# Exogenous cascades versus self-excitation: distinguishing Markov-Switching-Multifractal Cox processes from Hawkes processes, with an application to musical onsets

Evan Tabak Atlas (Independent)

ORCID 0009-0007-7374-2338, Evanatlas@gmail.com

Preprint draft. Target venue: Chaos / Physica A / Physical Review E (physics framing).

---

ABSTRACT

Bursty, hierarchically clustered event streams are almost universally modeled as self-exciting **Hawkes** processes. We study an alternative mechanism — **exogenous** hierarchical regime switching — realized as a doubly-stochastic (Cox) process whose intensity is a discrete-time Markov-Switching Multifractal (MSM): $N_t \sim \mathrm{Poisson}(\lambda_0 \cdot G_t^{\kappa})$ with $G_t$ the product of $k$ mirror-balanced two-state multipliers on a geometric rate ladder. We derive the closed-form stationary moments, overdispersion, and autocovariance of the MSM-Cox counts, and show the autocovariance is an exact sum of geometrically-spaced exponentials — the 1/f mechanism. We then prove a **mimicry** result: because a Hawkes count autocovariance is itself an exponential sum, a Hawkes kernel can be fit to match the MSM-Cox second-order structure to $R^2 > 0.999$. The processes nonetheless **separate at higher order**: MSM-Cox is genuinely multifractal (MFDFA width $h(-2)-h(2) > 0$), has heavier inter-burst quiet-time tails, and — decisively — has an intensity that is *autonomous* of realized events, which the time-rescaling residual test detects. A likelihood cross-fitting study yields a strongly diagonal model-selection matrix at $10^4$+ events, including for **near-critical** Hawkes (branching ratio 0.9), the regime where second-order mimicry is strongest. Applied to onset streams from the Fenophone generative instrument (an MSM-Cox by construction), the tools select MSM-Cox and attribute the decision to the discriminating statistics. Applied to real music (first corpus run 2026-07-12; seeded 120-piece samples per corpus on a 12-ticks/beat grid), min-BIC selection chooses **MSM-Cox for 84% of MAESTRO piano performances** — real piano onset streams prefer exogenous regime-switching over self-excitation — while Groove drum-kit streams on the beat grid select Poisson/Hawkes (58%/37%), consistent with the companion finding that the drum idiom's cascade lives in velocity, not count density. The earthquake positive control (run 2026-07-13, SCEDC M≥2.5 2010–2020): handed a Hawkes-true catalog, min-BIC selects Hawkes — decisively with a two-exponential (Omori-approximating) kernel — with the instructive twist that MSM-Cox out-fits a *single*-exponential Hawkes on those same aftershocks, quantifying on real data that likelihood selection between mechanism classes is fair only when each class is given comparable timescale flexibility (mimicry cuts both ways; §7a).

## 1. Introduction

Bursty, hierarchically clustered event streams — aftershock sequences, order flow in financial markets, neural spike trains, cascades of social activity, the onsets of a musical performance — are almost universally modeled as **self-exciting** point processes: the linear Hawkes process (Hawkes 1971) and, in seismology, its descendant the ETAS model (Ogata 1988). The Hawkes conditional intensity is a functional of the realized event history, so adopting it is more than a curve fit. It is a **mechanism** claim: events beget events; each arrival transiently raises the probability of further arrivals. In some applications that is exactly what is wanted — an aftershock really is triggered by a mainshock. In many others the Hawkes process is adopted because it is the convenient, estimable model of clustering, and the mechanism claim rides along silently.

This paper develops the natural alternative: **exogenous** hierarchical regime switching. Here a latent intensity moves on its own — a slow arc of activity with nested faster fluctuations — and events are emitted conditionally independently given that arc. A flurry does not mechanically cause the next flurry; it merely reveals that the latent intensity is currently high. The classical container for this alternative is the doubly stochastic, or Cox, process (Cox 1955). We give the latent intensity a specific, parsimonious, estimable form: the discrete-time **Markov-Switching Multifractal** (MSM) of Calvet and Fisher (2001, 2004, 2008), a product of $k$ independent two-state multipliers switching on a geometric ladder of timescales — the stationary, causal, Markov realization of a multiplicative cascade. The resulting **MSM-Cox** process is, we will argue, the sharpest available foil to the Hawkes process: it produces the same phenomenology — bursts, lulls, overdispersion, a 1/f-family spectrum — from a mechanism that contains no feedback whatsoever.

How distinguishable are the two mechanisms, in principle and in practice? The answer we establish is deliberately two-sided. At **second order** they are not distinguishable: the MSM-Cox count autocovariance is an exact sum of geometrically spaced exponentials (§3), a Hawkes count autocovariance is itself an exponential sum, and we exhibit fits matching one to the other to $R^2 > 0.999$ (§4). At **higher order** they separate — by multifractal width, by inter-burst quiet-time tails, and, most sharply, by whether the intensity is autonomous of the realized events, which the time-rescaling residual test detects (§5) — and a likelihood cross-fitting study shows that finite-sample model selection resolves the mechanisms even in the near-critical Hawkes regime where the second-order statistics converge (§6).

Musical onset streams are our arena. They are bursty and hierarchically structured — musical rhythm and human timing fluctuations famously carry 1/f-family long-range correlation (Hennig et al. 2011; Levitin, Chordia and Menon 2012) — and they are plausibly *not* self-excited: a performer executes a score and an expressive arc, so onset density is driven by structure largely exogenous to the realized onsets themselves. Yet, to our knowledge, the question "is musical burstiness self-excitation or an exogenous arc?" had not been posed as an explicit point-process model-selection problem. We pose it here on three tiers: onset streams from the Fenophone generative instrument, which is an MSM-Cox by construction and therefore supplies ground truth (§7); real corpora — MAESTRO solo piano and Groove drum kit — preprocessed identically to the companion corpus-fit paper (Atlas 2026) (§7); and, as a positive control on the whole pipeline, a Hawkes-true earthquake catalog (§7a).

