OPTIMAL DETECTION OF GENERAL MOMENT CHANGES: SIMULTANEOUS MEAN AND COVARIANCE CHANGE DETECTION AND BEYOND

Anonymous authors
Paper under double-blind review

ABSTRACT

We study multiple change-point detection in multivariate time series whose distributions change in a piecewise constant manner. Distributional changes can manifest across different moment orders, from shifts in the mean and covariance to changes in higher-order moments. Higher-order moments capture increasingly rich distributional features but become difficult to estimate in high dimensions. Our tensor representation unifies moments of different orders within a common linear algebraic framework, enabling a new method to detect changes in moments of all orders up to a prescribed fixed order p. The resulting procedure accommodates temporal dependence and allows the dimension of the time series to grow with the sample size. Under suitable regularity conditions, the proposed procedure achieves a localization error rate that matches a newly developed minimax lower bound. We further derive limiting distributions under both nonvanishing and vanishing moment jumps and construct asymptotically valid confidence intervals in the vanishing-jump regime. Numerical experiments and real-data analyses demonstrate the method's effectiveness in detecting moment changes and a range of distributional shifts.

1 INTRODUCTION

Multivariate time series analysis is central to learning from dynamic systems whose components evolve jointly. Machine learning applications range from clinical monitoring (Hyland et al., 2020) and epidemic growth rates inference (Dehning et al., 2020) to global weather forecasting (Lam et al., 2023) and electricity-demand and traffic prediction (Liu et al., 2024). Such data can span multiple statistical regime transitions. Fitting a single model across those changes can obscure their important transition structure and may lead to misleading conclusions. Change-point analysis partitions observations into several homogeneous segments and quantifies uncertainty about the change points.

We consider the scenario where $X_1, \ldots, X_n \in \mathbb{R}^D$ is a time series whose distribution may change at multiple unknown times. We allow the dimension D to grow with the sample size n. Under the sub-Gaussian tail condition in Assumption 1, the full collection of joint moments uniquely determines the distribution, so every distributional change must alter at least one moment. This observation motivates a unified approach to distributional change detection through moment changes.

Throughout, $p \geq 2$ is a fixed integer that does not grow with n. We study changes in the moments $\mathbb{E}X_t^{\otimes r}$, $r = 1, \ldots, p$, where $X_t^{\otimes r}$ denotes the r-fold tensor product of X_t. We assume that these moment tensors are piecewise constant, with an unknown number $K \geq 1$ of change points $0 < \eta_1 < \cdots < \eta_K < n$. Specifically,

$$\mathbb{E}X_t^{\otimes r} \neq \mathbb{E}X_{t+1}^{\otimes r} \quad \text{for at least one } r \in \{1, \ldots, p\} \qquad \text{if and only if} \quad t \in \{\eta_1, \ldots, \eta_K\} \tag{1}$$

To detect changes in moments of different orders simultaneously, we introduce

$$Y_t = \begin{pmatrix} 1 \\ X_t \end{pmatrix} \in \mathbb{R}^{D+1}, \qquad t = 1, \ldots, n,$$

and define its order-p moment tensor as

$$\mathcal{M}_t^{(p)} = \mathbb{E}Y_t^{\otimes p}.$$

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.

Remark 1 (Hierarchy of moment tensors). *For $1 \le r \le p$, write $\mathcal{M}_t^{(r)} = \mathbb{E}Y_t^{\otimes r}$. Since $Y_{t,1} = 1$,*

$$\left(\mathcal{M}_t^{(r+1)}\right)_{1,i_1,\ldots,i_r} = \left(\mathcal{M}_t^{(r)}\right)_{i_1,\ldots,i_r}, \qquad 1 \le r < p, \tag{2}$$

for $i_1, \ldots, i_r \in \{1, \ldots, D+1\}$. Thus, $\mathcal{M}_t^{(p)}$ contains all lower-order moment tensors of Y_t and all joint moments of X_t through order p. For example, when $p = 3$, the first three moment orders appear as

$$\textit{Mean (first-order moments):} \qquad \left(\mathcal{M}_t^{(3)}\right)_{1,1,i+1} = \mathbb{E}X_{t,i},$$

$$\textit{Second-order moments:} \qquad \left(\mathcal{M}_t^{(3)}\right)_{1,i+1,j+1} = \mathbb{E}(X_{t,i}X_{t,j}),$$

$$\textit{Third-order moments:} \qquad \left(\mathcal{M}_t^{(3)}\right)_{i+1,j+1,k+1} = \mathbb{E}(X_{t,i}X_{t,j}X_{t,k}),$$

where $i, j, k \in \{1, \ldots, D\}$.

The moment-change model (1) is therefore equivalent to

$$\mathcal{M}_t^{(p)} \ne \mathcal{M}_{t+1}^{(p)} \quad \text{if and only if} \quad t \in \{\eta_1, \ldots, \eta_K\}, \qquad t = 1, \ldots, n-1. \tag{3}$$

The difficulty of detecting a change depends on its magnitude and its separation from neighboring changes. For $k = 1, \ldots, K$, define the jump tensor

$$\Theta_k = \mathcal{M}_{\eta_k}^{(p)} - \mathcal{M}_{\eta_k+1}^{(p)} \tag{4}$$

and its magnitude

$$\kappa_k = \|\Theta_k\|_{\mathrm{op}}, \qquad k = 1, \ldots, K, \tag{5}$$

and let $\kappa_{\min} = \min_{1 \le k \le K} \kappa_k > 0$. Set $\eta_0 = 0$ and $\eta_{K+1} = n$. The segment lengths are

$$\Delta_k = \eta_{k+1} - \eta_k, \qquad k = 0, \ldots, K, \tag{6}$$

and the minimum spacing is $\Delta_{\min} = \min_{0 \le k \le K} \Delta_k$. Our goal is to estimate K and the locations $\eta_1, \ldots, \eta_K$, and to perform statistical inference on each change-point location. We impose the following assumption to allow temporal dependence in the observed time series.

Assumption 1. *Let $X_1, \ldots, X_n$ be a time series in $\mathbb{R}^D$ satisfying the following conditions.*

(A) Sub-Gaussian tails. *There exists a fixed constant $1 \le \sigma < \infty$, independent of n, such that*

$$\sup_{1 \le t \le n} \sup_{v \in \mathbb{S}^{D-1}} \left\| v^\top X_t \right\|_{\psi_2} \le \sigma.$$

(B) α-mixing. *The α-mixing coefficients (see the definition in (19)) of $\{X_t\}_{t=1}^n$ decay at the generalized-exponential rate*

$$\alpha_X(\ell) \le b_0 \exp(-b_1 \ell^\gamma), \qquad 1 \le \ell < n, \tag{7}$$

where $b_0, b_1 > 0$ and $\gamma \in [1, \infty]$ are fixed constants.

The case $\gamma = \infty$ corresponds to temporal independence, characterized by $\alpha_X(\ell) = 0$ for every positive lag. Under Assumption 1, Y_t has the same mixing coefficients as X_t and a sub-Gaussian scale bounded by a universal constant times σ.

1.1 Related literature

Moment-based methods often target means or covariance matrices. Wang & Samworth (2018) use CUSUM projections for multiple sparse mean changes under independence. For a single change under temporal dependence, their localization bound achieves the minimax optimality. Wang et al. (2021) use independent projections for multiple covariance changes under independence and achieving minimax optimal localization error. Dette et al. (2022) screen covariance entries and prove consistency for a single change under independence, assuming the mean is constant. For broader distributional changes, Kanrar et al. (2025) use classifier AUC to test for and locate both single change

and multiple changes under independence. Their localization error bound additionally requires negligible classifier error. Kernel-based methods for detecting distributional changes include Arlot et al. (2019) and Madrid Padilla et al. (2023). When changes take a parametric form, such as a shift in the mean or covariance, general-purpose kernel methods can be less effective than parametric methods. Indeed, Madrid Padilla et al. (2022) establish the minimax localization rate for kernel density-based change-point detection whose dependence on the observation dimension is exponential, illustrating the curse of dimensionality.

A closely related approach is developed by Cui et al. (2025) for the simultaneous detection of mean and covariance changes. They construct separate U-statistic scans for mean and covariance, and combine their p-values using Fisher's method. For location estimation, they compute the Fisher statistic at each candidate split and choose its maximizer. Their main focus is testing under independence and a single change point. They establish consistency of the location estimator but no localization rate or location limiting distribution. In contrast, our method addresses multiple changes in joint moments through order p under temporal dependence, with minimax-optimal localization and distributional inference for the change locations.

Change-point location limiting distributions have been established for multiple univariate mean changes with i.i.d. errors, together with bootstrap confidence intervals (Cho & Kirch, 2022); single functional mean changes (Aue et al., 2009); and multiple covariance changes in fragmented functional data (Xue et al., 2026). Under temporal dependence, change-point location limits are available for sparse high-dimensional regression (Xu et al., 2024), functional regression (Kumar et al., 2024). We derive change-point location limit laws for nonvanishing and vanishing regime under temporal dependence and construct data-driven confidence intervals with asymptotically valid coverage in the vanishing regime.

1.2 Contributions

1. **Simultaneous detection across moment orders.** Distributional changes can affect different moments. We develop a unified procedure that detects changes in all joint moments through any prescribed fixed order p. The framework can capture mean and covariance changes jointly and extend simultaneous to higher-order changes.

2. **Joint dimension reduction for moment-based detection.** Higher-order moments reveal more complex distributional features but become increasingly difficult to estimate in high dimensions. Our tensor representation brings different moment orders into a unified linear algebra framework, allowing us to construct powerful test statistics that combine information across moment orders to detect changes. The resulting method enables precise change-point localization at a rate that explicitly reflects the moment order p.

3. **Minimax-optimal localization.** We establish the localization rate and prove a newly developed minimax lower bound that matches this rate. These results reveal the statistical difficulty of locating changes using higher-order moment information.

4. **Data-driven confidence interval.** We quantify uncertainty in the estimated change-point locations by deriving limiting distributions for both vanishing and nonvanishing jumps under temporal dependence, see Appendix A. For vanishing regime, we further construct data-driven confidence inter- vals with asymptotically valid coverage, as established in Theorem 6.

1.3 Notation

For integer $0 \leq s < e \leq n$, write $(s, e] = \{s+1, \ldots, e\}$. We use $\|\cdot\|_2$ for the Euclidean norm, $\mathbb{S}^{d-1}$ for the unit sphere in $\mathbb{R}^d$, and $\otimes$ for the outer product, with $u^{\otimes p}$ denoting p identical factors. The Frobenius inner product and norm are $\langle \cdot, \cdot \rangle_{\mathrm{F}}$ and $\|\cdot\|_{\mathrm{F}}$. For an order-p tensor $\mathcal{A} \in \mathbb{R}^{d \times \cdots \times d}$, define

$$\|\mathcal{A}\|_{\mathrm{op}} = \sup_{u_1, \ldots, u_p \in \mathbb{S}^{d-1}} \left| \langle \mathcal{A}, u_1 \otimes \cdots \otimes u_p \rangle_{\mathrm{F}} \right|.$$

The tail scale $1 \leq \sigma < \infty$ is fixed independently of n. Unless stated otherwise, positive constants $C, C_1, C_2, \ldots$ are independent of n, D but may depend on p, γ, b_0, b_1. Moment and tail norms, mixing coefficients, and other conventions are given in Additional notation in the appendix.

2 CHANGE-POINT LOCALIZATION

We propose a two-stage procedure for change-point localization. The first stage, discussed in Section 2.1, evaluates tensor CUSUM statistics over seeded intervals to estimate the number of changes and obtain preliminary locations. The second stage, discussed in Section 2.2, uses these preliminary locations as proxies to estimate the adjacent moment tensors and their leading jump directions. These estimates guide a local refinement step that yields more precise change-point locations.

2.1 PRELIMINARY SEEDED DETECTION

We begin by defining the empirical moment tensor on an interval $(s, e]$,

$$\overline{\mathcal{M}}_{(s,e]}^{(p)} = \frac{1}{e-s} \sum_{t=s+1}^{e} Y_t^{\otimes p}.$$

For any candidate $t \in (s, e)$, define the tensor CUSUM statistic

$$\mathcal{T}_{(s,e]}(t) = \sqrt{\frac{(t-s)(e-t)}{e-s}} \left\{ \overline{\mathcal{M}}_{(s,t]}^{(p)} - \overline{\mathcal{M}}_{(t,e]}^{(p)} \right\}.$$

When the population moment tensor is constant on $(s, e]$, we have $\mathbb{E}\mathcal{T}_{(s,e]}(t) = 0$. If the interval contains a single change point $\eta \in (s, e-1]$, then $\left\|\mathbb{E}\mathcal{T}_{(s,e]}(t)\right\|_{\mathrm{op}}$ is uniquely maximized at $t = \eta$. These properties allow us to use $\left\|\mathcal{T}_{(s,e]}(t)\right\|_{\mathrm{op}}$ to measure the signal of a change.

To efficiently locate the change points using the tensor cusum statistics, we evaluate these statistics over a deterministic collection of overlapping intervals at multiple scales, called seeded intervals (Kovács et al., 2023). Define dyadic interval scale set as

$$\mathcal{L}_n = \{2^r : r = 4, \ldots, \lfloor \log_2 n \rfloor\}.$$

The intervals at scale $\ell \in \mathcal{L}_n$ are

$$\mathcal{I}_\ell = \left\{ \left(\frac{j\ell}{2}, \frac{j\ell}{2} + \ell \right] : j = 0, \ldots, \left\lfloor \frac{2(n-\ell)}{\ell} \right\rfloor \right\} \cup \{(n-\ell, n]\}.$$

At scale ℓ, $\mathcal{I}_\ell$ contains intervals of length ℓ starting at equally spaced grid points $0, \ell/2, \ell, \ldots$, together with the final interval $(n-\ell, n]$. Taking the union over scales gives the seeded collection

$$\mathcal{I} = \bigcup_{\ell \in \mathcal{L}_n} \mathcal{I}_\ell. \tag{8}$$

Seeded intervals have proved effective for univariate mean change detection (Kovács et al., 2023). We adapt this construction to detect changes in multivariate moment tensors.

To obtain preliminary change-point estimates, Algorithm 1 applies narrowest-over-threshold selection (Baranowski et al., 2019) to the tensor CUSUM statistics on seeded intervals. The statistic $G_{(s,e]}$, defined in (9), measures the strongest evidence of a moment change within the central half of $(s, e]$. This restriction avoids splits near interval boundaries, where empirical higher-order moments can be unstable because of few observations. The threshold in (10) adjusts for interval length, moment order, and temporal dependence, with the tuning parameter τ controlling detection sensitivity.

Among intervals exceeding the threshold, Algorithm 1 selects a shortest interval to favor isolating an individual change. Within the selected interval, we estimate the change point by the split with the largest test statistic. We then remove overlapping intervals to avoid detecting the same change. The resulting locations initialize the second-stage refinement. We next establish conditions under which the preliminary procedure consistently estimates the number and locations of the change points.

Assumption 2. *For a sufficiently large constant $C_{\mathrm{I}} > 0$,*

$$\frac{\Delta_{\min}\kappa_{\min}^2}{\sigma^{2p}} \geq C_{\mathrm{I}}(D + \log n), \qquad \frac{\Delta_{\min}\kappa_{\min}}{\sigma^p} \geq \sqrt{C_{\mathrm{I}}}\,(D + \log n)^{p/2+1/\gamma}.$$

Algorithm 1 Preliminary seeded detection

Require: Observations $X_1, \dots, X_n$, moment order p, and parameters $\gamma, \tau > 0$.

