated market makers. These studies show that arbitrage connects the two types of venues. However, they do not model the risk that arbitrageurs carry while the on-chain part of the trade is waiting.

Liquidity providers and the cost of arbitrage. Angeris et al. (2021) analyse price formation and slippage in constant-function market makers, and the Uniswap v3 white paper

describes concentrated liquidity (Adams et al. 2021). Milionis et al. (2024) show that liquidity providers systematically lose to arbitrageurs when pool prices are stale, and they propose a way to measure this loss. Chemaya et al. (2025) and Adams et al. (2025) measure the costs of DEX swaps and the role of slippage tolerance. Brönnimann et al. (2024) and Lim (2024) study the incentives and risks of liquidity providers.

Fee markets, priority and transaction ordering. Buterin et al. (2019) describe the EIP-1559 fee market, and Roughgarden (2020, 2021) analyses it. Liu et al. (2022) study its effect on fees and waiting times in the data. Daian et al. (2020) show how users compete for the order of transactions, and they introduce the idea of maximal extractable value (MEV). Capponi et al. (2026) show that high-fee DEX trades carry more private information. They also find that traders pay high fees both to win priority and to avoid execution risk during congestion. Other studies look at high-frequency jumps in crypto markets (Saef et al. 2024) and at manipulation and sandwich attacks on DEXs (Bai et al. 2025; Naviglio et al. 2026).

Execution under delay and risk control. The classical papers of Bertsimas and Lo (1998) and Almgren and Chriss (2001) treat order execution as a trade-off between expected cost and the variance of price changes before the order is completed. Cartea and Sánchez-Betancourt (2023) include random latency in the control problem and show that delay changes the optimal policy in a qualitative way. For the dynamic version of my model, I use the continuous-time portfolio framework of Merton (1971), the results of Kramkov and Schachermayer (1999) on when such problems are well posed, and standard textbooks on the control of Markov processes and jump diffusions (Fleming and Soner 2006; Øksendal and Sulem 2007; Pham 2009).

Each of these groups of studies is well developed, but they are rarely combined. As far as I know, there is no study that (i) treats CEX–DEX arbitrage as a risk-management problem with a public execution queue, (ii) links the priority fee directly to the distribution of the waiting time and chooses it together with a hedge, and (iii) estimates all parts of the model from public exchange and on-chain data. This paper tries to fill this gap.

3 Material and methods

3.1 Single-cycle arbitrage

At time t the arbitrageur sees a positive spread between the CEX and the DEX. The arbitrageur immediately fills the CEX trade at price S_t^C and sends the DEX transaction. The

DEX trade is confirmed after a random delay τ. The trade size $q > 0$ is fixed. The DEX mid-price S^D and the price of the hedging futures contract F follow

$$\frac{dS_u^D}{S_u^D} = \mu_d\, du + \sigma_d\, dW_u^{(d)}, \qquad \frac{dF_u}{F_u} = \mu_f\, du + \sigma_f\, dW_u^{(f)}, \qquad d\langle W^{(d)}, W^{(f)}\rangle_u = \rho_{df}\, du. \tag{1}$$

I assume that the delay is exponential, with an intensity $\lambda(g)$ that increases in the chosen fee g:

$$\tau \sim \mathrm{Exp}\big(\lambda(g)\big), \qquad \mathbb{E}[\tau \mid g] = 1/\lambda(g). \tag{2}$$

I use the exponential distribution as a simple reduced form, not as a model of how the mempool works. It has two useful properties. The probability of inclusion in a short interval du is known and increases with the priority payment, and $\mathbb{E}[\tau]$ can be written directly through $\lambda(g)$. This is similar to other reduced-form delay models (Cartea and Sánchez-Betancourt 2023; Liu et al. 2022).

With a futures hedge of size h, the profit of one cycle is

$$\Pi(g,h) = qS_t^C - qP_{t+\tau}^{D,\mathrm{exec}} + h\,(F_{t+\tau} - F_t) - c_{\mathrm{gas}}(g), \qquad P_{t+\tau}^{D,\mathrm{exec}} = S_{t+\tau}^D + c_{\mathrm{slip}}(q, L_{t+\tau}), \tag{3}$$

where $P^{D,\mathrm{exec}}$ is the realised DEX execution price, L is pool liquidity and $c_{\mathrm{slip}} \geq 0$ is slippage. Following the AMM literature, I separate slippage from the mid-price (Angeris et al. 2021; Milionis et al. 2024). Using the spread $\delta_t := S_t^C - S_t^D$ and the increments $\Delta S_\tau^D := S_{t+\tau}^D - S_t^D$ and $\Delta F_\tau := F_{t+\tau} - F_t$, I can rewrite the profit as

$$\Pi(g,h) = q\delta_t - q\,\Delta S_\tau^D - q\,c_{\mathrm{slip}}(q, L_{t+\tau}) + h\,\Delta F_\tau - c_{\mathrm{gas}}(g). \tag{4}$$

After the CEX trade is filled, there are two sources of risk: price changes during the wait and random slippage. To make the problem tractable, I use two assumptions. Both are checked against the data in Section 3.7.

Assumption 1 (conditional independence of the delay). Conditional on $\mathcal{F}_t$, the delay τ is independent of future Brownian increments.

Assumption 2 (reduced-form slippage). There is a function $A(q,t) > 0$ such that

$$\mathbb{E}\big[q\,c_{\mathrm{slip}}(q, L_{t+\tau}) \mid \mathcal{F}_t\big] = \frac{A(q,t)}{\lambda(g)}, \qquad \mathrm{Var}\big(c_{\mathrm{slip}}(q, L_{t+\tau}) \mid \mathcal{F}_t\big) = 0. \tag{5}$$

Assumption 1 is standard in execution models with delay. Assumption 2 puts the systematic losses from price impact, adverse selection and worse execution conditions during the wait into one expected cost. Execution costs depend strongly on trade size and liquidity (Adams et al. 2025; Angeris et al. 2021). In my data, slippage is well described by a power function whose parameters change slowly compared with the time of one transaction. For this reason, I ignore its conditional variance as a first approximation.

3.2 Mean–variance objective

The arbitrageur maximises

$$J(g,h) := \mathbb{E}[\Pi(g,h) \mid \mathcal{F}_t] - \frac{\gamma}{2}\,\mathrm{Var}(\Pi(g,h) \mid \mathcal{F}_t), \qquad \gamma > 0. \tag{6}$$

In words, the arbitrageur values expected profit but dislikes uncertainty about it, and γ measures how strong this dislike is. This criterion goes back to Markowitz (1952) and is often used in execution problems with timing risk (Almgren and Chriss 2001; Gollier 2001). Following Pratt (1964), J is a two-moment approximation of $\mathbb{E}[U(\Pi)]$ when U is smooth and the risk is small. It is exact for constant absolute risk aversion (CARA) utility and normally distributed profit. My horizons are from seconds to minutes, and positions are small compared with capital, so I consider this approximation acceptable. In Section 3.5 I compare it with a full formulation with constant relative risk aversion (CRRA).

Conditional mean. By Assumption 1, $\mathbb{E}[\Delta S^D_\tau \mid \mathcal{F}_t] = \mu_d S^D_t/\lambda(g)$ and $\mathbb{E}[\Delta F_\tau \mid \mathcal{F}_t] = \mu_f F_t/\lambda(g)$. On my horizons (tens of seconds at moderate fees), the drift is very small compared with the diffusion, $|\mu|\tau \ll \sigma\sqrt{\tau}$. I check this condition in Section 3.7 and set both expectations to zero. Using Assumption 2, I get

$$\mathbb{E}[\Pi \mid \mathcal{F}_t] = q\delta_t - c_{\mathrm{gas}}(g) - \frac{A(q,t)}{\lambda(g)}. \tag{7}$$

The hedge does not appear in the mean. It affects only the variance.

