DECAY-WEIGHTED SURVIVABILITY ESTIMATORS: A NONPARAMETRIC CLASS OF REGIME DETECTORS WITH BOUNDED MEMORY AND AN ASYMPTOTIC TRUST DIFFUSION LIMIT

Anonymous authors
Paper under double-blind review

ABSTRACT

Tracking regime shifts in non-stationary time series is difficult because standard filtering and changepoint detection algorithms typically discard inactive hypotheses to remain computationally efficient. When the environment shifts abruptly, or reverts to a regime seen before, these models are forced to reconstruct state estimates from limited recent observations. We introduce **Decay-Weighted Survivability Estimators (DWSE)**, a class of nonparametric regime detectors that instead retain inactive hypotheses through a decay-plus-floor trust recurrence: rather than deleting a hypothesis, its trust decays toward a small but nonzero floor, allowing rapid reactivation if the regime it describes recurs. This guarantee is distinct in kind from forgetting-factor results in adaptive filtering, which bound convergence of a parameter estimate; DWSE instead bounds the dynamics of hypothesis selection itself. We instantiate the class in three ways **DWSE-Mean** and **DWSE-Vol**, targeted to shifts in the first and second moments respectively, and **DWSE-Rank**, an omnibus detector sensitive to distributional change more broadly and prove finite-sample bounds on hypothesis forgetting time (T_{forget}), reactivation time (T_{recover}), and worst-case detection delay (T_{detect}), alongside a non-asymptotic concentration guarantee. Under stable regime conditions, we further show the centered trust process satisfies a functional central limit theorem, converging weakly to a scaled Brownian motion, giving a principled basis for threshold calibration.On Bitcoin and non-stationary climate data, all three DWSE instances achieve strong recall on labelled regime shifts, performing comparably to or better than classical baselines such as CUSUM, EWMA control charts, directional change, Kalman filtering, and particle filtering, all with the same fixed parameters across domains.

1 INTRODUCTION

Learning systems in non-stationary regimes face a fundamental challenge: when the environment shifts, the correct hypothesis may have already been discarded long before the shift occurs. Classical particle filters prune low-weight particles via resampling (Gordon et al., 1993), killing the very paths that become correct again after a regime reversion. Online learning algorithms track a shrinking active set of experts (Herbster & Warmuth, 1998), and adaptive-regret methods reweight based on recent loss (Hazan & Seshadhri, 2007). These approaches share a common limitation: they update reactively based on observed loss, with no mechanism to preserve a hypothesis through the interval where it is locally wrong but may later become right again. Furthermore, dominant frameworks including Kalman filters, GARCH, and related state-space approaches require Gaussian or conditionally Gaussian distributional assumptions, assumptions that are systematically violated in financial returns, cryptocurrency markets, and climate data, all of which exhibit fat tails, skewness, and volatility clustering. DWSE is fully distribution-free, requiring only a bounded compatibility score, with no Gaussian assumption at any stage.

We propose **Decay-Weighted Survivability Estimators (DWSE)**, built on one structural departure:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

A hypothesis's trust should decay under disconfirming evidence, but never reach zero.

DWSE is not a forecasting system. It tracks the survivability of each hypothesis under a decay-plus-floor recurrence,

$$T_t(h) = \max\big(\gamma\, T_{t-1}(h) + f(x_t, h),\ \varepsilon\big),$$

where $f(x_t, h) \in [0, 1]$ is any bounded compatibility score. Trust in an incompatible hypothesis decays but never reaches zero; when the environment reverts to a regime that hypothesis describes well, its trust reactivates from the floor rather than rebuilding from nothing. The framework is reproducible in any non-stationary environment, and we instantiate it in three concrete detectors, with further extensions planned in subsequent work.

Main contributions:

Framework: (1) decay-weighted survivability as a trust primitive for hypothesis tracking, distinct in kind from the forgetting-factor results of adaptive filtering such as SIFt-RLS (Lai & Bernstein, 2024), which bound convergence of a parameter estimate under decayed least squares, whereas DWSE bounds the dynamics of *hypothesis selection* among a fixed finite set; (2) fully distribution-free, requiring only a bounded compatibility score, in contrast to the Gaussian assumptions underlying Kalman filters and GARCH.

Theorems: (3) a finite-sample forgetting/reactivation bound, $T_{\text{forget}} \le \log(\varepsilon(1-\gamma))/\log(\gamma)$, with a bounded T_{recover} giving worst-case detection delay $T_{\text{detect}} \le T_{\text{forget}} + T_{\text{recover}}$, the first quantitative guarantee of this type; (4) a finite-sample concentration guarantee showing the true hypothesis's trust dominates all others with high probability after a bounded burn-in, under a floor-corrected concentration argument; (5) an asymptotic functional limit, showing the centered trust process converges weakly to a scaled Brownian motion under stationarity, giving a principled basis for threshold calibration.

Empirical validation:we evaluate all three DWSE instances (DWSE-Mean, DWSE-Vol, DWSE-Rank) against classical baselines (CUSUM, Kalman filtering, particle filtering, EWMA,directional change) on Bitcoin and non-stationary climate series, and report a systematic negative results study of variants that did not improve on the reported methods.

2 Related Work

2.1 Classical Change-Point Detection

Classical change-point methods, including CUSUM-type tests (Page, 1954) and their generalizations (Basseville & Nikiforov, 1993), as well as Bayesian online change-point detection, decide that a regime shift has occurred once sufficient post-shift evidence accumulates against the null. They detect change, but they do not maintain a population of never-deleted hypotheses: once a shift is declared, the pre-shift model is typically dropped rather than retained at reduced trust. Consequently these methods have no mechanism, and no corresponding bound, for how quickly a previously-active hypothesis can be reinstated if the environment reverts. DWSE is not a change-point detector in this sense; it is a hypothesis-trust process for which change-point-like behavior (crossing a threshold) is a derived quantity, and reactivation after reversion is a first-class guarantee rather than an afterthought.

2.2 Sequential Monte Carlo and Particle Filters

Classical particle filters (Gordon et al., 1993) maintain a weighted population of particles, resampled at each step according to observation likelihood; low-weight particles are discarded outright. Weight collapse and particle impoverishment are well-documented failure modes of this scheme. Recent differentiable particle filter methods (Brady et al., 2024; Li et al., 2023) and resampling-free variants (Boopathy et al., 2024) address specific instances of this problem, but resampling (or an approximation to it) remains a core operation, so a hypothesis that becomes locally incorrect is still driven toward extinction. DWSE uses a floor instead of resampling: trust in a hypothesis shrinks under disconfirming evidence but is bounded below by $\varepsilon > 0$, so no hypothesis is ever fully deleted. This preserves old hypotheses at bounded cost and gives a formal, finite-sample recovery bound rather than an empirical robustness improvement.

