DOCUMENT TYPE	DATE	STATUS	FRAMEWORK
Study Spec / Pre-Reg	12 Jul 2026	Locked pre-data	STORM / Block-Bootstrap
PRINCIPAL	COMPANION TO	TARGET PUBLICATION	DISTRIBUTION
B. Sullivan, CIO	"Release Valve" Editorial	w/o 27 Jul 2026	AQI Research / SSRN-track

Fragility Episodes: Resets or Regime Change?

A pre-registered block-bootstrap replication and extension of the J.P. Morgan KOSPI fragility framework across SOX, NVDA, and NQ — testing whether tail-clustering episodes after large trailing runs are positioning resets (as claimed) or early evidence of a deteriorating regime.

PRE-REGISTRATION · HYPOTHESES & PARAMETERS LOCKED BEFORE DATA CONTACT

§1 Motivation & Provenance

J.P. Morgan's cross-asset team (Global Markets Strategy, 10 Jul 2026) advanced an empirical claim about the AI-complex equity trade: that recent tail-clustering episodes in the KOSPI represent *positioning resets after outsized gains* rather than a progression toward more frequent or persistent fragility regimes. Their evidence is an episode table with three rows — drawdowns of −9%, −5%, and −15%, against trailing one-year returns at trough of 58%, 92%, and 151% — and a vol-adjusted tail-count gauge showing stress arriving in bursts rather than a grind higher.

The claim is important, tradeable, and resting on three observations. Our companion editorial ("The Release Valve Is the Crash," 11 Jul 2026, fn. 3) committed us publicly to a proper test. This document is that commitment made rigorous: we lock the episode definition, universe, hypotheses, test statistics, and falsification criteria **before** the analysis runs, so the result — whichever way it lands — carries evidentiary weight rather than narrative convenience.

> WHY PRE-REGISTER
>
> We are publicly critiquing a bulge-bracket desk for fitting a monotonic story to n=3. The critique only has standing if our own methodology forecloses the same sin. Locked parameters, declared nulls, stated falsification thresholds — then data. Any deviation from this spec in the published study will be disclosed and justified in a change log.

The intellectual frame is familiar territory for AQI: this is a conditional forward-return event study with bootstrap nulls, structurally identical to our NVDA failed-breakdown, META outside-reversal, and IWF/IWD ratio studies. The novelty is (a) the episode-based (rather than single-day) event definition, and (b) the explicit test of a *trend in fragility itself* — a claim about the second derivative of market stress, which almost nobody bothers to test formally.

§2 Research Questions & Hypotheses

Three questions, each mapped to a falsifiable hypothesis with a pre-declared null and a stated result that would count against it.

H1 RESET

The "altitude" claim. Forward returns following fragility-episode troughs that occur after large trailing runs (top-tercile trailing 1Y return at trough) are not statistically worse than forward returns following episodes after modest runs (bottom tercile), at horizons of 21, 63, 126, and 252 trading days.

AGAINST JPM / FALSIFIES H1: *top-tercile-run episodes show median forward returns below the bootstrap 5th percentile of the unconditional episode distribution at ≥2 horizons, consistently across ≥2 of 3 core instruments. That result would mean big-run drawdowns are not benign resets.*

H2 NO TREND

The "no regime deterioration" claim. Within each instrument's full history, there is no significant positive time trend in (i) episode frequency (episodes per rolling 3Y window), (ii) episode persistence (duration in trading days), or (iii) episode intensity (peak 20-day tail count within episode).

AGAINST JPM / FALSIFIES H2: *a positive trend coefficient outside the bootstrap 95% band for ≥2 of the 3 fragility dimensions, in ≥2 of 3 core instruments, over the post-2020 subperiod. That result would support the "fragility regimes are worsening" narrative JPM dismisses.*

H3 ENTRY

The "entry-point sensitivity" claim. The risk asymmetry between positioned holders and late joiners is real and quantifiable: conditional max-adverse-excursion (MAE) for a hypothetical entry at the pre-episode peak is significantly larger when trailing 1Y return at entry is in the top tercile, while trough-entry forward returns show no such penalty.

