Dominoes or weather? Separating self-excitation from latent-regime clustering, in three forms — with a power map and a crisis-panel application

Paper C of the Fenocrisis founding program (PROMPTS.md P-2.3; Arc 2). The Q2 mechanism report and the five-literatures identification synthesis. Aimed at an applied-statistics venue (Annals of Applied Statistics-shaped).

Author. Evan Tabak Atlas (ORCID 0009-0007-7374-2338).

Status. Complete draft, gated on owner review before submission (Owner queue). The three-form protocol is stated in §4 in enough detail that a third party can run it on their own data without this repository. Every applied number traces to a committed record (`out/v0/`, `out/v1/x2/`).

Abstract

"True contagion or shared propensity?" is a seventy-year-old identification question that lives, un-reconciled, in five non-communicating literatures: accident-proneness (Bates & Neyman 1952), econometric state-dependence-vs- heterogeneity (Heckman 1981), criminology's boost-vs-flag (Tseloni & Pease 2003), credit's contagion-vs-frailty arc (Das et al. 2007 → Azizpour, Giesecke & Schwenkler 2018 → Berentsen et al. 2022), and the point-process second-order equivalence (Hawkes 1971; Filimonov & Sornette 2015). Each silo asks whether a cluster of events reflects events *triggering* events (self-excitation, *dominoes*) or a shared latent propensity under which events fall together (frailty/regime, *weather*). Each has independently learned that the two are not identified from counts alone and that the estimation bias runs toward false self-excitation. We synthesize the five as one question and propose a resolution that is honest about when it cannot be resolved: a **three-form protocol** — (i) a nested decomposition measuring how much self-excitation survives conditioning on the latent regime and vice versa; (ii) multi-horizon held-out scoring, exploiting that latent-regime persistence and branching persistence separate at long horizons even when one-step fits tie; (iii) a panel identification statistic contrasting contemporaneous cross-stream synchronization against within-stream temporal ordering — each preceded by a **bidirectional recovery gate** that refuses a verdict where the generating mechanism is not recoverable at the available sample size. We map that recoverability as a **discrimination atlas**: the first power curve for Hawkes-vs-Cox discrimination, which shows the mechanism is undecidable below mechanism-specific sample-size onsets and identifiable in a parallel panel where a single stream is not. We apply the protocol to cross-domain crisis event streams. On a six-domain country-year corpus the mechanism is **undecidable at the available configuration** — a committed result, not a gap — and the three forms are read as descriptive decompositions. On a two-domain country-month panel it is decidable, and the three forms lean *weather* at long

horizons and *both* in the nest and the panel — reported as three forms, never as a single winner. **Scope: this concerns the mechanism of onset *timing*; magnitude/severity is separate.**

---

1. One question, five silos

The question is old and recurrent. Whenever events cluster in time — accidents, recidivism, defaults, earthquakes, crises — an analyst must decide whether the clustering means each event *raised the hazard* of the next (self-excitation, state dependence, contagion, *dominoes*) or whether an *unobserved shared propensity* made some units or periods high-hazard throughout (heterogeneity, frailty, a latent regime, *weather*). Five literatures discovered this independently and never cite one another:

- **Accident proneness (Bates & Neyman 1952).** True contagion and apparent contagion (heterogeneity in an underlying Poisson rate) produce the same overdispersed counts; they are **not identified from the count distribution alone.**
- **Econometric state dependence vs. heterogeneity (Heckman 1981).** Whether past participation raises future participation, or a persistent unobserved trait does, is confounded — acutely so through the **initial-conditions problem**, when a process is observed mid-history (as every crisis stream is).
- **Criminology's boost vs. flag (Tseloni & Pease 2003; Johnson 2008).** Does a burglary *boost* the risk of the same address (contagion), or does a stable *flag* (an attractive, poorly-secured target) explain repeats? Same structure, different vocabulary.
- **Credit's contagion vs. frailty (Das et al. 2007 → Duffie et al. 2009 → Lando & Nielsen 2010 → Azizpour, Giesecke & Schwenkler 2018 → Berentsen et al. 2022).** Do defaults cluster because one default triggers others, or because firms share an unobserved *frailty* factor? The arc's terminus is a **nested residual- clustering protocol** (Azizpour et al.) and the finding that adding a regime substantially *reduces* estimated contagion (Berentsen et al.).
- **Point-process second-order equivalence (Hawkes 1971; Filimonov & Sornette 2015; Wheatley, Filimonov & Sornette 2016).** Hawkes (1971) proved a self-exciting and a doubly-stochastic process can share identical second-order properties; Filimonov & Sornette (2015) showed a Hawkes fit to regime-switching Poisson data manufactures spurious near-critical branching ratios; Wheatley et al. (2016) showed misspecified immigration inflates the branching ratio. **The bias always runs toward false self-excitation.**

Cox's (1981) parameter-driven vs. observation-driven taxonomy names the two arms cleanly: a latent-regime Cox process is parameter-driven (the intensity is a function of an exogenous hidden state); a Hawkes process is observation-driven (the intensity is a function of the process's own past). The question "dominoes or weather?" is exactly "observation-driven or parameter-driven?", and the five silos are five discoveries that you cannot answer it from marginal counts.

The synthesis is itself a contribution: naming these as one problem, with one vocabulary and one honest apparatus, is what lets a crisis-domain application inherit seventy years of hard-won caution instead of rediscovering it.

---

2. The two arms, made estimable and raced fairly

We instantiate *weather* as an MSM-Cox process — a k-level Markov-switching-multifractal latent cascade (Calvet & Fisher 2001/2004) modulating a Cox intensity through an exact 2^k-state recursive filter — and *dominoes* as a discrete-time Hawkes process with an exponential kernel. The nesting vehicle is the **hybrid**: an MSM-Cox baseline plus a Hawkes excitation term, which lets the two be conditioned on each other rather than only compared. All three, plus the independent-Poisson null and flexibility controls (a flat K-state Poisson-HMM; Bayesian-blocks piecewise-constant Poisson), are fit under one fairness discipline (inherited verbatim from a sibling tournament): identical exposure and baseline covariates for every arm, identical held-out splits (temporal blocks *and* cross- sectional folds), BIC and held-out log-loss both reported, parametric bootstrap on every verdict, and depth-honest penalties. The MSM-Cox arm is a solved technology (Chen, Diebold & Schorfheide 2013 fit an MSM-modulated point-process intensity in finance) that has simply never met a Hawkes opponent or left its home domain.

---

3. The power map: a discrimination atlas (record v0)

Before any verdict, we map where a verdict is *possible*. No published power curve for Hawkes-vs-Cox discrimination exists anywhere; the atlas is that curve, and it is a standalone methods result. It is simulated known-truth only — no real data enters (the F6 principle): each arm generates streams across a pinned grid (event count N ∈ {100, 300, 1,000, 10,000, 100,000}, Hawkes branching ratio η ∈ {0.2, 0.5, 0.8, 0.9}, ladder depth k ∈ {2, 3, 4}), every arm is refit to every stream, and a cell is **decidable** when the generating arm is recovered (selected by BIC) in ≥ 80% of replications. Seeds and the bit-reproducible numeric record are in `out/v0/` (`atlas.json`).

Three results carry over into the protocol:

- **Discrimination is a story about sample size.** Of 45 single-stream cells, 26 are decidable and 19 are not. The onset (smallest N at which the true arm is recovered) is mechanism-specific: flat Poisson N ≈ 100, MSM-Cox N ≈ 300, self-exciting Hawkes N ≈ 1,000, regime-switching HMM N ≈ 10,000. **Below its onset, a mechanism is undecidable, and any in-sample "winner" there is an artefact of sample size.**
- **The bias is measured, not assumed.** The mandatory anchor cell reproduces Filimonov–Sornette: regime-switching Poisson truth with η = 0 (no self-excitation at

all) is fit by a Hawkes arm as a near-critical **$\hat{\eta}$ = 0.80**. Its mirror — a genuine Hawkes truth (η = 0.5) at N = 3,000 — is preferred over MSM-Cox by BIC in 8/8 replications: discrimination runs the right way when self-excitation is real.

