

Shock Engine

Strategy, execution and capacity of a systematic intraday strategy on Bitcoin and Ether

Eight years of simulated trading, all 1,150 executions since 2020 replayed on real market data, and an honest account of what remains to be proven.

Oreste Raynaud • ShadowMarketPro Research • 13 October 2026

2.02

Sharpe ratio

simulated, 2018–2026

–17.8%

Maximum drawdown

simulated, 2018–2026

1.92

Net Sharpe at \$1M

2020–2026, measured costs

> \$5M

Capacity at Sharpe 1.5

Binance, tested to \$5M

Status of the evidence

In-sample historical simulation after commissions, with execution costs measured on real Binance market data. The two entry conditions were set in October 2026 with knowledge of the data. A pre-registered forward test of the short-entry condition started on 12 October 2026. **There is no live track record on real capital yet.**

What is inside

The strategy and how it was tested; a framework for executing large notional drawn from the market-impact literature; the 1,150 executions; costs, capacity and forward-looking estimates; what we got wrong; risks.

Historical simulation after modelled and measured transaction costs. Not live performance. Simulated results have inherent limitations and benefit from hindsight in the design of the rules. Past simulated performance does not predict future results. This document is neither investment advice nor an offer of any financial product or service. Proprietary thresholds, window lengths and exit multiples are not disclosed; no test reported here depends on knowing them.

A note before the numbers

Every systematic strategy begins as a story the data seems to tell. Shock Engine's story is simple: when Bitcoin or Ether moves abnormally within fifteen minutes, and the move is confirmed by its surroundings, it tends to keep going for a few hours. The backtest says this story has paid well for eight years.

We have learned to distrust that sentence. A backtest is a hypothesis about the past, written by someone who already knew how the past ended. So this dossier is organised around the questions an allocator asks *after* reading the headline Sharpe ratio. Would the strategy have survived real execution? How much capital can it carry before its own orders eat the edge? What would prove it wrong, and when will we know?

We tried to answer them the way a careful trading desk would. For the validation studies, we wrote each test down before running it. We measured with market data rather than assumptions, and when a convenient model and a measurement disagreed, we kept the measurement. We tried hard to break our own result. And we say plainly, throughout, which numbers are *demonstrated*, which are *estimated*, and which only live trading can settle.

We were wrong several times along the way: about a three-second wait nobody had measured, about how much market impact to expect, about the cost of trading Ether, and even about a figure in the first draft of this very document. Those corrections are collected in Section 8, because how a team handles its mistakes says more about its process than any single figure.

What this document does not contain is a live track record on real capital. That is the next chapter. It will be reported from its own start date, separately from the simulation, whatever it shows.

Oreste Raynaud

ShadowMarketPro Research, 13 October 2026

Contents

A note before the numbers	1
At a glance	2
1 How we worked: the scientific approach	3
2 The strategy	5
3 Executing large notional: what the literature tells us	9
4 The 1,150 executions	10
5 Execution costs: measurement and control	12
6 Capacity	15
7 Forward-looking results	16
8 What we got wrong, and what it taught us	19
9 Risks and limitations	20
10 What would reduce the uncertainty most	21
A Definitions	21
B Sources and reproducibility	22
C Version history	22

At a glance

Shock Engine is a systematic intraday strategy on Bitcoin and Ether. Over 1 September 2018 to 30 September 2026 its 50/50 portfolio shows, in simulation, a Sharpe ratio of 2.02 and a maximum drawdown of -17.8% , against 0.78 and -76.9% for buy-and-hold. The result survives placebos, delays, parameter perturbations, a transfer to Ether and a regression on standard trend strategies. Replaying all 1,150 historical executions on real market data, with the bot's execution path as it is now coded, the net Sharpe at \$1M of capital over 2020–2026 is estimated at 1.92 (2.00 before execution costs) and stays above 1.5 at every capital tested on Binance, up to \$5M. The evidence remains in-sample, including the two entry conditions, which were set in October 2026 with knowledge of the data; the intended venue (Hyperliquid) has thinner books than Binance; and there is no live track record yet.

Demonstrated (data, reproducible)	Estimated (model, conditional)	Open (needs live trading)
The historical edge is not drift, exit shape or a knife-edge parameter: random entries lose to it 100% of the time, every perturbed neighbour is profitable.	Net Sharpe at \$1M: 1.92 [1.40; 2.41], CAGR 56.3%, funding excluded.	Persistence out of sample. A walk-forward on Bitcoin keeps a Sharpe of 0.83 (median 0.97 over 12 longer window designs), against 1.55 for the fixed rules.
It is not repackaged trend following: residual alpha 44.8% a year, $t = 4.94$, against six standard trend strategies (in-sample).	Capacity at Sharpe 1.5: above \$5M on Binance, the largest capital tested (central estimate, wide interval).	The two entry conditions (calm regime, bearish trend for shorts), in-sample like every rule; selection risk of the choice 19%. The short condition has a pre-registered forward test, about three years.
All 1,150 executions replayed on real trades: 91.7% demonstrably fillable at \$1M, none shown unfillable.	On Hyperliquid: above \$2.5M if its books are half as deep as Binance's. Bitcoin depth beyond its 20 visible levels is unknown.	Our own market impact; Hyperliquid depth beyond 20 levels; the bot's real latency on its server.
Each second of delay costs money: 0.4 bps at 0.5 s, 4.4 bps at 5 s after the signal.	Cost of the bot as now coded (not yet deployed or traded) at \$1M: 2.27 bps per execution, 15% of the budget compatible with a Sharpe of 1.5.	A live track record. One year cannot reliably tell a Sharpe of 1.5 from zero (Section 7).

How to read the numbers in this dossier

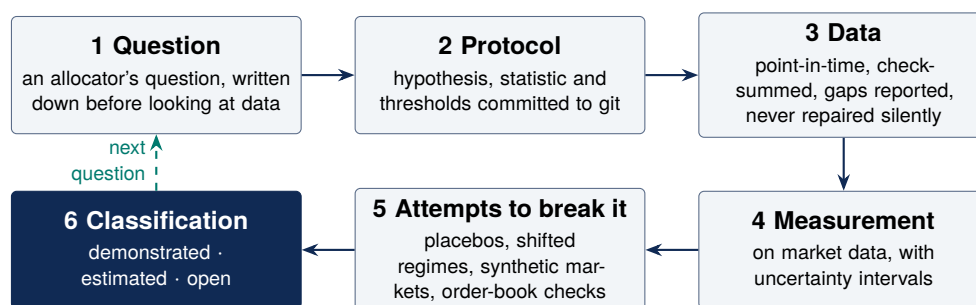
Simulated means a historical replay of fixed rules. **Measured** means observed in historical market data (real trades and order books), not orders sent by Shock Engine. **Estimated** means a model fitted to those measurements. Strategy figures cover 2018–2026; execution figures cover 2020–2026, the period with Binance perpetual futures. All costs are in basis points (bps) of the traded notional; Sharpe ratios are annualised from daily returns unless stated otherwise.

1 How we worked: the scientific approach

“Why should I believe any of this?”

The headline numbers are strong, which is precisely why they deserve suspicion. Searching many rules on one history inflates in-sample performance (White, 2000; Bailey and López de Prado, 2014; Harvey et al., 2016). We cannot remove that bias after the fact, but we can bound it, expose it and test around it. Six working rules shaped everything that follows.

- 1. Write the test before running it.** The validation studies of the last phase were pre-registered (Table 2): hypothesis, statistic, thresholds and data were committed to version control before the first computation, in the tradition of clinical trials (Nosek et al., 2018). The commit is the timestamp. The earlier robustness tests of Table 4 used criteria set during the original research, without a committed protocol.
- 2. Try to break the result, not to confirm it.** Random-entry placebos, delayed entries, shifted regimes and synthetic markets ask whether something with no edge could have done as well.
- 3. Measure before modelling.** When a measurement and a convenient model disagreed, we kept the measurement and reported the gap. Section 8 shows how often this mattered.
- 4. Never use the future.** Every simulated decision uses only information available at that instant: closed candles, the order book as it stood, the previous day’s regime.
- 5. Separate demonstrated, estimated and open.** A number’s status travels with it, all the way to the summary.
- 6. Publish failures and corrections.** Markets where the rules fail, tests that came out against us, and our own errors are reported next to the successes.