1: $Y_t \leftarrow (1, X_t^\top)^\top$, $t = 1, \dots, n$.
2: Construct $\mathcal{I}$ according to (8).
3: **for** $(s, e] \in \mathcal{I}$ **do**
4: Compute

$$G_{(s,e]} \leftarrow \max_{s+(e-s)/4 \le t \le e-(e-s)/4} \left\| \mathcal{T}_{(s,e]}(t) \right\|_{\mathrm{op}}. \tag{9}$$

5: **end for**
6: Set

$$\mathcal{I}_{\mathrm{act}} \leftarrow \left\{ (s, e] \in \mathcal{I} : G_{(s,e]} > \tau \left[\sqrt{D + \log n} + \frac{(D + \log n)^{p/2+1/\gamma}}{\sqrt{e - s}} \right] \right\}. \tag{10}$$

7: $\widetilde{\mathcal{E}} \leftarrow \varnothing$.
8: **while** $\mathcal{I}_{\mathrm{act}} \neq \varnothing$ **do**
9: Choose the shortest $(s, e] \in \mathcal{I}_{\mathrm{act}}$.
10: $\widetilde{\eta} \leftarrow \min \arg\max_{s+(e-s)/4 \le t \le e-(e-s)/4} \left\| \mathcal{T}_{(s,e]}(t) \right\|_{\mathrm{op}}$.
11: $\widetilde{\mathcal{E}} \leftarrow \widetilde{\mathcal{E}} \cup \{\widetilde{\eta}\}$.
12: $\mathcal{I}_{\mathrm{act}} \leftarrow \{(a, b] \in \mathcal{I}_{\mathrm{act}} : (a, b] \cap (s, e] = \varnothing\}$.
13: **end while**
14: **return** The estimated change points $\widetilde{\eta}_1 < \cdots < \widetilde{\eta}_{\widetilde{K}}$ of $\widetilde{\mathcal{E}}$ in increasing order.

Under temporal independence setting that $\gamma = \infty$, if $\Delta_{\min} \ge (D + \log n)^{p-1}$, then Assumption 2 reduces to

$$\frac{\Delta_{\min} \kappa_{\min}^2}{\sigma^{2p}} \ge C_{\mathrm{I}}(D + \log n). \tag{11}$$

For $p = 2$, (11) up to logarithmic factors, matches the covariance change-point detection condition of Wang et al. (2021), and is minimax optimal. In addition, for $p = 2$, the requirement $\Delta_{\min} \gtrsim D$ also appears in Wang et al. (2021).

Theorem 1. *Let $\{\widetilde{\eta}_k\}_{k=1}^{\widetilde{K}}$ be the estimated change points return by Algorithm 1 with input γ and $\tau = c_\tau \sigma^p$, where $c_\tau > 0$ is sufficiently large. Suppose Assumptions 1 and 2 hold. Then, with probability at least $1 - n^{-5}$, Algorithm 1 returns estimates satisfying*

$$\widetilde{K} = K, \qquad \max_{1 \le k \le K} |\widetilde{\eta}_k - \eta_k| \le \frac{3\Delta_{\min}}{64}. \tag{12}$$

Remark 2. *The constant $3/64$ in Theorem 1 is chosen for convenience. The preliminary estimates need only be sufficiently accurate relative to the minimum spacing to ensure that there is a one to one correpsondece between the estimate change points and the true ones. Theorem 1 provides sufficient accuracy for the subsequent refinement procedure.*

2.2 Local refinement

To obtain optimal change-point estimates, we use the preliminary estimates $\{\widetilde{\eta}_k\}_{k=1}^{K}$ from Algorithm 1 to estimate the moment jumps Θ_k defined in (4). We then estimate the most significant projection direction of each jump,

$$\nu_k \in \arg\max_{u \in \mathbb{S}^D} \left| \left\langle \Theta_k, u^{\otimes p} \right\rangle_{\mathrm{F}} \right|, \tag{13}$$

and use the estimated projection to refine the change-point location.

Set $\widetilde{\eta}_0 = 0$ and $\widetilde{\eta}_{\widetilde{K}+1} = n$. To avoid uncertain segment boundaries, we fit moments on the central half $(a_k, b_k]$ of each estimated segment, where, for $k = 0, \dots, \widetilde{K}$,

$$a_k = \left\lceil \frac{3\widetilde{\eta}_k + \widetilde{\eta}_{k+1}}{4} \right\rceil - 1, \qquad b_k = \left\lfloor \frac{\widetilde{\eta}_k + 3\widetilde{\eta}_{k+1}}{4} \right\rfloor. \tag{14}$$

Under (12), $\widetilde{K} = K$ and $[a_k, b_k] \subseteq (\eta_k, \eta_{k+1}]$ for $k = 0, \ldots, K$. For $k = 1, \ldots, K$, estimate the moment jump by

$$\widehat{\Theta}_k = \frac{1}{b_{k-1} - a_{k-1}} \sum_{t=a_{k-1}+1}^{b_{k-1}} Y_t^{\otimes p} - \frac{1}{b_k - a_k} \sum_{t=a_k+1}^{b_k} Y_t^{\otimes p}. \tag{15}$$

We then estimate ν_k in (13) by

$$\widehat{\nu}_k \in \arg\max_{u \in \mathbb{S}^D} \left| \left\langle \widehat{\Theta}_k, u^{\otimes p} \right\rangle_{\mathrm{F}} \right|. \tag{16}$$

Under standard identifiability and eigengap-type conditions on Θ_k (see Assumption 7 in the appendix), we establish an error bound for the estimated projection directions in Section F.3 of the appendix. For $k = 1, \ldots, K$, we refine the location within $(s_k, e_k]$, with endpoints

$$s_k = \left\lfloor \frac{9\widetilde{\eta}_{k-1} + \widetilde{\eta}_k}{10} \right\rfloor, \qquad e_k = \left\lfloor \frac{\widetilde{\eta}_k + 9\widetilde{\eta}_{k+1}}{10} \right\rfloor.$$

Under (12), each window $(s_k, e_k]$ contains exactly one true change point, η_k. For a candidate $t \in (s_k, e_k - 1]$, define

$$\widehat{Q}_k(t) = \sum_{i=s_k+1}^{t} \left\{ (\widehat{\nu}_k^\top Y_i)^p - \frac{\sum_{j=a_{k-1}+1}^{b_{k-1}} (\widehat{\nu}_k^\top Y_j)^p}{b_{k-1} - a_{k-1}} \right\}^2 + \sum_{i=t+1}^{e_k} \left\{ (\widehat{\nu}_k^\top Y_i)^p - \frac{\sum_{j=a_k+1}^{b_k} (\widehat{\nu}_k^\top Y_j)^p}{b_k - a_k} \right\}^2.$$

The refined estimate is

$$\widehat{\eta}_k = \min \arg\min_{t \in (s_k, e_k - 1]} \widehat{Q}_k(t). \tag{17}$$

To obtain theoretical guarantees of the refined change point estimators $\{\widehat{\eta}_k\}_{k=1}^K$, we impose the following signal-strength conditions.

Assumption 3. *As $n \to \infty$,*

$$\frac{\Delta_{\min} \kappa_{\min}^2}{\sigma^{2p} D(D + \log n)} \to \infty \quad \textit{and} \quad \frac{\Delta_{\min} \kappa_{\min}}{\sigma^p \sqrt{D}\,(D + \log n)^{p/2 + 1/\gamma}} \to \infty.$$

Assumption 3 strengthens Assumption 2 to ensure accurate estimation of the order-p moment tensors relative to the jump magnitudes. The conditions allow both a growing dimension and vanishing jump magnitudes, provided the segments contain sufficiently many observations to estimate the higher-order moments accurately.

Theorem 2. *Suppose the moment-change model* (3) *holds, $p \geq 2$ is fixed, and Assumptions 1, 3 and 7 hold. Use Algorithm 1 with $\tau = c_\tau \sigma^p$ for a sufficiently large fixed constant c_τ. Then the refined estimators in* (17) *satisfy*

$$\max_{1 \leq k \leq K} \frac{\kappa_k^2}{\sigma^{2p}} |\widehat{\eta}_k - \eta_k| = O_{\mathbb{P}}(1).$$

Theorem 2 shows that the refined localization error for each change point is $O_{\mathbb{P}}(\sigma^{2p}/\kappa_k^2)$. The constants and signal requirements depend on the fixed order p. Under additional assumptions, Appendix A establishes limiting distributions for the refined estimators and constructs asymptotically valid confidence intervals for vanishing jumps.

2.3 Minimax lower bound

We establish a minimax lower bound matching the localization rate in Theorem 2. For positive integers n, D, Δ with $2\Delta \leq n$ and $\kappa > 0$, let $\mathcal{P}(n, D, \Delta, \sigma, \kappa)$ denote the class of joint distributions of independent observations with the fixed tail scale σ from Assumption 1, satisfying

$$X_1, \ldots, X_\eta \overset{\text{i.i.d.}}{\sim} F_L, \qquad X_{\eta+1}, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F_R,$$

where F_L, F_R are distributions on $\mathbb{R}^D$ and the unknown change point satisfies $\Delta \le \eta \le n - \Delta$. The observations satisfy the uniform sub-Gaussian bound

$$\sup_{1\le t\le n} \sup_{u\in\mathbb{S}^{D-1}} \left\|u^\top X_t\right\|_{\psi_2} \le \sigma.$$

Define the two population moment tensors by

$$\mathcal{M}_L^{(p)} = \mathbb{E}_{X\sim F_L}\left[Y^{\otimes p}\right], \qquad \mathcal{M}_R^{(p)} = \mathbb{E}_{X\sim F_R}\left[Y^{\otimes p}\right],$$

where $Y = (1, X^\top)^\top$. Suppose that $\left\|\mathcal{M}_L^{(p)} - \mathcal{M}_R^{(p)}\right\|_{\mathrm{op}} = \kappa$.

Theorem 3. *Fix an integer $p \ge 2$ and a constant $1 \le \sigma < \infty$. Let $D = D_n$ and $\Delta = \Delta_n$ be positive integers with $\Delta \le n/4$, and let $\kappa = \kappa_n > 0$. Suppose Assumption 3 holds with $\Delta_{\min} = \Delta$, $\kappa_{\min} = \kappa$, and $\gamma = \infty$. There exist constants $c_0, c_1 > 0$, depending only on p, such that, whenever $0 < \kappa \le c_0\sigma^p$, the following bound holds for all sufficiently large n:*

$$\inf_{\check\eta} \sup_{P\in\mathcal{P}(n,D,\Delta,\sigma,\kappa)} \mathbb{E}_P\left|\check\eta - \eta\right| \ge c_1\frac{\sigma^{2p}}{\kappa^2}, \tag{18}$$

where the infimum ranges over all estimators based on $X_1, \ldots, X_n$.

3 Simulation and real data example

In this section, we study the numerical properties of our method and compare it with state-of-the-art methods targeting mean, covariance, and general distributional changes. In particular, we compare our method with CPWZ (Cui et al., 2025), changeAUC (Kanrar et al., 2025), a mean-change binary segmentation method, and MNSBS (Madrid Padilla et al., 2023) across various simulated settings and a real-data example. CPWZ jointly detects changes in the mean and covariance by combining statistics for the two types of changes. ChangeAUC detects general distributional changes by training a classifier and using its area under the receiver operating characteristic curve as the change statistic. The mean-change method applies binary segmentation to multivariate mean CUSUM statistics. MNSBS detects general multivariate distributional changes under temporal dependence through nonparametric density estimation and seeded binary segmentation.

3.1 Tuning parameter selection

For competing methods, we use default or paper-recommended tuning and their associated data-driven selection procedures. For Tensor, we fix $\gamma = 1$ and restrict the moment order to the small candidate set $p \in \{2, 3\}$. We jointly select p and τ by odd–even validation: estimate and refine change points on the odd-indexed observations, then score the resulting segments on the even-indexed observations. The pair minimizing the score is used for the full time series. Appendix G.1 gives the candidate grid and kernel score.

3.2 Simulation studies

We use $D = 100$ and 100 independent repetitions per setting. With independent standard normal initial values Z_{1j} and innovations ε_{tj}, generate stationary AR(1) sequences

$$Z_{tj} = \rho Z_{t-1,j} + \sqrt{1-\rho^2}\,\varepsilon_{tj}, \qquad t = 2, \ldots, n, \quad j = 1, \ldots, D.$$

Case 1: a single change point. Settings R1–R3 have $n = 1600$ and a single change at $\eta = 880$.

R1 (variance and skewness changes). With $\rho = 0.6$, set $X_{tj} = Z_{tj} + 0.25(Z_{tj}^2 - 1)$ for $t > \eta$ and $j \le 10$, and $X_{tj} = Z_{tj}$ otherwise. The mean is unchanged.

R2 (skewness changes). With $\rho = 0.4$, let W be an independent copy of Z. For $j \le 8$, define

$$B_{tj} = \mathbf{1}\{W_{tj} < \Phi^{-1}(0.2)\}, \qquad Y_{tj} = \frac{B_{tj} - 0.2}{\sqrt{0.2(1-0.2)}} + 0.3Z_{tj},$$

Table 1: Mean Hausdorff distance (SD) and percentages $[\widehat{K} < K, \widehat{K} = K, \widehat{K} > K]$ over 100 repetitions. Tensor reports refined estimates; lowest means are bold. Dashes: not applicable.

Setting	Moment Tensor	CPWZ	changeAUC	Mean-change	MNSBS
R1	**227.81** (180.77) $[0, 100, 0]$	304.43 (151.49) $[0, 100, 0]$	611.18 (152.48) $[61, 39, 0]$	510.53 (236.89) $[0, 100, 0]$	391.14 (219.32) $[0, 100, 0]$
R2	**7.15** (8.27) $[0, 100, 0]$	343.62 (152.84) $[0, 100, 0]$	665.37 (114.75) $[79, 21, 0]$	549.07 (228.91) $[0, 100, 0]$	416.62 (231.35) $[2, 98, 0]$
R3	**10.95** (19.33) $[0, 100, 0]$	49.95 (79.17) $[0, 100, 0]$	585.57 (181.78) $[58, 42, 0]$	356.99 (268.00) $[0, 100, 0]$	410.56 (228.78) $[1, 99, 0]$
R4	**1.82** (2.25) $[0, 100, 0]$	—	17.09 (24.09) $[0, 99, 1]$	12.19 (56.40) $[0, 97, 3]$	295.60 (103.20) $[5, 11, 84]$
R5	**1.00** (1.53) $[0, 100, 0]$	—	23.69 (33.24) $[0, 94, 6]$	57.53 (97.82) $[3, 65, 32]$	134.26 (154.40) $[0, 50, 50]$

and set $X_{tj} = Y_{tj}$ for $t \le \eta$ and $X_{tj} = -Y_{tj}$ for $t > \eta$. For $j > 8$, set $X_{tj} = Z_{tj}$. Reversing sign changes skewness but preserves the mean vector and covariance matrix.

R3 (sparse mean and variance changes). With $\rho = 0.4$, set $X_{t1} = 1.75Z_{t1} + 0.3$ for $t > \eta$; all other observations satisfy $X_{tj} = Z_{tj}$. Only one coordinate changes.

Case 2: multiple change points. Settings R4–R5 have $n = 1200$ and two changes. We omit CPWZ, which targets a single change point.

R4 (unequal jumps). Set $\rho = 0.2$ and $(\eta_1, \eta_2) = (400, 800)$. For $j \le 10$, let

$$X_{tj} = \begin{cases} Z_{tj}, & t \le \eta_1, \\ 0.65 + 1.3Z_{tj}, & \eta_1 < t \le \eta_2, \\ 1.55 + 1.8Z_{tj}, & t > \eta_2. \end{cases}$$

For $j > 10$, set $X_{tj} = Z_{tj}$. Both mean and variance increase, with unequal jump sizes at equally spaced locations.

R5 (unequal spacings). With $\rho = 0.1$ and $(\eta_1, \eta_2) = (450, 950)$, let H be an independent copy of Z. For $j \le 30$, define

$$U_{tj} = \frac{2 \cdot \mathbf{1}\{H_{tj} < 0\} - 1 + 0.8Z_{tj}}{\sqrt{1.64}}, \qquad X_{tj} = \begin{cases} 0.35 + 2U_{tj}, & \eta_1 < t \le \eta_2, \\ U_{tj}, & \text{otherwise.} \end{cases}$$

For $j > 30$, set $X_{tj} = Z_{tj}$. The Gaussian mixture undergoes a temporary mean and scale shift, yielding equal jump sizes and unequal segment lengths.

Table 1 shows that Tensor recovers the correct number of changes in all 100 repetitions of each setting, and its refined estimates attain the lowest mean Hausdorff distance. In R2 and R3, the mean distances are 7.15 and 10.95, compared with 343.62 and 49.95 for CPWZ, the closest competitor. R1 remains harder to localize, as the moment signal is weaker relative to its variability: on each affected coordinate, the variance and third-moment jumps are 0.125 and 1.625, whereas the third-moment jump in R2 is 3. These differences make R1 harder to localize even though Tensor detects the change in every repetition. For the two-change settings R4 and R5, Tensor attains mean distances of 1.82 and 1.00; competing methods have larger localization errors and less reliable estimates of the number of changes. The simulation-calibrated 95% intervals have empirical coverage of 95% in R4 and 96% in R5 at each change point, conditional on $\widehat{K} = K = 2$ (Appendix G.3).

