RESEARCH NOTE · SHORT PUBLIC EDITION

Alpha search in crypto perpetual futures

A counted search, July to October 2026, and one downgraded pass

Denys Shkola

8 November 2026. Short edition of a working paper dated 21 October 2026; not peer-reviewed.

Abstract. From July to October 2026 I searched for a trading edge in crypto perpetual futures that survives realistic costs; yield strategies such as funding carry did not count. By 21 October the search had counted 1,871 trials, about 1,700 of them estimated for an early phase that was not logged cell by cell. From 12 October every new test was written into a hash-chained ledger, with its pass and kill bars, before it ran. Order-book, arbitrage, pairs, event, carry, transformer and most other families failed, mostly because the predicted move was smaller than the fee. One dollar-neutral, cross-sectional long/short book that holds positions for about a week passed its single look at a 2026 holdout: $+23.0$ bp/day of capital net of costs and funding, HAC t 2.61, Sharpe ratio 3.45, worst drawdown 7.9% (backtest). A separate adversarial review that I ran downgraded the pass: part of its design had already seen the window, the three clean months earned $+9.1$ bp/day at t 0.90, five coins made 67% of the 2026 net return, and the deflated Sharpe ratio is 0.0047. No standalone edge was found. This edition reports the data, the protocol, the search accounting and the backtest results; it does not disclose how the surviving book is built.

Disclaimer. Research only: not investment advice, not a recommendation to trade and not an offer of any financial service. Every result is a backtest on historical data; it describes the past and is not a forecast. Nothing in this paper was traded with real money. Crypto derivatives are leveraged instruments, and positions can lose all the capital put in.

Headline numbers. Binance data; returns in basis points (bp, 0.01%) of fixed capital per day at 1x gross exposure, not compounded, net of fees, slippage and funding on the realistic maker-taker cost line unless marked.

Search	1,871 trials counted by 21 October 2026: about 1,700 estimated, 154 registered, 17 charged
Families	11 strategy categories; none produced a standalone edge that cleared its own rules
Survivor, 2026 holdout	$+23.0$ bp/day, HAC t 2.61, Sharpe 3.45, worst drawdown 7.9%; all-taker line $+22.1$ bp/day (t 2.53). One look, declared peeked
Survivor, clean slice	July to September 2026, 91 days: $+9.1$ bp/day, t 0.90
Survivor, development	2023–2025: $+11.1$ bp/day, t 4.46, Sharpe 2.44, worst drawdown 11.3%
Deflation	deflated Sharpe ratio 0.0047 at the 1,866 trials counted at confirmation
Concentration, 2026	top 5 coins made 67% of the net return; the best 20 days made 101%
What breaks it (development)	without small coins -2.02 bp/day (t -0.49); without the best 30 days $+4.08$ bp/day (t 1.73)
Verdict	pass, downgraded. No standalone edge; Binance data only; not a tradable result

1 The question

1.1 Alpha, not carry

By alpha I mean a price-prediction edge that survives fees, slippage and funding on data the strategy has not seen. Carry is excluded: positions whose return is a yield, such as the funding a perpetual pays to the less popular side. The exclusion is also empirical. An audited cash-and-carry on BTC and ETH earned 1.8% a year since 2022 and 0.4% a year in 2025–2026, about what cash pays (Section 4).

1.2 How the work ran

The work ran from 12 July to 21 October 2026: 39 days of research to 19 August, a 52-day pause with no recorded research, and 11 days from 11 to 21 October.

- **July and August, before the ledger.** A measurement rebuild on 15 July found gross alpha of about zero: the average trade lost 16.18 bp against round-trip costs of 13.5 to 16.5 bp, so the loss was the fee bill, and every "Sharpe ratio" in three test harnesses had been a t-statistic. Short-horizon signals failed out of sample on 22 July, and on 30 July a walk-forward go/no-go over nine families and 6.5 years promoted nothing. In mid-August a carry rule and a carry-momentum mix briefly looked like a breakthrough.
- **The re-audit, 11 October.** A re-audit overturned the August results: carry yielded about cash, the carry-momentum mix was not validated, and the real number of configurations searched so far was about 1,700. Every earlier "win" that was audited shrank or vanished.
- **The registered program, 12 to 21 October.** Round 1 designed 7 families with 50 hypotheses, froze one cost table and built the hash-chained ledger, a purged cross-validation splitter and a point-in-time universe. Round 2 recorded 187 claims, 121 of them kills; the one survivor was a maker-execution method that saves about 2.5 bp per order. Round 3 tested machine learning at 4 and 24 hours, two transformers and a Bybit panel, then lower-turnover books that hold positions for about a week; one of them was confirmed on the holdout and stress-tested.