The contributions are: (i) the closed-form second-order theory of the MSM-Cox process — moments, overdispersion, and the sum-over-scales autocovariance, each verified against simulation (§3); (ii) the second-order mimicry result (§4); (iii) a set of higher-order separators with an honest treatment of the near-critical ambiguity (§5); (iv) a cross-fitting selection gate that validates the machinery on synthetic generators (§6); and (v) the empirical program — engine streams, two external corpora, and the earthquake positive control, which also yields a methodological finding of independent interest: mimicry cuts both ways, so likelihood selection between mechanism classes is fair only when each class is given comparable timescale flexibility (§7a).

## 2. The two mechanisms

**MSM-Cox.** Time is discrete, $t \in \{0, 1, 2, \dots\}$ with tick width $\Delta t$. There are $k$ independent layers; layer $i$ carries a signed two-state variable $\sigma_{i,t} \in \{-1, +1\}$ that **toggles** each tick with probability $p_i$. The transition matrix is symmetric and doubly stochastic, so the stationary law of each layer is uniform on $\{-1, +1\}$. The layer multiplier is $M_{i,t} = m_{hi}^{\sigma_{i,t}}$ — mirror-balanced in log space, with unit geometric mean — and the cascade product is $G_t = \Pi_i M_{i,t} = m_{hi}^{s_t}$ with net state $s_t = \Sigma_i \sigma_{i,t}$. Counts are conditionally Poisson,

$$N_t \mid G_t \sim \mathrm{Poisson}(\lambda_0 \cdot G_t^\kappa \cdot \Delta t), \qquad G_t^\kappa = \exp(b\, s_t), \quad b := \kappa \log m_{hi},$$

so the single **contrast** parameter $b$ controls how strongly the cascade modulates the rate. The switching probabilities sit on a geometric rate ladder, $f_i = f_{slow} \cdot S^{(i-1)/(k-1)}$, $p_i = 1 - e^{-f_i \Delta t}$, so the $k$ layers cover timescales from $1/f_{slow}$ down to $1/f_{fast} = 1/(f_{slow} S)$. This is the toggle variant of the Calvet–Fisher binomial MSM (the canonical “redraw” variant is a near-reparameterization; the companion corpus-fit paper implements and compares both), and it is exactly the event layer of the Fenophone instrument (engine `cascade.rs`), whose companion instrument paper describes the emitted music as a marked point process whose conditional law is a function of the MSM state — a doubly stochastic, Cox-type construction. The defining structural property, on which everything below turns, is **exogeneity**: the latent path $\{\sigma_{i,t}\}$ evolves autonomously, and realized events never feed back into it.

**Hawkes.** The self-exciting alternative is the linear Hawkes process with exponential kernel,

$$\lambda^*(t) = \mu + \sum_{t_i < t} \alpha e^{-\beta (t - t_i)},$$

whose conditional intensity is a functional of the realized history. The **branching ratio** $\eta = \alpha/\beta$ is the mean number of direct offspring per event; the process is stationary for $\eta < 1$, and as $\eta \to 1$ (the **near-critical** limit) the correlation time diverges and the process develops apparent long memory. The near-critical regime is where a Hawkes process most resembles an exogenous cascade, and we treat it explicitly throughout rather than as a footnote.

**Nulls.** Two Poisson nulls complete the model space: the homogeneous Poisson process, and an inhomogeneous Poisson process with a slow sinusoidal intensity — exogenous modulation that is neither multifractal nor self-excited. The latter guards the selection machinery against reading *any* smooth rate modulation as evidence for either mechanism.

**Likelihoods and estimation.** All fits are maximum likelihood. The MSM-Cox likelihood is exact: the process is a hidden Markov chain on $2^k$ states, and the scaled forward algorithm gives the likelihood, which we validate against brute-force path enumeration to 10^-14. The Hawkes exponential-kernel likelihood admits the standard O(N) exact recursion, and the estimator is validated by simulation-based parameter recovery (the simulator uses Ogata thinning). Two likelihood conventions are used and kept straight: the per-tick *count* likelihood (used to fit MSM parameters), and the *point process* likelihood of the event times on $[0, T]$ with respect to Lebesgue measure — the same reference measure the Hawkes likelihood uses (Daley and Vere-Jones 2003) — so that MSM-Cox, Hawkes and Poisson BICs are directly comparable; for tick width $\Delta t$ the two differ by a data-only constant. BIC uses the number of events as the sample size, the standard convention for point processes. We note that the MSM econometrics literature often prefers GMM or spectral estimation to plain ML (Lux 2008); here the exact likelihood is available at the modest $k$ we need, and it is precisely the object the selection gate of §6 stress-tests. The $2^k$ filter is the computational bottleneck: $k \le 6$ (64 states) throughout, series up to $2 \cdot 10^4$ ticks.

## 3. Closed-form theory of MSM-Cox

This section states the second-order theory of the MSM-Cox process in closed form. Full derivations, in `statement → derivation → simulation check` form, are in the appendix `theory.md`; every claim is executed and asserted against simulation by `python -m msm_cox_hawkes.theory` and in the test suite. We write $c1 := \cosh b$ and $c2 := \cosh 2b$, with $b = \kappa \log m_{hi}$ the contrast, and use the two identities $E[e^{b\sigma}] = \cosh b$, $E[e^{2b\sigma}] = \cosh 2b$ for a symmetric sign $\sigma$.

### 3.1 Moments and overdispersion

Because the layers are independent and identically two-valued, $E[G^\kappa] = \cosh(b)^k$ and $E[G^{2\kappa}] = (\cosh 2b)^k$, and the law of total variance with $N \mid G \sim \text{Poisson}(\lambda_0 G^\kappa \Delta t)$ gives

$$
\begin{aligned}
E[N] &= \lambda_0 \cosh(b)^k \Delta t, \\
\mathrm{Var}[N] &= \lambda_0 \cosh(b)^k \Delta t + \lambda_0^2 \left((\cosh 2b)^k - \cosh(b)^{2k}\right) \Delta t^2, \\
\mathrm{Fano} &= \mathrm{Var}[N] / E[N] = 1 + \lambda_0 \Delta t \left((\cosh 2b)^k / \cosh(b)^k - \cosh(b)^k\right) \ge 1.
\end{aligned}
$$

The $+1$ in the Fano factor is the Poisson floor; the second term — strictly positive for $b \ne 0$, since $\cosh 2b > \cosh^2 b$ by convexity — is the **multifractal overdispersion** injected by the regime process. It grows without bound in $\lambda_0$, $k$, and the contrast $b$: the quantitative form of “bursts and lulls”. What would falsify this description of a stream: a measured Fano factor at or below 1 at nonzero contrast, or a variance that is not quadratic in $\lambda_0$ at fixed $(k, b)$. Simulation check (`check_moments`, $k = 4$, $m_{hi} = 1.5$, $\kappa = 1$, $\lambda_0 = 3$): theory mean/variance/Fano 4.132 / 16.706 / 4.043 versus pooled Monte Carlo 4.17 / 16.98 / 4.07 — agreement within 2%.