Any deviation from a protocol is logged in its report, with the reason.

Figure 1. The research cycle behind every study in this dossier.

The evidence ladder. Not all evidence is equal. Table 1 ranks what each kind of test can rule out and where Shock Engine stands on each rung. The honest summary is that the strategy has climbed the rungs that history allows and is now on the ones that only time can climb.

Table 1. The evidence ladder: what each kind of test rules out, and where Shock Engine stands.

Rung	Evidence	What it rules out	Shock Engine
1	In-sample simulation after costs	Nothing on its own	Sharpe 2.02, 2018–2026
2	Placebos, delays, parameter neighbourhoods	Market drift, exit artefacts, knife-edge parameters	Passed (Table 4)
3	Transfer to markets never used for design	Fitting the noise of one price series	Ether passed on two feeds; Solana and gold partial; TAO failed
4	Walk-forward re-selection	Look-ahead in the choice of parameters	Bitcoin: 0.83 against 1.55
5	Execution replay on real market data	The illusion of frictionless fills	1,150 executions, Sections 4–6
6	Pre-registered forward test	Hindsight in the rules themselves	Short condition: running since 12 Oct. 2026
7	Live track record on real capital	All of the above, plus our own impact	Not yet available

Pre-registered studies. Pre-registration was adopted for the validation phase. The strategy rules themselves were developed without it, which is why every historical performance figure in this dossier is treated as in-sample. Table 2 lists the protocols; their commit hashes are in Appendix B. When the entry rules were restricted to the calm volatility regime on 13 October 2026, every study was run again on the published rules with its protocol unchanged; the outcomes below are those of the published rules.

Table 2. Pre-registered studies, in the order their protocols were committed.

Committed (2026)	Study	Question	Outcome
8 Oct.	Short entries, 8 unseen assets	Do symmetric filters improve the short side?	Improvement not demonstrated ($p = 0.44$)
8 Oct.	Forward test of the short-entry condition	Do shorts in a bearish regime beat the others?	Running since 12 Oct. 2026
8 Oct.	Historical diagnostics of the short-entry condition	Can past data eliminate the condition?	Not eliminated; not a validation
9 Oct.	Falsification of the short-entry condition (tests A–C)	Is its historical gain exceptional?	Short quality: yes. Sharpe gain: largely generic
9 Oct.	Residual alpha	Is it repackaged trend following?	No: alpha 44.8%/yr, $t = 4.94$ (in-sample)
12 Oct.	Capacity, phases P0–P1	What cost can the strategy bear?	Budget of 14.7 bps per execution at Sharpe 1.5
12 Oct.	Latency drift	What does each second of delay cost?	4.4 bps at 5 s, 0.4 bps at 0.5 s
12 Oct.	Executability, \$500k and \$1M	Could every order have been filled, and at what price?	96.8% and 91.7% demonstrated; none shown unfillable
13 Oct.	Selection risk of the calm-regime restriction	How much of its advantage is the choice among the variants tried?	Overfitting probability 19%; gain not significant after deflation (18%)

2 The strategy

“What exactly is being traded, and does the edge survive scrutiny?”

2.1 Economic intuition

Two views compete about what follows a large price move. Under *underreaction*, information is absorbed gradually and prices drift in the direction of the move (Hong and Stein, 1999). Under *overreaction* or liquidity provision, the move is partly a temporary concession that reverts. Which dominates is an empirical question, and the answer may depend on the state of the market. Shock Engine bets on continuation, but only after the move has been qualified: shocks are measured against local volatility, in the spirit of nonparametric jump tests (Lee and Mykland, 2008), and their clustering reflects the self-exciting dynamics documented for jumps. Entries are conditioned on the market state because momentum payoffs are state dependent and crash in rebounds (Daniel and Moskowitz, 2016): new positions are opened only in calm volatility regimes, and shorts only in bearish trends.

2.2 The rules, in words

- ▶ **Signal.** A 15-minute return that is extreme relative to local volatility and breaks the recent range.
- ▶ **Qualification.** Higher-timeframe trend on the right side, a candle that closes near its extreme, abnormal volume, shocks that cluster, and a calm enough volatility state before the shock.
- ▶ **Regime.** New positions are opened only when the daily volatility regime, read from the last closed day, is calm. No position is opened in agitated markets; a position already open runs to its normal exit.
- ▶ **Exits.** A volatility-scaled initial stop, a wide trailing stop that rarely triggers, or a qualified shock in the opposite direction. No partial take-profit.
- ▶ **Short side.** A mirror of the long side, admitted only when the daily trend regime is bearish.
- ▶ **Exposure.** One position per market at a time, 100% of that market's capital, no leverage, no pyramiding, no averaging down. The portfolio holds half of the capital in each market and never rebalances.
- ▶ **Calibration.** Parameters set on Bitcoin only and applied unchanged to Ether; the two entry conditions were set with knowledge of both markets.

Design history: two entry conditions added with knowledge of the data

For most of the research, the strategy traded both volatility regimes with an unconditional short side mirroring the long side. The trend condition on shorts was adopted in October 2026 after a pre-registered study of short entries in which it was one of several candidate rules, not the primary hypothesis; the restriction to calm regimes was adopted on 13 October 2026 after an exploratory study of the agitated regime. Both were chosen with knowledge of the 2017–2026 Bitcoin and Ether data, so every historical figure in this dossier is in-sample with respect to them. Over the same period, the rules before both conditions show a Sharpe ratio of 1.73, a CAGR of 64.5% and a maximum drawdown of –23.5%; adding the trend condition alone gives 1.88. Across the 16 rule variants actually tried on the portfolio, the probability that the choice is overfitted is 19% (Bailey et al., 2017); the gain of the calm restriction is consistent over time but not significant once the number of variants is accounted for (deflated Sharpe ratio 18%; Bailey and López de Prado, 2014). Three pre-registered falsification tests (Table 4) find the improvement in the quality of shorts exceptional, while most of the trend condition's gain in Sharpe ratio is generic to removing shorts outside bearish regimes. The forward test of Section 7 tests the short condition out of sample; the calm restriction has no forward test of its own yet.

2.3 Historical results

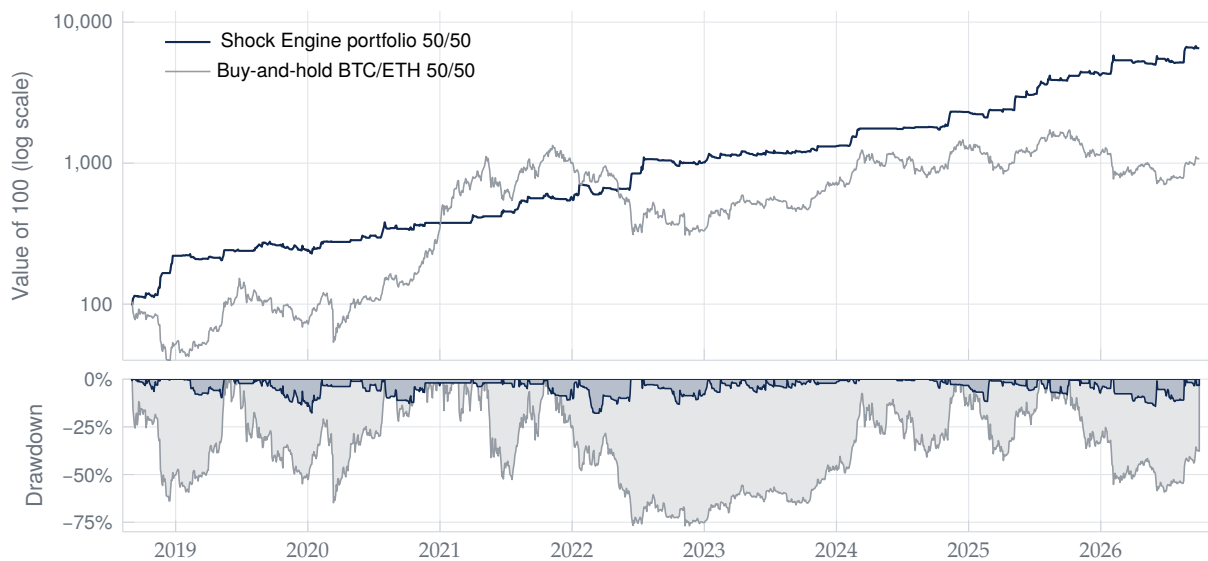


Figure 2. Value of 100 invested and drawdown from the running peak, daily closes, 1 September 2018 to 30 September 2026. Historical simulation after commissions; buy-and-hold without costs.