1.3 Why the bar is high

Every strategy variant is a lottery ticket. If N strategies with no edge are each scored by a standard normal z, the expected best score is

$$\mathbb{E}\Big[\max_{i\le N} z_i\Big] \approx (1-\gamma)\,\Phi^{-1}\Big(1-\frac{1}{N}\Big) + \gamma\,\Phi^{-1}\Big(1-\frac{1}{Ne}\Big), \qquad \gamma \approx 0.5772, \tag{1}$$

which is 2.51 at $N = 94$, 3.40 at $N = 1{,}700$ and 3.51 at $N = 2{,}500$; at the program's 1,871 counted trials it is 3.43 (Figure 1). A single t of 3 is what luck alone delivers after a search of this size. On the ledger's own spread of Sharpe ratios, the best of 1,840 null trials shows a daily Sharpe ratio of 0.2473, about 4.7 annualised. This is the multiple-testing problem of the factor literature [4] and the reason for the deflated Sharpe ratio [2]. What carries information is (a) a pass rule written before the data are seen, (b) a count of every variant, (c) a sealed sample looked at once, and (d) controls that should show no skill, reported whether or not they behave.



Figure 1: The luck penalty, equation (1): the expected best t-statistic among N strategies with no edge. At the 1,871 counted trials it is 3.43, far above the usual single-test pass mark of 2.

2 Data

All prices, volumes, funding payments and order books come from the exchanges' public archives and interfaces, without keys. The main evidence uses Binance USDT-margined perpetuals (USDT is a dollar-pegged token, so returns and margin are in near-dollars), from the checksum-verified public archive. Bybit is the second venue, through its public interface. No store exists for Hyperliquid, OKX or Kraken, so nothing here was tested on those venues.

Table 1: Main data stores. Bar frequencies from 100 milliseconds (order books) to one hour; the surviving book uses the Binance stores only.

Store	Content	Window
Binance 5-minute panel	500 symbols, 81,979,123 bars; 1,167,210 funding settlements; premium-index bars	Dec 2021 to Sep 2026
Binance hourly	498 symbols, 4,153,874 rows	Jan 2022 to Sep 2026
Bybit hourly	1,108 symbols including delisted ones, 13,895,270 rows (see correction 2, Section 7)	Jan 2021 to 12 Oct 2026
Bybit order books	100 ms top-10 books for BTC, ETH and SOL; 163 day-files	Aug and Sep 2026

One flow family also used Binance data from February 2020. Funding is the realised rate at each settlement, usually every 8 hours and for some contracts every 4 or 1, booked on the position open at that instant.

Point-in-time universe. At each month start, symbols are ranked by recent trading volume using only bars opened before the month began; stablecoins, index, stock and commodity contracts, leveraged products and dated futures are excluded. The traded universe is the top 100 plus BTC and ETH, in three liquidity tiers: tier 1 is BTC and ETH, tier 2 the rest of the top 20, tier 3 all smaller coins. The Binance universe draws on 829 symbols over 78 months, the Bybit universe on 699 over 75. Because membership uses earlier data only, coins later delisted stay in the universe while they traded: the share of each month's top 100 later delisted is 12% in July 2021, 5% in January 2023, 3% in January 2024 and 1% in January 2025.

Periods. Development: 1 January 2023 to 31 December 2025, 1,095 daily decisions, with 2022 as a training-only warm-up. Holdout: 1 January to 29 September 2026, 272 daily decisions, opened once. Clean slice: 1 July to 29 September 2026, 91 days: the only months that an earlier, related signal family had never seen. Part of the book's design had still been evaluated on them (Section 5.9). Older event hypotheses used an event holdout from January 2025 to September 2026. Three holdouts were never opened.

Known limits of the 2026 window. The August and September 2026 universe reuses the July ranking, and the liquidity tiers are one static list. For 109 of the 209 symbols traded in 2026, the prepared 2026 hourly series stop at month boundaries, which forces early exits that are optimistic but use no future data. Two coins have no funding events in the third quarter.

3 Method

3.1 Registration and the ledger

From 12 October, every task started with a registration file that fixed the cells (fully specified strategy variants), data windows, splitter, cost line, metric, pass bar and kill bar; a cell without a kill bar was refused. The file's SHA-256 hash went into the ledger, an append-only log in which each record carries the hash of the one before, so any edit to the past breaks the chain. The ledger refuses a result whose registration file has changed, whose cell was never registered or whose registration came later than the result. Work that is not a trial, such as feature censuses, goes to a separate diagnostics ledger. By 21 October there were 19 trial and 33 diagnostic registrations. An adversarial review fixed 16 infrastructure defects, among them an embargo that silently truncated to zero, and the ledger verified clean on 21 October.

