BTR ALM: An Empirical Anatomy of Concentrated-Liquidity Market Making

Measured Losses, Loss-Minimizing Geometry, and the Capacity Ceiling

Toni Rangers · Jake The Slug
BTR Research
btr.markets · contact@btr.markets

July 2026

Abstract

We present an empirical anatomy of why providing concentrated liquidity on Uniswap-v3-style venues fails for the liquidity provider: the mechanism, the measurements, the loss-minimizing geometry, and the capacity ceiling. The measurements cover five pairs on PancakeSwap v3 and Aerodrome Slipstream on Base (measurement campaign April–July 2026, with a two-year on-chain replay window), through three evaluators of increasing fidelity built around a deliberately small live deployment: a calibrated block-bootstrap Monte-Carlo engine; a tick-by-tick event-replay engine reconstructing pool state from complete on-chain histories (77.8M swaps over two years) and reproducing 0.97–1.00 of the fees actually collected by production positions; and live forward paper-trading arms. Four results emerge. (1) Liquidity provision "for the fees" is structurally negative-sum against arbitrageurs: pool-wide LP fees of $374k versus $662k of realized adverse selection (markout) over 90 days, and the same inequality in 14 of 14 pre-registered market regimes over two years. (2) Concentration aggravates the loss; the loss-minimizing geometry is a single wide (±10–15%) *lazy* range — the exact reversal of the calibrated simulator's recommendation, confirmed forward by the live arms. (3) The only positive revenue is protocol emissions, whose measured budget (≈$1.02M/yr across the farms studied) caps the opportunity: net-positive capacity versus buy-and-hold at product size is zero. (4) A reusable taxonomy of measurement artifacts for empirical DeFi. The detailed measurements are all from Base; BTR's internal research on other chains and venues reached the same qualitative conclusion. Negative results are the point of the paper.

1 Introduction

Concentrated-liquidity automated market makers (CLAMMs), introduced by Uniswap v3 [3], allow liquidity providers (LPs) to allocate capital to arbitrary price ranges rather than to the full price axis. The design multiplies the fee income earned per dollar of capital by a concentration factor that can exceed two orders of magnitude, and it has become the dominant venue architecture for spot trading on Ethereum rollups. It also created an industry of *automated liquidity managers* (ALMs): agents that choose a range, monitor the price, and rebalance the position when the price escapes, with the stated goal of turning fee income into a yield product for passive capital.

This paper is the concluding report of a two-year research program (2024–2026) on automated liquidity management. The program's earlier phases built the system whose measurements we report: the three-force market model at the mathematical core, the range optimizer and its fitness functional, and the multi-chain R&D and a previous engine implementation in Rust, both due to the second author, all predate the work described here. What this paper documents in detail is the program's final, intensive *measurement campaign* (April–July 2026), in which we operated the ALM continuously on Base, an Ethereum rollup, across five pairs on PancakeSwap v3 and Aerodrome Slipstream. The live deployment was kept deliberately at sandbox scale, because

its purpose was not income but *measurement*: every fee, emission, rebalance and gas cost was recorded on-chain and reconciled against benchmarks. Around this live anchor we built a hierarchy of three independent evaluators of increasing fidelity: (i) a Monte-Carlo engine driven by block-bootstrap resampling of five years of exchange returns, calibrated against the realized volatility and the realized divergence loss of our own production positions; (ii) a tick-by-tick *event-replay* engine that reconstructs the full state of a pool from its complete on-chain event history — 77.8 million swaps over two years for our two PancakeSwap pools — and whose fee accounting reproduces the fees actually collected by our production positions to within 0.97–1.00; and (iii) a set of live paper-trading arms that run candidate strategies forward, in parallel with production, on the same pools and the same prices.

The findings are negative, and they are the point of the paper. On the pools we studied, providing concentrated liquidity "for the fees" is a structurally negative-sum game against better-informed arbitrageurs. Measured over the entire pool (all LPs combined) on a 90-day window, LPs collected ≈ \$374k in fees while losing ≈ \$662k to adverse selection, measured swap-by-swap by markout; per dollar of volume, the LP fee of 0.67 bp is dominated by a realized toxicity of ≈ 1.2 bp, so that roughly half a basis point of every traded dollar flows from LPs to arbitrageurs. Over two years and 77.8 million swaps the picture is identical (ETH-USDC: \$3.30M of LP fees against \$6.41M of toxicity; cbBTC-USDC: \$0.93M against \$1.74M), and it holds in every one of 14 pre-registered market regimes — bull, bear and sideways alike. Concentration does not help; it hurts, because toxic flow strikes precisely at the active tick where concentrated positions hold their capital. The only genuinely positive cash-flow available to an LP on these pools is protocol emissions (CAKE, AERO), which transforms the range-selection problem from yield maximization into *loss minimization while farming emissions* — and under that objective the optimal geometry is the least sophisticated one we tested: a single, wide (±10–15%), *lazy* range that waits when the price leaves and shifts minimally instead of recentering. Finally, the emissions budget itself caps the opportunity: all our PancakeSwap farms together emit ≈ \$1M per year for all participants combined, and at a simulated position size of \$50k every strategy we measured — including the best one — loses to a buy-and-hold benchmark once emissions are included. The measured capacity of a scalable, net-positive-versus-holding product on these pools is zero. Our detailed measurements are all from Base, where our event data is complete; BTR's internal R&D on other chains and venues, conducted before and alongside this campaign, reached the same qualitative conclusion — passive concentrated liquidity loses to informed flow — so we do not regard the result as an artifact of one chain, though we quantify it only where we can measure it.

We believe these results are worth publishing precisely because the opposite claim is the premise of a large commercial ecosystem. Our contributions are:

1. **A negative-sum measurement.** Using complete on-chain event histories, we measure realized flow toxicity (markout) against LP fee income at the level of the whole pool, per market regime, and per position geometry. LP fee income is strictly dominated by adverse selection in 14/14 regimes over two years (Section 5). The result is model-free: it uses only recorded swaps and subsequent pool prices.
2. **An evaluator hierarchy and a calibration standard.** We document a three-tier measurement stack (calibrated Monte-Carlo simulation → historical event replay → live forward arms) together with the calibration criterion that made us trust it: the replay engine must reproduce the fees *actually collected* by live production positions (0.97–1.00 achieved). We show a concrete failure mode of the simulation tier — fee credit is assigned by range width while per-tick adverse selection is invisible — that led it to recommend a geometry the real data inverts (Section 4).
3. **A geometry reversal.** On replayed real events, a single wide lazy range dominates nested, center-weighted allocations in 14/14 regimes (pooled two-year advantage of +106 APR points on ETH-USDC and +68 on cbBTC-USDC against the previously locked nested configuration), with a monotone ordering — the less capital at the center, the better —

that is the exact inverse of the simulation-tier optimum. The live paper arms subsequently reproduced this hierarchy forward (Section 7).

4. **A capacity ceiling.** We combine the negative-sum fee result with the measured emissions budget to bound the size of a viable product: emissions of $\approx$ \$1.02M/yr across all our PancakeSwap farms, dilution-adjusted net returns at \$50k of -12 to -22%/yr on ETH-USDC (emissions included), and a scalable net-positive capacity of \$0 (Section 8).

Provenance labels. Because this paper mixes on-chain accounting, historical replay and forward sandbox results, every quantitative claim is tagged with one of three provenance labels, and we never blur them: MEASURED — computed from on-chain events or from the production database of the live bot (real dollars, real fills); SIMULATED — produced by an evaluator replaying history or sampling paths, typically at a virtual position size of \$50k; LIVE — produced by the forward sandbox or its paper arms running in real time at deliberately small, sandbox-scale position sizes. The distinction matters: for instance, position sizes that are viable LIVE at sandbox scale (a few hundred dollars) are demonstrably not viable SIMULATED at \$50k, and this size dependence is itself one of the paper's results.

Outline. Section 2 fixes notation and recalls the Uniswap v3 mechanics we rely on, including the divergence-loss closed form of Echenim et al. [1], the loss-versus-rebalancing (LVR) framework of Milionis et al. [2], and our markout definition of realized toxicity. Section 3 describes the deployed system, the venues and the datasets. Section 4 presents the three evaluators and their calibration. Section 5 establishes the negative-sum result and its regime robustness; Section 6 measures the winning side of the trade, the arbitrage ecosystem. Section 7 documents the geometry reversal and the loss-minimizing lazy range; Section 8 measures the emissions budget and derives the capacity ceiling. Section 9 catalogs every lever tested, Section 10 distills the methodological lessons into an artifact taxonomy, and Section 11 concludes; the team's ongoing pivot toward BTR Prime, a proprietary AMM design that internalizes the arbitrage flow measured here, is sketched there as future work.

2 Uniswap v3 mechanics and notation

This section collects the definitions used in the rest of the paper. It is deliberately minimal: for a complete mathematical treatment of Uniswap v3 we refer to Echenim, Gobet and Maurice [1], whose notation we follow (writing $\pi = \sqrt{P}$ for the square-root price), and to the protocol whitepaper [3]. Venue-specific mechanics (PancakeSwap v3 farms, Aerodrome Slipstream gauges) are described in the respective protocol documentations [5, 4].

2.1 Pools, price, and the tick grid

A pool trades a *base* (risky) token X against a *quote* token Y, at spot price P_t expressed in units of Y per unit of X.[1] Throughout,

$$\pi_t \;\coloneqq\; \sqrt{P_t} \tag{1}$$

denotes the square-root price, the variable in which all v3 formulas are affine. Prices live on a geometric grid of *ticks*: tick $i \in \mathbb{Z}$ corresponds to the price

$$p(i) \;=\; 1.0001^{\,i}, \tag{2}$$

i.e. adjacent ticks are one basis point apart. Each pool constrains range endpoints to multiples of a *tick spacing* δ ($\delta = 1$ for the 0.01% pools traded by our bot; up to $\delta = 200$ on other tiers). A pool charges a fee rate $\gamma \in (0, 1)$ on every swap input (e.g. $\gamma = 10^{-4}$ for a 0.01% pool); a

[1] For volatile/volatile pairs such as WETH-cbBTC the quote token is itself risky; dollar values are then obtained through the quote token's own dollar price. Our production pairs are ETH-USDC, EURC-USDC, SOL-USDC, cbBTC-USDC and WETH-cbBTC on Base.

fraction $\phi \in [0, 1)$ of collected fees may be diverted to the protocol rather than to LPs (on our PancakeSwap pools we measure $\phi = 0.33$; see Section 4).