2.3 Online Learning with Experts and Adaptive Regret

The tracking-the-best-expert framework (Herbster & Warmuth, 1998) and adaptive-regret methods such as Follow-the-Leading-History (Hazan & Seshadhri, 2007) allow a learner's active hypothesis set to shift over time, and more recent work (Zhao et al., 2025) gives efficient algorithms for the non-stationary online learning setting more broadly. These methods, including sleeping-experts variants, reweight hypotheses based on observed loss, and in the adversarial or worst-case settings they target, guarantees are stated in terms of cumulative regret rather than the state of any individual hypothesis. DWSE's trust recurrence is closer to a survival/selection process than a loss-based weight update: rather than bounding regret against the best hypothesis in hindsight, we bound how long an individual hypothesis takes to decay to its floor and how long it takes to recover once the environment reverts.

2.4 Forgetting-Factor and Regularized Filtering

Recursive least squares with a forgetting factor, and directional variants such as SIFt-RLS (Lai & Bernstein, 2024), use an exponential-decay mechanism formally similar to DWSE's recurrence: older information is down-weighted geometrically so that the estimator can track time-varying parameters. Floor-regularized particle filters and mixture-of-experts models have likewise used minimum weights to prevent full collapse of diversity. This is the closest prior work to DWSE in mechanism, and we are explicit about the distinction: SIFt-RLS and related forgetting-factor methods bound the convergence of a *parameter estimate* under decayed least squares, in a fixed-parameter or slowly time-varying setting. DWSE bounds the dynamics of *hypothesis selection* among a fixed finite set, a discrete, combinatorial object rather than a continuous parameter vector. The decay-plus-floor structure is shared; the guaranteed quantity is not.

2.5 Robust Statistics for the Compatibility Score

DWSE's compatibility score $f(x_t, h)$ is deliberately left unspecified at the axiomatic level, but our instantiations (DWSE-Mean, DWSE-Vol, DWSE-Rank)draw on standard tools from robust statistics for constructing compatibility thresholds that are insensitive to outliers and do not require distributional assumptions (Huber, 1981; Rousseeuw & Croux, 1993). This justifies the nonparametric compatibility threshold used throughout, while keeping the trust-decay mechanism itself entirely independent of the specific compatibility score chosen.

3 DWSE Framework

3.1 Formal Setting and Axioms

We consider a discrete-time sequence of observations $x_1, x_2, \ldots$ from an unknown, non-stationary process. At every time step, a finite set of hypotheses $\mathcal{H} = \{h_1, \ldots, h_N\}$ is maintained. Each hypothesis h carries a trust score $T_t(h) \in [\varepsilon, +\infty)$ that represents the accumulated empirical survivability of h under the observed sequence.

Assumption 1 (DWSE Axioms). For all time steps $t \geq 1$:

(A1) **No deletion.** The hypothesis set $\mathcal{H}$ is fixed; no hypothesis is ever removed.

(A2) **Decay-plus-floor trust update.** Each trust score obeys

$$T_t(h) = \max\big(\gamma\, T_{t-1}(h) + f(x_t, h),\ \varepsilon\big), \tag{1}$$

where $\gamma \in (0, 1)$ is a decay factor and $\varepsilon > 0$ is a global trust floor.

(A3) **Bounded compatibility.** The compatibility score $f(x_t, h)$ is a measurable function taking values in $[0, 1]$.

The floor ε is the defining feature of DWSE. It ensures that trust in a currently incompatible hypothesis decays toward, but never reaches, zero. This is what permits fast reactivation when the environment reverts to a previously observed regime.

3.2 ADAPTIVE OBSERVATION WINDOW

The compatibility score f is typically constructed from a local dispersion estimate. In our instantiations we use a rolling Median Absolute Deviation (MAD) computed over a window of length W:

$$s(t) = \text{median}\left\{|o_{t-i} - \text{median}(o_{t-W:t})| : i = 0, \ldots, W\right\}, \tag{2}$$

where o_t is the observation transformed as appropriate for the instance (e.g., raw return, absolute return, or rank score). To avoid manual tuning, we select W online by minimizing the coefficient of variation of $s(t)$ over a trailing evaluation window. This is a heuristic motivated by the observation that, within a stationary segment, the MAD estimate stabilizes as W grows, while mixing across a regime boundary inflates the MAD and increases its temporal variability. The adaptive window procedure is not claimed as a theorem, but it removes the main manually set parameter in practice.

3.3 SOFT COMPATIBILITY SCORES

Axiom (A3) only requires $f(x_t, h) \in [0, 1]$. All theoretical guarantees below remain valid if the compatibility score is continuous, probabilistic, or kernel-based, as long as boundedness is preserved. In the experiments we use indicator-based scores for clarity, but the class is broader.

4 THEORETICAL GUARANTEES

Throughout this section, we fix a single hypothesis h and write $C_t = f(x_t, h)$ for its compatibility score at time t. The results below apply to each hypothesis separately; in the concentration proposition we compare the true hypothesis h^* against the remaining hypotheses.

Theorem 1 (Bounded Forgetting)**.** *Let $T_0(h) \le 1/(1-\gamma)$. If $C_t = 0$ for all $t \ge 1$, then*

$$T_t(h) \le \varepsilon \quad \textit{for all } t \ge T_{\text{forget}} \coloneqq \left\lceil \frac{\log(\varepsilon(1-\gamma))}{\log \gamma} \right\rceil. \tag{3}$$

With the default parameters $\gamma = 0.8$, $\varepsilon = 0.001$, this bound is $T_{\text{forget}} = 39$ steps.

Proof. Under permanent incompatibility, $T_t = \gamma T_{t-1}$ after the first step. Unrolling from $T_0 \le 1/(1-\gamma)$ gives $T_t \le \gamma^t/(1-\gamma)$. Solving $\gamma^t/(1-\gamma) \le \varepsilon$ yields the bound. □

Remark 1. Theorem 1 does not require the exact value of $T_0(h)$; any initialization bounded by $1/(1-\gamma)$ suffices because the decay dominates after a finite burn-in. In practice we initialize $T_0(h) = \varepsilon$, which trivially satisfies this bound.

Theorem 2 (Bounded Reactivation)**.** *Let $T_0(h) = \varepsilon$. Suppose $C_t = 1$ for all $t \ge 1$. Then for any target trust level θ with $\varepsilon < \theta < 1/(1-\gamma)$,*

$$T_t(h) \ge \theta \quad \textit{for all } t \ge T_{\text{recover}}(\theta) \coloneqq \left\lceil \frac{\log\left(\frac{\theta - 1/(1-\gamma)}{\varepsilon - 1/(1-\gamma)}\right)}{\log \gamma} \right\rceil. \tag{4}$$

With the default parameters $\gamma = 0.8$, $\varepsilon = 0.001$, the bound gives $T_{\text{recover}}(3) = 5$.