3.3 Real data analysis

We analyze two panels of stocks that remained in the Nasdaq-100 during their respective study periods: February 1, 2007 through December 31, 2010, and January 2, 2018 through December 30, 2022. The respective panels contain $n = 988$ observations for $D = 58$ stocks and $n = 1258$

Table 2: Preliminary and refined Tensor change-point locations for moment orders $p = 2$ and $p = 3$ in the two Nasdaq-100 stock panels. Each configuration returns one estimated change in each period.

	$p = 2$		$p = 3$	
Period	Preliminary	Refined	Preliminary	Refined
2007–2010	Sep. 25, 2008	Sep. 15, 2008	Oct. 13, 2008	Oct. 10, 2008
2018–2022	Jan. 28, 2020	Feb. 21, 2020	Jan. 28, 2020	Feb. 21, 2020

observations for $D = 59$ stocks. For stock j, we calculate $r_{tj} = \log(P_{tj}/P_{t-1,j})$, where P_{tj} is its closing price on retained trading day t (Quantiacs). We analyze $X_{tj} = (r_{tj} - \bar{r}_j)/s_j$, where $\bar{r}_j$ and s_j are the mean and standard deviation of stock j's returns over that period. All methods use the same standardized matrix $X \in \mathbb{R}^{n \times D}$ within each period.

We follow the joint (p, τ) tuning framework of Section 3.1, using the candidate grid in Appendix G.1 and the implementation's moment-kernel validation score. It selects $(p, \tau) = (2, 16)$ in both periods. We also report the preliminary and refined locations obtained with $p = 3$ for comparison. Table 2 lists the Tensor locations, and Figure 1 displays the daily cross-sectional RMS $R_t = (D^{-1} \sum_{j=1}^{D} X_{tj}^2)^{1/2}$ and its average over the current and up to 20 preceding trading days. Peaks of the smoothed curve in November 2008 and April 2020 indicate sustained elevated return magnitudes; a peak may follow the initial shock and need not coincide with an estimated change-point date. For $p = 2$, refinement moves the earlier estimate from September 25 to September 15, 2008, the date of the Lehman Brothers bankruptcy filing, a pivotal event in the escalation of the financial crisis (U.S. Department of the Treasury). In the later period, the refined estimate is February 21, 2020, in line with the onset of the market reversal associated with the international spread of COVID-19 (The Market Intelligence Desk Team & MarketInsite, 2020).

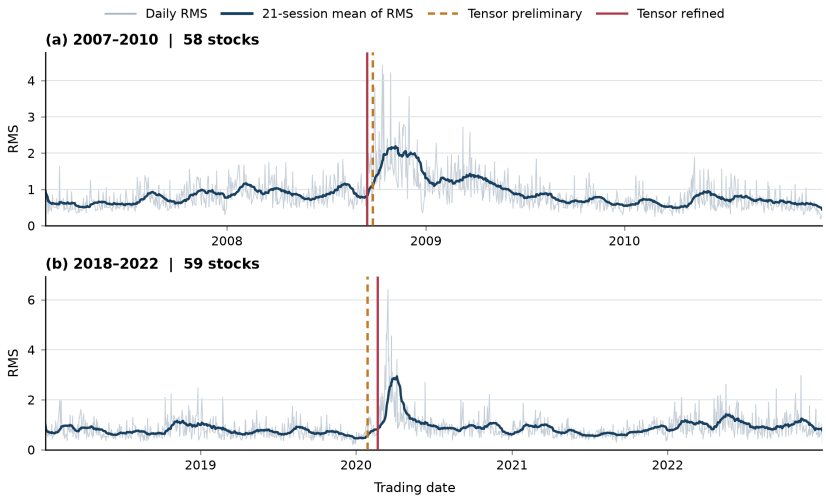


Figure 1: Daily cross-sectional RMS R_t (light curves) and its trailing 21-session average (dark curves) for the two Nasdaq-100 stock panels. Dashed and solid vertical lines mark the $p = 2$ preliminary and refined dates, respectively.

MNSBS estimates changes on July 17, 2007; August 29, 2008; August 18, 2009; February 26, April 27, and December 22, 2010, in the first panel, and on January 23, 2020 and September 27, 2021 in the second. In a recorded two-change changeAUC run for each panel, its respective dates are November 13, 2008 and April 21, 2009; and February 4, 2020 and December 30, 2021. Other recorded changeAUC runs return up to three dates in the later panel. These alternative estimates support the broader finding of changing market behavior during the 2008 crisis and the early COVID-19 disruption, although their exact locations differ from the Tensor estimates.

AI USE STATEMENT

We used generative AI tools to assist with writing and debugging code for simulations and real-data analyses, checking references, and polishing the writing of the manuscript. The authors take responsibility for all code, analyses, and text in this work.

REFERENCES

Joshua Abramson and Steven N. Evans. Lipschitz minorants of Brownian motion and Lévy processes. *Probability Theory and Related Fields*, 158(3–4):809–857, 2014.

Sylvain Arlot, Alain Celisse, and Zaid Harchaoui. A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research*, 20(162):1–56, 2019. URL `https://www.jmlr.org/papers/v20/16-155.html`.

Alexander Aue, Robertas Gabrys, Lajos Horváth, and Piotr Kokoszka. Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, 100(10):2254–2269, 2009.

Rafał Baranowski, Yining Chen, and Piotr Fryzlewicz. Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(3):649–672, 2019.

Haeran Cho and Claudia Kirch. Bootstrap confidence intervals for multiple change points based on moving sum procedures. *Computational Statistics & Data Analysis*, 175:107552, 2022.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006.

Junfeng Cui, Guangming Pan, Guanghui Wang, and Changliang Zou. Simultaneous detection and localization of mean and covariance changes in high dimensions. arXiv:2508.19523, 2025.

Yu. A. Davydov. Convergence of distributions generated by stationary stochastic processes. *Theory of Probability and Its Applications*, 13(4):691–696, 1968. English translation of the Russian original, pp. 730–737.

Jonas Dehning, Johannes Zierenberg, F. Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*, 369(6500):eabb9789, 2020.

Holger Dette, Guangming Pan, and Qing Yang. Estimating a change point in a sequence of very high-dimensional covariance matrices. *Journal of the American Statistical Association*, 117(537): 444–454, 2022.

Shmuel Friedland and Lek-Heng Lim. Nuclear norm of higher-order tensors. *Mathematics of Computation*, 87(311):1255–1281, 2018.

Stephanie L. Hyland, Martin Faltys, Matthias Hüser, Xinrui Lyu, Thomas Gumbsch, Cristóbal Esteban, Christian Bock, Max Horn, Michael Moor, Bastian Rieck, Marc Zimmermann, Dean Bodenham, Karsten Borgwardt, Gunnar Rätsch, and Tobias M. Merz. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature Medicine*, 26(3):364–373, 2020.

Ariel Jaffe, Roi Weiss, and Boaz Nadler. Newton correction methods for computing real eigenpairs of symmetric tensors. *SIAM Journal on Matrix Analysis and Applications*, 39(3):1071–1094, 2018. doi: 10.1137/17M1133312.

Rohit Kanrar, Feiyu Jiang, and Zhanrui Cai. Model-free change-point detection using AUC of a classifier. *Journal of Machine Learning Research*, 26(190):1–50, 2025.

Jeankyung Kim and David Pollard. Cube root asymptotics. *The Annals of Statistics*, 18(1):191–219, 1990.

Tamara G. Kolda and Jackson R. Mayo. Shifted power method for computing tensor eigenpairs. *SIAM Journal on Matrix Analysis and Applications*, 32(4):1095–1124, 2011.

Solt Kovács, Peter Bühlmann, Housen Li, and Axel Munk. Seeded binary segmentation: a general methodology for fast and optimal changepoint detection. *Biometrika*, 110(1):249–256, 2023.

Shivam Kumar, Haotian Xu, Haeran Cho, and Daren Wang. Estimation and inference for change points in functional regression time series. arXiv:2405.05459, 2024. Version 1; forthcoming in Statistica Sinica.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023.

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*, 2024.

Carlos Misael Madrid Padilla, Haotian Xu, Daren Wang, Oscar Hernan Madrid Padilla, and Yi Yu. Change point detection and inference in multivariate non-parametric models under mixing conditions. In *Advances in Neural Information Processing Systems*, volume 36, pp. 21081–21134, 2023.

Oscar Hernan Madrid Padilla, Yi Yu, Daren Wang, and Alessandro Rinaldo. Optimal nonparametric multivariate change point detection and localization. *IEEE Transactions on Information Theory*, 68(3):1922–1944, 2022. doi: 10.1109/TIT.2021.3130330. URL https://arxiv.org/pdf/1910.13289.

Florence Merlevède, Magda Peligrad, and Emmanuel Rio. A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151(3–4): 435–474, 2011.

Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

Yang Qi, Pierre Comon, and Lek-Heng Lim. Uniqueness of nonnegative tensor approximations. *IEEE Transactions on Information Theory*, 62(4):2170–2183, 2016. doi: 10.1109/TIT.2016.2532906.

Quantiacs. Financial data. URL https://quantiacs.com/documentation/en/user_guide/data.html. Accessed 24 September 2026.

Ingo Steinwart and Johanna F. Ziegel. Strictly proper kernel scores and characteristic kernels on compact spaces. *Applied and Computational Harmonic Analysis*, 51:510–542, 2021. doi: 10.1016/j.acha.2019.11.005.

The Market Intelligence Desk Team and MarketInsite. February 2020 review and outlook. Nasdaq, 3 March 2020. URL https://www.nasdaq.com/articles/february-2020-review-and-outlook-2020-03-03. Accessed 24 September 2026.

U.S. Department of the Treasury. About TARP. URL https://home.treasury.gov/data/troubled-assets-relief-program/about-tarp. Accessed 24 September 2026.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, New York, 1996.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018.

Daren Wang, Yi Yu, and Alessandro Rinaldo. Optimal covariance change point localization in high dimensions. *Bernoulli*, 27(1):554–575, 2021.

Tengyao Wang and Richard J. Samworth. High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(1):57–83, 2018.

C. S. Withers. Central limit theorems for dependent variables. I. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57(4):509–534, 1981. Corrigendum: volume 63, page 555, 1983.

Haotian Xu, Daren Wang, Zifeng Zhao, and Yi Yu. Change-point inference in high-dimensional regression models under temporal dependence. *The Annals of Statistics*, 52(3):999–1026, 2024.

Gengyu Xue, Haotian Xu, and Yi Yu. Covariance change point localisation and inference in fragmented functional data. *Bernoulli*, 32(2):1456–1480, 2026.

ADDITIONAL NOTATION

Vectors. For $v \in \mathbb{R}^d$, let v_i denote its ith coordinate and $\|v\|_2$ its Euclidean norm. Write $\mathbb{S}^{d-1} = \{v \in \mathbb{R}^d : \|v\|_2 = 1\}$. Throughout the paper, for $u, v \in \mathbb{S}^{d-1}$, the distance up to sign is $d_{\pm}(u, v) = \min\{\|u - v\|_2, \|u + v\|_2\}$.

Matrices and tensors. Matrix and tensor entries are denoted by A_{ij} and $\mathcal{A}_{i_1,\dots,i_p}$. Write $u_1 \otimes \cdots \otimes u_p$ for the outer product of vectors $u_1, \dots, u_p$, and $u^{\otimes p}$ when all factors equal u. The Frobenius inner product $\langle \mathcal{A}, \mathcal{B} \rangle_{\mathrm{F}}$ is the sum of products of corresponding entries, with $\|\mathcal{A}\|_{\mathrm{F}}^2 = \langle \mathcal{A}, \mathcal{A} \rangle_{\mathrm{F}}$.

Moments and tails. For a scalar random variable W, let $\|W\|_{L^q} = (\mathbb{E}\,|W|^q)^{1/q}$, $q \geq 1$. For $\beta > 0$, define

$$\|W\|_{\psi_\beta} = \sup_{q \geq 1} q^{-1/\beta} \|W\|_{L^q}.$$

Temporal dependence. For sigma-fields $\mathcal{A}, \mathcal{B}$, define

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A},\, B \in \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

For $W_1, \dots, W_n$, the α-mixing coefficients are

$$\alpha_W(\ell) = \sup_{1 \leq s \leq n-\ell} \alpha\{\sigma(W_1, \dots, W_s), \sigma(W_{s+\ell}, \dots, W_n)\}, \qquad 1 \leq \ell < n. \tag{19}$$

Asymptotic notation. We use the standard asymptotic notation O, o, $O_{\mathbb{P}}$, and $o_{\mathbb{P}}$, and write $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{d}$ for convergence in probability and distribution.

A LIMITING DISTRIBUTIONS AND CONFIDENCE INTERVALS

We quantify uncertainty in the refined change-point estimates. We establish their limiting distributions under nonvanishing and vanishing moment jumps, and use the vanishing-jump limit to construct data-driven confidence intervals.

A.1 THE NONVANISHING REGIME

We impose the following additional conditions on the observations around η_k.

Assumption 4. ***(A)*** *The moment order p is fixed, and, as $n \to \infty$,*

$$\kappa_k \to \kappa_k^* > 0.$$

The dimension D may grow with n.

(B) *There exists a real-valued process $\{\xi_{k,t}\}_{t \in \mathbb{Z}}$, whose law does not depend on n, such that, for every fixed integer $H \geq 1$,*

$$\big(\nu_k^\top Y_{\eta_k + t}\big)_{t=-H+1}^{H} \xrightarrow{d} \big(\xi_{k,t}\big)_{t=-H+1}^{H}.$$

Here ν_k is defined in (13)*, with its sign chosen consistently.*

(C) *For every fixed integer $H \geq 1$, the vector $(\xi_{k,-H+1}, \dots, \xi_{k,H})$ has a joint density with respect to Lebesgue measure on $\mathbb{R}^{2H}$.*

Condition (A) of Assumption 4 follows the usual formulation of the nonvanishing regime in change-point analysis. Condition (B) holds, for example, when D is fixed, ν_k converges, and the segment $(X_{\eta_{k-1}+1}, \dots, X_{\eta_k}, X_{\eta_k+1}, \dots, X_{\eta_{k+1}})$ admits an infinite extension whose joint law does not depend on n. This latter formulation is commonly used in the change-point literature. Condition (C) is a standard continuity condition that ensures uniqueness of the limiting random-walk minimizer.

Define the population limiting moment jump by

$$\rho_k = \mathbb{E}\big\{\xi_{k,0}^p\big\} - \mathbb{E}\big\{\xi_{k,1}^p\big\}. \tag{20}$$

Under Assumptions 1 and 4, it follows from (20) that $|\rho_k| = \kappa_k^*$. Define the two-sided random walk

$$\mathcal{L}_k(r) = \begin{cases} |r|(\kappa_k^*)^2 + 2\rho_k \sum_{t=r+1}^{0} \left[\xi_{k,t}^p - \mathbb{E}\{\xi_{k,t}^p\}\right], & r < 0, \\ 0, & r = 0, \\ r(\kappa_k^*)^2 - 2\rho_k \sum_{t=1}^{r} \left[\xi_{k,t}^p - \mathbb{E}\{\xi_{k,t}^p\}\right], & r > 0, \end{cases} \qquad r \in \mathbb{Z}. \tag{21}$$

Theorem 4. *Fix $k \in \{1, \dots, K\}$. Suppose Assumptions 1, 3, 4 and 7 hold. Then the refined estimator in* (17) *satisfies*

$$\widehat{\eta}_k - \eta_k \xrightarrow{d} \arg\min_{r \in \mathbb{Z}} \mathcal{L}_k(r).$$

A.2 The vanishing regime

We study the regime in which the moment jump size κ_k, defined in (5), tends to zero. As we show in this subsection, the rescaled localization error $\kappa_k^2(\widehat{\eta}_k - \eta_k)$ converges in distribution to the minimizer of a two-sided Brownian motion with drift. To establish this result, we impose the following assumptions.

