FLARE: FACTOR-LOADING-BREAK-TRIGGERED ADAPTIVE READOUT FOR HIGH-DIMENSIONAL TIME SERIES UNDER REGIME SHIFT

Anonymous authors
Paper under double-blind review

ABSTRACT

Time-series forecasters can degrade sharply when the shared low-rank structure of a high-dimensional panel changes across regimes. Yet common remedies such as per-series instance normalization primarily target marginal distribution shifts and can therefore miss changes in the joint dependence structure. To address this challenge, we formulate regime shifts as low-rank structural breaks, covering both span-changing shifts in the loading space and span-preserving changes in factor strength. Building on this formulation, we introduce FLARE, an online detector that uses a whitened amortized representation and a single operator statistic to jointly capture subspace geometry and factor strength. Each detected break then triggers lightweight readout-only adaptation while the forecasting backbone remains frozen. Theoretically, we establish a weak-factor detectability boundary and a detection-delay guarantee, and show that whitening lowers the detection threshold relative to window-wise PCA. Empirically, FLARE detects weak-factor breaks missed by classical factor detectors, localizes a real market regime shift, and matches the forecasting performance of always-on test-time adaptation at a fraction of the update cost.

1 INTRODUCTION

High-dimensional multivariate forecasting underlies macroeconomic nowcasting (McCracken & Ng, 2016; Stock & Watson, 2002a; Forni et al., 2000), portfolio risk (Fan et al., 2015; 2013), and brain-state monitoring (Hutchison et al., 2013; Gonzalez-Castillo & Bandettini, 2018), where dozens to thousands of series are driven by a small number of common factors (Bai, 2003; Stock & Watson, 2002b). Yet this cross-sectional dependence structure can change over time. Such changes may arise from shifts in loading directions, factor strengths, or effective rank, inducing structural breaks in the underlying factor model (Barigozzi et al., 2018; 2026; Bai & Han, 2016; Barigozzi & Trapani, 2020). As a result, a deep forecaster (Nie et al., 2023; Liu et al., 2024) trained under one regime may rely on a stale representation of this structure, degrading its predictions even when individual series remain well behaved. Because these shifts are joint rather than necessarily marginal, they can be difficult to detect before prediction quality deteriorates.

A common response to non-stationarity is normalization. RevIN (Kim et al., 2022) rescales each series using its recent statistics, primarily targeting marginal distribution shifts. Test-time adaptation goes further by updating lightweight model components online to track distributional changes (Kim et al., 2025; Wang et al., 2026). Despite these advances, neither approach explicitly detects changes in the shared low-rank dependence structure. In particular, marginal normalization can miss structural shifts that leave individual series largely unchanged, whereas always-on adaptation updates the model without distinguishing structural breaks from transient fluctuations. These limitations call for a structure-aware mechanism that detects shifts in the shared low-rank structure and adapts only when such a shift occurs.

Factor change-point detection provides a natural statistical formulation for such structural shifts. It separates the common component from idiosyncratic variation and localizes changes in the loading structure or factor dimension. Early work develops PCA- and QML-based procedures for break localization (Barigozzi et al., 2018; Duan et al., 2023), while subsequent studies characterize differ-

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.

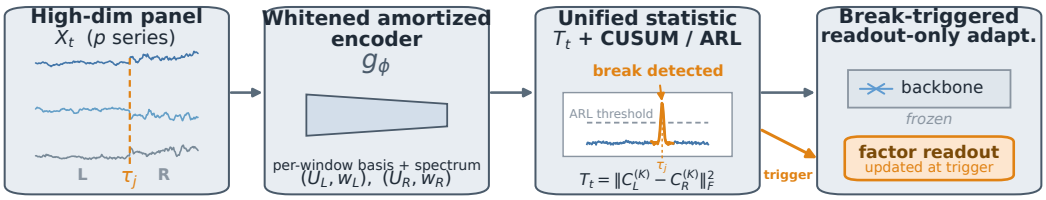


Figure 1: FLARE pipeline: a whitened, amortized encoder tracks the loading space online using two adjacent windows and returns the basis and spectrum of each window. A *single* unified statistic T_t, defined as the squared Frobenius distance between the two rank-truncated whitened-covariance operators, increases at a structural break at τ_j, which is then declared by a self-normalized CUSUM with an ARL-calibrated threshold. Each declared break triggers a readout-only update that refreshes the post-break low-rank representation while keeping the forecasting backbone frozen.

ent types of structural changes (Duan et al., 2026; Koo et al., 2026) and develop likelihood-ratio, moving-sum, and sequential procedures (Bai et al., 2024; Barigozzi et al., 2026; Barigozzi & Trapani, 2020). Within this formulation, existing taxonomies distinguish changes in the loading span from span-preserving changes in factor strength. However, these methods primarily target statistical detection or break localization rather than using detected breaks to trigger forecasting adaptation, and many rely on sufficiently strong factors for reliable eigenspace estimation. These limitations leave factor change-point methods poorly suited to online forecasting, particularly in weak-factor regimes, where structural changes must be detected promptly and translated into adaptation decisions.

To bridge this gap, we propose FLARE (Figure 1), an online detect-then-adapt approach for low-rank structural breaks. In our method, a whitened amortized encoder represents the common low-rank structure in each window, and a single statistic compares the resulting rank-truncated whitened-covariance operators across adjacent windows. This operator representation jointly encodes subspace geometry and factor strength, allowing the statistic to capture both span-changing and span-preserving structural shifts. Specifically, the statistic reduces to a weighted principal-angle distance when the spectrum is unchanged and to an eigenvalue-energy distance when the subspace is fixed, thereby unifying both cases within one calibrated detector.

This weak-factor regime is particularly challenging because structural signals can be difficult to distinguish from idiosyncratic variation. We characterize this difficulty through a factor-strength detectability boundary and a detection-delay guarantee, which quantify how factor strength, break magnitude, window size, and the false-alarm target affect online detection. When this detector signals a break, FLARE re-estimates the post-break basis and updates only a lightweight readout while keeping the forecasting backbone fixed, including for pretrained time-series foundation models (Ansari et al., 2024; Woo et al., 2024). Detection thus serves directly as an adaptation trigger, restricting model updates to detected structural breaks rather than stationary periods.

Overall, FLARE extends factor change-point analysis from break localization to adaptation-aware online forecasting. Our contributions are threefold.

1. We propose an online detector for low-rank structural breaks using a whitened amortized representation and a single operator statistic that captures both span-changing and span-preserving shifts.
2. Weak-factor detection guarantees are established, including a detection-delay bound and detectability boundaries for factor-strength and rank changes. Whitening is further shown to lower the threshold relative to per-window PCA.
3. Detection is coupled with break-triggered readout adaptation while the forecasting backbone remains fixed. Experiments on synthetic and real data support the predicted weak-factor detectability boundary, quantify the accuracy–adaptation-cost trade-off, and localize a real market regime shift.

Notations. For a matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|$ and $\|\boldsymbol{A}\|_F$ denote the spectral and Frobenius norms, respectively, $\boldsymbol{A}^\top$ its transpose, and $\mathrm{col}(\boldsymbol{A})$ its column space. For a vector $\boldsymbol{a}$, $\|\boldsymbol{a}\| = \|\boldsymbol{a}\|_2$ denotes the Euclidean

norm. The identity matrix is denoted by $\boldsymbol{I}$, $\langle \cdot, \cdot \rangle_F$ denotes the Frobenius inner product, and $\circ$ denotes the Hadamard product. For two subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$, $\boldsymbol{\Theta}(\mathcal{S}_1, \mathcal{S}_2)$ denotes their principal-angle matrix. We use $\mathcal{O}(\cdot)$, $\mathcal{O}_P(\cdot)$, $o(\cdot)$, $\asymp$, and $\gg$ in their standard asymptotic senses, and $\delta_{ij} = \mathbf{1}_{\{i=j\}}$ denotes the Kronecker delta, where $\mathbf{1}_{\{\cdot\}}$ is the indicator function.

2 Related Work

Retrospective detection of structural breaks in factor models. A mature literature studies structural breaks in factor models by separating the common component from idiosyncratic variation and detecting changes in the loading structure or factor dimension. Early work primarily addresses break localization: *Factorcpt* (Barigozzi et al., 2018) detects change points after PCA-based estimation of the common component, while Duan et al. (2023) develop a quasi-maximum likelihood (QML) estimator for break locations. Beyond localization, subsequent work characterizes the nature of structural changes, distinguishing rotational from singular loading breaks (Duan et al., 2026), separating loading from variance breaks (Koo et al., 2026), and developing likelihood-ratio and moving-sum procedures with asymptotic guarantees (Bai et al., 2024; Barigozzi et al., 2026). Related work also studies change-point latent factor models under potentially nonstationary factor and noise processes (Liu & Zhang, 2022), while earlier surveys provide a broader treatment of structural changes in high-dimensional factor models (Bai & Han, 2016). However, these retrospective methods are not directly designed for sequential detection, particularly in weak-factor regimes where local eigenspace estimation becomes statistically difficult.