2.2 Concentrated liquidity and position value

Definition 2.1 (Position)**.** *A position is a triple (L, p_a, p_b) with liquidity $L > 0$ deployed on the price range $[p_a, p_b]$, $0 < p_a < p_b$. Writing $\pi_a = \sqrt{p_a}$, $\pi_b = \sqrt{p_b}$, the token amounts held by the position at spot price P (with $\pi = \sqrt{P}$) are*

$$x(P) \;=\; L\left(\frac{1}{\pi^R} - \frac{1}{\pi_b}\right), \qquad y(P) \;=\; L\left(\pi^R - \pi_a\right), \tag{3}$$

where $\pi^R \coloneqq \min\big(\max(\pi, \pi_a),\, \pi_b\big)$ is the projection of π onto $[\pi_a, \pi_b]$.

The projection encodes the three regimes in one formula: below the range the position holds only X; above it, only Y; in between, a mix that continuously sells X as the price rises, swaps following the constant-product rule on *virtual* reserves of depth L.

Proposition 2.2 (Position value)**.** *The mark-to-market value of the position in quote units, $V(P) \coloneqq P\,x(P) + y(P)$, is*

$$V(P) \;=\; \begin{cases} L\,\pi^2\left(\dfrac{1}{\pi_a} - \dfrac{1}{\pi_b}\right), & P \le p_a, \\[2ex] L\left(2\pi - \pi_a - \dfrac{\pi^2}{\pi_b}\right), & p_a \le P \le p_b, \\[2ex] L\left(\pi_b - \pi_a\right), & P \ge p_b. \end{cases} \tag{4}$$

V is continuously differentiable, concave, and strictly concave on (p_a, p_b) with $V''(P) = -L/(2P^{3/2})$ there.

Concavity is the structural fact behind everything that follows: a concave payoff is short convexity, so the position systematically loses value under price movement relative to holding its instantaneous inventory — the loss fees are supposed to compensate. The *concentration factor* — fee income per dollar relative to a full-range position — grows like $(1 - \pi_a/\pi_b)^{-1}$ as the range narrows, which is the design's appeal; Section 5 shows that on our pools the toxicity borne per dollar grows *faster*.

2.3 Fee accounting

Uniswap v3 credits fees through global accumulators: each swap increases a per-unit-of-liquidity fee-growth accumulator on every tick interval it crosses, and a position's claimable fees are

$$\text{fees}(t_0, t_1) \;=\; L \cdot \big(f_r(t_1) - f_r(t_0)\big), \tag{5}$$

where f_r denotes the pool's `feeGrowthInside` accumulator restricted to the position's range $[p_a, p_b]$ (one accumulator per token; we aggregate in quote units). Two consequences matter for this paper. First, fee income is proportional to the position's share of liquidity *at the executed ticks*: fees are earned only while the spot price is inside the range, and are diluted one-for-one by competing in-range liquidity, including the position's own contribution — a self-dilution that becomes material at the \$50k sizes we simulate. Second, equation (5) is directly observable on-chain, which is what allows the replay engine of Section 4 to be validated against the fees actually collected by our production positions, and what the passive fee-capture *probes* of Section 3.3 exploit.

In addition to swap fees, both venues pay *emissions* in their native token to attract liquidity: CAKE on PancakeSwap v3 (positions staked in the `MasterChef` farm contract) and AERO on Aerodrome Slipstream (positions staked in per-pool *gauges*). We write $E(t_0, t_1)$ for the emissions

accrued by a position over a window, valued in quote units at collection time. On Slipstream the two income streams are mutually exclusive: a gauge-staked position earns emissions only, its swap fees being redirected to protocol voters [4] — an accounting subtlety that invalidates any naive "fees plus emissions" sum on that venue.

2.4 Divergence loss and LVR

Two standard notions of LP loss are used in this paper, against two different benchmarks.

Definition 2.3 (Divergence loss)**.** *Fix a position (L, p_a, p_b) minted at time t_0 and held to t_1, with no intervening liquidity change. Let $V_{\mathrm{LP}}(t_1)$ be the value (4) of the position at P_{t_1}, and let $V_{\mathrm{H}}(t_1)$ be the value at P_{t_1} of the initial inventory $\big(x(P_{t_0}), y(P_{t_0})\big)$ held unchanged (buy-and-hold). The divergence loss (or impermanent loss) is*

$$\mathrm{DL}(t_0, t_1) \;\coloneqq\; V_{\mathrm{LP}}(t_1) - V_{\mathrm{H}}(t_1) \;\le\; 0. \tag{6}$$

Proposition 2.4 (Closed form; Echenim et al. [1])**.** *With $\pi_i = \sqrt{P_{t_i}}$ and π_i^R the projection of π_i onto $[\pi_a, \pi_b]$ for $i \in \{0, 1\}$,*

$$\mathrm{DL}(t_0, t_1) \;=\; -\,L\left(\pi_0^R - \pi_1^R\right)\left(1 - \frac{\pi_1^2}{\pi_0^R\,\pi_1^R}\right). \tag{7}$$

The expression is path-independent: it depends only on the endpoint prices, whatever the intermediate trajectory, and it covers the in-range and out-of-range cases uniformly through the projections. When both endpoints are in range it reduces to $\mathrm{DL} = -L\,(\pi_0 - \pi_1)^2/\pi_0$.

Equation (7) is implemented verbatim in our production accounting: at every rebalance the realized DL of the closed position is computed from its mint and close prices, which is what lets us decompose live performance into fees, emissions, divergence loss and costs (Section 7.4).

Divergence loss benchmarks against buy-and-hold. The *loss-versus-rebalancing* (LVR) of Milionis et al. [2] benchmarks instead against a continuously rebalanced replicating portfolio, isolating the component of LP loss that is pure adverse selection: the profit of an arbitrageur who trades against the pool only when the pool price is stale relative to an external market.

Definition 2.5 (LVR)**.** *Let the external price follow a diffusion with instantaneous volatility σ_t. The LVR of a pool position with value function V over $[t_0, t_1]$ is*

$$\mathrm{LVR}(t_0, t_1) \;=\; \int_{t_0}^{t_1} \frac{\sigma_t^2 P_t^2}{2}\,\big|V''(P_t)\big|\,\mathrm{d}t \;\;\ge\; 0. \tag{8}$$

For the v3 position of Proposition 2.2 this rate equals $\frac{1}{4}\,\sigma_t^2\,L\,\pi_t$ while $P_t \in (p_a, p_b)$, and vanishes out of range.

LVR is a model quantity: it prices adverse selection through an assumed volatility. It is the right object for simulation (our Monte-Carlo tier charges an LVR-type cost along sampled paths) but it cannot, by itself, tell us what arbitrageurs actually extracted from a specific pool. For that we use a realized, model-free counterpart.

2.5 Markout: realized flow toxicity

Definition 2.6 (Markout)**.** *Consider swap k executed at time t_k, in which the pool's LPs (in aggregate) receive the signed inventory change $(\Delta x_k, \Delta y_k)$ — what the trader pays in, minus what the trader takes out — excluding the fee. For a horizon $\tau > 0$, the markout of the swap is its inventory change repriced at the pool mid-price τ later:*

$$m_\tau(k) \;\coloneqq\; \Delta x_k\,P_{t_k+\tau} + \Delta y_k. \tag{9}$$

We use $\tau \in \{300\,\text{s}, 1800\,\text{s}\}$ *and report both. The net flow PnL of the pool's LPs over a window* W *is*

$$\Pi_\tau(W) \;=\; \sum_{k \in W} \Big(\text{fee}_k + m_\tau(k)\Big), \tag{10}$$

where $\text{fee}_k = \gamma(1-\phi)\cdot$ (*swap notional*) *is the LP share of the fee. We call the flow toxic when* $\sum_k m_\tau(k) < 0$*, and the pool negative-sum for LPs when* $\Pi_\tau(W) < 0$*: fee income does not cover what the flow takes.*

Markout is the standard toxicity measure of market-making practice: an uninformed trader's inventory change has zero expected markout, whereas an arbitrageur trading on stale quotes leaves the LP with inventory systematically worth less moments later. It requires no volatility model and no assumption about trader identity — only the recorded swaps and the pool's own subsequent mid-price. When markout is attributed to a *specific* position rather than to the whole pool, fees and markout are allocated pro rata to the share of liquidity the position holds at the executed ticks — the allocation rule the chain itself applies to fees via (5).

Remark 2.7 (Relation between the three loss measures)**.** *DL is an accounting identity against buy-and-hold, exact per position* (7)*, recorded by our live ledger. LVR is the model-implied adverse-selection cost, used inside the Monte-Carlo tier. Markout is the realized adverse-selection cost, computable only from full event history, and is the quantity in which the central negative-sum result is stated. The three are consistent on our data — rankings agree, levels differ by* 5–20% *— but they are not interchangeable, and every number is labeled with the measure that produced it.*

2.6 Rebalancing, benchmarks, and notation summary

An ALM policy closes a position and opens a new one — a *rebalance* — incurring gas, swap fees and slippage on the inventory swap, and crystallizing the DL of the closed position. Unless stated otherwise, strategy performance is reported as the growth of position net asset value (NAV, fees and emissions included, all costs deducted) *relative to the same wallet's buy-and-hold value*, in percent (or annualized, in APR points); "vs. B&H" columns in later tables are exactly this quantity. Table 1 summarizes the notation.

Table 1: Notation used throughout the paper.

Symbol	Meaning
X, Y	base (risky) and quote token of a pool
P_t, $\pi_t = \sqrt{P_t}$	spot price (Y per X) and square-root price
$p(i) = 1.0001^i$, δ	tick grid and pool tick spacing
γ, ϕ	pool fee rate; protocol share of fees
(L, p_a, p_b)	position: liquidity and range bounds
π_a, π_b; π^R	sqrt-range bounds; projection of π onto $[\pi_a, \pi_b]$
$x(P), y(P)$; $V(P)$	position amounts (3); position value (4)
f_r	in-range fee-growth accumulator (`feeGrowthInside`)
$E(t_0, t_1)$	emissions (CAKE/AERO) collected over a window
DL	divergence loss vs. buy-and-hold (7)
LVR	loss-versus-rebalancing (Definition 2.5)
m_τ, Π_τ	markout at horizon τ; net LP flow PnL (10)
η, λ_τ	LP fee yield and toxicity per dollar of volume (bp)
MEASURED/SIMULATED/LIVE	provenance labels (Section 1)

3 Experimental Infrastructure

Our measurements come from three physically distinct layers, and we maintain a strict terminological separation between them throughout the paper:

- **Measured** — quantities read directly from on-chain events or from the production accounting database (collected fees, confirmed transactions, reward accruals, realized slippage);
- **Simulated** — quantities produced by replaying historical data against a virtual position, typically of $50,000 notional (Section 4.2);
- **Live** — quantities produced by the sandbox-scale production system and its paper-trading arms running forward in real time (Sections 3.1–3.2).

Every figure reported in Sections 5–8 is tagged with its layer. The distinction matters because the three layers have different failure modes: measured data is sparse but incontrovertible, simulation covers years but is a single draw of history, and live capital is the only forward-looking instrument but accumulates evidence slowly.