Assumption 5. ***(A)*** *The moment order p is fixed, and, as $n \to \infty$,*

$$\kappa_k \to 0.$$

(B) *For each n, the segment $(X_{\eta_{k-1}+1}, \dots, X_{\eta_{k+1}})$ admits an extension indexed by $\mathbb{Z}$ that is strictly stationary separately on $t \le \eta_k$ and $t > \eta_k$. The extended time series satisfies the tail and mixing bounds in Assumption 1, with the mixing bound imposed at every positive lag.*

Using the population direction ν_k in (13), define

$$Z_{k,t} = (\nu_k^\top Y_t)^p - \mathbb{E}\{(\nu_k^\top Y_t)^p\}, \qquad t = 1, \dots, n. \tag{22}$$

For the variance limits below, evaluate the same formula on the two-regime extension in Assumption 5. This defines $Z_{k,t}$ for every $t \in \mathbb{Z}$ and agrees with the observed projection on $(\eta_{k-1}, \eta_{k+1}]$.

Assumption 6. *There exist constants $\omega_{k,L}, \omega_{k,R} \in (0, \infty)$ such that, for every integer sequence $m = m_n \to \infty$, as $n \to \infty$,*

$$\frac{1}{m} \operatorname{Var}\left(\sum_{j=1}^{m} Z_{k,\eta_k+j}\right) \longrightarrow \omega_{k,R}^2 \quad \textit{and} \quad \frac{1}{m} \operatorname{Var}\left(\sum_{j=1}^{m} Z_{k,\eta_k-j+1}\right) \longrightarrow \omega_{k,L}^2.$$

Let $B_{k,L}$ and $B_{k,R}$ be independent standard Brownian motions. Define the two-sided Brownian motion with drift by

$$\mathcal{G}_k(r) = \begin{cases} |r| + 2\omega_{k,L} B_{k,L}(-r), & r < 0, \\ 0, & r = 0, \\ r + 2\omega_{k,R} B_{k,R}(r), & r > 0, \end{cases} \qquad r \in \mathbb{R}. \tag{23}$$

Theorem 5. *Fix $k \in \{1, \dots, K\}$. Suppose Assumptions 1, 3 and 5 to 7 hold. Then the refined estimator in* (17) *satisfies*

$$\kappa_k^2(\widehat{\eta}_k - \eta_k) \xrightarrow{d} \arg\min_{r \in \mathbb{R}} \mathcal{G}_k(r).$$

A.3 Confidence intervals

We propose a data-driven procedure with theoretical guarantees for constructing confidence intervals for η_k based on the refined estimator $\widehat{\eta}_k$. To do so, we first estimate the jump magnitude κ_k. Recall that $(a_k, b_k]$, with endpoints defined in (14), lies within the homogeneous segment $(\eta_k, \eta_{k+1}]$ with

high probability under the conditions of Theorem 1. Using the estimated leading direction $\widehat{\nu}_k$ defined in (16), we estimate κ_k by comparing the projected moment averages on the two adjacent fitting intervals:

$$\widehat{\kappa}_k = \left| \frac{1}{b_{k-1} - a_{k-1}} \sum_{t=a_{k-1}+1}^{b_{k-1}} (\widehat{\nu}_k^\top Y_t)^p - \frac{1}{b_k - a_k} \sum_{t=a_k+1}^{b_k} (\widehat{\nu}_k^\top Y_t)^p \right|.$$

Note that by (15),

$$\widehat{\kappa}_k = \left| \left\langle \widehat{\Theta}_k, \widehat{\nu}_k^{\otimes p} \right\rangle_{\mathrm{F}} \right| \approx \left| \langle \Theta_k, \nu_k^{\otimes p} \rangle_{\mathrm{F}} \right| = \kappa_k,$$

which motivates $\widehat{\kappa}_k$ as an estimator of κ_k.

We next estimate the long-run variances of $Z_{k,t}$, defined in (22), before and after the change at η_k. Denote

$$\ell_n = \left\lfloor (\log n)^{(1+1/\gamma)/2} \right\rfloor.$$

The Bartlett estimators (Newey & West, 1987) on the left and right fitting intervals are, respectively,

$$\begin{aligned} \widehat{\omega}_{k,L}^2 = \frac{1}{b_{k-1} - a_{k-1}} \sum_{\substack{a_{k-1} < t, r \le b_{k-1} \\ |t-r| \le \ell_n}} & \left(1 - \frac{|t-r|}{\ell_n + 1}\right) \\ & \times \left\{ (\widehat{\nu}_k^\top Y_t)^p - \frac{1}{b_{k-1} - a_{k-1}} \sum_{j=a_{k-1}+1}^{b_{k-1}} (\widehat{\nu}_k^\top Y_j)^p \right\} \\ & \times \left\{ (\widehat{\nu}_k^\top Y_r)^p - \frac{1}{b_{k-1} - a_{k-1}} \sum_{j=a_{k-1}+1}^{b_{k-1}} (\widehat{\nu}_k^\top Y_j)^p \right\} \end{aligned} \tag{24}$$

and

$$\begin{aligned} \widehat{\omega}_{k,R}^2 = \frac{1}{b_k - a_k} \sum_{\substack{a_k < t, r \le b_k \\ |t-r| \le \ell_n}} & \left(1 - \frac{|t-r|}{\ell_n + 1}\right) \\ & \times \left\{ (\widehat{\nu}_k^\top Y_t)^p - \frac{1}{b_k - a_k} \sum_{j=a_k+1}^{b_k} (\widehat{\nu}_k^\top Y_j)^p \right\} \left\{ (\widehat{\nu}_k^\top Y_r)^p - \frac{1}{b_k - a_k} \sum_{j=a_k+1}^{b_k} (\widehat{\nu}_k^\top Y_j)^p \right\}. \end{aligned} \tag{25}$$

In (24) and (25), the summands with $t = r$ give the sample variance of $(\widehat{\nu}_k^\top Y_t)^p$ on the corresponding fitting interval, while the summands with $t \neq r$ account for temporal dependence through weighted sample autocovariances. The estimators $\widehat{\omega}_{k,L}^2$ and $\widehat{\omega}_{k,R}^2$ estimate $\omega_{k,L}^2$ and $\omega_{k,R}^2$ in Assumption 6. Let $\widehat{\omega}_{k,L}$ and $\widehat{\omega}_{k,R}$ denote their nonnegative square roots. Let $B_{k,L}$ and $B_{k,R}$ be independent standard Brownian motions, independent of the observations. Define

$$\widehat{\mathcal{H}}_k(u) = \begin{cases} |u| + 2\widehat{\omega}_{k,L} B_{k,L}(-u), & u < 0, \\ 0, & u = 0, \\ u + 2\widehat{\omega}_{k,R} B_{k,R}(u), & u > 0, \end{cases} \qquad u \in \mathbb{R}. \tag{26}$$

For $a \in (0,1)$, let $\widehat{q}_{k,a}$ be the conditional a-quantile, given the observations, of the minimizer of (26). These quantiles can be approximated by simulating independent paths and taking empirical quantiles of their minimizers. For $\alpha \in (0,1)$, define the confidence interval by

$$\mathcal{J}_{k,\alpha} = \left[\widehat{\eta}_k - \frac{\widehat{q}_{k,1-\alpha/2}}{\widehat{\kappa}_k^2},\ \widehat{\eta}_k - \frac{\widehat{q}_{k,\alpha/2}}{\widehat{\kappa}_k^2} \right]. \tag{27}$$

Theorem 6. *Fix $k \in \{1, \dots, K\}$ and suppose the conditions of Theorem 5 hold. Then for every fixed $\alpha \in (0,1)$, the confidence interval in* (27) *satisfies*

$$\mathbb{P}\big(\eta_k \in \mathcal{J}_{k,\alpha}\big) \longrightarrow 1 - \alpha.$$

B PROOFS FOR SECTION 2.1

Recall the uniform CUSUM event from (72):

$$\mathcal{E}_{\mathrm{cusum}} = \left\{ \sup_{\substack{I=(s,e] \\ 1\le h\le |I|/2}} \max_{s+h\le q\le e-h} \frac{\|\mathcal{T}_I(q) - \mathbb{E}\mathcal{T}_I(q)\|_{\mathrm{op}}}{\sqrt{D+\log n} + (D+\log n)^{p/2+1/\gamma}/\sqrt{h}} \le C_2\sigma^p \right\}.$$

Here I ranges over all nonempty index intervals in $(0, n]$. The constant C_2 is from Theorem 18, which gives $\mathbb{P}(\mathcal{E}_{\mathrm{cusum}}) \ge 1 - n^{-5}$.

Use $\tau = c_\tau \sigma^p$ as in Theorem 1. For $I \in \mathcal{I}$, denote the threshold in (10) by

$$\iota_I = \tau \left\{ \sqrt{D+\log n} + \frac{(D+\log n)^{p/2+1/\gamma}}{\sqrt{|I|}} \right\}.$$

Set $C_0 = 2C_2$. Throughout this appendix, take $c_\tau > C_0$ and C_{I} sufficiently large depending on c_τ, as required by Theorem 1.

Lemma 7 (Detection on an isolating seeded interval)**.** *Suppose Assumptions 1 and 2 hold for fixed $p \ge 2$. Take $\tau = c_\tau \sigma^p$ for sufficiently large c_τ. Let $I_k^* = (a_k^*, b_k^*]$ be an isolating interval defined in Theorem 20. Then, on the event $\mathcal{E}_{\mathrm{cusum}}$ defined in* (72)*, $G_{I_k^*} > \iota_{I_k^*}$, where G_I is the maximum of $\|\mathcal{T}_I(t)\|_{\mathrm{op}}$ over the central half of I, as defined in* (9)*.*

Proof of Theorem 7. Choose $c_\tau > C_0 = 2C_2$, with C_2 from Theorem 18, and then C_{I} sufficiently large depending on c_τ. Because I_k^* contains exactly one change, the CUSUM definition gives

$$\mathbb{E}\mathcal{T}_{I_k^*}(\eta_k) = \sqrt{\frac{(\eta_k - a_k^*)(b_k^* - \eta_k)}{|I_k^*|}}\,\Theta_k.$$

By (75), both $\eta_k - a_k^*$ and $b_k^* - \eta_k$ are at least $|I_k^*|/4$. Taking operator norms and using $\|\Theta_k\|_{\mathrm{op}} = \kappa_k$ gives

$$\left\|\mathbb{E}\mathcal{T}_{I_k^*}(\eta_k)\right\|_{\mathrm{op}} \ge \frac{1}{4}\sqrt{|I_k^*|}\,\kappa_k > \frac{\sqrt{\Delta_{\min}}}{16\sqrt{2}}\,\kappa_k. \tag{28}$$

The last inequality follows from $|I_k^*| > \Delta_{\min}/32$ in (74). On $\mathcal{E}_{\mathrm{cusum}}$, (72) gives

$$\max_{a_k^*+|I_k^*|/4\le q\le b_k^*-|I_k^*|/4} \left\|\mathcal{T}_{I_k^*}(q) - \mathbb{E}\mathcal{T}_{I_k^*}(q)\right\|_{\mathrm{op}} \le C_0\sigma^p \left\{ \sqrt{D+\log n} + \frac{(D+\log n)^{p/2+1/\gamma}}{\sqrt{|I_k^*|}} \right\}.$$

Assumption 2 therefore yields

$$\begin{aligned}
&(c_\tau + C_0)\sigma^p \left\{ \sqrt{D+\log n} + \frac{(D+\log n)^{p/2+1/\gamma}}{\sqrt{|I_k^*|}} \right\} \\
&\quad\le (c_\tau + C_0)\left(1+4\sqrt{2}\right) C_{\mathrm{I}}^{-1/2}\sqrt{\Delta_{\min}}\,\kappa_{\min} \\
&\quad\le (c_\tau + C_0)\left(1+4\sqrt{2}\right) C_{\mathrm{I}}^{-1/2}\sqrt{\Delta_{\min}}\,\kappa_k < \frac{\sqrt{\Delta_{\min}}}{32\sqrt{2}}\,\kappa_k.
\end{aligned}$$

By (75),

$$a_k^* + \frac{|I_k^*|}{4} \le \eta_k \le b_k^* - \frac{|I_k^*|}{4},$$

so η_k is among the candidate splits defining $G_{I_k^*}$ in (9). The definition of $\iota_{I_k^*}$ and the preceding bound give

$$\left(1 + \frac{C_0}{c_\tau}\right)\iota_{I_k^*} < \frac{\sqrt{\Delta_{\min}}}{32\sqrt{2}}\,\kappa_k < \left\|\mathbb{E}\mathcal{T}_{I_k^*}(\eta_k)\right\|_{\mathrm{op}},$$

where the second inequality follows from (28). On $\mathcal{E}_{\mathrm{cusum}}$, the reverse triangle inequality and the deviation bound therefore yield

$$\begin{aligned} G_{I_k^*} &\geq \left\|\mathcal{T}_{I_k^*}(\eta_k)\right\|_{\mathrm{op}} \\ &\geq \left\|\mathbb{E}\mathcal{T}_{I_k^*}(\eta_k)\right\|_{\mathrm{op}} - \left\|\mathcal{T}_{I_k^*}(\eta_k) - \mathbb{E}\mathcal{T}_{I_k^*}(\eta_k)\right\|_{\mathrm{op}} \\ &\geq \left\|\mathbb{E}\mathcal{T}_{I_k^*}(\eta_k)\right\|_{\mathrm{op}} - \frac{C_0}{c_\tau}\iota_{I_k^*} \\ &> \left(1 + \frac{C_0}{c_\tau}\right)\iota_{I_k^*} - \frac{C_0}{c_\tau}\iota_{I_k^*} = \iota_{I_k^*}. \end{aligned}$$

□

Proof of Theorem 1. **Step 1: uniform control and significant isolating intervals.** By Theorem 18, $\mathbb{P}(\mathcal{E}_{\mathrm{unif}}) \geq 1 - n^{-5}$ and $\mathcal{E}_{\mathrm{unif}} \subseteq \mathcal{E}_{\mathrm{cusum}}$. Work on $\mathcal{E}_{\mathrm{unif}}$. Applying (72) with $h = |I|/4$ gives, simultaneously over $I = (s, e] \in \mathcal{I}$,

$$\max_{s+|I|/4 \leq q \leq e-|I|/4} \|\mathcal{T}_I(q) - \mathbb{E}\mathcal{T}_I(q)\|_{\mathrm{op}} \leq \frac{C_0}{c_\tau}\iota_I < \iota_I.$$

If I contains no true change in its interior, its population CUSUM vanishes, so $G_I < \iota_I$. Thus every significant interval contains an interior change.

The augmented projection bound used in Theorem 15 and Hölder's inequality give

$$\sup_{1 \leq t \leq n} \left\|\mathcal{M}_t^{(p)}\right\|_{\mathrm{op}} \leq (2p)^{p/2}\sigma^p, \qquad \kappa_{\min} \leq 2(2p)^{p/2}\sigma^p.$$

Consequently, Assumption 2 implies

$$\Delta_{\min} \geq \frac{C_{\mathrm{I}}}{4(2p)^p}(D + \log n) \geq 256 \tag{29}$$

for sufficiently large C_{I}. Theorems 7 and 20 therefore give fixed intervals $I_k^* \in \mathcal{I}$ satisfying

$$G_{I_k^*} > \iota_{I_k^*}, \qquad k = 1, \ldots, K. \tag{30}$$

Step 2: selection and pruning. After r selections, let $\mathcal{I}_{\mathrm{act}}^{(r)}$ and $\widetilde{\mathcal{E}}^{(r)}$ be the active intervals and recorded points. Initially,

$$\mathcal{I}_{\mathrm{act}}^{(0)} = \{I \in \mathcal{I} : G_I > \iota_I\}, \qquad \widetilde{\mathcal{E}}^{(0)} = \varnothing.$$

The active family only shrinks, so every active interval remains significant. We prove by induction that there is a set $\mathcal{K}_r \subseteq \{1, \ldots, K\}$ of matched changes such that:

(i) the r distinct recorded points can be labeled as $\widetilde{\mathcal{E}}^{(r)} = \{\check{\eta}_k : k \in \mathcal{K}_r\}$, where $|\mathcal{K}_r| = r$ and $|\check{\eta}_k - \eta_k| \leq 3\Delta_{\min}/64$;

(ii) no active interval contains a matched change in its interior;

(iii) $I_k^* \in \mathcal{I}_{\mathrm{act}}^{(r)}$ for every $k \notin \mathcal{K}_r$.

These statements hold at $r = 0$ by (30).