3.2 Counting the search

A trial is one distinct registered strategy cell: a hypothesis with fixed parameters, universe, horizon and cost line, identified by a hash of that content. It counts when registered, not when run, so skipping a bad-looking cell cannot lower the count; four registered cells were never run for lack of compute and still count. Diagnostics are not trials, but when a trial picks a parameter after seeing a diagnostic, that diagnostic's cells are charged to it. The count used for every deflation is

$$n_{\text{program}} = \underbrace{1{,}700}_{\text{estimate}} + \underbrace{154}_{\text{registered}} + \underbrace{17}_{\text{charged}} = 1{,}871. \tag{2}$$

In short: 1,871 trials counted, about 1,700 of them estimated. The 1,700 is an estimate, written once on 12 October, of the July and August search, which was never recorded cell by cell. Sixteen of the charged cells came from choosing a replacement leak control after its diagnostics were seen (Section 3.5), one from a sizing study. The 1,145 diagnostic cells are not trials, and 1,871 backtests were not run. Counts quoted for earlier moments, such as 1,840 or 1,866, are the count at that time.[1]

3.3 Out-of-sample protocol

Within the development period, models are evaluated by combinatorial purged cross-validation (CPCV) [1]: the period is cut into contiguous time groups, each fit holds out two, and the 36 fits of a cell assemble into 8 complete out-of-sample paths. Purging deletes training rows whose label window overlaps a test label, and an embargo at least as long as the label horizon drops the rows after each test block. The year 2022 is warm-up, always trained on and never tested. Headline statistics are medians over the 8 paths, which share fits and are therefore correlated. CPCV paths are out of sample but not a walk-forward: some of their models train on later data (Section 7).

After a holdout opens, each finalist, at most three chosen by a development-only rule written in advance, gets exactly one holdout result; a miss is a kill, never a reason to try the next cell. Every load of data past the development period is logged (13 on record). The Binance holdout was opened once, on 19 October, for the selection of a model book. The surviving book was registered and took its own single look at the same window on 21 October, after that result was known, so its look is declared peeked. Later sizing studies re-read the window, so their 2026 numbers are post-selection checks, not fresh holdouts.

3.4 Statistics and deflation

The HAC t is the Newey–West t-statistic of the mean daily net return with 10 lags [5]; an ordinary t would overstate the evidence, because a book of overlapping weekly positions has autocorrelated daily returns. Sharpe ratios are annualised with $\sqrt{365}$; the maximum drawdown is the largest fall of the cumulative sum from a previous high, in % of capital.

The deflated Sharpe ratio (DSR) [2] asks whether an observed Sharpe ratio beats the best of N null trials, with $N = n_{\text{program}}$ and the variance of Sharpe ratios across trials measured from the ledger, never defaulted; the final pass gives a variance of 0.005215 and a null benchmark of 0.2473 per day at $N = 1{,}840$. DSR was a pass bar for the short-horizon machine-learning cells, an event fade and an earlier daily flow book, which died on it at 0.4876 against 0.5. By registration it was not a bar for the weekly books. The reason was written before the run: at the variance then estimated (about 0.008), the null maximum at about 1,844 trials was about 5.8 annualised (the final ledger value is 4.7), which would kill any plausible low-turnover cross-sectional book whatever the data say, so t, sign and per-year bars plus one holdout carried the weight. The consequence is stated plainly in Section 5.9: the survivor does not survive deflation for the full search. The probability of backtest overfitting (PBO) [3] needs a grid of cells; the final weekly books are single cells and have none.

3.5 Controls that should fail

Score placebo. Each final book is re-run 100 times with its ranking shuffled across coins within each day, keeping the book's code, costs and funding; the real book should beat every run. **Leak control.** A leak control retrains the pipeline on labels that cannot carry information about the test period and requires the result to show no skill. The control first registered for the weekly model book failed, so its first four cells were declared void and the holdout stayed closed; it then proved defective, because a large share of its training labels fell inside the test spans. It was replaced by a within-day label permutation, which shuffles training labels across coins inside each day. Disclosure: the replacement was chosen after the original failed. The ledger records both, charges the 16 diagnostic cells involved and keeps the original verdict.

3.6 Costs

One cost table was frozen on 12 October and used by every trial. Base fees per side are 5.0 bp taker and 2.0 bp maker on Binance and 5.5 and 2.0 on Bybit, both from secondary sources. Slippage and latency grow from BTC and ETH to the smallest coins, so a round trip costs from about 4 bp, maker on both sides, to about 29 bp, all taker on the smallest coins. The maker figures are optimistic: adverse selection, being filled just before the price moves against the order, is costed at zero, while naive quoting earlier in the project lost 2.5 to 5.7 bp per fill to it. Results are judged on a realistic maker-taker line: position increases rest as limit orders for an hour and convert to market orders if the market does not trade through them; reductions pay taker. An all-taker line is the conservative check. Funding is booked at each settlement.

[1] Later rounds in November raised the count to 1,915. Nothing in this edition depends on them, and it keeps the 21 October count throughout.

3.7 Pass and kill bars

The bars were written before the data were touched and applied as written; near misses were not rounded up. Table 2 gives the bars that decided the weekly books.

Table 2: Registered bars for the weekly books. Nets are bp per day of fixed capital on the realistic line unless marked; HAC t uses 10 lags.