AGAINST JPM / FALSIFIES H3: *no significant difference in peak-entry MAE across trailing-return terciles — which would mean "entry-point sensitivity" is a rhetorical flourish, not a measurable property.*

PRIOR — STATED FOR THE RECORD

Our prior, informed by the failed-breakdown and deep-crush studies, leans toward H1 and H3 surviving (resets after strength have historically resolved upward; entry timing after parabolic runs genuinely is the dominant risk). We are genuinely uncertain on H2 — levered ETF AUM growth and record implied financing are new structural amplifiers with no clean historical analogue, and a detectable post-2024 trend in episode intensity would not surprise us. Stating the prior now so the published result can be scored against it.

§3 Universe & Data Requirements

TABLE 1 · INSTRUMENT UNIVERSE

INSTRUMENT	ROLE	HISTORY TARGET	SERIES	RATIONALE
SOX	Core — sector index	1994 →	Daily close, price index	The upstream AI-hardware complex itself; longest clean semi-sector history; the instrument our editorial cited (−11% from 22 Jun peak).
NVDA	Core — single name	1999 →	Daily close, split-adjusted total return	The crowding epicenter; tests whether episode behavior survives single-name idiosyncrasy; links to existing AQI failed-breakdown study.
NQ / NDX	Core — index future proxy	1996 →	Daily close (NDX cash as continuous proxy)	Where the levered-ETF and index-futures balance-sheet mechanics actually transmit; broadest expression of the trade.
KOSPI	Replication anchor	1998 →	Daily close, price index	Direct replication of JPM's construction. Success criterion: independently recover their three 2025–26 episodes within ±3 trading days of reported start/end dates before extending anywhere else.
SMH	Robustness / tradeable	2000 →	Daily adjusted close	Tradeable proxy check; ties findings to executable instruments for the Flow-Anticipation stack.

- **Source hierarchy:** primary pull via IBKR historical data where available; public daily series (Stooq/Yahoo-class) acceptable for indices with cross-validation of ≥20 randomly sampled dates against a second source. Any disagreement >5bp on a sampled close disqualifies the source for that instrument.
- **Returns:** log daily returns for all vol and threshold computations; simple returns for reported forward-return tables.
- **Corporate actions:** NVDA and SMH must be split/dividend adjusted end-to-end; indices used as published.
- **Data integrity gate:** no analysis proceeds on any series until a written integrity check (missing days vs. exchange calendar, zero-return anomalies, adjustment discontinuities) is logged. This gate exists because we have been burned before and institutionalized the lesson.

§4 Episode Definition — Locked Parameters

We replicate the JPM construction as disclosed, resolving ambiguities with explicit choices (flagged `AQI-resolved`) that are frozen here and swept in §7 robustness.

TABLE 2 · BASELINE EPISODE PARAMETERS

PARAMETER	VALUE	BASIS
Volatility estimator σ	`21-day rolling SD of daily log returns`	JPM: "1M historical vol." 21 trading days = 1 month. `AQI-resolved`: simple SD, not EWMA.
Tail-day trigger	`r_t < -2·σ_(t-1)`	JPM construction; σ lagged one day to avoid contaminating the trigger with the event day itself. `AQI-resolved`
Counting window	`20-day rolling count of tail days`	JPM: "20-day rolling -2σ tail count."
Episode	`Contiguous period where count > 0`	JPM note 1, replicated verbatim.
Episode merge rule	`Gap ≤ 5 trading days ⇒ merge`	`AQI-resolved`: prevents a single quiet week from splitting one economic event into two rows. Swept 0/5/10 in §7.
Minimum severity filter	`Peak-to-trough within episode ≥ 3%`	`AQI-resolved`: excludes trivial single-tail-day episodes that would dilute the event set. Swept off/3%/5%.
Peak / trough	`Max close in 63d pre-episode-start → min close within episode`	`AQI-resolved`: JPM's peak levels imply a lookback beyond episode start; 63d standardizes it.
Trailing run ("altitude")	`252-day simple return as of trough date`	JPM note 2, replicated verbatim.

