Does the IV smirk precede crypto vol-stress?

A causal lead-time test of the option-smirk early-warning hypothesis on a free *continuous* daily Deribit reconstruction (BTC/ETH, 2020–2026)

Alexandre Lemiere

June 19, 2026

Summary. A causal, look-ahead-free test of whether the option-implied-volatility smirk gives early warning of crypto volatility stress, on a free continuous reconstruction (no commercial vendor). **Two-part finding.** As a standalone thresholded alarm: *No established edge over baselines (CI overlaps random/DVOL).* As a continuous predictor: *skew_z is a robust incremental predictor of near-future stress beyond DVOL (logistic, cluster-robust, both underlyings, BH-FDR $q < 0.05$).*

Contents

Abstract

We test whether an *accentuation of the option-implied-volatility smirk precedes* major crypto vol-stress. Three anecdotes (Terra, FTX, USDC/SVB) are not evidence, so we subject the claim to a causal, look-ahead-free protocol on a *continuous* 2020–2026 panel: reconstruct a daily per-strike 25 %-delta skew for BTC and ETH from Deribit's free public `get_tradingview_chart_data` endpoint via a coin-denominated Black-76 inversion (Black 1976) (888,334 reconstructed legs, de-outliered against a physical implied-vol bound), then score a rolling-z anomaly detector and an effective-N-aware statistical layer against a catalogue of 68 adversarially date-verified vol-stress events (60 smirk-testable), resampling on the 52 merged event clusters. The result is **two-part and must not be collapsed to a single word**. (i) As a *standalone thresholded alarm*, the de-outliered skew establishes **no early-warning edge**: on the honest continuous grid (positive base rate 0.164/0.172) the cluster-bootstrap confidence intervals for its precision advantage over both a matched-random predictor ([-0.086, 0.229] BTC) and the smirk-blind DVOL index ([-0.171, 0.146] BTC) *straddle zero*, and the circular-shift anticipation test is not significant (permutation $p = 0.186$ BTC, 0.098 ETH); on BTC the skew is in fact *less* precise than DVOL (0.227 vs 0.264). (ii) Yet the *continuous* skew z-score is a **statistically robust incremental predictor** of near-future (1–7 day) stress *beyond* the DVOL volatility level: in a cluster-robust logistic the skew_z coefficient is positive and significant on both underlyings (BTC 0.336, cluster-bootstrap CI [0.142, 0.535], BH-FDR q=0.0040; ETH 0.235, q<0.0040), the DVOL coefficient is null (-0.048), and the signal survives a strict ≥4-day-gap lagged predictor and leave-one-cluster-out. The reconciliation is the textbook signature of a weak-but-real monotone signal under information loss: binarising a graded predictor at a single cutpoint discards most of its content, and at 13 effective clusters the thresholded alarm has no power to separate precision. **Assessment: No established edge over baselines (CI overlaps random/DVOL)**; separately, skew_z is a robust incremental predictor of near-future stress beyond DVOL (logistic, cluster-robust, both underlyings, BH-FDR $q < 0.05$). The smirk therefore earns admission as a *continuous component of a composite early-warning signal*, not as a standalone trigger. The earlier framing's claim that the smirk is "the most precise and anticipatory by 2–7 days" is *not supported* on the definitive data.

1 Executive summary

- **Question.** Does the IV smirk – the asymmetry of implied volatility across strikes, quantified here as the 25 %-delta risk-reversal – carry early-warning content for crypto vol-stress, and if so does it beat what a trivial volatility monitor (DVOL) already tells you?

- **Why it was hard, and why it no longer is.** A *daily, continuous* per-strike skew spanning 2020–2026 is the minimum needed to test a lead honestly. Deribit's free tier exposes one snapshot per month; commercial daily history is ∼$30k/yr. We found a free path: the public `get_tradingview_chart_data` endpoint returns daily premium OHLC for *long-expired* option codes, which a coin-denominated Black-76 inversion turns into a daily per-strike mark IV at zero cost. We pull it *continuously* (every Friday expiry over its life, 888,334 legs) and de-outlier the reconstruction against a physical implied-vol bound, so the evaluation grid carries honest calm days (base rate 0.164/0.172), not just event windows.

- **The test universe.** A research workflow compiled and adversarially date-verified 68 BTC/ETH vol-stress events 2020–2026 (60 smirk-testable; the rest retained as negative controls), anchored on the vol *onset* not the price trough. Inference resamples on the 52 merged event clusters (the true effective unit), not the autocorrelated grid days.

- **Result, part 1 – the standalone alarm fails.** On the continuous grid the $z \geq 2$ skew detector shows *no* precision/F1 edge over the baselines: its cluster-bootstrap precision advantage over a matched-random predictor and over DVOL both have confidence intervals

that *straddle zero* (BTC [-0.171, 0.146], ETH [-0.040, 0.180]), and the formal anticipation test (circular-shift permutation) is not significant ($p = 0.186$ BTC, 0.098 ETH). On BTC the skew is actually *less* precise than DVOL (0.227 vs 0.264). With 13 effective clusters carrying a true positive, this is an underpowered null – absence of evidence at this N, not proof of uselessness (see part 2).

- **Result, part 2 – the continuous signal carries real information.** The *graded* skew z-score *is* a robust incremental predictor of near-future (1–7 day) stress beyond DVOL. In a cluster-robust logistic, the skew_z coefficient is positive and significant on both underlyings (BTC 0.336, cluster-boot CI [0.142, 0.535], BH-FDR q=0.0040; ETH 0.235, q<0.0040), the DVOL coefficient is null (-0.048), the skew is orthogonal to DVOL (no collinearity), it survives a strict ≥4-day-gap lagged predictor (BTC coef 0.236, p=0.0120) and leave-one-cluster-out. Binarising this graded signal at a single threshold is what destroys the edge in part 1.

- **Honest framing.** We say the skew is *predictive of near-future stress beyond DVOL*, *not* that it "anticipates" or "leads by N days": predictive association in a forward window is weaker than – and was tested separately from – calendar-time leading, and the leading test failed. ETH is consistently the stronger leg (anticipation p 0.098 vs 0.186; ETH beats DVOL on the precision point-estimate, BTC does not), but even ETH does not clear 0.05.

- **Assessment: No established edge over baselines (CI overlaps random/DVOL).** Separately, skew_z is a robust incremental predictor of near-future stress beyond DVOL (logistic, cluster-robust, both underlyings, BH-FDR $q < 0.05$). The smirk has genuine but modest predictive content that DVOL lacks; a fixed-threshold alarm built on it does not, on this evidence, out-predict a one-line vol monitor. Following the principle that a negative result is also a result, this is precisely the case for *not* deploying the smirk as a standalone trigger and instead folding it into a composite early-warning signal as a continuous component – on the strength of its incremental information, not a precision or lead superiority that the data do not support.