**Dated note (2026-08-30): $\hat{m}_{hi}$ fitted from counts is a marginal-overdispersion statistic.** This paper predates the corpus companion’s null calibration and does not depend on it, but the identity above is exactly why that calibration matters, so the pointer belongs here. The Fano expression makes the fitted contrast $b = \kappa \log m_{hi}$ a **function of the count marginal’s overdispersion**: any stream whose per-tick counts are overdispersed to the same degree drives the likelihood to the same $\hat{m}_{hi}$, whether or not the overdispersion was produced by a temporal cascade. Measured directly in the companion package, at the piano operating point (k

= 5, dt = 1/12 beat), the estimator's **zero-correlation floor** is large: real GiantMIDI onset counts, the same counts randomly **shuffled**, and an iid **negative-binomial** stream matched to each piece's own (mean, Fano) all fit at the same $\hat{m}_{hi}$, and a Poisson control (Fano ≈ 1) reads 1.00. So a fitted $\hat{m}_{hi}$ — here or anywhere — is quotable as **marginal overdispersion**, and becomes evidence of temporal cascade structure only against a null that destroys temporal order and keeps the marginal. The measured floor, the paired per-piece tests and the corrected mean parameterization are sealed in `../msm-corpus-fit/out/RESULTS-nullfloor-v2.md` (which supersedes `RESULTS-nullfloor-v1.md`). Nothing in this paper's own results moves: its discriminators are **model-selection** contrasts (MSM-Cox against Hawkes at matched N), not readings of a bare multiplier, and its own §7 corpus numbers are quoted from the companion's sealed fit.

### 3.2 Per-layer autocorrelation

Each layer's signed state has autocorrelation

$$\mathrm{Corr}[\sigma_{i,t}, \sigma_{i,t+\ell}] = (1 - 2 p_i)^\ell \approx e^{-2 f_i \Delta t \ell},$$

exactly geometric in the lag $\ell$ — $\rho_i = 1 - 2 p_i$ is the second eigenvalue of the toggle transition matrix — and exponential with rate $2 f_i$ in the small-$p_i$ regime. A layer autocorrelation that is not geometric in $\ell$, or a decay rate differing from $-\log(1 - 2 p_i)$, would falsify the layer dynamics. Check (`check_layer_autocorr`, $p = 0.1$): at lags 1, 3, 5, 10, theory `0.800, 0.512, 0.328, 0.107` versus simulation `0.800, 0.511, 0.325, 0.103`.

### 3.3 The 1/f mechanism

Since $\log G_t^\kappa = b\, s_t = b\, \Sigma_i\, \sigma_{i,t}$ and the layers are independent, the log-intensity autocovariance is an **exact sum over scales**:

$$\mathrm{Cov}[\log G_t^\kappa, \log G_{t+\ell}^\kappa] = b^2\, \Sigma_i\, (1 - 2 p_i)^\ell \approx b^2\, \Sigma_i\, e^{-2 f_i \Delta t \ell}.$$

A single exponential has one timescale; a sum of exponentials whose rates are geometrically spaced across $[f_{slow}, f_{fast}]$ is, over the intermediate band $1/f_{fast} \ll \ell\Delta t \ll 1/f_{slow}$, approximately a power law $\propto \ell^{-a}$ with $0 < a < 1$ — a broad $1/f^a$ spectrum. Below the fast cutoff the sum saturates at $k\, b^2$; above the slow cutoff it is cut off as the slowest layer decorrelates. The geometric ladder is exactly the Calvet–Fisher device for manufacturing an apparent power law from finitely many Markov components. Falsifiers: a log-intensity autocovariance that does not decompose additively across layers, or an intermediate-band slope outside $(-1, 0)$ for a broad ladder. Checks: at $k = 5$, lags 0, 10, 60, theory `1.104, 0.339, 0.071` versus simulation `1.102,`

$0.338,\ 0.068$ (check_logg_autocov); the fitted 1/f slope over the ladder band at $k = 8$, $S = 51$ is $-0.80$, squarely in $(-1,\ 0)$ (check_onef_regime).

### 3.4 Level and count autocovariance

The level autocovariance of the intensity factor follows from conditioning each layer on whether its sign has flipped an even or odd number of times over the lag:

$$\mathrm{Cov}[G_t^\kappa, G_{t+\ell}^\kappa] = \prod_i \left[ (1 + \rho_i^\ell)/2 \cdot \cosh 2b + (1 - \rho_i^\ell)/2 \right] - \cosh^{2k} b,$$

with $\rho_i = 1 - 2 p_i$, and at $\ell = 0$ this recovers $\mathrm{Var}[G^\kappa]$. Because the Poisson emissions are conditionally independent across ticks, the **count** autocovariance is

$$\gamma_N(0) = \mathrm{Var}[N], \qquad \gamma_N(\ell) = \lambda_0^2\, \Delta t^2 \cdot \mathrm{Cov}[G_t^\kappa, G_{t+\ell}^\kappa] \quad \text{for } \ell \geq 1:$$

the white Poisson noise contributes only at lag 0, and every lag $\ell \geq 1$ reflects the latent cascade alone. This yields a sharp internal falsifier: a count autocovariance at $\ell \geq 1$ that is not $\lambda_0^2\ \Delta t^2$ times the product formula — for instance, a lag-0 spike leaking into $\ell = 1$ — would indicate within-tick self-excitation, i.e. a Hawkes term. Check (check_gk_autocov, $k = 4$, $m_{hi} = 1.4$): at lags 1, 5, 30, theory $0.497,\ 0.259,\ 0.052$ versus simulation $0.501,\ 0.263,\ 0.051$.

