A discrimination atlas for Hawkes-versus-Cox model selection: where self-excitation and latent-regime clustering can be told apart, and where they cannot

A short methods note of the Fenocrisis program (PROMPTS.md P-5.7). The discrimination atlas — record `out/v0` and its near-critical degenerate cell `out/v2` — presented on its own: the first published power map for Hawkes-versus-Cox model selection, a service to anyone racing a self-exciting against a latent-regime model on their own counts, and publishable independent of any coupling verdict.

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

Status. Complete draft, gated on owner review before submission (Owner queue). Aimed at a short methods/software venue or a statistics-methods letters format. Every number traces to a committed, regenerable record (`out/v0/`, `out/v2/`); the atlas is queried programmatically by later work in this program, and is offered here as a stand-alone instrument. Nothing in it forecasts, warns, or ranks (F7).

Abstract

Whenever events cluster in time — accidents, defaults, earthquakes, crises — an analyst must decide whether each event *raised the hazard* of the next (self-excitation, a Hawkes process — *dominoes*) or whether an *unobserved shared propensity* made some periods high-hazard throughout (a latent regime, a Cox/MSM-Cox process — *weather*). The two are famously hard to separate from counts, and the estimation bias runs toward false self-excitation (Filimonov & Sornette 2015). Practitioners nonetheless routinely fit one model, report a branching ratio or a regime, and move on — with **no published power curve** telling them whether their sample could have distinguished the alternatives at all. This note supplies one. We simulate from a five-arm roster (homogeneous Poisson, Poisson-HMM, exponential-kernel Hawkes, MSM-Cox multiplicative cascade, Bayesian Blocks), refit **every** arm to **every** stream, and map — across event count `N`, branching ratio `η`, cascade depth `k`, and single-stream-versus-parallel-panel design — the region where the true generating arm is recovered by BIC in at least 80 % of replications, gated by a bidirectional recovery check (a verdict is refused where the generating mechanism is not recoverable). The resulting **discrimination atlas** (`out/v0`) shows the mechanism is undecidable below arm-specific sample-size onsets and identifiable in a parallel panel where a single stream is not; it reproduces the Filimonov–Sornette spurious-self-excitation inflation as a labelled anchor. A companion **degenerate cell** (`out/v2`) adds a near-critical power-law Hawkes generator — the case the exponential roster cannot represent — and finds a clean *crossover*: near criticality a self-exciting process is selected as a cascade, and a ladder-span discriminator is confounded exactly as the branching

ratio is. The atlas is a queryable lookup (`decidable(N, η, k, design) → bool + margin`) and is released for reuse.

---

1. The gap: everyone fits, no one reports the power

"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-versus-heterogeneity (Heckman 1981), criminology's boost-versus- flag (Tseloni & Pease 2003), credit's contagion-versus-frailty arc (Das et al. 2007; Azizpour, Giesecke & Schwenkler 2018), and the point-process second-order equivalence (Hawkes 1971; Filimonov & Sornette 2015). Each independently learned that self-excitation and shared latent propensity are **not identified from the count distribution alone**, and that the estimation bias runs toward false self-excitation: a regime-switching Poisson stream fit by a Hawkes model reports a spurious near-critical branching ratio.

What none of the five supplies is the practitioner's actual question: *given my sample — this many events, this branching strength, this design — could the fit have told the two mechanisms apart in the first place?* Power curves are standard for detecting an effect against a null; there is, to our knowledge, **no published power curve for discriminating Hawkes from Cox/regime model selection**. Absent one, a reported branching ratio or regime is un-interpretable: it may be a finding or it may be the sample-size floor of an undecidable problem. This note maps that floor.

2. Construction

Roster (arms refit to every stream). Homogeneous Poisson; a two-state Poisson-HMM; an exponential-kernel Hawkes fit by exact likelihood; an MSM-Cox multiplicative-cascade Cox process (the Calvet–Fisher 2^k geometric ladder, marginalised exactly by the forward algorithm); and a Bayesian-Blocks change-point model. The Poisson and HMM arms are the degenerate corners; Hawkes and MSM-Cox are the two mechanisms of interest; Bayesian Blocks is the non-parametric foil. All are exact-likelihood or exact-enumeration; none is a heuristic.