Table 3. Simulated performance, 1 September 2018 to 30 September 2026.

	Portfolio 50/50	Bitcoin	Ether	Buy-and-hold
Sharpe ratio	2.02	1.46	1.93	0.78
CAGR	67.7%	39.6%	80.0%	34.1%
Annualised volatility	27.4%	25.0%	33.3%	69.0%
Maximum drawdown	−17.8%	−24.8%	−24.1%	−76.9%
Sortino ratio	4.73	3.23	4.62	–
Calmar ratio	3.79	1.60	3.31	0.44
Trades (long / short)	716 (466 / 250)	358	358	–
Win rate; average win / average loss	31%; 5.58	30%; 5.18	33%; 5.52	–
Time in market	25%	15%	17%	100%

Daily UTC close-to-close returns, annualised with $\sqrt{365.25}$, zero risk-free rate. Commission 0.045% per order; no slippage, funding or leverage. Bitcoin and Ether columns are the two halves of the portfolio, each on its own capital.

The two strategies are much less correlated (0.36) than the coins themselves (0.82): each is out of the market most of the time. Split in two halves, the portfolio Sharpe ratio is 2.12 and then 1.98. The rolling one-year Sharpe ratio is positive in 100% of windows and above one in 93%. A month-block bootstrap gives a 90% interval of 1.53 to 2.51; a bootstrap resamples history and is not a forecast. The worst calendar month lost −13.9%.

The profile is convex: the strategy wins 31% of its trades, and its average winner is 5.58 times its average loser. The best 5% of trades account for 99% of the portfolio's log growth; without them the CAGR would be −0.3%. Missing a handful of large moves in live trading would change the outcome materially, which is why execution reliability, not only execution cost, matters so much here.

2.4 Robustness

Table 4. Tests designed to break the result, on the published rules. Placebo, delay and transfer criteria come from the original research and were applied unchanged; walk-forward and neighbourhood results are descriptive; the three tests of the short-entry condition were pre-registered.

Test	Result
Random entries with identical exits (200 runs)	The strategy beats 100% of runs on Bitcoin and on Ether (criterion: 95%).
Every entry delayed by one bar	97% (Bitcoin) and 94% (Ether) of the average trade kept.
Transfer without recalibration	Ether passes on two independent feeds (Sharpe 2.05 Binance, 1.82 Dukascopy). Other markets: Solana and gold partial; TAO failed; Sharpe ratios Solana (0.57), gold (0.39), TAO (0.01).
All parameters perturbed jointly by $\pm 20\%$ (300 neighbours)	Every neighbour profitable; median Sharpe 1.35 (Bitcoin), 1.94 (Ether). On Bitcoin the chosen set sits at the top of its neighbourhood, on Ether at the 76th percentile: a plateau, not a peak.
Walk-forward, Bitcoin, 28 quarterly windows 2020–2026	Sharpe 0.83 against 1.55 for the fixed rules; maximum drawdown –27.6%. A large part of the headline result on Bitcoin is a selection effect.
Walk-forward with longer windows (12 designs, pre-registered after the result above)	Calibration of 36 to 60 months or expanding, tests of 3 to 12 months, same test period: median Sharpe 0.97 (range 0.83 to 1.36), all 12 below the fixed rules. Window length does not explain the gap.
Selection among the 16 rule variants tried (12,870 splits of the history)	Probability of backtest overfitting 19%; gain of the calm restriction not significant after deflation (18%).
Regression on buy-and-hold and six trend strategies	Residual alpha 44.8% a year, Newey–West $t = 4.94$; 81% of the mean return unexplained (Newey and West, 1987).
Short condition, test A: random selection of shorts	Quality of shorts kept: 96.8th percentile ($p = 0.032$). Portfolio Sharpe: $p = 0.055$.
Short condition, test B: shifted regimes	$p = 0.005$ for the portfolio, 0.093 for Bitcoin alone, 0.001 for Ether.
Short condition, test C: synthetic markets	Short quality above the synthetic 95th percentile at every block length; the Sharpe gain only at the 43rd to 78th percentile.

Sharpe ratios in the transfer, neighbourhood and walk-forward rows are annualised from 15-minute returns and differ slightly from the daily figures. The tests of the short condition were run before the calm-regime restriction, on rules that also traded agitated-regime longs; every short entry comes from the calm regime under both rule sets, so the short-trade statistics carry over, while the portfolio-level p -values include those longs. The rules are not a universal law of markets: they work on the two most liquid crypto assets and not reliably elsewhere, and we see no reason to hide it.

3 Executing large notional: what the literature tells us

“A backtest fills at the close. What would a real \$1M order have paid?”

A backtest assumes that each order is filled at the closing price of the bar that triggered it. A real order arrives later, crosses a spread, consumes liquidity and may move the price for the next one. Before measuring anything we fixed the vocabulary, borrowed from four decades of research on institutional execution.

3.1 Implementation shortfall

Perold (1988) defined the cost of trading as the gap between a “paper” portfolio, executed at the decision price, and the real one. For an order of side $s = \pm 1$, decided at price P_d (here the close used by the backtest), arriving at t_a when the mid-price is m_{t_a} and the best quote on our side is q_{t_a} , and filled at an average price $\bar{P}$:

$$IS = \underbrace{s \frac{m_{t_a} - P_d}{P_d}}_{\text{delay: drift before arrival}} + \underbrace{s \frac{q_{t_a} - m_{t_a}}{P_d}}_{\text{half-spread}} + \underbrace{s \frac{\bar{P} - q_{t_a}}{P_d}}_{\text{depth: walking the book}}. \quad (1)$$

Every cost in Sections 4–6 is measured in these three parts, in basis points, separately from exchange fees, which the backtest already pays once per order.

3.2 Urgency against impact

Almgren and Chriss (2000) framed optimal execution as a trade-off: trading slowly reduces impact but exposes the order to price risk. When the order is driven by a short-lived signal, the price risk has a sign: the market moves *with* the signal, and waiting costs the edge itself (Cartea et al., 2015). For a schedule of duration T , the expected cost is roughly

$$\mathbb{E}[C(T)] \approx D(T) + I(Q, T), \quad (2)$$

where D , the drift during execution, grows with T and I , the impact of quantity Q , shrinks with T . Shock Engine sits at the urgent extreme. At its signal times the price moves against a waiting order by 0.4bps in half a second and 4.4bps in five seconds (Figure 5); spreading the order over the next minute costs 3.9bps (Bitcoin) and 9.2bps (Ether) of drift. Taking the visible book at once costs a fraction of that, at the median, up to \$2.5M orders on the rebuilt books (Section 5). The optimal schedule collapses to a single immediate order, and engineering latency, not order slicing, is where the money is.

3.3 The square-root law, and why it overstated our costs

The best-documented regularity in market impact is the square-root law: a metaorder of size Q , executed incrementally, moves the price by about

$$I(Q) \approx Y \sigma_D \sqrt{Q/V_D}, \quad (3)$$

where σ_D is daily volatility, V_D daily volume and Y a constant of order one (Tóth et al., 2011; Bouchaud et al., 2018), including on Bitcoin (Donier and Bonart, 2015). Estimated exponents vary: about 0.6 on US equities in Almgren et al. (2005), a logarithmic shape in Zarinelli et al. (2015). Our first capacity study, pre-registered, used Equation (3) with $Y \in \{0.5, 1, 2\}$ because no calibration was available. It implied about 2.5bps for a \$500k order in the central case.

