

Do Cryptocurrency Markets Differentiate Infrastructure from Regulatory Shocks? A Multi-Moment Event Study with Dependence-Robust Inference

Murad Farzulla^{1,2*}

^{1*}King’s Business School, King’s College London, London, United Kingdom.

²Dissensus, London, United Kingdom.

Corresponding author(s). E-mail(s): murad@dissensus.ai;

Abstract

Do cryptocurrency markets process infrastructure failures (exchange outages, protocol exploits, network disruptions) differently from regulatory shocks (enforcement actions, policy changes)? A natural prior, and one an earlier version of this work reported, is that they do: a six-asset GJR-GARCH-X panel suggested infrastructure events generate a roughly fivefold larger conditional-variance response than regulatory events, apparently significant. We test the claim properly at *both* moments of the return distribution over a single shared sample (50 events, six assets, January 2019–August 2025), using one dependence-robust design instantiated twice: an event-level block bootstrap with the Ibragimov–Müller few-cluster test at the first moment, and a Student-*t*-copula CCC-GARCH-X bootstrap at the second. Neither moment supports a statistically distinguishable asymmetry. At the first moment, cumulative abnormal returns differ by **+7.19** percentage points (infrastructure less negative), with a block-bootstrap $p = \mathbf{0.283}$ and Ibragimov–Müller $p = \mathbf{0.295}$; the 95% confidence interval $[-\mathbf{5.97}, +\mathbf{20.48}]$ comfortably crosses zero. At the second moment, the point estimate is directional ($\approx \mathbf{2.6-5\times}$, infrastructure larger) but the Student-*t*-copula bootstrap—which redraws innovations from the fitted heavy-tailed margins and preserves cross-asset dependence—returns $p = \mathbf{0.323}$ at baseline and $p = \mathbf{0.321}$ with a crisis-regime control. The apparent significance of the earlier variance result was an inference artefact: treating six strongly cross-correlated per-asset coefficients ($\bar{\rho} \approx \mathbf{0.69}$) as independent observations, and drawing bootstrap innovations from a Gaussian rather than the fitted Student-*t*,

both manufactured precision the data do not contain. Correcting each turns the headline into a failure to reject.

We frame this self-correction as the paper’s lead contribution: a four-part inference cautionary tale (pseudoreplication, cross-asset dependence, heavy-tailed bootstrap innovations, and block-versus-parametric resampling at the first moment) drawn from overturning our own earlier estimate. The result is an honest dual failure-to-reject, not a claim of “no effect”: the study is underpowered at the first moment (minimum detectable effect $\approx 19\text{--}20$ percentage points against an observed ≈ 7), so directional point estimates that the sample cannot resolve are exactly what we report. Two findings survive. First, a scope condition: the directional variance differential ($\approx 5\times$ point estimate, not significant under the inference of record) concentrates in *curated, high-salience* event identification, not in a mechanical impact screen ($1.3\text{--}1.6\times$, also never significant) applied to a broad candidate pool—we measure selection bias as an object rather than asserting it away. Second, at the data’s native weekly frequency, GDELT news sentiment Granger-leads volatility for 10 of 18 asset-sentiment pairs (7 surviving a Benjamini–Hochberg correction at $q < 0.05$); a prior daily test that forward-filled the weekly series found nothing, an artefact of the step-function predictor the forward-fill creates. This lead survives Toda–Yamamoto lag-augmentation and a Bitcoin-co-movement control (7 of 7 FDR survivors) but is concentrated in SEC-litigation assets and substantially litigation-driven (3 of 7 survive a pre-litigation-window test; XRP’s lead is entirely a SEC v. Ripple-window phenomenon), so we report it as a litigation-concentrated lead-lag association requiring further validation rather than a general anticipatory channel.

Keywords: Cryptocurrency, Event study, Multi-moment analysis, Dependence-robust inference, Block bootstrap, GJR-GARCH-X, Pseudoreplication

1 Introduction

A recurring question in cryptocurrency risk management is whether markets differentiate between the *type* of bad news, or whether “all bad news is equivalent.” Two event categories are natural candidates for distinct processing. Infrastructure failures—exchange outages, protocol exploits, bridge hacks, network disruptions—create immediate mechanical impairment of trading and settlement [1, 2]. Regulatory shocks—enforcement actions, policy changes, jurisdictional rulings—operate through expectation channels, requiring participants to reassess legal constraints and compliance costs [3, 4]. If markets price these channels differently, the difference should show up somewhere in the return distribution: in the first moment (cumulative abnormal returns), the second moment (conditional variance), or both.

An earlier version of this work answered the question affirmatively at the second moment [5]. Estimating GJR-GARCH-X models on a six-asset panel, it reported that infrastructure events generate a $5.7\times$ larger conditional-variance response than regulatory events, with an apparently decisive significance level. The companion first-moment analysis [6] found no return-level difference, and the two were presented as a

tidy story: the market differentiates shock types through the risk channel, not expected returns—structure that lives in the second moment.

That story does not survive correct inference. This paper re-examines the same events at both moments under one dependence-robust design, and the asymmetry dissolves at *both*. The lead contribution is therefore not a finding about cryptocurrency but a finding about inference, told as a self-correction: when the cross-asset dependence and the heavy tails that pervade this data are handled properly, an estimate the author himself previously reported as a significant fivefold effect becomes a directional point estimate that the data cannot distinguish from zero. We think this is worth reporting plainly. The apparent volatility asymmetry in prior work—including our own—was an inference artefact, and the discipline of following the correct inference dissolved a headline we had every incentive to keep.

1.1 What we test, and what we find

We assemble a single shared sample: 50 carefully-identified events (26 infrastructure, 24 regulatory) across six cryptocurrencies (BTC, ETH, XRP, BNB, LTC, ADA), January 2019 to August 2025, with a decomposed GDELT sentiment series. We then test the infrastructure-regulatory difference at both moments of the return distribution, using one inference design instantiated twice—once for the mean, once for the variance—each calibrated to the same source of trouble: the six assets see the same events and are strongly cross-correlated, so any test that treats per-asset or per-observation quantities as independent will overstate precision.

First moment (returns). An event-level block bootstrap that resamples whole events (preserving the cross-sectional correlation of returns within an event) and an Ibragimov–Müller few-cluster t -test on event-level mean cumulative abnormal returns (CARs). On the shared six-asset/50-event basis, infrastructure CARs are -0.23% and regulatory CARs -7.42% , a difference of $+7.19$ percentage points (infrastructure *less* negative). The block-bootstrap two-sided p is 0.283 , the Ibragimov–Müller p is 0.295 , and the 95% confidence interval $[-5.97\%, +20.48\%]$ crosses zero comfortably. The observed effect (≈ 7 pp, Cohen’s $d = 0.30$) sits well below the minimum detectable effect for this sample (≈ 19 pp at 80% power)—an underpowered failure to reject, not a demonstration of equality.

Second moment (variance). A Student- t -copula CCC-GARCH-X parametric bootstrap. Standardised innovations are redrawn from the fitted Student- t margins (degrees of freedom $\nu \approx 3$) under the cross-asset standardised-residual correlation, and the infrastructure-regulatory variance multiplier is re-estimated on each replicate. The point estimate is directional— $\approx 4.9\times$ at baseline, attenuating to $\approx 2.6\text{--}5\times$ under structural-break controls—but the bootstrap returns $p = 0.323$ (baseline) and $p = 0.321$ (crisis-regime control). The multiplier is directional, not significant.

Both moments thus show directional point estimates that do not survive dependence-robust inference. We are deliberate about the framing. This is a *dual failure to reject*, not a claim that infrastructure and regulatory shocks are identical or that the effects are zero. At the first moment the study is plainly underpowered; at

the second the point estimate is non-trivial but cannot be distinguished from zero once dependence and heavy tails are accounted for. The honest summary is: we cannot reject equality of infrastructure and regulatory event impacts at either moment, on this sample, under correct inference.

1.2 Why the asymmetry was an artefact

The earlier variance headline rested on two inferential moves that, examined closely, manufacture precision. First, the significance test treated the six per-asset GJR-GARCH-X coefficients as six independent observations. They are not: all six assets are exposed to the same 50 events and correlate 0.54–0.83 (mean $\bar{\rho} \approx 0.69$), so the effective sample size is far below six. This is textbook pseudoreplication, and it is responsible for the apparent $p \approx 0.001$. Second, the cross-asset bootstrap that was meant to address dependence drew its standardised innovations from a Gaussian even though each model is fitted with Student- t errors at $\nu \approx 3$. A unit-variance Student- t at $\nu \approx 3$ concentrates its mass in the body and carries its variance in rare extreme tails; the Gaussian draw therefore mis-specifies the null distribution and biases the bootstrap p downward. Correcting the first move—re-testing at the event level and applying a design-effect correction with the right reference distribution—moves the significance from $p \approx 0.001$ to $p \approx 0.07$ –0.15 (marginal-to-null); correcting the second moves the cross-asset bootstrap from a spuriously-low $p \approx 0.06$ (a heavy-tail mis-specification artefact) to $p = 0.323$.

We organise this as an *inference ladder* (Section 6, Table 7): the *same* infrastructure-versus-regulatory comparison evaluated by seven progressively-more-correct inference procedures, from the dead headline ($t = 4.768$, $p = 0.0008$) down to the Student- t -copula bootstrap that is the inference of record ($p = 0.322$). Each rung is named and explained; the rungs *bracket* the comparison rather than climbing monotonically—the naive test far below at 0.0008, the low-power model-free rungs above at $p \approx 0.5$, and the calibrated bootstrap settling it at 0.322. We then turn the ladder from an anecdote into a demonstrated general result with a Monte-Carlo size study (Section 6.8): simulating a thousand panels under a fitted true null of no differential event effect, the naive i.i.d. rule that produced the headline rejects a true null 51% of the time at the 5% level—roughly ten times its nominal size—while the Student- t -copula bootstrap restores size to nominal (0.039 at 5%). The ladder and the size study together form the paper’s central methodological contribution.

1.3 What survives

Two substantive results do survive the re-examination and provide the paper’s ballast.

A scope condition on the variance differential. The directional second-moment point estimate ($\approx 5\times$, not significant under the inference of record) concentrates in *curated*, *high-salience* event identification, not in a mechanical impact threshold. Applying the same impact filter mechanically, at increasing strictness, to a broad reconstructed candidate pool yields a modest 1.3–1.6 $\times$ multiplier that is not significant at any threshold; the like-for-like two-asset screen gives 1.61 $\times$ ($p = 0.16$), not the curated $\approx 5\times$. We treat the event-inclusion rule as an explicit researcher

degree of freedom and trace the multiplier across it (Section 7.1), converting the standard selection-on-the-dependent-variable objection into a measured object. A drop-out census of the candidate pool further shows the mechanical screen does not favour infrastructure—if anything regulatory candidates pass at a higher rate (75.5% vs 64.6%).

A weekly, litigation-concentrated sentiment lead. At the data’s native weekly frequency, GDELT news sentiment Granger-leads volatility for 10 of 18 asset-sentiment pairs, 7 surviving a Benjamini–Hochberg correction at $q < 0.05$ (Section 7.4). A prior daily test found nothing, but that null was an artefact: forward-filling the weekly sentiment to daily turns the predictor into a step function constant within each week, which cannot carry daily lead-lag information. We stress-test the recovered lead with five corrections: it survives non-stationarity (Toda–Yamamoto lag-augmentation), Bitcoin co-movement, and zero-imputation (7 of 7 FDR survivors robust), but is concentrated in SEC-litigation assets and substantially litigation-driven—only 3 of 7 retain the lead outside litigation windows, and XRP’s vanishes outside the SEC v. Ripple period. We therefore report it as a litigation-concentrated lead-lag association requiring further validation, not a general anticipatory channel. It remains an instance of the paper’s broader theme—that inference choices, not the data, generated a spurious result—and we keep it as a methodological exhibit alongside its qualified substantive content.

1.4 Contribution and structure

The contribution is threefold. First, and centrally, an inference cautionary tale: a worked, self-refuting demonstration of how pseudoreplication and heavy-tailed misspecification can manufacture a significant cross-asset event-study result, and how the correct dependence-robust inference dissolves it—at both moments of the return distribution simultaneously. We make this concrete in two ways: an *inference ladder* that walks the identical comparison from $p = 0.0008$ to $p = 0.323$ as the inference is corrected rung by rung, and a *Monte-Carlo size study* that shows the naive rule over-rejects a true null by an order of magnitude while the inference of record restores nominal size. Second, the scope condition: a quantified mapping of how the variance differential responds to the event-inclusion screen, measuring selection bias rather than asserting it. Third, the weekly sentiment-leads-volatility result, recovered once the frequency artefact is removed. Section 2 situates the work; Section 3 describes the shared sample and the matched dependence-robust design; Section 4 reports the first moment and Section 5 the second; Section 6 sets out the inference ladder, the four lessons, and the Monte-Carlo size study; Section 7 reports the surviving robustness battery and the sentiment result; Sections 8–9 discuss and conclude.

2 Literature and Background

2.1 Cryptocurrency volatility and event studies

Cryptocurrency volatility is heavy-tailed, strongly clustered, and near-integrated. GARCH-family models outperform simpler estimators [7, 8], with asymmetric specifications often preferred given pronounced leverage-type effects [9]—though the sign

of the asymmetry is contested in crypto: Baur and Dimpfl [10] document an *inverted* leverage effect, in which positive return shocks amplify volatility more than negative ones. Cross-asset dependence is high: Makarov and Schoar [11] document persistent cross-exchange price gaps, Bouri et al. [12] significant volatility spillovers, and Kim et al. [13] increasing correlation with Bitcoin typically identified as the dominant transmitter; the same systemic cross-asset co-movement motivates aggregate crypto risk indices [14]. This dependence is precisely what makes naive cross-asset inference unreliable, a point this paper makes central.

