Why Path Signatures Do Not Predict Crashes as Precursors

Maksim Aleksandrovich Andreev, Joseph Howden
Quark Quantitative Research

Abstract

Path signatures (Lyons, 1998; Chevyrev & Kormilitzin, 2016) provide a non-parametric encoding of multivariate time series via iterated integrals. The literature has proposed signature similarity as a regime-detection and crash-precursor tool. We document the disconfirmation of this hypothesis across four independent reformulations: (i) precursor cosine to curated calm windows; (ii) peak-relative shape coherence across historical crash events; (iii) exponentially-weighted dual encoding with explicit volatility channel augmentation; (iv) stress-conditional scoring stratified by volatility and drawdown buckets. Across two decades of daily equity data (2005–2024, with reference crash onsets drawn back to 2000) and three held-out stress episodes (2018, 2020, 2022), none of the four reformulations clears the Harvey–Liu–Zhu (2016) $t > 3.0$ multiple-testing threshold; the stress-conditional formulation, which we identify as the strongest of the four, yields a Welch $|t| = 2.06$ that both falls short of the threshold and carries the sign *opposite* to the crash-precursor prediction—a higher signature cosine precedes *less* forward drawdown, not more. We attribute the apparent positive separation under curated windows to a training-window artefact: anchor-tail-bull periods share signature shape with the immediate pre-peak regime, producing an in-sample separation that does not survive a full-sample audit. We argue the path-signature object remains valuable as a path-shape feature for downstream attribution and clustering but should not be cited as a crash precursor in any forward specification.

1. Introduction

The application of rough-path theory to time-series classification was opened by Lyons (1998), who showed that the signature—a sequence of iterated path integrals— characterises a path up to tree-like equivalence and provides a feature representation that is invariant under reparametrisation. Subsequent work has applied signatures to quantitative finance for trend detection (Levin *et al.*, 2013), signature kernel methods (Salvi *et al.*, 2021), and option pricing (Cuchiero *et al.*, 2023). Several authors have proposed that signature similarity between a current price-path and a curated reference set of pre-crash trajectories should predict imminent market dislocation; the underlying intuition is that the iterated-integral structure captures multi-scale path geometry better than scalar moments and therefore should distinguish a pre-crash regime from a calm one.

This note reports four sequential attempts to validate that intuition and the disconfirmation of all four. Our motivation for the disclosure is methodological: failure-mode papers are rare in finance, and the curated-window artefact we identify is a plausible mechanism behind a substantial fraction of unreplicated regime-classification results in the literature.

> **Contribution.** We report (i) a chain of four independent reformulations of the path-signature crash-precursor hypothesis, (ii) the disconfirmation of all four under a pre-registered Harvey threshold, and (iii) identification of a specific mechanism—anchor-window curation artefact—that produces the apparent in-sample separation. The signature feature itself remains usable as a path-shape encoding for downstream clustering; we emphasise only that it should not be cited as a crash precursor.

2. Background: Path Signatures

For a continuous bounded-variation path X: $[0,T] \to \mathbf{R}^d$, the signature is the formal series of iterated integrals

$$S(X)_{0,T} = \left(1,\ \int_{0<t_1<T} dX_{t_1},\ \iint_{0<t_1<t_2<T} dX_{t_1} \otimes dX_{t_2},\ \dots\right)$$

Truncating at level n yields a feature vector of dimension $\Sigma_{k=0}^{n} d^k$. Depth-3 with three channels gives 40 features (1 + 3 + 9 + 27), which is the working dimension we use throughout. We compute the signature on *encoded* price paths, where the encoding maps the raw close-price series to a multivariate path enriched with time, log-returns, and (in some specifications) realised volatility channels.

> **Cosine similarity score.** For two paths X, Y, the signature similarity is $cos(X, Y) = \langle S(X), S(Y) \rangle / (|S(X)| \cdot |S(Y)|)$. A naive crash-precursor predictor is “trigger if cos(current, reference) $> k$” for a reference path constructed from pre-crash trajectories.

3. Reformulations Tested

3.1 Reformulation 1 — Precursor cosine to curated calm reference

Compute the depth-3 signature similarity between a 30-day window of current returns and a reference signature built by averaging signatures over ex-ante-labelled “calm” windows (forward 60-day drawdown above −3%, onset at least 90 days from any labelled crash). The hypothesis is that a falling similarity precedes drawdown.