Proof. When $C_t = 1$ and $T_t \ge \varepsilon$, the recurrence is $T_t = \gamma T_{t-1} + 1$. Unrolling yields $T_t = \gamma^t \varepsilon + (1-\gamma^t)/(1-\gamma)$; solving $T_t \ge \theta$ completes the proof. □

Remark 2. The target θ must satisfy $\varepsilon < \theta < 1/(1-\gamma)$. The upper bound $1/(1-\gamma)$ is the steady-state ceiling under permanent compatibility; if $\theta \ge 1/(1-\gamma)$, no finite recovery time exists.

Corollary 1 (Detection-Delay Bound)**.** *Let h_{old} be the previously dominant hypothesis and h_{new} the hypothesis compatible with the post-shift regime. Suppose at the shift time t_0, $C_t(h_{\text{old}}) = 0$ and $C_t(h_{\text{new}}) = 1$ for all $t > t_0$. Then there exists $t^* \le t_0 + T_{\text{forget}} + T_{\text{recover}}(\theta)$ such that, for all $t \ge t^*$, $T_t(h_{\text{old}}) \le \varepsilon$ and $T_t(h_{\text{new}}) \ge \theta$: any threshold-crossing detection rule fires within $T_{\text{forget}} + T_{\text{recover}}(\theta)$ steps of an idealized instantaneous, permanent shift.*

Proof. Immediate from Theorems 1 and 2, applied to h_{old} and h_{new} respectively, taking $t^* = t_0 + \max(T_{\text{forget}}, T_{\text{recover}}(\theta))$. □

This bound applies to idealized instantaneous and permanent regime shifts. Section 6 evaluates empirical detection delay under noisy transitions rather than assuming this clean-room model.

Proposition 1 (Trust Concentration). Assume a stationary regime with i.i.d. compatibility observations. Let $p_* = \Pr(C_t(h^*) = 1)$ and $q = \max_{h \neq h^*} \Pr(C_t(h) = 1)$, with separation $\Delta = p_* - q > 0$. Define the finite-time error probability

$$\delta_t = N \exp\left(-\frac{(1+\gamma)\big[\Delta(1-\gamma^t) - \varepsilon(1-\gamma)\big]^2}{2(1-\gamma)(1-\gamma^{2t})} \right). \tag{5}$$

Then, for any t such that $\Delta(1-\gamma^t) > \varepsilon(1-\gamma)$,

$$\Pr\big(\forall h \neq h^*,\ T_t(h^*) > T_t(h)\big) \geq 1 - \delta_t.$$

In particular, for any $\delta \in (0,1)$, after a burn-in period $T_0 = O\Big(\frac{\log(N/\delta)}{1-\gamma}\Big)$,

$$\Pr\big(\forall h \neq h^*,\ T_t(h^*) > T_t(h)\big) \geq 1 - \delta \quad \text{for all } t \geq T_0,$$

provided

$$\Delta \geq \varepsilon(1-\gamma) + \sqrt{\frac{2(1-\gamma)\log(N/\delta)}{1+\gamma}}.$$

Proof sketch. Let $h \neq h^*$. Define unfloored surrogate $U_t(h) = \gamma U_{t-1}(h) + C_t$, $U_0 = \varepsilon$. We first show $T_t(h) - U_t(h) \leq \varepsilon$ for all t. Since $U_{t-1}, C_t \geq 0$,

$$T_t(h) - U_t(h) = \max\big(\gamma(T_{t-1} - U_{t-1}),\ \varepsilon - \gamma U_{t-1} - C_t\big) \leq \max\big(\gamma(T_{t-1} - U_{t-1}),\ \varepsilon\big).$$

We claim $T_t(h) - U_t(h) \leq \varepsilon$. At $t = 0$ it is zero. If it holds at $t-1$, then

$$T_t(h) - U_t(h) \leq \max(\gamma\varepsilon, \varepsilon) = \varepsilon.$$

Hence the claim holds by induction. Similarly, $T_t(h^*) \geq U_t(h^*)$ since the floor only raises trust. Therefore $T_t(h^*) > T_t(h)$ follows whenever $U_t(h^*) - U_t(h) > \varepsilon$. The difference $D_t = U_t(h^*) - U_t(h)$ is a weighted sum of independent bounded terms. Applying weighted Hoeffding with mean $\mu_t \geq \Delta(1-\gamma^t)/(1-\gamma)$ and threshold ε, we get

$$\Pr(D_t \leq \varepsilon) \leq \exp\left(-\frac{(1+\gamma)\big[\Delta(1-\gamma^t) - \varepsilon(1-\gamma)\big]^2}{2(1-\gamma)(1-\gamma^{2t})} \right).$$

A union bound over $N-1$ incorrect hypotheses yields the result. The asymptotic condition follows by letting $t \to \infty$. □

Theorem 3 (Functional Central Limit Theorem)**.** *In a stationary regime where the floor is inactive for the dominant hypothesis, define the centered trust $X_t = T_t - \mu$ with $\mu = p/(1-\gamma)$. Then the partial-sum process*

$$S_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nu \rfloor} X_t, \qquad u \in [0,1],$$

converges weakly in the Skorokhod space $D[0,1]$ to a scaled Brownian motion:

$$S_n \Rightarrow \omega B(u),$$

where B is a standard Brownian motion and $\omega^2 = \mathrm{Var}(C)/(1-\gamma)^2$.

Proposition 1 shows dominance over other hypotheses; separately, if the true compatibility probability satisfies $p_* > 0$, then $T_t(h^*)$ concentrates around $p_*/(1-\gamma)$, which for default parameters is much larger than ε. Hence the floor is inactive for h^* with high probability after burn-in.

Proof sketch. Under stationarity, the centered trust obeys the linear AR(1) equation $X_t = \gamma X_{t-1} + \varepsilon_t$, where $\varepsilon_t = C_t - p$ is bounded, zero-mean, and i.i.d. Unrolling gives the linear process representation $X_t = \sum_{j=0}^{\infty} \gamma^j \varepsilon_{t-j}$. The weights $\psi_j = \gamma^j$ satisfy $\sum_{j=0}^{\infty} j|\psi_j| < \infty$, so the Phillips Solo functional central limit theorem for linear processes applies (Theorem 3.1, Phillips & Solo, 1992). This yields weak convergence to Brownian motion with long-run variance $\omega^2 = \mathrm{Var}(\varepsilon)(\sum_j \gamma^j)^2 = \mathrm{Var}(C)/(1-\gamma)^2$. Floor inactivity holds with high probability after the concentration burn-in. □

Remark 3 (Robustness under adversarial contamination). If a fraction η of observations are corrupted by an adversary, the effective separation becomes $\Delta' = p_*(1-\eta) - q - \eta(1-q)$. The same Hoeffding and union-bound argument suggests concentration persists whenever $\Delta' > 0$, with a weakened rate. A formal proof is left to future work.