Conditional variance. The $\mathcal{F}_t$-measurable terms drop out, and by Assumption 2 slippage also drops out:

$$\mathrm{Var}(\Pi \mid \mathcal{F}_t) = \mathrm{Var}(-q\,\Delta S^D_\tau + h\,\Delta F_\tau \mid \mathcal{F}_t). \tag{8}$$

The fee still matters here, because it changes the distribution of τ. Combining (7) and (8), I obtain

$$J(g,h) = q\delta_t - \underbrace{c_{\mathrm{gas}}(g) - \frac{A(q,t)}{\lambda(g)}}_{\text{depends on } g \text{ only}} - \frac{\gamma}{2}\underbrace{\mathrm{Var}(-q\,\Delta S^D_\tau + h\,\Delta F_\tau \mid \mathcal{F}_t)}_{\text{depends on } g \text{ and } h}. \tag{9}$$

This means the problem can be solved in two steps. First, for a given g, I choose h to minimise the variance. Second, I optimise over g.

3.3 Optimal hedge

Proposition 1. If $\mathrm{Var}(\Delta F_\tau \mid \mathcal{F}_t) > 0$, the variance-minimising hedge is $h^\star = q\,\mathrm{Cov}(\Delta S^D_\tau, \Delta F_\tau \mid \mathcal{F}_t)/\mathrm{Var}(\Delta F_\tau \mid \mathcal{F}_t)$. In the driftless diffusion approximation $\Delta S^D_\tau \approx \sigma_d S^D_t \Delta W^{(d)}_\tau$, $\Delta F_\tau \approx \sigma_f F_t \Delta W^{(f)}_\tau$,

$$h^\star = q\,\rho_{df}\,\frac{\sigma_d S_t^D}{\sigma_f F_t}. \tag{10}$$

In particular, $h^\star$ does not depend on $\lambda(g)$.

Proof. The objective in (9) is quadratic in h with a positive coefficient, and the first-order condition gives the covariance ratio. Conditional on τ, the increments are normal with zero mean, $\mathrm{Var}(\Delta S_\tau^D \mid \tau) = \sigma_d^2 (S_t^D)^2 \tau$ and $\mathrm{Cov}(\Delta S_\tau^D, \Delta F_\tau \mid \tau) = \rho_{df}\sigma_d\sigma_f S_t^D F_t \tau$. After averaging over τ, the common factor $\mathbb{E}[\tau \mid \mathcal{F}_t]$ cancels. ▫

The intuition is simple. Both increments grow with $\sqrt{\tau}$, and the correlation between them is constant. A longer delay therefore makes the risk larger, but it does not change the proportion between the two positions. If I substitute (10), the residual variance is

$$\mathrm{Var}(\Pi(g, h^\star) \mid \mathcal{F}_t) = \frac{q^2 (S_t^D)^2 \sigma_d^2 \left(1 - \rho_{df}^2\right)}{\lambda(g)}. \tag{11}$$

The factor $1 - \rho_{df}^2$ is the part of the DEX price variance that a linear hedge cannot move to the futures market. This has a clear economic meaning. The hedge moves most of the waiting risk to the CEX, where execution is immediate. Only the basis risk between the DEX price and the futures price remains exposed to the delay, and only this part can be reduced by paying for speed. If the drift is kept, there is an extra term proportional to $\mathrm{Var}(\tau) = 1/\lambda(g)^2$. For short horizons and moderate μ, this term is about ten times smaller than the main term.

3.4 Optimal priority fee

Substituting (11) into (9) gives

$$J(g, h^\star) = q\delta_t - c_{\mathrm{gas}}(g) - \frac{B_t}{\lambda(g)}, \qquad B_t := A(q,t) + \frac{\gamma}{2} q^2 (S_t^D)^2 \sigma_d^2 \left(1 - \rho_{df}^2\right). \tag{12}$$

Expected slippage and the risk premium for waiting are both multiplied by $1/\lambda(g)$. The fee problem therefore becomes $\max_{g \geq 0}\{-c_{\mathrm{gas}}(g) - B_t/\lambda(g)\}$. To get a closed-form solution, I use the following functional forms:

$$c_{\mathrm{gas}}(g) = p\,g, \quad p > 0; \qquad \lambda(g) = \kappa g^\beta, \quad \kappa > 0,\ \beta > 0, \tag{13}$$

Here $p = (\text{gas units}) \cdot S_t^C$ is the dollar cost of one unit of g. Under EIP-1559 the total payment is affine in the priority fee, so after I separate the exogenous base fee, the cost is linear in g (Buterin et al. 2019; Roughgarden 2020, 2021). The power form of the intensity is not a rule of the protocol. It is a convenient form that I later calibrate. For the theory, it is only important that λ is monotone and can be described with a small number of parameters.

Proposition 2. Under (13), the problem has the unique interior solution

$$g^{\star}_{\mathrm{nc}} = \left(\frac{\beta B_t}{\kappa p}\right)^{\frac{1}{\beta+1}}. \tag{14}$$

Proof. The objective is $\mathrm{const} - pg - B_t/\left(\kappa g^{\beta}\right)$. Its derivative $-p + \beta B_t \kappa^{-1} g^{-(\beta+1)}$ equals zero at (14), and the second derivative $-(\beta+1)\beta B_t \kappa^{-1} g^{-(\beta+2)} < 0$ shows that this is a strict maximum. ▫

Corner solution. In practice, the fee cannot be lower than the market minimum g_{lo} that still allows inclusion in the current state of the network. So the real problem is $\max_{g \geq g_{\mathrm{lo}}} J(g)$, and $g^{\star} = \max(g^{\star}_{\mathrm{nc}}, g_{\mathrm{lo}})$. If $g^{\star}_{\mathrm{nc}} < g_{\mathrm{lo}}$, the benefit of more speed is already smaller than its cost. This happens in calm periods, when β is large and p is small. As I show later, this is the relevant case for 2026.

Comparative statics. The hedge $h^{\star}$ increases with q, σ_d and ρ_{df}. When the correlation is close to one, the futures position moves almost all of the waiting risk to the CEX. The fee $g^{\star}_{\mathrm{nc}}$ increases with everything that raises B_t: expected slippage $A(q,t)$, risk aversion γ, volatility σ_d, size q and the unhedgeable share $1 - \rho^2_{df}$. In other words, it makes sense to pay more for speed when the trade is large or poorly hedged. A higher price of priority (p) or a more fee-sensitive inclusion rate (higher κ or β) lowers $g^{\star}$, because the same λ becomes cheaper to reach. Equation (14) can therefore be read as the arbitrageur's demand for speed: the willingness to pay for priority rises with the cost of waiting, B_t, and falls with the price of priority, p.

3.5 Dynamic formulation with controlled execution intensity

The model above is static, because one decision (g, h) is made at time t. In reality, the arbitrageur watches the market all the time and can change both controls. If g changes over time, I can no longer use a fixed distribution for τ. Instead, I model execution as the first jump of a counting process N_t with controlled intensity:

$$\mathbb{P}(dN_t = 1 \mid \mathcal{F}_{t-}) = \lambda(g_t, \zeta_t)\, dt + o(dt), \qquad \tau := \inf\{t \geq 0 \colon N_t - N_{t-} = 1\}, \tag{15}$$

where ζ_t is the exogenous level of network congestion. If $g_t \equiv g$ is constant, this gives (2) again. The prices of the CEX asset, the DEX asset and the futures follow correlated geometric Brownian motions. Before execution, wealth X_t changes only because of the hedge, $dX_t = h_t\, dF_t$. At the moment of execution, wealth jumps by the net gain from the trade:

$$X_{\tau} = X_{\tau-} + G(S^C_{\tau}, S^D_{\tau}, L_{\tau}; q, g_{\tau}), \qquad G(s_c, s_d, l; q, g) := q(s_c - s_d) - c_{\mathrm{slip}}(q, l) - c_{\mathrm{gas}}(g) - c_{\mathrm{fix}}. \tag{16}$$

