THANAAYA TECHNOLOGIES

RESEARCH NOTE

A Test That Could Say Yes

A pre-registered test of liquidation-cascade reversal in crypto futures found no tradable edge, and proved beforehand that it could have found one.

ABOUT THANAAYA TECHNOLOGIES

Thanaaya Technologies is a software company licensed by the Dubai Department of Economy and Tourism, building since 2021. We design, build and operate production systems across financial software, distributed ledger, gaming, commerce, media and quantitative research — and we run two of them as our own products rather than only delivering them for clients.

We take a small number of engagements at a time, and the engineer who designed a system is the one who gets called when it misbehaves. This paper comes out of our own research programme, which exists to find out cheaply whether an idea is real. We publish the negative results as well as the positive ones: an undocumented failure gets rediscovered, and a threshold nobody wrote down cannot be checked.

2021	24	80+	9
BUILDING SINCE	PRODUCTION DOMAINS	SERVICES IN PRODUCTION	INDUSTRIES SERVED

Subject: reversal after inferred liquidation cascades in Binance USDT-margined perpetual futures, ten contracts
Programme: Thanaaya internal quantitative research
Status: study closed on its pre-registered rule; holdout never opened
Published: November 2026 · **Licence:** Dubai Department of Economy and Tourism 964356 · **Unified licence:** BL1484

PART ONE

In plain English

One idea, tested once, with the rules fixed in advance. What we asked, how we proved the test could work, what it found, and what that is worth. No mathematics in this section.

1. The question

When a crypto futures market is hit by a wave of forced selling or buying, does the price snap back far enough, and for long enough, to trade at a profit after costs?

When traders are leveraged and the market moves against them, the exchange closes their positions by force. The forced orders can push the price further than anyone chose to, and in theory the price then reverts, a mechanism economists describe as a liquidity spiral (Brunnermeier and Pedersen, 2009). It is a popular idea and a plausible one. Our programme had already tested hundreds of signals built on a different kind of data, and all of them had come back null. This was a deliberate change of direction, to a mechanism that does not depend on the data those signals shared.

One hypothesis, pre-registered, and one verdict. We did not test variants, because every extra variant makes a lucky result more likely, and we had just spent months learning that lesson the hard way.

2. Making the test fair before looking

Before any result existed we fixed, wrote down and fingerprinted:

- **One idea.** A single hypothesis and a single holding rule, not a menu.
- **A twelve-month development window** on ten of the largest USDT-margined perpetual contracts on Binance, and a **sealed six-month holdout** that was listed but never downloaded. The holdout would open only on a written phrase, and only if development passed.
- **A cost bar set high.** The result had to clear a 28-basis-point round trip, twice the cost we think is realistic. A pass at the lower figure would not count.
- **A sentence that protects us from ourselves:** a result below the gate is a null, and the response is to close the study, not to re-parameterise.
- **An honest statement of power.** About forty market-wide days made up the real sample. A null was the expected outcome, and a null would show only that this sample could not detect an edge, not that none exists.

The exchange does not publish forced liquidations in its public archive, so the events are inferred from price, trading flow and open interest. We said so in the registration.

3. Proving the test works before running it

A test that always says “no” looks the same as a broken one. So before the real run, the pipeline had to pass four checks on itself.

CHECK	WHAT IT PROVES	RESULT
Replay test, random cuts	The pipeline cannot see the future: data after a random cut is replaced with noise and every earlier event must be unchanged.	**50 runs, 0 differences**
Entry-bar cut	The entry price is not contaminated by information from the entry bar itself.	**312 of 312 events unchanged**
Oracle	The gate can say yes: trades given deliberate foresight of the outcome must pass.	**Passed (Sharpe 6.1, t 7.7)**
Eight coin-flip sets	The gate cannot be fooled by noise: random entries and directions must fail.	**All 8 failed**

The replay test also earned its place in a way we did not plan. **The first version of the pipeline failed it.** A one-minute overlap between the window used to detect an event and the bar used to enter it meant that 215 of the 321 events then on file changed when the entry bar’s information was removed. No return had been computed. The fix was applied as a dated, hashed amendment, the event list was rebuilt (312 events), and the test then passed in full. A tool built to catch lookahead caught ours before it could flatter anything.

4. What came back

−1.64
annualised Sharpe ratio at 28 bp, against 1.0 needed

−1.95 vs 2.20
month-clustered t-statistic against the critical value

308
trades on 120 days across ten contracts

The strategy failed five of its six statistical criteria, and not narrowly. The study is closed, as the registration says. The six-month holdout was never opened and no file from it was ever downloaded.

The lower cost assumption did not rescue it either: at 14 basis points the result was still negative (Sharpe −0.67, t −0.79). There is no “it would have worked if we traded cheaper” story to tell.