Suppose they hold for $r < K$. By (iii), the active family is nonempty. Let $J = (s, e]$ be the selected shortest interval. Step 1 gives an interior change η_k in J, and (ii) implies $k \notin \mathcal{K}_r$. Since I_k^* is active by (iii),

$$|J| \leq |I_k^*| \leq \Delta_{\min}/16.$$

Thus J contains only η_k. Every point in its central half, including the recorded maximizer $\check{\eta}_k$, satisfies

$$|\check{\eta}_k - \eta_k| \leq 3|J|/4 \leq 3\Delta_{\min}/64.$$

The selected intervals are pairwise disjoint, because each selection removes all overlapping intervals. Their recorded points are therefore distinct, proving (i) for the new selection.

Pruning removes every active interval containing η_k in its interior, so (ii) is preserved. For any remaining unmatched change η_ℓ, an intersection $I_\ell^* \cap J \neq \varnothing$ would imply

$$\Delta_{\min} \le |\eta_\ell - \eta_k| \le |I_\ell^*| + |J| \le \Delta_{\min}/8,$$

a contradiction. Hence every such I_ℓ^* remains active, proving (iii). Taking $\mathcal{K}_{r+1} = \mathcal{K}_r \cup \{k\}$ completes the induction.

Step 3: termination and ordering. Property (iii) prevents termination before K selections. After K selections, (ii) implies that any remaining active interval has no interior change, contradicting Step 1. Thus the algorithm returns exactly K distinct points. Their true order follows from

$$\check{\eta}_{k+1} - \check{\eta}_k \ge \Delta_k - 3\Delta_{\min}/32 \ge 29\Delta_k/32 > 0, \qquad k = 1, \dots, K-1.$$

Sorting therefore gives $\widetilde{\eta}_k = \check{\eta}_k$, proving (12) on $\mathcal{E}_{\mathrm{unif}}$. Its probability bound from Step 1 completes the proof. □

C Proofs for Section 2.2

C.1 Leading jump direction estimation

Let $\mathcal{E}_{\mathrm{pre}}$ denote the preliminary-localization event in (12), established by Theorem 1 with probability at least $1 - n^{-5}$. Set $\widetilde{\eta}_0 = \eta_0 = 0$ and $\widetilde{\eta}_{K+1} = \eta_{K+1} = n$ on this event. The fitting endpoints a_j, b_j and refinement endpoints s_k, e_k are those of Section 2.2.

Lemma 8 (Fitting intervals and refinement windows)**.** *Suppose $\mathcal{E}_{\mathrm{pre}}$ holds and $\Delta_{\min} \ge 256$. For $j = 0, \dots, K$, write $\widetilde{\Delta}_j = \widetilde{\eta}_{j+1} - \widetilde{\eta}_j$. Then*

$$\frac{29}{32}\Delta_j \le \widetilde{\Delta}_j \le \frac{35}{32}\Delta_j, \tag{31}$$

$$(a_j, b_j] \subseteq (\eta_j, \eta_{j+1}] \quad \textit{and} \quad b_j - a_j \ge \frac{7}{16}\Delta_j. \tag{32}$$

For $k = 1, \dots, K$, we have $\eta_{k-1} \le s_k < \eta_k < e_k \le \eta_{k+1}$ and

$$\eta_k - s_k \ge \frac{4}{5}\Delta_{k-1} \quad \textit{and} \quad e_k - \eta_k \ge \frac{4}{5}\Delta_k. \tag{33}$$

Thus each refinement window contains exactly one true change point.

Proof of Theorem 8. The preliminary errors at either endpoint of the jth segment satisfy $|\widetilde{\eta}_j - \eta_j|$, $|\widetilde{\eta}_{j+1} - \eta_{j+1}| \le 3\Delta_j/64$. Hence $\left|\widetilde{\Delta}_j - \Delta_j\right| \le 3\Delta_j/32$, proving (31).

Recall the fitting endpoints in (14) are

$$a_j = \left\lceil \frac{3\widetilde{\eta}_j + \widetilde{\eta}_{j+1}}{4} \right\rceil - 1 \quad \text{and} \quad b_j = \left\lfloor \frac{\widetilde{\eta}_j + 3\widetilde{\eta}_{j+1}}{4} \right\rfloor.$$

On $\mathcal{E}_{\mathrm{pre}}$, both endpoint errors are at most $3\Delta_{\min}/64 \le 3\Delta_j/64$. The triangle inequality gives

$$\left|\frac{3\widetilde{\eta}_j + \widetilde{\eta}_{j+1}}{4} - \frac{3\eta_j + \eta_{j+1}}{4}\right| \le \frac{3}{4}|\widetilde{\eta}_j - \eta_j| + \frac{1}{4}|\widetilde{\eta}_{j+1} - \eta_{j+1}| \le \frac{3\Delta_j}{64},$$
$$\left|\frac{\widetilde{\eta}_j + 3\widetilde{\eta}_{j+1}}{4} - \frac{\eta_j + 3\eta_{j+1}}{4}\right| \le \frac{1}{4}|\widetilde{\eta}_j - \eta_j| + \frac{3}{4}|\widetilde{\eta}_{j+1} - \eta_{j+1}| \le \frac{3\Delta_j}{64}.$$

Since $\lceil x \rceil \ge x$ and $\lfloor x \rfloor \le x$, the first and last fitted indices satisfy

$$a_j + 1 - \eta_j \ge \frac{3\widetilde{\eta}_j + \widetilde{\eta}_{j+1}}{4} - \eta_j \ge \frac{\Delta_j}{4} - \frac{3\Delta_j}{64} = \frac{13\Delta_j}{64} > 0,$$
$$\eta_{j+1} - b_j \ge \eta_{j+1} - \frac{\widetilde{\eta}_j + 3\widetilde{\eta}_{j+1}}{4} \ge \frac{13\Delta_j}{64} > 0.$$

Thus $(a_j, b_j] \subseteq (\eta_j, \eta_{j+1}]$. The inequalities $\lceil x \rceil - 1 \le x$ and $\lfloor x \rfloor \ge x - 1$ also give

$$b_j - a_j \ge \frac{\widetilde{\Delta}_j}{2} - 1 \ge \frac{29\Delta_j}{64} - 1 = \frac{7}{16}\Delta_j + \frac{\Delta_j}{64} - 1 \ge \frac{7}{16}\Delta_j,$$

where the last inequality uses $\Delta_j \ge 256$. This proves (32).

Recall the refinement endpoints, including integer rounding, are

$$s_k = \left\lfloor \frac{9\widetilde{\eta}_{k-1} + \widetilde{\eta}_k}{10} \right\rfloor, \qquad e_k = \left\lfloor \frac{\widetilde{\eta}_k + 9\widetilde{\eta}_{k+1}}{10} \right\rfloor.$$

The corresponding true-endpoint averages are $\eta_{k-1} + \Delta_{k-1}/10$ and $\eta_k + 9\Delta_k/10$. Their preliminary versions differ by at most $3\Delta_{\min}/64$. Using $x - 1 \le \lfloor x \rfloor \le x$ and $\Delta_{\min} \le \Delta_{k-1}, \Delta_k$ therefore gives

$$\eta_k - s_k \ge \frac{273}{320}\Delta_{k-1}, \qquad s_k - \eta_{k-1} \ge \frac{17}{320}\Delta_{k-1} - 1,$$
$$e_k - \eta_k \ge \frac{273}{320}\Delta_k - 1, \qquad \eta_{k+1} - e_k \ge \frac{17}{320}\Delta_k.$$

For $\Delta \ge 256$, we have $17\Delta/320 - 1 > 0$ and $273\Delta/320 - 1 \ge 4\Delta/5$. Applying these inequalities with $\Delta = \Delta_{k-1}$ and $\Delta = \Delta_k$ proves both margins in (33). The outer distances also satisfy $s_k - \eta_{k-1} > 0$ and $\eta_{k+1} - e_k > 0$. Consequently, $\eta_{k-1} < s_k < \eta_k < e_k < \eta_{k+1}$, so the window contains η_k and no other true change point. □

For a sufficiently large fixed constant C, define

$$\epsilon_n = C\sigma^p \left\{ \sqrt{\frac{D + \log n}{\Delta_{\min}}} + \frac{(D + \log n)^{p/2 + 1/\gamma}}{\Delta_{\min}} \right\}. \tag{34}$$

For each change k, on $\mathcal{E}_{\mathrm{pre}}$, use the adjacent population and fitted tensors

$$\mathcal{M}_L^{(p)} = \mathcal{M}_{\eta_k}^{(p)}, \qquad \mathcal{M}_R^{(p)} = \mathcal{M}_{\eta_k+1}^{(p)},$$
$$\widehat{\mathcal{M}}_L^{(p)} = \overline{\mathcal{M}}_{(a_{k-1}, b_{k-1}]}^{(p)}, \qquad \widehat{\mathcal{M}}_R^{(p)} = \overline{\mathcal{M}}_{(a_k, b_k]}^{(p)},$$

where k is suppressed in these four symbols. Let $\mathcal{E}_{\mathrm{mom},k}$ be the event (82) for these tensors and radius ϵ_n, and define

$$\mathcal{E}_{\mathrm{ref},k} = \mathcal{E}_{\mathrm{mom},k} \cap \{\eta_{k-1} \le s_k < \eta_k < e_k \le \eta_{k+1}\}. \tag{35}$$

When a single change is fixed, suppress k in this event.

Lemma 9 (Simultaneous preliminary accuracy)**.** *Suppose the conditions of Theorem 1 hold, and choose C sufficiently large in* (34)*. For all sufficiently large n, on $\mathcal{E}_{\mathrm{unif}}$ from* (71)*,*

$$\max_{0 \le j \le K} \left\| \overline{\mathcal{M}}_{(a_j, b_j]}^{(p)} - \mathcal{M}_{\eta_j+1}^{(p)} \right\|_{\mathrm{op}} \le \epsilon_n. \tag{36}$$

Moreover,

$$\mathcal{E}_{\mathrm{unif}} \subseteq \mathcal{E}_{\mathrm{pre}} \cap \bigcap_{k=1}^{K} \mathcal{E}_{\mathrm{ref},k} \quad \textit{and} \quad \mathbb{P}(\mathcal{E}_{\mathrm{unif}}) \ge 1 - n^{-5}. \tag{37}$$

Proof of Theorem 9. The projection bound in the proof of Theorem 15 gives $\left\| u^\top Y_t \right\|_{L^p} \le \sqrt{2p}\,\sigma$ for every t and $u \in \mathbb{S}^D$. Hölder's inequality and the triangle inequality yield

$$\left\| \mathcal{M}_t^{(p)} \right\|_{\mathrm{op}} \le \sup_{u_1, \dots, u_p \in \mathbb{S}^D} \prod_{a=1}^{p} \left\| u_a^\top Y_t \right\|_{L^p} \le (2p)^{p/2}\sigma^p,$$
$$\kappa_{\min} \le \kappa_k \le \left\| \mathcal{M}_{\eta_k}^{(p)} \right\|_{\mathrm{op}} + \left\| \mathcal{M}_{\eta_k+1}^{(p)} \right\|_{\mathrm{op}} \le 2(2p)^{p/2}\sigma^p.$$

Combining this with Assumption 2 gives $\Delta_{\min} \geq c(D + \log n) \geq 256$ for sufficiently large n. On $\mathcal{E}_{\mathrm{unif}}$, the proof of Theorem 1 gives $\mathcal{E}_{\mathrm{pre}}$. Hence Theorem 8 places each fitting interval $(a_j, b_j]$ inside $(\eta_j, \eta_{j+1}]$, with $b_j - a_j \geq 7\Delta_{\min}/16$.

The simultaneous interval bound (71) therefore applies to these data-selected intervals and gives, for every $j = 0, \ldots, K$,

$$\left\| \overline{\mathcal{M}}^{(p)}_{(a_j, b_j]} - \mathcal{M}^{(p)}_{\eta_j + 1} \right\|_{\mathrm{op}} \leq C_1 \sigma^p \left\{ \sqrt{\frac{D + \log n}{b_j - a_j}} + \frac{(D + \log n)^{p/2 + 1/\gamma}}{b_j - a_j} \right\} \leq \epsilon_n,$$

where the last inequality uses the length bound and the choice of C. The two adjacent moment bounds give $\mathcal{E}_{\mathrm{mom},k}$ for every k. The window inclusions in Theorem 8 then give $\mathcal{E}_{\mathrm{ref},k}$. This proves the event inclusion in (37). The probability bound follows from Theorem 18. □

Lemma 10 (Direction accuracy for local refinement)**.** *Suppose Assumptions 3 and 7 hold, and use ϵ_n defined in* (34)*. For all sufficiently large n, on $\mathcal{E}_{\mathrm{ref},k}$,*

$$d_{\pm}(\widehat{\nu}_k, \nu_k) \leq C_\nu \frac{\epsilon_n}{\kappa_k} \quad \textit{with} \quad C_\nu = \frac{2p}{c_0}. \tag{38}$$

Furthermore,

$$\frac{\epsilon_n \sqrt{D+1}}{\kappa_{\min}} \to 0. \tag{39}$$

Proof of Theorem 10. Since $D + 1 \leq 2D$, Assumption 3 gives

$$\frac{\epsilon_n \sqrt{D+1}}{\kappa_{\min}} \leq C \left\{ \sqrt{\frac{\sigma^{2p} D(D + \log n)}{\Delta_{\min} \kappa_{\min}^2}} + \frac{\sigma^p \sqrt{D}\,(D + \log n)^{p/2+1/\gamma}}{\Delta_{\min} \kappa_{\min}} \right\} \longrightarrow 0.$$

Since $\epsilon_n / \kappa_{\min} \to 0$ and p, c_0, r_0 are fixed, for all sufficiently large n,

$$\max_{1 \leq k \leq K} \frac{\epsilon_n}{\kappa_k} = \frac{\epsilon_n}{\kappa_{\min}} \leq \min\left\{ \frac{1}{8}, \frac{c_0 r_0^2}{8}, \frac{c_0}{32 p^2} \right\}.$$

This verifies (83) in Theorem 23 for every change. Apply that theorem to the adjacent population and fitted tensors defined above, with maximizing directions ν_k and $\widehat{\nu}_k$. Assumption 7 supplies the separation condition, and $\mathcal{E}_{\mathrm{ref},k} \subseteq \mathcal{E}_{\mathrm{mom},k}$ gives (38). □

C.2 Proof of the refinement guarantee

Proof of Theorem 2. **Step 1: Preliminary accuracy and notation.** Since $D \geq 1$, Assumption 3 gives

$$\frac{\Delta_{\min} \kappa_{\min}^2}{\sigma^{2p}(D + \log n)} \geq \frac{\Delta_{\min} \kappa_{\min}^2}{\sigma^{2p} D (D + \log n)} \to \infty,$$
$$\frac{\Delta_{\min} \kappa_{\min}}{\sigma^p (D + \log n)^{p/2+1/\gamma}} \geq \frac{\Delta_{\min} \kappa_{\min}}{\sigma^p \sqrt{D}\,(D + \log n)^{p/2+1/\gamma}} \to \infty.$$

Thus Assumption 2 holds for all sufficiently large n. By Theorem 9,

$$\mathbb{P}(\mathcal{E}_{\mathrm{unif}}^c) \leq n^{-5}, \qquad \mathcal{E}_{\mathrm{unif}} \subseteq \{\widetilde{K} = K\} \cap \bigcap_{k=1}^{K} \mathcal{E}_{\mathrm{ref},k},$$

where $\mathcal{E}_{\mathrm{ref},k}$ is defined in (35). All moment and probability bounds below are unconditional; the deterministic comparisons are on $\mathcal{E}_{\mathrm{unif}}$.

Fix k and suppress its index in $\eta, \widehat{\eta}, \nu, \widehat{\nu}, \Theta, \widehat{\Theta}, \kappa, s, e, \widehat{Q}$. Use the population and fitted tensors $\mathcal{M}_\ell^{(p)}, \widehat{\mathcal{M}}_\ell^{(p)}$, $\ell \in \{L, R\}$, from Section C.1. On $\mathcal{E}_{\mathrm{unif}}$, Theorem 10 gives, after sign alignment,

$$\|\widehat{\nu} - \nu\|_2 \leq \rho_{n,k}, \qquad \rho_{n,k} = C_\nu \epsilon_n / \kappa_k. \tag{40}$$

By (39), take n sufficiently large that $\rho_{n,k}\sqrt{D+1} \le 1$ and $\epsilon_n \le \kappa_k/64$ for every k. For integers $r \ge 0$, define the centered forward and backward sums

$$S_r^+ = \sum_{t=\eta+1}^{\eta+r} \{Y_t^{\otimes p} - \mathbb{E}Y_t^{\otimes p}\}, \qquad S_r^- = \sum_{t=\eta-r+1}^{\eta} \{Y_t^{\otimes p} - \mathbb{E}Y_t^{\otimes p}\}.$$

Here $S_0^{\pm} = 0$, and centered summands outside $\{1,\dots,n\}$ are interpreted as zero. For $M \ge 1$, define $r_M = \lceil M\sigma^{2p}/\kappa^2 \rceil$.

