THANAAYA TECHNOLOGIES

RESEARCH NOTE

Measuring the Noise Floor

What a systematic search returns when there is nothing there — and why we published the number instead of the failure.

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 several 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 internal 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: directional prediction from market-microstructure signals, SOL perpetual futures
Programme: internal quantitative research — no client capital at any stage
Status: closed on a measured null
Published: September 2026 · **Licence:** Dubai Department of Economy and Tourism 964356 · **Unified licence:** BL1484

PART ONE

In plain English

No mathematics in this section. What we tested, on what, and what came back.

1. The question

Can you make money trading Solana futures by reading the order flow?

That is the whole premise, and it is a reasonable one. The instrument is the **SOL perpetual futures contract on Binance** — the largest venue for it, trading around the clock. Every trade and every change to the queue of waiting orders is published publicly, millisecond by millisecond.

The theory, which has decades of academic work behind it, is that this detail leaks. When someone who knows something is buying, the *pattern* of buying looks different from ordinary trading — and it should look different *before* the price moves. Read the pattern, get in front of the move.

We recorded a week of it: **5.7 million individual trades and 5.8 million snapshots of the order book**. Then we computed the standard indicators the literature recommends, plus several of our own, and asked whether any of them predicted the next move by enough to cover the cost of trading.

2. What the week actually looked like

Before any signal, it is worth knowing what you are trying to predict. Over that week, SOL went from **$75.69 to $76.00**. Thirty-one cents, over seven days.

Split into five-minute windows, here is every one of them:

919	967	133
five-minute windows where the price finished higher	windows where it finished lower	windows where it did not move at all

Nine hundred and nineteen up against nine hundred and sixty-seven down. **That is a coin.** Not approximately a coin — we later measured how close, and the answer was within 8% of a textbook coin across every holding period from ten minutes to seven hours.

So the thing being predicted has no trend to lean on. Whatever edge exists has to come entirely from calling individual flips.

3. The best signal we could find

Out of everything we tested, one indicator did best at the five-minute horizon: a measure of how one-sided the buying and selling had been. We took only the most extreme 5% of readings — the moments it was shouting loudest.

It fired 102 times in the week. Here is the entire scorecard:

Times it called the direction correctly	57
Times it was wrong	37
Average gain when right	0.081%
Average loss when wrong	0.076%
Average per trade, all 102	+0.018%
Cost to place one trade, in and out	0.036%

> READ THE LAST TWO ROWS TOGETHER
>
> The signal was right more often than not, and it still loses money. It earns 0.018% per trade and the trade costs 0.036% to place — **you would pay twice what you made.** And that is against the cheapest fee available anywhere on the venue; at ordinary fees the cost is 0.106%, or six times the gain.

4. Why we did not stop there

Being right 57 times out of 94 is 61% — which sounds like something, and is the point at which most people conclude they have found an edge and start looking for ways to reduce costs.

We did not, for one reason: **that was the best of 854 things we tried.**

> THE CORE PROBLEM
>
> When you test hundreds of possible rules and report the one that worked best, the winner will always look impressive — **even if none of them work at all**. Somebody wins the lottery every week. That is not evidence they had a system.

5. So we measured what "looking impressive" is worth

We took the real data and deliberately broke it. Every price movement kept exactly as it happened; every indicator kept exactly as calculated — but shifted apart in time, so each indicator was matched against a price move from a different moment. Same numbers, same volatility, same rhythms. Only cause and effect were severed.

In that version of the world, prediction is impossible by construction. Then we ran the whole 854-rule search on it, recorded how good the best one looked, and repeated that a thousand times.

t ≈ 3.4

What the best of 854 rules typically scored on data with nothing in it. On the standard scale, 2 is "statistically significant" and 3 is usually called strong evidence. One run in twenty exceeded 5.9.

Our real result scored 4.9 — which sits at the 88th percentile of what pure noise routinely produced. Unusual enough to feel exciting. Far too common to believe.

6. Why that number is worth more than the strategy

We could have written "we tested order-flow signals on SOL and they did not work". True, and nearly useless — the kind of sentence that gets quietly reopened in eighteen months by someone with a new indicator, because it names no threshold anyone can check.

Instead there is a number. Any future study on this desk gets measured against it *before* anyone becomes attached to a result. If a new idea scores 3, that is not a discovery. That is the floor.

The failed strategy cost a few weeks. The floor is permanent, and it applies to work not yet started.

7. Two other things worth knowing

Some of it was impossible before we began

SOL moves an average of 0.074% in five minutes. A fast round trip on Binance costs about 0.106%. **The cost of the trade is larger than the entire average move**, so at that speed you would need to be right more than 100% of the time. That is arithmetic, not a research finding, and it was knowable before we recorded a single trade.

We were wrong about something, and the test caught it

Partway through, one well-known indicator appeared to predict how *volatile* the market would be — not direction, but the size of the coming move — in perfect order: the more extreme the reading, the bigger the move. We argued that such a clean pattern was unlikely to be coincidence.

