

Cross-Source Price Integrity: A Reproducible Measurement Method

Delta Motion Analytics — Methods & Reproducibility Whitepaper

Version 1.0 · September 2026

1. What this document is (and is not)

This is a methods-transparency document. Its purpose is to let a technically-fluent reader understand *how* we measure cross-source price integrity — the exact sources, windows, statistics, and null construction — well enough to judge the work, without exposing the engine's source code.

It is not a sales sheet, a product manual, a trading signal, a securities research report, or a recommendation. It makes no claim about future prices or returns and gives no advice on what to buy, sell, or allocate. It measures whether reported price data is *genuine*, not whether it is *profitable*.

2. The problem we measure

An index or reference-rate issuer, an exchange, or a data vendor publishes a price assembled from multiple venues. A fraction of the inputs is not genuine — phantom or stale quotes, outlier prints. Contaminated inputs corrupt the published figure, and the issuer typically cannot independently demonstrate to a third party — an institutional client, an auditor, a regulator operating under IOSCO or benchmark-integrity expectations — that its inputs are clean or its consensus price is defensible.

(Non-economic or "wash" volume is a real, related contamination in these markets, but authenticating volume is a distinct capability this assessment does not perform — see the Roadmap, §10. This document does not claim to measure volume integrity.)

Existing tooling addresses adjacent problems, not this one:

- Pipeline observability watches schema, freshness, volume, and lineage — not whether a cross-source price is genuine.
- Surveillance and anomaly ML flag suspicious activity via black-box models on paid feeds — without a transparent, reproducible, honest-null method a third party can verify.
- Index administration constructs a rate on its own governed methodology — it is the thing being measured, not an independent second opinion on it.

The position this document occupies: an independent, reproducible, method-transparent measurement of cross-source price integrity, on free public data, with honest reporting of nulls.

3. What we measure

- 1 **Consensus / fair-price reconstruction** — a defensible reference price across a named set of venues.
- 2 **Cross-source divergence and lead-lag** — how far sources sit apart, and which venue leads, quantified.

- 3 **Artifact flagging** — phantom, stale, and outlier prints identified against a proper statistical null, with rates reported.
- 4 **Null-validated throughout** — every claim is tested against what chance alone would produce; where there is nothing genuine to find, we say so (§6).

Volume authenticity is not among the demonstrated capabilities (§10).

4. Method transparency (the reproducibility core)

Method is disclosed faithfully; no engine source code is exposed.

4.1 Data sources. Public 1-minute BTC data from first-party exchange APIs: Binance and Coinbase (the headline efficiency and convergence pair), with Kraken as the third venue in the live lead-lag measurement and Kraken USDT/USD as the de-margining rate. This is single-provider *derived* analysis, not redistributed raw data.

4.2 Sampling and windows. Six independent 7-day windows spanning January–August 2026, fixed in advance by a mechanical 42-day-step rule (each window at least 35 days clear of the next), with approximately 10,000 aligned 1-minute observations per window on the Binance–Coinbase pair. The lead-lag is a single live tick window across the three venues — not a historical or replicated measurement.

4.3 The statistics.

- *Divergence*: the per-minute basis-point deviation of each venue from the cross-venue consensus (the row-wise median), after de-margining USDT to USD; the headline pair is $(\text{Binance\_USD} - \text{Coinbase}) / \text{Coinbase} \times 10^4$.
- *Lead-lag*: Hayashi–Yoshida cross-correlation against a clock-shift grid on asynchronous ticks; the peak shift is the lead. This is the correct estimator for asynchronous ticks — naïve fixed-interval correlation is blind to asynchronicity and the Epps effect.
- *Artifact detection*: a frozen/stale mask, a single-venue corroboration requirement, and a low-information-minute mask.

4.4 Null construction and decision rule. An increment-shuffle surrogate: the gap's first differences are shuffled to build a random-walk surrogate, preserving the step-size distribution while destroying mean-reversion. The decision bar is pre-registered: $z \geq \sim 3$ and the real statistic $\geq \sim 1.5 \times$ the surrogate. Convergence parameters (threshold, horizon, shrink) are frozen before the windows are seen.

4.5 Phantom, stale, and outlier guard. A venue reprinting its prior price is flagged stale. A single-venue deviation survives only if the other venues corroborate it as the lone outlier (tolerance-banded). Materiality is set by round-trip cost bands (retail, mid, institutional). Low-information minutes are masked and the test is re-run.