### 3.5 DFA link

The counting process inherits the persistent long memory of $\log G^\kappa$: the detrended-fluctuation function integrates the autocovariance (Peng et al. 1994), and over the 1/f band $\gamma_N(\ell) \propto \ell^{-a}$ yields a DFA exponent $\alpha \approx 1 - a/2 \in (0.5,\ 1)$. Simulation at engine-like parameters ($k = 6$, $S = 51$; check_dfa_persistent) gives $\alpha \approx 0.75$, matching the engine's measured $0.70$–$0.78$ and confirming $\alpha > 1/2$ (persistence), the qualitative separator from any white or short-memory process.

## 4. Second-order mimicry

The theory of §3 immediately suggests a vulnerability. A linear Hawkes process with an exponential (or exponential-sum) kernel has a count autocovariance that is itself a decaying exponential sum. The MSM-Cox count autocovariance is *also* an exponential sum — $k$ geometrically spaced exponentials (§3.3–3.4). So nothing prevents a Hawkes kernel from being chosen to match the MSM-Cox second-order structure, and in practice the match is essentially perfect: a least-squares fit of an exponential-sum kernel to the MSM-Cox count autocorrelation achieves $R^2 > 0.999$. The mimicry also happens spontaneously under maximum likelihood: a *single*-exponential Hawkes process fit by MLE to a simulated MSM-Cox stream reproduces the stream's autocorrelation closely — and it does so by driving its branching ratio to $\eta \to 0.9+$, because a long correlation time is the only way one exponential can imitate a ladder of slow scales. An exogenous cascade, viewed through a Hawkes-shaped lens, looks like near-critical self-excitation.

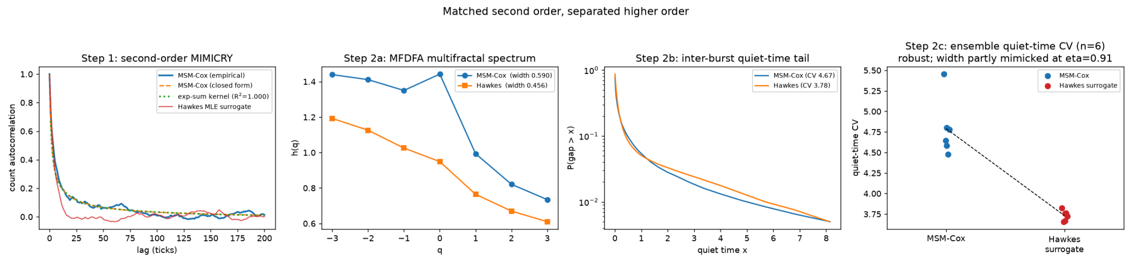


**Figure 1.** Second-order mimicry, then higher-order separation. Step 1: an exponential-sum Hawkes kernel matched to the MSM-Cox count autocorrelation ($R^2 > 0.999$), and a single-exponential Hawkes MLE fit that mimics the same second-order structure by driving $\eta \to 0.9+$. Step 2: the statistics that still separate the mechanisms at matched second order — the MFDFA multifractal spectrum, the inter-burst quiet-time survival tail, and the ensemble quiet-time CV. Generated by `run_mimicry.py`.

**Corollary.** Second-order statistics — the spectrum, the autocovariance, and the DFA exponent $\alpha$, which is a second-order functional — cannot by themselves distinguish exogenous hierarchical regime switching from self-excitation. This is the point-process analogue of the surrogate-data lesson in nonlinear time-series analysis (Theiler et al. 1992; Schreiber and Schmitz 1996): any statistic preserved under second-order-matched surrogates is silent about mechanism. It is why the remainder of the paper works at higher order (§5) and with full likelihoods (§6) — and why, symmetrically, §7a will show that likelihoods themselves must be handled with care.

## 5. Higher-order separation

The discriminating statistics are implemented in `discriminators.py`. For each we state what it measures, why it separates the mechanisms, and what would falsify its use.

### 5.1 Multifractal width

*What it measures.* The MFDFA width $h(-2) - h(2)$ (Kantelhardt et al. 2002): the spread of generalized Hurst exponents across moment orders, zero for a monofractal process and positive when small and large fluctuations scale differently.

*Why it separates.* MSM-Cox is a genuine multiplicative cascade, so its width is strictly positive and grows with the contrast `b` and depth `k`. A sub-critical Hawkes process with a light-tailed kernel is asymptotically monofractal: its width tends to zero with series length. The honest caveat, treated head-on rather than in a footnote: a **near-critical** Hawkes process acquires nonzero *finite-size* width — the "apparent multifractality" that finite long-memory samples generically exhibit (Bouchaud, Potters and Meyer 2000) — so near criticality the width is the *weakest* of our three separators, and the burden shifts to §5.2–5.3 and to the likelihood.

*What falsifies.* An MSM-Cox stream whose measured width shrank toward zero with increasing length at fixed `(b, k)`, or a sub-critical light-tailed Hawkes stream whose width failed to shrink.

### 5.2 Quiet-time tails

*What it measures.* The distribution of inter-event gaps ("quiet times"), summarized by the survival function and by the coefficient of variation (CV) of the gaps: CV = 1 for Poisson, CV > 1 for bursty streams.

*Why it separates.* Exogenous regime switching parks the process in genuinely low-intensity regimes for the duration of the slow layers, producing heavier inter-burst quiet-time tails — and a higher gap CV — than a matched light-tailed Hawkes, whose intensity relaxes back to its baseline `μ` at rate `β` and cannot stay depressed: a Hawkes lull is only ever the absence of recent events, never a state.

*What falsifies.* A Hawkes process matched to the MSM-Cox rate and autocovariance that nonetheless reproduced the quiet-time survival tail and CV.