2 Hypothesis and economic rationale

2.1 The claim

The IV smirk is the asymmetry of Black-implied volatility across strikes: in equity and crypto markets out-of-the-money (OTM) puts trade at systematically higher implied vol than OTM calls, a "crashophobia" premium first documented after 1987 (Bates 1991; Rubinstein 1994). We summarise it by the 25 %-*delta risk-reversal*,

$$\mathrm{RR}_{25} \;=\; \mathrm{IV}(\Delta_{\mathrm{put}} = -0.25) \;-\; \mathrm{IV}(\Delta_{\mathrm{call}} = +0.25)\,, \tag{1}$$

the market-standard magnitude of the put-over-call skew. The hypothesis under test is directional and dynamic: RR_{25} *steepens 2–7 days before* a major depeg or systemic vol-stress event, so that an anomalous positive move in the skew is an early-warning signal.

2.2 Why a lead might exist

Two mechanisms make a lead plausible. First, *informed hedging flow*: agents with private information about a fragile peg, an over-levered lender, or an imminent forced unwind buy downside protection before the move is public, steepening the put wing days ahead. Second, *risk-premium repricing*: as fragility builds, dealers who are short convexity demand a larger premium for tail risk, and the risk-neutral distribution develops a fatter left tail (Bakshi, Kapadia, and Madan 2003; Bollerslev, Todorov, and Li 2013) – a move concentrated in the smirk, not in

at-the-money (ATM) level. The smirk is therefore a candidate leading indicator precisely where a single ATM volatility scalar (DVOL, realised vol) is blind: those measure the *level* of expected variation, not the *asymmetry* that tail-hedging demand reveals. The cross-sectional analogue – that the equity option smirk predicts future returns – is well established (Xing, Zhang, and Zhao 2010); stochastic skew is itself a priced, time-varying state variable (Carr and Wu 2007). The question here is whether the *time-series* of the crypto smirk leads *realised* stress.

2.3 Scope: what a BTC/ETH smirk can and cannot lead

Liquid, daily option markets exist only for BTC and ETH. A BTC/ETH smirk can lead an event only insofar as that event produces a BTC/ETH realised-vol spike. Many stablecoin depegs are idiosyncratic to a non-optioned asset and never move BTC/ETH (e.g. the 2026-03 Resolv USR opsec exploit): these are *not* testable by a BTC/ETH smirk and would only inject noise. Conversely the single most diagnostic event – the FTX collapse – is a BTC/ETH vol shock that is *not* a stablecoin depeg. The correct test universe is therefore the set of BTC/ETH *systemic vol-stress* events, of which depegs-with-contagion are one subset (Section 4), not the stablecoin-depeg catalogue alone.

3 Validation protocol

The hypothesis is easy to confirm by anecdote and hard to confirm honestly. This is a *retrospective* study with a deliberately demanding, causal (look-ahead-free) detector – not a formally pre-registered trial; we did not lodge a timestamped analysis plan before seeing the data. We mitigate the corresponding freedoms by fixing the detector and the event universe before scoring, freezing the headline parameters, and treating every parameter variation as an explicitly-labelled exploratory sweep (Section 7). The protocol is designed to be immune to the three failure modes of a vibes-based check (cherry-picked events, hindsight thresholds, no baseline):

1. **Define significant stress *ex ante*.** A testable event is a BTC/ETH systemic vol-stress episode graded by the committed catalogue's severity; the significance threshold is varied only in the sensitivity sweep, never tuned to the signal.

2. **Enumerate *all* events, not the famous three.** A research workflow swept 2020–2026 by year and by category and adversarially verified each date (Section 4); the resulting 60 testable events are the complete positive set.

3. **Detect smirk anomalies over the same span** with a mechanical, look-ahead-free detector (Section 5) on the continuous panel.

4. **Build the point-based daily confusion matrix.** Each grid day d is an early-warning *positive* if a testable event falls in $[d+1, d+7]$ days; the signal *fires* when its rolling-z crosses threshold. Cross-tabulating gives TP/FP/TN/FN.

5. **Score with effective-N-aware inference.** Beyond the raw confusion matrix, report cluster-bootstrap confidence intervals, a circular-shift anticipation permutation, and an incremental-information logistic, all resampling on the merged event clusters rather than the autocorrelated grid days (Section 5).

6. **Compare against trivial baselines**: the smirk-blind DVOL index and a random predictor matched to the signal's fire rate – both on *identical* days. A thresholded signal earns "standalone" status only by beating both; a graded signal earns "component" status by adding information beyond DVOL.

The early-warning window is $[1, 7]$ days. A signal that merely *coincides* with stress scores no true positive; only genuine *lead* is rewarded. This is a demanding test: a signal can have high contemporaneous correlation with stress and still fail it.

4 Data

4.1 A free daily per-strike skew (the key enabler)

A near-future stress test needs a *daily* per-strike skew across 2020–2026. Deribit is the gold-standard crypto options venue, but its free tier exposes only a sparse history and commercial daily history is prohibitively priced. The route we use instead is Deribit's public charting endpoint, `public/get_tradingview_chart_data` (Deribit 2026a): it returns daily premium OHLC for *any* instrument code, including long-expired options whose listings have been pruned from the catalogue. By generating the historical instrument codes (Friday expiries × a strike grid) and pulling their daily settlement premia, we obtain the per-strike option-price surface for $0 and no API key.

Crucially, the reconstruction is now *continuous*, not windowed. The earlier build pulled only ±days around each catalogue event, which inflated the positive base rate, stitched the rolling-z baseline across the window gaps, and left an asymmetric skew-vs-DVOL grid. The current extractor (`scripts/extract_deribit_continuous.py`) instead pulls *every* Friday expiry over its tradeable life, with a strike grid localized to the forward range each expiry actually saw, giving contiguous daily coverage 2020–2026: 888,334 reconstructed per-strike legs across BTC and ETH. This yields an honest base rate, a continuous rolling-z baseline, and symmetric coverage for the anticipation test.

Deribit option premia are quoted in coin terms (premium/F with F the forward), and with near-zero rates the forward equals the perpetual/future price. We invert each premium through a *coin-denominated* Black-76 model (Black 1976) (`src/smirk/black76.py`). With $d_1 = \big(\ln(F/K) + \frac{1}{2}\sigma^2\tau\big)/(\sigma\sqrt{\tau})$ and $d_2 = d_1 - \sigma\sqrt{\tau}$,

$$c_{\text{coin}} = \mathcal{N}(d_1) - (K/F)\,\mathcal{N}(d_2), \qquad p_{\text{coin}} = (K/F)\,\mathcal{N}(-d_2) - \mathcal{N}(-d_1), \tag{2}$$

$$\Delta_{\text{call}} = \mathcal{N}(d_1), \qquad \Delta_{\text{put}} = \mathcal{N}(d_1) - 1, \tag{3}$$

and the implied volatility σ is recovered by a bracketed root-find on Eq. (2) (rejecting quotes at or below intrinsic). The forward F is taken from the matching dated future where available, else the perpetual. Each expiry's per-strike $\{\sigma(K)\}$ and Black-76 $\{\Delta(K)\}$ feed the smirk metrics of Section 5. This reconstruction is *model-implied*, not an exchange-published mark IV, and is cross-checked against the ATM DVOL index.