4.1 Practical Implication for Threshold Calibration

The FCLT provides an asymptotic Gaussian reference distribution for cumulative trust fluctuations under the null of no regime change. In practice, we use fixed z-score thresholds (e.g., $Z = 2$ or 3) on derived signals such as the MAP-trust drop or persistence z-score; the theoretical long-run variance ω^2 justifies this choice and allows, in principle, direct p-value computation.

5 DWSE Instances

The DWSE axioms specify the trust mechanism but leave the observation transformation, hypothesis space, and compatibility score unspecified. In this section we describe three concrete instances used in the experiments. All share the same trust recurrence, floor ε, and decay γ; they differ only in the observable fed into the compatibility function and the derived detection statistic.

5.1 DWSE-Mean: First-Moment Shifts

DWSE-Mean targets changes in the conditional mean of a return series. The observable is the raw return $o_t = r_t$, and the hypothesis space is a finite grid of mean levels $\mathcal{H}_{\text{mean}} = \{\mu_{\min}, \ldots, \mu_{\max}\}$, constructed from the mean and standard deviation of a burn-in period. As with the other instances, compatibility is judged by a binary MAD-band test,

$$f_{\text{mean}}(o_t, h) = \mathbf{1}\{|o_t - \mu_h| \le s(t)\},$$

with $s(t)$ the rolling MAD threshold from Section 3.2. For the detection statistic, we follow the same design as DWSE-Vol rather than introducing a separate mechanism: let $h^\star(t) = \arg\max_{h\in\mathcal{H}} T_t(h)$ denote the dominant mean hypothesis at time t, and take the z-score of the negative trust change, $-\Delta_t = -(T_t(h_t^\star) - T_{t-1}(h_{t-1}^\star))$, relative to a rolling baseline. Using the same MAP-trust-drop construction across both instances keeps the detection statistic tied directly to the trust recurrence, rather than relying on an auxiliary smoothing signal outside the shared framework.

5.2 DWSE-Vol: Second-Moment Shifts

DWSE-Vol targets changes in volatility. The observable is the absolute return

$$o_t = |r_t|,$$

and the hypothesis space is a grid of volatility levels

$$\mathcal{H}_{\text{vol}} = \{\sigma_{\min}, \ldots, \sigma_{\max}\}$$

again initialized from burn-in statistics. The compatibility score is

$$f_{\text{vol}}(o_t, h) = \mathbf{1}\{|o_t - \sigma_h| \le s(t)\}\,.$$

Because volatility has no directional sign, the mean-instance persistence signal is not meaningful. Instead we use the MAP-trust drop: let

$$h^\star(t) = \arg\max_{h\in\mathcal{H}} T_t(h)$$

be the dominant hypothesis at time t. The detection statistic is the z-score of the negative trust change

$$-\Delta_t = -\big(T_t(h_t^\star) - T_{t-1}(h_{t-1}^\star)\big),$$

computed relative to a rolling baseline. A sharp drop in the trust of the previously dominant volatility hypothesis indicates a volatility regime shift.

5.3 DWSE-RANK: OMNIBUS DISTRIBUTIONAL SHIFTS

The third instance is not targeted at a specific moment. Instead, it uses a rolling empirical-rank transformation

$$o_t = \frac{1}{W} \sum_{i=0}^{W-1} \mathbf{1}\{r_{t-i} \le r_t\},$$

so that $o_t \in [0, 1]$ is the fraction of observations in the trailing window that are less than or equal to the current return. The hypothesis space is a grid of rank levels

$$\mathcal{H}_{\text{rank}} = \{\rho_1, \ldots, \rho_N\} \subset [0, 1],$$

and the compatibility score is again a MAD-band test on the rank scale:

$$f_{\text{rank}}(o_t, h) = \mathbf{1}\{|o_t - \rho_h| \le s(t)\}\,.$$

Because the rank transform is sensitive to any distributional change, this instance acts as an omnibus detector rather than a moment-specific one. In the experiments we evaluate the same DWSE-Rank signal against both mean-shift and volatility-shift label sets. Identical recall and false-alarm rates on the two label sets are expected and are not interpreted as two independent detection capabilities; rather, they indicate a single generic sensitivity to distributional change.

6 EXPERIMENTS

6.1 SETUP

We evaluate DWSE on three datasets: Bitcoin hourly returns ($80{,}295$ observations), Pune daily temperature anomalies ($26{,}744$ observations), and additional climate series (Los Angeles, London) for cross-domain validation. All experiments use the same core parameters: decay $\gamma = 0.8$, trust floor $\varepsilon = 0.001$, $N = 100$ hypotheses, and detection threshold $Z = 2.0$. The observation window W is selected adaptively per dataset as described in Section 3.2, with candidate windows appropriate to the sampling frequency (hours for Bitcoin, days for climate). No domain-specific tuning is performed across instances or datasets.

Baselines are tuned fairly: for each baseline, we sweep its primary threshold parameter(s) and select the configuration that maximizes recall subject to a false-alarm rate (FAR) constraint of either FAR $\le 5\%$ or FAR $\le 10\%$. Unless otherwise stated, detection is evaluated within a strictly causal window of length $\max(W, 40)$ steps after the labelled event, ensuring no forward-looking leakage and respecting the worst-case detection-delay bound $T_{\text{detect}} \le T_{\text{forget}} + T_{\text{recover}}$ from Corollary 1.

We also attempted a comparison against Bayesian online changepoint detection (BOCPD; Adams & MacKay, 2007) using a Student-t observation model. Across a hazard-rate sweep spanning two orders of magnitude ($1/10$ to $1/100$) and outlier-robust (winsorized) preprocessing, the model's changepoint probability remained exactly equal to the hazard rate at every time step, independent of the observed data, indicating the run-length posterior failed to concentrate. This is consistent with known difficulties applying conjugate-prior BOCPD to heavy-tailed financial return series. We were unable to obtain a data-responsive BOCPD baseline within the scope of this work and exclude it from the main comparison.

6.2 BITCOIN REGIME DETECTION

Table 1 reports detection performance on the five labelled Bitcoin mean-shift events. DWSE-Mean achieves $4/5$ recall at a false-alarm rate of 1.28%, the lowest FAR among all methods tested under the FAR$\le 5\%$ protocol at comparable recall.

For volatility shifts, Table 2 shows DWSE-Vol obtains $5/5$ recall at FAR 2.26%. The best tuned baseline under FAR$\le 5\%$ is CUSUM on absolute returns with $4/5$ recall and 3.06% FAR; no other method matches DWSE-Vol's recall–FAR trade-off.