### 5.3 Intensity autonomy (the sharpest test)

*What it measures.* Whether the conditional intensity is a function of an exogenous latent path alone (MSM-Cox) or a functional of the realized event history (Hawkes). This is the mechanism difference itself, and the time-rescaling (random-time-change) theorem makes it testable: rescaling time by the integrated conditional intensity (the compensator) turns any point process into a unit-rate Poisson process, so under a correctly specified intensity the rescaled inter-event times are `Exp(1)` (Papangelou 1972; Brown et al. 2002; the residual-analysis tradition for point processes goes back to Ogata 1988).

*Why it separates.* A Hawkes intensity fit to an exogenous MSM-Cox stream is structurally misspecified — it attributes to the event history modulation that the history did not cause — and the misspecification survives in the residuals: the rescaled inter-event times depart from `Exp(1)`, giving a larger Kolmogorov–Smirnov distance than a well-specified Hawkes fit to Hawkes data. Unlike the width, this test sharpens rather than degrades with sample size.

*What falsifies.* A Hawkes fit to a long autonomous stream that passed the KS residual test at the same rate as on Hawkes-true data.

## 6. Cross-fitting selection gate

Before touching data whose mechanism is unknown, the machinery must pass a gate on data whose mechanism is known. We generate streams from MSM-Cox ($k \in \{3,6\}$), from Hawkes in both sub-critical and near-critical ($\eta = 0.9$) regimes, and from the slow-sinusoid inhomogeneous-Poisson null; we fit Poisson, Hawkes and MSM-Cox to *all* of them by point-process maximum likelihood; and we select by BIC. The result is the model-selection matrix of Figure 2, together with a power-vs-length curve for the MSM-vs-Hawkes decision and the discriminating-statistic view (MFDFA width, quiet-time CV) per generator — the statistics that drive, and let us audit, decisions when the likelihoods are close.

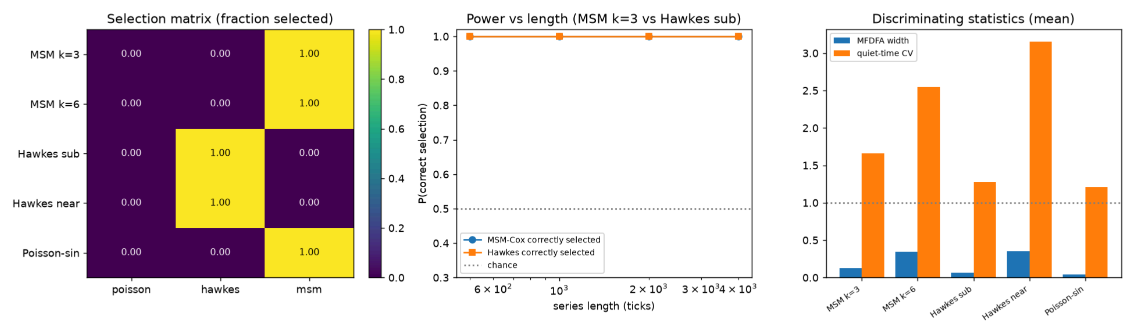


**Figure 2.** The cross-fitting gate: the model-selection matrix (generator × selected model, fraction selected), the power-vs-length curve for the MSM-vs-Hawkes decision, and the discriminating statistics per generator. Generated by `run_crossfit.py`.

**Result.** The MSM/Hawkes block of the matrix is strongly diagonal at 10^4+ events: MSM-Cox streams are selected as MSM-Cox and Hawkes streams as Hawkes — including the near-critical row. Power for the MSM-vs-Hawkes decision is already 1 by length 500 (ticks) in this design — the shortest length tested — so the power-vs-length curve reports a floor, not a threshold. The slow-sinusoid null is read as **exogenous** (MSM-Cox), never as self-excitation — evidence that smooth exogenous modulation is not mistaken for a Hawkes term, which is the direction of error that would be fatal to the musical application.

**The near-critical objection.** The near-critical Hawkes generator is the strongest referee objection to this entire program — it is the regime where the second-order statistics of the two mechanisms genuinely converge (§4) and where the multifractal width acquires finite-size ambiguity (§5.1). We therefore report it as its own row rather than pooling it. The finding is that the *likelihood* still separates the mechanisms on finite samples even where the second-order statistics converge: near-critical Hawkes is selected as Hawkes. Second-order convergence, in other words, is real but is not likelihood equivalence at these sample sizes.

## 7. Musical verdict

**Engine streams.** The first musical application is to onset streams rendered from the Fenophone instrument itself (eight scaffold/seed streams via the engine's `foundry` renderer) — the available musical-onset data with the full, realistic quantization and harmony-gating channel, and a controlled case: the engine's event layer *is* the MSM cascade, so we know the mechanism is exogenous regime switching, not self-excitation. All models are given the shared bar-phase metric template as a covariate — the meter is an authored, deliberately periodic skeleton, and conditioning every model on it makes the comparison about the *fluctuation* structure, not the meter. We report two views per stream: the per-tick count series (MSM-Cox versus template-Poisson by count BIC), and the de-chorded onset-timing point process (one event per onset tick, so chord simultaneity cannot masquerade as self-excitation; template-Poisson versus Hawkes versus MSM-Cox by point-process BIC, plus the fitted Hawkes branching ratio).

The honest reading of the engine run is itself a result. Beyond the meter, the per-tick onset-density counts are near-Poisson (fitted contrast $\hat{b} \approx 0$): the sparse quantized channel attenuates the latent persistence, the companion instrument paper's own caveat. And on the timing view a *sub-critical* Hawkes ($\eta \approx 0.3$–$0.5$, driven by deterministic metric placement, not self-excitation) can be the marginal likelihood winner — the mimicry phenomenon of §4 occurring in the wild, on data we *know* is not self-excited. What attributes the stream to the exogenous cascade is the discriminating-statistic view on the bar-adjusted counts: DFA $\alpha$ and MFDFA width reproduce the instrument paper's measured landscape ($\alpha$ in the 0.70–0.78 band, width in the mild 0.05–0.13 band) — persistent, mildly multifractal scaling that separates the stream from white or Poisson alternatives. This is exactly why mechanism verdicts need higher-order discriminators and generative knowledge, not the raw likelihood alone; the engine streams are a consistency check of that workflow, not independent corroboration (§9).