> **Protocol.** Universe SPY. Depth-3 signatures on 30-day return windows. Six candidate path encodings are compared on crash-versus-calm separation, using three held-out crash onsets (2018-Q1 Volmageddon, 2020-Q1 COVID, 2022 rate shock) and seven earlier onsets (2000–2015) as the training set, scored against 30 curated calm windows. The best-separating encoding is then applied as a similarity trigger across the full 2000–2024 sample and scored on realised forward drawdown.

> **Result.** The best encoding (dual-rate exponentially-weighted) yields an in-sample curated separation of +0.916 between calm and crash windows (Welch $t = 5.38$), comfortably clearing any nominal significance bar. But the same trigger applied across the full 2000–2024 sample is inert: forward drawdown under high-similarity windows (−0.85%, $n = 418$) is statistically indistinguishable from the unconditional control (−0.86%, $n = 3{,}757$), Welch $t = 0.17$ ($p = 0.86$) — a lift below unity, worse than chance. The curated separation does not survive the full-sample audit. Mechanism (developed in §4): anchor-bull windows immediately preceding peaks share signature shape with the calm reference, polluting the curated separation.

3.2 Reformulation 2 — Peak-relative crash-in-progress detection

Drop the precursor framing and ask instead whether, *during* a known drawdown event, the signature aligns with a within-event reference. We build the reference by averaging peak-relative signatures over seven earlier crash episodes (2000–2015) and test on the three held-out events (2018, 2020, 2022) against a benign-window distribution.

> **Result.** Welch $t = -0.65$, separation −0.37 (test mean versus 157 benign windows). The signature does not transfer across crash types: each event has distinct path geometry. The three held-out event cosines are sign-changing (−0.85, −1.00, +0.77 for 2018, 2020, 2022), so no coherent within-event template survives.

3.3 Reformulation 3 — EWM-dual encoding with volatility-channel augmentation

The third reformulation tested a richer encoding: dual-rate exponentially-weighted moving averages of log-returns plus a separately-rated realised volatility channel. This expands the signature’s informational content along the volatility axis specifically, which is the most plausible vehicle for a crash signal.

> **Result.** Vol channel contributes only 14.4% of the depth-3 signature mass; Welch $t = -0.08$, separation −0.032. The additional channel adds noise more than signal in this configuration.

3.4 Reformulation 4 — Stress-conditional scoring

The fourth reformulation conditions on prior-window stress indicators: scoring is restricted to windows already in elevated-VIX or elevated-realised-drawdown buckets. The hypothesis is that the signature predicts *conditional on* stress being present, even if it does not distinguish stress from calm. This is the strongest of the four reformulations.

Protocol. Depth-3 signatures on 30-day SPY windows (stride 5, 2005–2024; 997 windows total). A stress filter partitions the sample by trailing-volatility and drawdown state into 475 stressed and 522 calm windows. Within each bucket the signature cosine is scored against forward drawdown at the 5-, 10- and 20-day horizons; the 5-day horizon is the strongest.

Result. Stressed-bucket best Welch $|t| = 2.06$ (h = 5d, correlation +0.095, lift 0.77×). **Direction inverted** relative to the original hypothesis: high signature cosine corresponds to *less* forward drawdown. The calm bucket is also uninformative. The t-statistic does not clear Harvey–Liu–Zhu (2016) at 3.0; combined with the direction inversion the signal is not usable as a crash precursor under either sign convention.