The median measured depth cost of a \$500k order at our entry times was 0.24bps on Bitcoin and 0.60bps on Ether. The two numbers measure different things. The square-root law describes

metaorders spread over minutes to days, including the reaction of liquidity providers to repeated flow. A single marketable order that sweeps the book in milliseconds pays the shape of the book at that instant (Obizhaeva and Wang, 2013; Cont et al., 2014), and the books of Bitcoin and Ether perpetuals are deep near the touch. Frazzini et al. (2018), using a large manager's live executions, also found real costs well below model-based estimates. We take the measurement, with one caveat that the order-book check in Section 5 confirms: large real orders tend to arrive when liquidity is there, so a method built on them can flatter the cost of an order sent at a fixed time.

3.4 Capacity, two venues, and what remains unmeasured

- **Capacity.** Studies of anomalies measure how trading costs erode returns as size grows (Novy-Marx and Velikov, 2016; Landier et al., 2015; Frazzini et al., 2018). We define capacity as the capital at which the net Sharpe ratio falls below 1.5, a threshold fixed in the pre-registered protocol, not chosen after seeing the curve.
- **Two venues, one price?** Crypto prices can differ persistently across exchanges (Makarov and Schoar, 2020). Signals are computed on Bitstamp and Binance spot data, costs are measured on Binance perpetuals, and the bot trades Hyperliquid perpetuals. We measure the cross-venue gap separately from book quality.
- **Not measured.** Permanent impact and its decay (Gatheral, 2010; Bacry et al., 2015), and the reaction of market makers to a predictable flow at 15-minute closes. Only our own live fills can measure them.

4 The 1,150 executions

“Could every order have been filled, and at what price?”

Every execution of the portfolio from January 2020 to September 2026, without exception, was replayed on the real trades of Binance USD-M perpetual futures, the most liquid venue and the only one with free millisecond trade history covering the period. None is demonstrably unfillable at \$500k or \$1M of capital.

Table 5. The executions studied, January 2020 to September 2026.

Type	Bitcoin	Ether	Total	Order used by the bot
Entries	300	275	575	marketable limit (IOC)
Exits at the close (opposite shock)	103	119	222	marketable limit (IOC)
Stops and trailing stops	197	156	353	stop triggered by the exchange
Total	600	550	1,150	

By year: 485 executions in 2020–2022, 249 in 2023, 94 in 2024, 185 in 2025 and 137 from January to September 2026.

Data. 805 days of trades time-stamped to the millisecond from Binance's public archives, every file checked against its published SHA-256 checksum, none unreadable. The reference price was found for 1,149 of the 1,150 executions and, for all 353 stops, the moment the stop level was crossed. Binance's book-depth snapshots ($\pm 1\%$, since 2023) complement the trades, with snapshots frozen for more than 30 minutes excluded. For 9 executions falling on free data days, the full Binance order book was rebuilt to the millisecond (Section 5).

Method: what an order of our size actually paid. Rather than assume an impact model, we looked for real orders of our size around each execution (Figure 3). Aggressive orders were rebuilt from trades sharing a millisecond and a side. When a real order of at least Q dollars traded within two minutes of our arrival, the cost of its first Q dollars is what an order of our size really paid within two minutes of ours. When none did, the cost comes from a model fitted to these observations (log cost on size, volatility and volume; R^2 of 0.41 for Bitcoin and 0.32 for Ether; size elasticity 0.82 and 0.70).



Figure 3. Measuring execution cost from comparable real orders. Grey bars are aggressive orders rebuilt from Binance trades, with height proportional to size.

Table 6. Reconciliation of the 1,150 executions (entries, exits at the close, stops), initial bot.

Category, mutually exclusive	\$500k capital (\$250k orders)		\$1M capital (\$500k orders)	
	Count	Share	Count	Share
Comparable real order, first Q dollars within 50 bps	1,113	96.8%	1,054	91.7%
Depth within 1% only (book snapshots)	2	0.2%	8	0.7%
No direct evidence; cost estimated by the model	35	3.0%	88	7.7%
Total	1,150	100%	1,150	100%

50 bps is the price cap of the bot's market orders at entries and exits at the close; stops are exchange stop orders, capped at 5% from their trigger. Most executions without a comparable order are Ether (30 of 35 at \$500k; 63 of 88 at \$1M), mostly in 2020–2022, when real orders of that size were rare. At \$2M of capital, 77.8% of executions are demonstrated within 50 bps; at \$5M, none, because \$2.5M orders exceed the largest size observed directly (54.4% are demonstrated within 1% by the book snapshots). The rules take no partial profit, so there is no resting order to account for.

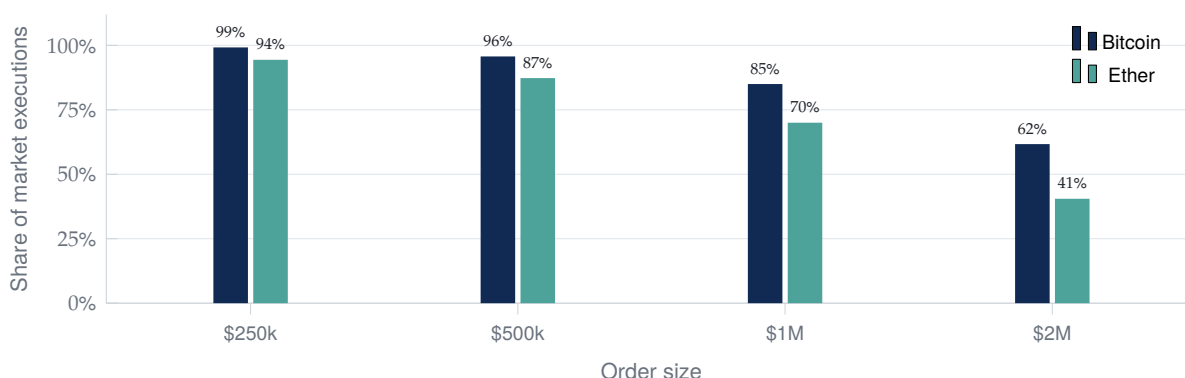


Figure 4. How often a real aggressive order of at least our size traded within two minutes of a Shock Engine execution, by order size. After our arrival, the first real \$500k order traded a median 5.7 s later on Bitcoin and 10.7 s later on Ether: the book absorbs our size at once; we would not have to wait for other trades.

5 Execution costs: measurement and control

“Where does the money go between the signal and the fill?”

The dominant cost is time, not size. With the bot as it is now built, a \$1M portfolio pays on average 2.27 bps per execution relative to the backtest price, 15% of the budget compatible with a Sharpe ratio of 1.5.

The cost budget. Adding one basis point of cost to every execution removes 0.034 from the Sharpe ratio. The ratio stays above 1.5 as long as the average added cost stays below 14.7 bps per execution (90% interval: 0 to 32.5 bps, because the history itself is uncertain), and above 1.0 up to 29.7 bps. Entries carry most of the sensitivity, then stops; exits at the close less.

