THANAAYA TECHNOLOGIES

RESEARCH NOTE

The Same Knob Twice

How a parameter repair and a threshold change turned out to be the same action, and the test that told them apart.

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: VPIN bucket geometry and threshold discipline, SOL perpetual futures
Programme: internal quantitative research — no client capital at any stage
Status: specification withdrawn before it was registered
Published: September 2026 · **Licence:** Dubai Department of Economy and Tourism 964356 · **Unified licence:** BL1484

PART ONE

In plain English

No mathematics in this section. It is about a way of accidentally cheating that looks exactly like fixing a broken instrument.

1. The rule that never fired

We had a rule for trading **SOL perpetual futures on Binance**: only take a position when a gauge called VPIN — a published academic measure of how much informed trading is going on — reads above **0.42**.

The number 0.42 was written down before we looked at any data. That is the whole point of writing it down: you cannot then move it to suit the result.

Over a full week of trading — 5.7 million trades — the gauge never once got above **0.18**. Not close. Less than half the level required.

0.42	0.18	0
reading required before the rule would trade	highest reading in the entire week	trades taken

So the strategy produced nothing. No wins, no losses, no evidence in either direction. And that kind of failure comes with an obvious temptation attached.

> THE TEMPTATION
>
> Lower 0.42 until something happens. Everyone knows that is cheating, so nobody does it openly. What they do instead is find a *principled-sounding reason* to change something else that has the same effect.

2. The fix that sounded principled

The gauge works by looking at a window of recent trading and measuring how lopsided the buying and selling were. Our window covered about **22 hours** of trading.

We proposed that this was too wide — that averaging over most of a day was smoothing away exactly the short bursts we wanted to catch. Narrowing it to about four hours would be a *repair to the instrument*, not a change to the threshold. The 0.42 would stay untouched, and we said so explicitly, in writing.

That argument is wrong. It is not obviously wrong, which is why it took a deliberate test to see it.

3. Why it is the same dial

Here is the thing that defeats it: **even in a market with no information in it at all, buying and selling never come out exactly even** — and the smaller the window, the more lopsided it looks by pure chance.

Flip a coin a million times and you get very close to half heads. Flip it four times and you often get three. The gauge cannot tell genuine lopsidedness from small-sample lopsidedness. So narrowing the window raises the reading whether or not anything real changed.

1.0 What the gauge reads if you narrow the window to a single trade. Every window is then 100% buy or 100% sell, so it reports maximum informed trading on a market with nothing happening in it.

Narrowing the window and lowering the threshold are **the same action described two ways**. One looks like tuning; one looks like a repair. They move the result in the same direction for the same reason.

4. The test

We built a fake version of the same week. Every trade kept its exact time, price and size — we changed only the record of whether each was a buy or a sell, shuffling those labels at random. The trading is real; the thing the gauge claims to detect is gone.

Then we ran the gauge on both versions across nine window widths, from 44 hours down to 13 minutes.

5. What came back

First, the mechanism is confirmed. On the shuffled week — where there is definitionally nothing to detect — the gauge still climbs steadily as the window narrows, from 0.02 to 0.27. That climb is pure arithmetic.

Second, and this settled it: narrowing the window makes the gauge *worse*, not sharper.

WINDOW WIDTH	REAL READING	READING ON SHUFFLED DATA	HOW MUCH IS REAL
22 hours *(the setting we were going to abandon)*	0.155	0.034	4.6×
4 hours *(the proposed "repair")*	0.258	0.068	3.8×
13 minutes	0.652	0.269	2.4×

THE FINDING

The original 22-hour setting — the one we were proposing to throw away as too wide — had **the best signal-to-noise ratio of every setting we tested.** The "repair" would have discarded information while appearing to sharpen the instrument, and raised the reading enough to make the rule fire on arithmetic alone.

6. What it would have cost

Nothing dramatic, and that is the point. It would have produced a strategy that traded, on a threshold that was still technically the pre-registered 0.42, using a gauge whose elevated readings were a property of the measurement rather than the market. Every document would have been consistent. It would have looked disciplined. And it would have been wrong.

The test took about an hour. We ran it before making the change, so the change was never made.

7. The general lesson

When a rule fixed in advance fails to fire, the honest options are to accept it or to run a genuinely new rule and count it as a new attempt. What is not available is changing a setting that is mathematically tied to the threshold and calling it a repair.

The way to tell the difference is not argument. Build a version of the data with the effect removed, and check whether your proposed change moves the number there too. If it does, the change is not measuring anything.

PART TWO

The technical record

The indicator is VPIN. The scaling argument, the permutation null and the full bucket-size ladder, as measured.