4.6 Independent reproduction. The delivered case-study was reproduced against a genuinely different generative family (jump-diffusion with stochastic-volatility clustering, Student-t innovations, and bid-ask-bounce microstructure), importing the fixed core and the published harness unchanged; the cross-exchange result survived. This different-family campaign covered the delivered case-studies only. The methodology nulls in §6 are not reproduced by this campaign; each is instead validated by its own offline differential self-test (see §6), and the two claims are kept strictly separate. One boundary must be stated plainly: supra-threshold bid-ask bounce can fake convergence on a less-liquid, wider-spread pair — the ~ 1.5 bps convergence threshold is what protects the liquid-BTC result. The guard's protection is

threshold- and liquidity-dependent, not universal.

5. Worked demonstration — cross-exchange price discovery (delivered case-study)

The headline result is a two-venue (Binance–Coinbase) measurement; Kraken enters only for de-margining and the live lead-lag.

5.1 Setup. BTC, 1-minute, six independent 7-day windows across January–August 2026 (\$4.2), Binance–Coinbase headline pair, de-margined via Kraken USDT/USD.

5.2 Divergence and efficiency. The de-margined Binance–Coinbase gap sits at approximately 1 bp every window. The retail and mid round-trip cost bands are cleared in roughly 0 of ~10,000 minutes across all six windows; the institutional band is touched only rarely and never in bursts (0–38 isolated minutes per window, at most ~0.4%).

5.3 Convergence. The gap mean-reverts far above the ~2% random-walk surrogate in all six windows (real $\approx$ 20–52%, tens of σ ; window 1 reproduces the original ~36%), and survives the low-information-minute mask in every window. The honest reading is that this is a real cross-venue binding force that is provably too small to trade — a monitoring and integrity-grade property, not a tradable signal. A genuine binding force exists *and* it is sub-cost, and that combination is exactly what makes it an integrity metric rather than an edge.

5.4 Lead-lag. All three venues move together; the measured lead sits at the estimator's resolution floor (~0.25–0.5 s, with Binance marginally first) — no economically meaningful lead. This is a live re-confirmation on a single window, not a temporally replicated finding (free tick history is unavailable), and is not presented as replicated.

5.5 Artifact rates. Reported as the cost-band clearing rates and mask-survival above (§5.2–5.3).

5.6 Reproduction. The cross-exchange result reproduced under the different generative family (§4.6), carrying the bid-ask-bounce and liquidity-threshold boundary.

6. Honest nulls — the instrument tests (methodology)

These are methodology results, not delivered client work. Each is validated by its own offline differential self-test — a known signal, a known artifact, and a placebo are injected, and the engine must keep the real one and reject the others. This is the correct instrument test for a methodology null; it is separate from the different-family reproduction campaign in §4.6.

6.1 Stablecoin peg integrity. Peg integrity resolves into separable properties, and the engine reports them independently. Cross-venue agreement is tight and sub-cost: over a ~12-hour two-venue window, Kraken and Coinbase agreed on USDT's price to within about half a basis point, matched sign on every larger deviation (497 of 497), and the cross-venue gap never came within reach of round-trip trading costs; USDC held inside ± 1 bp all window. But the peg *level* itself is a drifting, non-stationary discount, not a self-correcting one: over a ~7-day Coinbase window (~10,000 aligned minutes) USDT averaged about –4 bps and narrowed one-directionally from roughly –9 bps to –2 bps; a pre-registered augmented Dickey–Fuller test could not reject a unit root ($p \approx 0.13$), and the demeaned deviation reverted no faster than an increment-shuffle surrogate. One week of narrowing is an observation, not a proven reversion-to-peg force. This is integrity monitoring — venue concurrence and sub-cost gaps — not a demonstrated peg-restoring mechanism, and not a trade. (DAI was dropped before

analysis — Coinbase-delisted, Kraken ~97% zero-volume — a case of the instrument correctly rejecting a dead market.)

6.2 DEX-vs-CEX lead-lag. Binance ETHUSDT trades as the continuous reference against a Uniswap v3 pool sampled once per block; Hayashi-Yoshida lead-lag runs on asynchronous ticks, with the block cadence as a hard resolution floor and a per-window block-quantization control. On Ethereum mainnet (~12 s blocks) the floor is too coarse to resolve a one-block lead — a bounded null by construction. Escalated to Base (~2 s blocks, sixfold tighter): across eight powered windows the on-chain price tracked the exchange contemporaneously — every window's lead at or below the 2 s floor (0–2 s), consistently CEX-first (the correct direction) but never sharp or large enough to clear the gate. Zero of eight resolved: the DEX adjusts to the CEX inside a single block, a correctly-signed tendency below the resolution floor rather than a resolvable lead. Mid-build, the engine caught and discarded one of its own spurious readings — a truncated window that briefly faked a 33-second DEX lead — before it reached any record.