Gas appears in the jump and not in the drift of wealth, because under EIP-1559 the fee is paid only when the transaction is included (Roughgarden 2021). The agent maximises $\mathbb{E}[U(X_T)]$ for a concave increasing U. I split the interval $[t, t+\delta t]$ into the cases with and without a jump, apply Itô's formula when there is no jump and let $\delta t \to 0$ (see Appendix A). This gives the HJB equation

$$0 = \partial_t V + \sup_{(h,g)} \{\mathcal{L}^h V + \lambda(g,\zeta)\big[U\big(x + G(s_c, s_d, l; q, g)\big) - V\big]\}, \qquad V(T,\cdot) = U(x), \tag{17}$$

where $\mathcal{L}^h$ is a diffusion operator of the Merton (1971) type. The fee affects the value function only through the jump term. The Hamiltonian can be split by controls: h appears only in $\mathcal{L}^h V$, which is quadratic in h, and g appears only in the jump term. The first-order condition for the fee (Eq. (A.4) in Appendix A) says that, at the optimum, the marginal utility gain from faster execution equals the marginal cost of gas. Under CRRA utility, and when trade size and costs are proportional to wealth, the problem reduces to a one-dimensional equation in the log-spread $u := \log(S^C/S^D)$ (Eq. (A.5)), which I solve numerically (Section 3.9 and Appendix B). Near the terminal condition, its solution reduces to the closed-form fee (14) and hedge (10), and the corrections matter only when the residual volatility or the gain from the trade is large (Eq. (A.6)). In other words, the simple rules of Sections 3.3 and 3.4 are a first-order approximation of the dynamic solution. Appendix C gives extensions to several hedging instruments, the choice between options and futures, and a portfolio of opportunities that appear at the same time. I do not test these extensions empirically.

3.6 Data

I use three data sources:

1. Binance REST: hourly and minute open–high–low–close–volume (OHLCV) candles for ether (ETH) against Tether (USDT) from 1 January 2021 to 5 May 2026 (about 46,000 hourly bars, and 173,000 minute bars for January–April 2026).

2. The Graph and GeckoTerminal: hourly prices of the Uniswap v3 pool of ETH and USD Coin (USDC) with a 0.05% fee tier on Ethereum mainnet (43,841 bars, 5 May 2021 to 6 May 2026) and raw swap events (732,530 records, January–April 2026).

3. Etherscan and an Ethereum remote procedure call (RPC) node: the gas oracle's proposed price and baseFeePerGas sampled from blocks in nine historical periods, about 2,400 blocks in total.

After synchronisation, the hourly sample has 43,812 observations, with $\hat{\rho}_{cd} = 0.9691$ and a median absolute spread $|\delta_t|$ of USD 1.07. I call an hour an opportunity if $|\delta_t|$ is above the 75th percentile of its rolling 24-hour distribution. This is similar to threshold entry rules in spread trading, where a position is opened when the spread moves far from its recent history (Chen and Lin 2017; Gatev et al. 2006). Using a rolling percentile lets the rule adapt to changing noise. With this definition I get 10,966 opportunities. In the high-frequency (block-level) analysis, a swap is classified as arbitrage if $|\delta_t| \geq 3$ basis points (bps) and its size is at least 1 ETH (Section 4.7).

In all comparisons, the benchmark is a naive rule, $g_{\text{naive}} = g_{\text{propose}}$, which is the gas price proposed by the network oracle. This rule ignores both the position size and the current spread, and it represents an arbitrage bot without optimisation. All strategies (HJB, closed form (CF) and naive) are evaluated on the same opportunities. The absolute profit and loss (PnL) depends on how price dynamics during the wait are simulated, but differences between strategies are still valid comparisons. For this reason, I focus more on risk profiles (variance and tails) than on mean PnL.

I test three hypotheses. H1: the hedge (10) substantially reduces PnL variance. H2: the advantage of the model over the naive rule grows with network congestion. H3: the fees from the HJB solution and from the closed form (14) are the same.

3.7 Stylised facts and assumptions

Negligible drift. For horizons $\tau \leq 1$ hour, the ratio $|\hat{\mu}|\tau/(\hat{\sigma}\sqrt{\tau})$ estimated from hourly log-returns is not higher than 0.05 for any horizon from 12 seconds to one hour. This supports setting the drift terms to zero.

Diffusion approximation. Hourly CEX and DEX log-returns have excess kurtosis and heavier tails than the normal distribution, and the Jarque–Bera test rejects normality ($p < 10^{-6}$). I do not think this is critical, because the geometric Brownian motion (GBM) approximation is used only for the first two moments of increments over a short horizon. Rolling annualised volatility on a 24-hour window does not show a clear structural drift, which agrees with the assumption of locally constant volatility.

Stable correlation. Over the five-year period, the rolling 24-hour correlation $\hat{\rho}_{cd}$ has a mean of 0.9865 and a coefficient of variation of 4.0%, and it stays above 0.95 almost all the time. Therefore, I treat the correlation as constant within one arbitrage cycle.

3.8 Calibration

Inclusion intensity. With historical data, it is not possible to recover individual priority bids from the mempool cleanly in all regimes. For this reason, I estimate $\lambda(g)$ as an approximate relation between the fee and the rate of inclusion. In the empirical part, g is the total gas price in gwei (1 gwei = 10^{-9} ETH). I estimate the probability of inclusion in the next block from the empirical distribution of the base fee, $\widehat{\mathbb{P}}(\text{incl} \mid g) = \text{ECDF}(\text{baseFee})|_{\text{baseFee}\le g}$, and convert it into an inclusion intensity using the 12-second block time. This is not a microstructure model of how block builders choose transactions. It is only a monotone empirical approximation of the link between the fee and inclusion speed, and it agrees with theoretical and empirical studies of EIP-1559 (Liu et al. 2022; Roughgarden 2020). A log-linear regression of $\log\hat{\lambda}$ on $\log g$ for April 2026 gives an elasticity of the inclusion intensity with respect to the fee of $\hat{\beta} = 3.56$ ($R^2 = 0.76$). The high $\hat{\beta}$ and the low base fee show that the network was calm. I repeat the same procedure for eight historical periods with different congestion (Section 4.4).

Slippage. I calibrate the reduced form of Assumption 2 on realised price impact in pool swaps (Adams et al. 2025; Angeris et al. 2021; Chemaya et al. 2025). Impact in basis points is modelled as $\alpha_{\text{slip}}(Q \cdot S)^{\beta_{\text{slip}}}$. A log–log regression of the absolute price move on swap size gives $\hat{\alpha}_{\text{slip}} = 0.112$ and $\hat{\beta}_{\text{slip}} = 0.465$.

Table 1 shows the calibrated parameters. The position size $Q = 10$ ETH is about USD 24,000 at $S_0 \approx 2{,}390$.

Table 1 Calibrated parameters of the hourly model (April 2026)

Parameter	Value	Parameter	Value
$\hat{\sigma}_c$	0.740	$\hat{\beta}$	3.56
$\hat{\sigma}_d$	0.754	$\hat{g}_{\text{lo}}$	1.02 gwei
$\hat{\sigma}_f$	0.495	$\hat{g}_{\text{hi}}$	11.0 gwei
$\hat{\rho}_{cd}$	0.969	$\hat{\alpha}_{\text{slip}}, \hat{\beta}_{\text{slip}}$	0.112, 0.465
Q	10 ETH		

Note: volatilities are annualised. g_{lo} and g_{hi} bound the admissible fee range.

3.9 Numerical solution of the HJB equation

I solve the reduced equation (A.5) backwards in time with an operator-splitting scheme. At each time step, the jump part is updated exactly, with the fee chosen at each grid node from 150 candidate values, and then an implicit Crank–Nicolson step is made for the diffusion

part. Appendix B gives the details. The final grid has $N_z = 2{,}000$ spatial nodes and $N_t = 3{,}000$ time steps, and the computation takes about three minutes on a standard CPU.

3.10 Use of generative AI

A generative AI tool (Claude, Anthropic) was used to translate the text of the manuscript from Russian, to shorten and restructure it, and to improve its readability and language. The tool was not used for the model, the data, the computations or the interpretation of the results. I reviewed and edited all text produced with its assistance and take full responsibility for the content of the article.