Step 2: Localization error implies a boundary crossing. On $\mathcal{E}_{\mathrm{unif}}$, every $\mathcal{E}_{\mathrm{ref},k}$ holds. For $\ell \in \{L,R\}$, set

$$\mu_\ell = \left\langle \mathcal{M}_\ell^{(p)}, \widehat{\nu}^{\otimes p} \right\rangle_{\mathrm{F}}, \qquad \widehat{\mu}_\ell = \left\langle \widehat{\mathcal{M}}_\ell^{(p)}, \widehat{\nu}^{\otimes p} \right\rangle_{\mathrm{F}}.$$

The moment-error bounds imply $|\widehat{\mu}_\ell - \mu_\ell| \le \epsilon_n$ and

$$\left\|\widehat{\Theta} - \Theta\right\|_{\mathrm{op}} \le \left\|\widehat{\mathcal{M}}_L^{(p)} - \mathcal{M}_L^{(p)}\right\|_{\mathrm{op}} + \left\|\widehat{\mathcal{M}}_R^{(p)} - \mathcal{M}_R^{(p)}\right\|_{\mathrm{op}} \le 2\epsilon_n.$$

Exact maximization at $\widehat{\nu}$ and ν gives

$$\begin{aligned} |\mu_L - \mu_R| &= \left|\left\langle \Theta, \widehat{\nu}^{\otimes p}\right\rangle_{\mathrm{F}}\right| \\ &\ge \left|\left\langle \widehat{\Theta}, \widehat{\nu}^{\otimes p}\right\rangle_{\mathrm{F}}\right| - 2\epsilon_n \\ &\ge \left|\left\langle \widehat{\Theta}, \nu^{\otimes p}\right\rangle_{\mathrm{F}}\right| - 2\epsilon_n \\ &\ge \left|\left\langle \Theta, \nu^{\otimes p}\right\rangle_{\mathrm{F}}\right| - 4\epsilon_n = \kappa - 4\epsilon_n. \end{aligned}$$

The triangle inequalities then give

$$\begin{aligned} |\widehat{\mu}_L - \mu_R| &\ge |\mu_L - \mu_R| - |\widehat{\mu}_L - \mu_L| \ge \kappa - 5\epsilon_n, \\ |\widehat{\mu}_L - \widehat{\mu}_R| &\le \left\|\widehat{\Theta}\right\|_{\mathrm{op}} \le \|\Theta\|_{\mathrm{op}} + \left\|\widehat{\Theta} - \Theta\right\|_{\mathrm{op}} \le \kappa + 2\epsilon_n. \end{aligned}$$

Recall the refinement criterion from Section 2.2. With $Z_i = (\widehat{\nu}^\top Y_i)^p$,

$$\widehat{Q}(t) = \sum_{i=s+1}^{t} (Z_i - \widehat{\mu}_L)^2 + \sum_{i=t+1}^{e} (Z_i - \widehat{\mu}_R)^2, \qquad s < t \le e-1.$$

We detail the case $\widehat{\eta} - \eta \ge 0$. For an integer $r \ge 1$ with $s < \eta + r \le e-1$, $\mathbb{E}Y_i^{\otimes p} = \mathcal{M}_R^{(p)}$ for $\eta < i \le \eta + r$, so

$$\begin{aligned} \widehat{Q}(\eta+r) - \widehat{Q}(\eta) &= \sum_{i=\eta+1}^{\eta+r} \left\{(Z_i - \widehat{\mu}_L)^2 - (Z_i - \widehat{\mu}_R)^2\right\} \\ &= r\left\{(\widehat{\mu}_L - \mu_R)^2 - (\widehat{\mu}_R - \mu_R)^2\right\} - 2(\widehat{\mu}_L - \widehat{\mu}_R)\sum_{i=\eta+1}^{\eta+r}(Z_i - \mu_R) \\ &= r\left\{(\widehat{\mu}_L - \mu_R)^2 - (\widehat{\mu}_R - \mu_R)^2\right\} - 2(\widehat{\mu}_L - \widehat{\mu}_R)\left\langle S_r^+, \widehat{\nu}^{\otimes p}\right\rangle_{\mathrm{F}}. \end{aligned}$$

Suppose $\kappa^2(\widehat{\eta} - \eta)/\sigma^{2p} \ge M$ and set $r = \widehat{\eta} - \eta$. Then $r \ge r_M \ge 1$, since r is an integer. The minimizing property of $\widehat{\eta}$ and the preceding bounds imply

$$\begin{aligned} 0 &\ge \widehat{Q}(\widehat{\eta}) - \widehat{Q}(\eta) \\ &\ge r\{(\kappa - 5\epsilon_n)^2 - \epsilon_n^2\} - 2(\kappa + 2\epsilon_n)\left|\left\langle S_r^+, \widehat{\nu}^{\otimes p}\right\rangle_{\mathrm{F}}\right|. \end{aligned}$$

Consequently,

$$\left|\left\langle S_r^+, \widehat{\nu}^{\otimes p}\right\rangle_{\mathrm{F}}\right| \ge \frac{r\{(\kappa - 5\epsilon_n)^2 - \epsilon_n^2\}}{2(\kappa + 2\epsilon_n)} > \frac{r\kappa}{4},$$

where the last inequality uses $\epsilon_n \le \kappa/64$. For $\widehat{\eta} < \eta$, the same argument uses $r = \eta - \widehat{\eta}$, interchanges L and R, and replaces S_r^+ by S_r^-. Taking $h = \widehat{\nu} - \nu$, with $\|h\|_2 \le \rho_{n,k}$ on $\mathcal{E}_{\mathrm{unif}}$, gives the event inclusion

$$\begin{aligned}
&\mathcal{E}_{\mathrm{unif}} \cap \left\{ \frac{\kappa^2}{\sigma^{2p}} \left|\widehat{\eta} - \eta\right| \ge M \right\} \\
&\quad \subseteq \left\{ \exists\, r \ge r_M : \sup_{\|h\|_2 \le \rho_{n,k}} \left| \left\langle S_r^+, (\nu + h)^{\otimes p} \right\rangle_{\mathrm{F}} \right| > \frac{r\kappa}{4} \right\} \\
&\qquad \cup \left\{ \exists\, r \ge r_M : \sup_{\|h\|_2 \le \rho_{n,k}} \left| \left\langle S_r^-, (\nu + h)^{\otimes p} \right\rangle_{\mathrm{F}} \right| > \frac{r\kappa}{4} \right\}.
\end{aligned}$$

Step 3: Probability of a boundary crossing. For deterministic $\nu, \rho_{n,k}$, the triangle and Minkowski inequalities and (87)–(88) give, for either sign and every $m \ge 1$,

$$\begin{aligned}
\left\| \sup_{\|h\|_2 \le \rho_{n,k}} \max_{1 \le r \le m} \left| \left\langle S_r^{\pm}, (\nu + h)^{\otimes p} \right\rangle_{\mathrm{F}} \right| \right\|_{L^4} &\le \left\| \max_{1 \le r \le m} \left| \left\langle S_r^{\pm}, \nu^{\otimes p} \right\rangle_{\mathrm{F}} \right| \right\|_{L^4} \\
&\quad + \left\| \sup_{\|h\|_2 \le \rho_{n,k}} \max_{1 \le r \le m} \left| \left\langle S_r^{\pm}, (\nu + h)^{\otimes p} - \nu^{\otimes p} \right\rangle_{\mathrm{F}} \right| \right\|_{L^4} \\
&\le C\sigma^p \sqrt{m} \left[1 + \sum_{j=1}^{p} \{\rho_{n,k} \sqrt{D+1}\}^j \right] \\
&\le C(1+p)\sigma^p \sqrt{m}.
\end{aligned}$$

The last inequality uses $\rho_{n,k}\sqrt{D+1} \le 1$. Since p is fixed, taking fourth powers gives

$$\mathbb{E}\left[\sup_{\|h\|_2 \le \rho_{n,k}} \max_{1 \le r \le m} \left| \left\langle S_r^{\pm}, (\nu + h)^{\otimes p} \right\rangle_{\mathrm{F}} \right|^4 \right] \le C m^2 \sigma^{4p}. \tag{41}$$

Applying Markov's inequality to the fourth power in (41), with $m = 2^{q+1} r_M$, gives for every integer $q \ge 0$,

$$\begin{aligned}
\mathbb{P}\left(\sup_{\|h\|_2 \le \rho_{n,k}} \max_{1 \le r \le 2^{q+1} r_M} \left| \left\langle S_r^{\pm}, (\nu + h)^{\otimes p} \right\rangle_{\mathrm{F}} \right| > \tfrac{1}{4} 2^q r_M \kappa \right) &\le \frac{C\sigma^{4p} (2^{q+1} r_M)^2}{(2^q r_M \kappa / 4)^4} \\
&\le \frac{C\sigma^{4p}}{2^{2q} r_M^2 \kappa^4} \\
&\le C(2^q M)^{-2},
\end{aligned}$$

where the last inequality uses $r_M \ge M\sigma^{2p}/\kappa^2$.

For an integer $r \ge r_M$ with $\sup_{\|h\|_2 \le \rho_{n,k}} |\langle S_r^{\pm}, (\nu+h)^{\otimes p} \rangle_{\mathrm{F}}| > r\kappa/4$, choose $q \ge 0$ with $2^q r_M \le r < 2^{q+1} r_M$. Then

$$\begin{aligned}
&\sup_{\|h\|_2 \le \rho_{n,k}} \max_{1 \le u \le 2^{q+1} r_M} \left| \left\langle S_u^{\pm}, (\nu + h)^{\otimes p} \right\rangle_{\mathrm{F}} \right| \\
&\quad \ge \sup_{\|h\|_2 \le \rho_{n,k}} \left| \left\langle S_r^{\pm}, (\nu + h)^{\otimes p} \right\rangle_{\mathrm{F}} \right| > \frac{r\kappa}{4} \ge \frac{2^q r_M \kappa}{4}.
\end{aligned}$$

Taking the union over q and applying the preceding bound gives

$$\begin{aligned}
&\mathbb{P}\left(\exists\, r \ge r_M : \sup_{\|h\|_2 \le \rho_{n,k}} \left| \left\langle S_r^{\pm}, (\nu + h)^{\otimes p} \right\rangle_{\mathrm{F}} \right| > \tfrac{1}{4} r\kappa \right) \\
&\le \sum_{q=0}^{\infty} \mathbb{P}\left(\sup_{\|h\|_2 \le \rho_{n,k}} \max_{1 \le r \le 2^{q+1} r_M} \left| \left\langle S_r^{\pm}, (\nu + h)^{\otimes p} \right\rangle_{\mathrm{F}} \right| > \tfrac{1}{4} 2^q r_M \kappa \right) \\
&\le C M^{-2} \sum_{q=0}^{\infty} 4^{-q} \le C M^{-2}.
\end{aligned} \tag{42}$$

Step 4: Final union bound. The two events in the Step 2 inclusion each have probability at most CM^{-2} by (42). Their union therefore gives, for every k,

$$\mathbb{P}\left(\mathcal{E}_{\text{unif}} \cap \left\{\frac{\kappa_k^2}{\sigma^{2p}}\,|\widehat{\eta}_k - \eta_k| \geq M\right\}\right) \leq 2CM^{-2}.$$

Absorb the factor 2 into C. Since $\widetilde{K} = K$ on $\mathcal{E}_{\text{unif}}$, a union bound over k now gives, for all sufficiently large n and every $M \geq 1$,

$$\begin{aligned}
&\mathbb{P}\left(\{\widetilde{K} \neq K\} \cup \left\{\max_{1\leq k\leq K} \frac{\kappa_k^2}{\sigma^{2p}}\,|\widehat{\eta}_k - \eta_k| \geq M\right\}\right) \\
&\quad\leq \mathbb{P}(\mathcal{E}_{\text{unif}}^c) + \sum_{k=1}^{K} \mathbb{P}\left(\mathcal{E}_{\text{unif}} \cap \left\{\frac{\kappa_k^2}{\sigma^{2p}}\,|\widehat{\eta}_k - \eta_k| \geq M\right\}\right) \\
&\quad\leq n^{-5} + CKM^{-2}. \qquad (43)
\end{aligned}$$

As K is fixed, this proves

$$\max_{1\leq k\leq K} \frac{\kappa_k^2}{\sigma^{2p}}\,|\widehat{\eta}_k - \eta_k| = O_{\mathbb{P}}(1). \qquad (44)$$

□

D Proof for Section 2.3

Proof of Theorem 3. We first construct two regime laws with identical moments below order p and an order-p jump of size κ. We then place this jump at two locations in $[\Delta, n-\Delta]$ and bound the error of any estimator by Le Cam's two-point method. All constants depend only on p.

Step 1: Construct the left and right regime laws. We take $X = (Z, 0, \ldots, 0)^\top \in \mathbb{R}^D$ in both regimes and denote its laws before and after the change point by F_L and F_R, respectively. We specify these laws through two Gaussian mixtures for Z. Set

$$a = \frac{\sigma}{p+3}, \qquad c_0 = \frac{p!}{2(p+3)^p}, \qquad \theta = \frac{\kappa}{2p!a^p}.$$

Then $0 < \kappa \leq c_0\sigma^p$ implies $0 < \theta \leq 1/4$. For each component $j = 0, \ldots, p$, define

$$\pi_j = 2^{-p}\binom{p}{j}, \qquad s_j = (-1)^{p-j},$$
$$w_{L,j} = \pi_j(1+\theta s_j), \qquad w_{R,j} = \pi_j(1-\theta s_j).$$

These are probability weights: since $\sum_{j=0}^p \pi_j = 1$ and $\sum_{j=0}^p \pi_j s_j = 2^{-p}(1-1)^p = 0$,

$$\sum_{j=0}^{p} w_{\ell,j} = 1, \qquad w_{\ell,j} \geq \frac{3}{4}\pi_j > 0, \qquad \ell \in \{L, R\}.$$

The scalar distributions defining F_L and F_R are

$$\text{left regime:} \quad Z \sim \sum_{j=0}^{p} w_{L,j}\, N\big(a(2j-p), a^2\big),$$
$$\text{right regime:} \quad Z \sim \sum_{j=0}^{p} w_{R,j}\, N\big(a(2j-p), a^2\big).$$

Both regimes share the same $p+1$ Gaussian components; their weights differ by the sign of the perturbation $\theta\pi_j s_j$.

Equivalently, for $\ell \in \{L, R\}$, let $J \in \{0, \ldots, p\}$ be the component index with $\mathbb{P}(J = j) = w_{\ell,j}$. Draw $G \sim N(0,1)$ independently of J and set $Z = a(2J - p + G)$. The augmented vector is $Y = (1, X^\top)^\top = (1, Z, 0, \ldots, 0)^\top \in \mathbb{R}^{D+1}$.