Event-study methodology faces particular challenges in crypto: continuous 24/7 trading eliminates natural event windows, global participation creates timezone-dependent liquidity, and retail dominance changes information processing [15–17]. Chen et al. [1] adapt event studies for exchange failures, finding immediate liquidity contractions; Grobys [2] use GARCH-based event studies for DeFi exploits. The two literatures use incompatible windows—short for infrastructure (mechanical impact), long for regulation (gradual absorption)—which precludes direct comparison. The cleanest way to compare them is to hold the sample, the model, and the inference fixed and vary only the event type, which is what we do.

2.2 Dependence-robust and few-cluster inference

The methodological core of this paper is the correct treatment of dependence. When abnormal returns are cross-sectionally correlated, standard event-study t -tests over-reject [18], and event-induced variance compounds the problem [19]. The cluster-robust literature [20–22] recommends resampling at the level of the dependent unit and, with few clusters, conservative procedures such as the Ibragimov and Müller [23] t -test, which treats each cluster as a single observation. Recent work emphasises that event-study inference is fragile to exactly these choices [24, 25]. Our two instantiations—an event-level block bootstrap with Ibragimov–Müller at the first moment, and a dependence-preserving parametric bootstrap at the second—are applications of this literature to a setting where the dependence is severe and the heavy tails are extreme.

2.3 Sentiment and information processing

GDELT-based news sentiment provides a public alternative to social-media measures [26, 27]; Feinstein and Werbach [4] find regulatory-news coverage predicts price movements, and Liu and Tsyvinski [28] that attention proxies carry return information. No prior study decomposes GDELT into thematic (infrastructure vs regulatory) components or examines whether such coverage Granger-leads volatility at the appropriate frequency—a gap our weekly Granger analysis addresses.

3 Data and Methodology

The design principle is one matched dependence-robust inference, instantiated at two moments of the same return distribution, on one shared event sample. We first describe the sample, then the two instantiations.

3.1 Shared Sample

3.1.1 Assets and prices

Six cryptocurrencies are selected for continuous trading history over January 2019–August 2025: Bitcoin (BTC), Ethereum (ETH), XRP, Binance Coin (BNB), Litecoin (LTC), and Cardano (ADA). Daily closing prices from CoinGecko yield 2,350 observations per asset (14,100 total). Log returns are expressed in percentages, with extreme outliers winsorised. The sample exhibits the characteristic crypto features—excess kurtosis (5.23–8.91), negative skewness (−0.42 to −0.71), and annualised volatility from 54.3% (BTC) to 77.3% (XRP) (Table 1). Cross-asset correlation is high: BTC-ETH is the strongest pair (0.83), XRP the weakest-correlated (0.54–0.65), reflecting its distinct regulatory trajectory during the SEC litigation period (Figure 1). This dependence—mean pairwise $\bar{\rho} \approx 0.69$ across the panel—is the quantity that governs all the inference below.

3.1.2 Event sample

The event sample is the unified basis for both moments: 50 events, classified binary infrastructure (26) versus regulatory (24), identified by a systematic four-stage protocol. Candidates were drawn from five independent sources (blockchain news aggregators, mainstream financial media, official regulatory filings, security-incident databases, and exchange status pages), screened for verifiable UTC timestamps, public documentation, and demonstrable cross-sectional impact (at least two assets), classified by primary information channel, and reduced from 208 initial candidates to the final 50. Infrastructure events affect transaction and settlement mechanics (exchange failures, protocol exploits, network disruptions, crypto-exposed bank failures); regulatory events alter the legal environment (enforcement actions, legislative developments, policy announcements, jurisdictional changes). Boundary cases (e.g. exchange enforcement actions, stablecoin depegs) were assigned by primary channel; the resulting split is 26 infrastructure / 24 regulatory. Selected events appear in Appendix A.

Both moments are estimated on *this* sample. Every one of the 50 events produces a valid measurement on the six-asset basis—each asset has at least 120 estimation days for every event window, so the event counts are 26 versus 24 throughout, with no silent dropping. We note that the first-moment estimates are run on the full binary 50-event split (mixing positive- and negative-valence events) to match the variance basis exactly; a valence-controlled negative-only subset is retained as a robustness row (Section 4), where the CAR magnitudes are not directly comparable to the full-sample binary numbers.

3.1.3 Sentiment

GDELT-based sentiment indices are constructed by a three-stage decomposition: hierarchical keyword filtering separating regulatory content (“SEC,” “regulation,” “compliance”) from infrastructure content (“hack,” “exploit,” “outage”); volume-weighted tone aggregation with recursive z -score detrending over 52-week rolling

windows; and thematic decomposition weighted by article coverage,

$$S_t^{\text{REG}} = S_t^{\text{GDELT}} \times \text{Proportion}_t^{\text{REG}}, \quad (1)$$

$$S_t^{\text{INFRA}} = S_t^{\text{GDELT}} \times \text{Proportion}_t^{\text{INFRA}}. \quad (2)$$

The series is natively weekly, with up to a 7-day temporal mismatch to daily volatility and 7% missing values (zero-imputed after the z -score transform, the conservative choice). The native weekly frequency is decisive for the Granger analysis of Section 7.4.

3.2 First Moment: Event-Level Block Bootstrap

The first-moment instantiation follows standard event-study methodology [16, 29] with cryptocurrency adaptations and a dependence-robust inference layer.

Windows. Estimation window of 250 trading days (120 minimum for newer assets), a 30-day gap to avoid anticipation contamination, and an event window of $[-5, +30]$ days (with $[0, +1]$ to $[0, +5]$ as robustness).

Abnormal returns. The primary specification is the constant-mean model, $AR_{it} = R_{it} - \bar{R}_i$, with $\bar{R}_i$ estimated over the estimation window; a market model using BTC as proxy, $AR_{it} = R_{it} - (\hat{\alpha}_i + \hat{\beta}_i R_{\text{BTC},t})$, is the robustness alternative. Cumulative abnormal returns over $[\tau_1, \tau_2]$ are $CAR_i(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} AR_{it}$.

The inference problem. Pooling asset-event observations as i.i.d. inflates the degrees of freedom: with 50 events $\times$ 6 assets the naive count is 300, but the true unit of independence is the event, and the assets within an event are strongly cross-correlated. A naive t -test on the pooled observations would therefore overstate power—the first-moment analogue of the pseudoreplication that inflated the variance headline.

The dependence-robust solution. We resample at the event level. For each event we first average the per-asset CARs into a single event-level CAR (equal-weighting events regardless of asset coverage), then bootstrap by resampling *event IDs* with replacement, keeping all assets within an event together so that within-event cross-sectional correlation is preserved [19–21]. We take 5,000 bootstrap replications (seed fixed) for percentile confidence intervals. As a few-cluster cross-check we apply the Ibragimov–Müller test [22, 23]: a Welch t -test on the event-level mean CARs, treating each event as an independent cluster. We also report, for reference only, a Kolari–Pynnonen cross-sectional-correlation adjustment [18], but the block bootstrap and Ibragimov–Müller test are the inference of record.

3.3 Second Moment: Student- t -Copula CCC-GARCH-X Bootstrap

The second-moment instantiation estimates the conditional-variance response to each event type within a GJR-GARCH-X framework, with inference from a dependence-preserving parametric bootstrap calibrated to the same cross-asset correlation.

The model. The mean equation is a constant conditional mean, $r_t = \mu + \varepsilon_t$, justified empirically: Ljung–Box tests find serial correlation in only 8 of 18 (asset $\times$ lag) cells, the fitted AR(1) coefficient is tiny and negative everywhere (-0.025 to -0.067), and refitting with an AR(1) mean leaves the event coefficients essentially unchanged

(Section 7). The variance equation is the GJR-GARCH-X of Glosten et al. [30] with exogenous regressors,

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 \mathbb{I}(\varepsilon_{t-1} < 0) + \beta \sigma_{t-1}^2 + \delta_{\text{infra}} D_t^{\text{infra}} + \delta_{\text{reg}} D_t^{\text{reg}} + \theta_1 S_t^{\text{REG}} + \theta_2 S_t^{\text{INFRA}}, \quad (3)$$

where D_t^i are event dummies over the $[-3, +3]$ -day window and the coefficients δ are additions to conditional variance in squared percentage points. Parameters are estimated by quasi-maximum likelihood with Student- t innovations. Two near-integration concerns arise and are addressed by distinct results. First, the exogenous covariates may be persistent: Han and Kristensen [31] establish QMLE consistency for GARCH-X with stationary *and* nonstationary ($I(1)$) covariates, which covers the persistent sentiment regressors. Second, and separately, the GARCH *volatility persistence* itself is near-integrated ($\alpha + \beta \rightarrow 1$); the asymptotic theory for QMLE under near-unit-root GARCH persistence [32, 33] establishes consistency and asymptotic normality of the estimator but is known to degrade the reliability of Hessian-based standard errors as persistence approaches unity. Our inference of record sidesteps this entirely: the Student- t -copula bootstrap (Section 3.3) does not rely on Hessian standard errors at all, generating the null distribution by resampling under the fitted DGP, so the significance verdict is robust to the near-integrated-persistence concern even where analytic standard errors would be fragile. Covariance stationarity is enforced via $\alpha + \beta + \gamma/2 < 1$.

The quantity of interest is the cross-asset infrastructure-regulatory multiplier, $\bar{\delta}_{\text{infra}}/\bar{\delta}_{\text{reg}}$.

The inference problem. The six per-asset coefficients are not independent: the assets see the same events and correlate $\bar{\rho} \approx 0.69$. A test treating them as six i.i.d. observations is pseudoreplication and overstates precision (Section 6). The relevant innovation distribution is also heavy-tailed—the fitted Student- t degrees of freedom are $\nu \approx 3$ —so a bootstrap that draws Gaussian innovations mis-specifies the null.

The dependence-robust solution. We use a Student- t -copula CCC-GARCH-X parametric bootstrap ($B = 2000$). Standardised innovations are drawn from a multivariate- t copula—a latent multivariate normal at the standardised-residual correlation R_z , divided by a shared chi-square mixing variable (inducing joint tail dependence), mapped through uniforms to per-asset Student- t margins at the *fitted* ν , and rescaled to unit variance by $\sqrt{(\nu - 2)/\nu}$ (the variance recursion carries the scale). Each replicate refits the joint model and re-estimates the multiplier unrestricted; the null distribution is generated under a single common event coefficient. This is the inference of record for the second moment. Degenerate refits ($|\delta| > 50$) are dropped and counted (drop rate $\approx 7\%$ at baseline). Crucially, this corrects both inference defects in the earlier headline simultaneously: it respects the cross-asset dependence (so it is not pseudoreplicated) and it draws from the fitted heavy tails (so the null distribution is correctly heavy).

One design, two instantiations. Both instantiations share a single logic: identify the unit across which the data are dependent (the event, and the cross-asset correlation), resample at that level or simulate that dependence directly, and never

treat correlated quantities as independent. The first moment resamples events non-parametrically; the second simulates the joint heavy-tailed dependence parametrically. They are the same dependence-robust principle adapted to the two estimands.

3.4 Model selection and multiple testing

For the variance models, three nested specifications are compared per asset (GARCH, GJR-GARCH, GJR-GARCH-X) by AIC, and Benjamini–Hochberg corrections [34] at FDR 10% control Type I error across the per-asset event tests. For the structural-break analysis, change-points are detected via PELT [35] with Bai-Perron-style trimming [36, 37].

4 Results: First Moment (Returns)

4.1 Descriptive statistics

The panel comprises 2,350 daily observations per asset (Table 1). Heavy tails and negative skewness confirm the appropriateness of Student- t innovations for the variance models; the high cross-asset correlation (Figure 1) is the inference-relevant feature.

Table 1 Descriptive Statistics: Daily Returns (%). Annualised volatility uses the conventional $\sqrt{252}$ scaling for comparability and is illustrative only.

Stat	BTC	ETH	XRP	BNB	LTC	ADA
Mean	0.12	0.14	0.08	0.15	0.06	0.11
Std Dev	3.42	4.23	4.87	3.89	4.51	4.48
Skewness	−0.42	−0.58	−0.71	−0.45	−0.53	−0.67
Kurtosis	6.24	7.33	8.91	5.87	5.23	7.12
Ann. Vol	54.3	67.1	77.3	61.7	71.5	71.2

4.2 The first-moment null

Table 2 reports the infrastructure-regulatory CAR comparison on the shared six-asset/50-event basis, under both the block bootstrap and the Ibragimov–Müller test. Infrastructure CARs average -0.23% and regulatory CARs -7.42% , a difference of $+7.19$ percentage points: infrastructure returns hold up *less* negatively than regulatory returns. But the inference is unambiguous. The block-bootstrap two-sided p is 0.283 (one-sided 0.141), the Ibragimov–Müller p is 0.295, and the 95% confidence intervals— $[-5.97\%, +20.48\%]$ (block) and $[-6.46\%, +20.83\%]$ (Ibragimov–Müller)—cross zero with room to spare. We cannot reject equality of infrastructure and regulatory return responses.