Finally, DWSE-Rank uses the rolling empirical rank of returns as the observable and a grid of rank levels as hypotheses. The same signal is evaluated against both mean-shift and volatility-shift label sets; because it is a single generic change detector, the recall and FAR are identical for both label

Method	Best config	Recall	FAR
DWSE-Mean	MAP-trust drop	$4/5$	1.28%
CUSUM	$h = 8$	$4/5$	3.36%
EWMA control chart	λ=0.05, L=1.5	$4/5$	3.51%
Directional Change	thr= 0.005	$4/5$	4.77%
Kalman filter	$z = 3.0$	$1/5$	0.00%
Particle filter	n=50	$4/5$	1.30%
EWMA volatility	λ=0.9, z=3.0 (FAR$\leq 10\%$)	$5/5$	9.31%

Table 1: Bitcoin mean-shift detection under the corrected post-shift scoring window. Baselines are tuned to maximize recall subject to FAR $\leq 5\%$ unless noted otherwise.

Method	Best config	Recall	FAR
DWSE-Vol	MAP-trust drop	$\mathbf{5/5}$	$\mathbf{2.26}\%$
CUSUM (abs)	$h = 8$	$4/5$	3.06%
EWMA control (abs)	λ=0.05, L=1.5	$4/5$	4.65%
Directional Change (abs)	thr= 0.005	$4/5$	4.38%
Kalman filter (abs)	$z = 3.0$	$0/5$	0.00%
Particle filter (abs)	n=50	$2/5$	1.44%
EWMA volatility	λ=0.9, z=3.0 (FAR$\leq 10\%$)	$5/5$	9.31%

Table 2: Bitcoin volatility-shift detection. Baselines tuned as in Table 1.

sets. On Bitcoin, it achieves $5/5$ recall (100%) with FAR 3.72%. We emphasize that this is not evidence of two independent capabilities but rather of a single omnibus sensitivity to distributional change, consistent with the design described in Section 5.3.

6.3 Climate Regime Shifts

Table 3 reports cross-domain performance on daily temperature anomalies, using the identical parameters ($\gamma = 0.8$, $\varepsilon = 0.001$, $Z = 2.0$) selected once on Bitcoin, with no climate-specific retuning unlike the baselines in Tables 1 and 2, which were re-tuned per dataset. Recall here is naturally lower than on Bitcoin, which is expected rather than concerning: regime shifts in temperature anomalies tend to be gradual, and events like the onset of an El Niño episode do not announce themselves the way a market crash does. What matters more in this setting is that the false-alarm rate stays low (1.75–3.29%) across all three locations, showing that the detector correctly reads these series as calm most of the time and only flags when something genuinely shifts.

Dataset	Recall	FAR	Shifts
Pune temperature	67%	2.50%	6
Los Angeles temperature	40%	1.75%	5
London temperature	75%	3.29%	4

Table 3: DWSE climate detection results

6.4 Ablation: The Role of the Trust Floor

We ran an ablation with $\varepsilon = 0$ (no floor) to test the practical necessity of the floor. Across all datasets, removing the floor produced no measurable change in recall or FAR. This is not evidence against the floor's necessity: it reflects that these real regime recurrences happen before complete forgetting, not that the mechanism is inert.The controlled synthetic experiment in Appendix B isolates the regime where the floor is decisive, with absences exceeding T_{forget}, and shows the mechanism's benefit directly.

6.5 NEGATIVE RESULTS

We systematically tested thirteen alternative detection signals derived from or related to the DWSE trust process, including entropy-based signals, permutation entropy stagnation, predictive surprisal, H-trust gated persistence, heavy-tail trust variants, adaptive-K heuristics, and combined trust-pressure signals. With two exceptions, none matched the recall–FAR trade-off of the reported detectors. Two variants, predictive surprisal and magnitude-weighted persistence, matched or slightly exceeded DWSE-Mean's recall on Bitcoin alone; however, a brief cross-domain check found predictive surprisal's advantage does not hold on climate data (recall falling to 40–50% on London and Los Angeles temperature anomalies, versus 75% and 40% for DWSE-Mean/Vol under the same fixed parameters), consistent with it capturing something specific to Bitcoin's return distribution rather than a domain-general notion of regime change. The full list of thirteen variants, with recall and FAR values, appears in Appendix C.

6.6 STATISTICAL UNCERTAINTY AND SENSITIVITY

Because the number of labelled shifts is small(see Appendix E , we report exact binomial 95% confidence intervals for recall: $5/5$ recall gives CI $[0.478, 1.0]$; $6/6$ gives $[0.541, 1.0]$; $3/5$ gives $[0.147, 0.947]$; and $3/4$ gives $[0.194, 0.994]$. The false-alarm rate is computed over tens of thousands of observations, so its standard error is negligible; for example, Bitcoin mean FAR 1.28% over $80{,}295$ points has an approximate 95% CI within a few hundredths of a percent.

We also swept $\gamma \in \{0.7, 0.8, 0.9\}$, $\varepsilon \in \{0.0001, 0.001, 0.01\}$, and $Z \in \{1.5, 2.0, 2.5\}$. Recall and FAR shift meaningfully with the detection threshold Z, as expected for any threshold-based detector, while γ has a comparatively small effect within the tested range; ε produced no measurable change for either instance, since trust rarely approaches the floor during normal operation on this data and its effect is instead isolated directly in Appendix B. Full results are shown in Appendix D.

7 DISCUSSION AND LIMITATIONS

DWSE is a class, not a single detector. The three instances (DWSE-Mean, DWSE-Vol, DWSE-Rank) illustrate that the trust-floor mechanism operates at the axiomatic level and can be instantiated with different observables (moments vs. rank) and different detection statistics (persistence vs. MAP-trust drop). The adaptive observation window is a heuristic, not a theorem; it is motivated by the stable dispersion property and removes manual tuning, but we do not claim optimality.

Our empirical validation is limited to a small number of labelled events, chosen for being hard to call in real time rather than obvious in hindsight; recall estimates are therefore subject to wide confidence intervals and lack independent label verification. We deliberately exclude EUR/USD from the main results because its labelled macroeconomic events did not correspond to clean statistical regime breaks in daily returns; including it would obscure rather than clarify the framework's behavior. The FCLT assumes floor inactivity justified by the concentration result, but a fully finite-sample bound for the floor's inactivity remains open. Finally, the robustness remark is conjectural; a formal adversarial-noise analysis is left to future work.