4 Results

4.1 How much of the waiting risk can be hedged (H1)

For H1, with $\hat{\rho}_{cd} = 0.969$ the residual variance after hedging is $1 - \rho^2 \approx 6.1\%$ of the unhedged variance. This means the hedge removes 93.9% of the variance of PnL during the wait. The empirical hedge ratio $h^\star = \rho\,\sigma_d/\sigma_f \approx 1.475$ is stable over the whole price series. In the calm regime, Levene's test finds no significant difference between the standard deviations of CF and naive PnL (77.99 vs 78.02; $p = 0.989$). So, in this regime, the model does not add extra mean profit, but it also does not make anything worse: moving from the naive rule to the closed form does not worsen any of the risk measures. For an arbitrageur, this means that most of the risk of waiting can be removed without paying anything for speed.

4.2 The optimal fee in a calm market and consistency of the two solutions (H3)

For the base model, I compare the closed-form fee (14) with the numerical maximiser $g^\star_{\text{grid}} := \text{argmax}_g J(g)$ under the same parameters. The difference is $\left|g^\star_{\text{cf}} - g^\star_{\text{grid}}\right| = 0.008$ gwei, which is smaller than the grid step of 0.07 gwei.

The one-dimensional HJB solution calibrated to April 2026 gives a U-shaped value function. $\varphi(0, u)$ equals 0.997 at $u = 0$, where there is no spread and nothing to execute, and falls to 0.887 at the boundaries $|u| = u_{\max}$, where immediate execution is valuable. The optimal control is $g^\star(0, u) \equiv g_{\text{lo}} = 1.02$ gwei. With the calibrated intensity parameters and a low base fee, the interior optimum (14) is below the admissible minimum, so the solution is at the corner. This comes from the model itself and is not a numerical error. Because $\hat{\beta} = 3.56$ is high, λ is very sensitive to g, and the objective decreases in g on $[g_{\text{lo}}, g_{\text{hi}}]$. In a calm network, there is therefore no reason to overpay for priority: the remaining risk is small, and speed is already cheap. The weak dependence of $g^\star$ on u also agrees with the theory: the

current spread does not enter (14), and in the HJB solution the dependence on u is weak when residual diffusion is small.

I checked grid convergence by comparing $N_z = 2{,}000$ with $N_z = 4{,}000$: $\| \varphi^{(2000)} - \varphi^{(4000)} \|_\infty = 1.79 \cdot 10^{-7}$ and $\| g^{\star(2000)} - g^{\star(4000)} \|_\infty = 0$ gwei. To check the one-dimensional reduction, I also solved the full two-dimensional HJB in $(\log S^C, \log S^D)$ on a 200×200 grid with $N_t = 1{,}000$ (about three hours on a CPU). Along the diagonal section, the one-dimensional and two-dimensional value functions are the same in the centre of the domain and differ mainly near the grid boundary. The optimal fee is almost everywhere equal to g_{lo}, so the two-dimensional solution is mainly useful as a check of the reduction. Overall, H3 is supported: in the calm regime $g^\star_{\text{HJB}} \equiv g^\star_{\text{cf}} = g_{\text{lo}}$.

4.3 Backtest design

I compare the strategies on the 10,966 hourly opportunities. The strategies are the HJB policy $g^\star_{\text{HJB}}(u)$ from (A.5), the closed form (14) with updated inputs, and the naive rule. For each opportunity, the direction of the trade is given by the sign of the spread, and the hedge (10) is applied. PnL is calculated net of gas and slippage, using the realised price move over a delay $\tau \sim \text{Exp}\big(\lambda(g)\big)$. I measure performance with the per-trade Sharpe ratio.

4.4 When does paying for speed pay? Cross-regime backtest (H2)

My main test of H2 changes gas conditions but keeps prices fixed. The price data stay the same, and the gas-market parameters $(\kappa, \beta, \text{baseFee})$ are recalibrated from block samples of nine historical periods. In this way I separate the effect of congestion from the effect of price dynamics. The question is: for the same arbitrage opportunities, how much does the model gain over the naive rule in different network states?

Table 2 Cross-regime comparison of the closed-form and naive strategies

Regime	Median gas, gwei	Sharpe, CF	Sharpe, naive	ΔSharpe
Calm, Apr 2026	1.0	0.398	0.398	0.000
NFT boom, Sep 2021	59.8	0.089	0.035	+0.054
Congested, Dec 2021	72.6	0.016	−0.044	+0.060
Moderate, May 2022	42.6	0.183	0.070	+0.113
Terra collapse, Jun 2022	31.7	0.233	0.170	+0.063
FTX collapse, Nov 2022	10.5	0.326	0.276	+0.049
Banking stress, Mar 2023	21.0	0.293	0.259	+0.034
Bald token, Aug 2023	17.1	0.310	0.275	+0.035
Dencun upgrade, Mar 2024	44.1	0.163	0.121	+0.043

Note: NFT, non-fungible token. ΔSharpe = Sharpe(CF) − Sharpe(naive), per trade. Prices are the same in all regimes; only the gas-market parameters change.

The closed-form rule is better than the naive rule in eight of the nine regimes. In the calm regime the gain is zero because of the corner solution (Table 2). The Spearman rank correlation between median gas and ΔSharpe is 0.60 ($p = 0.088$). The sign supports H2, but with only nine regimes the result is significant only at the 10% level. In some regimes the effect is economically large. In the moderate regime of May 2022, $\Delta\text{Sharpe} = +0.113$, and the strategy changes from weakly profitable to clearly more stable (Sharpe 0.07 to 0.18). In the heavily congested regime of December 2021, the naive rule has a negative Sharpe ratio (-0.044), while the closed-form rule stays positive. I also ran a sensitivity analysis on a grid of position sizes and regimes. For larger Q, the gain is positive in all regimes except the calm one. This agrees with the theory, because B_t in (12) grows with q^2, so the risk premium becomes more important. Note that $g^\star$ is recalculated for each opportunity from the current model inputs; it is not one constant for all trades. Put simply, a smarter fee rule is worth most when the network is busy. Such periods are often also periods of market stress, such as the Terra and FTX collapses in Table 2.

4.5 Sensitivity to risk aversion and position size

Without the lower bound, the optimal fee $g^\star_{\text{nc}}$ increases with γ, because a more risk-averse agent is ready to pay more for speed. It also increases with Q, because a larger position makes waiting more costly. The expected delay $\mathbb{E}[\tau](g^\star)$ falls accordingly. An approximate Sharpe ratio on the same grid is not monotone in γ. It is lower for small γ, because volatility is not hedged enough, and it is lower for large γ, because too much is spent on gas. Under this calibration, it generally increases with Q.

4.6 Rolling out-of-sample check

In the rolling exercise, the training window is 14 days and the test window is 7 days. Across 12 folds, the mean out-of-sample Sharpe ratios of the HJB, CF and naive strategies are the same to two decimal places. I expected this, because most folds are in the calm regime, where the model and the naive rule are equivalent. The main conclusion is that the model does not get worse out of sample when it is recalibrated on a rolling basis.

4.7 Validation on real on-chain arbitrage swaps

The hourly analysis uses spreads built from hourly bars. To test the model on real arbitrage events, I take all 732,530 swaps in the Uniswap v3 ETH/USDC pool in January–April 2026 and classify as arbitrage the swaps with a spread of at least 3 bps and a size of at

least 1 ETH. This gives 182,562 on-chain arbitrage trades (24.9% of all swaps in the pool). For each of them I know the gas price and the realised PnL.

At minute frequency (prices resampled from the last swap in each minute), the calibration gives $\hat{\rho}_{cd} = 0.997$, $\hat{\sigma}_c = \hat{\sigma}_d = 0.640$ and $\hat{\beta} = 4.01$ ($R^2 = 0.90$). The correlation is higher because arbitrage is almost instant at short horizons and because last-trade resampling smooths the prices. Returns are again non-normal, and the distribution of gas prices is strongly right-skewed. The HJB solution with these parameters gives $\varphi \in [0.887, 0.999]$ and $g^\star \equiv g_{\text{lo}} = 0.22$ gwei, which is again a corner solution. The grid comparison gives the same results for $N_z = 2{,}000$ and $4{,}000$.