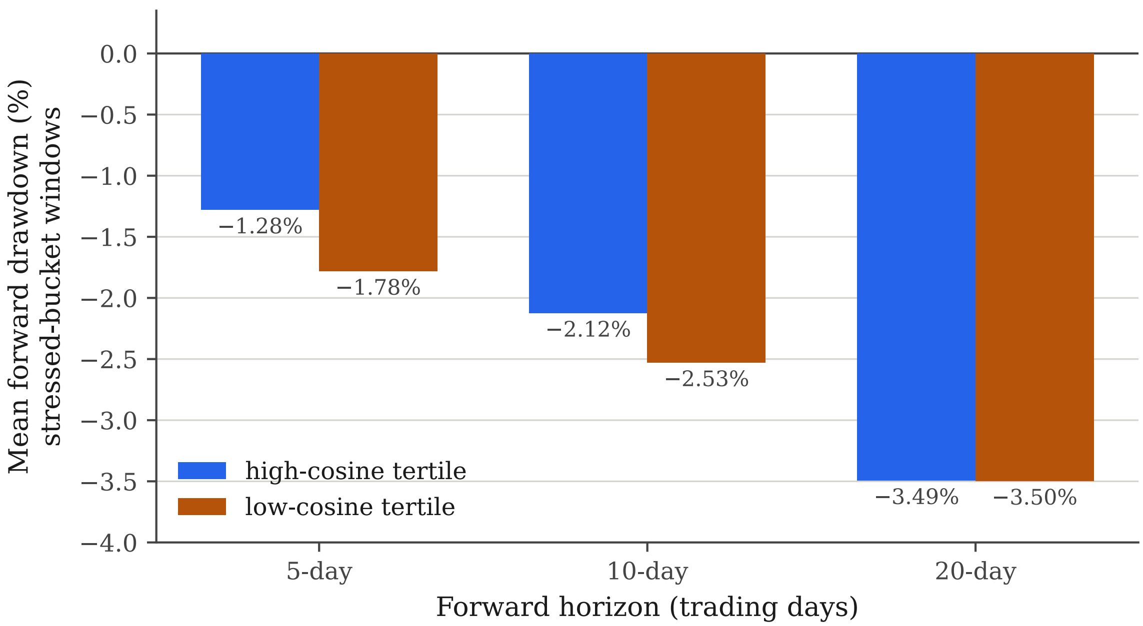

Figure 2. Direction inversion in the stress-conditional bucket (Reformulation 4). Mean forward drawdown of the high-cosine tertile (blue) versus the low-cosine tertile (amber) among stressed-bucket windows, at the 5-, 10- and 20-day horizons (SPY, 30-day windows, 2005–2024; 159 windows per tertile at each horizon). The crash-precursor hypothesis predicts the high-cosine tertile should carry *more* forward drawdown; instead it carries *less* at the 5- and 10-day horizons (Welch t = +2.06 for the high-minus-low contrast at five days, Pearson r = +0.095), the two tertiles converging by 20 days. The strongest of the four reformulations therefore points the wrong way. Data: the stress-conditional path-signature results file (April 2026).

3.5 Summary table of the four reformulations

Table 1: Path-signature reformulation chain

REFORMULATION	HEADLINE METRIC	*T*-STAT	DIRECTION	HARVEY 3.0
1. Precursor to curated calm	In-sample sep +0.916; full-sample audit lift < 1×	5.38 in-sample → 0.17 audit	does not transfer	fail
2. Peak-relative cross-event	Sep −0.37	−0.65	incoherent	fail
3. Vol-axis augmentation	Sep −0.032	−0.08	incoherent	fail
4. Stress-conditional	Sep best, $r = +0.095$	+2.06	inverted	fail