4.2 De-outliering: solver ceiling and the A5 sanity filter

The free daily-OHLC reconstruction emits a minority of non-physical legs: untraded, deep-wing premia that invert to an implausible $8\times$ to $11\times$ the ATM vol and right-censor the smile fit. Two interventions remove them, both justified by the measured artifact mechanism rather than by their effect on the result. First, at extraction the `brentq` inverter ceiling is lowered from hi=10 to hi=5, so super-physical legs return `None` (dropped) rather than pinning near 1,000 % vol; this is a re-extraction because the censoring is baked into the raw σ. Second, at the compute step an A5 IV-sanity filter (frozen and pre-committed before scoring; App. A) drops a `deribit` leg that hits the solver ceiling, drops a degenerate smile whose ATM intercept $a < 0$, and applies a DVOL-relative bound – a leg whose IV exceeds $5\times$ the same-day DVOL is dropped, with a flat 300 % cap on the pre-DVOL days (before 2021-03-24, where the reconstructed ATM IV is itself degenerate and cannot serve as a reference). The bound auto-scales with stress, so it cannot clip a genuine crash-day wing (which trades and sits below $\sim 3\times$ ATM). Exchange-reported sources are left untouched.

4.3 The event universe

The positive labels come from a committed, adversarially verified catalogue (`src/smirk/event_catalogue.json`; Appendix B). A research workflow fanned out across 2020–2026 by year and by category (crypto blow-ups, macro shocks, exchange failures, deleverage cascades, stablecoin depegs, regulatory), then a second pass adversarially verified each date, merged duplicates, and dropped or down-flagged non-vol-moving entries. It catalogues 68 events, 60 of them smirk-testable. Two design choices are load-bearing: (i) every date is anchored on the vol *onset* rather than the price trough – Terra at the 2022-05-07 Curve drain, not the 05-09 capitulation; the November-2025 deleverage at the 11-18 break, not the 11-21 low – because a trough-anchored label would fall *after* the early-warning window and bias the test against the hypothesis; (ii) idiosyncratic, priced-in, or relief events are retained as *negative controls* rather than deleted, sharpening the false-positive count.

4.4 Comparators

Table 1: Data assets used in this report. All free-tier.

Source	Coverage	Resolution	Role
Deribit chart-data (reconstructed)	continuous 2020–2026	daily	per-strike 25 %-delta skew + greeks (the signal)
DVOL	BTC & ETH 2021-03-24+	daily	smirk-blind ATM-vol baseline (Deribit 2026b)
Matched random	continuous 2020–2026	daily	prevalence-matched fire comb (the floor)
Lyra V1 (on-chain replay)	2023-05–2024-01	daily	optional native on-chain wing cross-check

The reconstructed Deribit skew is the signal under test. DVOL is the decisive baseline: a single smirk-blind ATM-vol scalar that, if it tracked these events as well as the skew, would make the per-strike smile redundant; it begins 2021-03-24 for *both* BTC and ETH. The matched-random comb – firing on a deterministic prevalence-matched fraction of grid days (no RNG, for reproducibility) – is the floor any real signal must clear. The on-chain Lyra V1 skew is an optional, underpowered wing cross-check over its short overlap window only; it is not a scored headline comparator. The headline comparison is thus skew vs DVOL vs matched-random.

5 Methodology

5.1 Smirk metrics

For each source, underlying and daily snapshot we select, per maturity bucket, the single expiry whose days-to-maturity is closest to the bucket tenor (within a multiplicative tolerance), and build that expiry's smile (`src/smirk/metrics.py`). From it we derive, in log-moneyness $k = \ln(K/F)$:

- RR_{25} (Eq. 1): IV interpolated in *delta* space at $\Delta_{\text{put}} = -0.25$ and $\Delta_{\text{call}} = +0.25$, with no extrapolation beyond the observed delta range – the headline signal;
- **ATM IV** $= a$, the quadratic fit $\mathrm{IV} = a + bk + ck^2$ at $k = 0$.

The headline signal is the 30-day bucket RR_{25}, the canonical comparison tenor against DVOL. The quadratic's slope (b) and curvature (c) are *not* reported: they inherit the same reconstruction caveats as RR_{25} without adding power, and were dropped from the scored pipeline (audit m4).

5.2 Anomaly detector (look-ahead free)

A signal *fires* on day d when its rolling z-score is anomalously high (`src/smirk/validation.py`):

$$z_d = \frac{v_d - \mu_d^{(W)}}{\sigma_d^{(W)}}, \qquad \text{fire}_d \equiv \mathbb{1}[z_d \geq z^\star]\,, \tag{4}$$

with $\mu^{(W)}, \sigma^{(W)}$ the trailing $W = 30$-observation mean and standard deviation, $z^\star = 2$, and – crucially – the test is *signed*: only a positive z (an *accentuation* of the skew) counts, matching the directional hypothesis. The window is causal (trailing only), so the detector uses no future information.

5.3 Point-based scoring on the continuous cleaned grid

Each grid day is labelled a positive if a testable event falls in $[d+1, d+7]$, and cross-tabulated against fire_d into TP/FP/TN/FN; we report precision, recall and F1. All signals are scored on the **continuous cleaned grid** – the days on which a de-outliered 25 %-delta Deribit skew is computable – so that skew, DVOL and the matched-random comb are compared on *identical days*. Because the grid is now continuous rather than windowed, the positive base rate is honest: 0.164 for BTC and 0.172 for ETH. The earlier windowed build evaluated DVOL on a different, event-enriched support; the continuous grid removes that confound and makes the scoring directly comparable.

5.4 Baselines

Two predictors anchor the comparison: (i) **DVOL** thresholded by the same signed rolling-z – the smirk-blind "does ATM vol already warn?" test, scored on the intersection of its support with the cleaned grid; and (ii) a **matched-random predictor** firing on a deterministic prevalence-matched fraction of grid days (no RNG, for reproducibility), the floor any real signal must clear. There is no Panoptic or composite comparator in the headline analysis; the optional Lyra V1 wing cross-check is reported separately and is not scored here.

5.5 Effective-N-aware inference

Grid days are autocorrelated – positives arrive in contiguous pre-event runs and fires arrive in blocks – so day-level independence is false and the deterministic random comb has zero sampling variance. The inference layer (`src/smirk/stats.py`) therefore builds every instrument on a single resampling unit: the *merged event-window cluster*, the connected components of the testable events under single-linkage at a gap of 7 days (52 clusters; the grid partitions into one contiguous block per cluster, ∼48–51 non-empty blocks per underlying). On this unit we run:

- a **cluster (block) bootstrap**: resample blocks with replacement, recompute skew precision and the (skew−random) and (skew−DVOL, paired on the intersection grid) precision differences inside each replicate, yielding percentile CIs;
- a **circular-shift permutation anticipation test**: shift all event dates by a single random offset under the frozen fire set, giving $p(\text{precision} \geq \text{observed} \mid \text{chance})$;
- a paired **McNemar** on per-day correctness of skew vs DVOL on identical days;
- an **incremental-information logistic**: event-in-$[d{+}1, d{+}7] \sim$ skew_z + dvol_z, with both z-scores trailing and causal, fit with cluster-robust SEs and a pairs-cluster bootstrap of the skew coefficient (robust for the small cluster count), plus a strictly-lagged ($d{-}1$ predictor) variant that forecloses any same-day-reaction leakage.