For each of the 182,562 trades, each strategy proposes a gas bid. If the bid is at least the realised winning gas price, I assume that the strategy would have been included in that block and would have earned the same PnL. Otherwise, the strategy skips the trade.

Table 3 High-frequency backtest on real arbitrage swaps (Uniswap v3 ETH/USDC, January–April 2026)

Strategy	n	Total PnL, USD m	Sharpe ratio	Mean g, gwei	Included, %
Actual arbitrageurs	182,562	16.31	0.062	10.53	—
Closed form	84,274	6.25	0.049	0.56	46.2
HJB	60,320	4.93	0.046	0.22	33.0
Naive	75,585	5.77	0.048	0.38	41.4

Note: a strategy is counted as included when its bid is at least the realised winning gas price, in which case it is assigned the realised PnL of that swap.

Actual arbitrageurs earn much more in absolute terms (USD 16.3m against USD 5–6m). This is because they take part in all opportunities and win some of them by bidding tens or hundreds of gwei. The agent in my model is a price-taker, and a fair comparison with these traders would require a model of the priority auction. For this reason, the comparison between the model strategies is more informative. A paired t-test on PnL per trade shows that CF earns USD 2.59 more than the naive rule ($t = 15.57$, $p \approx 0$). HJB is not statistically different from the naive rule in this calm period. The reason is that $g^\star_{\text{HJB}} = g_{\text{lo}} < g_{\text{naive}}$, so HJB skips slightly more opportunities.

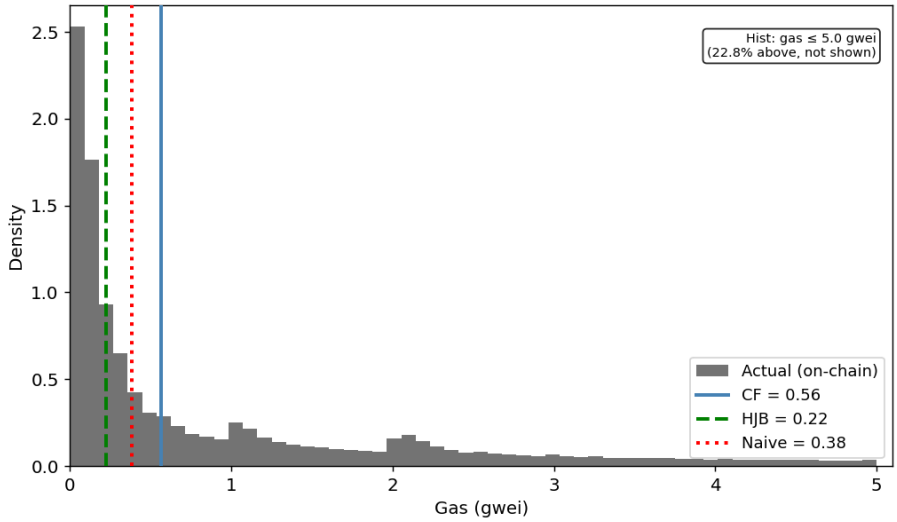


Fig. 2 Gas prices in real arbitrage swaps and model bids. The histogram shows the density of gas prices paid in on-chain arbitrage swaps in the Uniswap v3 ETH/USDC pool, January–April 2026 (values up to 5 gwei; 22.8% of swaps lie above this level and are not shown). Vertical lines mark the mean bids of the closed-form (CF, solid), HJB (dashed) and naive (dotted) strategies in gwei

The mean winning bid of actual arbitrageurs (10.5 gwei) is driven by a long right tail, with a maximum of 27,000 gwei for MEV bots. The median, however, is only 0.78 gwei, which is close to the model bids of 0.22–0.56 gwei (Fig. 2). The CF bid is higher than the realised winning bid in 46% of cases. This shows that the model does not behave like an MEV searcher. Instead, it behaves like a rational price-taking arbitrageur that bids near the median of the market distribution.

I also repeated the cross-regime exercise at block frequency, with $(\kappa, \beta, \text{baseFee})$ recalibrated to the historical regimes and inclusion recalculated for the same swaps. The result is the same as in the hourly data: ΔSharpe is larger in congested and moderate regimes and zero in the calm one.

5 Discussion

The results support H1 and H3. They also give some support for H2, but this support is statistically weak. In this section I explain what the results mean in economic terms and for different market participants. Then I discuss the limitations of the study.

5.1 Economic insights

Hedge first, then pay for speed. Execution risk in CEX–DEX arbitrage has two parts. Most of it is ordinary price risk, and a futures hedge can move it to the CEX. With the observed correlation, only about 6% of the variance remains after hedging. The rest is basis

risk between the pool price and the futures price. Only this part, together with expected slippage, is worth paying for with a higher fee, as Eq. (12) shows. This explains a result that may look surprising at first: the best hedge does not depend on the fee, but the best fee depends a lot on how well the trade is hedged. A trader who cannot hedge, or who hedges with a poorly correlated instrument, should pay much more for speed.

The value of speed depends on the state of the market. Equation (14) can be read as a demand curve for priority. The willingness to pay rises with the cost of waiting, that is, with trade size, risk aversion, volatility and the unhedged share of risk. It falls when priority is expensive or when inclusion reacts strongly to the fee. In a calm network, inclusion is fast even at the minimum fee, and the remaining risk is small. So the optimal fee is at the lower bound, and a simple rule is as good as an optimised one. In congested periods, the minimum fee no longer gives fast inclusion. Then the gap between a well-chosen fee and a fixed rule becomes economically important (Table 2). This agrees with the sharp rise in priority fees during congestion found in earlier work (Capponi et al. 2026; Liu et al. 2022).

Most very high fees are not paid for protection against waiting. On real swaps, the model bids are close to the median winning bid. They are far below the bids of MEV searchers in the right tail of the distribution (Fig. 2). This suggests that a large part of very high priority fees is paid for reasons that my model does not include. One example is competition with other traders for the same opportunity. Protection against price risk during the wait seems to be a smaller reason. This agrees with Capponi et al. (2026), who find that very high fees are often linked to competition for priority.

5.2 Practical implications

The implications for arbitrageurs below follow directly from the model. The implications for liquidity providers, protocol design and investors are not derived or tested in this paper; they are hypotheses that follow from the risk premium in Eq. (12) and that a model of the whole market would need to check.

Arbitrageurs and trading firms. A practical rule follows directly from the results: always hedge the waiting period on the CEX with the minimum-variance ratio, and optimise the fee actively only for large positions and in congested periods. The empirical hedge ratio is stable over the sample, so it does not need frequent recalibration. In the model, a trade is worth entering only if the spread covers gas, expected slippage and the risk premium $B_t/\lambda(g)$ in Eq. (12). This means that firms that cannot use CEX derivatives carry more execution risk and need wider spreads before they trade.

Liquidity providers and DEX design. Arbitrage keeps pool prices close to the market, but liquidity providers pay for it through their losses to arbitrageurs (Milionis et al. 2024). The model shows that the spread an arbitrageur needs before trading depends on execution risk. When this risk is higher, for example in congested periods or for traders who cannot hedge, larger price gaps are needed before arbitrage starts. For liquidity providers, this suggests that their losses to arbitrageurs may depend not only on volatility but also on the state of the fee market. Designers of pools and fee tiers could take this channel into account.

Fee-market and protocol design. For blockchain designers, the results show that the waiting time is a real economic cost for the traders who connect on-chain and off-chain markets. Any change that makes inclusion faster or more predictable, such as shorter block times or more stable base fees, lowers the risk premium in Eq. (12). Arbitrageurs would then need to pay lower fees for the same level of risk.

Investors and risk managers. For investors who trade on DEXs, the quality of pool prices depends on arbitrage. The results show that arbitrage becomes more costly and more risky in congested periods, and these periods often coincide with market stress (Table 2). In the most congested regime in my sample, the fixed-fee rule had a negative Sharpe ratio, while the optimised rule stayed positive. For risk managers and supervisors, the results show that hedging links on-chain risk to CEX derivatives markets. A large part of the risk taken by DEX arbitrageurs is moved to futures markets. So conditions in these markets also matter for how well DEX prices are kept in line.