- **The panel carries the power.** A shared latent regime that beats the independent null with probability only 0.25 on one short stream is recovered ever more reliably as parallel streams sharing that regime are added. The power lives in the cross-section, not the length of any one series — the design fact that makes form (iii) of the protocol worth running. **Dated correction (2026-08-25):** this bullet previously said the threshold is crossed "with **D = 3 parallel streams**". Record `out/v4` retracts that number. Rebuilt at **R = 8** replications (2× v0's R = 4, so recovery is not a quarter-granular fraction) and with per-domain heterogeneity, the measured onsets are **D = 4 for homogeneous replicates** — v0's "D = 3" cell reads recovery **0.75** at R = 8, just below the 0.8 threshold — **D = 2 for heterogeneous-mixed** domains (mean loading 0.6; the strongly-coupled domains concentrate the identification), and **not reached anywhere in the D grid for weakly coupled** domains (mean loading 0.3). The design fact survives; the count does not. **D-licensing is coupling-strength-conditioned**, which is why form (iii) is read against its atlas cell and not against a domain count.

(The near-critical corner, where a power-law Hawkes kernel and a multiplicative cascade mutually mimic and the choice between them degenerates, is mapped separately as a degenerate atlas cell — record v2 — and is the frontier of this power map, not part of the founding tournament.)

---

4. The three-form protocol (runnable without this repository)

The protocol answers "dominoes or weather?" in the only three forms the identification literature supports, and **never as a single winner-take-all number.** Where the forms disagree, the disagreement *is* the result. A third party can run all of it on their own event streams from the following specification.

Gate (run first, always) — the bidirectional recovery check. Fit the pure Hawkes arm and the pure MSM-Cox arm to the data. From each fitted arm, simulate B synthetic streams at the *actual* configuration (same N, same exposure, same baseline). Refit both arms to each synthetic stream. The mechanism is **decidable** iff each generating arm is recovered (lower BIC than its rival) in ≥ 80% of its own simulations. Cross-reference the operative cell to a discrimination atlas (§3). *If the gate fails, the committed result is "undecidable at this N/design," with the atlas cell cited — the three forms below are then run and reported as descriptive decompositions, not verdicts.* This is the load-bearing honesty rule (§5).

Form (i) — the nested decomposition (after Azizpour, Giesecke & Schwenkler 2018; Berentsen et al. 2022). Fit the hybrid (baseline + latent ladder + excitation). Report the *survival fraction*: how much of the pure excitation arm's likelihood gain over the baseline survives once the latent ladder is also present (and, symmetrically, how much of the ladder's gain survives once excitation is present). A large drop in the estimated branching ratio when the regime is added is the Berentsen signature of regime-clustering absorbing apparent contagion. After fitting the full nest, run a **residual-clustering test**: the within-stream lag-1 residual autocorrelation and the cross-stream contemporaneous correlation should fall inside a simulation reference band if the nest has captured the structure; residual cross-stream contemporaneous correlation that the one-step predictor cannot anticipate is a shared-regime signature (read with form iii), not a misspecification.

Form (ii) — multi-horizon held-out scoring. For each pure arm, compute predictive log-loss on a held-out block at horizons $h \in \{1, 5, 20\}$ (adapt the grid to the series length). The MSMD precedent: latent-regime persistence and branching persistence separate at 5–20-step horizons even when one-step fits tie — a latent-regime arm that ties at $h = 1$ but wins at $h = 20$ is *weather*. Report the Δlog-loss vs. the null at each horizon per arm; read the *ordering* across horizons, not one number. (Held-out *magnitudes* on trended, heavy-count streams are extrapolation-sensitive; cap the forecast at a physical ceiling and read robustness qualitatively — the audit of this is Paper B.)