Sequential detection and weak-factor inference. A related literature studies sequential detection in high-dimensional covariance models and inference under weak latent factors. Subspace-CUSUM (Xie et al., 2020) achieves first-order optimal detection of changes in a spiked subspace when the relevant eigenspace can be reliably estimated. Related work on high-dimensional factor models also studies the identification of factor dimension (Onatski, 2009) and the behavior of principal-component estimators under weak factors (Onatski, 2012; Bai & Ng, 2023). As factor strength weakens or idiosyncratic noise becomes heterogeneous, however, spectral separation deteriorates and eigenspace estimation becomes more difficult. In weak-factor regimes, Lam & Yao (2012) study factor-number estimation, while Fan et al. (2024) characterize when weak latent factors remain statistically identifiable. A related spectral line of work studies changes in spike magnitude, with Dörnemann & Paul (2026) deriving a functional limit for the largest eigenvalue and testing such changes above a spectral phase-transition threshold. A remaining challenge is therefore online detection of factor-structure breaks that may involve changes in either subspace geometry or factor strength, especially when the factors are weak.

Normalization and test-time adaptation for forecasting. A separate line of work addresses distribution shift by adapting deployed forecasters at test time. RevIN (Kim et al., 2022) normalizes each series using recent marginal statistics, while non-stationary Transformers (Liu et al., 2022) address non-stationarity through stationarization and de-stationary attention. TAFAS (Kim et al., 2025) and FAC (Wang et al., 2026) continuously update lightweight calibration modules, while COSA (Im & Kwon, 2026) adapts a frozen forecasting backbone through a lightweight output-space adapter. CALIPER (Fujiwara et al., 2026) instead uses detected drift to determine when retraining is needed. Such adaptation strategies can be deployed with modern forecasting backbones, including pretrained time-series models (Nie et al., 2023; Liu et al., 2024; Ansari et al., 2024; Woo et al., 2024). However, existing approaches generally respond to marginal shifts or generic drift rather than explicitly adapting to changes in the shared low-rank structure.

3 FLARE Framework

In this section, we present FLARE, an online detect-then-adapt approach for structural regime shifts in high-dimensional panels. The key idea is to represent each regime by its low-rank common structure, detect structural changes online, and trigger adaptation only upon a detected break. We first formalize the factor model and the identifiable break types. The three components of FLARE are then introduced: a whitened amortized encoder, a unified operator-based detector, and break-triggered readout adaptation. Algorithm 1 summarizes the resulting streaming procedure.

3.1 PROBLEM SETUP

We observe a panel $\boldsymbol{X}_t \in \mathbb{R}^p$ generated by a piecewise-stationary approximate factor model. Let $\tau_1 < \cdots < \tau_m$ denote the m unknown change points, with $\tau_0 = 0$ and $\tau_{m+1} = T$ defining the boundaries of the observation period. On each segment $\tau_{j-1} < t \leq \tau_j$,

$$\boldsymbol{X}_t = \boldsymbol{\Lambda}^{(j)} \boldsymbol{f}_t + \boldsymbol{e}_t, \tag{1}$$

where $\boldsymbol{\Lambda}^{(j)} \in \mathbb{R}^{p \times r_j}$ is the segment-specific loading matrix, $\boldsymbol{f}_t \in \mathbb{R}^{r_j}$ contains the latent common factors, and $\boldsymbol{e}_t \in \mathbb{R}^p$ is an idiosyncratic component that is weakly dependent across time and coordinates and may be heteroskedastic. The common component $\boldsymbol{\Lambda}^{(j)} \boldsymbol{f}_t$ lies in the loading space $\mathcal{S}^{(j)} = \mathrm{col}(\boldsymbol{\Lambda}^{(j)})$. We allow for *weak* factors, with signal energy scaling as $\|\boldsymbol{\Lambda}^{(j)\top} \boldsymbol{\Lambda}^{(j)}\| \asymp p^{\delta}$ for $\delta \in (0, 1]$. Smaller δ corresponds to weaker factors, making reliable recovery more difficult for per-window PCA.

Across regimes, the low-rank common structure may change in either geometry or strength. We define any such identifiable change as a structural break. Following the factor-model taxonomies of Duan et al. (2026) and Koo et al. (2026), we distinguish two classes according to whether the loading space itself changes.

- *span-changing break* — the loading space itself changes, so that $\mathcal{S}^{(j)} \neq \mathcal{S}^{(j+1)}$. Specifically, $\boldsymbol{\Lambda}^{(j+1)} = \boldsymbol{\Lambda}^{(j)} \boldsymbol{Q}$ with $\boldsymbol{Q}$ *rank-deficient* corresponds to a dying factor and hence a shrinking span, whereas $\boldsymbol{\Lambda}^{(j+1)} = \boldsymbol{\Lambda}^{(j)} \boldsymbol{Z} + \boldsymbol{W}$ with $\mathrm{col}(\boldsymbol{W}) \not\subseteq \mathrm{col}(\boldsymbol{\Lambda}^{(j)})$ corresponds to a factor birth or a tilt of the span into a new direction. This is the singular Type-1 break of Duan et al. (2026), corresponding to Koo's $\boldsymbol{W} \neq \boldsymbol{0}$ (Koo et al., 2026). We measure its size by the maximal principal angle $\overline{\theta}_j$ between $\mathcal{S}^{(j)}$ and $\mathcal{S}^{(j+1)}$. A rank change $r_{j+1} \neq r_j$ is the special case in which a direction is added or removed,with detectability governed by whether the born or dying factor exceeds a factor-strength threshold.
- *span-preserving break* — the loading space remains unchanged, so that $\mathcal{S}^{(j+1)} = \mathcal{S}^{(j)}$, with $\boldsymbol{\Lambda}^{(j+1)} = \boldsymbol{\Lambda}^{(j)} \boldsymbol{Q}$ for $\boldsymbol{Q}$ *full-rank*. This is the rotational Type-2 break of Duan et al. (2026), corresponding to Koo's within-span $\boldsymbol{Z}$. Since the loading space is preserved, such a break is invisible to any pure subspace statistic and is detectable only through changes in the second moment, which occur when $\boldsymbol{Q}$ is non-orthogonal or the factors are non-isotropic. When $\boldsymbol{Q}$ is orthogonal and the factor distribution is rotationally invariant, with $\mathrm{Cov}(\boldsymbol{f}_t) = \boldsymbol{I}$, the observed-data distribution remains unchanged, yielding an *identification null* that is statistically indistinguishable from no change.

The loading matrix $\boldsymbol{\Lambda}^{(j)}$ and the factors $\boldsymbol{f}_t$ are identifiable only up to an invertible transformation $\boldsymbol{R}$,

$$\boldsymbol{\Lambda}^{(j)} \boldsymbol{f}_t = (\boldsymbol{\Lambda}^{(j)} \boldsymbol{R})(\boldsymbol{R}^{-1} \boldsymbol{f}_t). \tag{2}$$

Accordingly, structural changes are characterized through identifiable low-rank quantities, including the loading space and the second-order structure of the common component, rather than through a particular parameterization of $\boldsymbol{\Lambda}^{(j)}$.

We target online, break-triggered adaptation of a deep forecaster $F_\theta(\boldsymbol{X}) = h_\psi(\mathrm{backbone}(\boldsymbol{X}); \hat{\boldsymbol{U}})$, where θ denotes the full parameterization of the composite forecaster, the backbone is frozen and may be a time-series foundation model, and ψ parameterizes the lightweight readout h_ψ, $\hat{\boldsymbol{U}}$ is defined in Section 3.2. At each step t, the incoming stream is monitored for changes in the low-rank common structure. Upon detection, only ψ is updated to adapt the forecaster to the current regime, while the backbone remains frozen and no adaptation is performed during stationary periods.

3.2 AMORTIZED LOADING ENCODER

A shared encoder g_ϕ maps a length-h window $\boldsymbol{X}_{t-h:t} \in \mathbb{R}^{p \times h}$ to an orthonormal basis

$$\widehat{\boldsymbol{U}}_t = g_\phi(\boldsymbol{X}_{t-h:t}) \in \mathbb{R}^{p \times K}$$

and the leading eigenvalues of its whitened covariance. The estimated common subspace is $\widehat{\mathcal{S}}_{t-h:t} = \mathrm{col}(\widehat{\boldsymbol{U}}_t)$. The encoder first applies an optional learned whitening step to reduce the effect of heteroskedastic idiosyncratic noise, then extracts a $p \times K$ feature map and orthogonalizes it onto the Stiefel manifold. The working rank K is selected as described in Section 3.4.

The encoder is trained without break labels by minimizing the masked projection loss $\|\boldsymbol{X}_t - \widehat{\boldsymbol{U}}_t\widehat{\boldsymbol{U}}_t^\top \boldsymbol{X}_t\|_2^2$ over windows. The whitening component is trained jointly, and the estimate depends only on incrementally updated window statistics, enabling an $\mathcal{O}(1)$ per-step streaming update. Sharing the encoder across windows supports joint optimization with the detector and improves robustness to heteroskedastic noise. The corresponding estimation condition is given in Assumption 2.