5.3 Limitations

The study has several limitations.

Optimistic inclusion rule in the on-chain test. A model strategy receives the realised PnL of a swap whenever its bid is at least the realised winning bid. This ignores how competing searchers would have reacted, and it favours opportunities where competition was weak. Therefore, the absolute PnL of the model strategies in Table 3 is an optimistic estimate. The comparison between the model strategies is less affected, because they all use the same rule. A stricter test would simulate the priority auction.

Price-taking agent. The model does not describe an MEV searcher that actively competes in the priority-fee auction. Calibrating $\lambda(g)$ from the empirical distribution of the base fee is suitable for a price-taker, but it does not capture strategic bidding or the block-builder market.

Cross-regime design. In Section 4.4, prices are fixed and only the gas-market parameters change. This isolates the effect of congestion. However, historical high-fee periods also had

higher price volatility. Since B_t increases with σ_d, the gains in Table 2 may understate the real advantage in those regimes, but I have not tested this.

Statistical power. The Spearman test of H2 is based on only nine regimes ($p = 0.088$). My conclusion relies on the sign pattern (a positive gain in eight of nine regimes) as much as on the correlation coefficient. More regimes, or bootstrap inference within regimes, would make the result stronger.

Reduced-form inclusion model. The inclusion intensity $\lambda(g)$ is estimated from the empirical distribution of the base fee and not from a structural model of how block builders select transactions. The model therefore gives decision rules for a single trader but no equilibrium or welfare analysis of the fee market.

Scope. The experiments cover only one pair, ETH/USDC in the 0.05% pool. Less liquid pools would need a separate calibration and could change the form of Assumption 2. The extensions in Appendix C are not tested, because a portfolio of opportunities needs parallel data from several pools.

6 Conclusions

In this paper I study execution risk in arbitrage between centralised and decentralised crypto exchanges. The CEX part of the trade is filled at once, while the DEX part waits for a random time that depends on the priority fee. I show how an arbitrageur should choose the fee and a hedge together. In a mean–variance model, the minimum-variance hedge does not depend on the fee. The optimal fee increases with slippage, risk aversion, volatility, position size and the share of risk that cannot be hedged. A dynamic model with controlled execution intensity confirms these closed-form rules as a first-order approximation.

After calibration on Binance, Uniswap v3 and Ethereum data, the hedge removes 93.9% of the price variance during the wait. In the calm market of 2026, the optimal fee is usually at its lower bound, so the optimised and simple fee rules give almost the same outcome. This is caused by the fee constraint and is not a failure of the method. Under historical congestion, the closed-form rule improves the per-trade Sharpe ratio in eight of nine gas regimes. On real arbitrage swaps, it earns USD 2.59 more per trade than a fixed-fee rule.

The main message is simple: hedge the wait first, and pay for speed only when the remaining risk is large, that is, for big positions and in congested networks. More generally, the cost of speed on a blockchain is part of the cost of keeping DEX prices in line with the wider market. Arbitrageurs, liquidity providers and, indirectly, DEX users all bear this cost. Possible directions for future work are a strategic model of the priority auction with

competing arbitrageurs, a structural model of the inclusion intensity, joint recalibration of prices and gas regimes, tests on several pools, pairs and hedging instruments, and an analysis of how execution risk affects the losses of liquidity providers and the efficiency of DEX prices.

Appendix A: The dynamic model

Let $\mathcal{L}^h$ be the diffusion operator acting on $V(t,x,s_c,s_d,f,l)$ under $dX_t = h_t\, dF_t$ and the GBM dynamics of S^C, S^D and F, with liquidity locally constant:

$$\begin{aligned}\mathcal{L}^h V = \quad & h\mu_f f V_x + \mu_c s_c V_{s_c} + \mu_d s_d V_{s_d} + \mu_f f V_f + 1/2\, h^2\sigma_f^2 f^2 V_{xx} + 1/2\, \sigma_c^2 s_c^2 V_{s_c s_c} \\ & + h\sigma_f^2 f^2 V_{xf} + h\rho_{cf}\sigma_c\sigma_f s_c f V_{x s_c} + h\rho_{df}\sigma_d\sigma_f s_d f V_{x s_d} + \rho_{cd}\sigma_c\sigma_d s_c s_d V_{s_c s_d}\end{aligned} \tag{A.1}$$

By the dynamic programming principle, $V(t,Y_t) = \sup\mathbb{E}_t[V(t+\delta t \wedge \tau,\, Y_{t+\delta t\wedge\tau})]$. Let $A_t^{\delta t} := \{\tau \in [t, t+\delta t]\}$. By (15), $\mathbb{P}\big(A_t^{\delta t} \mid \mathcal{F}_t\big) = \lambda(g_t,\zeta_t)\,\delta t + o(\delta t)$. Conditioning on the jump event, I get

$$\begin{aligned}&\mathbb{E}_t[\cdots] \\ &= \lambda\,\delta t\; U\big(x + G(s_c,s_d,l;q,g_t)\big) + (1-\lambda\,\delta t)\,\mathbb{E}_t[V(t+\delta t, Y_{t+\delta t})] + o(\delta t).\end{aligned} \tag{A.2}$$

When there is no jump, Itô's formula gives $dV = (V_t + \mathcal{L}^h V)\,dt + (\nabla_W V)\cdot dW$, and the conditional expectation removes the martingale term: $\mathbb{E}_t[V(t+\delta t, Y_{t+\delta t})] = V + (V_t + \mathcal{L}^h V)\,\delta t + o(\delta t)$. I substitute this into (A.2), cancel V, divide by δt and let $\delta t \to 0$, which gives (17). Equation (17) is a necessary condition if V is regular enough. A verification theorem needs additional conditions on admissible controls and on the growth of the value function (Fleming and Soner 2006; Øksendal and Sulem 2007; Pham 2009). Admissible controls are bounded, piecewise-continuous, predictable processes with $h \in [h_{\min}, h_{\max}]$ and $g \in [g_{\text{lo}}, g_{\text{hi}}]$.

First-order condition for h. The part of (A.1) that depends on h is quadratic, with coefficient $(1/2)\sigma_f^2 f^2 V_{xx} < 0$ because U is strictly concave. Setting its derivative to zero gives

$$h^\star = -\frac{\mu_f f V_x + \sigma_f^2 f^2 V_{xf} + \rho_{cf}\sigma_c\sigma_f s_c f V_{x s_c} + \rho_{df}\sigma_d\sigma_f s_d f V_{x s_d}}{\sigma_f^2 f^2 V_{xx}}. \tag{A.3}$$

If V does not depend on prices, (A.3) is the Merton (1971) rule. With the mixed derivatives and the CRRA factorisation, it becomes $\alpha^\star = \big(\mu_f\varphi + \sigma_f^2 f\varphi_f + \rho_{cf}\sigma_c\sigma_f s_c\varphi_{s_c} + \rho_{df}\sigma_d\sigma_f s_d\varphi_{s_d}\big)/\big(\gamma\sigma_f^2\varphi\big)$. This reduces to the minimum-variance rule (10) when φ depends only weakly on prices.

First-order condition for g. Differentiating the Hamiltonian in (17) with respect to g gives

$$\partial_g \lambda(g^\star, \zeta)\,[U(x+G) - V] = \lambda(g^\star, \zeta)\,U'(x+G)\,c_{\text{gas}}{}'(g^\star), \tag{A.4}$$

It says that, at the optimum, the marginal utility gain from faster execution equals the marginal cost of gas.

With $c_{\text{gas}}(g) = pg$ and $\lambda(g) = \kappa g^\beta$, condition (A.4) becomes $\kappa\beta g^{\beta-1}[U(x+G) - V] = \kappa g^\beta U'(x+G)\,p$. Where V and $U(x+G)$ can be expressed through B_t, this is equivalent to (14).