Form (iii) — panel identification. For a panel of parallel streams (countries), compute two statistics: a **cross-stream synchronization** statistic (contemporaneous coincidence across streams, above each stream's own rate) and a **within-stream ordering** statistic (temporal lead–lag structure within a stream). A shared latent regime synchronizes streams contemporaneously with no within-stream ordering; self-excitation is ordered within streams. Test each against a **per-stream circular-shift surrogate** (shift each stream's fluctuation around its fitted baseline, destroying cross-stream alignment while preserving the trend and marginal autocorrelation), and report the excess over the surrogate with its tail p-value, per domain. Detrend each stream before measuring co-movement (a low-degree polynomial strips the reporting ramp); audit the detrend degree separately (Paper B).

The protocol's outputs are three excess-with-uncertainty statements plus a gate outcome — never a scalar "Hawkes vs. Cox" verdict.

---

5. Undecidability is a result, not a gap

The gate makes "we cannot tell" a first-class, publishable outcome rather than an embarrassment papered over with an in-sample winner. This is the methodological core. The point-process degeneracy (§1) guarantees that in some (N, η, k, design) region the two arms are second-order equivalent; the atlas (§3) maps that region on known truth; the gate (§4) refuses a verdict inside it on real data and commits the undecidability statement with the atlas cell cited. A field that reports a branching ratio wherever it can fit one will report false self-excitation exactly where the bias is strongest (near criticality, small N); a field that runs the gate first reports *nothing* there, and says so. The committed "undecidable at this N/design" is the honest-practice deliverable, and it is the one the crisis application below actually yields on its larger corpus.

---

6. Application: cross-domain crisis event streams (record v1 / x2)

We apply the protocol to two crisis corpora (Paper A): Corpus-A (Delannoy six-domain country-year, 173 countries, 1970–2019) and Corpus-B (UCDP GED conflict + WHO-DON epidemic country-month, 192 countries, 1990–2024). Seed `20260902`, profile `default`; gate simulations B = 40, synchronization surrogates 300; records `out/v1/x2/` (`x2.json`, `x2_three_form.png`, `x2_summary_table.png`). The protocol is scoped by the Q1 entanglement result (Paper A): form applies "where coupling exists."

6.1 Corpus-A — the gate fails; undecidable

The bidirectional recovery gate **fails**: MSM-Cox is recovered over Hawkes in 1.0 of its simulations, but **Hawkes is recovered over MSM-Cox in 0.0** (n = 40, threshold 0.8) — the Filimonov–Sornette bias confirmed at the actual N and design, the tournament always preferring the latent-regime arm. And Q1 was negative here (Paper A: no cross-domain excess survives the surrogate). **The committed result is undecidability at this configuration.** The three forms are therefore run and reported as *descriptive decompositions of within-domain structure*, not as a cross-domain mechanism verdict (none was established):

- **(i)** excitation surviving conditioning on the ladder = 0.104 of its raw gain (ΔBIC +4,556); the ladder surviving conditioning on excitation = 0.118 (ΔBIC +5,471); the branching ratio drops 0.917 → 0.590 (shrinkage 0.643) when the regime is added. Residual clustering after the full nest (reference band ±0.283): within-stream lag-1 0.736 → 0.555, cross-domain contemporaneous 0.259 → 0.134.
- **(ii)** Δlog-loss vs. null at $h \in \{1, 3, 5\}$: hawkes {457.0, 194.5, −77.7}; msm_cox {152.3, 55.2, −28.8} — both decay to negative by h = 5 (a country-year series is short; the horizon separation is weak here).
- **(iii)** cross-sync excess is significant (p = 0.003) in five of six domains — CLIMATIC 0.047, CONFLICTS 0.112, ECOLOGICAL 0.081, ECONOMIC 0.064, TECHNOLOGICAL 0.046 — and null in GEOPHYSICAL (0.001, p = 0.435); within-stream ordering excess is near zero throughout. The contemporaneous-without-ordering signature is a within-domain regime pattern here, not cross-domain entanglement.