3.1 The production system

The automated liquidity manager (ALM) operates five concentrated-liquidity positions on Base: ETH–USDC, cbBTC–USDC and WETH–cbBTC on PancakeSwap V3 pools [5], and SOL–USDC and EURC–USDC on Aerodrome Slipstream pools [4] (unstaked). This venue assignment is the configuration that held for the bulk of the April–July 2026 campaign, and the emission rates quoted in Section 8 were measured under it. Each pair is deliberately capitalized at sandbox scale (a few hundred dollars). This *sandbox sizing* was a design decision, not a constraint: execution defects on a money path are discovered by losing real—but small—dollars, while all size-dependent conclusions are delegated to the simulated $50,000 layer, where the position's own dilution of pool liquidity is modeled explicitly (Section 4.2).

The control loop condenses two years of design iteration and is worth describing. Every 15 minutes the system snapshots each pool and position into an append-only database and evaluates a signal pipeline built on three market *forces*, each estimated on several timeframes (15-minute, hourly, four-hourly) and blended into a weighted composite: a volatility force based on the Parkinson high–low range estimator, a momentum force based on RSI, and a trend force based on moving-average geometry. The composite proposes a candidate range — width driven by the volatility force, asymmetry by the trend bias — and a Nelder–Mead optimizer then refines the range parameters online against a fitness functional that combines expected fee income with an LVR-type adverse-selection penalty in the spirit of [2], the exact impermanent-loss expression of [1], a first-exit-time probability, and a Hausdorff-distance penalty against the incumbent range (pricing the cost of moving). Fee income inside the fitness is not assumed but supplied by the *empirical fee program* f_0: the realized fee APR as a function of range width, measured continuously per pool by the on-chain probes of Section 3.3. A proposed rebalance executes only if it clears a battery of gates — projected gas cost against expected benefit, minimum range shift, signal confidence and breadth, and a cooldown since the last action — whose combined effect is that most cycles end in *hold*. Execution is a withdraw–swap–mint–stake sequence with aggregator routing (LI.FI primary, KyberSwap fallback) and explicit on-chain confirmation of every transaction; a submitted hash is never trusted as a completed action. Rewards (CAKE) are harvested, valued and sold automatically above a $3 threshold; external flows are journaled separately so that performance attribution is never contaminated. Benchmarks are computed on the same 15-minute grid: *HODL* (the entry inventory held passively) and *HODL+Yield* (HODL plus the realized reward stream).

3.2 The paper-trading fleet

A separate process runs *shadow arms*: virtual portfolios that consume the same 15-minute price and pool feed as production, apply a candidate strategy's rebalancing logic, and maintain the same NAV-versus-HODL accounting, including modeled swap costs and gas and the same reward-rate overlay per pair. Because every arm sees identical prices at identical instants, arm-versus-arm comparisons are matched by construction; arm-versus-production comparisons are clipped to a common start (*same-start* accounting) before any delta is reported.

The fleet grew with the research (Table 2). Four *channels* carried variants of the production mathematics: `v1` (a patch channel re-exporting the production math verbatim), `v2` (the quant channel, carrying stable-pair overrides and later serving as the frozen concentrated reference), `v3` (an experimental regime channel fed by a Renko bar feed from NX Rates (partner infrastructure) and a boosted-tree classifier, eventually decommissioned), and `v3fit` (a fitness-realism experiment aligning the optimizer's internal rebalance model with the live gates). Three *arms* isolated single mechanisms against a control: `armA` (a width damper flooring the volatility force off its steep low-volatility knee), `armB` (a volatility-scaled rebalance threshold for calm pairs, stable pairs excluded), and `armC` (an unmodified control). As the evaluator verdicts of Sections 4 and later arrived, a *constant-mix family* was deployed (`cmix`, equal-weight nested ranges; `cmix_ih`, the simulator's inner-heavy weighting; `cmix_dln`, derived log-normal leg weights; `cmix_1range`, a single range), followed by the *lazy family* implementing the replay evaluator's recommendation (`cmix_lazyshift`, minimal shift after a patient delay; `cmix_w15`, a single wide ±15% range). Ten arms ran concurrently by the end of the campaign. The fleet is the delivery vehicle of the forward A/B test that serves as final judge in the evaluator hierarchy (Section 4.3).

Table 2: The paper-trading fleet and the passive on-chain probes deployed over the measurement campaign.

Instrument	Role
Channels (variants of the production mathematics)	
`v1`	patch channel; re-exports the production math verbatim
`v2`	quant channel (stable-pair overrides); frozen concentrated reference
`v3`	Renko-bar feed + boosted-tree regime classifier; decommissioned
`v3fit`	fitness realism: optimizer rebalance model aligned to live gates
Mechanism arms (single-lever A/B against a control)	
`armA`	width damper: volatility force floored off its low-vol knee
`armB`	vol-scaled rebalance threshold (calm pairs; stables excluded)
`armC`	unmodified control
Constant-mix family (post-tournament)	
`cmix` / `cmix_ih` / `cmix_dln`	nested ranges: equal-weight / inner-heavy / log-normal leg weights
`cmix_1range`	single-range constant mix
Lazy family (post-replay reversal)	
`cmix_lazyshift`	minimal shift after a patient delay
`cmix_w15`	single wide ±15% lazy range
Passive on-chain probes (Section 3.3)	
f_0 width probe	fee capture at five widths simultaneously; the empirical fee program
passive arm	fixed ±10% range, never rebalanced (prototype of the lazy optimum)
out-of-range arm	recenters only on range exit; the passive/active contrast
σ overlay	static width vs. vol-timed width, same pool, identical instants
new-pools probe	fee-capture screening of newly listed pools before any capital

3.3 On-chain fee-capture probes

Between the sandbox (real but tiny) and the simulator (large but virtual) sits a third instrument: *capture probes*. A probe is a fictitious position of $50,000 notional whose bounds are registered off-chain; at each 15-minute cycle the probe reads the pool's `feeGrowthInside` accumulators on-chain for its tick range and converts the increment into the fees the position *would have* collected.

This is ground truth for fee capture — no fee model is involved, only the pool's own accounting — at zero capital cost. Five probe programs ran over the campaign (Table 2): the f_0 width probe, reading fee capture at five widths simultaneously (from a few tenths of a percent to ±10–15%) and producing the empirical fee density consumed by the optimizer and the Monte-Carlo tier; a passive ±10% arm that never rebalances — in hindsight, the prototype of the lazy optimum of Section 7; an out-of-range arm that recenters only when the price exits, instrumenting the passive-versus-active contrast on-chain; a σ-overlay probe holding a static-width arm and a volatility-timed arm on the same pool at identical instants, which is what refuted volatility timing at zero capital cost; and a new-pools probe applying the same fee-capture reading to newly listed pools before any capital decision. For venue comparisons, probes were also run on up to four pools of the same pair at identical instants, which removes regime effects from cross-pool comparisons. As an example of resolution, the EURC–USDC probe (661 samples, 12.87 days, 99–100% in-range) measured a fee-capture gradient from 8.29% APR at the tightest width to 0.78% at the widest. The probes' blind spot is their window (days to weeks); they are used as calibration targets and venue arbiters, never as strategy verdicts on their own.

4 The Evaluator Hierarchy

The central methodological lesson of this project is that *the choice of evaluator is the result.* Strategy questions on concentrated liquidity were answered differently — sometimes with opposite sign — by a calibrated Monte-Carlo engine and by an exact replay of on-chain events, and only the latter agreed with live capital. We therefore present the three evaluators in the order in which trust was ultimately assigned, with each one's calibration evidence and known blind spot.

4.1 Calibrated Monte-Carlo engine (secondary)

The first evaluator is a Monte-Carlo path engine over centrally-traded prices. Price paths are generated by *block bootstrap* of real Binance hourly returns (43,793 hourly bars per symbol, roughly five years), which preserves heavy tails and volatility clustering that a parametric GBM misses; paths are drawn within four disjoint historical regimes (bull 2021, bear 2022, chop 2023, bull 2024+) so that verdicts can be reported per regime as well as pooled. On each path, a candidate range strategy is evolved and scored as fees minus impermanent loss minus costs, with an optional reward (CAKE) overlay. Fee income is modeled through an *empirical fee density* $f_0(w)$ — the fee APR as a function of range width — measured by the live probes of Section 3.3 rather than assumed. A typical verdict consumes roughly 9×10^5 path evaluations.

The engine was accepted only after passing three calibration gates (Table 3): bootstrap volatility must reproduce realized Binance volatility; the internal impermanent-loss computation must agree exactly with an independent empirical evaluator; and modeled IL must match the IL realized by the production positions, which it does on all four pairs with a ratio of 0.63–0.85, i.e. the model is conservative (reality is slightly worse for the LP).

Table 3: Calibration gates of the Monte-Carlo engine.

Gate	Criterion	Result
Volatility	bootstrap vol vs. realized Binance vol	4/4 pairs, $< 1\%$ rel. error
IL consistency	engine IL vs. independent empirical IL	exact (zero difference)
IL realism	modeled IL vs. production-realized IL	4/4 pairs, ratio 0.63–0.85

Why it failed for geometry. The engine credits fee income to a position through the width-level density $f_0(w)$, but it does not *debit* adverse selection at the tick level. As Section 5 quantifies, toxic order flow — arbitrage against stale pool prices, the flow underlying loss-versus-rebalancing [2] — strikes precisely where the fees are earned, at the ticks adjacent to the current price,

and its marginal cost grows *faster* there than the fee density. An evaluator blind to per-tick toxicity therefore systematically overvalues capital concentrated near the price. Concretely, the engine recommended an inner-heavy weighting of a nested four-range structure; the event-replay evaluator reversed that recommendation in 14 out of 14 dated market regimes, and the live A/B subsequently sided with the replay. The engine was demoted to *secondary* status: it remains useful for aggregate fee/IL trade-offs and for synthetic multi-regime stress, and is never again used alone for a placement or geometry question.

4.2 Tick-level event replay (primary)

The second evaluator removes the price model entirely: it reconstructs the pool itself. Using HyperSync as the extraction layer, we fetched *every* `Swap`, `Mint` and `Burn` event of the studied pools into a local database — a 90-day pilot (3 April–2 July 2026, 90.03 days) comprising 10.7M events (6,227,440 swaps on ETH–USDC, 2,503,896 on cbBTC–USDC, zero empty days), then a two-year extension totalling 77.8M swaps — together with the full mint/burn history since pool creation, so that the liquidity resident in every tick is known exactly at every block. Each historical swap is then re-marched through the tick map against the *actual competing liquidity of that moment*, and a virtual position is inserted into the reconstruction.