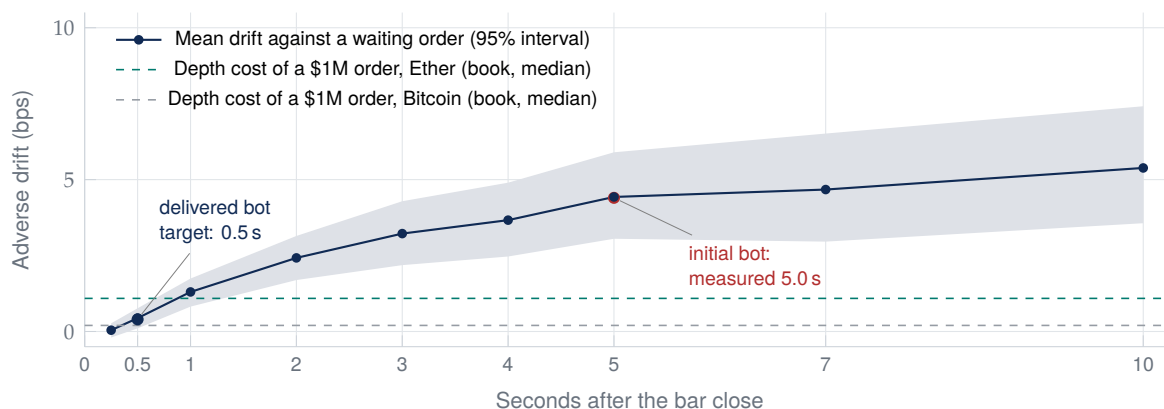


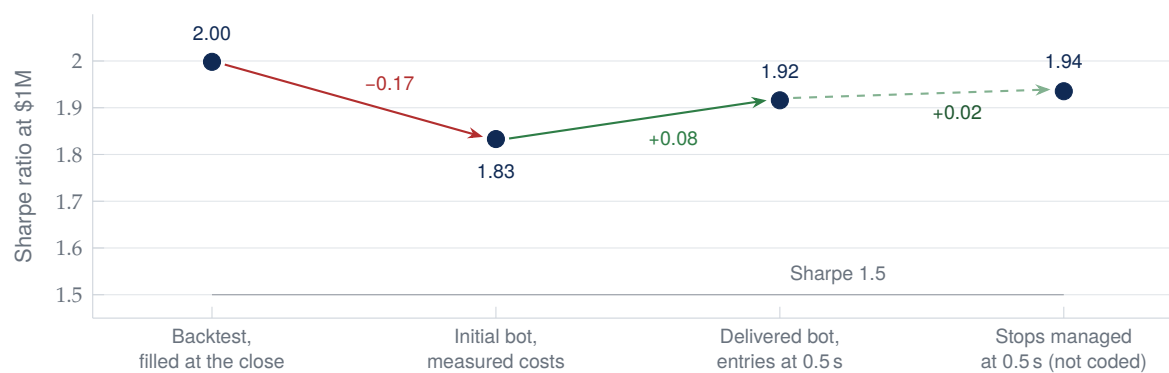
Figure 5. One second of delay costs more than sweeping a million dollars of book depth. Mean drift of the last traded price against the order after the signal’s bar close, 110 executions drawn at random and measured on Binance trades (pre-registered, bootstrap 95% interval). Depth costs: median over the rebuilt order books (two for Bitcoin, seven for Ether).

Latency, measured and engineered. The first strategy trade of the bot on Hyperliquid’s test network carries an exchange time stamp 5.01 s after the bar close. Reading the code explained it: a fixed three-second wait, then seven to nine network round trips made one after the other. Without touching the signals, the code was rewritten to read the closed candle as soon as the exchange opens the next one, to run its network calls in parallel and to keep connections warm, leaving two round trips between the close and the order. In the replay, this moves the Sharpe ratio at \$1M up by 0.08 [0.06; 0.10], positive in 100% of paired bootstrap draws. The rewrite was committed on 12 October 2026; it is not yet confirmed deployed and has not traded. The 0.5 s target is the optimistic end of the 0.5 to 0.9 s expected from a server in Europe (0.1 to 0.3 s from Tokyo); the new time stamps journalled by the bot will measure it. Throughout this dossier, “initial bot” and “delivered bot” describe only the execution path, the timing of orders, applied to the published rules: the bot’s own entry rules do not yet apply the calm-regime restriction (Table 12).

Table 7. Execution cost per execution relative to the backtest price, exchange fees excluded (bps; Binance, 2020–2026).

Capital, version of the bot	Median	P95	P99	Mean	Share of budget
\$500k, initial bot (order at $T + 5$ s)	2.64	20.5	39.4	4.25	29%
\$500k, delivered bot (entries at $T + 0.5$ s)	0.62	11.9	34.3	1.87	13%
\$1M, initial bot	2.81	21.9	45.9	4.65	32%
\$1M, delivered bot	0.81	13.1	35.0	2.27	15%

The delivered bot accelerates entries and exits at the close; stops remain native exchange stops, assumed to reach the book 3 s after the price crosses their level (Hyperliquid triggers them on the mark price, whose history we do not have). Costs cover the 1,149 executions with a reference price. Decomposition at \$1M for the initial bot (medians): half-spread 0.02 bps, drift during the delay 2.30 bps (2.76 bps on Bitcoin entries, 4.14 bps on Ether entries), depth 0.40 bps. Budget: 14.7 bps per execution for a Sharpe ratio of 1.5.

**Figure 6.** From backtest to net: the Sharpe ratio at \$1M, 2020–2026. Measured execution costs remove -0.17 ; faster entries recover 0.08 . Not included: the fee correction (next volume tier: about $+0.02$) and perpetual funding (about -0.06).

5.1 An order-book check on 9 executions

The method of comparable orders has a known weakness, so we checked it against the full order book where the data were free. On 9 executions falling on the first day of a month (two for Bitcoin, seven for Ether), the Binance book was rebuilt from the updates of the ten minutes before each instant, a method validated against a full replay from a 1,000-level snapshot (zero sequence breaks on a full day; one break among the 9 short windows). The Hyperliquid book at the same instants comes from 20-level snapshots every 0.55 s. Each simulated order used only the book as it stood at its own instant.

Table 8. Binance against Hyperliquid at the same 9 instants, 0.5 s after the signal (medians; minimum in brackets).

	BTC Binance	BTC Hyperliquid	ETH Binance	ETH Hyperliquid
Spread	0.01 bps	0.10 bps	0.05 bps	0.44 bps
Depth within 3 bps	\$5.8M (4.7)	\$0.2M (0)	\$2.3M (0.4)	\$0.7M (0.01)
Depth within 10 bps	\$22.1M (14.9)	\$0.2M (0)	\$6.5M (2.5)	\$5.8M (2.1)
\$250k order: filled; cost	2/2; 0.00 bps	1/2; 1.05 bps	7/7; 0.17 bps	7/7; 0.60 bps
\$500k order	2/2; 0.00 bps	0/2; –	7/7; 0.54 bps	7/7; 0.97 bps
\$1M order	2/2; 0.20 bps	0/2; –	7/7; 1.09 bps	7/7; 2.05 bps
\$2.5M order	2/2; 0.67 bps	0/2; –	7/7; 1.47 bps	6/7; 3.02 bps
\$1M if half the displayed size is withdrawn	0.59 bps	–	1.38 bps	–

Cost: average fill price beyond each venue's own best price. On Hyperliquid, "filled" means within the 20 visible levels; an order not filled there is unobservable, not unfillable. Two Bitcoin and seven Ether executions; hard cases kept in the medians. With two cases, the Bitcoin columns illustrate rather than estimate.

- **Bitcoin: consistent, on two cases.** The order book and the comparable-orders method agree to within about a third of a basis point from \$250k to \$1M.
- **Ether: the method was too optimistic.** The book costs 0.35 bps more for a \$500k order and 0.71 bps more for \$1M, in five cases out of seven. Large real orders trade when liquidity is there; an order sent at a fixed time sometimes meets a thinner book. Applied to Ether's executions, this removes about 0.01 from the Sharpe ratio. We keep it as a correction to carry, not a reason to change the method.
- **Hyperliquid is thinner.** Its spread is ten times wider on Bitcoin; its depth cost is about 2 to 3.5 times higher on Ether. On Bitcoin the 20 published levels cover only 1.7 to 2.1 bps and held about \$0.2M, so orders of \$500k or more cannot be costed from its archives.
- **The cross-venue gap is real.** Hyperliquid's mid-price sat above Binance's by a median 2.9 bps (Bitcoin, two cases) and 1.5 bps (Ether), up to 8.3 bps. It is measured apart from book quality.