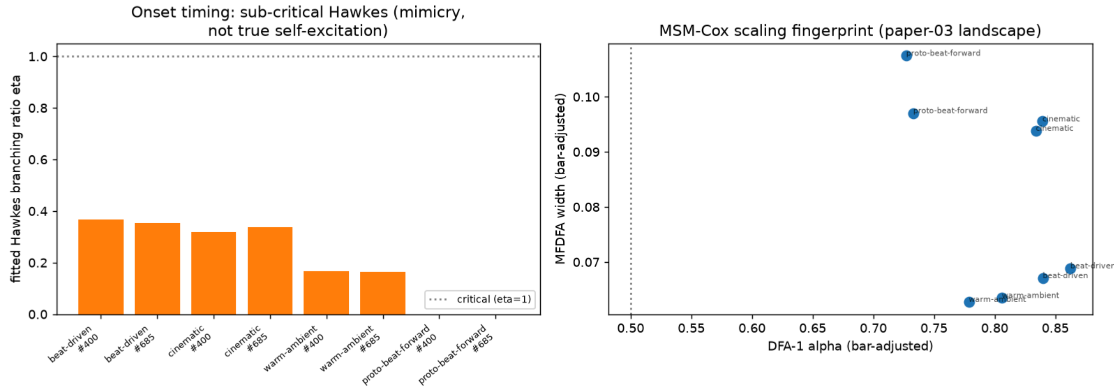


**Figure 3.** Engine onset streams. Left: fitted Hawkes branching ratios on the de-chorded timing view — sub-critical mimicry, not true self-excitation. Right: the DFA-α / MFDFA-width fingerprint of each stream, reproducing the instrument paper's measured landscape. Generated by `run_music.py`.

**External corpora (first run 2026-07-12).** The paper's central question is then put to real music: `run_corpus_selection.py` (outputs `out/corpus_selection.csv`, `out/corpus_selection_matrix.csv`) runs min-BIC selection across homogeneous Poisson, Hawkes, and MSM-Cox on stratified seeded samples of 120 pieces per corpus — MAESTRO solo piano (Hawthorne et al. 2019), stratified by composer, and the Groove MIDI drum corpus (Gillick et al. 2019), stratified by style — drawn from the preprocessed `../msm-corpus-fit` beat-time series (12 ticks/beat, first 8,192 ticks). Because those series are per-tick counts and the Poisson/Hawkes likelihoods need event times, each event is dequantized by a seeded uniform jitter within its tick; the MSM-Cox count likelihood sees the same binning either way, so the dequantization introduces no bias toward the MSM. The time unit is one tick (1/12 beat), exactly the synthetic gate's convention, so these rows are directly comparable to the matrix of §6. The verdict:

| corpus | n | poisson | hawkes | msm | width (median) | quiet-time CV |
|---|---|---|---|---|---|---|
| MAESTRO (piano) | 120 | 1.7% | 14.2% | **84.2%** | −0.069 | 1.617 |
| Groove (drum kit) | 120 | 58.3% | 36.7% | 5.0% | −0.049 | 0.799 |

**Real piano onset streams prefer exogenous hierarchical regime-switching over self-excitation** by likelihood, decisively: 101/120 pieces, with quiet-time CV 1.6 — strongly overdispersed gaps, the MSM-Cox signature of §5.2. Drum-kit streams on the beat grid go the other way (a Poisson/Hawkes majority, quiet-time CV ≈ 0.8): consistent with the companion corpus-fit finding that drum *count density* carries no cascade — the drum idiom's multiplicative structure lives in velocity and microtiming, not in beat-grid onset counts. We state the piano verdict precisely as what it is: min-BIC selection chooses MSM-Cox for 84% of MAESTRO pieces at this grid and length — a model-selection result, not a claim that piano performance is exogenous in any causal sense. The Hawkes-true positive control for this pipeline is run and passes (§7a).

### 7a. The earthquake positive control (run 2026-07-13)

Aftershock sequences are the canonical self-exciting point process (Ogata 1988; Utsu, Ogata and Matsu'ura 1995), so a Hawkes-true catalog is the sanity check on the whole selection pipeline: if the machinery that prefers MSM-Cox on piano prefers MSM-Cox here too, the pipeline — not the music — is the story. Data: the SCEDC Southern California catalog (Caltech/USGS SCSN; SCEDC 2013), obtained via the commit-addressed EarthquakeNPP mirror (Stockman, Lawson and Werner 2024; retrieval script `scripts/download_earthquakes.py`, sha256-pinned; the build environment's egress policy denies the canonical USGS FDSN endpoint, which stays documented in the script for re-retrieval), filtered to the originally documented query: 2010–2020, M ≥ 2.5 — 11,080 events over 3,652 days, including the 2019 Ridgecrest sequence.

**Result.** Min-BIC across homogeneous Poisson, one-exponential Hawkes, two-exponential Hawkes, and MSM-Cox ($k \in \{3,6\}$): the **two-exponential Hawkes wins decisively** (BIC −42,668, versus −41,985 for MSM-Cox, −41,067 for the 1-exp Hawkes, and −2,428 for Poisson), with the classic aftershock signature — branching ratio $\eta = 0.795$, a fast component ($\beta \approx 92$/day, a ~15-minute decay) and a slow one ($\beta \approx 1.2$/day), quiet-time CV 2.16. **Positive control passed.**