CRRA reduction. Let $U(x) = x^{1-\gamma}/(1-\gamma)$, and let trade size, hedge and costs be proportional to wealth ($q_t = \nu_t X_t$, $h_t = \alpha_t X_t / F_t$, $c_{\text{gas}} = X_t \tilde{c}_{\text{gas}}$). Then the value function can be written as $V = \left(x^{1-\gamma}/(1-\gamma)\right)\varphi(t, s_c, s_d, f, l)$ (Merton 1971; Pham 2009). The jump term depends on prices only through the log-spread $u := \log(S^C/S^D)$. The dynamics of u are closed, with residual volatility $\sigma_{\text{res}}^2 := \sigma_c^2 - 2\rho_{cd}\sigma_c\sigma_d + \sigma_d^2$. If the hedge is locally optimal and the drift is negligible, the problem becomes the one-dimensional equation

$$\partial_t \varphi + 1/2\,\sigma_{\text{res}}^2\,\varphi_{uu} + \sup_{g \geq 0}\left\{\lambda(g,\zeta)\left[(1+\tilde{G}(u;\nu,g))^{1-\gamma} - \varphi\right]\right\} = 0, \qquad \tilde{G} = \nu(1-e^{-u}) - \tilde{c}_{\text{slip}} - \tilde{c}_{\text{gas}}(g), \tag{A.5}$$

with $\varphi(T,u) = 1$. The supremum is written for $\gamma > 1$. In this case $\varphi \geq 0$, and a smaller φ means a larger V. In the numerical solution, the fee is chosen so that φ moves towards $\left(1+\tilde{G}\right)^{1-\gamma}$ as fast as possible (Appendix B). Note that this reduction works only when all value quantities scale with wealth. With a fixed absolute position, homotheticity does not hold, and the problem has to be solved in the original state space.

Link to the closed form. I expand $\left(1+\tilde{G}\right)^{1-\gamma} = 1 - (\gamma-1)\tilde{G} + O\left(\tilde{G}^2\right)$ and set $\varphi \approx 1$ near the terminal condition. Then (A.4) becomes the mean–variance condition behind (14), with a remainder of order $O\left(\tilde{G}^2\right) + O(\sigma_{\text{res}}^2 T)$:

$$g^\star_{\text{HJB}}(y) = g^\star_{\text{cf}} + O(\| R \|). \tag{A.6}$$

In the same way, the HJB hedge reduces to (10) when φ depends only weakly on prices. So the closed form is a first-order approximation of the dynamic solution, and the corrections are important only when σ_{res} or $\tilde{G}$ is large.

Appendix B: Numerical solution of the HJB equation

I solve Equation (A.5) backwards in time from $\varphi(T,u) = 1$ using operator splitting. At each time step and grid node, I scan a discrete set of fees $g \in [g_{\text{lo}}, g_{\text{hi}}]$ and choose the value that maximises $R(g) = \lambda(g)\,[\varphi_i^n - \text{target}(g)]$, where $\text{target}(g) = \left(1+\tilde{G}\right)^{1-\gamma}$. In this way interior and corner solutions are treated in the same way; at a corner, the Kuhn–Tucker

condition applies. The jump part is then integrated exactly as a one-step ordinary differential equation,

$$\varphi_i^{n+1/2} = \text{target}_i^\star + (\varphi_i^n - \text{target}_i^\star)\, e^{-\lambda_i^\star dt}, \tag{B.1}$$

This is followed by an implicit Crank–Nicolson diffusion step, $(I - (dt/2)\mathcal{L})\varphi^{n+1} = (I + (dt/2)\mathcal{L})\varphi^{n+1/2}$, where $\mathcal{L} = (1/2)\sigma_{\text{res}}^2\, \partial_{uu}$ is discretised with a three-point stencil and Dirichlet conditions $\varphi = 1$ at $|u| = u_{\max}$. Unlike an explicit jump step, the update (B.1) is unconditionally stable and keeps $\varphi \in (0,1]$ for $\gamma > 1$ and $\tilde{G} \geq 0$. As a result, I do not need clipping or a Courant–Friedrichs–Lewy restriction for the fee. Other methods could also be used, such as value iteration, Howard policy iteration or semi-Lagrangian and monotone schemes. In this one-dimensional setting, however, they do not give better accuracy for the same computing time.

The full problem (17) has a five-dimensional state space (x, s_c, s_d, f, l). Even with 100 grid points per dimension, this would need about 10^8 nodes. The CRRA factorisation removes wealth, and the change of variable to u combines the price variables into one, given that liquidity is locally constant. The final grid has $N_z = 2{,}000$ spatial nodes, $N_t = 3{,}000$ time steps and 150 candidate fees. The computation takes about three minutes on a standard CPU.

Appendix C: Extensions

Vector hedge. Suppose n instruments with increments $\Delta H_\tau \in \mathbb{R}^n$ are available, and let $Y_\tau := q\, \Delta S_\tau^D$. With a quadratic penalty $h^\top K h$ ($K \succcurlyeq 0$) for transaction and margin costs and a linear cost $k^\top h$, the problem is $\min_h \text{Var}(-Y_\tau + h^\top \Delta H_\tau \mid \mathcal{F}_t) + h^\top K h + k^\top h$. If $\Sigma_{HH} + K \succ 0$,

$$h^\star = (\Sigma_{HH} + K)^{-1}(\Sigma_{HY} - 1/2\, k), \tag{C.1}$$

where $\Sigma_{HH} := \text{Cov}(\Delta H_\tau, \Delta H_\tau \mid \mathcal{F}_t)$ and $\Sigma_{HY} := \text{Cov}(\Delta H_\tau, Y_\tau \mid \mathcal{F}_t)$. For $n = 1$, $K = 0$ and $k = 0$, this gives (10). The matrix K works as a regulariser when the instruments are almost collinear.

Option versus futures. Using a delta–gamma approximation of the option over a short horizon (Hull 2018), the delta-hedge positions are $h_F^\star = q\rho_{df}\sigma_d S_t^D / (\sigma_f F_t)$ and $n_O^\star = q\rho_{dc}\sigma_d S_t^D / (\Delta_t \sigma_c S_t^C)$, with residual variances $q^2 (S_t^D)^2 \sigma_d^2 (1 - \rho^2)\, \mathbb{E}[\tau]$. Under criterion (6), the option delta hedge is better than the futures hedge if and only if the reduction in the variance penalty is larger than the option premium:

$$\frac{\gamma}{2} q^2 (S_t^D)^2 \sigma_d^2 \left(\rho_{dc}^2 - \rho_{df}^2\right) \mathbb{E}[\tau] > n_O^\star\, C_t. \tag{C.2}$$

On a single CEX, usually $\rho_{dc} \approx \rho_{df}$, so the left-hand side is close to zero and futures are preferred. The gamma term adds $(1/2)\Gamma_t^2\sigma_c^4(S_t^C)^4\tau^2$ to the residual variance. This is negligible for $\tau < 10$ s and becomes important only when $\mathbb{E}[\tau]$ is longer than about a minute.

Portfolio of opportunities. Suppose there are m opportunities at the same time, with expected reduced profits $\mu(g)$ and covariance $\Sigma_{\mathrm{red}}(g)$. The diagonal of $\Sigma_{\mathrm{red}}(g)$ is proportional to $\sigma_i^2/\lambda_i(g_i)$, and the off-diagonal terms come from common congestion. The agent solves $\max_w\{\mu(g)^\top w - (\gamma/2)\, w^\top \Sigma_{\mathrm{red}}(g) w\}$. This is a Markowitz (1952) problem with endogenous moments. For fixed g,

$$w^\star(g) = \frac{1}{\gamma}\Sigma_{\mathrm{red}}(g)^{-1}\mu(g). \tag{C.3}$$

By the envelope theorem, the outer problem in g needs only the explicit partial derivatives, $w_i^\star \mu_i'(g_i) - (\gamma/2)(w^\star)^\top(\partial\Sigma_{\mathrm{red}}/\,\partial g_i)w^\star$. A two-level iteration that alternates (C.3) with the first-order conditions in g does not decrease the objective at any step. For $m = 1$ it reduces to (14), and with diagonal Σ_{red} and fee-independent μ_i it splits into m independent problems.