Robustness within the first moment. Dropping XRP (the asset with the distinct regulatory trajectory) leaves the conclusion intact: a $+6.06$ -pp difference,

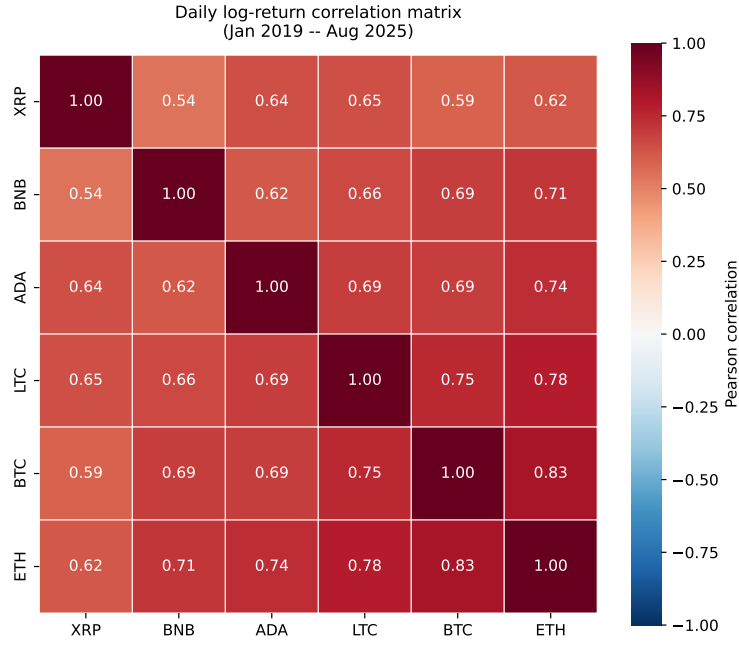


Fig. 1 Daily log-return correlation matrix across the six-asset panel, January 2019–August 2025. BTC and ETH form the most tightly correlated pair (0.83); XRP correlates least with the others (0.54–0.65). The mean pairwise correlation $\bar{\rho} \approx 0.69$ is the quantity that governs the dependence-robust inference throughout.

Table 2 First-moment comparison: infrastructure vs regulatory cumulative abnormal returns, shared six-asset/50-event basis. Block bootstrap resamples whole events ($B = 5,000$); Ibragimov–Müller (IM) is a Welch t -test on event-level mean CARs. The five-asset (ex-XRP) row and the valence-controlled negative-only row are robustness checks.

Specification	CAR _{infra}	CAR _{reg}	Diff (pp)	Block p	IM p
Six-asset (incl. XRP), $n = 26/24$	−0.23%	−7.42%	+7.19	0.283	0.295
Five-asset (ex-XRP), $n = 26/24$	+0.91%	−5.15%	+6.06	0.400	0.405
Negative-only (valence-controlled), $n = 8/7$	−7.9%	−9.4%	+1.5	0.916	0.927

Six-asset 95% CI: $[-5.97\%, +20.48\%]$ (block), $[-6.46\%, +20.83\%]$ (IM). Observed Cohen’s $d = 0.30$; pooled SD 24.1 pp; minimum detectable effect ≈ 19.1 pp at 80% power, $\alpha = 0.05$. The negative-only row is the legacy first-moment basis carried over for comparability: it uses a different four-asset universe (BTC, ETH, SOL, ADA—adding Solana and omitting XRP, BNB, and LTC, the latter for want of continuous early-period coverage in that cache) and restricts to the negative-valence events only (those whose primary price impact is adverse, coded `Infra.Negative/Reg.Negative` by `type.detailed`, as distinct from positive-valence events such as the Bitcoin ETF approval). Of the 50 binary events, 15 (30%) are negative-valence (8 infrastructure, 7 regulatory); the remaining 70% are positive- or mixed-valence and are excluded from this row. Its CAR magnitudes are therefore not directly comparable to the full binary 50-event rows.

block $p = 0.400$, IM $p = 0.405$. A valence-controlled negative-only subset (eight infrastructure vs seven regulatory negative-valence events, on the legacy four-asset BTC/ETH/SOL/ADA basis; see Table 2 note for the universe and valence-coding detail) gives a +1.5-pp difference with $p \approx 0.92$ – 0.93 across both bootstrap weighting schemes and an exact permutation test ($\binom{15}{8} = 6,435$ assignments, two-tailed $p = 0.93$). Under the market-model specification the point estimate is somewhat larger ($\Delta_{\text{MM}} \approx +7.3$ pp on the negative subset) but remains economically small and statistically indistinguishable from zero. The null does not depend on the weighting scheme, the asset universe, the valence control, the inference method, or the return model. One caveat on the constant-mean normal-return benchmark: because the infrastructure events cluster temporally in the 2022–2023 crisis window, an event’s 250-day estimation window can embed returns from prior nearby crisis episodes, depressing the estimated $\bar{R}_i$ and thereby shrinking the measured abnormal return toward zero; if anything this biases the already-insignificant infrastructure CAR conservatively (toward the null), so it does not change the failure-to-reject conclusion, and the market-model specification—which conditions out the common Bitcoin path rather than a static mean—is the natural cross-check, and it agrees.

4.3 This is an underpowered failure to reject, not “no effect”

We are careful not to over-read the null. The observed six-asset difference of +7.19 pp is not negligible—it corresponds to Cohen’s $d = 0.30$ —but with a pooled standard deviation of 24.1 pp and an event-level sample of 26 versus 24, the minimum detectable effect at 80% power is approximately 19.1 pp (ex-XRP, ≈ 20.3 pp). The observed effect is roughly a third of what this sample could resolve. The honest statement is therefore “no statistically significant difference; wide confidence intervals; underpowered,” not “infrastructure and regulatory returns are identical.” The directional point estimate (infrastructure less negative) is real and worth reporting; the sample simply cannot determine whether it reflects a genuine difference or sampling noise. This caveat is symmetric with the second-moment finding, to which we now turn.

5 Results: Second Moment (Conditional Variance)

5.1 Model estimates

Table 3 reports the three nested variance specifications. GJR-GARCH-X attains the lowest AIC for four of six assets (BTC, ETH, LTC, XRP), with GARCH(1,1) marginally preferred for ADA and BNB. Persistence is near-integrated ($\alpha + \beta$ between 0.954 and 0.999, mean 0.985), a documented stylised fact of cryptocurrency volatility [7, 9, 12]; the structural-break analysis of Section 7 shows within-regime persistence is somewhat lower but imprecisely estimated. The GJR leverage parameter γ is uniformly small and slightly negative (-0.066 to -0.001), consistent with the inverted crypto leverage effect [10] rather than with conventional equity-style asymmetry, and present already in the event-free GJR-GARCH column.

Table 3 Model Comparison: GARCH vs GJR-GARCH vs GJR-GARCH-X (AIC; lowest in bold)

Asset	Model	AIC	BIC	LogLik	γ
BTC	GARCH(1,1)	11904	11933	−5947	−
	GJR-GARCH	11906	11940	−5947	−0.012
	GJR-GARCH-X	11900	11964	−5939	−0.001
ETH	GARCH(1,1)	13345	13374	−6667	−
	GJR-GARCH	13347	13381	−6667	−0.010
	GJR-GARCH-X	13329	13393	−6654	−0.013
XRP	GARCH(1,1)	13324	13353	−6657	−
	GJR-GARCH	13325	13360	−6657	−0.056
	GJR-GARCH-X	13323	13387	−6651	−0.027
BNB	GARCH(1,1)	11400	11429	−5695	−
	GJR-GARCH	11401	11435	−5694	−0.039
	GJR-GARCH-X	11400	11463	−5689	−0.017
LTC	GARCH(1,1)	13780	13809	−6885	−
	GJR-GARCH	13774	13808	−6881	−0.066
	GJR-GARCH-X	13772	13836	−6875	−0.040
ADA	GARCH(1,1)	14091	14120	−7041	−
	GJR-GARCH	14093	14128	−7041	−0.010
	GJR-GARCH-X	14092	14156	−7035	−0.033

γ = GJR leverage parameter; uniformly small and negative, consistent with the inverted crypto leverage effect.

5.1.1 Variance-equation adequacy: no residual ARCH

Before interpreting the event coefficients, we verify that the GJR-GARCH-X variance equation fully absorbs the conditional heteroskedasticity—otherwise unmodelled volatility clustering, which is concentrated in the 2022–2023 window where the infrastructure events cluster, could be soaked up by the D^{infra} dummy and mechanically inflate δ_{infra} . Table 4 reports the McLeod–Li portmanteau (Ljung–Box Q on squared standardised residuals z_t^2 at lags 5/10/20, with the Li–Mak [38] degrees-of-freedom adjustment $\text{df} = \text{lag} - 2$) and the Engle ARCH-LM test, per asset. No asset shows significant residual ARCH at the 5% level on *any* test (0 of 6 on every lag, under both adjusted and naive degrees of freedom; the closest cell, Bitcoin’s lag-5 Ljung–Box, is $p = 0.086$, and its ARCH-LM(5) $p = 0.247$ confirms no actual residual ARCH). The standardised residuals are also well-behaved at the level— $\text{mean}(z) \approx 0$ throughout—with residual excess kurtosis (2.3–5.9) expected from the Student- t margins ($\nu \approx 3.1$ –4.6) rather than an ARCH symptom. The variance equation is adequate: the event coefficients are estimated on properly whitened squared residuals, so the curated $\approx 4.88\times$ multiplier is not an artefact of unmodelled 2022–2023 volatility clustering being absorbed by the infrastructure dummies, and does not require flagging as

an upper bound on heteroskedasticity grounds. (The portmanteau rules out remaining *linear* ARCH in z^2 , not higher-order nonlinearity or regime structure—handled separately by the break analysis of Section 7.3.3—and the separately documented caveats on the multiplier, pseudoreplication and the curated-versus-mechanical scope condition, are independent of this and unaffected.)

Table 4 Variance-equation diagnostics on the baseline GJR-GARCH-X fit: Ljung–Box Q on squared standardised residuals (Li–Mak adjusted p , $\text{df} = \text{lag} - 2$) and Engle ARCH-LM p . No asset shows significant residual ARCH at 5% on any test (0/6 throughout).

Asset	$\text{LB}(z^2) \ p_5$	$\text{LB}(z^2) \ p_{10}$	$\text{LB}(z^2) \ p_{20}$	ARCH-LM p_5	ARCH-LM p_{10}
BTC	0.086	0.167	0.171	0.247	0.303
ETH	0.734	0.463	0.623	0.941	0.521
XRP	0.483	0.745	0.911	0.780	0.874
BNB	0.317	0.272	0.390	0.628	0.475
LTC	0.304	0.535	0.721	0.610	0.718
ADA	0.275	0.337	0.498	0.558	0.495

Significant flags at 5%: 0/6 assets on every test. Li–Mak df subtract the ARCH+GARCH lag orders ($p + q = 2$); γ is a within-lag asymmetry term, not an extra lag order. Naive- df p -values (in the replication output) tell the same story.

The per-asset event coefficients (Table 5) point uniformly in the same direction—infrastructure larger than regulatory for all six assets—with a cross-asset mean of $\bar{\delta}_{\text{infra}} \approx 1.98$ versus $\bar{\delta}_{\text{reg}} \approx 0.41$ (squared percentage points). The point-estimate multiplier is $\approx 4.9\times$. The question is whether that direction is statistically distinguishable from zero once the cross-asset dependence and heavy tails are respected.

Table 5 Per-asset baseline event-impact coefficients (squared % points). All six assets show infrastructure $>$ regulatory; the direction is uniform, the multiplier per asset varies. The cross-asset mean multiplier is $4.88\times$, consistent with the baseline of Table 6.

Asset	Infra	Reg	Ratio
ADA	2.79	0.10	$28.2\times$
LTC	2.38	0.14	$16.5\times$
ETH	2.25	0.49	$4.6\times$
XRP	2.24	1.22	$1.8\times$
BNB	1.21	0.18	$6.7\times$
BTC	1.01	0.30	$3.4\times$
Mean	1.98	0.41	$4.88\times$

Ratios computed from full-precision coefficients. These are baseline point estimates, not the inference of record: per-asset significance tests treat the cross-correlated assets as independent (the pseudoreplication Section 6 corrects). The cross-asset mean multiplier ($4.88\times$) is the quantity whose significance is assessed by the copula bootstrap of Table 6.

5.2 The second-moment null

Table 6 reports the Student- t -copula CCC-GARCH-X bootstrap, the inference of record for the second moment. At baseline the multiplier is $4.88\times$ (the point estimate, unchanged from prior work), but the bootstrap p is 0.323: not significant at the 10% level, directional only. Adding a single FTX-crisis-regime control gives $3.97\times$ with $p = 0.321$ —the point estimate attenuates but the significance is essentially identical, so the differential is not merely the crisis regime’s elevated variance. Under the full set of regime dummies the multiplier is $2.64\times$ with $p = 0.284$ (this specification is numerically less reliable: $\approx 23\%$ of null refits hit degenerate optima). Across all three specifications the conclusion is the same: the variance differential is directional ($\approx 2.6\text{--}5\times$, infrastructure larger) but cannot be distinguished from zero under dependence-robust, heavy-tailed inference.

Table 6 Second-moment comparison: Student- t -copula CCC-GARCH-X bootstrap ($B = 2000$, curated $n = 50$ sample). Innovations are redrawn from the fitted Student- t margins ($\nu \approx 3$) under the cross-asset standardised-residual correlation. This is the inference of record. The point estimate is directional; none of the three specifications is significant.

Specification	$\bar{\delta}_{\text{infra}}$	$\bar{\delta}_{\text{reg}}$	Multiplier	Bootstrap p (2-sided)
Baseline (no break controls)	1.978	0.405	$4.88\times$	0.323
+ Crisis-regime control	1.914	0.483	$3.97\times$	0.321
+ Full regime controls [†]	2.443	0.926	$2.64\times$	0.284

One-sided p -values are 0.322, 0.318, and 0.249 respectively. [†]The full-regime bootstrap is numerically less reliable— $\approx 23\%$ of null refits reach degenerate optima, versus $\approx 7\%$ for baseline and $\approx 7\%$ for crisis—so its p should be read with caution. The point estimate ($4.88\times$) and direction are unchanged from prior work; what changes is the significance.

5.3 Both moments agree: directional, not significant

Reading the two moments together gives the paper’s central empirical result. At the first moment the infrastructure-regulatory CAR difference is $+7.19$ pp ($p = 0.283$); at the second the variance multiplier is $\approx 5\times$ ($p = 0.323$). Both point estimates run in the direction one might expect—infrastructure events distinct from regulatory ones—and *neither* survives dependence-robust inference. This is a symmetric dual failure to reject. It is not the original story (a return-level null sharpened by a significant variance asymmetry); the variance “signal” was, on correct inference, no more distinguishable from zero than the return-level difference. We turn now to why the earlier estimate appeared significant.