Grid (`out/v0`). Event count `N ∈ {100, 300, 1000, 10 000, 100 000}`; Hawkes branching ratio `η ∈ {0.2, 0.5, 0.8, 0.9}`; MSM-Cox truth depth `k ∈ {2, 3, 4}` with the fitted arm capped at `k = 3` (a stated honesty ceiling — a deeper truth is still detected as *some* cascade, which is the discrimination signal); a regime-persistence axis; and a **panel** configuration of many short parallel streams sharing one regime. Series length is capped at `T ≤ 1500` ticks, the rate rising with `N` above the cap, so cost scales with `N` rather than an unbounded tick count — the whole grid runs at laptop scale (≈ 19 min).

Selection rule and recovery gate. In each cell every arm is refit to each replication; the cell is **decidable** if the true generating arm is selected by BIC in ≥ 80 % of replications, with held-out log-loss and the mean BIC margin reported as cross-checks (not gates).

Decidability is a statement about *recovery of the generating mechanism* under a bidirectional check — simulate from each arm, refit, confirm the generator is recovered — not merely about one model beating a null. Where a configuration is not recoverable, the honest output is "undecidable at this `N`", not a winner.

3. What the atlas shows (`out/v0`)

- **A decidable region and an undecidable region.** Of the 45 single-stream cells, **26 are decidable and 19 are not.** Discrimination is not a property of a model; it is a property of a model *at a sample size and design*.
- **Mechanism-specific sample-size onsets.** The smallest `N` at which each arm's own streams are recovered: homogeneous Poisson `N = 100`; MSM-Cox `N = 300`; exponential Hawkes `N = 1 000`; Poisson-HMM `N = 10 000`. A latent cascade is recoverable at an order of magnitude fewer events than a two-state HMM — the cascade's multi-scale signature is easier to read than a flat two-state regime.
- **The Filimonov–Sornette anchor, reproduced and labelled.** A regime-switching Poisson truth (no self-excitation, `η = 0`) fit by the Hawkes arm infers `η̂ = 0.80` — spurious near-critical self-excitation, the field's canonical failure, here reproduced as a fixed anchor cell rather than a footnote.
- **The panel-design gain.** A *single* short stream recovers a shared latent regime with probability only **0.25**; adding parallel streams that share the regime raises the recovery until it crosses the 80 % decidability threshold. Power for latent-regime discrimination lives in the cross-section, not in a longer single series — the design lesson that motivates panel, rather than aggregate-series, tests. **Dated correction (2026-08-25):** this bullet previously said **three** parallel streams cross the threshold. Record `out/v4` retracts that count. Re-run at **R = 8** replications — 2× the R = 4 the v0 panel cell used, so recovery is no longer quarter-granular — and extended to per-domain heterogeneity, the measured onsets are **D = 4** for homogeneous replicate streams (the D = 3 cell reads recovery **0.75** at R = 8, just below threshold: the original "three" was the quarter-granular artifact), **D = 2** for heterogeneous-mixed domains (mean loading 0.6 — earlier, because the strongly-coupled domains concentrate the identification), and **not reached anywhere in the D grid** for weakly coupled domains (mean loading 0.3). The threshold is therefore **conditioned on coupling strength**, not on a domain count — itself an instance of this note's thesis, that decidability is a property of a model *at a sample size and design*.

4. The degenerate cell: near-critical power-law Hawkes (`out/v2`)

The `out/v0` roster is exponential-kernel throughout. The one case it cannot represent is the process that most threatens the whole exercise: a **nearly-unstable Hawkes with a power-law kernel tail** `x^{-(1+α)}`, `α ∈ (½, 1)`, which converges (Jaisson & Rosenbaum 2015) to a *rough* volatility limit with Hurst exponent `H = α − ½`. If "rough drift" and "near-critical dominoes" are one scaling cell, the trichotomy of §2 should degenerate there.

`out/v2` adds that generator (Ogata thinning with the kernel as a finite sum of exponentials) over `α ∈ {0.55, 0.6, 0.7, 0.9} × η ∈ {0.8, 0.9, 0.95, 0.99, 0.995, 0.999}`, and scores four pre-registered predictions:

- **A crossover, not the naive limit.** At `η ≥ 0.99`, MSM-Cox has the lower BIC than the true exponential-Hawkes arm in **0.81** of replications; the true Hawkes family is the outright roster winner only **0.05** of the time (the Bayesian-Blocks arm is the modal near-critical pick). A near-critical self-exciting process is **read as a multiplicative cascade** — the crossover begins around `η ≈ 0.8` across the sampled `α`. Which mechanism a fit reports near criticality is an artefact of the sample, not the generator.
- **The loop closes.** The `α = 0.611` generator, passed through a log-scale variogram roughness estimator, reads `Ĥ = 0.157` — squarely rough (`H < ½`) and in the neighbourhood of the predicted `H = α − ½ = 0.111`. Reported in the identifiable amplitude coordinate `λ² = H(1 − 2H)ν² = 0.103` (never as a bare `Ĥ`), it places near-critical endogeneity and rough drift in one cell.
- **A confounded ladder span.** MSM-Cox fitted to single-mechanism power-law-Hawkes truth reports a broad, genuinely-loaded ladder span **`S = 1.14` decades — 87 % of a real multi-scale cascade's `S = 1.31`** — on data with *no* hierarchy at all. Any span-based discriminator inherits the same caveat as the branching ratio: a span is only ever an excess over a matched null.
- **A discriminator that does not survive, either.** A time-asymmetry statistic `τ` (exactly odd under time reversal, and provably zero for the reversible MSM-Cox counts) has real power on fine-resolution signals but **washes out on the coarse binned near-critical counts**: its sign is a near-coin-flip (sign-consistency ≈ 0.60) and `|t| < 2`. Even the discriminator that keys on endogeneity itself does not separate the two at the resolution crisis counts actually have.

5. Using the atlas

The atlas is a queryable JSON lookup, not just a set of figures. The intended use is **pre-registration**: before fitting a Hawkes-versus-regime comparison on real counts, query `decidable(N, η, k, design)` (returning a boolean and a margin) at the study's own configuration, and commit in advance to reporting "undecidable at this `N`" if the cell says so. A verdict is licensed only inside the decidable region; outside it, a fitted branching ratio or regime is reported as descriptive, never as a mechanism finding. This is exactly how the rest of the Fenocrisis program consumes the atlas — every coupling verdict cites its operative cell — and the same discipline transfers to any counts domain.

A companion **marked atlas** (`out/v6`) extends the same construction to the *severity* (mark) channel — where a latent regime or self-excitation modulates event *magnitude* rather than timing — with its own decidability floors; it is a natural sequel for magnitude-clustering applications and is described in the program's severity records.

6. Availability

The atlas, its regeneration commands and pinned seeds, the selection matrices, and the machine-readable lookup are in the committed records `out/v0/` and `out/v2/` of the Fenocrisis repository (CC BY 4.0). Both are deterministic from the recorded seeds and reproduce bit-for-bit; the full grid runs at laptop scale. Because the atlas is known-truth simulation throughout, it carries no data-licensing restriction and can be regenerated, extended, or re-gridded by any reader.

References (to verify against sources before submission — Standing Rule 11)

- Bates, G. E. & Neyman, J. (1952). Contributions to the theory of accident proneness. *University of California Publications in Statistics*.
- Heckman, J. J. (1981). Heterogeneity and state dependence. In *Studies in Labor Markets*.
- Tseloni, A. & Pease, K. (2003). Repeat personal victimization: boosts or flags? *British Journal of Criminology* 43(1).
- Das, S. R., Duffie, D., Kapadia, N. & Saita, L. (2007). Common failings: how corporate defaults are correlated. *Journal of Finance* 62(1).
- Azizpour, S., Giesecke, K. & Schwenkler, G. (2018). Exploring the sources of default clustering. *Journal of Financial Economics* 129(1).
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58(1).
- Filimonov, V. & Sornette, D. (2015). Apparent criticality and calibration issues in the Hawkes self-excited point process model. *Quantitative Finance* 15(8).
- Calvet, L. E. & Fisher, A. J. (2001, 2004). Forecasting multifractal volatility; How to forecast long-run volatility. *Journal of Econometrics* / *Journal of Financial Econometrics*.
- Jaisson, T. & Rosenbaum, M. (2015). Limit theorems for nearly unstable Hawkes processes. *Annals of Applied Probability* 25(2).
- Scargle, J. D. et al. (2013). Studies in astronomical time series analysis. VI. Bayesian Blocks. *Astrophysical Journal* 764(2).

Every claim in this note traces to `out/v0/RECORD.md` or `out/v2/RECORD.md`; the reference list is verified against the actual papers before release (Standing Rule 11). No forecasting, warning, or ranking language appears anywhere (F7). License: CC BY 4.0.

Deposit DOIs (published 2026-08-31). This note and the two atlases it reports are deposited separately.

- This note — `10.5281/zenodo.22180513`.
- Record `out/v0` (the discrimination atlas) — `10.5281/zenodo.22180515`.

- Record `out/v2` (atlas v2, the near-critical power-law 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.)*