3.3 Online unified-statistic break detector

At each candidate split t, the detector compares a *left* window $(t-h, t]$ and a *right* window $(t, t+h]$. Their whitened covariance estimates have orthonormal bases $\boldsymbol{U}_L, \boldsymbol{U}_R \in \mathbb{R}^{p\times K}$ and leading eigenvalues $\lambda_{L,1} \geq \cdots \geq \lambda_{L,K}$ and $\lambda_{R,1} \geq \cdots \geq \lambda_{R,K}$. The two adjacent windows form a standard moving-sum (MOSUM) contrast around t (Barigozzi et al., 2026). In the online setting, the split at t is evaluated at calendar time $t+h$, once the right window has been observed, ensuring causal detection.

For each window, we form the rank-K truncated whitened-covariance operator

$$\widehat{\boldsymbol{C}}^{(K)} = \sum_{i=1}^{K} w_i\, \boldsymbol{u}_i\boldsymbol{u}_i^\top, \qquad w_i = \log(1+\lambda_i), \tag{3}$$

where $\boldsymbol{u}_i$ and λ_i denote the corresponding eigenvector and eigenvalue. This operator jointly represents subspace geometry through its eigenvectors and spectral strength through its weights. The *unified statistic* is the squared Frobenius distance between the left and right operators,

$$T_t = \big\|\widehat{\boldsymbol{C}}_L^{(K)} - \widehat{\boldsymbol{C}}_R^{(K)}\big\|_F^2 = \|\boldsymbol{w}_L\|^2 + \|\boldsymbol{w}_R\|^2 - 2\,\boldsymbol{w}_L^\top(\boldsymbol{M}\circ\boldsymbol{M})\,\boldsymbol{w}_R, \qquad \boldsymbol{M} = \boldsymbol{U}_L^\top\boldsymbol{U}_R, \tag{4}$$

where $\boldsymbol{w}_L = (w_{L,1},\ldots,w_{L,K})^\top$ and $\boldsymbol{w}_R = (w_{R,1},\ldots,w_{R,K})^\top$ collect the spectral weights, $\boldsymbol{M} \in \mathbb{R}^{K\times K}$ is the cross-Gram matrix of the two eigenbases, and $\boldsymbol{M}\circ\boldsymbol{M}$ denotes its Hadamard square. The closed form follows from the orthonormality of the eigenvectors, with $\|\boldsymbol{u}_i\boldsymbol{u}_i^\top\|_F = 1$ and $\langle \boldsymbol{u}_{L,i}\boldsymbol{u}_{L,i}^\top, \boldsymbol{u}_{R,j}\boldsymbol{u}_{R,j}^\top\rangle_F = M_{ij}^2$. The weighting $w_i = \log(1+\lambda_i)$ defines the deployed statistic $T_{\log} = T_t$, while the plain-eigenvalue weighting $w_i = \lambda_i$ defines the raw variant T_{raw}.

Two familiar statistics arise as limiting cases of equation 4: the eigenvalue-difference statistic $E_t = \|\boldsymbol{w}_L - \boldsymbol{w}_R\|^2$ and the weighted squared sine-$\boldsymbol{\Theta}$ statistic $D_t = 2\sum_{i,j} w_i w_j(\delta_{ij} - M_{ij}^2)$. Detailed derivations and interpretations are provided in the Appendix.

For sequential monitoring, we self-normalize T_t using a heteroskedasticity- and autocorrelation-consistent (HAC) null scale $\hat{\sigma}_T$ estimated from the left window, and accumulate the normalized statistic through CUSUM:

$$\tilde{T}_t = T_t/\hat{\sigma}_T, \qquad g_t = \max\big(0,\, g_{t-1} + \tilde{T}_t - k\big), \tag{5}$$

A break is declared when g_t first exceeds the threshold c_γ, calibrated to a target average run length (ARL) γ under the no-change null. The change point is localized as $\hat{\tau} = \arg\max_s \tilde{T}_s$ over the active window, and $(\hat{\tau}, \widehat{\boldsymbol{U}}_{\hat{\tau}})$ is passed to the adaptation module. The HAC self-normalization stabilizes the null calibration under heteroskedasticity, serial dependence, and varying ambient dimension p. Detection-delay guarantees for rank-preserving span-changing breaks are established in Theorem 3.

3.4 Estimating an unknown number of factors

The truncation rank K in equation 3 is the detector's only structural parameter. We estimate the number of factors $\hat{r}$ from the pre-break window using the Bai–Ng information criterion (Bai & Ng, 2002) and set $K = \hat{r} + 1$. This upward adjustment accounts for the asymmetric effects of rank misspecification: mild over-specification introduces an additional near-null direction, whereas under-specification may omit a genuine factor and weaken sensitivity to structural changes. The extra $(\hat{r}+1)$-th direction also enables detection of a factor *birth*, whose new direction lies outside the estimated pre-break span. Because weak factors may lead to underestimation of $\hat{r}$, the upward adjustment also provides a buffer against rank underestimation.

3.5 BREAK-TRIGGERED READOUT-ONLY ADAPTATION

The readout $h_\psi(\cdot; \widehat{\boldsymbol{U}})$ operates on the factor coordinates $\widehat{\boldsymbol{U}}^\top \mathrm{backbone}(\boldsymbol{X})$. When no break is detected, both the backbone and readout remain frozen. When a break is declared at $\hat{\tau}$, we re-estimate $\widehat{\boldsymbol{U}}$ from the post-break window and update only the readout using either a closed-form ridge update or one gradient step with the refreshed factor coordinates. The update involves $\mathcal{O}(rp)$ parameters while keeping the backbone fixed, thereby aligning the forecaster with the new low-rank structure using few post-break samples. This event-triggered update avoids unnecessary adaptation during stationary periods and reduces the risk of catastrophic forgetting relative to full test-time fine-tuning (Kim et al., 2025).

Algorithm 1: FLARE online detection-and-adaptation procedure

Input: stream $\{\boldsymbol{X}_t\}$; encoder g_ϕ; window h; rank $K = \hat{r} + 1$; ARL target γ (threshold c_γ); frozen backbone; readout h_ψ

Initialize $g_0 \leftarrow 0$ and readout basis $\widehat{\boldsymbol{U}} \leftarrow g_\phi(\boldsymbol{X}_{0:h})$

for $t = 2h, 2h+1, \ldots$ **do**

- $s \leftarrow t - h$
- Encode the split's left/right windows: $(\boldsymbol{U}_L, \boldsymbol{w}_L) \leftarrow g_\phi(\boldsymbol{X}_{s-h:s})$, $(\boldsymbol{U}_R, \boldsymbol{w}_R) \leftarrow g_\phi(\boldsymbol{X}_{s:t})$
- $\boldsymbol{M} \leftarrow \boldsymbol{U}_L^\top \boldsymbol{U}_R$; $T_s \leftarrow \|\boldsymbol{w}_L\|^2 + \|\boldsymbol{w}_R\|^2 - 2\boldsymbol{w}_L^\top(\boldsymbol{M} \circ \boldsymbol{M})\boldsymbol{w}_R$
- Update HAC null scale $\hat{\sigma}_T$; $\tilde{T}_s \leftarrow T_s/\hat{\sigma}_T$; $g_t \leftarrow \max(0, g_{t-1} + \tilde{T}_s - k)$
- **if** $g_t > c_\gamma$ **then**
  - $\hat{\tau} \leftarrow \arg\max_{s'} \tilde{T}_{s'}$ over the active window
  - $\widehat{\boldsymbol{U}} \leftarrow g_\phi(\boldsymbol{X}_{\hat{\tau}:\hat{\tau}+h})$
  - Update readout ψ (ridge / one-step gradient), with backbone frozen
  - Reset $g_t \leftarrow 0$ after the triggered update
- Forecast $\hat{\boldsymbol{X}}_{t+1} \leftarrow h_\psi(\mathrm{backbone}(\boldsymbol{X}_{t-h:t}); \widehat{\boldsymbol{U}})$

4 THEORETICAL GUARANTEES

Under the piecewise-stationary factor model, we establish false-alarm control, detection-delay guarantees, a weak-factor detectability boundary for rank-preserving span-changing breaks, and a spectral boundary for rank-changing breaks. The unified statistic detects identifiable span-preserving breaks through second-moment changes. We begin by stating the assumptions underlying our theoretical analysis.

Assumption 1 (Regularity and weak-factor regime)**.** *For the rank-preserving analysis, the factor covariance satisfies $\boldsymbol{\Sigma}_f \to \boldsymbol{\Sigma}_f^\star \succ 0$, with fixed r and $\|\boldsymbol{\Sigma}_f\| \asymp 1$. The idiosyncratic process is mean zero, stationary, and geometrically α-mixing, with covariance $\boldsymbol{\Sigma}_e$ satisfying $\|\boldsymbol{\Sigma}_e\| \le C$. The factors and idiosyncratic components have uniformly bounded eighth moments, and the whitened scores are sub-Weibull. The segment lengths satisfy $\min_j(\tau_j - \tau_{j-1}) \ge \Delta$ with $\Delta/h \to \infty$, and the factor-strength exponent satisfies $\delta \in (\delta_0, 1]$ for some fixed $\delta_0 > 0$.*

Assumption 2 (Amortized subspace rate)**.** *On any stationary window of length h, the encoder satisfies*

$$\|\sin\boldsymbol{\Theta}(\widehat{\mathcal{S}}_{a:b}, \mathcal{S})\| = \mathcal{O}_P(\rho_{p,h}), \qquad \rho_{p,h} = \mathcal{O}\big(p^{-\delta/2}h^{-1/2}\big), \qquad \rho_{p,h} \to 0. \tag{6}$$

Under heteroskedasticity, define $\boldsymbol{D} := \mathrm{diag}\big(\mathrm{var}(e_{t,1})^{1/2}, \ldots, \mathrm{var}(e_{t,p})^{1/2}\big)$ for the diagonal matrix of marginal idiosyncratic scales, per-window PCA attains the rate $\mathcal{O}_P(\kappa\rho_{p,h})$, where $\kappa = \|\boldsymbol{D}\|\|\boldsymbol{D}^{-1}\|$ is the condition number associated with $\boldsymbol{D}$, whereas the whitened amortized encoder removes this conditioning factor.