6 Why the Asymmetry Was an Artefact: The Inference Ladder

The earlier version of this work reported the variance asymmetry as a significant five-fold effect. The point estimate has not changed—infrastructure events do, on average,

carry a larger conditional-variance increment in this curated sample. What changed is the significance, and the change is entirely attributable to inference. This section is the paper’s central methodological contribution. We set it out in three parts: an *inference ladder* that walks the identical infrastructure-versus-regulatory comparison through seven progressively-more-correct inference procedures (Section 6.1); the four *lessons* the ladder teaches, stated as transferable principles (Sections 6.2–6.5); and a *Monte-Carlo size study* that converts the ladder from a single worked anecdote into a demonstrated general result (Section 6.8). We present all of it as self-correction: each rung and each lesson is one we learned by overturning our own published estimate.

6.1 The inference ladder

The cleanest way to see what happened is to hold the comparison fixed—the same six assets, the same 50 events, the same point estimate (infrastructure variance increment $\approx 5\times$ regulatory at the second moment, +7.19 pp at the first)—and vary *only* the inference procedure, from the most naive to the most correct. Table 7 is that walk. The point estimate is constant down every rung; the inference is not. The p -value does not climb monotonically. The naive test sits far below at 0.0008, the low-power model-free corrections overshoot above to $p \approx 0.5$, and the dependence-honest rungs cluster in between: the corrected design-effect correction at $p \approx 0.07$ –0.15, the calibrated Student- t -copula bootstrap at 0.322, and the model-free event-level proxies above them at $p \approx 0.5$. The one rung that sits *below* this dependence-honest cluster, at $p = 0.057$, is the Gaussian-copula bootstrap (rung 5)—and that is an artefact: it draws Gaussian innovations under fitted Student- $t(\nu \approx 3)$ margins and so over-rejects (Lesson 3, Section 6.8), pulling its p spuriously low. Reading the ladder top to bottom is reading the entire gap between the dead headline and the present null.

The ladder has a clear internal structure. Rungs (2)–(3) are model-free event-level proxies: they treat the event as the unit of independence (correct) but discard the GARCH structure (lossy), so they are low-powered and *understate* the effect, sitting high at $p \approx 0.5$. Rung (1) sits at the opposite extreme, far below at $p = 0.0008$. The dependence-honest rungs cluster in between: the corrected design-effect correction on the actual estimator at $p \approx 0.07$ –0.15 (rung 4), and the Student- t -copula bootstrap—once the innovations are drawn from the fitted heavy tails rather than a Gaussian—at $p = 0.322$ (rung 6, the inference of record). These three (corrected rung 4, the t -copula, and the event-level proxies) all point the same way: directional, not significant. The lone outlier among the dependence-aware rungs is the Gaussian-copula bootstrap (rung 5) at $p = 0.057$, which gets the dependence right but the tails wrong and so over-rejects (Lesson 3); the size study (Section 6.8) shows it rejects a true null 31% of the time, so its low p is the same under-dispersion error the naive rung (1) commits, relocated into the copula tails. The first-moment companion (rung 7) was, correctly, never significant to begin with. The four lessons that follow name the four corrections that move the p -value down the ladder.

Table 7 The inference ladder: the same infrastructure-versus-regulatory comparison, evaluated by progressively-more-correct inference procedures. The point estimate is unchanged at every rung; only the inference—and therefore the p -value—changes. Rungs (1)–(6) are the second-moment (variance) comparison; rung (7) is the first-moment (returns) companion on the shared basis. The naive test understates the p -value and the low-power model-free rungs (2)–(3) overstate it; rung (6), the Student- t -copula bootstrap, is the calibrated inference of record.

#	Inference procedure	Statistic	p
1	Naive i.i.d. t -test (pseudoreplication)	$t = 4.768$	0.0008
2	Event-level Welch test	—	0.52
	Event-level Mann–Whitney test	—	0.54
3	Cluster-robust panel (SE clustered by event)	—	0.50
4	Design-effect correction (eff. $N \approx 1.4$ –2.8)	$t \approx 2.3$	0.07–0.15
5	Gaussian-copula CCC-GARCH-X bootstrap	—	0.057
6	Student-t-copula CCC-GARCH-X bootstrap	—	0.322
7	First-moment event-level block bootstrap	+7.19 pp	0.283

Rungs (2)–(3) are model-free event-level proxies (low-powered, they understate the effect); rung (4) is the design-effect correction on the actual estimator, reported as a range because it is reference-distribution-sensitive (Section 6.2): the matching degrees of freedom are $df_{\text{eff}} = (N - 1)/\text{DEFF}$, not $N - 1$, and the correct correlation object is the cross-asset correlation of the per-asset *differences* $d_i = \delta_{\text{infra},i} - \delta_{\text{reg},i}$ ($\bar{\rho}_d = 0.23$), not the raw-return correlation (0.69); under these corrections rung (4) lands at $p \approx 0.07$ –0.15, marginal-to-null. Rungs (5)–(6) are the cross-asset parametric bootstrap, Gaussian-innovation versus the correct Student- t -margin draw. Rung (5) is the lone anomaly: it draws Gaussian innovations under a fitted Student- t ($\nu \approx 3$) margin and so over-rejects, pulling its p spuriously low (Lesson 3); rung (6) corrects this and is the inference of record (one-sided 0.322, two-sided 0.323). Rung (7) is the first-moment block bootstrap on the shared six-asset/50-event basis (Table 2).

6.2 Lesson 1: Pseudoreplication across cross-correlated assets

The original significance test was a t -test (and Mann-Whitney, and inverse-variance-weighted Z) on the six per-asset GJR-GARCH-X coefficients, treating them as six independent observations. This is pseudoreplication [39]: replicating the measurement (six assets) without replicating the unit that carries the independent information (the event). The six assets are exposed to the *same* 50 events and are strongly cross-correlated ($\bar{\rho} \approx 0.69$); their coefficients are tightly clustered not because the effect is precisely estimated but because they average the same correlated events. The tight clustering is the symptom most likely to be mistaken for precision: six near-identical numbers look like six confirmations when they are closer to one measurement repeated six times.

This is rung (1) of the ladder. An uncorrected t -test on the six coefficients reported $t = 4.768$, $p = 0.0008$. The correction is to re-test at the event level—the genuine unit of independence—which collapses the significance completely: an event-level Welch test gives $p = 0.52$ (rung 2), a Mann-Whitney $p = 0.54$, and a cluster-robust panel with standard errors clustered by event $p = 0.50$ (rung 3). These model-free proxies are low-powered: by averaging over the GARCH structure they discard information and so *understate* the effect, sitting high at $p \approx 0.5$, while the uncorrected test sits at the opposite extreme at $p = 0.0008$. The honest intermediate is rung (4): a design-effect correction on the *actual* estimator. Two refinements matter here, and they pull in opposite directions. First, a design-effect adjustment shrinks the *effective* sample

size to $N_{\text{eff}} = N/\text{DEFF}$, so the matching reference distribution is a t with $\text{df}_{\text{eff}} = (N - 1)/\text{DEFF}$ degrees of freedom—not $t(N - 1)$ and not the normal—which is a heavier-tailed reference that *raises* p .¹ Applying both, the corrected rung sits at $p \approx 0.07\text{--}0.15$ depending on the statistic and correlation input (the df_{eff} reference under $\bar{\rho}_{\text{ret}}$ gives $p \approx 0.12$; the correct corr-of-differences gives $p \approx 0.06\text{--}0.10$), marginal-to-null—never the 5% the uncorrected test implied, and the earlier crisp 0.067 over-states the rung by reading a normal tail where a $t(\text{df}_{\text{eff}} \approx 1\text{--}2.3)$ is required. The size study of Section 6.8 quantifies the damage directly: the rung-(1) rule rejects a true null 51% of the time at the nominal 5% level, and the design-effect rung’s apparent over-rejection is itself reference-dependent—read against the $t(N - 1)$ tail it over-rejects (43%), and read against its correct $t(\text{df}_{\text{eff}})$ critical value it improves sharply (to 22%) but remains oversized, confirming that the analytic correction narrows the reference-distribution gap without closing it (which is why the ladder does not stop at rung 4 but proceeds to the fully-calibrated copula bootstrap). The lesson is elementary but consequential: when N correlated series see the same events, N is not the sample size, and the gap between the nominal and the effective N is large enough here to turn $p \approx 0.1$ into $p = 0.0008$.

6.3 Lesson 2: Cross-asset dependence concentrates, it does not dilute

A natural hope is that the dependence penalty is mild—that the assets are “mostly” independent and that filtering out the common volatility dynamics will leave near-independent residuals to average over. The hope is wrong, and wrong in the unfavourable direction. The standardised-residual correlation, computed after the GJR-GARCH-X has stripped out each asset’s own conditional-volatility path, is if anything *higher* than the raw-return correlation: $\bar{\rho}_z = 0.705$ versus $\bar{\rho}_{\text{ret}} = 0.688$ (and stable at 0.701–0.703 across the break-control specifications). Conditioning on the GARCH dynamics therefore *concentrates* the cross-asset dependence rather than diluting it. The intuition is that the common component the per-asset GARCH filters is partly idiosyncratic volatility clustering; what remains—the contemporaneous comovement of shocks across assets on the same calendar day—is precisely the systematic part, and it is what binds the six event responses together.

The consequence is that the effective sample size implied by this correlation is ≈ 1.4 , not six, and no amount of averaging recovers the missing independence: a mean of six numbers that move together is barely more informative than one of them. Any valid inference procedure must therefore *simulate* the dependence directly rather than assume it averages out. This is what motivates the copula bootstrap of Section 3.3: drawing the six assets’ innovations from a multivariate- t copula at R_z reproduces the joint dependence, including the joint tail dependence (assets crashing together) that a Gaussian dependence structure would miss and that the per-asset tests ignore entirely. The size study makes the cost of ignoring it concrete: a procedure that gets the point

¹The averaged statistic is the per-asset signed difference $d_i = \delta_{\text{infra},i} - \delta_{\text{reg},i}$, so the design effect must be computed on the cross-asset correlation of the *differences*, $\bar{\rho}_d$, not on the raw-return correlation. Estimated from the Student- t -copula bootstrap draws of the d_i , $\bar{\rho}_d = 0.23$ (bootstrap-covariance $\text{DEFF} = 1.70$, Kish-form 2.16), far below the raw-return 0.69: differencing out the common market component within each asset leaves the infrastructure-versus-regulatory contrast largely idiosyncratic. The design-effect penalty is therefore milder than the raw-return correlation would imply, which *lowers* p relative to using $\bar{\rho}_{\text{ret}}$.

estimate right but the dependence wrong (the naive rule) cannot control its false-positive rate at all, rejecting a true null 51% of the time. Dependence is not a nuisance to be averaged down; on this data it is the dominant determinant of the inference.

6.4 Lesson 3: Bootstrap innovations must match the fitted tails

Even after moving to a cross-asset bootstrap that respects the dependence (rung 5), a second, subtler defect remained. The bootstrap drew its standardised innovations from a *Gaussian* while each model was fitted with Student- t errors at $\nu \approx 3$ (the per-asset fitted degrees of freedom are 3.18, 3.70, 3.14, 4.20, 3.95, 4.56, median 3.82). With $\nu \approx 3$ the true innovation is far heavier-tailed than Gaussian, so the Gaussian draw mis-specifies the shape of the null distribution. Correcting the draw to the fitted Student- t margins—a multivariate- t copula with per-asset ν , rescaled to unit variance by $\sqrt{(\nu - 2)/\nu}$ —moves the baseline p from 0.057 (rung 5, Gaussian) to 0.322 (rung 6, Student- t). The same correction moves the crisis-control spec from 0.061 to 0.318 and the full-regime spec from 0.109 to 0.249 (Table 8). The gap looks large, but the right way to read it is that the Gaussian rung is the *anomaly*, not the t -copula rung the mover: the 0.057 sits below the cluster of dependence-honest rungs (corrected rung 4 at ≈ 0.07 –0.15, the t -copula at 0.322, the event-level proxies at ≈ 0.5), and the size study (Section 6.8) shows why—the Gaussian-copula bootstrap over-rejects a true null at 31% versus the t -copula’s nominal 3.9%. The Gaussian rung’s low p is therefore the same too-thin-tails under-dispersion error the naive rung (1) commits, now in the copula’s tails; the t -copula is doing correct work and only appears to move the p -value sharply because it is measured against a rung the size study has already shown to be broken. This is the correction most likely to be missed, because the intuition runs the wrong way.

The mechanism is worth stating precisely, because the obvious heuristic is a false tripwire here. A *unit-variance* Student- t at $\nu \approx 3$ is *more* concentrated in its body than a Gaussian—about 80% of its mass lies within ± 1 versus 68% for the Gaussian—while still carrying its full unit variance, which it must therefore make up in rare, extreme tails. The bulk of bootstrap refits is consequently *tighter* under the t draw, not wider. The naive expectation (“heavier tails widen the null, so p falls”) holds only for an *un*-rescaled t , whose raw variance is $\nu/(\nu - 2) \approx 3$; but the specification correctly feeds unit-variance innovations, because the variance recursion σ_t^2 already carries the scale. What the heavy tails actually do, under unit-variance margins, is occasionally produce an extreme joint draw that shifts the null distribution’s *location* upward, toward the observed difference—which *raises* p . The null standard deviation does widen too (from 0.45 to 0.79 at baseline), but the location shift is the decisive channel. The correct diagnostic is therefore not “does the null widen” but “does p rise,” and it rises in every specification. The size study (Section 6.8) confirms this is not a coincidence of one dataset: the Gaussian-copula bootstrap, with the right dependence but the wrong tails, still over-rejects a true null at 31% (versus nominal 5%), whereas the Student- t -copula bootstrap—right dependence *and* right tails—lands at 3.9%, nominal up to Monte-Carlo error. Matching the bootstrap innovation distribution to the fitted tails is what turns the marginal $p = 0.057$ into the non-significant $p = 0.322$, and it is what restores correct size.