Case study: Bitcoin, 1 May 2025, a buy signal at the close

Half a second after the close, Binance offered 5.7 BTC (\$0.55M) at its best ask of 96,291.1 and \$29.3M within 10 bps: a \$2.5M buy would have cost 0.60 bps beyond the best price. On Hyperliquid the 20 visible ask levels held only dust up to 2.1 bps away, and its best *bid*, at 96,328, sat above Binance's best ask. Hyperliquid was quoting 4.8 bps higher: its sellers had just been lifted. A desk executing there would have paid both the price gap and the missing depth. This is the situation a capacity estimate must survive, and the reason Hyperliquid figures in this dossier are conditional.

Fees and funding. The backtest pays 0.045% per order, Hyperliquid's base taker fee, once per execution. Reaching the next volume tier at \$1M would add about 0.02 to the Sharpe ratio; we

leave it out. Historical perpetual funding, applied with its sign to long and short positions, lowers the 2018–2026 Sharpe ratio from 2.02 to 1.97 (1.95 with Hyperliquid’s own funding from May 2023). The net projections of Section 7 exclude funding and say so.

Safeguards in the bot.

- Entries and exits at the close are sent as marketable limit orders (IOC) capped at 0.5% from the best price: never a fill beyond it. Stops are native exchange stop orders, triggered on the mark price and capped at 5% from the trigger; at \$1M their cost reaches 49 to 54 bps at the 99th percentile in the replay.
- An incomplete exit is retried, then the remainder is protected by a stop and the bot stops itself.
- The position is re-read from the exchange before any halt; every order carries an identifier that tells its purpose.
- Every step between the close and the fill is time-stamped, so that live latency and costs can be reported as P50 to P99 distributions.

6 Capacity

“How much capital can it carry before its own orders eat the edge?”

On Binance, the bot’s execution path as built keeps a net Sharpe ratio above 1.5 at every capital tested, up to \$5M, in the central estimate. The uncertainty is wide, and on Hyperliquid the answer depends on a depth that history does not fully show.

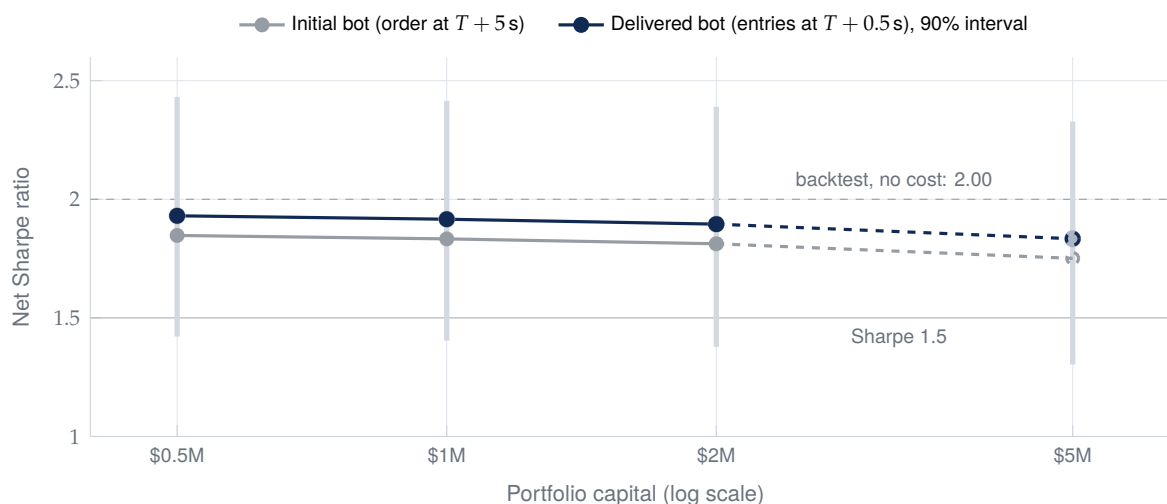


Figure 7. Net Sharpe ratio after measured execution costs, by capital, Binance perptuals, 2020–2026; month-block bootstrap intervals. Solid grey line: the threshold of 1.5, which neither version crosses up to \$5M. Hollow points at \$5M are estimates: \$2.5M orders exceed the largest size observed directly. Funding excluded.

- **Probability, not a line.** The chance that the resampled net Sharpe ratio exceeds 1.5 is 91% at \$500k, 90% at \$1M, 89% at \$2M and 85% at \$5M. No crossing below 1.5 is observed up to \$5M, the largest capital tested; the \$5M point is itself an estimate. We do not extrapolate beyond it.

- **Stops are the binding constraint.** Stops fire in fast markets. At \$1M of capital, the stop order exceeds 10% of the volume of its trigger minute in 3.5% of cases; at \$5M, in 31%, and the whole minute's volume in 2.6%.
- **What was gained.** Managing stops at 0.5 s, not yet coded, would add about 0.02 at \$1M. The pre-registered model-based study, which spread each order over the first minute and used an uncalibrated square-root impact, gave a net Sharpe ratio of 1.63 at \$5M in its central case and 1.14 in its prudent case, against 1.83 measured; most of that gap comes from the one-minute execution assumption.

From Binance to Hyperliquid. For the depth part of the cost, a book ρ times as deep as Binance's is equivalent to a capital $1/\rho$ times larger; the drift does not depend on size. The net Sharpe ratio of \$1M on Hyperliquid is therefore the Binance result at \$1M/ ρ , and capacity scales with ρ .

Table 9. Net Sharpe ratio at \$1M and capacity at a Sharpe ratio of 1.5 for the bot's execution path as built, by relative depth ρ of Hyperliquid against Binance.

Relative depth ρ	Sharpe at \$1M	Capacity	Basis
1	1.92	> \$5M	Binance, measured costs
0.5	1.90	> \$2.5M	hypothesis (Binance at \$2M)
0.2	1.83	> \$1M	hypothesis (Binance at \$5M, estimated)

On the rebuilt executions, the median depth visible within 5 bps was \$2.3M on Hyperliquid against \$3.0M on Binance for Ether ($\rho \approx 0.77$, seven cases). For Bitcoin, Hyperliquid's 20 published levels held at most \$0.4M, all within about 2 bps, against \$8.7M within 5 bps on Binance (median of two cases): the ratio cannot be measured from the archive, and if Bitcoin's real depth on Hyperliquid were that thin, a \$1M portfolio would sit far outside the range of Table 9. Hyperliquid's real capacity will be measured by the bot's own first orders and by a deeper book archive (Section 10).

7 Forward-looking results

"What should I expect, and when will we know if it is real?"

No result on real capital exists yet. The projections below apply measured costs to the historical record; the forward validation started in October 2026.

Table 10. Projection after measured execution costs, published rules with the delivered execution path, January 2020 to September 2026.

Capital	Execution venue	Sharpe [90%]	CAGR [90%]	Max. DD	P(Sharpe > 1.5)
Reference	backtest, no execution cost	2.00 [1.49; 2.50]	59.9%	−17.8%	—
\$500k	Binance, measured	1.93 [1.42; 2.43]	57.0% [34%; 90%]	−18.1%	91%
\$1M	Binance, measured	1.92 [1.40; 2.41]	56.3% [34%; 89%]	−18.1%	90%
\$1M	Hyperliquid if $\rho = 0.5$	1.90 [1.38; 2.39]	55.3% [33%; 88%]	−18.1%	89%

Commissions included; perpetual funding excluded (about −0.06 of Sharpe ratio over 2018–2026). Intervals: month-block bootstrap, 2,000 draws; they resample history and are not forecasts. The Hyperliquid row equals the Binance result at \$2M, by the depth equivalence of Section 6.

A haircut for selection. The record is in-sample. On Bitcoin, the walk-forward kept 54% of the fixed rules' Sharpe ratio (0.83 against 1.55); across 12 longer window designs the median kept 63%. Applying these ratios to the net Sharpe at \$1M gives about 1.0 to 1.2. This is an order of magnitude to keep in mind, not an estimate; it does not cover the selection of the two entry conditions, measured separately (Section 2).