Assumption 1 specifies the factor, idiosyncratic, moment, segment-length, and weak-factor conditions used in the analysis. Assumption 2 characterizes the subspace-estimation rate of the amortized encoder.

Theorem 3 (Online detection delay under weak factors)**.** *Under Assumptions 1–2, consider a rank-preserving span-changing break at τ_j with maximal principal angle $\overline{\theta}_j$. If*

$$\rho_{p,h}^2 \overline{\theta}_j^{-2} \log \gamma = o(h), \tag{7}$$

then the detector has the following properties:

(i) under the no-change null H_0, $\mathrm{ARL} \geq \gamma$;

(ii) the detection delay satisfies

$$\hat{\tau}_j - \tau_j = h + \mathcal{O}_P\Big(\rho_{p,h}^2 \, \overline{\theta}_j^{-2} \log \gamma\Big). \tag{8}$$

Consequently, the change point is localized consistently when $h = o(\Delta)$.

Corollary 4 (Strong-factor recovery)**.** *If $\delta = 1$, then $\rho_{p,h} = \mathcal{O}((ph)^{-1/2})$, and the detection-delay bound becomes $h + \mathcal{O}_P\big((ph)^{-1}\overline{\theta}_j^{-2} \log \gamma\big)$ for a rank-preserving span-changing break.*

Theorem 3 separates the detection delay into the window latency h and an additional accumulation term governed by the subspace-estimation error and break magnitude. Corollary 4 shows that under strong factors, this additional delay scales as $(ph)^{-1}\overline{\theta}_j^{-2} \log \gamma$, bringing the overall delay closer to the intrinsic window latency h.

Theorem 5 (Detection boundary)**.** *Consider rank-preserving span-changing breaks with maximal principal angle $\overline{\theta}$ and factor-strength exponent δ, where $\rho_{p,h} = p^{-\delta/2}h^{-1/2}$. Then:*

(i) Impossibility. *If $\overline{\theta}^2 = o(\rho_{p,h}^2)$, no online test can distinguish the no-change null from a least-favorable one-direction span-changing alternative with vanishing total error. Hence, consistent localization is information-theoretically impossible.*

(ii) Achievability. *If*

$$\overline{\theta}^2 \gg \rho_{p,h}^2 \log \gamma / \Delta, \qquad \rho_{p,h}^2 \overline{\theta}^{-2} \log \gamma = o(h), \tag{9}$$

then the detector localizes the break with the delay in Theorem 3.

(iii) Estimator ordering. *Under heteroskedasticity, let $\delta^\star_{\mathrm{FLARE}}$ denote the critical factor-strength boundary induced by the achievability condition, and $\delta^\star_{\mathrm{PCA}}$ analogously after replacing $\rho_{p,h}$ by the PCA rate $\kappa\rho_{p,h}$. Consequently, when $\kappa > 1$, we have*

$$\delta^\star_{\mathrm{PCA}} > \delta^\star_{FLARE}. \tag{10}$$

Theorem 5 characterizes a weak-factor phase transition in detectability and shows how whitening and the window length h shift the achievable boundary. These predictions are supported by the experiments in Section 5.

5 Experiments

Our experiments examine weak-factor break detection, the effect of whitening on the detectability boundary, and the efficiency of break-triggered adaptation. Synthetic studies evaluate detection performance and key design choices, while real-data experiments assess regime localization and adaptation. All synthetic results are averaged over repeated runs, with the experimental setup summarized in Section 5.1.

5.1 Setup

We summarize the datasets, forecasting backbone, and evaluation metrics used in our experiments.

Data. *Synthetic:* We generate piecewise-stationary factor panels from the model in Section 3.1, varying factor strength δ, break type (span-changing, span-preserving, or rank change), heteroskedastic idiosyncratic scale, and tail behavior. Ground-truth change points are known by construction. *Real:* We use the Ken French 100 size-and-value portfolio panel ($p = 95$, 678 months)

Table 1: Weak-factor detection against external online baselines under a common target ARL $\gamma = 1000$. Recall and F_1 detectability boundaries and the realized null ARL are given. A dash indicates that the corresponding metric did not reach 0.8 within the evaluated factor strength range, so no finite detectability boundary was obtained. Lower $\delta^\star$ indicates detection under weaker factors.

Method	$\delta^\star_{\text{recall}}$	$\delta^\star_{F_1}$	ARL
FLARE $T_{\log}$ (ours)	**0.50**	**0.50**	996
Subspace-CUSUM (raw)	—	—	1045
Subspace-CUSUM (whitened)	0.90	—	965
Leading-eig CUSUM (whitened)	—	—	936
BOCPD (whitened PC scores)	—	—	1045
Streaming-PCA angle (recursive)	0.70	—	1183

as the primary benchmark and the FRED-MD monthly macroeconomic panel (McCracken & Ng, 2016) ($p = 121$, 770 months, 1960–2024) as secondary corroboration. Both are canonical high-dimensional approximate-factor benchmarks, on which we apply the whitened detector to localize documented structural shifts.

Backbone. Forecasting experiments use a frozen lightweight temporal backbone with the factor-readout head of Section 3.5. The adaptation mechanism is backbone-agnostic and also applies to frozen time-series foundation models.

Metrics. Detection performance is evaluated by tolerance-windowed recall, F_1, average run length (ARL), and expected detection delay (EDD). For a metric m, we define $\delta^\star_m$ as the smallest factor strength at which its mean reaches 0.8, and report the resulting recall and F_1 boundaries, $\delta^\star_{\text{recall}}$ and $\delta^\star_{F_1}$. Forecasting performance under shift is measured by post-break MSE, decomposed into common and idiosyncratic components.

5.2 Simulation Results

We compare FLARE with five online baselines covering complementary detection principles. Subspace-CUSUM monitors changes in the estimated loading space, using either the raw panel covariance or the whitened encoder representation. Leading-eig CUSUM tracks changes in the leading eigenvalue and therefore focuses on spectral strength rather than the full subspace geometry. BOCPD applies Bayesian online change-point detection to the whitened principal-component scores. Streaming-PCA angle uses a recursive PCA estimator and monitors the angle between consecutive estimated subspaces. All baselines use the same two-window online protocol and are calibrated to the common target ARL $\gamma = 1000$.

Under matched ARL calibration, $T_{\log}$ achieves both $\delta^\star_{\text{recall}} = 0.50$ and $\delta^\star_{F_1} = 0.50$, outperforming all external online baselines in the weak-factor regime. The strongest baseline reaches only $\delta^\star_{\text{recall}} = 0.70$, while none attains a finite F_1 boundary. Holding the detector fixed, replacing per-window PCA with the whitened encoder further improves the recall boundary from 0.70 to 0.50 and the F_1 boundary from 0.90 to 0.50, consistent with the estimator ordering in Theorem 5. Additional baseline and estimator results are reported in Appendix C.1 and C.2.

5.3 Ablation and Sensitivity Analysis

Under matched ARL calibration, $T_{\log}$ achieves full recall on the span-changing break while remaining sensitive to span-preserving changes, whereas the angle limit D_t and energy limit E_t each capture only one break type. The raw-eigenvalue variant T_{raw} performs poorly on both. Across the weak-factor sweep, $T_{\log}$ attains $\delta^\star_{\text{recall}} = \delta^\star_{F_1} = 0.50$, compared with $\delta^\star_{\text{recall}} = 0.60$ and no finite F_1 boundary for D_t.

The detector uses a single structural parameter, $K = \hat{r} + 1$, with $\hat{r}$ estimated by the Bai–Ng criterion. Rank under-specification is substantially more harmful than mild over-specification, motivating the one-step upward adjustment, which also preserves sensitivity to factor births without introducing additional structural tuning. See results in table C.2 in Appendix C.