6.5 Lesson 4: Block versus parametric resampling at the first moment

The same dependence discipline governs the first moment (rung 7), where the failure mode is the converse of the variance case but the principle identical. Pooling the asset-event observations as i.i.d. would treat the panel as $50 \times 6 = 300$ independent draws when the genuine unit of independence is the ≈ 50 events; a pooled-observation or parametric test would therefore *understate* the standard error by inflating the degrees of freedom sixfold—the first-moment analogue of the pseudoreplication that inflated the variance headline. The block bootstrap that resamples whole events (keeping all six assets within an event together, so within-event cross-sectional correlation is preserved) restores the correct uncertainty: it widens the 95% confidence interval to $[-5.97\%, +20.48\%]$ where a naive pooled bootstrap would give a spuriously tight band. The Ibragimov–Müller few-cluster test, which goes further and treats each event as a single observation, triangulates the same conclusion from the conservative end ($p = 0.295$, CI $[-6.46\%, +20.83\%]$). Both land on the same verdict: +7.19 pp, not distinguishable from zero.

The general principle across both moments is one sentence: resample, or simulate, at the level of genuine independence—the event, and the cross-asset dependence structure—and never at the level of the correlated observation. The two moments are mirror images of the same error. At the first moment, applying the principle *widens* an interval that was already insignificant (so the conclusion is unchanged but the uncertainty is honestly stated); at the second, applying it *dissolves* an interval that had appeared significant (so the conclusion itself changes). In both cases the correct inference is more conservative than the naive one. The symmetry is the point: the original paper’s tidy story—a conservative, correctly-specified return-level null beside a sharp, mis-specified variance asymmetry—was never two findings. It was one well-specified inference next to one mis-specified inference, and the contrast between them was mistaken for substantive structure.

6.6 The Gaussian-versus-Student- t correction across specifications

Lesson 3 separates the lone anomalous rung from the dependence-honest cluster, so it is worth showing that the Gaussian-versus-Student- t gap operates uniformly, not just at baseline. Table 8 reports the same bootstrap under the Gaussian innovation draw (rung 5) and the corrected Student- t -copula draw (rung 6) for all three break-control specifications. The correction raises p in every case, and by a similar amount: baseline $0.057 \rightarrow 0.322$, crisis-control $0.061 \rightarrow 0.318$, full-regime $0.109 \rightarrow 0.249$. The null standard deviation widens in tandem (the t -copula’s heavier tails), and the standardised-residual correlation that drives the dependence is stable at $\bar{\rho}_z \approx 0.70$ throughout—higher than the raw-return 0.688, as Lesson 2 noted. The Gaussian column is the “right dependence, wrong tails” comparator, and (as the size study confirms) the over-rejecting one; the Student- t column is the fully-correct inference of record. The gap between the two columns is the entire content of Lesson 3: the Gaussian rung is the anomaly the t -copula corrects, consistent across specifications.

Table 8 The heavy-tail correction (Lesson 3) across all three specifications: Gaussian-innovation bootstrap (rung 5) versus the corrected Student- t -copula bootstrap (rung 6). The point estimate is unchanged; correcting the innovation draw raises the p -value in every specification. $B = 2000$; one-sided p reported.

Specification	Mult.	Gaussian p	t -copula p	Null SD ($G \rightarrow t$)	$\bar{\rho}_z$
Baseline (no break controls)	4.88×	0.057	0.322	0.455 $\rightarrow$ 0.787	0.705
+ Crisis-regime control	3.97×	0.061	0.318	0.473 $\rightarrow$ 0.841	0.703
+ Full regime controls [†]	2.64×	0.109	0.249	0.839 $\rightarrow$ 1.400	0.701

Two-sided t -copula p are 0.323, 0.321, 0.284. [†]The full-regime bootstrap is numerically less reliable ($\approx 23\%$ of null refits degenerate, versus $\approx 7\%$ for baseline and crisis), so its p should be read as “ ≥ 0.10 , not significant” rather than as a precise figure. Across all three the Student- t correction raises p and none is significant at the 10% level.

6.7 The cautionary tale, stated plainly

Taken together, the four lessons explain the entire gap between the earlier headline and the present null, and the ladder (Table 7) is the gap made visible. The $p \approx 0.001$ was pseudoreplication (Lesson 1); the dependence it ignored is severe and concentrating, not mild and diluting (Lesson 2); the $p \approx 0.06$ that a Gaussian cross-asset bootstrap then produced was a heavy-tail mis-specification (Lesson 3); and the first-moment companion’s null was, correctly, never significant to begin with (Lesson 4). The point estimate was always $\approx 5\times$; the significance was an artefact of inference at every stage. We report this not to flagellate the earlier estimate but because the demonstration is the most transferable thing in the paper: cross-asset event studies in heavy-tailed markets are exceptionally prone to manufacturing significance, and the corrections are specific and reproducible. The next subsection shows that the ladder is not a property of this one dataset but a general feature of these inference procedures, by simulating their behaviour under a known true null.

6.8 A Monte-Carlo size study: the ladder is a general result

The ladder shows what happened to *one* comparison as the inference was corrected. A sceptic could reply that we have simply walked a single dataset from a low p to a high one and that nothing general follows. The decisive test of that objection is a size study: simulate many datasets under a *known* true null of no differential event effect, run each rung of the ladder on every simulated dataset, and count how often each rung rejects. A correctly-sized test rejects a true null exactly α of the time; a test that rejects far more often manufactures false “significance” as a matter of construction, independently of any particular sample. This converts the ladder’s central claim—that the naive rule over-rejects and the dependence-robust rule does not—from an assertion about our data into a measured property of the procedures.

Design. We fit the true-null data-generating process to the actual data: per-asset GJR-GARCH-X dynamics estimated under a *null-imposed* combined event dummy ($D_{\text{event}} = D_{\text{infra}} + D_{\text{reg}}$), so that $\delta_{\text{infra}} = \delta_{\text{reg}}$ by construction and there is genuinely no

differential effect in the truth. Cross-asset dependence is imposed at the standardised-residual correlation $\bar{\rho}_z = 0.705$; innovations are Student- t at each asset’s fitted ν (median 3.82) with joint tail dependence from a true multivariate- t copula—the identical DGP as the inference-of-record bootstrap. The real 26-infrastructure / 24-regulatory event structure and the real event-window dummies are reused unchanged; only the returns are re-simulated. We draw $N = 1,000$ panels (923 usable after convergence screening), calibrate each bootstrap rung’s critical value from $B_{\text{ref}} = 4,000$ reference draws, and record the empirical rejection rate of each rung at the 5% and 10% nominal levels. The full DGP and screening rules are in Appendix D.

Result. Table 9 is the size table. The naive i.i.d. t -test—the rule that produced the headline $t = 4.768$, $p = 0.0008$ —rejects a true null **51.2%** of the time at the nominal 5% level and **70.1%** at the 10% level: roughly *ten times* and *seven times* its nominal size. Under the realistic null, a “significant” headline from this rule is a manufactured false positive a majority of the time. The design-effect rule’s size is reference-dependent, and this is itself instructive. Read against the $t(N-1=5)$ critical value the size study originally used—which applies the design-effect point penalty but ignores its consequence for the reference distribution—it rejects 43.4% / 65.5% of the time, apparently barely better than the naive rule. But that is the wrong-reference error of Lesson 1: the design effect shrinks the effective degrees of freedom to $\text{df}_{\text{eff}} \approx 2.3$ (computed on the cross-asset correlation of the per-asset *differences*, $\bar{\rho}_d = 0.23$, not the raw-return 0.69), and re-deciding the identical statistics against the correct $t(\text{df}_{\text{eff}})$ critical value brings its size down to **22.0%** / **53.5%**—a large improvement over both the naive rule and the $t(5)$ reference, but still over-sized. The analytic design-effect correction therefore narrows the gap without closing it: a closed-form effective-df adjustment cannot reproduce the joint GARCH-and-heavy-tail structure, which is why size control requires simulating the dependence and the tails directly rather than approximating them, as the copula bootstrap does. The Gaussian-copula bootstrap, which gets the dependence right but the tails wrong, rejects 30.8% / 39.8% of the time—still badly over-sized, the too-thin-tails defect Lesson 3 diagnosed, and the reason rung (5)’s $p = 0.057$ is the ladder’s lone anomaly. Only the Student- t -copula bootstrap restores nominal size: **3.9%** at the 5% level and **9.5%** at the 10% level, both within Monte-Carlo error of nominal.

The self-validation that makes this trustworthy. The Student- t -copula bootstrap is calibrated from the *same* DGP the panels are drawn from, so its size *must* land near nominal up to Monte-Carlo error—if it did not, the implementation would contain a bug. It comes out 0.039 / 0.095 against targets 0.05 / 0.10, both within ≈ 1.7 standard errors of nominal (mildly conservative at 5%, essentially exact at 10%). This is the tripwire that certifies the simulation: the procedure we adopt as the inference of record is demonstrably correctly-sized on data we know to be null. The over-rejection of rungs (i)–(iii) is therefore not an artefact of a broken simulator but a real property of those procedures.

What the size study establishes. The ladder is not a story about one dataset that happened to walk from $p = 0.0008$ to $p = 0.322$. It is the visible trace of a general fact: under a realistic heavy-tailed, cross-correlated null, the naive i.i.d. rule cannot control its false-positive rate—it rejects a true null half the time at the 5% level—and

Table 9 Monte-Carlo size study: empirical rejection rate of a *true null* of no differential event effect, by inference procedure. $N = 923$ usable simulated panels; one-sided test (H1: infrastructure > regulatory); nominal sizes 0.05 and 0.10. The naive rule (rung 1 of the ladder) over-rejects by an order of magnitude; only the Student- t -copula bootstrap (rung 6, the inference of record) controls size.

Inference procedure	Size @ 0.05	Size @ 0.10
(i) Naive i.i.d. t -test	0.512 ± 0.016	0.701 ± 0.015
(ii) Design-effect, $t(N-1=5)$ critical value <i>same rule, correct $t(df_{\text{eff}} \approx 2.3)$ crit.</i>	0.434 ± 0.016 <i>0.220</i>	0.655 ± 0.016 <i>0.535</i>
(iii) Gaussian-copula bootstrap	0.308 ± 0.015	0.398 ± 0.016
(iv) Student-t-copula bootstrap	0.039 ± 0.006	0.095 ± 0.010
<i>Nominal (target)</i>	0.05	0.10

$\pm$ is the Monte-Carlo binomial standard error. Procedures (i)–(iv) correspond to rungs (1), (4), (5), (6) of Table 7. The design-effect rule (ii) is reported under two reference distributions: the $t(N-1=5)$ critical value the size study originally used (0.434 / 0.655, over-sized because it applies the design-effect point penalty but ignores its consequence for the degrees of freedom) and the correct $t(df_{\text{eff}} \approx 2.3)$ critical value the design effect implies (0.220 / 0.535). The corrected reference improves the size sharply but does not by itself reach nominal: a closed-form effective-df adjustment cannot reproduce the GARCH-plus-heavy-tail structure that only the copula bootstrap simulates. Panel drop rate 7.7%; reference-draw drop rates 5.8% (Gaussian) / 7.3% (t -copula), all under the 10% reliability threshold.

the corrections walk it back toward nominal in exactly the order the lessons predict, with two distinct failure modes diagnosed. The design-effect rule improves sharply once its reference distribution is corrected for the effective degrees of freedom (from 43.4% under the $t(5)$ tail to 22.0% under the correct $t(df_{\text{eff}})$), but the analytic correction alone does not reach nominal size; the Gaussian-copula bootstrap fails for a different reason, its tails (it gets the dependence right but draws too-thin innovations and over-rejects at 31%). The Student- t -copula bootstrap is the only rung that gets both the dependence and the tails right and controls size, which is the formal justification for adopting it as the inference of record. The headline $p = 0.0008$ was, with high probability, a draw from the 51% of true nulls that the naive rule falsely rejects.

7 What Survives: Scope Condition, Robustness, and the Weekly Sentiment Lead

The dual null is the headline, but several results survive the re-examination and bound or qualify it. We report the verified battery here; analyses that did not reproduce under audit have been removed.

7.1 Scope condition: curated identification versus mechanical screening

The directional second-moment point estimate ($\approx 5\times$, not significant under the inference of record) concentrates in how events are identified, not in a mechanical impact threshold. We treat the event-inclusion rule as an explicit researcher degree of freedom—a specification-curve analysis [40], equivalently a multiverse analysis [41],

applied to the inclusion screen—and trace the multiplier across it. This converts the standard selection-on-the-dependent-variable objection into a measured object.

7.1.1 Drop-out census

We reconstruct a systematic candidate pool of 135 events (82 infrastructure-categorised, 53 regulatory-categorised) from public-record sources: Rekt/DeFiLlama leaderboards, SEC and CFTC enforcement actions, major-jurisdiction regulatory announcements, and exchange post-mortem pages. Applying the Stage-2 impact filter (≥ 2 assets with $|r| > 1$ sample-SD in $[-1, +1]$) yields a pass rate of 64.6% for infrastructure candidates (53/82) and 75.5% for regulatory candidates (40/53). Contrary to the selection-bias hypothesis, the filter is *less* demanding for regulatory candidates; a two-proportion test gives $z = -1.33$ (not significant, and opposite in sign to the concern). The mechanical screen, applied identically to both legs, does not favour infrastructure. We are precise about scope: this rules out the specific mechanism (the impact threshold admitting infrastructure more readily); it does not adjudicate whether the upstream manual curation was category-neutral, a separate question the census cannot settle.