Test	Pass if all hold	Kill if any holds
Weekly model book, development	net > 0; HAC $t \geq 2.5$; at least 2 of 3 years positive; all-taker net > -2 bp/day	net ≤ 0; Brier skill > 0 on at most 3 of 8 paths
Weekly model book, holdout	net > 0; HAC $t \geq 2.0$; all-taker net > -2 bp/day	anything else
Surviving book, holdout	net > 0; HAC $t \geq 2.0$; clean-slice net > 0; all-taker net > -2 bp/day	net ≤ 0; otherwise inconclusive

3.8 Why a strong backtest is not enough

Besides multiple testing [4], backtests mislead through peeking, overlapping returns, survivor-only universes, look-ahead and fills at prices nobody could trade, and this program met each. In round 1 about 40% of the roughly 1,200 designed cells were already decidable from earlier looks and were marked as peeked. A funding-settlement drift, the best raw result of the program at $+65$ to $+104$ bp per event, was matched by a clock-matched placebo (t 5.39): it selected on a print known only afterwards.

4 What failed

Figure 2 shows where the 154 registered and 17 charged trials went. Microstructure and events took the most trials, and microstructure also took 759 of the 1,145 diagnostic cells, mostly a census of order-book features. Pairs, arbitrage and lead-lag have no registered trials: their diagnostics showed costs above any gross edge, so the gates that would have released trial cells never opened. Carry, about 130 variants, was searched before the ledger existed and sits inside the 1,700 estimate.

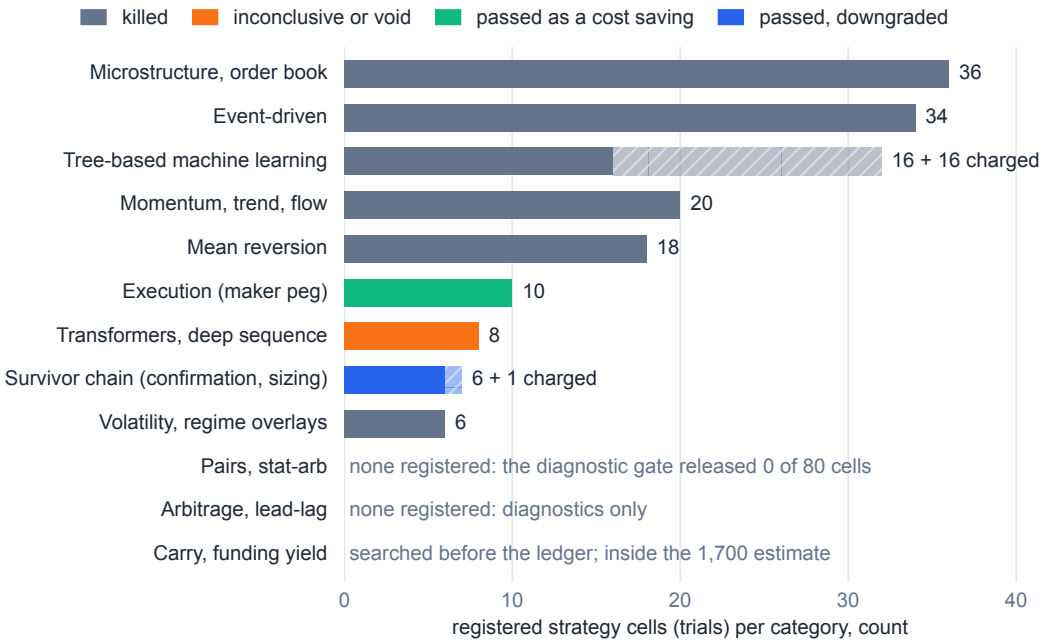


Figure 2: Registered strategy cells (trials) per category in the ledger of 21 October 2026, coloured by the category's verdict. Hatched: diagnostic cells charged as trials. The estimated 1,700 earlier configurations are not drawn.

1. **Transformers.** The cross-sectional transformer had real ranking skill, an information coefficient (IC, the daily rank correlation of forecast and outcome) of 0.075 at t 14 at 24 hours, but netted only $+5.03$ bp per trade (t 0.955); it is void because its shuffle control kept skill, and a plain gradient-boosted model did as well on identical rows.

Table 3: Best cell per category, net of the category's own costs on the frozen table. Net is per day of fixed capital for books and per trade, basket, event or order otherwise; units differ and the t-statistics are not comparable across rows. Event holdout: January 2025 to September 2026.

#	Category	Best cell; sample	Net, bp	t	Verdict
1	Transformers	cross-sectional transformer, 24 h; dev 2023–25	$+5.03$/trade	0.95	inconclusive, void
2	Tree-based ML	lower-turnover weekly book; 2026 holdout	$+21.4$/day	1.90	killed by rule
3	Pairs, stat-arb	basket reversion gate cell; dev 2023–25	-0.25/basket	0.88	killed
4	Arbitrage, lead-lag	SOL Binance–Bybit latency arbitrage; 21 days	-6.3/trade	n/a	killed
5	Mean reversion	open-interest flush fade; event holdout	-8.07/event	-0.35	killed
6	Momentum, flow	daily cross-sectional flow book; Binance 2020 to mid-2026	$+14.6$/day	3.57	killed, DSR 0.488
7	Carry	BTC and ETH cash-and-carry; 2022–2026	$+0.49$/day	n/a	killed as alpha
8	Order book	book imbalance, 1 to 10 s; Sep 2026	≈ -3.0/trade	22.8 (IC)	killed
9	Events	funding-settlement drift; event holdout	$+103.6$/event	3.82	killed, placebo
10	Vol. overlays	volatility-forecast scaling; dev	ΔSharpe < 0	n/a	killed
11	Maker execution	peg-to-touch, BTC on Bybit; Sep 2026	$+2.5$/order saved	88.5	pass, cost saving
S	Surviving book	construction not disclosed; 2026 holdout	$+23.0$/day	2.61	pass, downgraded

The order-book transformer never beat 2.8 bp gross per fill against a 4.0 bp cost. Why it failed: the fee, not the model class.