REPLICATION GATE

Before any extension: apply this definition to KOSPI 2024–26 and confirm recovery of JPM's three episodes (4 Nov–18 Dec 25; 2 Feb–17 Apr 26; 8 Jun–7 Jul 26) with start/end within ±3 trading days and drawdowns within ±2pp of (−9%, −5%, −15%). Failure to replicate halts the study and triggers a definition audit — published either way. A failed replication of a bulge-bracket table is itself a publishable result.

§5 Outcome Variables & Conditioning

5.1 Outcomes (per episode)

- **Forward returns from trough:** +21d, +63d, +126d, +252d simple returns.
- **Forward returns from episode end** (count returns to zero): same horizons — the realistic entry for a signal-driven allocator who cannot call the trough.
- **Drawdown extension probability:** P(new low > 2% below episode trough within +63d) — does the "trough" hold?
- **Peak-entry MAE:** for hypothetical entry at pre-episode peak, maximum adverse excursion over the following 126d (H3).
- **Time to recovery:** trading days from trough to reclaim of pre-episode peak (censored at 252d).

5.2 Conditioning variables

- **Altitude tercile:** trailing 252d return at trough, terciled within-instrument. The primary conditioner — this is JPM's claim.
- **Drawdown depth** and **episode duration**: secondary conditioners for interaction checks.
- **Era flag:** pre/post 1 Jan 2023 (the AI-supercycle boundary) — tests whether the current structural regime (levered ETF stack, record financing) behaves like history at all.

§6 Statistical Methodology

6.1 Bootstrap engine

All inference is against **stationary block-bootstrap** nulls (Politis–Romano), resampling each instrument's full daily log-return history with expected block length set by the Politis–White automatic selector, floored at 21 days to preserve monthly-scale volatility clustering. `B = 10,000` resamples per instrument per test. Random seed fixed at `20260712` and published with the code.

The null construction is the critical design choice: for each bootstrap world, we *re-run the full episode-detection pipeline* on the resampled series and recompute every statistic. The null therefore embeds the question "what do fragility episodes look like in a market with these unconditional dynamics but no special AI-era structure?" — not merely "what do random forward returns look like?" This is the same architecture as the IWF/IWD ratio study and is what makes episode-frequency and trend tests (H2) meaningful.

6.2 Test statistics

TABLE 3 · HYPOTHESIS → STATISTIC MAPPING

HYP	STATISTIC	NULL DISTRIBUTION	DECISION RULE
H1	Difference in median forward return, top vs. bottom altitude tercile, per horizon	Tercile labels shuffled within bootstrap-world episode sets	Two-sided, α=0.05 per horizon; family-wise via Holm across 4 horizons
H2	OLS slope of episode frequency / duration / intensity on episode start date	Slopes from bootstrap worlds (no true trend by construction)	One-sided (worsening), α=0.05 per dimension; Holm across 3 dimensions
H3	Difference in median peak-entry MAE, top vs. bottom altitude tercile	As H1	One-sided (top tercile worse), α=0.05

6.3 Honesty constraints

- **Overlap:** forward-return windows overlap across nearby episodes; the block bootstrap handles serial dependence in the null, and we additionally report Newey–West adjusted point estimates as a cross-check.
- **Small n, declared:** expected episode counts per instrument are ~15–40 over full histories. Every reported result carries its n. No result is promoted to a headline finding unless directionally consistent in ≥2 of 3 core instruments.
- **No peeking:** parameter sweeps (§7) are reported in full grid form — no selective emphasis of the cell that flatters the narrative.

- **Multiple testing:** Holm corrections within hypothesis families as specified; the three hypotheses are treated as separate pre-registered questions, not corrected across.