Definition 4.1 (Pro-rata fee attribution per tick crossed)**.** *Consider a swap that consumes input across tick segments $s = 1, \dots, m$, with x_s the input amount consumed in segment s, L_s the aggregate active liquidity in s recorded on-chain, γ the pool fee rate, and ϕ the protocol fee share. A virtual position holding liquidity L_v active in segment s collects*

$$\mathrm{fee}_v(s) \;=\; (1-\phi)\,\gamma\,x_s\,\frac{L_v}{L_s + L_v}, \tag{11}$$

i.e. fees are split per tick crossed, pro rata of liquidity, and the virtual position dilutes the pool it is measuring: its own liquidity appears in the denominator.

Equation (11) is what makes the \$50,000 simulations honest at size — at that notional the virtual position represents 25–43% of in-range liquidity on our pools, and self-dilution is a first-order effect (Section 8.3).

Adverse selection is measured on the reconstruction by the markout of Definition 2.6: each swap's LP inventory change (fee excluded) is revalued at the pool price $\tau \in \{300\,\mathrm{s}, 1800\,\mathrm{s}\}$ later. Aggregated over swaps, negative markout is the realized counterpart of loss-versus-rebalancing [2]: the cost of trading against better-informed flow.

Strategy backtests on the reconstruction run at the same 15-minute decision cadence as production, with costs of \$0.05 gas per rebalance plus \$0.05 per range leg and 2 bps on re-swapped notional, in-sample/out-of-sample splits, and — for the two-year runs — 14 pre-registered, dated market regimes. Geometry comparisons are *matched-control*: competing structures are forced to identical recentering instants and identical swap costs, so that width and trading-frequency confounds are excluded by construction.

Calibration. Table 4 lists the validation checks, all re-derived independently from the raw artifacts. The decisive one is the last: the fees *actually collected on-chain* by the production positions over the pilot window (29 matched positions on ETH–USDC, 49 on cbBTC–USDC) match the reconstruction's prediction to 0.990 and 0.969 respectively. Together with the exact liquidity walk and the external volume cross-check, this establishes the replay as a measurement instrument rather than a model.

Blind spot and status. The replay is the past: one sample of history, with no counterfactual ("what if the market had done otherwise"). Within that limitation it is exact, and it is the *primary* evaluator for every placement and geometry question in this paper. All headline results of Sections 5–7 — the negative-sum measurement, the 14/14-regime reversal of the nested/inner-heavy geometry, and the lazy wide-range optimum — are its output.

Table 4: Validation of the event-replay evaluator. Pilot = 90.03 days (3 Apr–2 Jul 2026); 2y = two-year extension (77.8M swaps).

Check	Result
Liquidity walk vs. event-reported liquidity	exact on 6.2M swaps (pilot; max err. 10^{-4}); 100.0000% on 61M swaps (2y, ETH–USDC)
Reconstructed per-tick input vs. actual swap input	ratio 0.9999991–1.0000004
Protocol fee share (measured from events)	0.330 (venue documentation: 32–33%)
90-day volume vs. independent aggregator	ratio 1.003 (both pools); median daily ratio 1.0015
Pilot ↔ 2y splice consistency	1.00000
Live f_0 probe cross-check (common 7.6 h window, 21 samples)	$< 1\%$ at wide widths; worst 15.4% at the finest widths (recentering granularity)
Production-collected fees vs. reconstruction (90 d)	**0.990** (ETH–USDC, 29 positions) / **0.969** (cbBTC–USDC, 49 positions)

4.3 Live forward A/B (final judge)

The third evaluator is the slowest and the only one that cannot overfit history: the shadow arms of Section 3.2 running forward against production and buy-and-hold under identical accounting. Its role is adjudication. When the replay contradicted the Monte-Carlo engine on geometry, arms implementing both sides (`cmix_ih` for the engine's inner-heavy recommendation; `cmix_1range`, `cmix_lazyshift` and `cmix_w15` for the replay's single-wide-lazy recommendation) were deployed and left to accumulate a common window. At the end of the campaign (windows of 13–17 days), the live ranking reproduced the two-year backtest ranking — wide lazy arms at the top, concentrated variants at the bottom (Section 7.4). A 13–17 day window in a single regime does not prove a law; it does discriminate between two evaluators that disagree, which is all that was asked of it.

4.4 The resulting epistemic protocol

The campaign's history of reversals (fee-density optimum → early arms → inner-heavy constant-mix → single wide lazy range) was not random walk: each stage used the best evaluator then available, and each reversal was caused by a specific, subsequently identified blind spot. The protocol we converged to, and recommend for any empirical DeFi strategy work, is:

1. No conclusion is *locked* on the word of a single evaluator. Locking requires two independent evaluators in agreement *plus* forward confirmation by live capital.
2. Every verdict must cite its evaluator *and* that evaluator's known blind spot: the Monte-Carlo engine is blind to per-tick toxicity; the event replay is a single draw of the past; the probes are short-windowed; the live A/B is slow.
3. Any evaluator making size-dependent claims must model the position's own dilution (Definition 4.1); any evaluator making placement claims must price adverse selection at tick resolution (Definition 2.6).

Sections 5–8 apply this protocol to the economics of the pools themselves.

5 Main result: concentrated liquidity provision is a negative-sum game on the pools studied

This section presents the central empirical finding of the project. On the Base-chain pools we operated (PancakeSwap V3, 1 bp fee tier), fee income earned by the aggregate LP population is systematically smaller than the adverse-selection cost imposed by informed order flow — in every dated market regime of a two-year window, at every horizon we measured, and for every

range geometry we replayed. Fee-only liquidity provision on these venues is therefore a structurally losing trade: range design can only minimize the loss, never reverse its sign. All headline numbers in this section are MEASURED quantities derived from raw on-chain events (Section 5.1), except where explicitly tagged SIMULATED (the $50k virtual-LP replays of Section 5.5).

5.1 Markout: a realized, model-free measure of adverse selection

Rather than the model-implied LVR of [2], we measure its *realized* counterpart: the markout of Definition 2.6, evaluated on the pool's own event stream (the horizon price $P_{t+\tau}$ is the price of the last pool swap with timestamp $\leq t+\tau$, derived from `sqrtPriceX96`). When reported *trader-side* the sign is flipped: the trader-side markout is the value of what the trader took out of the pool, revalued τ seconds later, net of the pool fee, and the LP-side *toxicity* of the swap is its negative.

We evaluate m_τ at $\tau = 300\,\text{s}$ and $\tau = 1800\,\text{s}$ for every swap in the window; results are insensitive to the choice (Table 6). Because the revaluation uses the *pool's own* subsequent price rather than an external oracle, the measure requires no model of the efficient price and no off-chain data. Two caveats are recorded for honesty. First, trader-side markout measures the trader's gain *against the pool*; any CEX hedging leg is invisible, so aggregate positive markout is an *upper bound* on the arbitrageurs' realized profit (this matters in Section 6). Second, markout at short horizons and realized impermanent loss per rebalancing cycle are distinct estimators of the same underlying cost; in our data they agree in ordering across all strategies and in magnitude to within 5–20%, which is the concordance one expects from two estimators with different windows.

The measurement infrastructure is the event-replay engine of Section 4.2, whose validation — exact liquidity walk, measured protocol fee share, external volume cross-check, and the binding reproduction of production-collected fees at 0.97–0.99 — is summarized in Table 4.

5.2 Pool-level accounting: the 90-day pilot

We first ran the accounting on a 90.0-day pilot window (2026-04-03 to 2026-07-02; 8.7M swaps across the two production pools). Define the pool-level fee yield and toxicity per dollar of volume:

$$\eta \;=\; \frac{\text{LP fees}}{\text{volume}}, \qquad \lambda_\tau \;=\; \frac{-\sum_{\text{swaps}} m_\tau}{\text{volume}}.$$

On a 1 bp fee tier with a measured 33% protocol share, $\eta = 0.67\,\text{bp}$ by construction (the symbol ϕ being reserved for the protocol share); the empirical question is λ.

Table 5: Pool-wide LP accounting, 90-day pilot window (MEASURED, all LPs aggregated, PancakeSwap V3 on Base, 1 bp fee tier). Markout at $\tau = 300\,\text{s}$.

Pool	Swaps	LP fees (90d)	Toxicity (90d)	Net
ETH-USDC	6,227,440	$247k	$428k	−$181k
CBBTC-USDC	2,503,896	$127k	$234k	−$107k
Both pools	8,731,336	$374k	$662k	−$288k

Per dollar traded, the LP population collects $\eta = 0.67\,\text{bp}$ and pays $\lambda_{300} = 1.16\,\text{bp}$ (ETH-USDC) to $1.25\,\text{bp}$ (CBBTC-USDC): roughly $0.5\,\text{bp}$ of every dollar of volume is transferred from LPs to informed takers. The toxicity is not driven by large trades: the most toxic notional bucket is $1k–$10k (-1.24 to $-1.37\,\text{bp}$), and the worst 1% of swaps account for 26–32% of total negative markout — consistent with high-frequency arbitrage rather than institutional block flow.

5.3 The two-year replay

A 90-day window is a single market regime. We therefore extended the fetch to the full available history and replayed 77.8M swaps: 730 days for ETH-USDC and 654 days for CBBTC-USDC (the latter pool's entire life since its creation in September 2024).

Table 6: Pool-wide LP accounting over the full replay window (MEASURED). "Gross + markout" adds back the 33% protocol share, i.e. it is the counterfactual in which LPs keep 100% of swap fees.

	ETH-USDC (730 d)	CBBTC-USDC (654 d)
Volume	\$49.29B	\$13.86B
Swaps	60,975,796	16,867,874
LP fees	\$3.30M	\$0.93M
Markout ($\tau = 300$ s)	−\$6.41M	−\$1.74M
Markout ($\tau = 1800$ s)	−\$6.67M	−\$1.74M
LP fees + markout$_{300}$	−\$3.11M	−\$0.81M
Gross fees + markout$_{300}$	−\$1.48M	−\$0.35M
η (bp of volume)	0.67	0.67
λ_{300} (bp of volume)	1.30	1.25

Proposition 5.1 (Negative-sum result)**.** *Over the full replay window, the aggregate LP population of both pools paid more in adverse selection than it collected in fees: $\eta < \lambda_\tau$ for both pools and both horizons. The result is robust to the fee split: even crediting LPs with 100% of gross fees (protocol share included), the net remains negative (Table 6, last accounting row). Fee-only liquidity provision on these pools is thus negative-sum against the arbitrage sector regardless of any conceivable fee-rebate policy.*

This is not a statement about our bot or our parameters — it is a property of the venue: *every* dollar of passive liquidity in these pools, however deployed, participated in a losing aggregate trade.