Then we ran it through the same scrambling test. Meaningless data produced that same perfect ordering **14% of the time.** We withdrew the finding. It is recorded as withdrawn, with the reasoning error stated plainly.

8. What we did not settle

Everything above assumes you are *paying* to trade immediately — taking the price on offer. The opposite approach, posting a patient order and waiting for someone to trade against you, has completely different economics, and **we have not measured it.** We are not claiming it works. We are declining to claim it does not.

PART TWO

The technical record

Instrument, construction, results, corrections and boundary conditions. Figures are as measured; nothing here is rounded in a favourable direction.

1. Subject and data

Instrument	SOLUSDT perpetual, Binance USDⓈ-M
Registered window	`2026-08-10 17:03:55 → 2026-08-17 23:59:56 UTC (174.93 h)`
Trades	`5,702,722 (after placeholder exclusion) · 79,795,693 SOL`
Book snapshots	`5,781,551 at five levels per side`
Held back, untouched	`~24 days / ~54.2M trades`
Feature families	VPIN · multi-level order-flow imbalance · leverage stress · liquidation-cluster gravity · order imbalance · cumulative volume delta
Decision clock	volume-time (VPIN bucket closes), backward-only as-of joins

Two bucket geometries were evaluated: V = 199,500 SOL (26.2 min/bucket, 399 rows) and V = 39,500 SOL (5.19 min/bucket, 2,020 rows). The latter is the geometry of record; the former produced fewer than twenty entries at a 5% conditioning threshold and therefore has no directional power.

2. The measured null

2.1 Construction

The null is a **random circular rotation of the forward-return block against the feature block**. An offset `k` is drawn uniformly and every forward-return column is rotated by the same `k`, pairing features at time `t` with returns from `t+k`.

Rotation destroys the feature→future-return relationship exactly, while preserving: each feature's distribution and autocorrelation; the return series' distribution, autocorrelation and volatility clustering; and the overlap structure among forward-return columns at different horizons. A block bootstrap would have required a block-length parameter. Rotation has no free parameter — which matters in a study whose subject is the cost of free parameters.

For each replication the *entire* search is re-run and only the extremes retained, so the resulting distribution is that of the *maximum over the whole grid* — the correct comparison, because the observed result was itself selected as a maximum. 1,000 replications, seed 7.

2.2 Result — geometry of record (1,940 usable rows; 287 magnitude + 567 direction cells)

Observed maxima against the distribution of maxima under rotation.

STATISTIC	OBSERVED	NULL P50	NULL P95	NULL P99	NULL MAX	P
Max conditional-volatility lift	1.713	1.569	2.284	2.406	2.749	0.367
Max direction edge (bps)	56.75	44.49	74.49	86.74	93.42	0.256
Max episode-corrected *t*	4.938	3.405	5.955	8.454	11.714	0.122

Observed maxima sit at the 63rd, 74th and 88th percentiles of the null. **Nothing in the search is distinguishable from rotation noise.**

The reported *t* is not the nominal statistic. Overlapping forward windows make nominal *t* meaningless — a return over 80 buckets shares 79 of them with its neighbour — so entries closer together than the horizon are collapsed into a single non-overlapping episode before the statistic is computed. **The 3.405 null median is therefore the honest statistic, not the inflated one.**

3. Volume-time diffusion

Unconditional mean |forward return| against a √t prediction anchored at h = 1.

HORIZON (BUCKETS)	1	2	5	10	20	40	80
Minutes	5	10	26	52	104	208	416
Observed (bps)	7.37	10.29	15.61	22.44	32.90	46.50	60.68
√t prediction	7.37	10.43	16.49	23.32	32.97	46.63	65.95
Ratio	1.000	0.987	0.947	0.962	0.998	0.997	0.920

Anchored at h = 1 and verified across the remaining 40× span of horizons, the volume-time price path is diffusive to within 8%. This is a stronger and more portable statement than any individual failed test, because it does not depend on which features were tried.

4. Cost floors

Every result was scored against five candidate cost floors rather than one, because the venue's published taker rate is inconsistently reported. The cheapest conceivable floor is a 3.60 bps round trip. The best edge found at the only overlap-free horizon was 1.78 bps — short by a factor of two — and it did not survive the null in any case, so the cost comparison is the second line of defence rather than the first. The full floor arithmetic is in *Impossible Before We Started*.

5. The four attempts

5.1 Four-gate conjunction (registered, scored)

Suspended on a pre-registered insufficient-events condition: 399 decision buckets, zero distressed rows, zero completed trades against a floor of thirty. The conjunction never fired — VPIN ceilings at `0.1821` against a `0.42` entry gate, unreachable by a factor of 2.3 before any other condition was evaluated.

5.2 Bucket-size "measurement repair" (drafted, never tagged, withdrawn)

A second specification proposed shrinking the volume bucket while holding the entry gate fixed, on the stated ground that lowering the gate would have been the illegitimate response. The two turned out to be the same knob, and the repair would have destroyed information rather than recovering it. That result is its own study: see the companion paper *The Same Knob Twice*, which measures VPIN against bucket size on this tape with an uninformative-labels null.