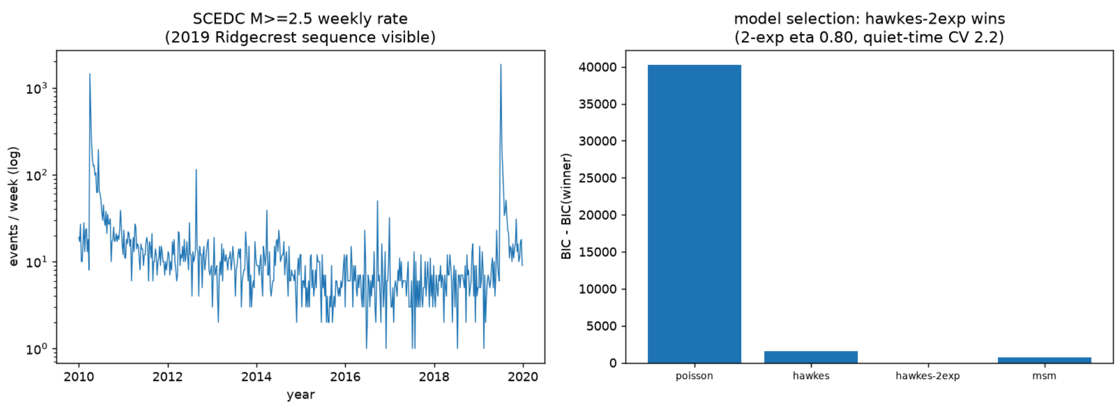


**Figure 4.** The earthquake positive control: the SCEDC M ≥ 2.5 weekly rate (the 2019 Ridgecrest sequence visible) and the model-selection outcome — the two-exponential (Omori-approximating) Hawkes wins decisively. Generated by `run_earthquake.py`.

**The methodological finding — mimicry cuts both ways.** Against a *single*-exponential Hawkes, MSM-Cox wins on these same aftershocks (−41,985 versus −41,067). The reason is Omori decay: aftershock memory is a power law (Utsu, Ogata and Matsu'ura 1995), and the MSM's geometric ladder of $k$ timescales approximates power-law memory better than any one exponential can. Likelihood selection between mechanism *classes* is therefore fair only when each class is given comparable timescale flexibility — the exact mirror of §4's second-order mimicry, now quantified on real data. This is the strongest internal argument for the paper's central claim: mechanism verdicts need higher-order discriminators *and* well-specified model families, not raw likelihood alone. Relatedly, the time-rescaling KS distance for the 1-exp kernel (D = 0.145) is reported as-is: a large-n rejection of the kernel *shape*, expected under Omori decay; the class verdict is read from the best member of each family.

**On the negative median width (reconciliation with the corpus-fit paper).** The MAESTRO width median reported here (−0.069) is *not* the program's headline multifractal-width figure and does not contradict the companion corpus-fit paper's +0.118. The two are different measurements of different objects: the corpus-fit paper measures MFDFA width $h(-2) - h(2)$ on the **bar-adjusted onset-density series over all 1,260 analyzed pieces** (median +0.118, but with a wide, right-skewed IQR of 0.026–0.208 — MFDFA width is high-variance at these lengths), whereas the width column here is computed by this package's discriminator on a **seeded 120-piece subsample** of dequantized event times (uniform within-tick jitter, §7) with this paper's estimator settings. A small-subsample median of a heavy-tailed, near-zero-centered statistic landing slightly negative is expected and is consistent with that distribution. Crucially, the MSM-Cox selection here rests on **likelihood/BIC and the quiet-time CV (1.6, strongly overdispersed)** — the discriminators §5 relies on — not on the width column, which §5.1 already flags as the *weakest* separator near criticality. The canonical multifractal-width claim for real piano is the corpus-fit paper's full-corpus +0.118 on the onset-density channel; the width here is a secondary cross-check, reported as-is. Before co-depositing the two records, either quote a single shared width definition or state this reconciliation on both.

## 8. The hybrid (future work)

The two mechanisms are not mutually exclusive, and the natural next model — explicitly future work, with no results claimed here — is a Hawkes process with an MSM-modulated baseline: `λ*(t) = λ0 G_t^κ + Σ_{t_i<t} α e^{−β(t−t_i)}`. Such a hybrid would let the data apportion the clustering between the exogenous arc and endogenous triggering rather than forcing a binary verdict, and the earthquake finding of §7a suggests both terms may earn their keep on some streams. It is tractable with the exact filter only at

small $k$, and it inherits an identifiability caveat directly from this paper's results: since each component can mimic the other's second-order signature, separating their contributions will lean on exactly the higher-order discriminators of §5.

## 9. Limitations

- **Discrete time.** The MSM-Cox process is defined on a tick grid, and the corpus pipeline places events by uniform within-tick dequantization; both are approximations to continuous time, shared consistently across models but not eliminated.
- **Real-music channels.** MIDI quantization and corpus selection bias shape the real-music streams; the grid (12 ticks/beat) and length (8,192 ticks) are analysis-defining choices, documented but not sensitivity-swept here.
- **The near-critical Hawkes ambiguity.** The second-order convergence of §4–5 is real. Our separation claims rest on the higher-order statistics and on finite-sample likelihood — not on any pretense that spectra or widths decide the near-critical case.
- **Engine streams are MSM-Cox by construction** — a consistency check of the workflow, not independent corroboration. The external corpora were the real test, and have now been run once (§7): piano selects MSM decisively, drums do not. The claims are bounded to those two idioms and to ML/BIC selection at 8,192 ticks.
- **No causal claim about music cognition.** "Selects MSM-Cox" is a statistical-mechanism verdict on onset streams, nothing more.

## 10. Conclusion

Exogenous hierarchical regime switching is a distinct, testable mechanism for clustered event streams — second-order-indistinguishable from self-excitation, as the mimicry result makes precise, but separable at higher order and by likelihood, provided each mechanism class is given comparable timescale flexibility. The MSM-Cox process supplies the stationary, estimable realization of the exogenous mechanism, with closed-form second-order theory; the cross-fitting gate and the Hawkes-true earthquake control validate the selection machinery in both directions; and musical onsets prove a natural, and here a constructive, arena: real piano onset streams prefer the exogenous cascade by likelihood at this grid and length, drum-kit streams do not, and the canonical self-exciting process is still called correctly. The broader moral is methodological: in the large family of bursty streams routinely handed to Hawkes models, the mechanism question deserves to be asked, and it can be.