5.4 Regime robustness: a pre-registered segmentation protocol

A pooled two-year total could in principle hide profitable sub-periods. To rule this out without inviting the opposite bias — gerrymandering regime boundaries until the desired result appears — we fixed the segmentation *before* running any strategy backtest, with a mechanical rule and a timestamped artifact:

1. Boundaries are produced by a **zigzag algorithm with a 22% reversal threshold** applied to the daily first-swap pool price: boundaries are local price extrema of the actual traded series, not hand-picked dates.
2. Each leg is labeled mechanically: $|\text{return}| \geq 30\% \Rightarrow$ *bull* or *bear* by sign, otherwise *chop*.
3. The boundary file (`regimes_2y.json`) was frozen on 2026-07-03, *prior* to the strategy replays; the segmentation is an exact partition of the window (no gaps, no overlaps).

The rule yields 8 dated regimes for ETH-USDC (returns ranging from −63% to +228%) and 6 for CBBTC-USDC (from −50% to +84%): 14 regimes in total covering bull, bear and chop conditions in both directions.

The negative-sum inequality $\eta < \lambda_{300}$ holds in 14 of 14 dated regimes, with no exception — in the +228% ETH bull leg, in the −63% bear leg, and in every chop segment, at both markout horizons, and also under the 100%-fee-rebate counterfactual. Adverse selection on these pools is not a bear-market phenomenon; it is the steady-state operating mode of the venue.

5.5 Concentration makes it worse, not better

The intuitive response to thin fees is to concentrate: fees accrue where the price is, so a narrow range multiplies fee capture per dollar. The first half of this intuition is correct and we measured it: at a one-hour horizon, 72.5% (ETH) to 81% (CBBTC) of all LP fees accrue within ±0.5% of the prevailing price, and fee density falls by a factor of ∼10 between the center and the 0.5–1% band, and by ∼40 beyond.

The second half is where the intuition fails: *toxic flow strikes at exactly the same place.* Arbitrage trades are, by construction, executed at the current price; a range concentrated at the tick collects a larger share of the fees *and* a larger share of the adverse selection — and the marginal toxicity grows *faster* than the marginal fee as the range narrows. On the 90-day window, a single ±0.5% position of $50k on ETH-USDC collects fees at an annualized $202k rate but absorbs toxicity at $345k/yr; the four-band nested geometry our production optimizer had converged to absorbs $114k/yr against $64k/yr of fees. Concentration scales both legs of a losing trade.

Table 9 in Section 7 quantifies this over the full two-year window with the SIMULATED virtual-LP replay (a $50k constant-notional position walked through the recorded event stream, self-dilution included): every geometry is negative — consistent with Proposition 5.1 — and the net APR is *monotone in center weight*, from the single wide lazy band at the top to the tight actively-recentered band at the bottom (−347 to −1149%/yr), which collects 2.3–4× more fees than the wide band and still loses an order of magnitude more. The full geometry analysis, the 14/14-regime robustness of this ordering, and the reversal of the Monte-Carlo tier's recommendation that it entailed are the subject of Section 7.

5.6 The oracle bound: the money exists, the forecast does not

Is the negative sum a physical inevitability — fees simply too small relative to volatility — or a prediction problem? We separate the two with a clairvoyance experiment (SIMULATED, calibrated Monte-Carlo engine, 4 pairs).

A *clairvoyant sidestepper* — an LP allowed to see 72 hours into the future and step out of the pool exactly when the position is about to lose — beats buy-and-hold on **4 of 4 pairs**. The gap between the oracle and the best causal strategy, i.e. the prize a perfect forecaster would capture, is large: +164 APR points on ETH-USDC, +129 on CBBTC-USDC, +100 on WETH-CBBTC, +351 on SOL-USDC. The wall is therefore not arithmetic: conditional on perfect foresight, timed liquidity provision is profitable.

Every implementable forecaster we then tested failed to capture any of this gap:

- *Forward volatility forecasts* (EWMA and trend variants driving a volatility gate): negative on 4 of 4 pairs; best cell −59 APR points versus doing nothing.
- *A perfect regime label*: an oracle that knows the true chop/trend label at every instant — hindsight information no real strategy can have — still loses versus HODL; only the P&L timing itself pays.
- *Reactive volatility sizing*: trailing (Parkinson) volatility widens the range *after* the move, precisely too late; vol-scaled width lost −25.5 points pooled against a fixed width, despite a genuine but unexploitable +8-point timing residue under a shuffled control.

Remark 5.2. *The pattern is exactly what an information-asymmetry account predicts. Toxicity arrives with the arbitrageurs' private signal — the CEX price a few hundred milliseconds ahead of the pool — which is, by construction, absent from the pool's public state. Public-state functionals (trailing volatility, regime labels, band occupancy) therefore cannot anticipate it; they can only describe the damage after it is done. The oracle experiments bound the value of the missing information; Section 6 identifies who possesses it.*

6 Who wins: measuring the arbitrage ecosystem

If LPs pay ∼0.5 bp of volume into the game, someone collects it. The markout pipeline attributes every swap to its `sender` (the contract or account that called the pool), which lets us measure the winning side of the trade directly. All quantities in this section are MEASURED on the 90-day pilot window (8.7M swaps, both pools); an independent adversarial re-implementation (separate code, disjoint receipt samples) reproduced the totals to the dollar (e.g. ETH-USDC positive markout $284,864 vs. $284,863) and confirmed every structural claim below.

6.1 Actor identification and concentration

Aggregating trader-side markout by sender yields a sharply bimodal population: a small set of dedicated contracts with systematically positive markout, and a long tail of routers and retail flow with negative markout. Three checks support the "dedicated arbitrage bot" reading of the top senders: (i) all top-10 positive senders are contracts (bytecode 0.3–22.7kB, verified on-chain); (ii) each has exactly *one* distinct swap recipient over the entire 90 days — a fixed treasury — whereas a public aggregation router would fan out to thousands of end-user recipients; and (iii) their activity is continuous, 24/7, at 10^4–10^6 swaps per actor.

Table 7: Concentration of gross positive markout ($\tau = 300\,\text{s}$), 90-day window (MEASURED). "Actors" are distinct swap senders; "winners" are senders with positive aggregate markout.

	ETH-USDC	CBBTC-USDC
Distinct actors	1,507	720
Winners (positive markout)	569	292
Gross positive markout (90d)	\$285k	\$113k
Top-1 share	38.2%	57.5%
Top-3 share	58.8%	74.9%
Top-10 share	83.9%	93.5%

The distribution is winner-take-most. The same contract (`0x83d5…90d5`, 22.7kB of bytecode) is the top-1 actor on *both* pools simultaneously and alone captures **43.6% of the combined gross markout**: \$173.6k gross over 90 days (\$704k/yr annualized) on 1.99M transactions — **22,121 transactions per day** at a mean notional of ∼\$1,100, a per-swap gross edge of \$0.07–\$0.11, and a swap-level win rate of ≈50.5%: not clairvoyance per trade but a small positive drift harvested at enormous frequency. Concentration is stable across two disjoint half-windows (top-1 share 0.40/0.36 on ETH, 0.58/0.56 on CBBTC): a structural feature, not a regime artifact.

6.2 The latency race

Base produces blocks every 2 seconds with first-come-first-served ordering by the sequencer, so stale-price capture is a pure latency race: the first swap in a block trades against the pool state left by the previous block. The top-1 bot lands its swap **first in the block in 45.5% (ETH) to 59.0% (CBBTC) of its transactions** — against a baseline that would be negligible for an unprivileged sender given a median in-block transaction index of 129 — and its winning markout is concentrated in those first-in-block swaps. Priority fees are not the instrument of the race (median 0.066 gwei, p99 1.87 gwei, i.e. fractions of a cent): the competition is fought in propagation latency to the sequencer and sub-second CEX price feeds, not in the fee market.

6.3 The pie, net of costs

Sampling 420 transaction receipts across the top actors (plus 240 disjoint receipts by the independent verifier) gives the cost side: median cost \$0.024 per transaction (Base L1 data fees are negligible). Costs consume 6.6% of gross markout in the pooled sample but vary strongly by actor (15–60% for the followers; 29% for the top-1 bot, which nets ≈\$472k/yr across both pools).

Three consequences close the analysis. First, *the whole game is small*: the net-of-gas arbitrage pie of both pools is \$0.2–1.1M/yr — the same order as the LP emissions budget of Section 8, and a ceiling, since the CEX leg is uncounted. Second, *the marginal entrant is priced out by the hedge*: the dominant bot's gross edge of 1–1.5 bp of notional is smaller than the 1–2 bp CEX taker fee needed to hedge; a profitable operation requires maker-tier or zero-fee CEX execution or warehoused inventory risk — infrastructure and balance-sheet advantages, not strategy advantages — and an entrant without them projects to under \$10k/yr. Third, aggregation-router flow — the closest observable retail proxy — has *negative* aggregate markout on both pools: retail takers and passive LPs jointly fund a small, latency-selected set of professional operators.

Table 8: The arbitrage economy of the two pools, 90-day window annualized (MEASURED; net = gross markout − sampled gas; the CEX hedging leg is not observable on-chain and would reduce all net figures further).

Quantity	Value
Gross positive markout, both pools	\$1.61M/yr
of which arbitrage-bot class	\$1.54M/yr
Net of gas, arbitrage-bot class (range)	\$0.20–1.10M/yr
Top-1 bot alone, net of gas	≈\$0.47M/yr
Top-1 gross edge per notional	1.5 bp (ETH) / 0.98 bp (CBBTC)
CEX taker fee (hedging benchmark)	1–2 bp
Aggregation-router flow markout (retail proxy)	−\$10.3k (ETH) / −\$3.8k (CBBTC) per 90d

This measurement also settles the obvious strategic alternative: pivoting from liquidity provision to arbitrage is unattractive on these numbers. The durable output is the markout pipeline itself — a general instrument for measuring the toxicity of any V3-style venue *before* deploying capital to it.

7 The loss-minimizing geometry: from nested ladders to a single lazy range

Given the structural negativity of fee income established in Section 5, the role of range geometry changes: it can no longer be chosen to *maximize a profit*, only to *minimize a structural loss* while the position farms protocol emissions (Section 8). This section documents how our answer to the geometry question reversed when we moved from the calibrated Monte-Carlo evaluator to the event-replay evaluator, and what the final, thrice-confirmed answer is. We regard this reversal itself as a result: it quantifies the model error committed by an evaluator that is blind to per-tick toxicity, a blind spot we expect to be shared by most simulation-based LP backtests in the literature.

7.1 The simulated optimum: a nested, inner-weighted ladder

Our geometry search first ran on the calibrated block-bootstrap Monte-Carlo engine (Section 4), whose fee, impermanent-loss and volatility calibrations all passed against realized production data. On that evaluator, a systematic sweep (4 recentering policies × 3 weight profiles × 8 widths × 4 pairs, multi-regime, out-of-sample, with adversarial re-runs under honest all-leg fee accounting) selected:

- a *nested ladder* of four concentric ranges with half-widths $\{1, 2, 4, 8\} \times w$ around the current price, recentered only when price exits the outermost band ("outer-only" policy) — recycling the inner legs on their own exits was shown to multiply impermanent loss by 4–7× even at zero gas;
- *inner-heavy* capital weights $[0.40, 0.30, 0.20, 0.10]$ from the tightest to the widest leg, which improved fee-only net APR by +12.9 percentage points (pp) over equal weights, positive on all four pairs under honest fee accounting (ETH +16.7, CBBTC +9.7, SOL +18.3, WETH-CBBTC +6.5 pp).