Why a live Sharpe ratio takes years. For independent daily returns, the standard error of an annualised Sharpe ratio estimated over T years is approximately (Lo, 2002)

$$\widehat{\text{SE}}(\widehat{\text{SR}}) \approx \sqrt{365.25} \sqrt{\frac{1 + \frac{1}{2} \text{SR}_d^2}{N}} \approx \frac{1}{\sqrt{T}}, \quad \text{SR}_d = \frac{\text{SR}}{\sqrt{365.25}}, \quad N = 365.25 T. \quad (4)$$

Departures from independence and normality change it; the month-block bootstrap of our own history implies about $0.85/\sqrt{T}$.

Table 11. What a live record can show (Equation 4, independent daily returns).

Years of live trading	1	2	3	5	10
Standard error of the measured Sharpe ratio	1.00	0.71	0.58	0.45	0.32

If the true Sharpe ratio is 1.5, about 1.2 years of live trading are needed to be 95% confident that it is above zero, and about 11 years to tell it from 1.0. A first year of live trading will therefore establish *implementation fidelity*: that the live system takes the same decisions as the simulation on the same bars, at a measured cost. It cannot establish the Sharpe ratio, and we will not present it as if it did.

Table 12. Forward validation, status on 13 October 2026.

Component	Started	Status, and what it will establish
Bot on Hyperliquid's test network (Bitcoin, play money)	9 Oct.	Still runs the rules in force before 13 October: it does not yet apply the calm-regime restriction, and will be aligned before any live capital. No trade since its restart. One earlier strategy trade (a short entered on 7 October with an exchange time stamp at $T + 5.01$ s, closed by hand at the restart). Establishes that live decisions match the simulation, time-stamped by the exchange; not performance.
Pre-registered forward test of the short-entry condition	12 Oct.	No short closed yet; first evaluation after 1 November 2026. Tests whether shorts entered in a bearish regime beat the others. The calm-regime restriction has no forward test of its own.
Latency instrumentation of the bot	Oct. 2026	In the code; deployment on the server to be confirmed. Will give the real distribution of the delay from the close to the order.

The rules of the forward test. All committed before the first observation: interim looks after 25, 50 and 75 shorts closed in a bearish regime, each requiring $p < 0.001$; a final look when both groups reach 100 closed shorts, requiring $p < 0.047$, and no later than 1 October 2031, for a total type-I error of at most 0.05 (Haybittle, 1971; Peto et al., 1976). At roughly 89 shorts a year, the final look should come in about three years. Its power is 44% to 54% if the true difference in mean

risk-normalised return R is 0.4 (about 0.2 standard deviation of R), less once the correlation of shorts within a month is counted. An inconclusive result will often mean “not enough data”; the protocol accepts this, preferring a clean test to a fast one.

Disclosure: rule changes after the forward protocol was written

The forward protocol was written while the research reference still traded an unconditional short mirror, with the trend condition on shorts tracked in shadow, and it stated that an inconclusive result would mean no change. On 9 October 2026 the bot was aligned with the trend condition, and on 13 October 2026 the published rules were restricted to entries in the calm volatility regime. The test itself is unaffected: it replays the unconditional rules offline on new data and classifies each short by the trend regime. But its “no change” rule was written for the unconditional rules: an inconclusive result would now leave the trend condition in place without out-of-sample confirmation, and the calm restriction is outside the test altogether. We record this as a deviation from the protocol’s governance, not from its test.

8 What we got wrong, and what it taught us

“How does this team behave when the data disagree with it?”

Research is mostly the correction of earlier beliefs. These are the ones that changed numbers in this dossier.

1. A three-second wait that nobody had measured

The bot waited three seconds after each close “so that the exchange could finalise the candle”. No measurement had set that margin. On the first live test trade the order left 5.01 s after the close, and the pre-registered drift study priced that delay at about 4.4 bps per execution. An audit showed that candle integrity came from a different rule entirely, so the wait protected nothing. **What we did:** rewrote the critical path without touching the signals; Figure 6 shows the value of that work.

2. An early capacity model built on assumptions, not measurements

Our first capacity study, lacking data, assumed each order would be spread over the first minute after the signal and priced impact with the square-root law and an uncalibrated constant. Both assumptions were costly. The drift over one minute is 3.9 to 9.2 bps, against 0.4 bps for an order sent half a second after the close; and the model’s 2.5 bps for a \$500k order is several times the median depth cost real orders of that size paid at our entry times (0.24 bps on Bitcoin, 0.60 bps on Ether). At \$5M of capital the model’s prudent case left a net Sharpe ratio of 1.14; the measurement gives 1.83. **What we did:** kept the model’s result in its report, labelled as superseded, and built the capacity curve from measurements.

3. A measurement method that flattered Ether

Measuring costs from comparable real orders is powerful but carries a selection bias: big orders trade when liquidity allows. Rebuilding the full order book at our execution instants was consistent with the method on Bitcoin and showed Ether costs understated by 0.35 to 0.71 bps. **What we did:** report the correction and its effect (about -0.01 of Sharpe ratio), and recommend a larger, pre-registered order-book sample.

4. Entry conditions added late, with knowledge of the data

For most of the research the strategy traded both volatility regimes with an unconditional short side. Its Sharpe ratio fell from 2.13 in the first half of the sample to 1.29 in the second, almost entirely because of shorts; with the trend condition on shorts and entries restricted to calm regimes, the two halves read 2.12 and 1.98. Both conditions were chosen on data already used to build the strategy, so the comparison explains the decline but cannot validate them. **What we did:** froze the conditions; measured the selection risk of the choice across the 16 variants actually tried (probability of backtest overfitting 19%; gain of the calm restriction not significant after deflation); subjected the short condition to three pre-registered falsification tests and disclosed that most of its Sharpe gain is generic; and launched a forward test that can say no.

5. A wrong figure in the first draft of this dossier

The first draft quoted the execution cost of a version of the bot that manages stops itself; that version is not coded. An independent line-by-line check against the source reports caught it. The same check found that our first summary left out the executions that sat in an intermediate category of evidence; Table 6 now shows all three categories. **What we did:** corrected every affected figure and kept the check as a standing step before publication; the figures of this edition are generated from the result files, and the build refuses to run if they disagree.

9 Risks and limitations

These results show that the historical edge is not a simple artefact and that, on Binance, it survives the measured execution costs up to a few million dollars in the central estimate. They do not show that it will persist, nor what Hyperliquid's depth allows.

- ▶ **Selection.** Bitcoin's parameters were selected on data covering the whole period; Ether is a transfer across markets, not across time. Walk-forward: 0.83 against 1.55, with a maximum drawdown of -27.6% ; median 0.97 with longer windows.
- ▶ **Entry conditions in-sample.** The calm-regime restriction and the short-entry condition were set in October 2026 with knowledge of both markets; the probability that the choice is overfitted is 19%. Only the forward test can validate the short condition out of sample, and nothing yet tests the calm restriction out of sample.
- ▶ **Right tail.** The best 5% of trades make 99% of the growth. Missed trades in live operation would weigh heavily.
- ▶ **Different venues.** Signals come from Bitstamp (Bitcoin) and Binance spot (Ether); execution is planned on Hyperliquid perpetuals, which quoted a median 1.5 to 2.9 bps above Binance at the executions studied. The effect of this gap on performance is not modelled.
- ▶ **Hyperliquid liquidity.** Its historical book shows 20 levels only, about 2 bps deep on Bitcoin. Capacity there is conditional on a depth ratio that the 9 cases studied place below one, far below on Bitcoin.
- ▶ **Own impact.** Permanent impact, information leakage and the adaptation of market makers to a regular flow are not measured. Above about \$4M of capital, single orders exceed the largest size observed.
- ▶ **Stop mechanics.** Hyperliquid stops trigger on the mark price, not the last trade, and fill up to 5% from the trigger. As in the backtest, no exit order rests during a position's first 15-minute bar; an added protective stop would change realised results.
- ▶ **Execution assumptions.** The bot as now coded has not traded; its 0.5 s latency is a target; the replay assumes stops reach the book 3 s after the price crosses their level.
- ▶ **Bot not yet aligned.** The test-network bot does not yet apply the calm-regime restriction; until it does, its decisions match the earlier rules, not the published ones.
- ▶ **Funding.** Excluded from headline and projected figures; about -0.06 of Sharpe ratio historically.
- ▶ **Venue and operational risk.** Counterparty, technical and regulatory risks of the execution venue are not assessed in this dossier.