Step 2: Verify the tail and moment conditions. For either regime, the Gaussian even-moment formula and monotonicity of L^q norms give, for $q \geq 1$ and $m = \lceil q/2 \rceil$,

$$\|G\|_{L^q} \leq \|G\|_{L^{2m}} = \{(2m-1)!!\}^{1/(2m)} \leq \sqrt{2m} \leq 2\sqrt{q},$$

where $(2m-1)!! \leq (2m)^m$ and $2m \leq q+2 \leq 4q$. Since $|2J-p| \leq p$, Minkowski's inequality gives

$$q^{-1/2} \|Z\|_{L^q} \leq a\left(\frac{p}{\sqrt{q}} + 2\right) \leq a(p+2) \leq \sigma.$$

Thus $\|Z\|_{\psi_2} \leq \sigma$, and under both F_L and F_R,

$$\sup_{v \in \mathbb{S}^{D-1}} \left\|v^\top X\right\|_{\psi_2} = \sup_{v \in \mathbb{S}^{D-1}} |v_1| \, \|Z\|_{\psi_2} \leq \sigma. \tag{45}$$

For the moment comparison, let $\mathbb{E}_L$ and $\mathbb{E}_R$ denote expectations when $X \sim F_L$ and $X \sim F_R$, respectively; in particular, Z has the left and right scalar mixture distributions from Step 1. Independence of J and G and the binomial theorem yield

$$\begin{aligned}
\mathbb{E}_L e^{tZ} - \mathbb{E}_R e^{tZ} &= e^{a^2t^2/2} \sum_{j=0}^{p} (w_{L,j} - w_{R,j}) e^{at(2j-p)} \\
&= 2\theta e^{a^2t^2/2} \, 2^{-p} e^{-apt} (e^{2at} - 1)^p \\
&= 2\theta e^{a^2t^2/2} \sinh(at)^p \\
&= 2\theta a^p t^p + O(t^{p+2}) \qquad (t \to 0).
\end{aligned}$$

Since finite Gaussian mixtures have analytic moment generating functions, the coefficient of t^r is $(\mathbb{E}_L Z^r - \mathbb{E}_R Z^r)/r!$. Comparing coefficients therefore gives

$$\mathbb{E}_L Z^r - \mathbb{E}_R Z^r = 0 \quad (0 \leq r < p), \qquad \mathbb{E}_L Z^p - \mathbb{E}_R Z^p = 2p! a^p \theta = \kappa.$$

Write $\mathcal{M}_\ell^{(p)} = \mathbb{E}_\ell Y^{\otimes p}$ for $\ell \in \{L, R\}$. Since $Y_1 = 1$, $Y_2 = Z$, and $Y_i = 0$ for $i > 2$, each potentially nonzero tensor entry equals $\mathbb{E}_\ell Z^r$, where r counts the indices equal to 2. The entries therefore agree between regimes except at $(2, \ldots, 2)$, where the difference is κ. Consequently,

$$\Theta = \mathcal{M}_L^{(p)} - \mathcal{M}_R^{(p)} = \kappa e_2^{\otimes p}, \qquad \|\Theta\|_{\mathrm{op}} = \kappa \sup_{u_1, \ldots, u_p \in \mathbb{S}^D} \prod_{j=1}^{p} |(u_j)_2| = \kappa. \tag{46}$$

Here $e_2 = (0, 1, 0, \ldots, 0)^\top \in \mathbb{R}^{D+1}$ is the second canonical basis vector, corresponding to the coordinate Z in Y. The same argument shows that all moment tensors of order less than p agree.

The maximizing direction is $\nu = e_2$, up to sign. For $u \in \mathbb{S}^D$,

$$\kappa - \left|\left\langle \Theta, u^{\otimes p} \right\rangle_{\mathrm{F}}\right| = \kappa(1 - |u_2|^p) \geq \kappa(1 - |u_2|) = \frac{\kappa}{2} d_\pm(u, e_2)^2. \tag{47}$$

Thus Assumption 7 holds with curvature constant $1/2$ and radius $r_0 = 1$.

Step 3: Bound the divergence between the regimes. Let ϕ_j be the density of $N(a(2j-p), a^2)$, and let $f_\ell(z) = \sum_{j=0}^{p} w_{\ell,j} \phi_j(z)$ be the scalar density of Z in regime $\ell \in \{L, R\}$. For positive numbers u_j, v_j, the log-sum inequality (Cover & Thomas, 2006, Theorem 2.7.1) states that

$$\left(\sum_{j=0}^{p} u_j\right) \log \frac{\sum_{j=0}^{p} u_j}{\sum_{j=0}^{p} v_j} \leq \sum_{j=0}^{p} u_j \log \frac{u_j}{v_j}.$$

It follows from convexity of $x \mapsto x \log x$, applied with weights $v_j / \sum_{i=0}^{p} v_i$ to the values u_j / v_j. Apply this inequality at each z with $u_j = w_{L,j} \phi_j(z)$ and $v_j = w_{R,j} \phi_j(z)$. The component densities

cancel inside the logarithm and integrate to one. Also, $X = (Z, 0, \dots, 0)^\top$ determines Z through its first coordinate, so embedding Z in X preserves KL divergence. Integration therefore gives

$$\begin{aligned}
\mathrm{KL}(F_L, F_R) &= \int_{\mathbb{R}} f_L(z) \log \frac{f_L(z)}{f_R(z)}\, dz \\
&\le \sum_{j=0}^{p} w_{L,j} \log \frac{w_{L,j}}{w_{R,j}} \\
&= (1+\theta) \log \frac{1+\theta}{1-\theta} \sum_{j:s_j=1} \pi_j + (1-\theta) \log \frac{1-\theta}{1+\theta} \sum_{j:s_j=-1} \pi_j \\
&= \frac{1+\theta}{2} \log \frac{1+\theta}{1-\theta} + \frac{1-\theta}{2} \log \frac{1-\theta}{1+\theta} \\
&= \theta \log \frac{1+\theta}{1-\theta} \le 4\theta^2.
\end{aligned}$$

Here $\sum_{j:s_j=1} \pi_j = \sum_{j:s_j=-1} \pi_j = 1/2$, because $\sum_j \pi_j = 1$ and $\sum_j \pi_j s_j = 0$. The final inequality follows from

$$\log \frac{1+\theta}{1-\theta} = 2 \int_0^\theta \frac{du}{1-u^2} \le 4\theta.$$

Interchanging L and R gives the same bound, so

$$\max\{\mathrm{KL}(F_L, F_R), \mathrm{KL}(F_R, F_L)\} \le 4\theta^2 = \frac{(p+3)^{2p}}{(p!)^2} \frac{\kappa^2}{\sigma^{2p}}. \tag{48}$$

Step 4: Construct the two change-point laws. The integer h will be the distance, measured in observation indices, between the two alternative true change-point locations. Set

$$\beta_p = \frac{(p!)^2}{2(p+3)^{2p}} = 2c_0^2, \qquad h = \left\lfloor \beta_p \frac{\sigma^{2p}}{\kappa^2} \right\rfloor. \tag{49}$$

Since $\kappa \le c_0 \sigma^p$ and $\beta_p = 2c_0^2$, we have $\beta_p \sigma^{2p}/\kappa^2 \ge 2$. Applying $x/2 \le \lfloor x \rfloor \le x$ for $x \ge 2$ gives

$$\frac{\beta_p}{2} \frac{\sigma^{2p}}{\kappa^2} \le h \le \beta_p \frac{\sigma^{2p}}{\kappa^2}.$$

The first condition in Assumption 3 gives

$$\frac{\Delta \kappa^2}{\sigma^{2p}} \ge \frac{\Delta \kappa^2}{\sigma^{2p} D(D + \log n)} \longrightarrow \infty.$$

Hence $h = o(\Delta)$, so $1 \le h < \Delta$ for all sufficiently large n. Choose two possible true change-point locations,

$$\eta^{(0)} = \Delta, \qquad \eta^{(1)} = \Delta + h.$$

For $i \in \{0, 1\}$, let $P_{i,n}$ be the joint law of n independent observations with law F_L up to $\eta^{(i)}$ and F_R afterward:

$$\begin{aligned}
P_{0,n} &= F_L^{\otimes \Delta} \otimes F_R^{\otimes (n-\Delta)}, \\
P_{1,n} &= F_L^{\otimes (\Delta+h)} \otimes F_R^{\otimes (n-\Delta-h)}.
\end{aligned}$$

Because $\Delta \le n/4$,

$$\Delta \le \eta^{(0)} < \eta^{(1)} < 2\Delta \le n - \Delta.$$

Together with (45) and (46), this proves $P_{0,n}, P_{1,n} \in \mathcal{P}(n, D, \Delta, \sigma, \kappa)$ for all sufficiently large n.

Only indices $\Delta + 1, \dots, \Delta + h$ have different marginals: F_R under $P_{0,n}$ and F_L under $P_{1,n}$. By independence and (48)–(49),

$$\begin{aligned}
\mathrm{KL}(P_{0,n}, P_{1,n}) &= h\, \mathrm{KL}(F_R, F_L) \\
&\le h \frac{(p+3)^{2p}}{(p!)^2} \frac{\kappa^2}{\sigma^{2p}} \le \frac{1}{2}.
\end{aligned} \tag{50}$$

With $\mathrm{TV}(P,Q) = \sup_A |P(A) - Q(A)|$, Pinsker's inequality and (50) give

$$\mathrm{TV}(P_{0,n}, P_{1,n}) \le \sqrt{\frac{\mathrm{KL}(P_{0,n}, P_{1,n})}{2}} \le \frac{1}{2}.$$

Step 5: Convert testing error into localization risk. For any estimator $\check{\eta}$, define

$$\psi = \mathbf{1}\{\check{\eta} \ge \Delta + h/2\},$$

where $\psi = i$ selects $P_{i,n}$. Since the change points are h apart,

$$\{\psi = 1\} \subseteq \big\{|\check{\eta} - \eta^{(0)}| \ge h/2\big\}, \qquad \{\psi = 0\} \subseteq \big\{|\check{\eta} - \eta^{(1)}| \ge h/2\big\}.$$

By the definition of total variation,

$$\begin{aligned} P_{0,n}(\psi = 1) + P_{1,n}(\psi = 0) &= 1 - \{P_{1,n}(\psi = 1) - P_{0,n}(\psi = 1)\} \\ &\ge 1 - \mathrm{TV}(P_{0,n}, P_{1,n}) \ge \frac{1}{2}. \end{aligned}$$

Thus

$$\max_{i=0,1} P_{i,n}\big(|\check{\eta} - \eta^{(i)}| \ge h/2\big) \ge \frac{1}{2}\big\{P_{0,n}(\psi = 1) + P_{1,n}(\psi = 0)\big\} \ge \frac{1}{4}.$$

Using $\mathbb{E}W \ge t\mathbb{P}(W \ge t)$ for $W \ge 0$, we obtain

$$\begin{aligned} \sup_{P \in \mathcal{P}(n,D,\Delta,\sigma,\kappa)} \mathbb{E}_P\, |\check{\eta} - \eta| &\ge \max_{i=0,1} \mathbb{E}_{P_{i,n}} |\check{\eta} - \eta^{(i)}| \\ &\ge \frac{h}{2} \max_{i=0,1} P_{i,n}\big(|\check{\eta} - \eta^{(i)}| \ge h/2\big) \\ &\ge \frac{h}{8} \ge \frac{\beta_p}{16} \frac{\sigma^{2p}}{\kappa^2}. \end{aligned} \tag{51}$$

Taking the infimum over $\check{\eta}$ in (51) proves (18) for all sufficiently large n, with

$$c_1 = \frac{\beta_p}{16} = \frac{(p!)^2}{32(p+3)^{2p}}.$$

Finally, the constructed laws satisfy Assumption 1 with $\gamma = \infty$, the separation condition by (47), and the signal conditions because their minimum spacings are at least Δ. Since $h/2 \ge (\beta_p/4)\sigma^{2p}/\kappa^2$, every estimator has probability at least $1/4$ of an error at least $(\beta_p/4)\sigma^{2p}/\kappa^2$ under at least one alternative. This matches the localization scale in Theorem 2. □

E Proofs for Section A

Fix $k \in \{1, \dots, K\}$ and write $\eta = \eta_k$, $\kappa = \kappa_k$, and $(s, e] = (s_k, e_k]$. The population and fitted directions remain ν_k and $\widehat{\nu}_k$. Let $\mathcal{M}_L^{(p)} = \mathcal{M}_\eta^{(p)}$ and $\mathcal{M}_R^{(p)} = \mathcal{M}_{\eta+1}^{(p)}$, with fitted counterparts $\widehat{\mathcal{M}}_L^{(p)}, \widehat{\mathcal{M}}_R^{(p)}$ from Section C.1. Define

$$\mu_\ell = \left\langle \mathcal{M}_\ell^{(p)}, \nu_k^{\otimes p} \right\rangle_{\mathrm{F}}, \quad \ell \in \{L, R\}, \qquad \delta = \mu_L - \mu_R,$$

so $|\delta| = \kappa$. Constants in this appendix may also depend on the fixed σ. Recall the uniform empirical-moment event from (71):

$$\mathcal{E}_{\mathrm{unif}} = \left\{ \sup_{|I| \ge 1} \frac{\left\| \overline{\mathcal{M}}_I^{(p)} - \mathbb{E}\overline{\mathcal{M}}_I^{(p)} \right\|_{\mathrm{op}}}{\sqrt{(D + \log n)/|I|} + (D + \log n)^{p/2+1/\gamma}/|I|} \le C_1 \sigma^p \right\}.$$

Here I ranges over all nonempty index intervals in $(0, n]$, and C_1 is the constant from Theorem 18. Its complement has probability at most n^{-5} by Theorem 9. All moment and probability bounds below use the original law; comparisons involving fitted windows are on $\mathcal{E}_{\mathrm{unif}}$.

Let ϵ_n be the deterministic radius in (34), and set $\rho_{n,k} = C_\nu \epsilon_n / \kappa_k$. By (39) and Assumption 3,

$$\rho_{n,k}\sqrt{D+1} \to 0, \qquad \frac{\sigma^{2p}}{\kappa_k^2} = o(\Delta_{\min}). \tag{52}$$

Since $C_\nu > 0$ is fixed, the first limit also gives $\epsilon_n/\kappa_k = \rho_{n,k}/C_\nu \le \rho_{n,k}\sqrt{D+1}/C_\nu \to 0$. These statements hold uniformly over the fixed set of changes. The moment bound $\kappa_k \le C_p \sigma^p$ also implies $\Delta_{\min} \to \infty$. On $\mathcal{E}_{\mathrm{unif}}$, choose the sign of $\widehat{\nu}_k$ so that $\|\widehat{\nu}_k - \nu_k\|_2 \le \rho_{n,k}$.

E.1 Uniform oracle approximation

For integer candidates q with $s < q \le e-1$, define

$$Q_k^*(q) = \sum_{\substack{t\in\mathbb{Z}\\ s<t\le q}} \{(\nu_k^\top Y_t)^p - \mu_L\}^2 + \sum_{\substack{t\in\mathbb{Z}\\ q<t\le e}} \{(\nu_k^\top Y_t)^p - \mu_R\}^2.$$

Define the centered sums within the adjacent true regimes by

$$S_r^+ = \sum_{t=\eta+1}^{\eta+r} \{Y_t^{\otimes p} - \mathbb{E}Y_t^{\otimes p}\}, \qquad 0 \le r \le \eta_{k+1} - \eta,$$
$$S_r^- = \sum_{t=\eta-r+1}^{\eta} \{Y_t^{\otimes p} - \mathbb{E}Y_t^{\otimes p}\}, \qquad 0 \le r \le \eta - \eta_{k-1},$$

for integer r, with $S_0^\pm = 0$. For larger r, extend the centered summands by zero beyond the respective adjacent regime. Expansion at admissible candidates gives

$$Q_k^*(\eta+r) - Q_k^*(\eta) = r\kappa^2 - 2\delta \left\langle S_r^+, \nu_k^{\otimes p} \right\rangle_{\mathrm{F}},$$
$$Q_k^*(\eta-r) - Q_k^*(\eta) = r\kappa^2 + 2\delta \left\langle S_r^-, \nu_k^{\otimes p} \right\rangle_{\mathrm{F}}. \tag{53}$$

Lemma 11 (Uniform oracle equivalence)**.** *Consider the moment-change model* (3)*, with $p \ge 2$ and K fixed. Suppose Assumptions 1, 3 and 7 hold. Use the estimator and tuning specified in Theorem 2. Then, for each $k \in \{1,\dots,K\}$ and every fixed $M, z > 0$,*

$$\mathbb{P}\left(\mathcal{E}_{\mathrm{unif}} \cap \left\{ \frac{1}{\sigma^{2p}} \max_{\substack{r\in\mathbb{Z}\\ |r|\le M\sigma^{2p}/\kappa^2}} \left| \{\widehat{Q}_k(\eta+r) - \widehat{Q}_k(\eta)\} - \{Q_k^*(\eta+r) - Q_k^*(\eta)\} \right| > z \right\} \right) \to 0. \tag{54}$$

Here $\mathcal{E}_{\mathrm{unif}}$ is the event defined in (71)*.*

Proof. On $\mathcal{E}_{\mathrm{unif}}$, (33) and (52) give $M\sigma^{2p}/\kappa^2 + 1 < \min\{\eta - s, e - \eta\}$ for all sufficiently large n. Thus every $\eta + r$ in the maximum in (54) lies in $(s, e-1]$, where both criteria are defined.