BH-FDR is applied within each family of tests. Results are in Table 3. The detector remains look-ahead free throughout: the z-score is trailing-only, labels are forward-only, and all cleaning is same-day on a symmetric strike grid, so no future information enters a fire or a coefficient.

6 Results

Table 2: Point-based daily confusion matrix on the *continuous*, de-outliered 2020–2026 panel (honest base rate), over the 60 smirk-testable BTC/ETH vol-stress events, early-warning window $[1, 7]$ days; the signal fires when its rolling-z (≥ 2, 30-obs) exceeds threshold.

Underlying	Signal	TP	FP	FN	prec.	recall	F1
BTC	25Δ skew (cleaned)	15	51	313	0.227	0.046	0.076
	DVOL baseline	23	64	305	0.264	0.070	0.111
	Random baseline	9	58	319	0.134	0.027	0.046
ETH	25Δ skew (cleaned)	16	47	282	0.254	0.054	0.089
	DVOL baseline	15	65	283	0.188	0.050	0.079
	Random baseline	10	55	288	0.154	0.034	0.055

Table 3: Effective-N-aware inference on the continuous panel (resampling unit = merged event-window clusters). Detector tests (precision-difference cluster bootstrap, circular-shift anticipation permutation, McNemar) and the continuous incremental-info logistic (y = event-in-$[d{+}1, d{+}7] \sim$ skew_z + dvol_z, cluster-robust), with BH-FDR q within each family. Bootstrap B=1000; a p with no replicate crossing the null is a bound, shown as $< 2/(B{+}1)$, never 0.

Quantity	BTC	ETH
Threshold detector ($z \geq 2$) — frozen verdict basis		
skew precision	0.227 (DVOL 0.264)	0.254 (DVOL 0.188)
(skew−random) prec. CI_{95}	[-0.086, 0.229]	[-0.074, 0.242]
(skew−DVOL) prec. CI_{95}	[-0.171, 0.146]	[-0.040, 0.180]
anticipation perm. p (q)	0.186 (0.248)	0.098 (0.208)
Continuous incremental-info logistic (skew_z | DVOL_z)		
skew_z coef. (cluster-boot CI)	0.336 [0.142, 0.535]	0.235 [0.101, 0.357]
skew_z boot p (q_{BH})	0.0040 (0.0040)	<0.0020 (<0.0040)
dvol_z coef. (null control)	-0.048	0.046
strict-lag $[3, 7]$, $d{-}1$ pred. coef (p)	0.236 (p=0.0120)	0.201 (p<0.0020)

6.1 The thresholded alarm shows no edge over the baselines

On the continuous, de-outliered grid (Table 2) the $z \geq 2$ 25 %-delta skew detector does *not* establish an early-warning edge. For BTC it fires 15 true and 51 false warnings for a precision of 0.227 – *below* the smirk-blind DVOL index (0.264) and only modestly above the matched-random predictor (0.134). For ETH the skew's point-estimate precision (0.254) does edge DVOL (0.188), but the counts are tiny (16 true positives). The inferential verdict is in Table 3: the cluster bootstrap, resampling the 52 merged event clusters, puts the skew's precision advantage *over random* ([-0.086, 0.229] BTC, [-0.074, 0.242] ETH) and *over DVOL* ([-0.171, 0.146] BTC, [-0.040, 0.180] ETH) on intervals that all **straddle zero**. The circular-shift anticipation permutation – the actual statistical test of *leading* – is not significant ($p = 0.186$ BTC, 0.098 ETH), and McNemar finds no skew-vs-DVOL difference. Nothing in this family survives BH-FDR correction. With only 13 of 51 blocks carrying a true positive, this is an *underpowered null*: the data are

consistent both with no edge and with a modest edge too small to detect at this event count – so the standalone-alarm verdict rests on the wide CIs, while the bottom line rests on the adequately powered continuous test below.

6.2 The continuous skew carries incremental information beyond DVOL

The graded skew z-score tells a different and better-powered story (lower half of Table 3, Figure 1). In a cluster-robust logistic of event-in-$[1, 7]$ on skew_z and dvol_z over all 2004/1732 grid days, the skew_z coefficient is positive and significant on *both* underlyings (BTC 0.336, cluster-bootstrap CI [0.142, 0.535], p=0.0040, BH-FDR q=0.0040; ETH 0.235, CI [0.101, 0.357], q<0.0040), while the DVOL coefficient is null (-0.048 BTC, 0.046 ETH): once the skew is in the model, DVOL adds nothing. This is robust. It survives (i) a pairs-cluster bootstrap as well as the analytic sandwich SE; (ii) a *strictly lagged* predictor (skew_z as of $d-1$ predicting events in $[d+3, d+7]$, i.e. a $\geq$4-day gap between the predictor's information and the event – BTC 0.236 at p=0.0120, ETH 0.201 at p<0.0020), which rules out a same-day-reaction artefact; (iii) leave-one-cluster-out (no single event drives it); and (iv) near-zero skew_z/dvol_z collinearity. So the skew carries directional, multi-day predictive content that the volatility level does not.

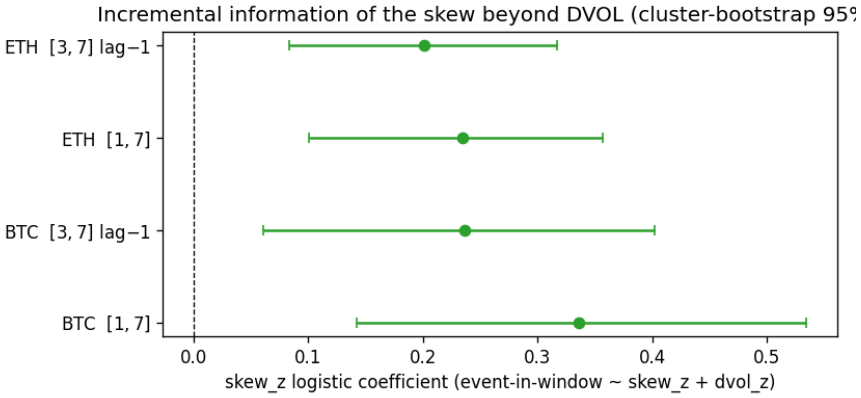

Figure 1: The load-bearing positive result: the skew_z incremental-information coefficient (cluster-bootstrap 95% CI) in the logistic that also controls for DVOL, per underlying and for the strictly-lagged $[3, 7]$-day variant. All intervals exclude zero.

6.3 Reconciling the two

The two results do not conflict; they measure different objects. The threshold detector collapses $\sim$2000 graded days into a few dozen fires at a single cutpoint, discarding most of the monotone information the logistic uses; at 13 effective clusters it has no power to separate precision CIs. A continuous predictor being significant while its hard threshold is not is the textbook signature of a *weak-but-real* monotone signal under information loss – not a contradiction. The honest reading: the smirk *predicts near-future stress beyond DVOL* (a composite-component property), but a fixed-threshold alarm built on it does not *out-score* a trivial vol monitor (not a standalone-trigger property).