7.1.2 Relaxed-threshold sweep

Table 10 re-estimates the full GJR-GARCH-X under five event definitions: the primary curated sample, then a nested mechanical sweep (135 $\rightarrow$ 115 $\rightarrow$ 93 $\rightarrow$ 77) at increasing strictness. The curated sample gives $4.88\times$ (tracking the $5.7\times$ canonical figure; the difference is a winsorisation routine, see note). On the reconstructed pool the multiplier is modest and insignificant at every threshold: $0.58\times$ (no filter), $1.49\times$ (1-asset), $1.61\times$ (2-asset—the like-for-like analogue of the curated criterion), and $1.31\times$ (3-asset). The two-asset screen at $1.61\times$ ($p = 0.16$) is the cleanest comparison, and it does not approach the curated $\approx 5\times$.

Table 10 Infrastructure-regulatory variance multiplier: curated sample versus a nested mechanical impact-screen sweep on the reconstructed 135-candidate pool. Only the curated sample shows a non-negligible point estimate (and even that is not significant under the inference of record, Table 6); the mechanical screens are all small and non-significant.

Spec	n_{infra}	n_{reg}	$\bar{\delta}_{\text{infra}}$	$\bar{\delta}_{\text{reg}}$	Multiplier	Welch p
Primary curated ($n = 50$)	26	24	1.978	0.405	$4.88\times$	0.0015
No filter (135)	82	53	0.209	0.364	$0.58\times$	0.446
≥ 1 asset (115)	69	46	1.403	0.944	$1.49\times$	0.321
≥ 2 assets (93)	53	40	2.889	1.793	$1.61\times$	0.164
≥ 3 assets (77)	42	35	3.834	2.924	$1.31\times$	0.403

The Welch p -values here treat the per-asset coefficients as independent and so are *not* the inference of record (they replicate the pseudoreplication of Lesson 1); they are reported for comparability with the sweep. The curated-row significance evaporates under the copula bootstrap (Table 6). The curated $4.88\times$ uses a global-quantile winsorisation; the canonical rolling-bound winsorisation gives $5.24\times$ curated versus $1.63\times$ at the two-asset screen—the same $\approx 3\times$ gap, so the curated-vs-mechanical gap is an identification effect, not a winsorisation artefact.

Holding membership fixed and re-running every specification under one consistent winsorisation rule reproduces the same $\approx 3\times$ curated-vs-mechanical gap (curated $5.24\times$, two-asset $1.63\times$ under rolling winsorisation), so the gap is a property of event identification, not of the winsorisation choice. The reading is a scope condition: the strong differential—such as it is, given that even the curated estimate is not significant under correct inference—is concentrated in carefully-identified, high-salience events and washes out under mechanical screening of a broad pool. The natural mechanism is that the curated protocol selects severe, cleanly-attributable events where the mechanical-disruption channel dominates, whereas a broad mechanical filter admits many marginal events whose responses are noise. We state the competing reading explicitly—that the curation *is* the selection bias and the mechanical $1.61\times$ is the unbiased estimate—and let the reader weigh it; the drop-out census, which shows the screen does not favour infrastructure, supports the former.

7.2 Control battery for the second-moment point estimate

The directional point estimate ($\approx 5\times$) is not significant under the inference of record, but it is worth establishing that the *point estimate itself* is not an artefact of an obvious confound—a crisis coincidence, anticipation leakage, a winsorisation choice, or a misspecified mean. We ran a battery of controls; each leaves the direction intact, and the one control that meaningfully moves the magnitude (break-regime baseline variance) is reported with its own cross-asset-robust significance. Throughout, these checks concern the point estimate; the significance verdict is always the copula bootstrap’s, and it is null.

7.2.1 Break-regime controls

Infrastructure events cluster in the 2022–2023 crisis window, so the differential could in principle be crisis-period baseline variance that infrastructure events happen to coincide with. We add structural-break regime dummies (from the PELT change-point detection of Section 7.3.3) to the variance equation and refit. Two variants: a single FTX-crisis-regime dummy (the 2021-break-to-2022-break high-variance segment), and the full set of per-segment regime dummies. Table 11 reports both, with the cross-asset-robust copula-bootstrap p as the inference of record (the naive Welch p is shown only to make the pseudoreplication gap visible once more).

The reading is that a meaningful chunk of the headline magnitude is crisis-regime baseline variance: the full-regime control roughly halves the multiplier (to $2.64\times$) by raising both coefficients. But the single FTX-crisis dummy—the cleanest control, with only 6% degenerate refits—barely moves either the point estimate ($4.88 \rightarrow 3.97\times$) or the robust p ($0.323 \rightarrow 0.318$): if the differential were *merely* crisis variance, the crisis dummy would absorb it, and it does not. The honest summary is that the controlled multiplier is $\approx 2.6\text{--}4\times$, directional, and not significant under correct inference at any of the three specifications—consistent with the baseline never having cleared significance either.

Table 11 Break-regime controls on the curated sample. Regime dummies are added to the variance equation. The multiplier attenuates but survives in direction; the cross-asset-robust copula-bootstrap p is the inference of record. The naive Welch p (which treats the six per-asset coefficients as independent) is shown only to illustrate, again, the pseudoreplication gap.

Specification	$\bar{\delta}_{\text{infra}}$	$\bar{\delta}_{\text{reg}}$	Mult.	Naive Welch p	Robust p
Baseline (no controls)	1.978	0.405	$4.88\times$	0.0015	0.323
+ Crisis-regime dummy	1.914	0.483	$3.97\times$	0.0033	0.318
+ Full regime dummies [†]	2.443	0.926	$2.64\times$	0.0115	0.249

Robust p is the Student- t -copula CCC-GARCH-X bootstrap (one-sided). [†]Full-regime refits are numerically less reliable ($\approx 23\%$ degenerate, above the 10% reliability bar); its p reads as “ ≥ 0.10 , not significant.” The full-regime control raises *both* coefficients (regime dummies absorb the variance level, and $\bar{\delta}_{\text{reg}}$ rises proportionally more), which is what halves the multiplier; it also inverts XRP alone ($1.83\times \rightarrow 0.80\times$, the one asset where $\bar{\delta}_{\text{reg}} > \bar{\delta}_{\text{infra}}$ under full controls). The single crisis dummy barely moves the point estimate ($4.88 \rightarrow 3.97\times$) or the robust p ($0.323 \rightarrow 0.318$).

7.2.2 Anticipation runs against the asymmetry

If regulatory events were anticipated—leaked through political cycles or prior supervisory signalling—then volatility would build before the announcement date, and a symmetric $[-3, +3]$ window would understate the regulatory response by excluding the leaked pre-event variance. Lengthening the regulatory pre-window should then *absorb* that leaked volatility into $\bar{\delta}_{\text{reg}}$, raising it and shrinking the multiplier toward one. Table 12 runs that test: the regulatory pre-window is lengthened from 3 to 10 days while infrastructure stays fixed at $[-3, +3]$.

Table 12 Anticipation sweep: the regulatory pre-window is progressively lengthened (infrastructure held at $[-3, +3]$). If regulatory events were anticipated, $\bar{\delta}_{\text{reg}}$ would *rise* and the multiplier shrink toward one. The opposite occurs.

Reg. pre-window	$\bar{\delta}_{\text{infra}}$	$\bar{\delta}_{\text{reg}}$	Multiplier	Naive Welch p
3 days (published, symmetric)	1.978	0.405	$4.88\times$	0.0015
5 days	1.962	0.272	$7.21\times$	0.0012
7 days	1.976	0.279	$7.09\times$	0.0012
10 days	1.966	0.224	$8.77\times$	0.0012

The infrastructure coefficient is invariant (≈ 1.97) throughout; the regulatory coefficient *falls* as the window lengthens (the extra pre-event days are quiet, diluting the dummy), so the multiplier *grows*. Naive Welch p shown for comparability only; not the inference of record.

The result runs decisively against the anticipation hypothesis. The infrastructure coefficient is invariant (≈ 1.97) across all windows. The regulatory coefficient *falls* ($0.405 \rightarrow 0.224$) as the window lengthens, because the extra pre-event days are quiet rather than carrying leaked volatility—widening the window dilutes the dummy rather than capturing anticipation. The point-estimate multiplier therefore *grows* ($4.88\times \rightarrow 8.77\times$). Anticipation, were it present, would shrink the asymmetry; instead the data show no pre-event regulatory volatility to absorb. (We note the data limitation: the

event census carries no per-event anticipated/surprise flag, so a surprise-only subset would require hand-coding each event and is out of scope; the window sweep is the feasible test.)

7.3 Specification stability

7.3.1 The constant mean is justified

Ljung–Box tests on the winsorised returns flag serial correlation in 8 of 18 (asset $\times$ lag $\in \{5, 10, 20\}$) cells, concentrated in ETH (all three lags) and BNB (all three lags), marginal in XRP and ADA, and absent in BTC and LTC. Because some serial correlation exists, we run the AR(1)-in-mean robustness refit rather than assert the constant mean. The fitted AR(1) coefficient is tiny and negative everywhere ($\phi \in [-0.067, -0.025]$)—bid-ask-bounce-scale dependence, economically negligible—and filtering the mean leaves the event coefficients essentially unchanged: the largest per-asset move is $\approx 3\%$, and the cross-asset multiplier is $5.04\times$ under AR(1)-in-mean versus $4.88\times$ under the constant mean. The constant-mean specification is not displacing mean dynamics into the variance equation in any material way.

7.3.2 Winsorisation invariance

The curated-versus-mechanical gap is the same under global-quantile winsorisation ($4.88\times$ curated, $1.61\times$ at the two-asset screen) and under the canonical rolling-bound (30-day, $\pm 5\sigma$) winsorisation ($5.24\times$ curated, $1.63\times$ at the two-asset screen)—a $\approx 3\times$ gap either way (Table 10 note). Membership held fixed, swapping the winsorisation rule moves the curated multiplier only from $4.88\times$ to $5.24\times$ (closer to the $5.7\times$ canonical figure) and all four mechanical screens stay at $\approx 0.6\text{--}1.6\times$ and non-significant under either rule. The curated-versus-mechanical gap is an identification effect, not a winsorisation artefact.

7.3.3 Persistence is descriptive only

Bai-Perron-style structural-break tests [35–37] detect 2–4 breaks per asset, clustering around late 2020, late 2021, and most consistently the November 2022 FTX collapse (the only episode producing aligned breaks in four of six assets). Within-sub-sample GARCH(1,1) persistence averages $\alpha + \beta \approx 0.905$, below the full-sample 0.985 (Table 13). We deliberately do not over-read this. When the sub-segment persistence parameters are estimated with their own standard errors, *none* of the 27 sub-segments falls more than two standard errors below its full-sample value: the regimes are short, their persistence SEs run from ± 0.06 to ± 0.31 , and the apparent drop lies entirely within estimation noise. Nor is the drop a small-sample artefact—mean sub-segment persistence is 0.891 for segments with $n \geq 500$ and 0.887 for $n < 500$, essentially identical. The near-integration framing is therefore qualified but not overturned: the within-regime persistence drop is directional and suggestive, not a sharply identified structural fact, and we present it as such. The break regimes themselves are used substantively only as the controls of Section 7.2.1.

Table 13 Within-sub-sample GARCH(1,1) persistence by detected regime. The sub-sample drop is directional; no sub-segment falls more than two standard errors below its full-sample value, so the drop is within estimation noise rather than sharply identified.

Asset	Full-sample $\alpha + \beta$	Sub-sample mean $\alpha + \beta$	Number of sub-samples
BTC	0.999	0.979	3
ETH	0.999	0.864	5
XRP	0.985	0.906	5
BNB	0.994	0.920	5
LTC	0.976	0.910	5
ADA	0.954	0.872	4
Mean	0.985	0.905	—

7.4 Sentiment leads volatility at the weekly frequency

The paper’s one surviving positive result concerns the sentiment-volatility lead-lag relationship—and both its recovery and its limits are inference cautionary tales. GDELT sentiment is a *weekly* series. An earlier draft ran Granger causality [42] at daily frequency by forward-filling the weekly sentiment to daily, and found no sentiment-to-volatility lead for any of the 18 asset-sentiment pairs. That null was an artefact: forward-filling repeats each weekly value across six intervening days, turning the predictor into a step function constant within every week, which mechanically cannot carry daily lead-lag information.

Run at the native weekly frequency—aggregating daily returns to the GDELT weekly grid (volatility proxy $\sum |r_t|$) and aligning to the native weekly sentiment with no forward-fill, both directions, weekly lags 1–8—the conclusion reverses. Sentiment Granger-*leads* weekly volatility for **10 of 18** asset-sentiment pairs at the 5% level (9 of these with no reverse volatility-to-sentiment lead, i.e. a clean one-directional lead). The strongest leads concentrate in the regulatory and aggregate sentiment channels for XRP, ADA, and BNB (Table 14). Applying a Benjamini–Hochberg correction [34] to the 18-pair sentiment-to-volatility family, **7 pairs survive at $q < 0.05$ and 12 at $q < 0.10$** ; for contrast, the daily forward-filled family yields 0 raw-significant pairs and 0 survivors at any q (and the daily both-directions family of 36 tests yields 4 raw-significant, only 1 surviving FDR). The weekly lead is robust to multiplicity; the daily “null” was the forward-fill suppressing the signal, not the absence of one.

7.4.1 Stress-testing the lead: five corrections

The 7 FDR survivors (XRP×3, ADA×3, BNB-regulatory) are all SEC-litigation assets and all high-Bitcoin-co-movement, which raises three statistical objections (spurious regression from non-stationarity, Bitcoin-co-movement masquerading as an asset-specific lead, and zero-imputation of the warm-up weeks) and one economic one (the lead being a litigation-window phenomenon rather than a general anticipatory channel). We address each.

Stationarity. Augmented Dickey–Fuller [43] and KPSS [44] tests place the three sentiment series cleanly at $I(0)$ (ADF $p < 10^{-13}$, KPSS $p = 0.10$), but the weekly volatility proxies are persistent and near-integrated—ADF rejects a unit root while KPSS rejects stationarity for five of six assets. Mixed and near-unit-root regressors invalidate the standard Granger F -test’s asymptotics, which is precisely why the lead must be re-tested by a method robust to integration order.