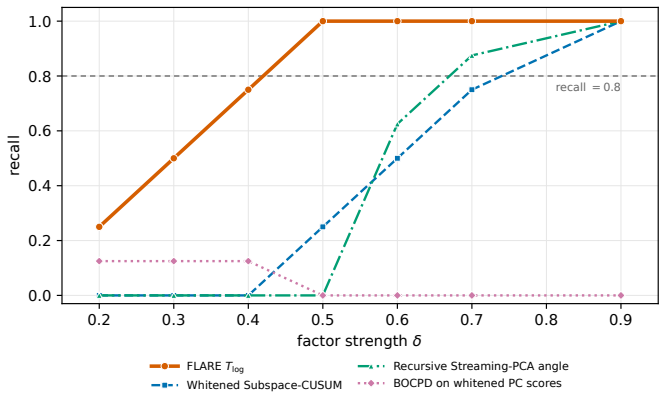


Figure 2: Weak-factor detectability under target ARL $\gamma = 1000$. Curves show tolerance-windowed recall versus factor strength δ, with the dashed line at recall $= 0.8$ defining $\delta^{\star}_{\text{recall}}$. FLARE reaches the boundary at $\delta = 0.50$, compared with 0.70 for recursive Streaming-PCA angle CUSUM and 0.90 for whitened Subspace-CUSUM.

Table 2: Statistic ablation under ARL-matched calibration ($p = 100$, $\delta = 0.5$, $\gamma = 1000$). Reported are recall for span-changing breaks (S1), span-preserving rescaling (S2, with a subscript denote the strength multiplier), false-alarm rate under the identification null (S3), and realized null ARL under corrected and default thresholds. D_t and E_t are the angle and energy limits, T_{raw} uses $w_i = \lambda_i$, sym-max denotes $D_t \vee E_t$, and $T_{\log}$ is the deployed log-weighted statistic.

Statistic	S1	$S2_{1.5}$	$S2_2$	$S2_3$	S3 null FA	ARL (corr.)	(default)
D_t (sinΘ limit)	0.58	0.17	0.17	0.17	**0.00**	1094	(2851)
E_t (energy limit)	0.00	0.33	0.50	0.83	0.125	950	(1775)
T_{raw} (λ)	0.25	0.33	0.33	0.25	0.06	922	(1711)
sym-max ($D_t \vee E_t$)	0.42	0.42	0.50	0.83	0.06	1012	(2543)
$T_{\log}$ (deployed)	**1.00**	0.25	0.50	0.75	0.06	1001	(2613)

5.4 Empirical Study

On the Ken French portfolio panel, the whitened detector identifies a single break in November 2008, closely aligned with the global financial crisis, while remaining inactive during the 2020 pandemic episode. On FRED-MD, it detects breaks in March 1983 and September 2003, with the former associated with the Great Moderation transition.

The break-triggered adaptation mechanism achieves a favorable accuracy–efficiency tradeoff. On the synthetic multi-break panel, FLARE attains post-break MSE 2.356, closely matching always-on TAFAS at 2.352 and improving over the frozen backbone at 2.490, while requiring only about one readout update per detected break compared with 24–30 updates for continuous adaptation. A similarly sparse update pattern is observed across the real panels.

6 Conclusion

In this paper, we proposed FLARE, an online and differentiable method for detecting and adapting to regime shifts in the low-rank common structure of high-dimensional time series. It uses a unified statistic to detect both span-changing and span-preserving breaks, while triggering lightweight adaptation only when a break is detected. Our theoretical analysis establishes weak-factor detectability and detection-delay guarantees, showing that whitening lowers the factor-strength threshold for break detection. Numerical studies further demonstrate the detection accuracy and adaptation efficiency of FLARE. A limitation of the current formulation is its reliance on a linear low-rank factor structure. Extending the method to nonlinear common structures and more general nonstationary factor processes is an important direction for future work.

AI USE STATEMENT

Generative AI tools assisted with language editing to improve the clarity and readability of author-written text. The authors reviewed all resulting edits and are responsible for the final manuscript.

ETHICS STATEMENT

This work is methodological and uses synthetic and existing datasets for evaluation. It does not involve human subjects or the collection of sensitive personal information. We are not aware of any substantial ethical concerns arising from this study.

REPRODUCIBILITY STATEMENT

The experimental setup and evaluation metrics are summarized in Section 5.1, with the main protocols and results reported in Section 5. Additional comparisons and sensitivity analyses are provided in Appendix C. Theoretical assumptions and guarantees are stated in Section 4, with supporting derivations and proofs in Appendix A.

REFERENCES

Abdul Fatir Ansari, Lorenzo Stella, Ali Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Hao Wang, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Bernie Wang. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024.

Jushan Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171, 2003. doi: 10.1111/1468-0262.00392.

Jushan Bai and Xu Han. Structural changes in high dimensional factor models. *Frontiers of Economics in China*, 11(1):9–39, 2016. doi: 10.3868/s060-005-016-0003-9.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002. doi: 10.1111/1468-0262.00273.

Jushan Bai and Serena Ng. Approximate factor models with weaker loadings. *Journal of Econometrics*, 235(2):1893–1916, 2023. doi: 10.1016/j.jeconom.2023.01.027.

Jushan Bai, Jiangtao Duan, and Xu Han. The likelihood ratio test for structural changes in factor models. *Journal of Econometrics*, 238(2):105631, 2024. doi: 10.1016/j.jeconom.2023.105631.

Matteo Barigozzi and Lorenzo Trapani. Sequential testing for structural stability in approximate factor models. *Stochastic Processes and their Applications*, 130(8):5149–5187, 2020. doi: 10.1016/j.spa.2020.03.003.

Matteo Barigozzi, Haeran Cho, and Piotr Fryzlewicz. Simultaneous multiple change-point and factor analysis for high-dimensional time series. *Journal of Econometrics*, 206(1):187–225, 2018. doi: 10.1016/j.jeconom.2018.05.003.

Matteo Barigozzi, Haeran Cho, and Lorenzo Trapani. Moving sum procedure for multiple change point detection in large factor models. *Journal of Time Series Analysis*, 47(3):450–464, 2026. doi: 10.1111/jtsa.70028.

Nina Dörnemann and Debashis Paul. Detecting spectral breaks in spiked covariance models. *Bernoulli*, 32(2):1243 – 1266, 2026. doi: 10.3150/25-BEJ1900.

Jiangtao Duan, Jushan Bai, and Xu Han. Quasi-maximum likelihood estimation of break point in high-dimensional factor models. *Journal of Econometrics*, 233(1):209–236, 2023. doi: 10.1016/j.jeconom.2021.12.011.

Jiangtao Duan, Jushan Bai, and Xu Han. Taxonomy and estimation of multiple breakpoints in high-dimensional factor models. *Journal of the American Statistical Association*, 0(ja):1–123, 2026. doi: 10.1080/01621459.2026.2719782.

Jianqing Fan, Yuan Liao, and Martina Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(4):603–680, 09 2013. doi: 10.1111/rssb.12016.

Jianqing Fan, Yuan Liao, and Xiaofeng Shi. Risks of large portfolios. *Journal of Econometrics*, 186 (2):367–387, 2015. doi: 10.1016/j.jeconom.2015.02.015.

Jianqing Fan, Yuling Yan, and Yuheng Zheng. When can weak latent factors be statistically inferred? *arXiv preprint arXiv:2407.03616*, 2024.

Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554, 2000. doi: 10.1162/003465300559037.

Ren Fujiwara, Yasuko Matsubara, and Yasushi Sakurai. When to retrain after drift: A data-only test of post-drift data size sufficiency. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 131925–131945, 2026.

Javier Gonzalez-Castillo and Peter A. Bandettini. Task-based dynamic functional connectivity: Recent findings and open questions. *NeuroImage*, 180:526–533, 2018. doi: 10.1016/j.neuroimage.2017.08.006.

R. Matthew Hutchison, Thilo Womelsdorf, Elena A. Allen, Peter A. Bandettini, Vince D. Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H. Duyn, Gary H. Glover, Javier Gonzalez-Castillo, Daniel A. Handwerker, Shella Keilholz, Vesa Kiviniemi, David A. Leopold, Francesco de Pasquale, Olaf Sporns, Martin Walter, and Catie Chang. Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage*, 80:360–378, 2013. doi: 10.1016/j.neuroimage.2013.05.079.

Jeonghwan Im and Hyuk-Yoon Kwon. Cosa: Context-aware output-space adapter for test-time adaptation in time series forecasting. In *International Conference on Learning Representations*, 2026.

HyunGi Kim, Siwon Kim, Jisoo Mok, and Sungroh Yoon. Battling the non-stationarity in time series forecasting via test-time adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17868–17876, 2025. doi: 10.1609/aaai.v39i17.33965.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*, 2022.

Bonsoo Koo, Benjamin Wong, and Ze-Yu Zhong. Disentangling structural breaks in factor models for macroeconomic data. *Journal of Business & Economic Statistics*, 0(0):1–12, 2026. doi: 10.1080/07350015.2025.2583205.

Clifford Lam and Qiwei Yao. Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 40(2):694–726, 2012.

Xialu Liu and Ting Zhang. Estimating change-point latent factor models for high-dimensional time series. *Journal of Statistical Planning and Inference*, 217:69–91, 2022. doi: 10.1016/j.jspi.2021.07.006.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. In *Advances in Neural Information Processing Systems*, volume 35, pp. 9881–9893, 2022. doi: 10.52202/068431-0718.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations*, volume 2024, pp. 11116–11140, 2024.

Michael W. McCracken and Serena Ng. Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589, 2016. doi: 10.1080/07350015.2015.1086655.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023.

Alexei Onatski. Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479, 2009. doi: 10.3982/ECTA6964.