This configuration survived roughly a dozen adversarial attacks (threshold and asymmetric policies, multi-horizon bands, vol-scaled widths, wider outer legs, derived log-normal weights), each refuted by matched controls, and was provisionally locked as the production target. Two of these refutations supplied a methodological tool that proved decisive later:

Definition 7.1 (Matched-control protocol)**.** *Two geometries are compared only under (i) identical recentering instants (both driven by the same outer-band exits), (ii) identical outer reach (the envelope* $[P_0(1-8w),\ P_0(1+8w)]$ *is the same), and (iii) identical gas and swap-cost accounting.*

Any comparison violating (ii) is discarded as a width confound: in every evaluator we used, a wider envelope mechanically reduces rebalancing frequency and impermanent loss, so an apparent gain of a "smarter" shape can be reproduced by a plain scale-up of the baseline. This confound alone explained the apparent high-quantile gain of multi-horizon bands (the "thesis-E" episode, where a $+0.118$ *APR-fraction advantage vanished — and reversed to* -0.10 *pooled — once the geometric baseline was rescaled to the same outer reach).*

7.2 The event-replay reversal

The event-replay evaluator (Section 4), calibrated to 0.97–1.00 of the fees actually collected by our production positions, was then run on the same question: first on the 90-day pilot (10.7M events), then on the full two-year history of the two PancakeSwap v3 production pools on Base (77.8M swaps, 14 dated market regimes fixed *before* the runs). Both runs are SIMULATED in the sense of Section 3 (a virtual \$50,000 LP replayed against the recorded event stream), but the prices, the competing liquidity and the order flow are the measured on-chain history, not a model.

The result inverted the simulated optimum, with no exception in any regime.

Table 9: Two-year event replay (July 2024–July 2026), fee-only net APR of a \$50,000 position, matched-control protocol (Definition 7.1): identical recenter instants, outer reach and costs within each column. $w05/w10$ denote base half-widths such that the outer envelope is $\pm4\%/\pm8\%$. "Tight-active" is a single $\pm w$ range recentering on its own exits. Source: `backtest_2y_{ETH,CBBTC}-USDC.json` (values stored as fractions; shown here in %/yr).

	ETH-USDC		CBBTC-USDC	
Geometry	$w05$	$w10$	$w05$	$w10$
Single $\pm8w$ (lazy)	**−164**	**−93**	**−93**	**−55**
Nested, outer-heavy	−217	−124	−128	−74
Nested, equal weights	−242	−138	−144	−83
Nested, inner-heavy (lock)	−270	−153	−161	−92
Nested, super-inner	−287	−162	−172	−97
Tight-active (single $\pm w$)	−1149	−647	−667	−347

Three facts stand out from Table 9 and the per-regime decompositions:

1. *A single wide range beats the nested ladder in 14/14 regimes*, with zero exceptions across bull, bear and sideways markets, two pools, two widths and both years. Pooled over two years the advantage is +106 pp of APR on ETH at $w05$ (+60 pp at $w10$) and +68/+38 pp on CBBTC. The largest advantage occurs in the deepest bear regime (+178 pp during a −63% drawdown). Gas does not explain it (both sides pay the same, roughly \$70 per two years, against annual deltas of \$500–\$1000).
2. *Equal weights beat inner-heavy weights in 14/14 regimes* as well (+28 pp ETH, +17 pp CBBTC pooled) — the exact opposite of the Monte-Carlo lock, on an identical-geometry comparison where only the weights differ.
3. *The ordering is monotone in center concentration, everywhere*: single > outer-heavy > equal > inner-heavy > super-inner ≫ tight-active. The intuitive "stay glued to the price to farm fees" strategy is the catastrophic extreme of the family, at −347% to −1149%/yr net.

The mechanism is the one measured in Section 5.5: fees are indeed concentrated at the price, but toxic flow strikes at the same ticks and its marginal cost grows faster than the marginal fee as the range narrows — the per-tick blind spot of the Monte-Carlo tier (Section 4). Following this episode we adopted the rule that no geometry conclusion is locked unless two independent evaluators agree and a live forward arm confirms.

7.3 Searching within the lazy family

With the single-range family established, a second sweep explored 23 configurations of the *lazy* single range on the two-year event replay: base half-widths from ±3% to ±15%, out-of-range patience timers (6 h/24 h/72 h), distance hysteresis, minimal-shift versus full recentering, and static asymmetry — each accompanied by negative controls (randomized triggers, sign-shuffled skews, seed repetitions to estimate the noise floor, which is ≈ 0.10–0.13 APR-fraction).

Only three components survived their controls:

1. **Width** ±10–15% is the dominant lever (ETH +42 pp, CBBTC +38 pp versus the ±8% baseline, monotone across both years and both pools); fine-tuning *within* the 10–15% band is indistinguishable from noise ($w10$ and $w11$ under the best policy differ by 2.9 pp).
2. **Patience in general** (+22/+31 pp): waiting out-of-range before acting. Critically, a *randomized* delay of matched frequency performs the same as the calibrated 24-hour trigger (ETH −35.6% random vs −41.9% timed, within the noise floor; CBBTC likewise), so the precise timer carries no information — what pays is rebalancing rarely and letting the price come back, not any signal.
3. **Minimal shift instead of full recentering** (+18 pp ETH, +4 pp CBBTC, positive in all four year-halves), but *only when combined with patience*; without it, shifting is harmful on ETH-USDC (−19 pp, a rebalance-spam effect).

Everything else died under its control: the exact 24-hour value, static and reactive asymmetry (sign-shuffled skew performs identically), distance hysteresis, and sophisticated weight profiles. The best configuration, `lz_w10_t24h_shift` (±10%, 24-h patience, minimal shift), achieves a fee-only net of **−42%/yr** on ETH at simulated $50,000 — against −93%/yr for the ±8% immediate-full-recenter reference — and −15.5%/yr on CBBTC (vs −54.5%) (`lazy_swarm_verify.json`, re-derived independently). We emphasize the sign: even the best geometry found by an exhaustive search on two years of real events remains *negative* before emissions. The search optimizes the minimal loss, not a profit.

7.4 Live forward confirmation

The final check is LIVE: ten shadow strategy arms run forward on the five production pairs with real prices, simulated execution costs and the same NAV-versus-HODL accounting as production (sandbox scale). Table 10 shows the state at the end of the campaign. Because the arms were seeded on different dates, the honest comparison column is production's performance *over each arm's exact window.*

Table 10: Live forward arms at campaign end, July 2026 (LIVE, sandbox scale). "vs B&H" is growth of the wallet/HODL ratio since seeding, averaged over the five pairs; "Δ prod" subtracts production over the same window. Source: production paper database via the dashboard API, re-derived from `shadow_epochs`.

Arm	Window (d)	vs B&H (%)	Δ prod (pp)
`cmix_w15` — single ±15%, lazy	13.0	**−0.30**	**+0.24**
`cmix_lazyshift` — ±10%, patient, shift	13.0	−0.48	+0.06
`cmix` — nested, equal weights	17.9	−1.30	−0.59
`armC`	17.9	−1.77	−1.06
`armB`	17.9	−1.78	−1.07
`cmix_1range` — single, optimiser width	16.8	−1.80	−1.34
`armA`	17.9	−2.13	−1.42
`cmix_ih` — nested, inner-heavy (ex-lock)	16.9	−2.02	−1.49
`cmix_dln` — derived log-normal weights	16.8	−2.17	−1.71
`v2` — legacy engine, tight active	67.5	−8.44	−7.51

The live ranking reproduces the two-year replay ranking exactly: the two wide lazy arms are

the only ones that beat production; everything that concentrates at the center trails; and the legacy tight-active engine is eliminated beyond doubt (−7.5 pp below production over 67 days). A telling detail: `cmix_1range` (single range, but at the optimiser's tighter width) *loses*, while `cmix_w15` (same idea forced to ±15%) wins — confirming that the operative lever is *width*, not range count, exactly as in the replay.

We state the limits of this confirmation plainly. Thirteen to seventeen days cover a single market regime; +0.24 pp is not statistically significant (nearly all of `cmix_w15`'s residual loss traces to a single pair over the window); and the fine ordering between the top two arms (0.18 pp apart) is noise. What the live data does establish is (i) the *hierarchy*, identical to the replay's, obtained by an independent forward-looking evaluator, and (ii) that even the best arm remains negative against buy-and-hold, consistent with the central negative-sum result.

8 Emissions as the product, and the capacity ceiling

If fee income is structurally negative and geometry can only minimize the loss, the only genuinely positive revenue stream on these venues is the protocol's liquidity-mining emissions (CAKE on PancakeSwap, AERO on Aerodrome). This section measures that stream, derives its dilution law, and computes the capacity of the resulting "product" — which turns out to be the binding constraint on the whole trade.

8.1 Emission mechanics and the concentration law

On PancakeSwap v3, staked positions receive CAKE through the MasterChef v3 farming contract, which accrues rewards per unit of liquidity only while the position is in range [5].

Proposition 8.1 (Emission accrual)**.** *Let L be a position's liquidity and* $\mathrm{IR} \in [0,1]$ *its in-range time fraction. The CAKE accrued over a period is proportional to $L \times \mathrm{IR}$, at a per-L rate common to all positions in the farm. Consequently, for a fixed dollar value V, since $L \propto V/w$ for a symmetric range of half-width w to first order (cf. the position-value formulas in [1, 3]), the emission APR of a position scales as* $\mathrm{IR}(w)/w$.

This is a mechanical property of the contract, but we verified it empirically (MEASURED, on-chain): across our three staked positions the predicted-to-realized ratio was 0.98–1.05, and a counter-check on four unrelated third-party positions with widths spanning 86 to 10,987 ticks found the *identical* per-L accrual rate to six significant digits. The law $k = \mathrm{APR} \times w$ held constant across a 7× width span on our own data.

Proposition 8.1 raises an obvious temptation: concentrate to farm more CAKE. It fails, because the adversarial loss of Section 7 grows *faster* than $1/w$ as the range narrows: on the two-year replay, tightening from the lazy optimum destroys more in toxicity and impermanent loss than the extra emissions are worth (the tight-active configuration nets roughly −530 to −740%/yr on ETH even with its elevated CAKE accrual). The concentration law therefore *reinforces* the wide-lazy geometry rather than reversing it.