8 CONCLUSION

We introduced Decay-Weighted Survivability Estimators (DWSE), a nonparametric class of regime detectors defined by three axioms: no hypothesis deletion, a decay-plus-floor trust recurrence, and bounded compatibility. We proved finite-sample bounds on forgetting and reactivation, leading to a worst-case detection-delay guarantee, and derived a floor-corrected concentration bound showing that the true hypothesis dominates after a burn-in period. An asymptotic Brownian functional central limit theorem justifies threshold calibration.Three instances mean, volatility, and rank demonstrate the class's breadth, and empirical results on Bitcoin and climate series show strong recall with controlled false alarms, matching or exceeding tuned classical baselines without domain-specific retuning. DWSE provides a principled, distribution-free primitive for online adaptation under non-stationarity.

AI USE STATEMENT

In this work, we used generative AI tools for prose editing, LaTeX formatting, and as an auxiliary check on the correctness of derived proofs, where AI-suggested corrections were reviewed and verified by the authors before inclusion. We have not used generative AI tools to generate synthetic data, formulate the paper's core theoretical claims, propose research hypotheses, design methodology or experiments, or interpret results; these were conceived and carried out by the authors. Additionally, we used generative AI tools for brainstorming and literature search support. We have reviewed all AI-assisted work, including verifying by hand a proof correction identified with AI assistance. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

All experiments were run with the code and parameter settings described in Section 6 and Appendix D. The datasets are publicly available; links are provided in the supplementary material. The theoretical proofs are fully stated in the main text and Appendix B. We will release an anonymized code repository upon acceptance.

REFERENCES

Ryan Prescott Adams and David J.C. MacKay. Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*, 2007.

M. Basseville and I.V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, 1993.

A. Boopathy et al. Resampling-free particle filters in high-dimensions. In *ICRA*, 2024.

J.J. Brady et al. Regime learning for differentiable particle filters. *arXiv:2405.04865*, 2024.

N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, 140(2):107–113, 1993.

E. Hazan and C. Seshadhri. Adaptive algorithms for online decision problems. Technical Report TR-798-07, Princeton University, 2007.

M. Herbster and M.K. Warmuth. Tracking the best expert. *Machine Learning*, 32(2):151–178, 1998.

Peter J. Huber. *Robust Statistics*. Wiley, 1981.

Brian Lai and Dennis S. Bernstein. SIFt-RLS: Subspace of information forgetting recursive least squares. *arXiv preprint arXiv:2404.10844*, 2024.

W. Li et al. Differentiable bootstrap particle filters for regime-switching models. *arXiv:2302.10319*, 2023.

E.S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

Peter C. B. Phillips and Victor Solo. Asymptotics for linear processes. *Annals of Statistics*, 20(2): 971–1001, 1992.

Peter J. Rousseeuw and Christophe Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424):1273–1283, 1993.

P. Zhao et al. Efficient methods for non-stationary online learning. *Journal of Machine Learning Research*, 26(208):1–66, 2025.

A FULL PROOFS

A.1 PROOF OF PROPOSITION 1 (TRUST CONCENTRATION)

Proof. **Step 1: Floor correction via unfloored surrogate.** Define the unfloored process $U_t(h) = \gamma U_{t-1}(h) + C_t$, $U_0(h) = \varepsilon$. We first show that for all t,

$$T_t(h) - U_t(h) \le \varepsilon.$$

At $t = 0$, both equal ε. Assume the inequality holds at $t-1$. Then

$$\begin{aligned} T_t(h) - U_t(h) &= \max\big(\gamma T_{t-1}(h) + C_t, \varepsilon\big) - \big(\gamma U_{t-1}(h) + C_t\big) \\ &= \max\big(\gamma(T_{t-1} - U_{t-1}),\ \varepsilon - \gamma U_{t-1} - C_t\big) \\ &\le \max(\gamma\varepsilon, \varepsilon) = \varepsilon, \end{aligned}$$

since $U_{t-1} \ge 0$ and $C_t \ge 0$. By induction the claim holds for all t. Similarly, $T_t(h^*) \ge U_t(h^*)$, since the floor only raises trust. Therefore

$$T_t(h^*) > T_t(h) \quad \text{whenever} \quad U_t(h^*) - U_t(h) > \varepsilon.$$

Step 2: Expectation of the trust gap. Unrolling the surrogate,

$$U_t(h) = \gamma^t \varepsilon + \sum_{i=1}^{t} \gamma^{t-i} C_i.$$

The difference between the true and an incorrect hypothesis is

$$D_t(h) = U_t(h^*) - U_t(h) = \sum_{i=1}^{t} \gamma^{t-i}\big(C_i(h^*) - C_i(h)\big) = \sum_{i=1}^{t} w_i Z_i,$$

where $w_i = \gamma^{t-i}$ are deterministic weights and $Z_i = C_i(h^*) - C_i(h) \in [-1, 1]$ are i.i.d. Since $\mathbb{E}[Z_i] = p_* - p_h \ge \Delta$,

$$\mu_t = \mathbb{E}[D_t(h)] \ge \Delta \sum_{i=1}^{t} \gamma^{t-i} = \Delta\, \frac{1-\gamma^t}{1-\gamma}.$$

Step 3: Weighted Hoeffding bound. For independent Z_i with range $b_i - a_i = 2$, weighted Hoeffding gives

$$\Pr\left(\sum_i w_i Z_i - \mu_t \le -r\right) \le \exp\left(-\frac{2r^2}{\sum_i w_i^2 (b_i - a_i)^2}\right).$$

Setting $r = \mu_t - \varepsilon$, the denominator is

$$\sum_{i=1}^{t} w_i^2 \cdot 4 = 4 \sum_{i=1}^{t} \gamma^{2(t-i)} = \frac{4(1-\gamma^{2t})}{(1-\gamma)(1+\gamma)}.$$

Hence

$$\Pr(D_t(h) \le \varepsilon) \le \exp\left(-\frac{(1-\gamma)(1+\gamma)(\mu_t - \varepsilon)^2}{2(1-\gamma^{2t})}\right).$$

Substituting $\mu_t \ge \Delta(1-\gamma^t)/(1-\gamma)$ gives

$$\Pr(D_t(h) \le \varepsilon) \le \exp\left(-\frac{(1+\gamma)\big[\Delta(1-\gamma^t) - \varepsilon(1-\gamma)\big]^2}{2(1-\gamma)(1-\gamma^{2t})}\right).$$

Step 4: Union bound and finite burn-in. With $N-1$ incorrect hypotheses, a union bound gives

$$\Pr\big(\exists\, h \ne h^* : T_t(h^*) \le T_t(h)\big) \le N \exp\left(-\frac{(1+\gamma)\big[\Delta(1-\gamma^t) - \varepsilon(1-\gamma)\big]^2}{2(1-\gamma)(1-\gamma^{2t})}\right) =: \delta_t.$$

Taking the complement gives $\Pr(\forall h \ne h^*,\ T_t(h^*) > T_t(h)) \ge 1 - \delta_t$.