6.2 Corpus-B — the gate passes; three forms, never one winner

The gate **passes**: MSM-Cox recovered over Hawkes 1.0, Hawkes recovered over MSM-Cox 0.975 (n = 40, threshold 0.8) — decidable. Q1 was positive here (a cross-domain excess survived the surrogate; Paper A), so the three forms are a cross-domain mechanism *attribution*:

- **(i) — both.** Excitation surviving conditioning on the ladder = 0.242 of its raw gain (ΔBIC +20,881); the ladder surviving conditioning on excitation = 0.106 (ΔBIC +7,694); the branching ratio drops only 0.993 → 0.959 (shrinkage 0.965). Residual clustering after the full nest (band ±0.098): within-stream lag-1 0.772 → −0.012 (excitation captured), cross-domain contemporaneous 0.210 → 0.016 (a shared contemporaneous shock the one-step predictor cannot anticipate — the regime signature).

- **(ii) — weather.** Δlog-loss vs. null at h ∈ {1, 5, 20}: hawkes {152.6, 102.2, **30.5**}; msm_cox {90.5, 67.5, **42.1**}. Hawkes wins at the short horizon; the latent-regime arm **crosses over and wins at h = 20** (42.1 vs. 30.5) — the MSMD long-horizon separation, leaning *weather*.
- **(iii) — both.** Cross-sync excess significant in both domains (conflict 0.018, epidemic 0.019; both p = 0.003), within-stream ordering near zero (conflict 0.003, epidemic 0.000) — contemporaneous synchronization dominates ordering, a regime signature, but on a coupling that Paper A's audit then found fragile.

The three-form reading is mixed: (i) both, (ii) weather, (iii) both. No single number is reported as the winner (the house rule). And it is reported subject to the audit: Paper A found the Corpus-B cross-domain coupling fragile to the μ(t) baseline and absorbable by a shared attention covariate — so the mechanism attribution here decomposes a coupling that does not survive as a robust claim. What *does* survive the audit is the within-domain mechanism structure (the excitation-given-ladder term is baseline-robust; Paper B §3.1): the leaning-weather long-horizon signature is a property of the streams, even where the cross-domain coupling it attributes is not robust.

---

7. What the protocol buys, and its limits

The protocol converts a seventy-year-old confound into a reproducible procedure with an honest failure mode. Its guarantees: it never reports a mechanism where the gate says the mechanism is unrecoverable; it never collapses the answer to one number; and it inherits the identification caution of five literatures instead of rediscovering their errors. Its limits: the gate's threshold (80%) and simulation count are choices, and near-boundary cells carry Monte-Carlo noise; the multi- horizon magnitudes are extrapolation-sensitive on trended count streams (audited in Paper B); and the whole apparatus concerns the mechanism of **onset timing** — the clock — not **severity or magnitude**, whose home is the multiplicative cascade and whose mechanism tournament is a separate deliverable (a marked-process atlas and tournament, forthcoming). Nothing here forecasts, warns, ranks, or scores; every number is a retrospective, dataset-level model-comparison on public, aggregate data (F7).

---

8. Data and code availability

```
python3 -m venv venv && venv/bin/pip install -r requirements.txt
venv/bin/python x0_verify.py            # the F1 anchor gate — must print X0 PASS
venv/bin/python x0_atlas.py             # record v0, the power map (the atlas §3)
./fetch_data.sh                         # open-tier data (Delannoy, UCDP GED, WHO-DON)
venv/bin/python x2_mechanism.py --profile default --seed 20260902   # → out/v1/x2/
```

Records: `out/v0/` (the discrimination atlas; `atlas.json`, `decidability_atlas.png`, `filimonov_sornette.png`, `panel_gain.png`) and `out/v1/x2/` (the three-form mechanism report; `x2.json`, `x2_three_form.png`, `x2_summary_table.png`). Environment: Python 3.11.15, NumPy 2.4.6, SciPy 1.17.1, Matplotlib 3.11.1. License: CC BY 4.0.