Toda–Yamamoto. We re-run every pair as a lag-augmented bivariate VAR [45]: select the lag order p by AIC, fit $p + 1$ lags, and apply an HC1-robust Wald test to the first p sentiment lags only (the extra lag restores $\chi^2(p)$ asymptotics under a unit root). All 7 FDR survivors remain significant (Toda–Yamamoto $p \approx 0.0001$ – 0.012), while marginal naive pairs wash out (LTC-regulatory $0.042 \rightarrow 0.122$). The lead is not a spurious-regression artefact.

Bitcoin co-movement. Re-running each sentiment-to-volatility test with lagged co-movement volatility as an exogenous control (lagged Bitcoin $\sum |r|$ for altcoins, lagged market-ex-asset volatility for BTC/ETH) leaves all 7 survivors essentially unchanged to the fourth decimal (XRP-aggregate $0.0009 \rightarrow 0.0010$; ADA-regulatory $0.0005 \rightarrow 0.0006$). The lead is asset-specific, not a Bitcoin-co-movement shadow—a test we expected might eliminate it and which it survived.

Missingness. The native weekly series zero-imputes the first ≈ 24 warm-up weeks (the pre-2019 cutoff plus the 52-week z -score window). A logit of the imputed-week indicator on lagged volatility is non-significant for every asset (all $p \geq 0.10$), so zero-imputation is not systematically tied to volatility and does not manufacture the lead.

Litigation sub-period—the decisive test. The survivors are all SEC-litigation assets, so we split each on its litigation onset (SEC v. Ripple, 2020-12-22; SEC v. Binance naming Binance Coin and Cardano, 2023-06-05) and re-test on the pre-litigation window. **XRP’s entire lead is a litigation-window phenomenon:** all three XRP pairs are non-significant before the suit ($p = 0.15$ – 0.22 , $n = 79$, with no Granger structure in *either* direction) but significant during it ($p = 0.005$ – 0.013). The lead switches on when the SEC sues. Cardano (regulatory, aggregate) and Binance Coin (regulatory) retain $p < 0.05$ in their pre-onset windows, but their June-2023 onsets leave the pre-period at ≈ 60 – 63% of the sample, so their split is a weaker test. Of the 7 survivors, **3 retain a pre-litigation lead**; XRP’s 3 do not.

7.4.2 What the lead is, and is not

Putting the five corrections together: the weekly sentiment lead survives every *statistical* objection (7/7 of the FDR survivors are robust to Toda–Yamamoto lag-augmentation and to Bitcoin-co-movement control, and the zero-imputation does not drive it) but fails the *economic* one (only 3/7 retain the lead outside litigation windows, and XRP—the cleanest single result—vanishes entirely outside the SEC v. Ripple period). The honest characterisation is therefore not a general anticipatory channel but *a weekly lead-lag association between GDELT sentiment and realised volatility, concentrated in SEC-litigation assets and substantially attributable to litigation windows (most starkly for XRP), robust to Toda–Yamamoto and to Bitcoin co-movement, but requiring further validation before it can be read as a general anticipatory mechanism.* We keep it for its substantive content—a degraded, weekly, 7%-missing series

nonetheless carries a litigation-period lead that survives non-trivial robustness—and, equally, for its methodological content: the daily forward-fill that originally hid it is a third instance of the paper’s theme, an inference choice (here the wrong frequency) inverting the substantive answer, just as pseudoreplication and the Gaussian bootstrap inverted it at the second moment. The direction of the artefact differs—here a false negative from forward-filling, there a false positive from mis-specified inference—but the moral is the same: run at the wrong level, or with the wrong distribution or frequency, the inference can invert the answer.

Table 14 Weekly Granger sentiment-to-volatility leads (p -values, weekly lags 1–8), by asset and sentiment channel. Run at the native weekly frequency with no forward-fill. Bold = significant at 5%. The daily forward-filled version of these tests produced no significant leads at all.

Asset	Regulatory sent.	Infrastructure sent.	Aggregate (GDELT) sent.
XRP	0.001	0.010	0.001
ADA	0.001	0.005	0.001
BNB	0.006	0.049	0.023
LTC	0.042	—	—
ETH	0.055	—	—
BTC	—	—	—

“—” = not significant at 5%. ETH-regulatory (0.055) is borderline. Family of 18 sentiment-to-volatility tests: 10 raw-significant at 5%; 7 survive Benjamini–Hochberg at $q < 0.05$, 12 at $q < 0.10$.

Beyond the litigation confound, the natural caveat is the frequency mismatch itself: the lead is established at the weekly horizon and cannot be sharpened to a daily lead without a daily-frequency GDELT extraction, which we flag as the natural next step. The pre-litigation windows are also short for XRP ($n = 79$), so its null—though corroborated by the absence of Granger structure in both directions—would be more decisive on a longer pre-period.

8 Discussion

8.1 What the dual null means

On this sample, under correct inference, cryptocurrency markets do not differentiate infrastructure from regulatory shocks at either moment of the return distribution. At the first moment the difference is +7.19 pp ($p = 0.283$); at the second the variance multiplier is $\approx 5\times$ ($p = 0.323$). Both are directional, neither significant. The economically interesting reading is not “the market sees them as the same”—we cannot establish that, and at the first moment we are plainly underpowered—but rather that the differential processing of shock types, if it exists, is below the resolution this data affords once its dependence and heavy tails are respected. The earlier claim that “the structure lives in the second moment” was an inference artefact: the second moment is no more distinguishable from zero than the first.

This matters for how the field reads cross-asset crypto event studies. The pattern that generated the original story—a clean return-level null beside an apparently sharp variance asymmetry—is precisely the pattern that pseudoreplication and heavy-tailed mis-specification produce. A return-level test is naturally conservative (few events, wide intervals), while a per-asset variance test that treats correlated assets as independent is naturally over-confident. The contrast between them can look like substantive structure when it is in fact a contrast between a well-specified and a mis-specified inference. We suspect this pattern is not unique to our data.

8.2 The directional point estimates

We do not claim the effects are zero. The first-moment difference (infrastructure CARs less negative than regulatory) and the second-moment multiplier (infrastructure variance larger) both point the same way across most specifications, and the second-moment direction is uniform across all six assets. A mechanical-disruption account—infrastructure events causing immediate liquidity impairment and sharp variance spikes, regulatory events operating through slower information channels—would predict exactly this direction. The honest position is that the direction is suggestive and the magnitude unresolved: a substantially larger curated event database, or a daily-frequency design, would be needed to determine whether the direction reflects a genuine difference or sampling noise. The scope condition (Section 7.1) further locates whatever effect exists in carefully-identified, high-salience events rather than in a population-level constant.

8.3 The sentiment channel

The surviving positive result is the weekly sentiment lead, but its scope is narrower than it first appears. That degraded, weekly, 7%-missing GDELT sentiment Granger-leads volatility for 10 of 18 pairs (7 surviving FDR) is real and robust to Toda–Yamamoto lag-augmentation and to Bitcoin-co-movement control (Section 7.4.1). But the lead is concentrated in SEC-litigation assets and is substantially litigation-driven: only 3 of the 7 FDR survivors retain a lead in their pre-litigation windows, and XRP’s lead—the cleanest single result—is entirely a SEC v. Ripple-window phenomenon, absent before the suit. The honest reading is therefore a weekly lead-lag association concentrated in litigation episodes, not yet a demonstrated general anticipatory channel. Sentiment intensity has been linked to liquidity withdrawal and adverse selection in cryptocurrency markets [46], a mechanism consistent with sentiment leading volatility around high-salience regulatory episodes; whether the lead generalises beyond litigation windows requires a longer pre-period and a daily-frequency extraction, the natural extensions.

8.4 Limitations

Statistical power. The first moment is underpowered ($\text{MDE} \approx 19\text{--}20$ pp against an observed ≈ 7 pp); the dual null is a failure to reject, not a demonstration of

equality. The second moment has a non-trivial point estimate that the dependence-robust bootstrap cannot distinguish from zero. Both caveats are stated symmetrically throughout.

Event classification. The infrastructure/regulatory taxonomy involves judgment calls; some events have elements of both. Alternative classifications could yield different point estimates.

Scope: systemic versus idiosyncratic. The Stage-2 criterion admits only events with cross-sectional impact (at least two assets), so the comparison is established for systemic, market-wide shocks; idiosyncratic single-asset events are out of scope.

Identification quality differs across legs. Infrastructure events—unannounced outages, externally-discovered exploits—are more plausibly exogenous to contemporaneous market conditions than regulatory announcements, which respond to supervisory cycles and can leak. The two legs differ in identification quality as well as in point estimate, which bears on interpretation of the (insignificant) difference.

Sentiment lead: frequency and litigation confound. The lead is established weekly and cannot be resolved to a daily horizon without a daily GDELT extraction. It is also concentrated in SEC-litigation assets and substantially litigation-driven—XRP’s lead is absent outside the SEC v. Ripple window, and only 3 of 7 FDR survivors retain the lead pre-litigation—so it is a litigation-concentrated association, not yet a demonstrated general anticipatory channel. The XRP pre-litigation window is short ($n = 79$), which limits the power of that decisive test.

Six assets. The panel may not represent the broader ecosystem; smaller-cap and DeFi-native assets may behave differently.

9 Conclusion

We asked whether cryptocurrency markets differentiate infrastructure from regulatory shocks, and tested the question at both moments of the return distribution on one shared sample under one dependence-robust design. The answer, under correct inference, is that we cannot reject equality at either moment: the first-moment CAR difference is +7.19 pp ($p = 0.283$, Ibragimov–Müller $p = 0.295$) and the second-moment variance multiplier is directional at $\approx 2.6\text{--}5\times$ but not significant (t -copula bootstrap $p = 0.323$). Both are underpowered or dependence-limited failures to reject, not claims of identity.

The lead contribution is the inference cautionary tale that produced this conclusion. An earlier version of this work reported the variance asymmetry as a significant fivefold effect; the present paper overturns that estimate by its author, tracing the apparent significance to four specific inferential defects—pseudoreplication across cross-correlated assets, the concentrating (not diluting) nature of that dependence, a Gaussian bootstrap draw where the fitted innovations are Student- t at $\nu \approx 3$, and the block-versus-parametric choice at the first moment. Each correction is specific and reproducible; together they convert a headline ($p \approx 0.001$) into a null ($p = 0.323$) with the point estimate unchanged. We report this as methodological maturity: we followed the correct inference and it dissolved our own result, and the demonstration

transfers directly to other cross-asset event studies in heavy-tailed markets, which are exceptionally prone to manufacturing significance in just this way.

Two findings survive and bound the result. The variance differential, such as it is, is a property of curated, high-salience event identification ($\approx 5\times$) and not of a mechanical impact screen ($1.3\text{--}1.6\times$, never significant)—a scope condition we measure rather than assert, with a drop-out census confirming the screen does not favour infrastructure. And at the data’s native weekly frequency, GDELT sentiment Granger-leads volatility for 10 of 18 asset-sentiment pairs (7 surviving FDR at $q < 0.05$), a relationship a daily forward-fill had spuriously hidden; on stress-testing, the lead is robust to Toda–Yamamoto lag-augmentation and Bitcoin co-movement but concentrated in SEC-litigation assets and substantially litigation-driven (3 of 7 survive outside litigation windows, XRP’s lead entirely a SEC v. Ripple-window phenomenon), so it stands as a litigation-concentrated lead-lag association requiring further validation rather than a general anticipatory channel. Future work fielding a larger curated database, daily-frequency sentiment, and a longer panel will be positioned to determine whether the directional point estimates we report reflect a genuine, if small, difference, or the sampling noise that this sample cannot rule out, and whether the sentiment lead generalises beyond litigation episodes.

Supplementary information. Complete replication materials—the data, the full analysis pipeline, and figure-generation scripts—are available at <https://github.com/studiofarzulla/crypto-event-study> and archived on Zenodo (DOI: 10.5281/zenodo.18099608); the scripts implementing the event-level block bootstrap, the Student-*t*-copula CCC-GARCH-X bootstrap, the pseudoreplication re-tests, the selection-bias sweep, the Monte-Carlo size study, and the weekly Granger analysis are included.

Acknowledgements. The author acknowledges King’s College London for resources and library access during the initial stages of this research. Computational analysis was carried out on independent hardware; this work forms part of a wider research programme under the Adversarial Systems and Complexity Research Initiative (ASCRI; systems.ac). AI assistants (Claude, Anthropic) were used for code development, literature-review synthesis, and manuscript preparation under the author’s direction and review.

Declarations

- **Funding:** This research was supported by resources provided by King’s College London. No external grants or funding were received.
- **Conflict of interest:** The author declares no conflict of interest.
- **Ethics approval:** Not applicable.
- **Consent to participate:** Not applicable.
- **Consent for publication:** Not applicable.
- **Data availability:** All data are publicly available from the CoinGecko API and the GDELT Project.
- **Code availability:** The complete verified pipeline (scripts `c1–c14`, estimators, and outputs) is available at <https://github.com/studiofarzulla/crypto-event-study> and archived on Zenodo (DOI: 10.5281/zenodo.18099608).

- **Author contribution:** Sole author.

Appendix A Event Classification Summary

Table A1 provides selected examples from the 50-event shared sample. The complete event list with timestamps, sources, and classification rationale is available in the replication materials.

Table A1 Selected Events by Category

Date	Type	Event
<i>Infrastructure Events (n=26)</i>		
2022-11-11	Infra	FTX exchange collapse
2022-05-09	Infra	Terra/Luna depegging
2023-03-10	Infra	Silvergate/SVB banking crisis
2022-06-13	Infra	Celsius withdrawal halt
2020-03-12	Infra	COVID “Black Thursday”
<i>Regulatory Events (n=24)</i>		
2023-06-05	Reg	SEC sues Binance
2023-06-06	Reg	SEC sues Coinbase
2020-12-23	Reg	SEC sues Ripple
2021-05-18	Reg	China bans crypto mining
2024-01-10	Reg	Bitcoin ETF approval

Appendix B GJR-GARCH-X Technical Specification

The GJR-GARCH-X model extends the GJR-GARCH(1,1) of Glosten et al. [30] with exogenous variance regressors.