Fix any finite burn-in $T_0 \geq 1$ and set $x_0 := 1 - \gamma^{T_0} \in (0, 1)$. For all $t \geq T_0$, $(1 - \gamma^t) \geq x_0$ and $(1 - \gamma^{2t}) \leq 1$, so

$$\delta_t \leq N \exp\left(-\frac{(1+\gamma)\big[\Delta x_0 - \varepsilon(1-\gamma)\big]^2}{2(1-\gamma)} \right)$$

uniformly for all $t \geq T_0$. Requiring the right-hand side to be at most δ gives the finite, non-asymptotic requirement

$$\Delta \geq \frac{\varepsilon(1-\gamma) + \sqrt{2(1-\gamma)\log(N/\delta)/(1+\gamma)}}{1-\gamma^{T_0}}.$$

For example, with $\gamma = 0.8$ and $T_0 = 5$, $x_0 = 1 - 0.8^5 \approx 0.672$, giving a concrete burn-in of 5 steps at a correspondingly larger Δ requirement than the asymptotic bound. As $T_0 \to \infty$, $x_0 \to 1$ and this recovers the limiting bound as a special case. □

A.2 Proof of Theorem 3 (Functional Central Limit Theorem)

Preconditions. Stationary regime, i.i.d. compatibility observations for the dominant hypothesis h^*, with $p_* = \Pr(C_t(h^*) = 1)$ satisfying

$$p_* > \varepsilon(1-\gamma).$$

Proof. **Step 1: Linear process representation.** Under the theorem's assumption the floor never binds for h^*, so $T_t = \gamma T_{t-1} + C_t$. Centering with $\epsilon_t = C_t - p_*$ gives $X_t = \gamma X_{t-1} + \epsilon_t$, which unrolls to the linear process

$$X_t = \sum_{j=0}^{\infty} \gamma^j \epsilon_{t-j}, \qquad \psi_j = \gamma^j.$$

Step 2: Verification of Phillips–Solo (1992) conditions.

1. *Innovations i.i.d. with finite variance:* $\epsilon_t = C_t - p_*$ are i.i.d., bounded in $[-p_*, 1 - p_*]$, hence of finite variance.
2. *Summability:* $\sum_{j=0}^{\infty} j|\psi_j| = \sum_{j=0}^{\infty} j\gamma^j = \dfrac{\gamma}{(1-\gamma)^2} < \infty$ for $0 < \gamma < 1$.

Both conditions hold, so the FCLT for linear processes applies.

Step 3: Long-run variance.

$$\omega^2 = \mathrm{Var}(\epsilon_t) \left(\sum_{j=0}^{\infty} \psi_j \right)^2 = \mathrm{Var}(C) \left(\frac{1}{1-\gamma} \right)^2 = \frac{\mathrm{Var}(C)}{(1-\gamma)^2}.$$

Hence $S_n \Rightarrow \omega B(u)$.

Step 4: Formal floor-inactivity probability bound. Let $U_t = \gamma U_{t-1} + C_t$, $U_0 = \varepsilon$, be the unfloored surrogate for h^*. If $U_t > \varepsilon$ then $T_t \geq U_t > \varepsilon$, so the floor is inactive at t. The mean of U_t is

$$\mu_U = \gamma^t \varepsilon + p_* \frac{1-\gamma^t}{1-\gamma} \xrightarrow{t\to\infty} \frac{p_*}{1-\gamma} > \varepsilon,$$

where the final inequality follows from the precondition $p_* > \varepsilon(1-\gamma)$. Choose T_{floor} such that $\mu_U > \varepsilon$ for all $t \geq T_{\text{floor}}$. For such t, Hoeffding's inequality gives

$$\Pr(U_t \leq \varepsilon) \leq \exp\left(-\frac{2(\mu_U - \varepsilon)^2 (1-\gamma^2)}{1-\gamma^{2t}} \right),$$

using $\sum_{i=1}^{t} \gamma^{2(t-i)} = (1-\gamma^{2t})/(1-\gamma^2)$. This bound decays exponentially in t, so for any $\delta_{\text{floor}} > 0$ there exists T_{floor} with $\Pr(U_t \leq \varepsilon) \leq \delta_{\text{floor}}$ for all $t \geq T_{\text{floor}}$.

Combining with the concentration bound δ_t from Proposition 1, for $t \geq \max(T_{\text{floor}}, T_0)$, with probability at least $1 - \delta_{\text{floor}} - \delta_t$ the dominant hypothesis's trust remains above ε *and* dominates all incorrect hypotheses. On this event the trust recurrence is exactly linear, and the FCLT of Steps 1-3 applies. □

B SYNTHETIC RECOVERY EXPERIMENT

To directly test the necessity of the trust floor, we construct a synthetic compatibility sequence in which a hypothesis h is compatible, then becomes incompatible for a fixed duration exceeding $T_{\text{forget}} = 39$ under the default parameters $\gamma = 0.8$, $\varepsilon = 0.001$, and then becomes compatible again, simulating a regime that disappears and later recurs. We repeat this for incompatible durations of 50, 100, 150, 200, 300, and 500 steps.

We compare two schemes on each sequence:

- **DWSE with floor**, following the recurrence in Equation (1), where trust decays toward but never reaches ε;
- **Delete-on-death**, modeling classical particle-filter and change-point behavior directly: the instant trust falls below a small kill threshold, the hypothesis is removed from the tracked set entirely and its trust is fixed at zero, with no further updates, until an external mechanism reinstates it.

Table 4 reports whether each scheme survives the incompatible period and, if so, how many steps it takes to recover to a target trust level $\theta = 3$ once the regime reverts. The floor condition survives every tested duration and recovers in a small, finite number of steps in every case. The delete-on-death condition never survives: once its trust is zeroed, it remains at zero for the rest of the simulation, including after the old regime returns, since deletion is permanent by construction and nothing in the simulation triggers its reintroduction.

This is the practical consequence of Axiom (A1) (no deletion) combined with the reactivation bound of Theorem 2: because DWSE never removes a hypothesis, it is always available to reactivate, and does so within the bound $T_{\text{recover}}(\theta)$ once compatibility resumes. A classical delete-on-death policy, by contrast, removes the hypothesis permanently and cannot recover without an explicit external reinstatement step, regardless of how long the simulation runs. This complements the ablation in Section 6, where no real dataset contained an absence long enough for the floor to be the limiting factor; here, by construction, every tested absence exceeds that threshold.

Incompatible duration	T_{forget} exceeded?	Floor survived?	Delete survived?	Floor recovery (steps)
50	Yes	Yes	No	7
100	Yes	Yes	No	2
150	Yes	Yes	No	1
200	Yes	Yes	No	20
300	Yes	Yes	No	13
500	Yes	Yes	No	1

Table 4: Survival and recovery under an incompatible period exceeding $T_{\text{forget}} = 39$, for a hypothesis tracked with the floor versus a delete-on-death policy. The floor condition always survives and recovers in finite time; the delete-on-death condition never recovers within the simulation, since deletion is permanent and no external reinstatement mechanism is present. Recovery time under the floor varies with the exact post-reversion compatibility sequence, not with absence duration, which is expected since Theorem 2's bound does not depend on how long the absence was.