6.3 Triangular no-arbitrage. Six pre-registered one-week windows (January–August 2026, mechanical 42-day step), ETHBTC against ETHUSDT ÷ BTCUSDT on Binance 1-minute closes — 10,080 aligned minutes per window, 60,480 in total. The residual (traded minus implied, in bps) is tested against two nulls with all parameters fixed in advance: an increment-shuffle surrogate and a tick-quantization null (the implied rate rounded to the direct leg's ~3.2 bp tick). The identity holds to within one quote tick and is decisively untradable: the largest single excursion ran 6.2–13.8 bps across the six windows and never reached the mid (15 bps) or retail (30 bps) fee band in any of the 60,480 minutes; only the fee-only institutional band was touched (at most 0.8% of minutes in the most volatile week, 0.1% or less elsewhere), reverting mostly within three minutes; a spread-charged institutional band cleared on three minutes in the entire study. The apparent mean-reversion is tick quantization, not a binding force: the residual reverts ~42–43% per window versus ~2% for the random-walk surrogate (tens of σ), but the quantization null reverts *more* (46–52%) — so the real residual reverts slightly *less* than a coarsely-rounded shadow, failing the stricter null in all six windows. The honest statement is that the identity is consistent to within a tick, not dynamically enforced faster than that tick.

Why nulls belong in a trust document. A method that only ever "finds something" is unfalsifiable. Publishing the cases where the engine correctly finds nothing — or catches and rejects its own spurious reading — is the strongest available evidence that a positive finding is real and not manufactured. It is the direct answer to "how do I know this isn't cherry-picked."

7. What this is not

- Not an index administrator. We do not build or sell your rate; we independently measure whether it is genuine.
 - Not a data-observability tool. We do not monitor your pipeline; we measure whether the price is real.
 - Not black-box ML surveillance. We pre-register, reproduce on free public data, and publish our nulls; the method is auditable.
 - Not a trading signal or securities research. No prediction of prices or returns; no buy, sell, hold, or allocation; no "alpha." Measurement of data integrity only.
-

8. How an engagement works

- 1 **Free first read** — an integrity snapshot on a subset of your public feed, at no cost and no commitment, so your team sees the method live.
- 2 **Integrity assessment** — a fixed-scope, one-time assessment on defined venues, pairs, and windows, delivered as a report you can hand to your auditor, institutional client, or regulator. It includes a per-engagement reproducibility appendix (sources, windows, statistics, null — no engine code).
- 3 **Monitoring** (*optional, later*) — a recurring integrity read once an assessment has landed.

Deeper-scope engagements (tick-level, more venues, intraday) use client-supplied or engagement-specific data, procured only for that engagement.

9. Scope and limitations

- 1 Volume authenticity is not demonstrated — volume enters only as a stale/low-information filter input, not as a measured capability.
 - 2 The lead-lag is a single live window, not replicated, measured at the estimator's resolution floor; there is no economically meaningful lead.
 - 3 The headline result is a two-venue (Binance-Coinbase) measurement; three-venue consensus is used only in the phantom stage and the live lead-lag.
 - 4 The increment-shuffle null certifies convergence-beyond-random-walk, not tradability — the point is that the convergence is sub-cost.
 - 5 The artifact-guard's protection is threshold- and liquidity-dependent (the bid-ask-bounce boundary; the ~1.5 bps threshold protects the liquid-BTC result).
 - 6 The \$6 nulls are validated by offline differential self-tests, not by the different-family reproduction campaign (which covered the case-studies).
 - 7 This is exploratory public-data analysis, single-provider derived — measurement, not investment advice.
 - 8 "No exploitable edge" is a null (a successful instrument test), not a forward-looking guarantee.
-

10. Roadmap — not demonstrated in this document

Volume authenticity (distinguishing genuine from non-economic volume) is a distinct capability, not built or demonstrated here. It is the natural next proof-point and, once independently validated, the basis of a future cross-source *volume* integrity extension. Nothing in this document claims it.

Delta Motion Analytics is a cross-border analytics studio providing measurement and reproducibility infrastructure. This document describes method, not source code, and does not constitute investment advice or a recommendation regarding any security or asset.