2. **Tree-based machine learning.** At 4 and 24 hours, gradient-boosted trees grossed 4.5 and about 22 bp per trade against 8.35 to 16.4 bp of cost. A lower-turnover book with daily decisions and positions held for about a week came closest: $+9.75$ bp/day in development (HAC t 2.89), then $+21.4$ bp/day on the 2026 holdout with HAC t 1.90 against a bar of 2.0, killed by rule. A Bybit-trained variant lost money in development and was void under its leak control (correction 3, Section 7). Why it failed: costs at short horizons, and a t 0.10 short of the bar at a week.
3. **Pairs and statistical arbitrage.** A diagnostic gate had to show reversion larger than cost before any of 80 planned cells was registered; it released none (best: 16.2 bp gross against 16.4 of cost). Why it failed: 5 to 16 bp of reversion cannot pay 16 to 51 bp round trips on two or more legs.
4. **Arbitrage and lead-lag.** Binance does lead Bybit, and BTC leads other coins, by 100 to 500 milliseconds, but after a 300-millisecond round trip only 0.2 to 1.5 bp of the move is left against 13 to 18 bp of cost; latency arbitrage lost 6.3 to 10.5 bp per trade. Why it failed: the gap lasts less than a retail round trip.
5. **Mean reversion.** The open-interest flush fade made $+43.8$ bp per event in development and -8.07 on its holdout (t -0.35). Why it failed: forced-flow spikes do not revert on average, and an event costs 20 to 37 bp.
6. **Momentum, trend and flow.** The strongest single signal of the program, a daily cross-sectional flow book, netted $+14.6$ bp/day with HAC t 3.57 from 2020 to mid-2026, but it was killed by its registered DSR bar (0.4876 against 0.5); its Sharpe ratio without its top 5 coins was 0.49, and trend on the majors fell to a Sharpe ratio of 0.29 on Bybit. Why it failed: deflation, concentration and venue transfer.
7. **Carry.** About 130 variants were searched before the ledger; the best, a cash-and-carry on BTC and ETH, earned 1.8% a year since 2022 and 0.4% a year in 2025–2026, and reverse carry made 58% of its gains from 27 squeezes. Why it failed as alpha: it yields about cash.
8. **Order book.** A 528-cell census of book and flow imbalance from 1 to 300 seconds found real signals (IC t 22.8 for BTC at one second) but predicted moves of only 0.4 to 2.7 bp against a 4 bp maker-maker floor, and following or fading large sweeps, the category's 36 registered trials, grossed only 0.24 to 0.87 bp against a 25.8 bp cost line. Why it failed: right about the next tick, but the tick is smaller than the fee.
9. **Events.** The funding-settlement drift produced the largest raw numbers of the program, $+103.6$ bp per event on its holdout (t 3.82), but a clock-matched placebo showed the same drift (t 5.39). Shorting new listings lost 270 bp per event. Why it failed: look-ahead artefacts and unbounded squeeze tails.
10. **Volatility overlays.** A volatility forecaster beat an exponential average by 22 to 25%, yet scaling a book by it changed the Sharpe ratio by -0.03 to -0.16. Why it failed: scaling cuts the sudden best weeks that make a book's return.
11. **Maker execution.** Not a strategy but a cheaper way to place orders: a peg-to-touch executor saved 2.36 to 2.56 bp per BTC order on Bybit against taker execution, simulated on 10,080 orders from 1 to 21 September 2026 (development data; its August holdout was never opened). Why it is not alpha: it needs an edge to apply to.

Only the week pays. Predictability is real at every horizon; what changes is whether the predicted move exceeds the cost of trading it. Table 4 follows one model family from 4 hours to 7 days, and Figure 3 sets the best gross moves of the short-horizon families against their costs.

Table 4: One model family by horizon, development 2023–2025. Upper rows: bp per trade, cost per round trip. Lower rows: bp per day of capital for a daily netted dollar-neutral book, medians over the 8 CPCV paths, funding paid by the book.