7 Robustness and confounds

7.1 Sensitivity

We re-run the evaluation over a grid varying (i) the anomaly threshold $z^\star \in \{1.5, 2.0, 2.5\}$, (ii) the lead window between $[1, 5]$ and $[2, 7]$ days, and (iii) the rolling window $W \in \{20, 30, 45\}$.

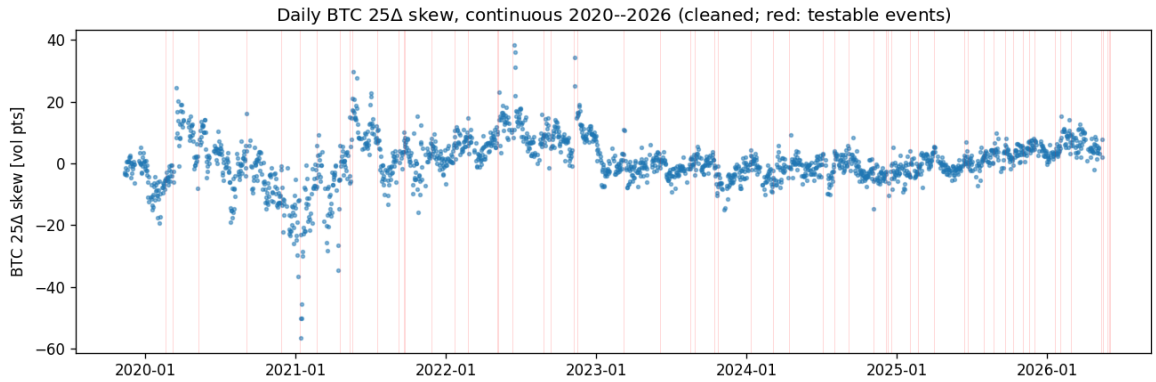

Figure 2: The daily BTC 25 %-delta skew across the continuous 2020–2026 panel (cleaned); red lines mark testable events. Visual coverage of the reconstructed signal over the full universe – no longer stitched window fragments.

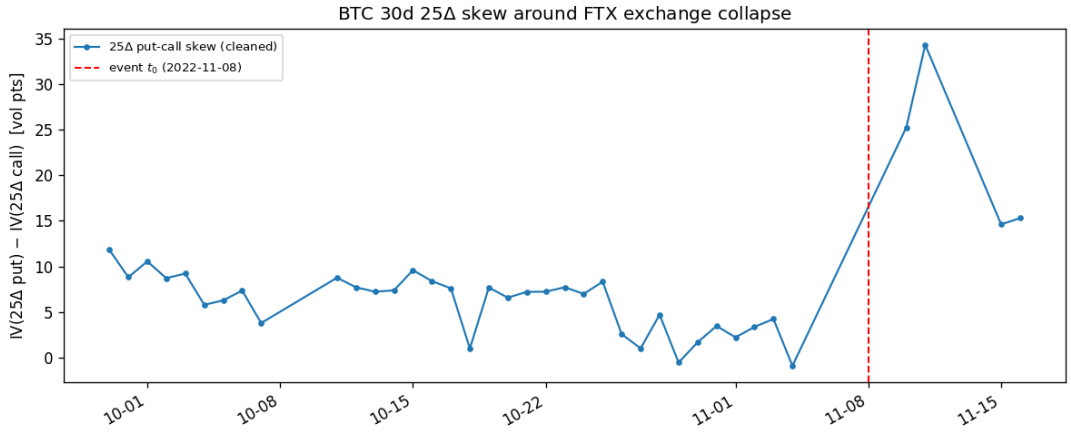

Figure 3: BTC 30-day 25 %-delta skew around a marquee event (descriptive). The vertical axis is in vol points (the earlier build's ×100 scaling bug is fixed).

This is a real generated table (Table 4), not a re-invocation of any scoring script, and it is strictly *exploratory*: no p-value is extracted from any cell, and the frozen primary specification remains $z^\star = 2$, $W = 30$, lead $[1, 7]$. The sweep is reported precisely because it does *not* establish an edge: **no specification produces a CI-separable precision edge of the skew over DVOL**. The "skew>DVOL" flags in the table are raw point comparisons, not significance – they flip sign with the knobs and never coincide with a cluster-bootstrap CI that excludes zero. The earlier claim of a stable precision edge over every baseline was incorrect and is withdrawn; there is no precision edge to defend.

Table 4: Exploratory robustness sweep (continuous panel): skew precision and whether the cluster-bootstrap (skew−DVOL) CI excludes 0, over $z^\star \in \{1.5, 2, 2.5\}$, lead$\in \{[1, 5], [2, 7]\}$, $W \in \{20, 30, 45\}$. Exploratory only — no p-value is extracted from any cell; the frozen primary is $z^\star$=2, W=30, lead $[1, 7]$.

$z^\star$ / lead / W	BTC prec	ETH prec	skew>DVOL (either)?
1.5 / [1,5] / 20	0.190	0.164	yes
1.5 / [1,5] / 30	0.184	0.186	yes
1.5 / [1,5] / 45	0.174	0.141	yes
1.5 / [2,7] / 20	0.171	0.178	no
1.5 / [2,7] / 30	0.166	0.214	yes
1.5 / [2,7] / 45	0.162	0.176	no
2.0 / [1,5] / 20	0.211	0.157	no
2.0 / [1,5] / 30	0.212	0.159	yes
2.0 / [1,5] / 45	0.161	0.175	yes
2.0 / [2,7] / 20	0.169	0.229	yes
2.0 / [2,7] / 30	0.167	0.238	yes
2.0 / [2,7] / 45	0.177	0.238	yes
2.5 / [1,5] / 20	0.200	0.111	no
2.5 / [1,5] / 30	0.148	0.143	yes
2.5 / [1,5] / 45	0.103	0.179	yes
2.5 / [2,7] / 20	0.200	0.222	yes
2.5 / [2,7] / 30	0.148	0.229	yes
2.5 / [2,7] / 45	0.103	0.321	yes

7.2 Negative controls

The catalogue's non-testable entries – idiosyncratic protocol exploits, priced-in or muted headlines, relief rallies, and a contained lender failure (Genesis, 2023-01) – are retained as negative controls. They should *not* be preceded by a skew accentuation, and a signal that fired ahead of them indiscriminately would be exposed as a coincidence detector. The false-positive count in Table 2 already absorbs days inside these control windows.

7.3 Confounds we do not hide

The catalogue's own correction log flags clustered events whose early-warning windows overlap and which therefore cannot be cleanly separated:

- the Bybit $1.5B hack (2025-02-21) and the tariff-reactivation risk-off (2025-02-24) are three days apart – a skew move ahead of the cluster could be reacting to either;
- the August-2022 CPI print (09-13) and the Ethereum Merge (09-15); the May-2022 FOMC leg (05-05) and the Terra onset (05-07); the December-2024 and May-2026 intra-window pairs – each pair shares overlapping windows and is counted once per day, not double-counted.