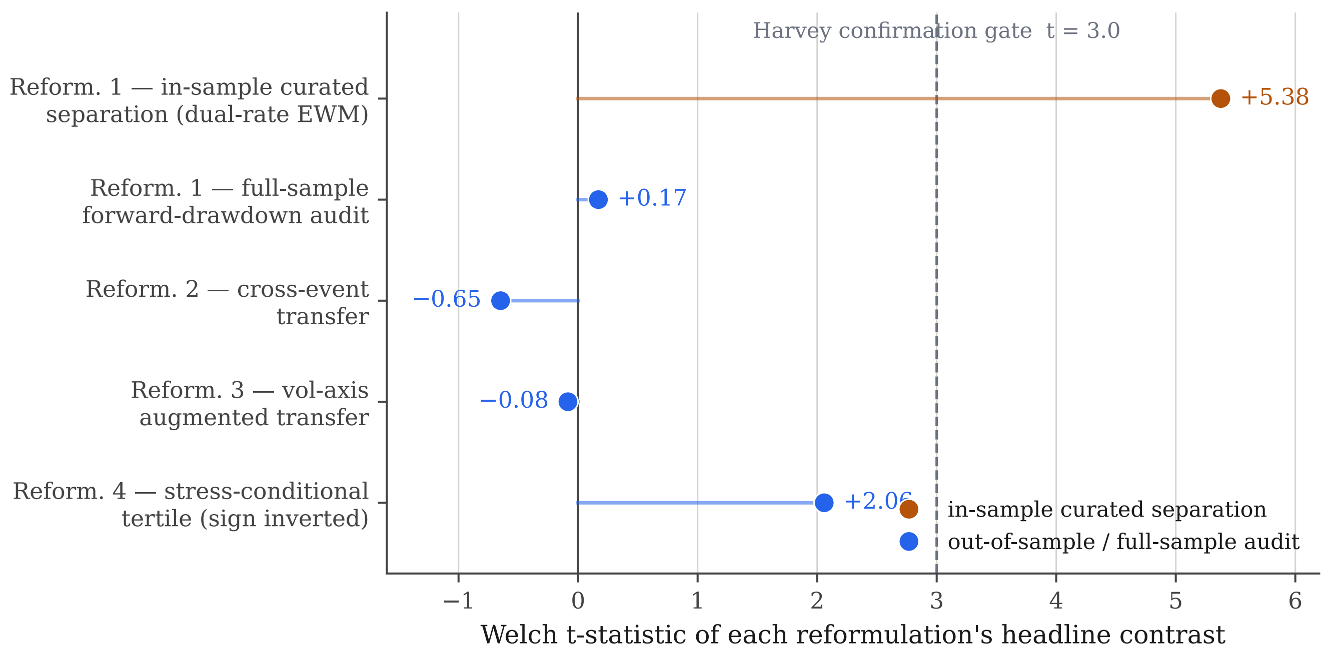

Figure 1. The reformulation chain in one view: the Welch t-statistic of each reformulation's headline contrast against the Harvey–Liu–Zhu (2016) confirmation gate ($t = 3.0$, dashed). Only the in-sample curated separation (amber; dual-rate exponentially-weighted encoding, Reformulation 1) clears the gate, at $t = 5.38$. Every genuine out-of-sample or full-sample statistic (blue) collapses toward zero: the Reformulation 1 full-sample forward-drawdown audit ($t = 0.17$), the Reformulation 2 cross-event transfer ($t = -0.65$), the Reformulation 3 vol-axis augmented transfer ($t = -0.08$), and the Reformulation 4 stress-conditional tertile contrast ($t = +2.06$, its sign opposite to the crash-precursor prediction). The contrast between the lone amber point and the blue cluster is the paper's thesis. Data: the four path-signature results files plus the full-sample template audit (2026).

4. Mechanism: Anchor-Window Curation Artefact

The apparent in-sample separation observed in the original (and now disconfirmed) reformulation 1 admits a specific mechanism we found while auditing the result: the calm reference window contained anchor periods immediately preceding all-time-high prints. Such windows share geometric features—low realised vol, gentle upward drift, low-amplitude oscillation— with the *immediate pre-peak* regime that occurs just before a drawdown. The signature therefore separates “calm” from “crash” in-sample not because it distinguishes regimes but because the curated calm set is over-represented by pre-peak tail segments.

> **Definition (anchor-window artefact).** A reference window that satisfies a nominal "calm" criterion (low VIX, low drawdown) but is selected from periods immediately preceding peaks is statistically confounded with the pre-crash regime under any shape-based encoding that is invariant to absolute price level. The cosine similarity between a current-window signature and the average curated-calm signature is then artificially elevated whenever the current window also sits near an all-time high, regardless of whether a crash actually follows.

This mechanism yields a testable prediction: removing the anchor windows immediately preceding the top decile of cumulative price from the calm reference should collapse the apparent in-sample separation. The prediction is consistent with what the full-sample audit already shows — that the curated separation carries no forward-drawdown edge (§3.1, Welch t = 0.17 across 418 high-similarity windows) — and we flag the exclusion as a mandatory control for any future signature–crash specification (§5.2).