Alexei Onatski. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258, 2012. doi: 10.1016/j.jeconom.2012.01.034.

James H. Stock and Mark W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162, 2002a. doi: 10.1198/073500102317351921.

James H Stock and Mark W Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, 2002b. doi: 10.1198/016214502388618960.

Haochun Wang, Ruichen Xu, Georgios Kementzidis, Karen Cho, Sebastian Ramirez Villarreal, and Yuefan Deng. Towards principled test-time adaptation for time series forecasting. *arXiv preprint arXiv:2605.17250*, 2026.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *Proceedings of the 41st International Conference on Machine Learning*, 2024.

Liyan Xie, Yao Xie, and George V. Moustakides. Sequential subspace change point detection. *Sequential Analysis*, 39(3):307–335, 2020. doi: 10.1080/07474946.2020.1823191.

A Technical Details

A.1 Supplements for Method and Theory

Relationship between T_t, E_t and D_t in Section 3.3: *(i) Matching eigenvectors:* if the windows share an eigenbasis, $\boldsymbol{U}_L = \boldsymbol{U}_R$ so $\boldsymbol{M} = \boldsymbol{M} \circ \boldsymbol{M} = \boldsymbol{I}_K$, and equation 4 collapses to $T_t = \|\boldsymbol{w}_L - \boldsymbol{w}_R\|^2$, the eigenvalue-difference or *energy* statistic E_t, reading span-preserving spectral change and rank change off the spectrum. *(ii) Matching eigenvalues:* if the windows share a spectrum, $\boldsymbol{w}_L = \boldsymbol{w}_R = \boldsymbol{w}$, then $T_t = 2\sum_{i,j} w_i w_j \, (\delta_{ij} - M_{ij}^2)$, a weighted squared sine-$\boldsymbol{\Theta}$ distance D_t; with equal weights this is exactly $2w^2\big(K - \|\boldsymbol{M}\|_F^2\big) = 2w^2 \|\sin\boldsymbol{\Theta}(\widehat{\mathcal{S}}_L, \widehat{\mathcal{S}}_R)\|_F^2$, reading a rank-preserving span-changing movement off the principal angles. In general equation 4 interpolates between these limits, so D_t and E_t are two facets of one statistic, carried by a single null, threshold, and boundary. The combiner ablation (Section 5) confirms that splitting the two into separate alarms misses the joint detection-and-precision bar and over-fires under the null. The same operator distance also subsumes per-series instance normalization (RevIN; Kim et al., 2022) as the diagonal-rescaling case.

Proposition A.1 (Rank-change spectral detectability)**.** *A rank-changing break is a span-changing subcase in which the factor dimension changes. Suppose a new factor contributes a whitened population spike λ_{new} to the post-break covariance, and let λ_{edge} denote the upper edge of the noise spectrum and $a_{p,h}$ the stochastic resolution scale of the $(r+1)$-th sample eigenvalue. If*

$$\lambda_{\text{new}} - \lambda_{\text{edge}} \gg a_{p,h},$$

the new spike separates from the noise bulk with probability tending to one and the spectral-energy component of T_t can detect the factor birth once the separated spike exceeds the detector's calibrated threshold. If

$$\lambda_{\text{new}} - \lambda_{\text{edge}} = o(a_{p,h}),$$

this spectral argument provides no consistent detection guarantee. Under the standard high-dimensional spiked-covariance normalization, the separation condition is the familiar BBP-type supercritical-spike condition; its exact constant depends on the normalization, noise spectrum, and aspect ratio.

Remark A.2 (Scope and coverage)**.** *The theoretical angle results concern rank-preserving span-changing breaks. Rank-changing breaks remain within the span-changing class but require the separate spectral argument in Proposition A.1. Span-preserving breaks satisfy $\mathcal{S}^{(j+1)} = \mathcal{S}^{(j)}$ and are detected by the spectral component of T_t when the change perturbs the second moment (e.g., non-orthogonal loading transformations or non-isotropic factors). The case of an orthogonal loading rotation with isotropic factors and $\mathrm{Cov}(\boldsymbol{f}_t) = \boldsymbol{I}$ is an identification null because the data distribution is unchanged. The detailed perturbation, null calibration, two-point lower bound, angle–energy bridge, and rank-change arguments are given in Appendix B.*

B Technical Proofs

Throughout, c, C denote positive constants that may change from line to line, and all limits are as $p, h \to \infty$ with r fixed. Recall the segment model equation 1, the common subspace $\mathcal{S} = \mathrm{col}(\boldsymbol{\Lambda})$ with factor strength $\|\boldsymbol{\Lambda}^\top \boldsymbol{\Lambda}\| \asymp p^\delta$, the window sample covariance $\widehat{\boldsymbol{\Sigma}} = \frac{1}{h}\sum_t (\boldsymbol{X}_t - \bar{\boldsymbol{X}})(\boldsymbol{X}_t - \bar{\boldsymbol{X}})^\top$ on a length-h stationary window, and the population $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_f\boldsymbol{\Lambda}^\top + \boldsymbol{\Sigma}_e$.

This appendix proves Theorem 3, Corollary 4, Theorem 5, and Proposition A.1. Throughout, on a stationary segment we write $\mathcal{S} = \mathrm{col}(\boldsymbol{\Lambda})$ and let $\widehat{\mathcal{S}}_{a:b}$ denote the encoder estimate from a window $(a, b]$ of length h. We use

$$\rho_{p,h} = p^{-\delta/2} h^{-1/2}$$

for the rate in Assumption 2; replacing this expression by the upper bound in that assumption leaves all conclusions unchanged. The delay proof is first written for the angle statistic

$$D_t = \big\| \sin \boldsymbol{\Theta}(\widehat{\mathcal{S}}_{t-h:t}, \widehat{\mathcal{S}}_{t:t+h}) \big\|_F^2.$$

Lemma B.3 below connects this angle statistic to the deployed operator statistic T_t in equation 4.

B.1 Window perturbation and subspace error

Lemma B.1 (Window subspace error)**.** *For a stationary window of length h contained in one segment, under Assumptions 1–2,*

$$\big\| \sin \boldsymbol{\Theta}(\widehat{\mathcal{S}}_{a:b}, \mathcal{S}) \big\| = \mathcal{O}_P(\rho_{p,h}).$$

Proof. Let $\widehat{\boldsymbol{\Sigma}}_{a:b}$ be the window covariance and write

$$\widehat{\boldsymbol{\Sigma}}_{a:b} = \boldsymbol{\Lambda}\widehat{\boldsymbol{\Sigma}}_f\boldsymbol{\Lambda}^\top + \boldsymbol{\Lambda}\widehat{\boldsymbol{C}}_{fe} + \widehat{\boldsymbol{C}}_{fe}^\top\boldsymbol{\Lambda}^\top + \widehat{\boldsymbol{\Sigma}}_e.$$

The population signal covariance is $\boldsymbol{\Sigma}_s = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_f\boldsymbol{\Lambda}^\top$. Since r is fixed and $\boldsymbol{\Sigma}_f \to \boldsymbol{\Sigma}_f^\star \succ 0$, its r nonzero eigenvalues are of order p^δ, while the $(r+1)$-st eigenvalue is zero. Thus the eigengap between $\mathcal{S}$ and its orthogonal complement is of order p^δ.

The component of the sample perturbation that can rotate the signal space is the normal-space component of $\boldsymbol{\Lambda}\widehat{\boldsymbol{C}}_{fe} + \widehat{\boldsymbol{C}}_{fe}^\top\boldsymbol{\Lambda}^\top + \widehat{\boldsymbol{\Sigma}}_e$. Under the mixing and moment conditions in Assumption 1, its operator norm is $\mathcal{O}_P(p^{\delta/2}h^{-1/2} + 1)$. Fluctuations wholly within $\mathcal{S}$ change the eigenvalues but do not rotate the population space and therefore do not enter the leading normal-space perturbation. Applying the Davis–Kahan $\sin\boldsymbol{\Theta}$ bound gives

$$\big\| \sin \boldsymbol{\Theta}(\widehat{\mathcal{S}}_{a:b}, \mathcal{S}) \big\| \lesssim \frac{p^{\delta/2}h^{-1/2} + 1}{p^\delta} = \mathcal{O}_P\Big(p^{-\delta/2}h^{-1/2}\Big) + \mathcal{O}_P(p^{-\delta}).$$

In the regime of Assumption 2, the first term is the stated rate; the remaining term is lower order. This proves the claim. The same argument after whitening removes the idiosyncratic conditioning factor κ, whereas per-window PCA has rate $\mathcal{O}_P(\kappa\rho_{p,h})$ as stated in Assumption 2. □

B.2 NULL CALIBRATION AND DETECTION DELAY

Lemma B.2 (Null order and ARL calibration). *Under the no-change null and Assumptions 1–2,*

$$D_t = \mathcal{O}_P(\rho_{p,h}^2).$$

After HAC self-normalization, calibration of the scan threshold to the $(1 - 1/\gamma)$ *quantile gives* $\mathrm{ARL} \geq \gamma + o(\gamma)$.