Mean equation:

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad z_t \sim t_\nu. \quad (\text{B1})$$

Variance equation:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 \mathbb{I}(\varepsilon_{t-1} < 0) + \beta \sigma_{t-1}^2 + \mathbf{X}_t' \boldsymbol{\delta}, \quad (\text{B2})$$

where $\mathbf{X}_t = [D_t^{\text{infra}}, D_t^{\text{reg}}, S_t^{\text{infra}}, S_t^{\text{reg}}]'$.

Estimation: quasi-maximum likelihood with Student- t innovations; log-likelihood

$$\ell(\theta) = \sum_{t=1}^T \left[\log \Gamma\left(\frac{\nu+1}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2} \log((\nu-2)\pi\sigma_t^2) - \frac{\nu+1}{2} \log\left(1 + \frac{\varepsilon_t^2}{(\nu-2)\sigma_t^2}\right) \right]. \quad (\text{B3})$$

Constraints: covariance stationarity via $\alpha + \beta + \gamma/2 < 1$; non-negativity $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \gamma \geq 0$.

Units: returns in percentages, so σ_t^2 and the event coefficients δ_j are in squared percentage points.

Appendix C Dependence-Robust Bootstrap Detail

First moment (event-level block bootstrap). For each event, per-asset CARs are averaged into one event-level CAR. Bootstrap replicates resample event IDs with replacement (all assets within an event kept together, so within-event cross-sectional correlation is preserved), and the difference of group means is recomputed; $B = 5,000$, percentile CIs, seed fixed. The Ibragimov–Müller test is a Welch t -test on the event-level mean CARs, treating each event as one cluster.

Second moment (Student- t -copula CCC-GARCH-X bootstrap). The procedure, for each of $B = 2000$ replicates, is:

1. Draw a latent Gaussian vector $\mathbf{w} \sim \text{MVN}(\mathbf{0}, R_z)$ at the standardised-residual correlation matrix R_z (the cross-asset dependence; $\bar{\rho}_z = 0.705$).
2. Draw a single shared mixing variable $s \sim \chi_{\nu_c}^2 / \nu_c$ with ν_c the median fitted degrees of freedom, and form $\mathbf{t} = \mathbf{w} / \sqrt{s}$. This is a draw from a multivariate- t copula—the shared s induces *joint* tail dependence (assets taking extreme values together), which a Gaussian copula would not.
3. Map each component to a uniform via its t_{ν_c} CDF, then to the per-asset Student- t margin at that asset’s *fitted* ν_i via the inverse CDF, and rescale by $\sqrt{(\nu_i - 2)/\nu_i}$ so the innovation has unit variance (the variance recursion σ_t^2 carries the scale).
4. Simulate each asset’s return path through its fitted GJR-GARCH-X variance recursion driven by these innovations, with the event coefficient set to a single common value (the null of no differential), and refit the joint model unrestricted, re-estimating $\bar{\delta}_{\text{infra}} / \bar{\delta}_{\text{reg}}$.

The bootstrap p is the fraction of null replicates whose multiplier statistic $\bar{d} = \bar{\delta}_{\text{infra}} - \bar{\delta}_{\text{reg}}$ equals or exceeds the observed $\bar{d}$. Degenerate refits ($|\delta| > 50$, SLSQP boundary optima) are dropped and counted; drop rates are $\approx 7\%$ at baseline and crisis, $\approx 23\%$ for the full-regime spec (which is therefore reliability-flagged). Non-common-window positions receive independent unit-variance Student- t draws at the fitted ν_i . The Gaussian-copula comparator (rung 5 of the ladder) is identical except that step 3 draws Gaussian innovations; the gap between the two is the entire content of Lesson 3.

Appendix D Monte-Carlo Size-Study Design

The size study (Section 6.8) simulates panels under a fitted true null and measures each inference procedure’s rejection rate.

True-null DGP. Per-asset GJR-GARCH-X dynamics are estimated under a *null-imposed* combined event dummy $D_{\text{event}} = D_{\text{infra}} + D_{\text{reg}}$, so $\delta_{\text{infra}} = \delta_{\text{reg}}$ by construction

and there is genuinely no differential event effect in the truth. Returns are then re-simulated from these fitted dynamics, with cross-asset innovations drawn from the same multivariate- t copula as the inference-of-record bootstrap: latent $\text{MVN}(\mathbf{0}, R_z)$ at $\bar{\rho}_z = 0.705$, shared chi-square mixing at $\nu_c = 3.82$ (the median fitted ν), per-asset Student- t margins at the fitted $\nu_i = (3.18, 3.70, 3.14, 4.20, 3.95, 4.56)$, unit-variance-rescaled. The real 26-infrastructure / 24-regulatory event-label structure and the real event-window dummies are reused unchanged for every panel; only the returns are re-simulated.

Procedure. $N = 1,000$ panels are drawn (923 usable after a convergence screen: a panel is dropped if a per-asset refit fails to converge, hits a degenerate $|\delta| > 50$ optimum, 7.7% dropped). Each bootstrap rung’s one-sided critical value is calibrated once from $B_{\text{ref}} = 4,000$ reference draws under the same DGP (the critical value is a property of the DGP under the null, so a redundant nested refit per panel is unnecessary; this is exact for size). For each panel we record whether each of the four procedures rejects at the 5% and 10% levels, and report the empirical rejection rate with its binomial Monte-Carlo standard error. The one-sided critical values are: Gaussian copula $\bar{d} = 1.615$ (5%) / 1.454 (10%); t -copula $\bar{d} = 2.445$ (5%) / 2.166 (10%).

Audit-only standard errors. Per-asset numerical-Hessian standard errors are computed for reporting (positive-definite for 98.7% of accepted panels) but are never on the decision path: neither the naive Welch rule (on the six coefficients) nor the design-effect rule (on the dispersion of the six paired differences) uses the model SE, so an SE failure never drops a panel or flips a decision. Panel acceptance therefore matches the reference-draw rule exactly, which is what makes the t -copula rung’s size nominal by construction—the self-validation in Section 6.8.

References

- [1] Chen, Y.-L., Chang, Y.-C., Yang, J.J.: Cryptocurrency hacking incidents and the price dynamics of bitcoin spot and futures. *Finance Research Letters* **55**, 103456 (2023)
- [2] Grobys, K.: When the blockchain does not block: On hackings and uncertainty in the cryptocurrency market. *Quantitative Finance* **21**(8), 1267–1279 (2021)
- [3] Auer, R., Claessens, S.: Regulating cryptocurrencies: Assessing market reactions. *BIS Quarterly Review*, 51–65 (2018)
- [4] Feinstein, B.D., Werbach, K.: The impact of cryptocurrency regulation on trading markets. *Journal of Financial Regulation* **7**(1), 48–99 (2021)
- [5] Farzulla, M.: Infrastructure vs Regulatory Shocks: Asymmetric Volatility Response in Cryptocurrency Markets. Research Square preprint, DOI 10.21203/rs.3.rs-8323026/v1 (2025)
- [6] Farzulla, M.: Same Returns, Different Risks: How Cryptocurrency Markets Process Infrastructure vs Regulatory Shocks. arXiv:2602.07046 (2025)

- [7] Katsiampa, P.: Volatility estimation for bitcoin: A comparison of GARCH models. *Economics Letters* **158**, 3–6 (2017)
- [8] Chu, J., Chan, S., Nadarajah, S., Osterrieder, J.: GARCH modelling of cryptocurrencies. *Journal of Risk and Financial Management* **10**(4), 17 (2017)
- [9] Cheikh, N.B., Zaied, Y.B., Chevallier, J.: Asymmetric volatility in cryptocurrency markets: New evidence from smooth transition GARCH models. *Finance Research Letters* **35**, 101293 (2020)
- [10] Baur, D.G., Dimpfl, T.: Asymmetric volatility in cryptocurrencies. *Economics Letters* **173**, 148–151 (2018)
- [11] Makarov, I., Schoar, A.: Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics* **135**(2), 293–319 (2020)
- [12] Bouri, E., Molnár, P., Azzi, G., Roubaud, D., Hagfors, L.I.: On the hedge and safe haven properties of bitcoin. *Finance Research Letters* **20**, 192–198 (2017)
- [13] Kim, A., Trimborn, S., Härdle, W.K.: VCRIX – a volatility index for cryptocurrencies. *International Review of Financial Analysis* **78**, 101915 (2021)
- [14] Farzulla, M., Maksakov, A.: ASRI: An Aggregated Systemic Risk Index for Cryptocurrency Markets. *arXiv:2602.03874* (2025)
- [15] Campbell, J.Y., Lo, A.W., MacKinlay, A.C.: *The Econometrics of Financial Markets*. Princeton University Press, Princeton (1997)
- [16] MacKinlay, A.C.: Event studies in economics and finance. *Journal of Economic Literature* **35**(1), 13–39 (1997)
- [17] Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M.C., Siering, M.: Bitcoin – asset or currency? Revealing users’ hidden intentions. In: *Proceedings of ECIS 2014*, Tel Aviv (2014)
- [18] Kolari, J.W., Pynnonen, S.: Event study testing with cross-sectional correlation of abnormal returns. *Review of Financial Studies* **23**(11), 3996–4025 (2010) <https://doi.org/10.1093/rfs/hhq072>
- [19] Boehmer, E., Musumeci, J., Poulsen, A.B.: Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics* **30**(2), 253–272 (1991) [https://doi.org/10.1016/0304-405X\(91\)90032-F](https://doi.org/10.1016/0304-405X(91)90032-F)
- [20] Cameron, A.C., Gelbach, J.B., Miller, D.L.: Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics* **90**(3), 414–427 (2008) <https://doi.org/10.1162/rest.90.3.414>
- [21] Petersen, M.A.: Estimating standard errors in finance panel data sets: Comparing

- approaches. *The Review of Financial Studies* **22**(1), 435–480 (2009) <https://doi.org/10.1093/rfs/hhn053>
- [22] MacKinnon, J.G., Nielsen, M.Ø., Webb, M.D.: Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics* **232**(2), 272–299 (2023) <https://doi.org/10.1016/j.jeconom.2022.04.001>
 - [23] Ibragimov, R., Müller, U.K.: t-statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics* **28**(4), 453–468 (2010) <https://doi.org/10.1198/jbes.2009.08046>
 - [24] Goldsmith-Pinkham, P., Lyu, T.: Causal inference in financial event studies. Working Paper, Yale University (2025) [arXiv:2511.15123](https://arxiv.org/abs/2511.15123)
 - [25] Casini, A., McCloskey, A.: Identification, estimation and inference in high-frequency event study regressions. CEIS Research Paper No. 608, Tor Vergata University (2025)
 - [26] Shen, D., Urquhart, A., Wang, P.: Does twitter predict bitcoin? *Economics Letters* **174**, 118–122 (2019)
 - [27] Phillips, R.C., Gorse, D.: Cryptocurrency price drivers: Wavelet coherence analysis revisited. *PLOS ONE* **13**(4), 0195200 (2018)
 - [28] Liu, Y., Tsyvinski, A.: Risks and returns of cryptocurrency. *Review of Financial Studies* **34**(6), 2689–2727 (2021)
 - [29] Kothari, S.P., Warner, J.B.: Econometrics of event studies. In: *Handbook of Empirical Corporate Finance*, pp. 3–36. Elsevier, ??? (2007). <https://doi.org/10.1016/B978-0-444-53265-7.50015-9>
 - [30] Glosten, L.R., Jagannathan, R., Runkle, D.: On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**(5), 1779–1801 (1993)
 - [31] Han, H., Kristensen, D.: Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates. *Journal of Business & Economic Statistics* **32**(3), 416–429 (2014) <https://doi.org/10.1080/07350015.2014.897954>
 - [32] Nelson, D.B.: Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**(3), 318–334 (1990) <https://doi.org/10.1017/S0266466600005296>
 - [33] Lumsdaine, R.L.: Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* **64**(3), 575–596 (1996) <https://doi.org/10.2307/2171862>
 - [34] Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society:*

Series B **57**(1), 289–300 (1995)

- [35] Killick, R., Fearnhead, P., Eckley, I.A.: Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* **107**(500), 1590–1598 (2012)
- [36] Bai, J., Perron, P.: Estimating and testing linear models with multiple structural changes. *Econometrica* **66**(1), 47–78 (1998)
- [37] Bai, J., Perron, P.: Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* **18**(1), 1–22 (2003)
- [38] Li, W.K., Mak, T.K.: On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis* **15**(6), 627–636 (1994) <https://doi.org/10.1111/j.1467-9892.1994.tb00217.x>
- [39] Hurlbert, S.H.: Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**(2), 187–211 (1984)
- [40] Simonsohn, U., Simmons, J.P., Nelson, L.D.: Specification curve analysis. *Nature Human Behaviour* **4**, 1208–1214 (2020) <https://doi.org/10.1038/s41562-020-0912-z>
- [41] Steegen, S., Tuerlinckx, F., Gelman, A., Vanpaemel, W.: Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science* **11**(5), 702–712 (2016) <https://doi.org/10.1177/1745691616658637>
- [42] Granger, C.W.J.: Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**(3), 424–438 (1969)
- [43] Dickey, D.A., Fuller, W.A.: Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**(366a), 427–431 (1979) <https://doi.org/10.1080/01621459.1979.10482531>
- [44] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y.: Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* **54**(1–3), 159–178 (1992) [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- [45] Toda, H.Y., Yamamoto, T.: Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics* **66**(1–2), 225–250 (1995) [https://doi.org/10.1016/0304-4076\(94\)01616-8](https://doi.org/10.1016/0304-4076(94)01616-8)
- [46] Farzulla, M.: The Extremity Premium: Sentiment Regimes and Adverse Selection in Cryptocurrency Markets. arXiv:2602.07018 (2025)