Statements and Declarations

Competing interests The author has no relevant financial or non-financial interests to disclose.

Funding No funding was received for conducting this study.

Ethics approval Not applicable. The study uses only public market and blockchain data and involves no human participants or animals.

Data availability All data come from public sources: the Binance REST API, The Graph and GeckoTerminal for Uniswap v3, and Etherscan and an Ethereum RPC node for fee data. The processed data are available from the author on reasonable request.

Code availability The code used for calibration, the numerical solution and the backtests is available from the author on reasonable request.

Author contributions Kirill Proskurin is the sole author. He designed the study, collected and processed the data, wrote the code, carried out the analysis and wrote the manuscript.

References

Adams, A., Chan, B.Y., Markovich, S., Wan, X.: Don't let MEV slip: the costs of swapping on the Uniswap protocol. In: Financial Cryptography and Data Security. Lecture Notes in

Computer Science, pp. 172–191. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-78676-1_10

Adams, H., Zinsmeister, N., Salem, M., Keefer, R., Robinson, D.: Uniswap v3 core. Uniswap Labs white paper. https://uniswap.org/whitepaper-v3.pdf (2021). Accessed 5 October 2026

Almgren, R., Chriss, N.: Optimal execution of portfolio transactions. Journal of Risk 3, 5–39 (2001). https://doi.org/10.21314/JOR.2001.041

Angeris, G., Kao, H.-T., Chiang, R., Noyes, C., Chitra, T.: An analysis of Uniswap markets. Cryptoeconomic Systems 1 (2021). https://doi.org/10.21428/58320208.c9738e64

Bai, J., Li, D., Jiang, Z., Du, G.: Design and implementation of a real-time detection system for multi-token sandwich attacks in Ethereum based on Geth client. Discover Computing 28, 180 (2025). https://doi.org/10.1007/s10791-025-09694-z

Bertsimas, D., Lo, A.W.: Optimal control of execution costs. Journal of Financial Markets 1, 1–50 (1998). https://doi.org/10.1016/S1386-4181(97)00012-8

Brönnimann, W., Egloff, P., Krabichler, T.: Automated market makers and their implications for liquidity providers. Digital Finance 6, 573–604 (2024). https://doi.org/10.1007/s42521-024-00117-0

Buterin, V., Conner, E., Dudley, R., Slipper, M., Norden, I., Bakhta, A.: EIP-1559: fee market change for ETH 1.0 chain. Ethereum Improvement Proposal No. 1559. https://eips.ethereum.org/EIPS/eip-1559 (2019). Accessed 5 October 2026

Capponi, A., Jia, R., Yu, S.: Price discovery on decentralized exchanges. The Review of Financial Studies (2026). https://doi.org/10.1093/rfs/hhag002

Cartea, Á., Sánchez-Betancourt, L.: Optimal execution with stochastic delay. Finance and Stochastics 27, 1–47 (2023). https://doi.org/10.1007/s00780-022-00491-w

Chemaya, N., Liu, D., McLaughlin, R., Ruaro, N., Kruegel, C., Vigna, G.: The power of default: measuring the effect of slippage tolerance in decentralized exchanges. In: Financial Cryptography and Data Security. Lecture Notes in Computer Science, pp. 192–208. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-78676-1_11

Chen, C.W.S., Lin, T.-Y.: Nonparametric tolerance limits for pair trading. Finance Research Letters 21, 1–9 (2017). https://doi.org/10.1016/j.frl.2016.11.002

Daian, P., Goldfeder, S., Kell, T., Li, Y., Zhao, X., Bentov, I., Breidenbach, L., Juels, A.: Flash boys 2.0: frontrunning in decentralized exchanges, miner extractable value, and

consensus instability. In: 2020 IEEE Symposium on Security and Privacy, pp. 910–927. IEEE, San Francisco (2020). https://doi.org/10.1109/SP40000.2020.00040

Fleming, W.H., Soner, H.M.: Controlled Markov Processes and Viscosity Solutions, 2nd edn. Springer, New York (2006)

Gatev, E., Goetzmann, W.N., Rouwenhorst, K.G.: Pairs trading: performance of a relative-value arbitrage rule. The Review of Financial Studies 19, 797–827 (2006). https://doi.org/10.1093/rfs/hhj020

Gollier, C.: The Economics of Risk and Time. MIT Press, Cambridge, MA (2001)

Hägele, S.: Centralized exchanges vs. decentralized exchanges in cryptocurrency markets: a systematic literature review. Electronic Markets 34, 33 (2024). https://doi.org/10.1007/s12525-024-00714-2

Hull, J.C.: Options, Futures, and Other Derivatives, 10th edn. Pearson, New York (2018)

Kirste, D., Poddey, A., Kannengießer, N., Sunyaev, A.: On the influence of conventional and automated market makers on market quality in cryptoeconomic systems. Electronic Markets 34, 42 (2024). https://doi.org/10.1007/s12525-024-00723-1

Kramkov, D., Schachermayer, W.: The asymptotic elasticity of utility functions and optimal investment in incomplete markets. The Annals of Applied Probability 9, 904–950 (1999). https://doi.org/10.1214/aoap/1029962818

Lehar, A., Parlour, C.: Decentralized exchange: the Uniswap automated market maker. The Journal of Finance 80, 321–374 (2025). https://doi.org/10.1111/jofi.13405

Lim, T.: Predictive crypto-asset automated market maker architecture for decentralized finance using deep reinforcement learning. Financial Innovation 10, 144 (2024). https://doi.org/10.1186/s40854-024-00660-0

Liu, Y., Lu, Y., Nayak, K., Zhang, F., Zhang, L., Zhao, Y.: Empirical analysis of EIP-1559: transaction fees, waiting times, and consensus security. In: Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. ACM, New York (2022). https://doi.org/10.1145/3548606.3559341

Makarov, I., Schoar, A.: Trading and arbitrage in cryptocurrency markets. Journal of Financial Economics 135, 293–319 (2020). https://doi.org/10.1016/j.jfineco.2019.07.001

Markowitz, H.: Portfolio selection. The Journal of Finance 7, 77–91 (1952). https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

Merton, R.C.: Optimum consumption and portfolio rules in a continuous-time model. Journal of Economic Theory 3, 373–413 (1971). https://doi.org/10.1016/0022-0531(71)90038-X

Milionis, J., Moallemi, C.C., Roughgarden, T., Zhang, A.L.: Automated market making and loss-versus-rebalancing. arXiv preprint arXiv:2208.06046 (2024). https://doi.org/10.48550/arXiv.2208.06046

Naviglio, M., Tarantelli, F., Lillo, F.: Price manipulation schemes of new crypto-tokens in decentralized exchanges. EPJ Data Science 15, 10 (2026). https://doi.org/10.1140/epjds/s13688-025-00602-5

Øksendal, B., Sulem, A.: Applied Stochastic Control of Jump Diffusions, 2nd edn. Springer, Berlin (2007)

Pham, H.: Continuous-time Stochastic Control and Optimization with Financial Applications. Springer, Berlin (2009)

Pratt, J.W.: Risk aversion in the small and in the large. Econometrica 32, 122–136 (1964). https://doi.org/10.2307/1913738

Roughgarden, T.: Transaction fee mechanism design for the Ethereum blockchain: an economic analysis of EIP-1559. arXiv preprint arXiv:2012.00854 (2020). https://doi.org/10.48550/arXiv.2012.00854

Roughgarden, T.: Transaction fee mechanism design. ACM SIGecom Exchanges 19, 52–55 (2021). https://doi.org/10.1145/3476436.3476445

Saef, D., Nagy, O., Sizov, S., Härdle, W.K.: Understanding temporal dynamics of jumps in cryptocurrency markets: evidence from tick-by-tick data. Digital Finance 6, 605–638 (2024). https://doi.org/10.1007/s42521-024-00116-1