Proof. Under H_0, both windows lie in the same segment and target the same population space $\mathcal{S}$. By the triangle inequality for the principal-angle metric and Lemma B.1,

$$\begin{aligned} D_t^{1/2} &\leq \left\| \sin \boldsymbol{\Theta}(\widehat{\mathcal{S}}_{t-h:t}, \mathcal{S}) \right\|_F + \left\| \sin \boldsymbol{\Theta}(\widehat{\mathcal{S}}_{t:t+h}, \mathcal{S}) \right\|_F \\ &= \mathcal{O}_P(\rho_{p,h}), \end{aligned}$$

which yields the asserted order. The HAC scale consistently estimates the long-run scale of the dependent score sequence under the geometric mixing and moment conditions in Assumption 1. Consequently the normalized scan statistic is asymptotically pivotal. The usual Gaussian approximation for a geometrically mixing, sub-Weibull sequence controls the scan maximum; choosing c_γ as its $(1 - 1/\gamma)$ quantile makes the false-alarm probability per run at most $1/\gamma + o(1/\gamma)$, which is equivalent to the stated ARL bound. □

Lemma B.3 (Angle–energy bridge). *Let T_t be the statistic in equation 4 and let D_t be the angle statistic above. If the two windows have the same retained spectrum, $\boldsymbol{w}_L = \boldsymbol{w}_R = \boldsymbol{w}$, then*

$$T_t = 2\boldsymbol{w}^\top (\boldsymbol{I} - \boldsymbol{M} \circ \boldsymbol{M})\boldsymbol{w}.$$

In particular, if the retained weights are equal, $w_i \equiv w$, then

$$T_t = 2w^2 D_t.$$

More generally, if the retained weights are bounded above and below by positive constants independent of (p, h), the angle contribution to T_t has the same order as D_t. If the subspace is fixed, $\boldsymbol{M} = \boldsymbol{I}$ and $T_t = \|\boldsymbol{w}_L - \boldsymbol{w}_R\|^2$, so spectral changes are detected even when $D_t = 0$.

Proof. With $\boldsymbol{w}_L = \boldsymbol{w}_R = \boldsymbol{w}$, expanding the Frobenius norm in equation 4 gives

$$T_t = 2\|\boldsymbol{w}\|^2 - 2\boldsymbol{w}^\top (\boldsymbol{M} \circ \boldsymbol{M})\boldsymbol{w} = 2\boldsymbol{w}^\top (\boldsymbol{I} - \boldsymbol{M} \circ \boldsymbol{M})\boldsymbol{w}.$$

If $w_i \equiv w$, then

$$T_t = 2w^2 \sum_{i,j} (\delta_{ij} - M_{ij}^2) = 2w^2 \big(K - \|\boldsymbol{M}\|_F^2\big) = 2w^2 D_t,$$

using the principal-angle identity. Bounded spectral spread changes only the constant in this comparison. Finally, when $\boldsymbol{M} = \boldsymbol{I}$, the two operators have the same eigenvectors and their squared Frobenius distance is exactly $\|\boldsymbol{w}_L - \boldsymbol{w}_R\|^2$. □

Proof of Theorem 3. Part (i) follows from Lemma B.2. For part (ii), consider a rank-preserving span-changing break at τ_j, and write

$$\overline{\theta}_j = \theta_{\max}(\mathcal{S}^{(j)}, \mathcal{S}^{(j+1)}).$$

Once the candidate split is sufficiently far from the change that the left and right windows lie in their respective regimes, the population angle satisfies

$$D_j^\star = \left\| \sin \boldsymbol{\Theta}(\mathcal{S}^{(j)}, \mathcal{S}^{(j+1)}) \right\|_F^2 \geq \sin^2 \overline{\theta}_j.$$

Applying Lemma B.1 to both windows and using the triangle inequality gives

$$D_t^{1/2} \geq \sin \overline{\theta}_j - \mathcal{O}_P(\rho_{p,h}).$$

Thus, when $\overline{\theta}_j \gg \rho_{p,h}$, the post-change statistic has drift of order $\overline{\theta}_j^2$ above a null noise floor of order $\rho_{p,h}^2$.

The split at τ_j becomes observable only after the right window is completed, producing the unavoidable initial delay h. Thereafter, the self-normalized CUSUM accumulates drift of order $\overline{\theta}_j^2/\rho_{p,h}^2$ per effective increment, while its calibrated threshold is of order $\log\gamma$. Therefore the additional accumulation delay is

$$\mathcal{O}_P\Big(\rho_{p,h}^2\overline{\theta}_j^{-2}\log\gamma\Big),$$

and hence

$$\hat{\tau}_j - \tau_j = h + \mathcal{O}_P\Big(\rho_{p,h}^2\overline{\theta}_j^{-2}\log\gamma\Big).$$

Lemma B.3 transfers the same rate to T_t in the matched-spectrum or bounded-spread regime used by the deployed detector. The condition $\rho_{p,h}^2\overline{\theta}_j^{-2}\log\gamma = o(h)$ ensures that the accumulation term is smaller than the window-completion delay. If additionally $h = o(\Delta)$ and $\overline{\theta}_j^2 \gg \rho_{p,h}^2\log\gamma/\Delta$, the total delay is $o(\Delta)$ and localization is consistent. □

Proof of Corollary 4. For $\delta = 1$, $\rho_{p,h} = O((ph)^{-1/2})$. Substitution into Theorem 3 yields

$$\hat{\tau}_j - \tau_j = h + \mathcal{O}_P\Big((ph)^{-1}\overline{\theta}_j^{-2}\log\gamma\Big).$$

□

B.3 Detection boundary

Lemma B.4 (Two-point divergence)**.** *Consider a least-favorable rank-preserving alternative in which exactly one principal direction rotates by $\overline{\theta}$ and the remaining directions are unchanged. Let P_0 and P_1 denote the corresponding two-window laws under the null and alternative. In the Gaussian submodel satisfying Assumption 1,*

$$\mathrm{KL}(P_1 \,\|\, P_0) \le Chp^{\delta}\overline{\theta}^2 = C\frac{\overline{\theta}^2}{\rho_{p,h}^2}.$$

Proof. The one-direction rotation changes the covariance in the right window by a rank-two perturbation whose squared operator size is $O(\overline{\theta}^2)$. The factor covariance contributes signal strength of order p^{δ}, and the right window contains h observations. The Gaussian covariance formula for the Kullback–Leibler divergence, expanded to second order in the perturbation, thus gives

$$\mathrm{KL}(P_1 \,\|\, P_0) \le Chp^{\delta}\overline{\theta}^2.$$

Since $\rho_{p,h}^2 = p^{-\delta}h^{-1}$, this is the stated bound. □

Proof of Theorem 5. For part (i), if $\overline{\theta}^2 = o(\rho_{p,h}^2)$, Lemma B.4 gives $\mathrm{KL}(P_1\|P_0) \to 0$. Pinsker's inequality then implies $\|P_1 - P_0\|_{\mathrm{TV}} \to 0$, and Le Cam's two-point inequality gives, for any test φ,

$$\alpha(\varphi) + \beta(\varphi) \ge 1 - \|P_1 - P_0\|_{\mathrm{TV}} \longrightarrow 1.$$

Thus no online detector can separate the null and the least-favorable alternative with vanishing total error, and consistent localization is impossible.

For part (ii), Theorem 3 applies whenever

$$\overline{\theta}^2 \gg \rho_{p,h}^2\log\gamma/\Delta \quad \text{and} \quad \rho_{p,h}^2\overline{\theta}^{-2}\log\gamma = o(h),$$

which gives the stated achievable boundary and delay.

For the estimator comparison, per-window PCA replaces $\rho_{p,h}$ by $\kappa\rho_{p,h}$ under heteroskedasticity. Its sufficient condition therefore becomes $\overline{\theta}^2 \gg \kappa^2\rho_{p,h}^2\log\gamma/\Delta$. At fixed $(\overline{\theta}, h, \gamma, \Delta)$ this requires a stronger factor signal than the whitened encoder, yielding $\delta^\star_{\mathrm{PCA}} > \delta^\star_{FLARE}$ when $\kappa > 1$. □

Table C.1: Full comparison with external online detectors on the synthetic weak-factor benchmark under target ARL $\gamma = 1000$. Reported are recall for the span-changing break (S1) and span-preserving rescaling with $s = 2$ ($S2_2$), false-alarm rate under the identification null (S3), recall and F_1 detectability boundaries, realized null ARL, and online EDD at $\delta = 0.6$.