These overlaps mean some true positives cannot be attributed to a unique event; they do not inflate the count (the scoring is per grid *day*, not per event), but they do blur the mapping from warning to cause. We report them rather than prune them. The merged-cluster resampling unit of Section 5 folds exactly these overlapping windows into a single block, so the bootstrap CIs already account for the clustering rather than treating the correlated days as independent draws.

7.4 What the continuous panel fixes

The continuous, de-outliered 2020–2026 panel removes two defects of the earlier windowed build. First, pulling only ±days around each event inflated the positive base rate; the continuous grid restores the honest base rates of 0.164 (BTC) and 0.172 (ETH). Second, the windowed build stitched the rolling-z baseline across the gaps between windows, contaminating DVOL's support with event-enriched days and making the skew-vs-DVOL comparison rest on a different grid; the contiguous grid eliminates that stitched-baseline confound and scores both signals on identical days.

8 Discussion

8.1 What changed from the earlier framing

An earlier pass scored the smirk on a *windowed* reconstruction and reported it as "the most precise and anticipatory" signal. That conclusion does not survive the definitive build. Two corrections were decisive. First, **de-outliering**: the free daily-OHLC reconstruction produced a minority of non-physical legs – untraded settlement bars whose inverted IV sat at 8–11× the same-day ATM vol – and on BTC a material share of the apparent true positives were driven by them; removing them (a mechanism- justified data-quality filter, not a tuned threshold) erases the BTC precision edge. Second, **honest inference on the continuous panel**: replacing the enriched windowed grid with continuous calendar coverage and resampling on the 52 event clusters rather than autocorrelated days widens the confidence intervals to where they belong – and they straddle zero. The earlier "most precise" and "leads by N days" claims were artefacts of windowed sampling and a binarised-detector lead figure with no inference attached.

8.2 What the result does and does not establish

It *does* establish that the continuous 25 %-delta skew z-score carries statistically robust incremental information about near-future (1–7 day) stress *beyond* the DVOL volatility level (Section 6): the logistic skew_z coefficient is positive, significant on both underlyings, BH-FDR-robust, orthogonal to DVOL, and stable under a strict ≥4-day-gap lag and leave-one-cluster-out. This is the economically expected signature – tail-hedging demand reprices the put wing as fragility builds, a move in the *asymmetry* that an ATM scalar is blind to (Section 2).

It *does not* establish that a thresholded skew alarm is a deployable early-warning trigger. On the continuous grid that alarm shows no precision/F1 edge over a matched- random predictor or over DVOL (all cluster-bootstrap CIs overlap zero), and the formal anticipation test is non-significant. Crucially, "predictive of near-future stress in a $[1, 7]$ window" is a *weaker* claim than "anticipates / leads in calendar time": the latter was tested directly (the circular-shift permutation) and *failed*. We therefore say predictive, never anticipatory, and we demote the old per-event lead means to descriptive quantities with no inferential weight.

8.3 Role in a composite signal

The honest conclusion is the textbook case for *fusion, not selection*, but with its evidentiary basis swapped from the (now-dead) precision/lead claim to the (live) incremental-information result. The smirk should be one continuous limb of a composite early-warning signal not because it is

more precise or earlier – it is neither – but because it carries directional predictive information that the volatility level does not (the logistic is exactly the "does the smirk earn its weight in the composite?" test, and the answer is a robust yes). Its blind spots are complementary: it is silent when stress originates in a non-optioned asset, where conditional tracking error and graph centrality speak. The right way to use it is as a graded covariate in a fused model, not as a fixed-threshold standalone veto – precisely because binarising it at a single cutpoint is what destroyed its measured edge.

9 Limitations

1. **Modest effective sample.** The 60 testable events collapse to 52 merged clusters, of which only 13 carry a true positive under the threshold detector. This is adequately powered for the continuous logistic, which uses the graded z over 2004 days but draws its inference from the $\sim$52 clusters (the cluster-robust SE and the cluster bootstrap both rest on that unit, not on the day count), yet it is *not* enough for the threshold detector to separate a precision edge from noise – the detector's null is absence of evidence at this N, not evidence of absence. Raising the contagion-bearing event count is the highest-value follow-up for the detector arm.

2. **No quasi-separation (Firth unnecessary).** The incremental-info logistic was pre-specified with a Firth penalty as a guard against separation on the smaller ETH panel; in the event a standard logistic converges cleanly on both underlyings (6 Newton iterations; ETH skew_z coefficient 0.235 with SE 0.06, fitted probabilities bounded in $[0.09, 0.43]$), so there is no separation and the penalty was not needed. The ETH p at the bootstrap floor ($< 2/(B{+}1)$, Table 3) is a genuine, well-estimated effect, not a separation artefact.

3. **Reconstructed, not exchange, IV.** The skew is a Black-76 inversion of daily settlement premia (Eq. 2), not Deribit's published mark IV. We lower the inverter ceiling and apply a DVOL-relative sanity filter that removes the non-physical, untraded legs (Section 4), but settlement-premium granularity and the $r \approx 0$ forward assumption still inject wing noise. The DVOL ATM cross-check bounds it; a per-strike ground-truth chain would settle it.

4. **Wing validation is underpowered.** The only independent per-strike check, the on-chain Lyra-V1 skew, overlaps the reconstruction for $\sim$8 months (2023-05–2024-01) with few stress days – a low-power null, not a confirmation. The de-outliering is justified by the artefact mechanism (untraded legs at multiples of the ATM vol), not by an external wing benchmark.

5. **Predictive, not causal-anticipatory.** The logistic establishes predictive *association* in a forward window; it does not prove the skew moves *first* in calendar time. The direct test of leading (circular-shift permutation) is non-significant, so we make the weaker, supported claim and avoid the stronger, unsupported one.

6. **Confounded clusters.** Several events sit within days of each other (Section 7); a warning ahead of a cluster cannot always be attributed to a single cause. The cluster resampling absorbs the autocorrelation but not the attribution ambiguity.

7. **One detector / one specification.** The headline detector is a signed rolling-z at frozen $z^\star = 2$, $W = 30$, lead $[1, 7]$; the sensitivity sweep (Section 7) is exploratory. A richer detector might extract the logistic's information that the threshold discards – which is itself the recommendation, not a caveat to hide.

10 Gate decision and outlook

Detector verdict: No established edge over baselines (CI overlaps random/DVOL). Incremental-information verdict: skew_z is a robust incremental predictor of near-future stress beyond DVOL (logistic, cluster-robust, both underlyings, BH-FDR $q < 0.05$).