5. What is worth reading in the numbers

+0.28% average gross return per trade, before costs: about the size of the cost bar

Averaged per trade, the gross return was positive and roughly equal to the 28-basis-point cost, so net it came out at about zero. The typical trade lost: the median net trade was −0.30%, and the average was carried by a few very large winners. One violent market-wide crash day produced

the three biggest winners in the whole sample, one of them above 60% in a single trade. Take away the ten best trades and the rest average −0.45% each.

That is a descriptive finding, and a modest one. It says the mechanism may appear at the extremes and is invisible in the ordinary run of events, which is where a day-level test with one dominant day cannot reach a verdict. It is **not** a reason to re-tune the event definition to isolate the extremes. That is the variant we registered against.

The long side (after a drop) earned a small gross positive return and the short side (after a spike) a small negative one. There were only 48 short trades, too few to explain anything, so we report the difference and leave it unexplained.

6. What the study leaves behind

- **A validated test harness.** It detects a planted effect and rejects random noise on this data with these statistics. Every later study starts from a machine already known to work.
- **A defect found by design.** The replay test caught a real lookahead bug on its first outing, before any result could be flattered.
- **A clean, citable null.** It is pre-registered, hashed, placebo-validated and oracle-validated, with a power statement written beforehand. It will stop us re-asking the same question without new information.
- **A way to close.** The response to a null was written down before the data. The study ended on that rule, without a rescue attempt.

PART TWO

The technical record

Data, protocol, amendments, validation, results and limits. Figures are as measured. Event definitions and trading parameters are not published.

1. Data and universe

Venue	Binance USDT-margined perpetual futures, ten large contracts (BTC, ETH, SOL, BNB, XRP, DOGE, ADA, AVAX, LINK, LTC).
Source	The exchange’s public archive: one-minute bars with taker volumes and five-minute open-interest history. Liquidation records are not available in the public archive, so events are inferred.
Windows	Twelve months of development data (seven extra days for warm-up). Six months of sealed holdout: the 3,680 archive files that cover it were listed in a manifest, hashed, and never downloaded.
Effective sample	312 events, 121 distinct event days, 41 market-wide days (three or more contracts the same day). All twelve months contain events.

2. Protocol

Costs	Gate at a 28 bp round trip (twice a realistic taker cost on this venue). 14 bp reported for context only. A fail at 28 bp that passed at 14 bp would have been a descriptive finding, not a certification.
Holding period	Set from the measured duration of the events (the median of how long abnormal flow persisted), rounded to five minutes and fixed before any return was computed. The rule, not a chosen number, produced it.
Entry	The open of the one-minute bar after the event window closes. One trade per contract at a time; overlapping events skipped and counted.
Daily series	Mean net return across each day’s trades, zero on days with none, 365 days.

The gate required, together: at least 100 trades on 25 days; a month-clustered t-statistic of at least 2.20; annualised Sharpe of at least 1.0; positive after removing the largest day, with t of at least 1.0; a day-block sign-flip permutation p of at most 0.01; at least six of ten contracts positive; and the lookahead tests above. The 2.20 is the 5% two-sided critical value for 11 degrees of freedom (twelve month clusters), not the conventional 2.0. The clustered standard error is the CR0 estimator over the twelve months (Cameron and Miller, 2015). The permutation flips the sign of all gross returns on a calendar day jointly, which keeps the same-day clustering

and the cost term and destroys only the link between cascade direction and the subsequent return. B = 20,000.

3. Amendments, all before any return

AMENDMENT	CHANGE AND REASON
1	Operational definitions of every gate statistic added as an appendix (the registration named the statistics but not their formulas); entry-side window exclusion added; power statement added to the main text.
2	Bar alignment fixed after the replay test failed. Event count 321 → 312; event days 135 → 121; market-wide days 40 → 41; thin-sample test still passed. Only 196 of the original 321 events appeared in the corrected set. The holding period was re-measured on the corrected events by the unchanged rule.
2 (addendum)	The replay requirement was made two-part: the per-event entry-bar cut and the registered random-cut test, both required, before the oracle.

Every amendment was dated, hashed and written before the first return was computed. The pipeline code was hashed before the first validation stage ran, and each stage refused to run unless the previous stage had passed.

4. Validation

CHECK	RESULT
Replay on the original pipeline: events changed when the entry bar was masked	**215 of 321**
Replay on the corrected pipeline: events changed	**0 of 312**
Random-cut replay (5 cuts × 10 contracts)	**0 differences in 50 runs**
Oracle: 308 trades; Sharpe; t; permutation p; breadth	**6.11; 7.74; 0.00005; 10 of 10**
Coin-flip sets: Sharpe range; t range; permutation p range	−5.85 to −3.84; −6.76 to −3.65; 0.24 to 0.68
Coin-flip sets passing the gate	**0 of 8**

The coin-flip sets are strongly negative because they pay the full cost on every random trade, which is also why the permutation null in the next section is centred below zero.