Horizon	Construction	Gross	Cost	Funding	Net	HAC t
4 hours	hourly decisions, per trade	4.5	8.35–16.4		−9.3	−4.13
24 hours	hourly decisions, per trade	≈ 22	8.35–16.4		+8.79	1.33
72 hours	daily book, turnover 0.395 of capital a day	10.2	3.1	4.0	+2.89	0.64
7 days	daily book, turnover 0.183 of capital a day	14.0	1.44	2.75	+9.75	2.89

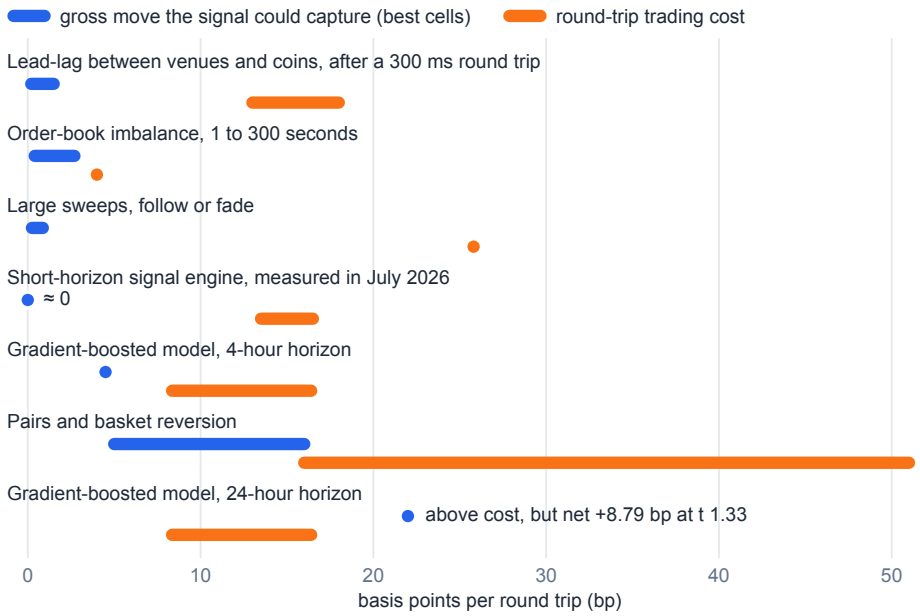


Figure 3: Best gross move per round trip (blue) against the cost of trading it (orange), short-horizon families and model cells. Only the 24-hour model beat its cost, and not significantly.

Not tested. A negative result is only as wide as its search. Not covered: the two largest registered model panels at 4 and 24 hours (not run for compute); cost-aware training objectives; targets such as funding or realised volatility; model classes beyond gradient-boosted trees, a flat multilayer perceptron and the two transformers; news, on-chain or options data; and every venue other than Binance and Bybit.

Takeaway. No category produced a standalone edge that cleared its own rules. Ranking skill is real at every horizon but pays for itself only at about a week; at shorter horizons fees take all of it.

5 The surviving book: backtest results

5.1 What is disclosed

The surviving book is a daily-rebalanced, dollar-neutral, cross-sectional long/short book on the point-in-time Binance universe of Section 2. It holds each position for about a week, and its long and short notional are equal. On an average development day it holds about 95 coins after netting, about 18% of capital changes hands, and its beta to BTC is slightly negative (−0.08). Its signals, inputs, weights, filters and execution rules are not disclosed in this edition. Returns are bp of fixed capital per day at 1x gross exposure, not compounded, net of fees, slippage and funding. Development figures use CPCV path 6, the ledger's median path, and are subject to correction 1 in Section 7.

5.2 Main results

Over the three development years the book adds up to +122% of capital, not compounded. Over the 2026 holdout it adds +62%, of which +8% in the clean slice. It passed its registered holdout bar; Section 5.9 explains why the pass is downgraded.

Table 5: The surviving book. Net return in bp/day on the realistic maker-taker line and the all-taker line, Newey–West HAC t with 10 lags, annualised Sharpe ratio and maximum drawdown in % of capital. Verdict against the registered bar (Table 2).

Sample	Days	Net	HAC t	Sharpe	Max DD	All-taker net (t)
Development, 2023–2025	1,095	11.1	4.46	2.44	11.3%	10.9 (4.33)
Holdout, 1 Jan to 29 Sep 2026, one look	272	23.0	2.61	3.45	7.9%	22.1 (2.53)
Clean slice, Jul to Sep 2026	91	9.1	0.90	1.52	–	8.4

5.3 By year and by month

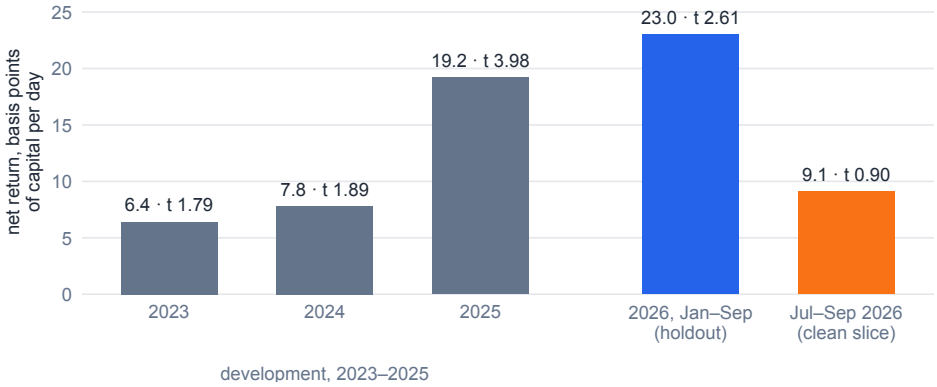


Figure 4: Net return by period, bp of capital per day, realistic line, with HAC t. 2023–2025: development, CPCV path 6. 2026: holdout, January to September, and its clean slice, July to September.