§7 Robustness Grid

TABLE 4 · PRE-COMMITTED PARAMETER SWEEPS

AXIS	BASELINE	SWEEP
Tail threshold	−2.0σ	−1.5σ · −2.5σ
Vol window	21d	10d · 63d · EWMA(λ=0.94)
Merge gap	5d	0d · 10d
Severity filter	3%	off · 5%
Altitude window	252d	126d · 504d
Subperiods	Full history	Pre/post-2010 · pre/post-2023

Headline findings must survive (same sign, ≥80% of grid cells significant where baseline is significant) or the fragility of the finding itself becomes the finding — reported as such.

§8 Deliverables, Timeline & Integration

13-15 JUL

Data acquisition & integrity gates

All five series pulled, cross-validated, integrity logs written. KOSPI replication gate attempted.

16-18 JUL

Pipeline build & replication sign-off

Episode detector, bootstrap engine, seed-locked. JPM three-episode table reproduced or halt-and-audit triggered.

19-23 JUL

Core runs H1–H3 + robustness grid

10k bootstrap worlds × 5 instruments; full grid archived before any writing begins.

24-26 JUL

STORM writeup & dashboard

Branded HTML dashboard + PDF; results scored against §2 priors; change log vs. this spec appended.

W/O 27 JUL

Publication

AQI-WP-2026-002 released; editorial follow-up note linking back to "Release Valve" fn. 3. Clear of the 1 Aug Query Breadth beta.

8.1 Downstream integration

- **Flow-Anticipation pipeline:** if H2 rejects (fragility is trending), episode intensity becomes a candidate sixth-pillar input or a gate modifier on the momentum pillar. If H1/H3 hold, the altitude tercile becomes a position-sizing conditioner for the Kelly-Thorp tranche machine — fractional-Kelly de-rating for top-tercile-altitude entries.
- **Monitoring stack:** the 20-day tail-count gauge ships as a standing indicator on the AlphaQuery Terminal regardless of hypothesis outcomes — it is a useful state variable even if JPM's interpretation of it is wrong.
- **Editorial:** one follow-up piece either way. "We checked, and JPM was right" is an underused genre and builds more credibility than the alternative.

§9 Known Limitations — Stated in Advance

- **The structural-novelty problem.** The current amplification stack (levered ETF AUM at records, implied financing at the 95th–100th percentile) has no historical analogue in our sample. History can tell us how tail-cluster episodes resolved *under past market structure*; it cannot certify the current one. We will say so, prominently.
- **Episode counts are small.** Even pooled, we expect <150 episodes. The bootstrap gives honest uncertainty bands, but power against subtle trend alternatives (H2) is limited — a non-rejection of H2 is weak evidence of absence, and will be framed as such.
- **NVDA idiosyncrasy.** Single-name episodes embed earnings and product-cycle events that index episodes do not; NVDA results are reported separately, never pooled silently.
- **KOSPI market-structure breaks.** Sidecar/curb rule changes and the 2020 short-sale ban episodes alter tail dynamics exogenously; Korean results carry an events appendix.
- **We are not neutral.** AQI carries positions and published views adjacent to this question. Pre-registration is the mitigation; the reader should still know.

ALPHAQUERY INTELLIGENCE · RESEARCH SPECIFICATION AQI-WP-2026-002 v1.0 · 12 JUL 2026. This document pre-registers a study design and constitutes neither investment advice nor a research finding. References to J.P. Morgan Global Markets Strategy (10 Jul 2026) are paraphrased characterizations of publicly reported research for the purpose of independent replication and critique. Episode data cited (three KOSPI episodes, 2025–26) are as reported by that publication and will be independently re-derived before use. All analysis code, seeds, and parameter grids will be archived at publication. Deviations from this specification will be disclosed in a change log appended to the final study.

◆ ◆ ◆

ALPHAQUERY INTELLIGENCE · TRAVERSE CITY, MICHIGAN · MMXXVI