5. Implications for Quantitative Research

5.1 Path signatures as features, not predictors

The signature is mathematically well-defined and information-rich. Our result does not refute the signature as a representation; it refutes its use as a binary crash-precursor classifier under the four specifications tested. The signature remains usable for:

> **Continuing valid uses of the signature feature:**
> • Path-shape clustering (e.g., regime taxonomies that do not commit to a forward outcome)
> • Cross-asset path comparison for spread or pairs construction
> • Drift estimation in randomised-signature non-parametric methods (Cuchiero et al., 2023)
> • Trend-stage attribution and feature engineering downstream of a separately validated predictor
> • Trajectory-quality metrics for execution analysis

5.2 Methodological recommendations

> **For any future signature–crash specification, we recommend:**
> • **Anchor-window exclusion.** Strip windows immediately preceding the top decile of cumulative price from any reference set. The artefact identified here is otherwise transparent only after the fact.
> • **Pre-registered Harvey thresholds.** Commit to a t-statistic cutoff and direction before the reformulation is run, not after the data is examined. The path-signature literature contains many positive results computed on training windows whose construction was informed by visual inspection of the test period.
> • **Cross-event transfer test.** A signature that predicts the 2008 crash should be tested on 2020 and 2022 without re-fitting. Failure of transfer should be reported as a primary result, not omitted.
> • **Direction discipline.** If the empirical direction inverts under a re-formulation, the original hypothesis is refuted; do not silently swap signs to recover a nominal t-statistic.

5.3 Connection to validation-limits framework

This note operationally instantiates the methodological position taken in Andreev & Howden (2026e) on the structural inadequacy of historical backtesting for wavefunction-calibrated financial systems: the disconfirmation here required a four-reformulation chain plus an auditable mechanism description. Each reformulation produced a different t-statistic on a different stratification of the same underlying data; only the forward-window protocol with frozen thresholds gave a result that could not be retro-fit. We endorse Harvey–Liu–Zhu (2016) in this paper's context and refer the reader to (2026e) for the broader argument.

6. Discussion

The negative result reported here is informative on three levels. First, at the empirical level, it removes path signatures from the candidate set of binary crash-precursor signals in equity markets under any of the four specifications we tested; this should reduce the research-allocation cost for downstream teams considering the same direction. Second, at the methodological level, it demonstrates that signature-similarity scores under curated reference windows can produce in-sample t-statistics that exceed 5.0 yet fail to survive direction- and transfer-testing—a textbook instance of the multiple-testing hazard emphasised by Harvey *et al.* (2016). Third, at the theoretical level, the anchor-window artefact identified in §4 is a generic mechanism that may apply to many shape-based features beyond signatures—wavelet coefficients, persistent-homology summaries, fractal dimension estimates—whenever the reference set is selected via thresholding criteria correlated with the underlying price level.

We deliberately publish this null. Failure-mode papers are systematically under-produced in the quantitative finance literature, which biases the public record toward apparent discoveries that do not replicate. Documenting *why* a method fails is at least as valuable as documenting *that* a method succeeds.

References

[1] T.J. Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310, 1998.

[2] I. Chevyrev & A. Kormilitzin. A primer on the signature method in machine learning. *arXiv preprint*, 1603.03788, 2016.

[3] D. Levin, T. Lyons & H. Ni. Learning from the past, predicting the statistics for the future, learning an evolving system. *arXiv preprint*, 1309.0260, 2013.

[4] C. Salvi *et al.*. The signature kernel is the solution of a Goursat PDE. *SIAM Journal on Mathematics of Data Science*, 3(3):873–899, 2021.

[5] C. Cuchiero, P. Schmocker & J. Teichmann. Global universal approximation of functional input maps on weighted spaces. *arXiv preprint*, 2305.18464, 2023.

[6] C.R. Harvey, Y. Liu & H. Zhu. …and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68, 2016.

[7] M.A. Andreev & J. Howden. Stochastic mechanical methods for quantitative portfolio management. *Quark Quantitative Research Working Paper*, 2026a.

[8] M.A. Andreev & J. Howden. Ontological limits of historical validation for wavefunction-calibrated financial systems. *Quark Quantitative Research Working Paper*, 2026e.

[9] M.L. Welch. The generalization of Student's problem when several different population variances are involved. *Biometrika*, 34(1–2):28–35, 1947.

Quark Quantitative Research Platform — Working Paper

Generated 2026-07-26 00:58 UTC

This paper documents the disconfirmation of path signatures as crash precursors across four sequential reformulations and identifies the anchor-window curation artefact responsible for the original apparent positive result. The signature feature remains valuable for path-shape representation; it should not be cited as a binary crash-precursor predictor. Correspondence: M.A. Andreev, Quark Quantitative Research.