2023 and 2024 are individually weak: $+6.4$ and $+7.8$ bp/day at single-year HAC t of 1.79 and 1.89, against $+19.2$ bp/day at t 3.98 in 2025 (Figure 4). Most of the edge came from 2025 and 2026. By month, the book was positive in 26 of 36 development months, 7 of 9 holdout months and 11 of 12 development quarters, the exception being the first quarter of 2023 at -7.0 bp/day. Its 2026 monthly means, January to September and rounded, were -14, $+60$, $+34$, $+6$, $+65$, $+30$, $+14$, -2 and $+15$ bp/day. It did better in the 13 development months when BTC fell, $+14.0$ bp/day, than in the 23 when it rose, $+9.5$ bp/day, in line with its mildly negative beta.

5.4 Where the return comes from

Fees and slippage cost 1.41 to 1.48 bp/day in both samples, because only about 18% of capital trades a day. Both legs earn: long 5.9 and short 5.2 bp/day in development, 12.8 and 10.1 bp/day in 2026. Funding is close to neutral for the book as a whole: it received 0.39 bp/day in development and paid 0.61 bp/day in 2026. The all-taker line is within 1 bp/day of the realistic line in both samples, so the result does not hinge on the maker fill model.

5.5 Concentration

In 2026 the top 5 coins made 67% of the net return and the top 10 made 92%; in development the top 5 made 25%. Small coins carry the book: tier 3 contributed 17.5 of 23.0 bp/day in 2026 and 9.7 of 11.1 bp/day in development, BTC and ETH together 0.16 and 0.57 bp/day. The best 20 days made 101% of the 2026 net return, so zeroing them leaves -0.2 bp/day; in development the best 30 days made 63%. This is the shape of a book that harvests squeezes and dispersion in small coins: a slow drift punctuated by a few large days. A year without such episodes is plausibly flat.

5.6 Drawdowns and tails

The development maximum drawdown was 11.3% (15 February 2023), recovered 80 days after its trough; in 2026 it was 7.9% (2 February 2026). In development the worst day was -5.9% of capital, the best $+7.2\%$, the worst 30 days -9.4%, and the longest spell below a previous high 125 days; daily skewness was 0.17. These are daily closes. On 10 October 2025, a day the book (then at 0.89x gross) closed up 3.1%, taking every position at its adverse 5-minute extreme in the same bar gives a worst-case bound of -14.5%, and every position's worst move of the day at once -27.7%. Intraday risk runs 2 to 4.6 times the risk visible in daily closes.

5.7 Trying to break it

Each stress test reruns the same book on development 2023–2025 with one change. A test counts as survived when the stressed book still nets more than zero at HAC $t \geq 2$; this is my convention, not a registered bar.

Table 6: The surviving book under stress, development 2023–2025, realistic line unless stated. Base: 11.1 bp/day at HAC t 4.46. Realistic bundle: costs $\times 1.5$, a one-day-stale ranking, taker entries an hour late and square-root impact at $5,000.

Test	Net	HAC t	Result	What it tests
All-taker line	10.9	4.33	survives	the maker fill model
Costs $\times 2$	9.67	3.88	survives	cost model too optimistic
Costs $\times 3$	8.22	3.29	survives	cost model far too optimistic
Costs $\times 3$, all-taker	6.02	2.38	survives	both at once
Ranking one day old	11.2	4.18	survives	data or compute delay
Taker entries an hour late	10.3	4.15	survives	no maker fills, one hour late
Stale ranking and late entries	10.5	3.94	survives	both delays
Funding off	10.7	4.40	survives	whether funding drives the result
Short-side funding $\times 2$	10.4	4.17	survives	worse squeeze funding
Top 5 coins banned	9.45	4.10	survives	a few names drive it
Coins younger than 180 days out	7.85	3.59	survives	new-listing effects
Realistic bundle	9.28	3.47	survives	several frictions at once
Best 10 days removed	8.01	3.48	survives	a few days drive it
Top 50 coins only	5.92	1.89	breaks	dependence on smaller coins
No tier-3 coins	-2.02	-0.49	breaks	dependence on small coins
Best 30 days removed	4.08	1.73	breaks	dependence on rare days

The book is robust to how it trades and fragile in what it trades. Without tier-3 coins only about 17 coins a day remain and the book loses money; top 50 coins only leaves 5.92 bp/day at t 1.89 with a 19.3% drawdown; without its best 30 days, 2.7% of development days, it keeps about 4.1 bp/day at t 1.73. The 2026 checks, post-selection because the window had already been read, show the same pattern: removing the 3, 5 or 10 best coins leaves 11.5, 7.5 or 1.7 bp/day, and dropping the best 10 days leaves 9.0 bp/day at t 1.37. No 2026 tier or cost stress was run. If real fills in the smallest coins are a few bp worse than modelled, or a year passes without large dispersed moves, the edge shrinks fast. Market impact does not bind: with square-root impact the book still nets 10.9 bp/day at t 4.36 at $500,000 of capital.