1. Subject

Indicator	VPIN — Volume-synchronised Probability of Informed Trading (Easley, López de Prado & O'Hara)
Instrument	SOLUSDT perpetual, Binance USDⓈ-M
Window	`2026-08-10 → 2026-08-17 UTC · 5,702,722 trades · 79,795,693 SOL`
Entry gate under test	`VPIN ≥ 0.42, fixed in advance and never moved`
VPIN window	`n = 50 buckets throughout; only bucket volume V varies`

2. The scaling argument

VPIN over n trailing equal-volume buckets of size V is

$$VPIN = \Sigma\ |V_{buy} - V_{sell}|\ /\ (n \cdot V)$$

Under order flow with any independent component, $E|V_{buy} - V_{sell}|$ grows like √V. The ratio therefore grows like 1/√V: **VPIN is a decreasing function of bucket size irrespective of information content**, and tends to 1.0 in the one-trade limit by construction. This is the substance of the Andersen & Bondarenko critique, and it is the reason the original specification read VPIN through its own trailing distribution rather than as a raw level.

3. Null construction

The null is a random permutation of the aggressor flags (seed 7). Timestamps, prices and sizes are untouched, and the marginal buy rate is preserved by construction, so bucket boundaries and bucket count are identical between the real and null tapes. The only thing destroyed is the sequencing of buy and sell — precisely the quantity VPIN claims to measure. Whatever VPIN the null produces is bucket arithmetic.

4. The ladder

VPIN against bucket size, real tape and permuted-label null. Highlighted rows are the two registered geometries.

V (SOL)	50-BUCKET WINDOW	REAL MEDIAN	NULL MEDIAN	REAL / NULL	REAL MAX	% ≥ 0.42
400,000	43.7 h	0.1153	0.0249	4.63	0.1568	0.00%
199,500	**21.8 h**	**0.1554**	**0.0338**	**4.60**	**0.1821**	**0.00%**
100,000	10.9 h	0.1954	0.0459	4.26	0.2417	0.00%
60,000	6.6 h	0.2232	0.0567	3.94	0.3035	0.00%
39,500	**4.3 h**	**0.2577**	**0.0683**	**3.77**	**0.3717**	**0.00%**
20,000	2.2 h	0.3163	0.0940	3.36	0.4963	5.81%
10,000	1.1 h	0.3990	0.1320	3.02	0.6394	35.93%
5,000	0.55 h	0.4990	0.1815	2.75	0.7965	89.37%
2,000	0.22 h	0.6516	0.2693	2.42	1.0000	99.94%

5. Three conclusions

5.1 The revised specification was already known to fail

At V = 39,500 the maximum VPIN over 1,970 buckets is **0.3717** against a 0.42 gate — zero crossings. The trial would have been suspended on the same insufficient-events condition as the first, and this was knowable before the specification was tagged. It was withdrawn as a draft rather than run.

5.2 The threshold discipline was already defeated

Reaching 0.42 with any regularity requires V ≈ 10,000, at which point 35.9% of buckets cross. Selecting a bucket size to make a fixed gate reachable is a threshold walk conducted through a different parameter. Holding 0.42 constant while choosing V is not discipline; it is the appearance of discipline.

5.3 The narrowing destroys information

The real-to-null ratio falls monotonically from **4.63 to 2.42** as V shrinks across the ladder. At the 21.8-hour window, roughly 78% of the reading sits above the no-information floor; at the 0.22-hour window only about 59% does. The original geometry — approximately one day, which is the canonical construction in the source literature — had the highest signal-to-arithmetic ratio of every setting tested. The proposed 4.3-hour window is five times shorter than the specification cited as its own authority.

6. What this does not say

It does not say VPIN is worthless, and it does not say the 0.42 gate was correctly chosen. A separate test on the same tape found that VPIN's conditional-volatility ordering — its original published claim — fails its own null at p = 0.143, and that result is recorded as withdrawn in the companion paper *Measuring the Noise Floor*. What this paper establishes is narrower and more durable: **the level of VPIN is not comparable across bucket sizes, so any threshold on the raw level is a statement about the geometry as much as about the market.**

The practical consequence is that VPIN should be read through its own trailing distribution — a percentile, not a level. That is what the original authors do, and it is what the implementation already computed and left unused.

7. Reusable constant

CONSTANT	VALUE
VPIN geometry law	`VPIN ~ 1/√V; permuted-label null median 0.0249 → 0.2693 as V falls 400k → 2k; real:null ratio 4.63 → 2.42`

Any VPIN threshold quoted without its bucket size is unfalsifiable. The pairing is the specification.