Method	S1	$S2_2$	S3 null FA ↓	$\delta^\star_{\text{recall}}$	$\delta^\star_{F_1}$	ARL	EDD@0.6
FLARE $T_{\log}$ (ours)	**1.00**	**0.375**	**0.00**	**0.50**	**0.50**	996	21.0
Subspace-CUSUM (raw)	0.00	0.125	0.125	—	—	1045	159.4
Subspace-CUSUM (whitened)	0.25	0.00	0.00	0.90	—	965	26.7
Leading-eig CUSUM (whitened)	0.125	0.125	0.00	—	—	936	158.7
BOCPD (whitened PC scores)	0.00	0.25	0.00	—	—	1045	—
Streaming-PCA angle (recursive)	0.00	0.00	0.00	0.70	—	1183	17.6

B.4 Rank-change detectability

Proof of Proposition A.1. A rank-changing break adds or removes a direction, so the rank-preserving principal-angle argument does not apply directly. Consider a factor birth and write the whitened post-break covariance as

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_0 + \lambda_{\text{new}} \boldsymbol{v}\boldsymbol{v}^\top, \qquad \|\boldsymbol{v}\| = 1.$$

Let λ_{edge} be the upper edge of the noise spectrum and let $a_{p,h}$ denote the stochastic resolution of the $(r+1)$-st sample eigenvalue. If

$$\lambda_{\text{new}} - \lambda_{\text{edge}} \gg a_{p,h},$$

standard spiked-covariance perturbation theory separates the new eigenvalue from the noise bulk with probability tending to one. Since the detector uses $K = \hat{r} + 1$, this separated eigenvalue contributes to the spectral-energy part of T_t, and the rank change is consistently detectable once that contribution exceeds the calibrated threshold.

Conversely, if $\lambda_{\text{new}} - \lambda_{\text{edge}} = o(a_{p,h})$, the new spike is below the resolution of the window covariance estimator. The two spectral laws are then contiguous at this scale, so no consistent rank-change guarantee is available. The resulting boundary is therefore BBP-type and depends on the whitened spike strength, as claimed. □

Remark B.5 (Scope of the proof)**.** *The angle-based delay and information-theoretic boundary concern rank-preserving span-changing breaks. Rank-changing breaks are handled separately through the spectral-energy argument above. Span-preserving breaks are outside the angle boundary because their subspaces coincide; they are detected by the spectral component of T_t when the second-order common structure changes. An orthogonal rotation with isotropic factors and covariance $\boldsymbol{I}$ is an identification null, because it leaves the observed-data distribution unchanged.*

C Additional Experiments

C.1 Full External-Baseline Results

Table C.1 provides the full comparison of $T_{\log}$ with five external online detectors on the synthetic weak-factor benchmark. All methods are calibrated to the same target ARL $\gamma = 1000$, with realized null ARLs ranging from 936 to 1183. Beyond the detectability boundaries reported in the main text, we report recall for the span-changing break (S1) and span-preserving rescaling with $s = 2$ ($S2_2$), false-alarm rate under the identification null (S3), and online expected detection delay (EDD) at $\delta = 0.6$. $T_{\log}$ is the only method attaining finite recall and F_1 boundaries while remaining sensitive to both break types. Streaming-PCA achieves a shorter EDD at $\delta = 0.6$ but does not attain a finite F_1 boundary.

The external detectors exhibit complementary sensitivities across break types. BOCPD is more responsive to span-preserving changes, whereas recursive Streaming-PCA attains the strongest external weak-factor recall boundary at $\delta^\star_{\text{recall}} = 0.70$ but no finite F_1 boundary. These results suggest that the advantage of $T_{\log}$ arises from both whitening and the unified two-window operator statistic.

Table C.2: Working-rank sensitivity at $p = 100$ and target ARL $\gamma = 1000$. Reported are S1 recall, recall and F_1 detectability boundaries, precision at $\delta = 0.6$, and realized null ARL. Under-specification degrades detection, whereas mild over-specification preserves span-change sensitivity. The plug-in rule uses $K = \hat{r} + 1$.

Rank	S1 recall	$\delta^\star_{\text{recall}}$	$\delta^\star_{F_1}$	prec@0.6	null ARL
$\hat{r}$=2, K=3 (under)	0.50	0.60	—	0.65	1206
$\hat{r}$=3, K=4 (deployed)	**1.00**	0.50	0.50	0.69	1001
$\hat{r}$=4, K=5 (over)	**1.00**	0.30	0.60	0.79	990
$\hat{r}$=5, K=6 (over)	**1.00**	0.20	0.70	0.63	950
$\hat{r}$ plug-in, K=$\hat{r}+1$	0.50	0.60	0.70	0.69	1176

Table C.3: Detector sensitivity across break types at $\delta = 0.5$ and target ARL $\gamma = 1000$. Reported are recall for the span-changing break (S1) and span-preserving rescalings (S2), and false-alarm rate under the identification null (S3). The unified statistic $T_{\log}$ remains sensitive to both break types, whereas the angle and energy limits are each selective to one.

Break class	D_t	E_t	sym-max	T_{raw}	$T_{\log}$
S1 orthogonal (span move)	0.58	0.00	0.42	0.25	**1.00**
S2 within-span, s=1.5	0.17	0.33	0.42	0.33	0.25
S2 within-span, s=2	0.17	0.50	0.50	0.33	0.50
S2 within-span, s=3	0.17	0.83	0.83	0.25	0.75
S3 identification null (FA ↓)	**0.00**	0.125	0.06	0.06	0.06

In the strong-factor regime, the two-window design incurs a modest delay cost. At $\delta = 0.9$, $T_{\log}$ and whitened Subspace-CUSUM both achieve full recall, while the recursive detector achieves a lower EDD (8.5 versus 12). This reflects the tradeoff between weak-factor sensitivity and detection speed under strong signals.

C.2 Additional Estimator Analysis

Using the same $T_{\log}$ statistic and CUSUM detector, the whitened encoder improves $\delta^\star_{\text{recall}}$ from 0.70 to 0.50 and $\delta^\star_{F_1}$ from 0.90 to 0.50 relative to per-window PCA. This ordering agrees with Theorem 5, confirming that whitening extends reliable detection further into the weak-factor regime.

C.3 The taxonomy is supported across all three break classes

Table C.3 confirms the expected detector–class correspondence. $T_{\log}$ detects the span-changing break while remaining sensitive to increasingly strong span-preserving rescaling, whereas the angle and energy limits are each selective to only one break type. Under the identification null, all statistics remain near zero, with $T_{\log}$ exhibiting only a low false-alarm rate.

Combining the angle and energy limits as separate channels does not recover the performance of the unified statistic. Neither the maximum rule nor alarm-budget splitting attains a finite F_1 boundary, and the split rules also yield poorer null calibration. In contrast, $T_{\log}$ achieves both $\delta^\star_{\text{recall}} = 0.50$ and $\delta^\star_{F_1} = 0.50$, supporting a single unified operator with one calibration and threshold.

C.4 Further Analysis

Figure C.1 illustrates the delay tradeoff between recursive Subspace-CUSUM and the two-window $T_{\log}$. Recursive detection is faster under strong factors, whereas $T_{\log}$ becomes more effective as the signal weakens. At $\delta = 0.5$, Subspace-CUSUM recall drops to 0.25, while $T_{\log}$ retains full recall. This reflects the tradeoff between strong-signal detection speed and weak-factor sensitivity induced by two-window evidence accumulation.

Subtle within-span rescalings remain challenging for all statistics, with $T_{\log}$ achieving only 0.25 recall at $s = 1.5$, where the second-moment perturbation is close to the null. Accordingly, the more

Table C.4: Combiner ablation at $\delta = 0.5$ and target ARL $\gamma = 1000$. Reported are recall for the span-changing break (S1) and span-preserving rescaling with $s = 2$ ($S2_2$), the recall detectability boundary, and realized null ARL. Sym-max combines D_t and E_t by their maximum, while split rules allocate the alarm budget between the two channels. None attains a finite F_1 boundary, and the split rules exhibit poorer null calibration.

Combiner	S1 recall	$S2_2$ recall	$\delta^\star_{\text{recall}}$	null ARL
D_t only	0.63	0.13	0.60	988
E_t only	0.00	0.63	—	1091
sym-max	0.63	0.75	—	1217
split a=0.5	0.75	0.63	—	624
split a=0.2	0.75	0.50	—	753
split a=0.1	0.75	0.50	—	702
split a=0.05	0.63	0.50	—	676

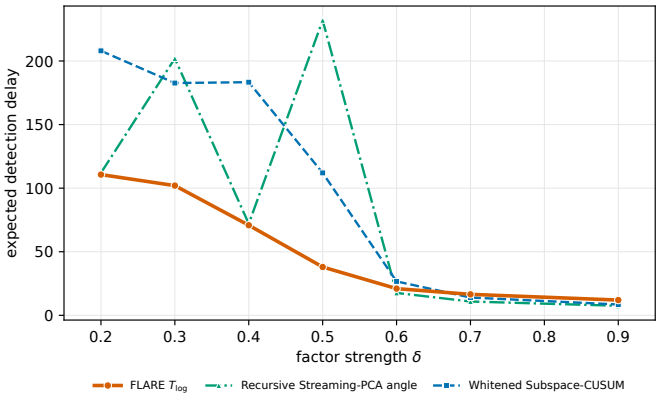


Figure C.1: Detection-delay tradeoff on the synthetic weak-factor benchmark under target ARL $\gamma = 1000$. Online expected detection delay is shown across factor strength δ for FLARE $T_{\log}$, recursive Streaming-PCA angle CUSUM, and whitened Subspace-CUSUM. Recursive detectors are faster under strong factors, whereas $T_{\log}$ exhibits more favorable delay behavior as the signal weakens toward the detectability boundary.

robust empirical evidence lies in the whitening-versus-PCA boundary gap and the relative ordering of the statistics rather than any single absolute $\delta^\star$.