8.2 Measured emission rates

Table 11 gives the forward CAKE rates measured on our production positions (MEASURED: incremental on-chain accrual, not protocol-advertised APRs). The production positions' liquidity-weighted mean half-widths at the time of measurement were ±9.3% (ETH), ±8.9% (CBBTC) and ±6.0% (WETH-CBBTC) — i.e. these rates are already approximately the rates of the lazy-width geometry.

8.3 The dilution law and the capacity ceiling

Emissions are a fixed budget shared among stakers, so any entrant dilutes the rate. Let E_{an} be a farm's annual emission budget (in USD, at current token price), D the incumbent in-range staked

Table 11: Measured forward CAKE emission rates on production positions (PancakeSwap v3, Base, 0.01% pools; SOL and EURC pairs sat on Aerodrome unstaked and earn no emissions).

Farm	APR (%/yr)	Half-width
ETH-USDC	30.7	±9.3%
CBBTC-USDC	16.3	±8.9%
WETH-CBBTC	14.0	±6.0%

Table 12: Net performance versus buy-and-hold at $50,000, CAKE included, two-year multi-regime event replay (SIMULATED on measured events), lazy ±10–11% geometry (the optimum). Ranges reflect the two markout horizons and policy variants.

Pair	Net vs B&H (%/yr)
ETH-USDC	−12 to −22
CBBTC-USDC	−2 to −4
WETH-CBBTC	≈ −8

capital, and C the entrant's capital deployed at the same effective concentration. The entrant's gross emission APR is

$$\mathrm{APR}(C) \;=\; \mathrm{IR}\cdot\frac{E_{\mathrm{an}}}{C+D}, \tag{12}$$

with IR its in-range fraction. Both E_{an} and D were measured on-chain per farm. The total emission budget of our three PancakeSwap farms is

$$E_{\mathrm{an}}^{\mathrm{tot}} \approx \$1.02\mathrm{M/yr} \quad (\text{ETH } \$519\mathrm{k} + \text{CBBTC } \$176\mathrm{k} + \text{WETH-CBBTC } \$329\mathrm{k}),$$

for *all market participants combined* (re-verified to 0.2%). Inverting Eq. (12) with the measured incumbent bases gives the aggregate capital that can be deployed at a given gross emission yield: approximately $0.7M at 20%/yr, $3M at 10%/yr, and $10M at 5%/yr — and fee dilution compounds the decline (about −12 pp on ETH at $1M deployed).

Gross capacity, however, is not the relevant number. Combining the emission overlay with the loss-minimizing geometry of Section 7 on the two-year replay gives the *net* result versus buy-and-hold at $50,000 per pair:

Table 12 reports the result; only CBBTC-USDC comes near breakeven.

The capacity of a net-positive-versus-HODL product on these farms is therefore $0 on the two-year window: no size exists at which the strategy both matters commercially and beats holding the tokens. Size is the crucial variable here. At $50,000 the simulated position already represents 25–43% of the in-range liquidity and dilutes its own fees and emissions (Eq. (12) with $C \sim D$); at sandbox scale (a few hundred dollars per pair), self-dilution vanishes and the live system indeed floats around buy-and-hold plus CAKE — which is exactly what production measured (Section 7.4). The strategy is viable as a toy and impossible as a product: a retail offering of even a few million dollars has, mathematically, no room to exist on a $1M/yr emission budget shared with all incumbents.

8.4 The Aerodrome door

Aerodrome's concentrated-liquidity pools (Slipstream) are the natural candidate for a larger emission budget, and we measured them with the same event pipeline (MEASURED: 5.87M swaps over 71–91 days of complete on-chain events across six pools, fee growth read from `ΔfeeGrowthGlobal`, toxicity by per-swap markout at 5 and 30 minutes; independently re-measured by a second agent). Three findings:

(a) Fees are net-negative on all six pools. Table 13 summarizes: on every pool measured, realized fees per dollar of volume are smaller than the adverse-selection cost, so an unstaked LP loses on net — the same structural conclusion as on PancakeSwap. Between 58.6% and 81.7% of volume arrives as the first transaction of its block, the signature of arbitrage bots.

Table 13: Aerodrome Slipstream pools, complete on-chain events, 71–91-day windows (April–July 2026). "Net LP" includes fees and markout toxicity ($\tau = 5$ min horizon; the 30-min horizon is the same sign everywhere). Emissions are annualized from gauge `NotifyReward` events at AERO = \$0.56. Pools are identified by pair and short address prefix. Source: `counter_*.json`, gauge-event database.

Pool	Pair	Fee (bp/vol)	Net LP (bp/vol)	Emissions (\$/yr)
`0x4e96…`	CBBTC-USDC	0.75	−0.45	4.35M
`0x160d…`	CBBTC-USDC	1.80	−0.53	3.71M
ETH pool	ETH-USDC	3.96	−0.53	8.18M
SOL pool	SOL-USDC	3.50	−0.96	1.21M
`0xf39b…`	EURC-USDC	0.66	−0.34	0.41M
`0xc5e5…`	EURC-USDC	∼0.70	−0.55	0 (gauge dead 16 Apr)

(b) The dynamic fee is a snapshot trap. Slipstream's `fee()` is dynamic; a spot read is not a parameter. On the CBBTC pool `0x4e96…`, a fee snapshot of 2.1 bp (210 pips read at a spike; the fee cycled 30 → 150 → 30 pips within 16 seconds) overstated the *realized* fee of 0.75 bp by 2.8×, because 75.6% of volume executes at the base fee of 30 pips. A venue-migration candidate modeled on the snapshot (∼ 3× better capture) was refuted by the realized tick-level measurement (0.77× actual). What attracts volume to a pool — a cheaper effective fee, favored by aggregator routing — is precisely what destroys fee capture; only realized fee growth over a window is admissible evidence.

(c) Emissions are 8–18× the CAKE budget; the staked question is out of scope. The six pools distribute ≈ \$17.9M/yr of AERO emissions (the ETH pool alone, \$8.2M/yr). Slipstream imposes an exclusive-or: a staked position earns *emissions only* (its fees are redirected to the gauge), an unstaked position earns fees only — so, fees being net-negative everywhere (finding (a)), staking is the only rational mode there, and the position becomes a purchase of emission flow against toxicity, i.e. exposure to the AERO price rather than a yield. Whether that trade clears was not concluded to this paper's evidentiary standard (single window, no forward test, exact staked-share replay incomplete) and lies outside the subject of this paper — the economics of the *range* itself, which findings (a) and (b) show behave on Slipstream exactly as on PancakeSwap. The staked door remains *open but unproven*, with a capacity ceiling of the same nature as Section 8.3: a fixed budget shared among all participants.

The measurements of Sections 5–8 thus compose into a single statement — negative-sum fees, emissions as the only real revenue, a wide lazy range as the loss-minimizing geometry, and an emission budget too small to support a scalable product — whose consequences are drawn in Section 11. The rational response is to stop supplying passive concentrated liquidity to these venues and to redirect the tooling (the event-replay evaluator and the toxicity pipeline) toward AMM designs that do not leak the spread to arbitrageurs by construction.

9 A Catalog of Negative Results

A central contribution of this study is a systematic record of what does *not* work. Over the campaign we tested approximately twenty distinct levers — prediction signals, hedges, range geometries, reward-timing policies, venue migrations, and one full strategic pivot — each with a stated protocol, a matched control where applicable, and an identified evaluator. Table 14 summarizes every lever, its verdict (*dead* = refuted with evidence, *kept* = survived controls, *open* = not concluded), the key measured figure, and the proximate cause of death. Unless stated otherwise, "pts" denotes percentage points of annualized return relative to the stated benchmark, at the \$50 k reference notional.

Three observations structure the catalog. First, *every* intelligence-type lever — prediction, timing, hedging, sophisticated geometry — died under controls; the only survivors are levers of

Table 14: Negative-results catalog: every lever tested, April–July 2026; each row is traceable to the project results ledger. "E" = event replay, "G" = calibrated GBM Monte-Carlo, "L" = live paper arms, "P" = on-chain probes.

Lever	Verdict	Key figure	Cause of death
Ranging tournament, 13 methods (constant-mix, wide-passive, vol-predictive, Bollinger, barbell, …)	dead*	constant-mix ranked #1 (−4.2 pts fee-only, +12.4 with rewards); *the live bot ranked 7th of 12* (G)	concentrating geometries lose to impermanent loss; only constant-mix survived
Volatility-gate / vol-forecast ML	dead	forward prediction fails 4/4 pairs; best variant −59 pts (G)	signal is reactive, not predictive: widens *after* the move
Perfect regime oracle (chop/trend)	dead	loses to HODL even with oracled regime labels (G)	knowing the regime is insufficient; toxicity arrives within regimes
Clairvoyant P&L oracle	kept (theor.)	beats HODL 4/4 pairs; capturable gap +164 pts (ETH), +351 pts (SOL) (G)	— (positive control: the wall is prediction, not physics)
Option hedge (long gamma)	dead	−48 pts at no-arbitrage premia (G)	premium costs ≈ what it saves; viable only at ∼100% fee APR
Delta-hedge / AAVE-blend / regime-timing	dead	−90 to −120 pts; worst of tournament (G)	hedging costs compound the structural fee deficit
Multi-timescale bands (four bands from {12 h, 3.5 d, 7 d, 30 d} volatility)	dead	−3.9 pts pooled vs. geometric at q_{68}; apparent q_{90} gain reproduced by a pure width scale-up (G)	width confound: the "gain" was a wider envelope, not the band shape (matched-outer-reach control)
Derived log-normal leg weights	dead	+1.1 pts pooled only; loses on ETH (G)	statistically indistinguishable refinement, not a lever
Vol-scaled width	dead	−25.5 pts pooled, loses 4/4; genuine vol-timing residue +8.4 pts vs. shuffled control, unexploitable (G)	breathing around the calibrated optimum costs more than the timing earns
Asymmetric / trend-skewed placement	dead	attributable directional signal +0.3 pts pooled; random-sign shuffle performs equally (G)	width artifact, no directional information
Wider outer leg (×1.5/2/3 scale-up)	dead	−9.8 pts pooled, monotonically worse with scale (G)	confirms the geometric ladder within the GBM world
Threshold / asymmetric recentering policies	dead	apparent win collapses under honest all-leg fee accounting (e.g. ETH +0.724 → +0.576 vs. outer-only) (G)	idle-leg fee over-credit artifact
CAKE harvest timing (sell-in-calm; $10/$30 thresholds)	dead	timing delta vs. random control = 0.00 exactly; high thresholds: 41–47% of paths lose (G)	no exploitable timing; threshold "gain" is reward-inventory beta
USD–USD stable-pool niche	dead	realized fee APR 1.35–5.4%, all below the 6% lending hurdle	microscopic fee tiers: "disguised lending"
Exotic stable + peg-guard (GHO–USDC v4)	dead	headline 15.3% fee APR was a 24 h snapshot; sustained medians 0.09–2.2%; two spike days = 27% of 99-day volume	snapshot artifact; no live pool clears the lending hurdle on sustained volume
Venue migration, CBBTC → Aerodrome (`fee()` snapshot, twice)	dead	modeled "∼3× capture, +64%" → realized ΔfeeGrowth: 0.77× incumbent; `fee()` observed cycling 30→150→30 in 16 s (E, P)	Slipstream `fee()` is dynamic: a spot read overstated realized fees 2.8×; cheap fees attract volume but kill capture
Heatmap-suggested low-tilt placement	dead	loses even in-sample once look-ahead removed (−1.44 vs. −1.32) (E)	look-ahead: a visual suggestion is not a strategy
Tight active range ("hug the price")	dead	2.3–4× more fees, net −350 to −1150 pts over 2 y at $50 k (E)	toxicity and IL at the active tick grow faster than fee capture
Concentrated live family (v2, arms A/B/C)	dead	v2 loses vs. HODL 5/5 pairs and vs. production 5/5; live: −8.44% vs. B&H, −7.5 pts below production over 67.5 d (L)	the concentrated family loses in simulation *and* forward
Arbitrageur pivot (become the informed flow)	dead	one HFT bot takes 44% of gross (22 121 tx/day); net-of-gas pie $0.2–1.1 M/yr; CEX taker hedge erases the 1–1.5 bp edge (E)	latency infrastructure race, not a capital game; entrant expectation < $10 k/yr
Inter-pair capital allocation	**kept**	+5 pts net APR forward, OOS in both directions; SOL zeroed in 200/200 resamples (G)	— (sole positive lever: zero SOL, equal-weight rewarded pairs, 40% cap)
Lazy-range swarm survivors	**kept**	width 10–15% (+42/+38 pts vs. ±8% full-recenter baseline), general patience (random delay matches timed), minimal shift (E)	— (the loss-minimizing geometry; exact timer, hysteresis and asymmetry all noise)
Aerodrome staked net open yield	open	emissions ≈ $17.9 M/yr across six pools; marginal staker estimates positive on 4 pools	not concluded: position-level staked-share replay incomplete; no forward test