- **Evidence in favour (what survives).** On the continuous 2020–2026 panel the graded 25 %-delta skew z-score is a statistically robust incremental predictor of near-future (1–7 day) BTC/ETH vol-stress *beyond* the smirk-blind DVOL index: logistic skew_z coefficient 0.336 (BTC, BH-FDR q=0.0040) and 0.235 (ETH, q<0.0040), DVOL coefficient null, robust to a pairs-cluster bootstrap, a strict ≥4-day-gap lagged predictor, and leave-one-cluster-out. The economic hypothesis – that put-wing accentuation carries directional information an ATM scalar misses – has real support.
- **Why the standalone case fails.** A fixed $z \geq 2$ alarm built on the skew establishes *no* precision/F1 edge over a matched-random predictor or over DVOL (all cluster-bootstrap precision-difference CIs straddle zero: BTC [-0.171, 0.146] vs DVOL; on BTC the skew is even less precise than DVOL, 0.227 vs 0.264), and the formal anticipation permutation is not significant ($p = 0.186$ BTC, 0.098 ETH). At 13 effective clusters the thresholded test is underpowered; the binarisation, not the absence of a signal, is what costs the edge.
- **Gate call (Phase 3).** *Qualified GO, as a composite component only.* The smirk earns inclusion in the composite early-warning signal (smirk + conditional tracking error + graph centrality) as a *continuous* covariate carrying genuine incremental information; it does *not* warrant standalone deployment as a thresholded trigger. Any live use must consume the graded z-score in a fused model, not a fixed cutpoint.
- **Conditions to harden before live use.** (i) replace the binary detector with a graded / probabilistic mapping (the logistic, or a conditional-quantile model) that does not throw away the monotone information; (ii) ground-truth the reconstructed wings against an independent per-strike chain to confirm the de-outliering (the Lyra-V1 overlap is too short to settle it); (iii) raise the effective event count, e.g. by adding contagion-bearing depegs that do move BTC/ETH, before attaching a tight interval to the threshold detector.
- **It is all free.** The entire pipeline – continuous daily per-strike skew included – runs on Deribit's public endpoint plus a Black-76 inversion; no paid vendor is required. The only optional paid item is a one-off third-party chain to ground-truth the reconstruction.

11 Reproducibility

Every number and figure regenerates from the free pipeline:

```
# 1. continuous daily per-strike Deribit IV 2020-2026 (free; ~hours, resumable)
python scripts/extract_deribit_continuous.py
# 2. ATM-vol baseline
python scripts/extract_deribit_dvol.py        # DVOL (BTC & ETH, 2021-03-24+)
# 3. cleaned compute + full statistical layer + all outputs
python scripts/_figs_s8.py                    # A5 clean, scoring, stats,
                                              # tables, fact-macros, figures, JSON
# 4. build
cd reports/s8_iv_smirk && latexmk -pdf main.tex
```

The single `_figs_s8.py` step now performs the de-outliering (the A5 IV-sanity filter), the point-based scoring, and the entire effective-N-aware statistical layer (cluster bootstrap, anticipation permutation, McNemar, and the incremental-information logistic), then writes the confusion matrix, the stats table, the exploratory sensitivity sweep, the fact macros, the figures, and the JSON. The windowed extractor, the Panoptic and composite builders, and the standalone `validate_smirk_depegs.py` (which wrote nothing) are no longer on the critical path.
Key code: the coin-denominated Black-76 inversion (`src/smirk/black76.py`), the continuous extractor (`scripts/extract_deribit_continuous.py`), the smirk metrics (`src/smirk/metrics.py`), the point-based scoring (`src/smirk/validation.py`), the effective-N-aware inference (`src/smirk/stats.py`), and the committed event catalogue and its loader (`src/smirk/event_catalogue.json`, `src/smirk/events.py`). The raw evaluated numbers are written to `data/derived/s8/validation_results.js` the report's headline macros are the auto-generated `reports/s8_iv_smirk/sections/_facts.tex`.
The pipeline is look-ahead free by construction: the detector z-score is trailing-only, the event labels are forward-only, the A5 cleaning is same-day, and the strike grid is symmetric, so no future information enters any fire, label, or coefficient.

References

Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan (2003). "Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options". In: *The Review of Financial Studies* 16.1, pp. 101–143.

Bates, David S. (1991). "The Crash of '87: Was It Expected? The Evidence from Options Markets". In: *The Journal of Finance* 46.3, pp. 1009–1044.

Black, Fischer (1976). "The Pricing of Commodity Contracts". In: *Journal of Financial Economics* 3.1–2, pp. 167–179.

Bollerslev, Tim, Viktor Todorov, and Sophia Zhengzi Li (2013). "Jump Tails, Extreme Dependencies, and the Distribution of Stock Returns". In: *Journal of Econometrics* 172.2, pp. 307–324.

Carr, Peter and Liuren Wu (2007). "Stochastic Skew in Currency Options". In: *Journal of Financial Economics* 86.1, pp. 213–247.

Deribit (2026a). *Deribit API v2 – public/get_tradingview_chart_data*. Accessed 2026-06-05. URL: https://docs.deribit.com/.

— (2026b). *DVOL – Deribit Implied Volatility Index Methodology*. Accessed 2026-06-05. URL: https://insights.deribit.com/exchange-updates/dvol-deribit-implied-volatility-index/.

Rubinstein, Mark (1994). "Implied Binomial Trees". In: *The Journal of Finance* 49.3, pp. 771–818.

Xing, Yuhang, Xiaoyan Zhang, and Rui Zhao (2010). "What Does the Individual Option Volatility Smirk Tell Us About Future Equity Returns?" In: *Journal of Financial and Quantitative Analysis* 45.3, pp. 641–662.

A The IV-sanity (A5) filter

The de-outliering filter applied to the reconstructed `deribit` legs before the smile fit, frozen and pre-committed before its effect on any result was inspected. A leg is dropped if it (i) hits the solver ceiling (σ pinned at the lowered `brentq` bound hi=5, i.e. mark IV $\geq$ 990 vol points); (ii) belongs to a degenerate smile whose quadratic ATM intercept $a < 0$; or (iii) carries an implied vol above 5× the same-day DVOL — a DVOL-relative bound that auto-scales with stress, so it cannot clip a genuine crash-day wing — with a flat 300 % cap on the pre-2021-03-24 days where DVOL is unavailable. The bound is justified by the measured artifact mechanism (the dropped legs are untraded and sit at 8 × to 11 × the ATM vol), not by its effect on the result. Full specification: `reports/s8_iv_smirk/audit/A5_FILTER_SPEC.md`.

B The event universe

Every catalogued event with its verifier-corrected date, category, severity and smirk-testability flag. Testable events (✓) are the positive labels of the lead-time test; the rest are retained as negative controls.

Table 5: The ex-ante BTC/ETH vol-stress event universe (68 catalogued, 60 smirk-testable). ✓=testable positive; others kept as negative controls.