- **Multiple testing.** Criteria were fixed in advance and failures are reported, but the research process still involved choices that no single test accounts for.

10 What would reduce the uncertainty most

In order of information gained per unit of cost and time:

1. **Align the bot with the published rules** (no entry in agitated regimes), so that live decisions can be compared with the simulation one for one.
2. **Six to twelve months of live execution with small capital on Hyperliquid**, with the new time stamps: real latency, fill rates, stop slippage and cost against the decision price, the only way to measure our own impact.
3. **A server close to the exchange.** According to Tardis, Hyperliquid's servers are in Tokyo; every round trip saved is drift not paid.
4. **Hyperliquid's official book archive** (requester-pays, about one dollar of transfer; coverage not yet verified), to measure ρ near the best price at our own signal times. It probably carries the same 20-level limit, so Bitcoin's deeper liquidity would still need another source, such as node or order-level data.
5. **A pre-registered sample of about 300 executions on full Binance order books** (paid data), to fix the Ether correction and the worst cases at \$1M to \$2.5M orders.
6. **Stops managed by the bot at 0.5 s** on the last traded price, worth about +0.02 of Sharpe ratio in the replay.
7. **The forward test of the short-entry condition**, which will run its course whatever the other results.

A Definitions

- **Returns and ratios.** Daily UTC close-to-close returns r_d with mean $\bar{r}$ and standard deviation s ; $SR = \sqrt{365.25} \bar{r} / s$; Sortino uses the downside deviation; the maximum drawdown is $\min_t V_t / \max_{u \leq t} V_u - 1$; Calmar is CAGR over the absolute maximum drawdown.
- **Risk-normalised trade return.** $R_i = (\pi_i / E_i) / \sigma_{d_i}$: profit over equity at entry, divided by the standard deviation of the 30 daily log returns before the entry day.
- **Implementation shortfall.** Equation (1); costs in basis points of notional, fees excluded.
- **Depth equivalence.** For the depth cost, an order Q in a book ρ times as deep costs what an order Q/ρ costs in the reference book.
- **Bootstrap.** Calendar months drawn with replacement, the same months for every series, so that cross-sectional dependence is preserved (Künsch, 1989).

B Sources and reproducibility

Every figure comes from versioned code and data in the project's research repository. Protocols were committed before the corresponding computations; the commit is the evidence.

Subject	Report	Pre-registration (commit)
Strategy, robustness, residual alpha	Working paper on the strategy, https://www.shadowmarketpro.com/research/btc-eth-portfolio-calm/	residual alpha (fa7e88b)
Selection risk of the rules	calm-only/	a2ae2e0 (plan 40fae7f)
Short entries; short-condition diagnostics and falsification	short-entry study, diagnostics and falsification reports	6b91ec7; a356c2d; 81e89f4
Cost budget, model capacity	capacity/capacity-report.md	capacity-p0p1.md (513836e, addendum 88d7960)
Latency and drift	execution/phase1-audit.md, latency-drift.json	execution-phase1-latency.md (2c1dd24)
1,150 executions, capacity	execution/fills-500k-1m.json, fills-2m-5m.json, fills-500k-1m.md	execution-fills-500k-1m.md (63f08d5); \$2M and \$5M not pre-registered
Order books	execution/binance-l2-vs-hl.md, hl-books-feasibility.md	–
Forward test of the short condition	forward-test report, updated at each look	f59ee70

Market data: Bitstamp and Binance candles; Binance USD-M trades and book depth from <https://data.binance.vision>; Tardis order books on free days; Hyperliquid's public order history. The figures and tables marked in the source are generated by `research/dossier/build-dossier.ts` from the published result files; the build refuses to run if the portfolio's consistency checks fail.

C Version history

- ▶ **3.1, 13 October 2026.** Walk-forward with longer windows (12 pre-registered designs) added to the robustness evidence and the selection haircut.
- ▶ **3.0, 13 October 2026.** Published rules updated: new positions are opened only in the calm volatility regime. Every figure recomputed on the new rules, including the 1,150 executions, costs, capacity and the order-book check; selection risk of the rule choice reported; design history of the two entry conditions corrected, since both were adopted in October 2026.
- ▶ **2.1, 12 October 2026.** The short-entry condition presented with the history of its implementation in the research code.
- ▶ **2.0, 12 October 2026.** Typeset edition; scientific approach, execution framework and literature, figures, lessons learned. Corrections: the cost of the bot as built; the three-way reconciliation of market executions; funding stated as excluded from projections.
- ▶ **1.0, 12 October 2026.** First dossier.

References

- Almgren, R., and Chriss, N. (2000). Optimal execution of portfolio transactions. *Journal of Risk*, 3(2), 5–39.
- Almgren, R., Thum, C., Hauptmann, E., and Li, H. (2005). Direct estimation of equity market impact. *Risk*, 18(7), 57–62.
- Bacry, E., Iuga, A., Lasnier, M., and Lehalle, C.-A. (2015). Market impacts and the life cycle of investors orders. *Market Microstructure and Liquidity*, 1(2), 1550009.
- Bailey, D. H., and López de Prado, M. (2014). The deflated Sharpe ratio: correcting for selection bias, backtest overfitting, and non-normality. *Journal of Portfolio Management*, 40(5), 94–107.
- Bailey, D. H., Borwein, J., López de Prado, M., and Zhu, Q. J. (2017). The probability of backtest overfitting. *Journal of Computational Finance*, 20(4), 39–69.
- Bouchaud, J.-P., Bonart, J., Donier, J., and Gould, M. (2018). *Trades, Quotes and Prices: Financial Markets Under the Microscope*. Cambridge University Press.
- Cartea, Á., Jaimungal, S., and Penalva, J. (2015). *Algorithmic and High-Frequency Trading*. Cambridge University Press.
- Cont, R., Kukanov, A., and Stoikov, S. (2014). The price impact of order book events. *Journal of Financial Econometrics*, 12(1), 47–88.
- Daniel, K., and Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, 122(2), 221–247.
- Donier, J., and Bonart, J. (2015). A million metaorder analysis of market impact on the Bitcoin. *Market Microstructure and Liquidity*, 1(2), 1550008.
- Frazzini, A., Israel, R., and Moskowitz, T. J. (2018). Trading costs. Working paper, AQR Capital Management and Yale University.
- Gatheral, J. (2010). No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10(7), 749–759.
- Harvey, C. R., Liu, Y., and Zhu, H. (2016). ... and the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5–68.
- Haybittle, J. L. (1971). Repeated assessment of results in clinical trials of cancer treatment. *British Journal of Radiology*, 44(526), 793–797.
- Hong, H., and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, 54(6), 2143–2184.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17(3), 1217–1241.
- Landier, A., Simon, G., and Thesmar, D. (2015). The capacity of trading strategies. Working paper, HEC Paris.
- Lee, S. S., and Mykland, P. A. (2008). Jumps in financial markets: a new nonparametric test and jump dynamics. *Review of Financial Studies*, 21(6), 2535–2563.
- Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58(4), 36–52.
- Makarov, I., and Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2), 293–319.
- Newey, W. K., and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., and Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606.
- Novy-Marx, R., and Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *Review of Financial Studies*, 29(1), 104–147.
- Obizhaeva, A. A., and Wang, J. (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1), 1–32.
- Perold, A. F. (1988). The implementation shortfall: paper versus reality. *Journal of Portfolio Management*, 14(3), 4–9.
- Peto, R., Pike, M. C., Armitage, P., et al. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *British Journal of Cancer*, 34(6), 585–612.
- Tóth, B., Lempérière, Y., Deremble, C., de Lataillade, J., Kockelkoren, J., and Bouchaud, J.-P. (2011). Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 1(2), 021006.
- White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5), 1097–1126.

Zarinelli, E., Treccani, M., Farmer, J. D., and Lillo, F. (2015). Beyond the square root: evidence for logarithmic dependence of market impact on size and participation rate. *Market Microstructure and Liquidity*, 1(2), 1550004.