5. Development-window result at 28 bp

CRITERION	REQUIRED	OBSERVED	MARGIN	RESULT
Sample (trades; entry days)	100; 25	308; 120	–	pass
Month-clustered t (CR0, 11 df)	≥ 2.20	−1.95	−4.15	fail
Annualised Sharpe	≥ 1.0	−1.64	−2.64	fail
Leave the largest day out	mean > 0, t ≥ 1.0	mean < 0, t −2.66	–	fail
Permutation p	≤ 0.01	0.376	+0.366	fail
Breadth (contracts positive)	≥ 6 of 10	3 of 10	−3	fail

Counts: 312 events; 1 removed by the entry-side window rule, none by the exit-side rule, no missing bars, 3 skipped as overlapping; 308 trades. The 20,000-permutation Sharpe distribution had percentiles of −4.16 (1st), −3.53 (5th), −1.95 (50th), −0.32 (95th) and +0.31 (99th). The observed −1.64 is unremarkable among shuffled directions: 37.6% of the shuffles did as well or better. At 14 bp the result was Sharpe −0.67, t −0.79 and breadth 5 of 10: also a fail.

6. Descriptive breakdowns (reported regardless of outcome)

Net return per trade at 28 bp.

DISTRIBUTION	VALUE
Mean / median	+0.003% / −0.304%
1st / 5th / 25th percentile	−5.08% / −3.16% / −1.23%
75th / 95th / 99th percentile	+0.60% / +2.42% / +4.76%
Skew	+10.4
Average gross per trade, before costs	+0.283%
Mean net after the ten best trades are removed (descriptive only)	−0.45%

By contract (net at 28 bp) and by side.

CONTRACT	TRADES	MEAN NET	CONTRACT	TRADES	MEAN NET
BTC	38	−0.08%	XRP	22	−0.56%
ETH	44	−0.37%	DOGE	42	+1.39%
SOL	36	+0.32%	ADA	23	−0.40%
BNB	26	−0.03%	AVAX	27	−0.89%
LINK	22	−0.34%	LTC	28	+0.16%
Long after a drop	**260**	**+0.086%**	**Short after a spike**	**48**	**−0.446%**

By month of the development window (net at 28 bp, percent).

MONTH	1	2	3	4	5	6	7	8	9	10	11	12
Trades	20	28	21	21	18	27	25	33	33	23	28	31
Mean net	+0.1	−1.3	−0.3	+0.7	−0.3	+2.9	+0.5	0.0	−0.7	−0.2	−0.5	−0.5

Month 6 contains the market-wide crash day behind the three largest winners.

7. Limits

- **Power.** A 12-month daily series gives t ≈ Sharpe × √1, so a clustered t of 2.20 needs a measured Sharpe near 2.2. A genuine edge of Sharpe 1 would be indistinguishable from luck here. The correct reading of the null is inconclusive at this power for modest effects, and clearly not a large effect, not “disproven”.
- **Twelve clusters.** Month-clustered inference with twelve clusters is fragile, which is why the critical value is taken at 11 degrees of freedom.
- **Inferred events.** Event flags are a proxy for forced liquidations, not a record of them. The pipeline checks reproducibility and the absence of lookahead, not that every flagged event was a liquidation.
- **One dominant day.** A single market-wide crash day supplies the largest gains. Day-level tests are fair to that, and they are also why the descriptive gross average cannot be separated from luck.
- **Short side.** 48 trades is too few to interpret the long/short difference.
- **No holdout.** The final six months stayed sealed. The study makes no claim about them.

8. Status

Closed. The response to a result below the gate was fixed in advance: close the study, do not re-parameterise. No variant, window change or re-cut of the event definition is proposed. Any new idea in this area would be a new registration with its own gate and its own sealed holdout.

References

Brunnermeier, M. K. and Pedersen, L. H. (2009). Market Liquidity and Funding Liquidity. Review of Financial Studies 22(6), 2201–2238. https://doi.org/10.1093/rfs/hhn098

Cameron, A. C. and Miller, D. L. (2015). A Practitioner’s Guide to Cluster-Robust Inference. Journal of Human Resources 50(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

Scope. Thanaaya Technologies is a Dubai Department of Economy and Tourism licensed company (licence 964356). This paper describes our own research programme, carried out by the company itself; no client is involved. No strategy or signal design is published, and nothing here constitutes investment advice or an offer of any service relating to trading. Enquiries: info@thanaaya.com · thanaaya.com