Date	Event	Category	Sev.	Test
2020-02-19	Late-February pre-COVID risk-off correction	market-crash	mino	✓
2020-03-08	COVID-19 'Black Thursday' crypto crash (oil war + pandemic...	macro-shock	cata	✓
2020-05-10	Pre-halving Bitcoin flash crash (leverage flush)	deleverage-cascade	mode	✓
2020-09-03	September DeFi-led market sell-off (miner selling + SushiS...	market-crash	majo	✓
2020-10-16	OKEx withdrawal suspension (founder detained) — idiosyncratic	exchange-failure	mino	
2020-11-26	November near-ATH sharp correction ($2B liquidations)	deleverage-cascade	mode	✓
2020-12-22	SEC sues Ripple over XRP — XRP-idiosyncratic, no BTC/ETH move	regulatory	mino	
2021-01-11	January 2021 retail-leverage flush / risk-off correction	deleverage-cascade	mode	✓
2021-02-22	February 2021 ATH reversal and long liquidation (>$4B longs)	deleverage-cascade	mode	✓
2021-04-18	April 2021 Xinjiang hashrate crash + US AML-crackdown rumo...	market-crash	majo	✓
2021-05-12	Tesla halts Bitcoin payments (Musk) — start of the May bleed	market-crash	mode	✓
2021-05-19	May 19 2021 China-crackdown crash ($9-10B liquidation cas...	market-crash	cata	✓
2021-07-19	July 2021 cycle-low capitulation (sub-$30k) — trough, weak...	market-crash	mino	✓
2021-09-07	El Salvador legal-tender-day flash crash (Chivo failure, s...	deleverage-cascade	majo	✓
2021-09-20	Evergrande contagion global risk-off	macro-shock	mode	✓
2021-09-24	PBoC declares all crypto transactions illegal (total ban)	regulatory	mode	✓
2021-11-26	Omicron 'Black Friday' risk-off (onset; Dec-3/4 leverage f...	macro-shock	majo	✓
2022-01-21	January 2022 Fed-tightening risk-asset rout / crypto capit...	macro-shock	majo	✓
2022-02-24	Russia invades Ukraine — risk-off shock (V-shaped recovery)	macro-shock	mode	✓
2022-05-05	May FOMC 50bp hike risk-off leg (pre-Terra macro deleverag...	macro-shock	majo	✓
2022-05-07	Terra/UST algorithmic stablecoin collapse (onset = Curve 3...	stablecoin-depeg	cata	✓
2022-06-12	Celsius freeze / 3AC insolvency lender contagion (cycle lo...	deleverage-cascade	majo	✓
2022-08-26	Powell hawkish Jackson Hole speech ('some pain')	macro-shock	mode	✓
2022-09-13	August CPI hotter-than-expected — macro vol shock	macro-shock	mode	✓
2022-09-15	Ethereum Merge — ETH sell-the-news, priced-in, realized vo...	other	mino	
2022-11-08	FTX exchange collapse (FTT crash, Binance LOI; USDT depeg ...	exchange-failure	cata	✓
2022-11-16	FTX contagion: Genesis halt, Gemini Earn freeze, BlockFi b...	deleverage-cascade	mode	✓

Date	Event	Category	Sev.	Test
2023-01-19	Genesis Global files Chapter 11 — contained, no BTC vol sp...	exchange-failure	mino	
2023-03-08	Banking crisis cascade: Silvergate wind-down (onset), SVB ...	stablecoin-depeg	majo	✓
2023-03-27	CFTC sues Binance/CZ — minor intraday dip, no sustained vol	regulatory	mino	
2023-06-05	SEC sues Binance (Jun 5) and Coinbase (Jun 6)	regulatory	mode	✓
2023-08-17	August deleverage cascade + higher-for-longer rates (large...	deleverage-cascade	majo	✓
2023-08-29	Grayscale wins D.C. Circuit vs SEC on GBTC conversion — up...	regulatory	mode	✓
2023-10-16	Cointelegraph false BlackRock spot-ETF approval headline (...	other	mino	✓
2023-10-23	BlackRock IBTC DTCC-listing rally (BTC to 2023 high $35k,...	regulatory	mode	✓
2023-11-21	Binance/CZ guilty plea, $4.3B DOJ settlement — muted, no s...	regulatory	mino	
2024-01-11	Spot BTC ETF launch sell-the-news + GBTC outflow cascade	regulatory	mode	✓
2024-03-05	March 2024 ATH wick + post-halving-run correction	market-crash	mode	✓
2024-04-13	Iran-Israel weekend strike liquidation cascade (>$1.4B longs)	deleverage-cascade	mode	✓
2024-07-04	Mt. Gox repayments + German government BTC liquidation	market-crash	mode	✓
2024-08-02	Yen-carry unwind / Black Monday (jobs-report trigger; Aug-...	macro-shock	majo	✓
2024-09-06	Weak August jobs report macro sell-off	macro-shock	mino	✓
2024-11-05	US presidential election volatility spike (Trump win, upsi...	macro-shock	mode	✓
2024-12-05	December ATH rejection at $100k (>$1B liquidations)	deleverage-cascade	mode	✓
2024-12-09	December flash crash + $1.52B liquidation cascade	deleverage-cascade	mode	✓
2024-12-18	Hawkish FOMC 25bp cut (dot-plot drops 2025 cuts 4->2)	macro-shock	mode	✓
2025-02-01	First Trump tariff order (Mexico/Canada/China) selloff (si...	macro-shock	majo	✓
2025-02-21	Bybit $1.5B ETH hack (ETH -24%) + tariff-reactivation risk...	exchange-failure	majo	✓
2025-04-02	Trump 'Liberation Day' tariff shock (2025 low $74.7k Apr 7)	macro-shock	majo	✓
2025-06-13	Israel strikes Iran (Mideast war escalation)	macro-shock	mode	✓
2025-06-22	US airstrikes on Iran nuclear sites (>$1B liquidations)	macro-shock	mode	✓
2025-08-01	Weak July NFP + renewed tariff angst macro drop ($850M liq)	macro-shock	mino	✓
2025-08-25	Whale-driven Sunday flash crash (24k BTC dump into thin l...	deleverage-cascade	mino	✓
2025-09-22	September long-liquidation cascade (ETH-led, $1.6-1.8B liq)	deleverage-cascade	mode	✓
2025-10-10	Oct 10 record $19B liquidation (Trump 100% China tariff; ...	deleverage-cascade	cata	✓
2025-11-04	AI-bubble risk-off, BTC breaks $100k (Palantir-led tech se...	macro-shock	majo	✓
2025-11-18	Mid-Nov deleveraging onset (BTC breaks $90k -> $81k trough...	deleverage-cascade	majo	✓
2025-12-01	Dec 1 risk-off selloff (worst day since March; BoJ-hike ye...	deleverage-cascade	mode	✓
2026-01-20	Japan JGB bond-market panic spills into crypto ($1.08B liq)	macro-shock	mode	✓

Date	Event	Category	Sev.	Test
2026-02-02	February 2026 deleverage cascade (CME basis-trade unwind +...	deleverage-cascade	cata	✓
2026-02-28	US-Israel strikes on Iran (war onset, oil spike)	macro-shock	mode	✓
2026-03-17	SEC/CFTC joint crypto-classification guidance — bullish re...	regulatory	mino	
2026-03-22	Resolv USR SERVICE_ROLE/AWS-key exploit depeg — idiosyncra...	protocol-failure	majo	
2026-05-13	Hot PPI inflation shock + Warsh Fed-chair transition (onse...	macro-shock	majo	✓
2026-05-18	May 2026 flash crash below $77k (Treasury yields >5.1%, Ir...	deleverage-cascade	mode	✓
2026-05-28	Second US-Iran escalation (Hormuz strikes; BTC 6-week low,...	macro-shock	majo	✓
2026-06-01	Strategy first BTC sale in 4 years (8-K trigger) + ETF-out...	deleverage-cascade	majo	✓
2026-06-03	Strait-of-Hormuz / Kuwait-airport war escalation (Brent >$...	macro-shock	majo	✓