5.3 Percentile / distress-score architecture

Equal-weight percentile ranks of four features, signed by one of them. Measured end to end at h = 1 it returns **+0.11 bps**. Its one strong-looking cell — nominal t = 5.16 at a seven-hour horizon — dissolves on honest counting: 99 entries resolve to 13 non-overlapping episodes (episode t = 1.53) and 99 shorts to 14 (episode t = 1.94), with individual episodes running to −100.3 bps. A nominal 5.16 was 194 restatements of 27 events.

Two structural defects are worth recording independently. The gravity feature never exceeds 1.7% of its nominal range on this window, so percentile-ranking inflates a near-constant series into a full spread. And the feature carrying the entire directional decision returns hit rates of 47.1%, 55.0%, 58.5%, 48.9% and 44.3% across its five conditioning variants — indistinguishable from a coin — while a natively signed feature is discarded by taking its absolute value.

5.4 Cumulative volume delta and real-time delta

CVD is a monotone transform of signed order flow already present in the feature set. Measuring it cumulatively, across three windows, and as a price/CVD divergence does not create information that is not in the tape. Across 154 cells, **43.5% had positive edge — below the 50% a coin would give.**

> A NOTE ON DIVERGENCE TESTS
>
> Divergence is structurally incapable of producing evidence. When price and CVD disagree, following CVD is by construction the exact negation of following price — measured literally here as `+2.654 bps` and `-2.654 bps`. One side is always positive, so a study reporting the winning reading *always* finds a result. It means something only if a side beats its null. Neither does.

6. Two corrections

6.1 VPIN volatility finding — withdrawn

An earlier reading noted that conditional |return| rises monotonically with VPIN extremeness — the indicator doing what its authors actually claim for it, predicting volatility rather than direction — and argued that such ordering across nested cut-points is not easily produced by a noise search.

STATISTIC	OBSERVED	NULL P50	P
Monotonicity ρ (geometry of record)	1.000	0.000	0.143
Lift gap, most extreme – broadest	0.252	-0.001	0.140

The null produces that ordering 14% of the time. The reasoning was wrong for a structural reason: the cut-points are *nested* (top 2% ⊂ top 5% ⊂ top 10%), so they are not independent tests and monotonicity is cheap. **Withdrawn.** It was the only quantity in the programme that leaned the right way, and it still does not clear its own null.

6.2 Fee clarification

A working assumption of a 0.04% taker rate was checked against the venue. Binance USDⓈ-M Futures VIP 0 is **maker 0.0200% / taker 0.0500%**; 0.04% is the legacy rate and still circulates in secondary sources. Because the discrepancy was live during analysis, all five floors were carried and no conclusion depends on which is correct.

7. What would have to be true

Closure is only useful if it is falsifiable, so the conditions under which this conclusion would flip — the directional accuracy required at each horizon, the fee tier, and the latency budget — are measured in the companion paper *Impossible Before We Started*. Its headline: at a five-minute horizon the round-trip fee exceeds the entire average move, so one whole class of strategy is excluded by arithmetic rather than by evidence.

8. Reusable constants

CONSTANT	VALUE
Search null floor	`median episode-corrected t = 3.405; p95 5.955; p99 8.454 (854-cell grid, 1,940 rows)`
Volume-time diffusion	`√t to within 8% across a 40× horizon span`
VPIN geometry law	`VPIN ~ 1/√V` — see *The Same Knob Twice*
Venue latency and cost	see *Impossible Before We Started*

The first row is the one that carries forward furthest. It is not a property of this strategy: it is a property of searching this dataset with a grid this size, and it applies to work that has not been started yet.

9. What is not closed

The branch is closed on *directional* alpha from microstructure features at retail fee tiers. It is not closed on whether anything is on this tape.

Maker-side adverse selection is untested. Every measurement here assumes crossing the spread. The naive spread-capture arithmetic is already unfavourable at VIP0 — a 0.548 bps spread against a 4.00 bps maker round trip — but that is not the maker question. The maker question is: *conditional on a resting order at level k being filled, what does the price do next?* That quantity is the missing input for evaluating any passive strategy at any fee tier, including rebate tiers where the spread-capture arithmetic inverts. It has not been measured, and this tape supports measuring it directly. Nothing in this document licenses a claim about it in either direction.

10. The held-back data

Approximately 24 days and 54.2 million trades remain sealed and were never touched by any diagnostic here. The reason they are not being spent is precise, and it is not that we ran out of ideas:

> **No specification survived its own null, so there is nothing worth confirming out of sample.**

Out-of-sample data has exactly one legitimate use: confirming a specification fixed in advance. Spending it to re-confirm a negative destroys the only clean test set the programme has and buys nothing, since a second failure adds no information once the first has been shown indistinguishable from noise. The data remains available for the maker question, which is a different specification and has not been searched.