*All 13 methods dead except constant-mix; predates the reward correction and the event-replay reversal (Sec. 7).

simplicity (one wide lazy range, patience) and of capital common sense (allocation tilts). Second, the clairvoyant-oracle result shows the binding constraint is predictability, not physics (Section 5.6). Third, a large fraction of the initially "positive" results were artifacts, which motivates the taxonomy that follows.

10 Methodological Lessons: An Artifact Taxonomy for Empirical DeFi

The campaign's most transferable output is a taxonomy of measurement artifacts, each of which produced a spurious positive result before being caught by a specific guard.

(i) Width confound → matched-reach control. Any change to a range structure that incidentally widens the envelope inherits a mechanical benefit (fewer recenterings, less gas, less realized IL) that masquerades as a strategy edge. The multi-timescale band thesis "won" at the q_{90} quantile (+11.8 pts) until the control was rescaled to the same outer reach, after which it lost everywhere. *Guard:* never compare structures of different envelope width; rescale the control to matched outer reach.

(ii) Idle-leg fee over-credit → honest all-leg accounting. A multi-leg simulator that tracks occupancy only for the innermost leg silently credits fees to legs the price has exited. Threshold and asymmetric recentering policies appeared to win solely through this channel; under all-leg occupancy accounting the ordering reversed. *Guard:* fee accrual gated on per-leg, per-period in-range occupancy; any advantage that disappears under honest accounting is rejected.

(iii) Snapshot metrics → sustained medians and realized growth. Two independent incidents: a stable pool advertised 15.3% fee APR that was two spike days out of ninety-nine (sustained median 0.09%); and the dynamic Slipstream `fee()` getter whose spot read overstated realized fees by 2.8× (Section 8.4). *Guard:* venue and pool comparisons use only *realized* quantities — sustained medians over multi-week windows, fee income reconstructed from ΔfeeGrowth — never an instantaneous getter multiplied by a headline volume.

(iv) Look-ahead → in/out-of-sample discipline. A fee-density heatmap computed over a window suggests placements optimal *for that window*; deployed as a strategy, the suggested tilt lost even in-sample once the leakage was removed. The allocation study's "all-in on the winning pair" variant was pure survivor bias, killed by resampling. *Guard:* pre-registered regime boundaries, IS/OOS splits, shuffle and random-sign controls, and machine-verifiable no-look-ahead self-checks in every backtest.

(v) Unit errors → re-derivation from raw data. The most expensive analysis error of the project: the `net_apr` fields of the event-replay outputs are fractions (−0.42 means −42%/yr). One relayed headline ("lazy beats buy-and-hold by +30%") was wrong purely on units; corrected, nothing beats buy-and-hold at $50 k. *Guard:* any load-bearing number is re-derived from raw data with one independent control division before it is relayed; rankings survive unit errors, levels do not.

(vi) Counter freshness → liveness checks. A long-running data fetch died silently while its progress file continued to be read as evidence of progress for nine days. *Guard:* a progress counter is only evidence when paired with its modification time and a verified live process.

(vii) Locking rule: two evaluators plus forward confirmation. The deepest lesson subsumes the others: a calibrated evaluator that passes all its gates can still confidently lock a wrong answer (Sections 4 and 7.2). Since no evaluator is artifact-free, no conclusion is locked on the authority of a single one: a locked verdict requires two independent evaluators in agreement plus forward confirmation from live arms, every verdict citing its evaluator's known blind spot. Adversarial verification of every load-bearing result caught a stale headline figure, a double-counted position, and two sub-claims that failed direct on-chain re-reading.

We suggest this taxonomy applies well beyond our pools: DeFi pipelines are unusually exposed to snapshot getters, dynamic parameters, unit-convention drift, and simulators well calibrated on aggregates yet blind to per-tick adverse selection.

11 Conclusion and Future Work

This paper asked whether Uniswap-v3-style concentrated liquidity provision can be made profitable for the provider, and answered it with the strongest instruments we could build: a calibrated simulator, an exact event replay validated against production-collected fees, and live forward arms. The measured answer is negative and, we argue, definitive for this class of pool: liquidity provision "for the fees" is a structurally negative-sum game against better-informed arbitrage flow, in all 14 dated market regimes with no exception. Concentration aggravates the deficit; prediction of adverse flow failed under every signal we tested; the loss-minimizing geometry is the least sophisticated one — a single wide (±10–15%) lazy range that shifts minimally after patient delays — confirmed independently by the event-replay evaluator, the calibrated simulator (after correction), and forward live arms whose ranking reproduced the two-year backtest; and the emissions budget of our farms supports *zero* scalable capacity net-positive against buy-and-hold at product size. Publishing this as a negative result is the point: the measurement pipeline that established it cost a fraction of what deploying the apparent "optimizations" of Table 14 would have lost.

The scope of these claims is the scope of the data: one chain (Base), two venues, low-fee-tier major pairs, a two-year replay window, and a live deployment that is small by design. We measured, not assumed, that the same structure holds on every pool we examined on both venues; we have not measured other chains, higher fee tiers, or pools whose flow is majority-retail, and the markout pipeline exists precisely so that such pools can be tested before capital is deployed to them. We note, however, that BTR's internal R&D on other chains and venues, conducted before and alongside this campaign, reached the same qualitative conclusion; we cite it as corroboration only, since those measurements do not meet the provenance standard applied to the numbers in this paper.

We record three explicit conditions under which the verdict could change and should be remeasured: (1) a pool whose flow is demonstrably majority-retail, i.e. realized fee income exceeding measured toxicity — a condition the markout pipeline can now test *ex ante*; (2) an emissions budget an order of magnitude larger whose mechanism does not dilute instantly — the Aerodrome staked-gauge door remains open but unproven (Section 8.4); and (3) a deliberately sandbox-scale deployment, at which measured behavior already floats near buy-and-hold plus emissions and the final lazy configuration is approximately optimal.

Future work: pricing the flow. The structural conclusion admits a constructive reading. If a passive maker quoting a static fee against informed flow must lose — as both the LVR decomposition of Milionis et al. [2] and our per-swap markout measurements indicate — then the designer's remaining degree of freedom is not the maker's *geometry* but the venue's *pricing*. Every result in this paper points at the same asymmetry: the arbitrageur knows the fair price and the pool does not, and a constant-function market maker has no instrument with which to charge for that information gap. A venue that closes the gap must price the flow itself: quote around an explicit fair-value mark rather than around its own inventory state, widen its effective spread when volatility, quote staleness or price uncertainty make informed flow more likely, and

skew its quotes to shed or attract inventory rather than waiting for arbitrage to rebalance it at the LPs' expense — the classical inventory-based market-making response, brought on-chain.

This is the direction of *BTR Prime*, a DEX design under development by our team: an adaptive inventory market maker in which depth and curve shape adapt to measured volatility, the spread dynamically charges precisely the components this paper measured as LP losses, and inventory-aware quoting is combined with structural safeguards against one-sided depletion. In that design the toxicity estimates of this paper's pipeline become an input to pricing rather than a post-mortem: the markout instrument built to explain why passive LPs lose is the same instrument a venue needs in order to charge informed flow its true cost. The design and its analysis are left to a dedicated report.

References

[1] M. Echenim, E. Gobet, and A.-C. Maurice, "Thorough mathematical modelling and analysis of Uniswap v3," HAL preprint hal-04214315, version 2, 2023 (revised 2025). Available: `https://hal.science/hal-04214315`.

[2] J. Milionis, C. C. Moallemi, T. Roughgarden, and A. L. Zhang, "Automated market making and loss-versus-rebalancing," arXiv preprint arXiv:2208.06046, 2022.

[3] H. Adams, N. Zinsmeister, M. Salem, R. Keefer, and D. Robinson, "Uniswap v3 core," Uniswap Labs whitepaper, March 2021. Available: `https://uniswap.org/whitepaper-v3.pdf`.

[4] Aerodrome Finance, "Aerodrome documentation: Slipstream concentrated liquidity and gauge emissions," 2024–2026. Available: `https://aerodrome.finance/docs`.

[5] PancakeSwap, "PancakeSwap v3 documentation: concentrated liquidity and MasterChef v3 farming," 2023–2026. Available: `https://docs.pancakeswap.finance`.

Data and code availability. The complete evidence base — the results ledger (every lever of Table 14 with protocol, controls and raw outputs), the event-replay pipeline and its calibration checks, the two-year backtest JSONs, the Monte-Carlo engine, probe time series, and the production and paper-arm databases — is maintained in the project repository, together with a verified final audit report in which the twenty load-bearing figures cited here were independently re-derived from raw sources (20/20 confirmed). Public pool addresses appear in the data registry; private operational details are withheld.