C NEGATIVE RESULTS

We systematically tested thirteen alternative detection signals derived from or related to the DWSE trust process. With two exceptions discussed in Section 6, none matched the recall,FAR trade-off of the sign-persistence (DWSE-Mean) or MAP-trust-drop (DWSE-Vol) detectors reported in the main text. Table 5 summarizes recall and false-alarm rate for each variant on the Bitcoin mean-shift label set.

Detector variant	Recall	FAR	Notes
H-trust-gated persistence (MAD clip)	$4/5$	5.30%	Higher FAR than DWSE-Mean at comparable recall
H-trust-gated persistence (heavy-tail)	$3/5$	5.20%	Lower recall, no FAR benefit
Permutation entropy (order 3, daily)	$1/5$	16.00%	Unstable at daily resolution
Permutation entropy (order 4, daily)	$2/5$	13.62%	High FAR
Permutation entropy (order 5, daily)	$2/5$	12.63%	High FAR
Predictive surprisal	$5/5$	1.91%	Strong on Bitcoin; does not generalize (see below)
Raw entropy drift	$1/5$	2.91%	Low recall
Autocorrelation drift	$1/5$	4.23%	Low recall
Adaptive-K via θ-instability	$0/5$	0.22%	Rarely fires
H-trust-modulated threshold	$5/5$	5.59%	Ties DWSE-Mean on recall, worse FAR
Combined adaptive-K + H-trust	$2/5$	14.09%	High FAR, no recall benefit
θ-instability gating	$0/5$	0.00%	Gate essentially never opens
Magnitude-weighted persistence	$5/5$	5.34%	Ties DWSE-Mean; same generalization concern as predictive surprisal

Table 5: Negative results: alternative detection signals tested on the Bitcoin mean-shift label set under identical post-shift scoring. Two variants (predictive surprisal, magnitude-weighted persistence) match or exceed DWSE-Mean's recall on Bitcoin alone but do not generalize as well to climate data, as discussed in Section 6.

D SENSITIVITY ANALYSIS

We swept $\gamma \in \{0.7, 0.8, 0.9\}$, $\varepsilon \in \{0.0001, 0.001, 0.01\}$, and $Z \in \{1.5, 2.0, 2.5\}$, holding all other parameters at their default values. Across every tested combination, ε had no observable effect on recall or FAR for either instance: trust rarely approaches the floor during normal operation on this dataset, so its effect is only visible in the extended-absence regime tested directly in Appendix B. Table 6 therefore reports the more informative dependence on γ and Z, at $\varepsilon = 0.001$ fixed.

γ	Z	Recall (Mean)	FAR (Mean)	Recall (Vol)	FAR (Vol)
0.7	1.5	5/5	12.04%	5/5	12.85%
0.7	2.0	5/5	3.38%	5/5	4.72%
0.7	2.5	0/5	0.11%	2/5	0.20%
0.8	1.5	5/5	10.03%	5/5	11.86%
0.8	2.0	**4/5**	**1.28%**	**5/5**	**2.26%**
0.8	2.5	0/5	0.07%	0/5	0.08%
0.9	1.5	5/5	6.24%	5/5	8.78%
0.9	2.0	3/5	0.28%	4/5	0.60%
0.9	2.5	0/5	0.06%	0/5	0.06%

Table 6: Sensitivity of DWSE-Mean and DWSE-Vol to γ and Z on Bitcoin label sets, at $\varepsilon = 0.001$. Both recall and FAR shift substantially with the detection threshold Z for both instances, as expected for any threshold-based detector; DWSE-Mean also shows pronounced sensitivity to γ within the tested range, while DWSE-Vol is comparatively less sensitive to γ at fixed Z. ε is omitted as it produced no measurable change (see text).

E LABELLED EVENT SETS

Tables 7 and 8 list the ground-truth regime-shift events used to compute the recall figures reported in Tables 1 through 3.

Events were not chosen for how large they look in hindsight. They were chosen because at the time, nobody could confidently say which way things were about to move, and getting an early signal there would actually have mattered. That is the situation a regime detector is supposed to be useful in, so we tried to test it there rather than on the easy, obvious cases that are only clear after the fact. All dates were fixed by the author from public price and weather records before any detector was run, using the same criterion across all three datasets.

E.1 BITCOIN

Bitcoin was chosen as the primary financial series because it is one of the most volatile widely-traded assets, making it a demanding testbed for regime detection under heavy-tailed, non-stationary conditions. Three of the five Bitcoin events appear in both the mean-shift and volatility-shift label sets, since large market dislocations tend to move both the mean and the variance at once. Because of this overlap, the recall numbers in Tables 1 and 2 should be read as related to each other rather than as two fully separate pieces of evidence.

Table 7: Labelled Bitcoin events used for the recall figures in Tables 1 and 2.

Date	Event	Mean-shift	Volatility-shift
2013-04-10	2013 first bubble peak	Yes	Yes
2013-11-01	2013 Q4 volatility rise		Yes
2013-12-17	2013 all-time high	Yes	
2017-12-01	2017 ATH volatility spike		Yes
2017-12-17	2017 all-time high	Yes	
2020-03-12	COVID crash	Yes	Yes
2020-10-01	2020 bull market start	Yes	Yes

E.2 CLIMATE

Climate events were chosen as documented starts of major climate-oscillation phases, such as El Niño and La Niña episodes, or as extreme, well-recorded temperature anomalies such as heatwaves, severe winters, and storm-driven cold snaps, one series at a time.

Table 8: Labelled climate shift events used for the recall figures in Table 3.

Location	Data source	Date	Event
Pune	Pune daily temperature records, 1951–2024	1976-06	1976–77 climate shift
		1983-01	1982–83 El Niño
		1998-05	1998 super El Niño
		2009-06	2009–10 El Niño
		2015-03	2015–16 El Niño
		2020-01	2020 recent warming
Los Angeles	NOAA records, 2015-07-08 to 2020-05-12	2016-01	2015–16 strong El Niño
		2017-01	2017 La Niña cool shift
		2018-06	2018 summer heatwave
		2019-03	2019 weak El Niño
		2020-03	2020 COVID-period shift
London	London daily weather records	1987-10	1987 North Atlantic cold and storm shift
		2003-08	2003 European heatwave
		2010-12	2010 severe UK winter
		2018-02	2018 Beast from the East

Note: with only 5 Bitcoin events per label set and 4 to 6 climate events per location, these recall estimates carry wide confidence intervals. See Section 6.6 for the exact binomial values.