5.8 Is the ranking real?

Score placebo. Shuffling the ranking between coins every day keeps everything except the information in the ranking. The structure plus costs and funding then loses about 2.8 bp/day on average, the best of 100 placebo runs makes -0.19 bp/day, and the real book lies above all 100 runs at z 12.7. This z is against one specification; correcting for the number tried is the deflated Sharpe ratio's job. **Leak control.** Retrained on labels permuted across coins within each day, the model behind the book has no skill on Binance: median IC $+0.006$ at 72 hours and -0.020 at 168 hours, inside bands of two standard errors (0.015 and 0.026), against 0.131 for the real model. The literal Brier-skill limit of 0.005 fails (0.0094 and 0.0093), but the excess is day-level class mix and the cross-sectional part is negative. This control was chosen after the originally registered one failed (Section 3.5).

5.9 The verdict on the pass

The book passed its registered holdout bar. A separate adversarial review, which I ran, graded the pass *downgraded* for four reasons:

1. part of its design had already been evaluated on this window, January loss included, when the book was registered, and a related signal from an earlier, killed book had been seen on the first half of 2026, so a hedge for January was choosable in hindsight;
2. on the clean third quarter it made $+9.1$ bp/day at HAC t 0.90, while the bar only asked for a positive number;
3. the top 5 coins made 67% of its 2026 net return and its best 20 days made 101%;
4. its deflated Sharpe ratio is 0.0047 at the 1,866 trials counted at confirmation, so it does not survive deflation for the full search; this was information, not a bar (Section 3.4).

The review found no hard leak: HAC t stays between 2.50 and 2.98 for lags 0 to 60, and dropping any one month leaves t between 2.13 and 2.99.

6 Verdict and limitations

No standalone edge was found. One book passed one look at the 2026 window, a look informed by an earlier result on the same window, and it was graded pass, downgraded. It rests on 2025–2026 and on small coins, and it is not a tradable result.

- **One venue.** All evidence for the book is on Binance data. It has not been tested on any other venue, and a Bybit-trained variant failed in development.
- **No forward evidence.** Every number in this paper is a backtest. The stress tests reuse the same development days, so they are not separate confirmations.
- **Selection.** The book was chosen from a counted search of 1,871 trials and does not survive deflation for it; the only clean test is new data, registered in advance.
- **Regime.** 2023 and 2024 were individually weak; most of the edge came from 2025 and 2026.
- **Small coins and rare days.** The result breaks without tier-3 coins or without its best 30 development days, and small-coin depth and slippage are coarsely modelled.
- **Costs.** Maker adverse selection is costed at zero and some fees come from secondary sources, although the all-taker line is within 1 bp/day and triple costs still leave 8.22 bp/day.
- **Data.** Order rules are a 21 October 2026 snapshot, the tier lists are static, and the 2026 window has the limits listed in Section 2.
- **Tails.** Daily closes understate intraday risk by a factor of 2 to 4.6 (Section 5.6).

7 Corrections since the 21 October edition

A review on 30 October found three errors in the 21 October working paper. This edition carries them as follows.

1. **Development scores are not a walk-forward.** The development-period model scores came from CPCV path 6, whose models trained partly on later data. Refitting quarterly on finished labels only lowered the model's rank IC from 0.125 to 0.114. In a re-run of a related, reduced book on Bybit prices, the refits lowered the mean by about 2.3 bp/day over January 2023 to September 2026 and by 2.9 bp/day over 2023–2025 (20.7 to 17.8), with 2026 unchanged. The development figures in Section 5 were not re-run and are likely optimistic. The 2026 holdout is unaffected: its model was trained on development data only.
2. **The Bybit hourly store is not survivorship-free.** A paging bug kept only the last 998 hours of 149 of 307 delisted contracts. The Bybit results in Section 4 carry that bias; the Binance results, including every number for the surviving book, do not use this store.
3. **A Bybit loss was misattributed.** The first edition said the model book lost 12.2 bp/day on Bybit because its shorts paid funding. That figure belongs to a different, Bybit-trained, model-only book. It reproduces exactly, but on repaired data that book loses 9.74 bp/day, and the funding it paid was a symptom, not the cause.

References

[1] M. López de Prado. *Advances in Financial Machine Learning*. Wiley, 2018.

[2] D. H. Bailey and M. López de Prado. The deflated Sharpe ratio: correcting for selection bias, backtest overfitting, and non-normality. *Journal of Portfolio Management* 40(5):94–107, 2014.

[3] D. H. Bailey, J. Borwein, M. López de Prado and Q. J. Zhu. The probability of backtest overfitting. *Journal of Computational Finance* 20(4):39–69, 2017.

[4] C. R. Harvey, Y. Liu and H. Zhu. … and the cross-section of expected returns. *Review of Financial Studies* 29(1):5–68, 2016.

[5] W. K. Newey and K. D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3):703–708, 1987.