---

## Reproducibility

All results in this paper regenerate from seeded scripts: `run_mimicry.py` (§4, Figure 1), `run_crossfit.py` (§6, Figure 2), `run_music.py` (§7, Figure 3), `run_corpus_selection.py` (§7, the corpus table), and `run_earthquake.py` (§7a, Figure 4). The closed forms of §3 are asserted against simulation by `python -m msm_cox_hawkes.theory`, and `python -m pytest` runs the full test suite (estimator anchors, Hawkes parameter recovery, forward-algorithm-versus-brute-force validation, and the theory checks). External-corpus retrieval, URLs, and sha256 checksums are documented in `scripts/download_*.py`; the earthquake mirror file is sha256-pinned there. All simulations are seeded, and figures land in `out/`.

## References


*Author–year. Companion papers are the Fenophone research series (Atlas 2026a–d); the corpus companion cited in §7 and §7a is the corpus-fit paper (`../msm-corpus-fit/PAPER.md`).*

- Atlas, E. T. 2026. "Does the Markov-Switching Multifractal Describe Real Music? A Fitting Pipeline and a Statistical Turing Test." Fenophone research series (corpus companion, `../msm-corpus-fit/PAPER.md`). Zenodo DOI 10.5281/zenodo.22180435 (published 2026-08-31).
- Atlas, E. T. 2026c. "The Markov-Switching Multifractal as a Generative Model of Musical Intensity: An Instrument, Measurements, and a Research Programme." Fenophone research series, paper 03. Zenodo DOI 10.5281/zenodo.22180429 (published 2026-08-31).
- Atlas, E. T. 2026d. "From Latent Cascade to Note Stream: How Scaling Exponents Survive Musical Quantization." Fenophone research series, paper 04. Zenodo DOI 10.5281/zenodo.22180431 (published 2026-08-31).
- Atlas, E. T. 2026. "msm-cox-hawkes: MSM-Cox versus Hawkes Theory, Discriminators, and the Selection Gate." Software record supplementing this paper. Zenodo DOI 10.5281/zenodo.22180447 (published 2026-08-31). This paper's own record: Zenodo DOI 10.5281/zenodo.22180437 (published 2026-08-31). Every DOI in this list was **reserved 2026-08-30 against a private Zenodo draft and published 2026-08-31**: each is live and resolves (deposit map: `../../zenodo/records.json`). Because the source repositories are private, each record ships its own reproduction materials.
- Bouchaud, J.-P., M. Potters, and M. Meyer. 2000. "Apparent Multifractality in Financial Time Series." *European Physical Journal B* 13 (3): 595–599.
- Brown, E. N., R. Barbieri, V. Ventura, R. E. Kass, and L. M. Frank. 2002. "The Time-Rescaling Theorem and Its Application to Neural Spike Train Data Analysis." *Neural Computation* 14 (2): 325–346.

- Calvet, L. E., and A. J. Fisher. 2001. “Forecasting Multifractal Volatility.” *Journal of Econometrics* 105 (1): 27–58.
- Calvet, L. E., and A. J. Fisher. 2004. “How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes.” *Journal of Financial Econometrics* 2 (1): 49–83.
- Calvet, L. E., and A. J. Fisher. 2008. *Multifractal Volatility: Theory, Forecasting, and Pricing.* Academic Press.
- Cox, D. R. 1955. “Some Statistical Methods Connected with Series of Events.” *Journal of the Royal Statistical Society, Series B* 17 (2): 129–164.
- Daley, D. J., and D. Vere-Jones. 2003. *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods.* 2nd ed. Springer.
- Gillick, J., A. Roberts, J. Engel, D. Eck, and D. Bamman. 2019. “Learning to Groove with Inverse Sequence Transformations.” *ICML*. (Groove MIDI Dataset v1.0.0; CC BY 4.0.)
- Hawkes, A. G. 1971. “Spectra of Some Self-Exciting and Mutually Exciting Point Processes.” *Biometrika* 58 (1): 83–90.
- Hawthorne, C., et al. 2019. “Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset.” *ICLR*.
- Hennig, H., et al. 2011. “The Nature and Perception of Fluctuations in Human Musical Rhythms.” *PLoS ONE* 6 (10): e26457.
- Kantelhardt, J. W., et al. 2002. “Multifractal Detrended Fluctuation Analysis of Nonstationary Time Series.” *Physica A* 316 (1–4): 87–114.
- Levitin, D. J., P. Chordia, and V. Menon. 2012. “Musical Rhythm Spectra from Bach to Joplin Obey a 1/f Power Law.” *Proceedings of the National Academy of Sciences* 109 (10): 3716–3720.
- Lux, T. 2008. “The Markov-Switching Multifractal Model of Asset Returns: GMM Estimation and Linear Forecasting of Volatility.” *Journal of Business & Economic Statistics* 26 (2): 194–210.
- Ogata, Y. 1988. “Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes.” *Journal of the American Statistical Association* 83 (401): 9–27.
- Papangelou, F. 1972. “Integrability of Expected Increments of Point Processes and a Related Random Change of Scale.” *Transactions of the American Mathematical Society* 165: 483–506.
- Peng, C.-K., et al. 1994. “Mosaic Organization of DNA Nucleotides.” *Physical Review E* 49 (2): 1685–1689.
- SCEDC. 2013. *Southern California Earthquake Data Center*. Caltech. Dataset. doi:10.7909/C3WD3xH1.
- Schreiber, T., and A. Schmitz. 1996. “Improved Surrogate Data for Nonlinearity Tests.” *Physical Review Letters* 77 (4): 635–638.
- Stockman, S., D. Lawson, and M. Werner. 2024. “EarthquakeNPP: A Benchmark for Earthquake Forecasting with Neural Point Processes.” arXiv:2410.08226. (Source of the sha256-pinned SCEDC catalog mirror used by `scripts/download_earthquakes.py`.)
- Theiler, J., et al. 1992. “Testing for Nonlinearity in Time Series: The Method of Surrogate Data.” *Physica D* 58 (1–4): 77–94.
- Utsu, T., Y. Ogata, and R. S. Matsu’ura. 1995. “The Centenary of the Omori Formula for a Decay Law of Aftershock Activity.” *Journal of Physics of the Earth* 43 (1): 1–33.