Deposit DOIs (published 2026-08-31). This paper and the two records it reads are deposited separately.

- This paper — `10.5281/zenodo.22180509`.
- Record `out/v0` (the discrimination atlas) — `10.5281/zenodo.22180515`.
- Record `out/v2` (atlas v2, the degenerate cell) — `10.5281/zenodo.22180517`.

Each DOI above is **published**: the deposits went live on Zenodo on **2026-08-31** and every identifier resolves. The source repository is private and will remain so, so each deposit is standalone and carries its own reproduction materials; deposit metadata deliberately omits repository links. *(Corrected 2026-08-31 — publication day: until today this paragraph read "reserved on Zenodo and its deposit is a private draft … the record is not yet published", with "(Zenodo DOI reserved; publication pending)" wherever a DOI appeared. That staging state ended when the owner pressed Publish on 2026-08-31.)* The three-form protocol (§4) is specified independently of this codebase and can be run on any parallel event-stream panel.

References

(Verified against the actual papers before release — Standing Rule 11.)

- Azizpour, S., Giesecke, K., & Schwenkler, G. 2018. Exploring the sources of default clustering. *Journal of Financial Economics* 129(1), 154–183.
- Bates, G. E., & Neyman, J. 1952. Contributions to the theory of accident proneness. II. True or false contagion. *University of California Publications in Statistics* 1(10), 255–275.
- Berentsen, G. D., Bulla, J., Maruotti, A., & Støve, B. 2022. Modelling clusters of corporate defaults: regime-switching models significantly reduce the contagion source. *Journal of the Royal Statistical Society Series C (Applied Statistics)* 71(3), 698–722.
- Calvet, L. E., & Fisher, A. J. 2001. Forecasting multifractal volatility. *Journal of Econometrics* 105(1), 27–58; 2004. How to forecast long-run volatility: regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics* 2(1), 49–83.
- Chen, F., Diebold, F. X., & Schorfheide, F. 2013. A Markov-switching multifractal inter-trade duration model, with application to US equities. *Journal of Econometrics* 177(2), 320–342.
- Cox, D. R. 1981. Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* 8(2), 93–115.

- Das, S. R., Duffie, D., Kapadia, N., & Saita, L. 2007. Common failings: how corporate defaults are correlated. *Journal of Finance* 62(1), 93–117.
- Duffie, D., Eckner, A., Horel, G., & Saita, L. 2009. Frailty correlated default. *Journal of Finance* 64(5), 2089–2123.
- Filimonov, V., & Sornette, D. 2015. Apparent criticality and calibration issues in the Hawkes self-excited point process model. *Quantitative Finance* 15(8), 1293–1314.
- Hawkes, A. G. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58(1), 83–90.
- Heckman, J. J. 1981. Statistical models for discrete panel data (pp. 114–178), and The incidental parameters problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process (pp. 179–195). In *Structural Analysis of Discrete Data with Econometric Applications* (Manski & McFadden, eds.), MIT Press.
- Lando, D., & Nielsen, M. S. 2010. Correlation in corporate defaults: contagion or conditional independence? *Journal of Financial Intermediation* 19(3), 355–372.
- Tseloni, A., & Pease, K. 2003. Repeat personal victimization: "boosts" or "flags"? *British Journal of Criminology* 43(1), 196–212.
- Wheatley, S., Filimonov, V., & Sornette, D. 2016. The Hawkes process with renewal immigration & its estimation with an EM algorithm. *Computational Statistics & Data Analysis* 94, 120–135.

Full program reference list: SPECS.md §10.