Step 1: Projected jump error. On $\mathcal{E}_{\mathrm{unif}}$, write $\widehat{\delta} = \left\langle \widehat{\Theta}_k, \widehat{\nu}_k^{\otimes p} \right\rangle_{\mathrm{F}}$. Since $\|\widehat{\nu}_k\|_2 = 1$, the tensor bounds in Theorem 9 imply, for $\ell \in \{L, R\}$,

$$\left| \left\langle \widehat{\mathcal{M}}_\ell^{(p)} - \mathcal{M}_\ell^{(p)}, \widehat{\nu}_k^{\otimes p} \right\rangle_{\mathrm{F}} \right| \le \|\widehat{\mathcal{M}}_\ell^{(p)} - \mathcal{M}_\ell^{(p)}\|_{\mathrm{op}} \le \epsilon_n,$$
$$\|\widehat{\Theta}_k - \Theta_k\|_{\mathrm{op}} \le \|\widehat{\mathcal{M}}_L^{(p)} - \mathcal{M}_L^{(p)}\|_{\mathrm{op}} + \|\widehat{\mathcal{M}}_R^{(p)} - \mathcal{M}_R^{(p)}\|_{\mathrm{op}} \le 2\epsilon_n.$$

For the signed jump error, add and subtract $\langle \Theta_k, \widehat{\nu}_k^{\otimes p} \rangle_{\mathrm{F}}$ and apply (81):

$$\begin{aligned} |\widehat{\delta} - \delta| &\le \left| \left\langle \widehat{\Theta}_k - \Theta_k, \widehat{\nu}_k^{\otimes p} \right\rangle_{\mathrm{F}} \right| + \left| \left\langle \Theta_k, \widehat{\nu}_k^{\otimes p} - \nu_k^{\otimes p} \right\rangle_{\mathrm{F}} \right| \\ &\le \|\widehat{\Theta}_k - \Theta_k\|_{\mathrm{op}} + p\|\Theta_k\|_{\mathrm{op}} \|\widehat{\nu}_k - \nu_k\|_2 \\ &\le 2\epsilon_n + p\kappa\rho_{n,k} = (2 + pC_\nu)\epsilon_n, \end{aligned}$$

where $\|\Theta_k\|_{\mathrm{op}} = \kappa$ and $\rho_{n,k} = C_\nu \epsilon_n / \kappa$. Finally, the unit norm of $\widehat{\nu}_k$ and the operator-norm triangle inequality give

$$|\widehat{\delta}| \le \|\widehat{\Theta}_k\|_{\mathrm{op}} \le \|\Theta_k\|_{\mathrm{op}} + \|\widehat{\Theta}_k - \Theta_k\|_{\mathrm{op}} \le \kappa + 2\epsilon_n.$$

Since $\delta^2 = \kappa^2$, it follows that

$$|\widehat{\delta}^2 - \kappa^2| = |\widehat{\delta} - \delta|\,|\widehat{\delta} + \delta| \le C(\kappa\epsilon_n + \epsilon_n^2).$$

Step 2: Uniform comparison. Expanding the fitted criterion and collecting the inner products gives, for positive integer shifts within its search region,

$$\widehat{Q}_k(\eta+r)-\widehat{Q}_k(\eta)=r\widehat{\delta}^2+2\widehat{\delta}\left\langle r(\widehat{\mathcal{M}}_R^{(p)}-\mathcal{M}_R^{(p)})-S_r^+,\widehat{\nu}_k^{\otimes p}\right\rangle_{\mathrm{F}},$$
$$\widehat{Q}_k(\eta-r)-\widehat{Q}_k(\eta)=r\widehat{\delta}^2-2\widehat{\delta}\left\langle r(\widehat{\mathcal{M}}_L^{(p)}-\mathcal{M}_L^{(p)})-S_r^-,\widehat{\nu}_k^{\otimes p}\right\rangle_{\mathrm{F}}.$$

Subtracting (53) and using Step 1 yields, on $\mathcal{E}_{\mathrm{unif}}$ and for either sign,

$$\left|\{\widehat{Q}_k(\eta\pm r)-\widehat{Q}_k(\eta)\}-\{Q_k^*(\eta\pm r)-Q_k^*(\eta)\}\right|\le Cr(\kappa\epsilon_n+\epsilon_n^2)+C\epsilon_n\left|\left\langle S_r^\pm,\nu_k^{\otimes p}\right\rangle_{\mathrm{F}}\right|$$
$$+C(\kappa+\epsilon_n)\left|\left\langle S_r^\pm,\widehat{\nu}_k^{\otimes p}-\nu_k^{\otimes p}\right\rangle_{\mathrm{F}}\right|.$$

The difference is zero at $r=0$.

Set $m=\lceil M\sigma^{2p}/\kappa^2\rceil$. On $\mathcal{E}_{\mathrm{unif}}$, the fitted direction lies in the deterministic ball $\|\widehat{\nu}_k-\nu_k\|_2\le\rho_{n,k}$. We therefore bound the last term uniformly over this ball before applying Theorem 25. Since the sums use deterministic blocks, the resulting moment bounds are unconditional:

$$\left\|\max_{1\le r\le m}\left|\left\langle S_r^\pm,\nu_k^{\otimes p}\right\rangle_{\mathrm{F}}\right|\right\|_{L^4}\le C\sigma^p\sqrt{m},$$

and

$$\left\|\sup_{\|h\|_2\le\rho_{n,k}}\max_{1\le r\le m}\left|\left\langle S_r^\pm,(\nu_k+h)^{\otimes p}-\nu_k^{\otimes p}\right\rangle_{\mathrm{F}}\right|\right\|_{L^4}\le C\sigma^p\sqrt{m}\sum_{j=1}^{p}\{\rho_{n,k}\sqrt{D+1}\}^j.$$

Since $\kappa\le C_p\sigma^p$, we have $m\le C_M\sigma^{2p}/\kappa^2$. Taking maxima over $|r|\le M\sigma^{2p}/\kappa^2$ in the comparison and applying these bounds, the resulting upper bound, divided by σ^{2p}, has L^4 norm at most

$$C_M\left\{\frac{\epsilon_n}{\kappa}+\frac{\epsilon_n^2}{\kappa^2}+\left(1+\frac{\epsilon_n}{\kappa}\right)\sum_{j=1}^{p}\{\rho_{n,k}\sqrt{D+1}\}^j\right\}\to 0$$

by (52). Markov's inequality proves (54). □

E.2 Nonvanishing-jump limit

Proof of Theorem 4. **Step 1: Projected moments and oracle increments.** For every fixed $q\ge 1$, the augmented projection bound in the proof of Theorem 15 gives, with fixed $\sigma<\infty$,

$$\sup_{1\le t\le n}\|\nu_k^\top Y_t\|_{L^q}\le C_q\sigma<\infty.$$

In particular, the pth powers are uniformly integrable, even when D grows. The projected finite-dimensional convergence in Assumption 4 therefore yields

$$\mu_L\to\mathbb{E}\xi_{k,0}^p,\qquad \mu_R\to\mathbb{E}\xi_{k,1}^p,\qquad \delta\to\rho_k,\qquad |\rho_k|=\kappa_k^*.$$

Because $\Delta_{\min}\to\infty$, every fixed local index eventually lies in its corresponding true regime. The moment model then gives $\mathbb{E}\xi_{k,t}^p=\mathbb{E}\xi_{k,0}^p$ for $t\le 0$ and $\mathbb{E}\xi_{k,t}^p=\mathbb{E}\xi_{k,1}^p$ for $t\ge 1$. Write $\zeta_{k,t}=\xi_{k,t}^p-\mathbb{E}\xi_{k,t}^p$.

Define the finite-sample oracle contrast by

$$\mathcal{L}_{n,k}(r)=\begin{cases}|r|\kappa^2+2\delta\left\langle S_{|r|}^-,\nu_k^{\otimes p}\right\rangle_{\mathrm{F}}, & r<0,\\ 0, & r=0,\\ r\kappa^2-2\delta\left\langle S_r^+,\nu_k^{\otimes p}\right\rangle_{\mathrm{F}}, & r>0,\end{cases}\qquad r\in\mathbb{Z}.$$

On $\mathcal{E}_{\mathrm{unif}}$, (53) gives $\mathcal{L}_{n,k}(r)=Q_k^*(\eta+r)-Q_k^*(\eta)$ whenever $s<\eta+r\le e-1$.

Fix an integer $H \geq 1$ independently of n. By Assumption 4 and the deterministic mean and coefficient limits above, continuous mapping through coordinatewise pth powers and finite sums gives jointly

$$\left(\left\langle S_r^+, \nu_k^{\otimes p}\right\rangle_{\mathrm{F}}, \left\langle S_r^-, \nu_k^{\otimes p}\right\rangle_{\mathrm{F}}\right)_{r=1}^{H} \xrightarrow{d} \left(\sum_{t=1}^{r} \zeta_{k,t}, \sum_{t=-r+1}^{0} \zeta_{k,t}\right)_{r=1}^{H}$$

and

$$\left(\mathcal{L}_{n,k}(r)\right)_{|r|\leq H} \xrightarrow{d} \left(\mathcal{L}_k(r)\right)_{|r|\leq H} \qquad \text{in } \mathbb{R}^{2H+1}, \tag{55}$$

where $\mathcal{L}_k$ is defined in (21).

Step 2: A finite, unique limiting minimizer. Set $T_r^+ = \sum_{t=1}^{r} \zeta_{k,t}$ and $T_r^- = \sum_{t=-r+1}^{0} \zeta_{k,t}$ for integers $r \geq 1$. Fix an integer $m \geq 1$. For all sufficiently large n, the forward block $(Y_{\eta+1}, \ldots, Y_{\eta+m})$ and the reverse block $(Y_\eta, \ldots, Y_{\eta-m+1})$ lie in the adjacent true regimes. Apply (87) to these deterministic blocks with the deterministic unit direction ν_k. Step 1 also gives joint convergence of $(\langle S_r^\pm, \nu_k^{\otimes p}\rangle_{\mathrm{F}})_{r=1}^m$ to $(T_r^\pm)_{r=1}^m$. For $A > 0$, the function $x \mapsto \min\{A, \max_{1\leq r\leq m} |x_r|^4\}$ is bounded and continuous on $\mathbb{R}^m$, so

$$\mathbb{E}\min\left\{A, \max_{1\leq r\leq m} |T_r^\pm|^4\right\} = \lim_{n\to\infty} \mathbb{E}\min\left\{A, \max_{1\leq r\leq m} \left|\left\langle S_r^\pm, \nu_k^{\otimes p}\right\rangle_{\mathrm{F}}\right|^4\right\} \leq C\sigma^{4p}m^2.$$

The last inequality follows by removing the truncation and taking fourth powers in (87). Letting $A \uparrow \infty$ and applying monotone convergence yields

$$\mathbb{E}\max_{1\leq r\leq m} |T_r^\pm|^4 \leq C\sigma^{4p}m^2.$$

For each $b > 0$, Markov's inequality implies

$$\sum_{j=0}^{\infty} \mathbb{P}\left(\max_{1\leq r\leq 2^j} |T_r^\pm| > b2^j\right) \leq C\sigma^{4p}b^{-4}\sum_{j=0}^{\infty} 2^{-2j} < \infty.$$

By the first Borel–Cantelli lemma, for each positive rational b and either sign, almost surely

$$\max_{1\leq s\leq 2^j} |T_s^\pm| \leq b2^j \qquad \text{for all sufficiently large } j.$$

The choices of b and sign are countable, so these statements hold simultaneously on one event of probability one. On this event, fix such a b and, for each integer $r \geq 1$, choose j with $2^{j-1} < r \leq 2^j$. For all sufficiently large r,

$$\frac{|T_r^\pm|}{r} \leq \frac{\max_{1\leq s\leq 2^j} |T_s^\pm|}{2^{j-1}} \leq 2b.$$

Taking $\limsup_{r\to\infty}$ and then letting $b \downarrow 0$ through positive rational values proves $T_r^\pm/r \to 0$ almost surely. Hence

$$\frac{\mathcal{L}_k(r)}{|r|} \longrightarrow (\kappa_k^*)^2 > 0 \qquad \text{as } |r| \to \infty.$$

The positive limit implies $\mathcal{L}_k(r) \to +\infty$ almost surely. Since $\mathcal{L}_k(0) = 0$, the set of integers with $\mathcal{L}_k(r) \leq 0$ is nonempty and finite. Minimizing over this set therefore gives the global minimum.

To rule out ties and prove uniqueness of the minimizer, fix integers $r < s$. Then

$$\mathcal{L}_k(s) - \mathcal{L}_k(r) = (|s| - |r|)(\kappa_k^*)^2 - 2\rho_k \sum_{t=r+1}^{s} \{\xi_{k,t}^p - \mathbb{E}\xi_{k,t}^p\}.$$

Since $\rho_k \neq 0$, the difference is a nonconstant polynomial with a Lebesgue-null zero set. The joint-density condition in Assumption 4 therefore rules out ties for each $r < s$, and a countable union gives uniqueness. Write $r_* = \arg\min_{r\in\mathbb{Z}} \mathcal{L}_k(r)$.

Step 3: Argmin convergence. Since $\mathbb{P}(\mathcal{E}_{\mathrm{unif}}^c) \to 0$, restriction to $\mathcal{E}_{\mathrm{unif}}$ preserves weak limits. On this event, (33) and $\Delta_{\min} \to \infty$ ensure that every fixed finite set of shifts is eventually allowed. With σ fixed and $\kappa \to \kappa_k^* > 0$, Theorem 11 and (55) therefore give weak convergence of $r \mapsto \widehat{Q}_k(\eta + r) - \widehat{Q}_k(\eta)$ to $\mathcal{L}_k$ in the supremum norm on every compact subset of $\mathbb{Z}$, since these subsets are finite.

The limit paths are continuous on the discrete space $\mathbb{Z}$, and Step 2 gives a unique finite, hence tight, minimizer r_*. The error $\widehat{\eta}_k - \eta_k$ exactly minimizes the centered criterion over its candidate set and is tight by Theorem 2. Extending the criterion outside this set by its minimum plus one preserves both the minimizers and the compact limits above. Applying the argmax theorem (van der Vaart & Wellner, 1996, Theorem 3.2.2) to the negative criterion gives

$$\widehat{\eta}_k - \eta_k \xrightarrow{d} r_* = \arg\min_{r \in \mathbb{Z}} \mathcal{L}_k(r).$$

□

E.3 VANISHING-JUMP LIMIT

Recall from (22) that the centered projected noise is

$$Z_{k,t} = (\nu_k^\top Y_t)^p - \mathbb{E}\{(\nu_k^\top Y_t)^p\} = \left\langle Y_t^{\otimes p} - \mathbb{E} Y_t^{\otimes p}, \nu_k^{\otimes p} \right\rangle_{\mathrm{F}}.$$

Write $\omega_L = \omega_{k,L}$ and $\omega_R = \omega_{k,R}$ for its positive limiting long-run standard deviations before and after η_k, respectively, as specified in Assumption 6. We use the extension in Assumption 5 to define $Z_{k,t}$ beyond the observed adjacent segments.

Recall also the signed jump in the projected pth moment:

$$\delta = \mu_L - \mu_R = \mathbb{E}\{(\nu_k^\top Y_\eta)^p\} - \mathbb{E}\{(\nu_k^\top Y_{\eta+1})^p\}, \qquad |\delta| = \kappa.$$

Lemma 12 (Two-sided projected invariance principle)**.** *Under the conditions of Theorem 5, define*

$$S_{L,n}(t) = \kappa \operatorname{sgn}(\delta) \sum_{j=1}^{\lfloor t/\kappa^2 \rfloor} Z_{k,\eta-j+1}, \qquad S_{R,n}(t) = \kappa \operatorname{sgn}(\delta) \sum_{j=1}^{\lfloor t/\kappa^2 \rfloor} Z_{k,\eta+j}, \qquad t \geq 0.$$

For every fixed $T > 0$,

$$(S_{L,n}, S_{R,n}) \xrightarrow{d} (\omega_L B_{k,L}, \omega_R B_{k,R}) \qquad \text{in } \ell^\infty([0,T])^2,$$

where $\ell^\infty([0,T])$ is the space of bounded functions equipped with the supremum norm, and $B_{k,L}, B_{k,R}$ are independent standard Brownian motions.

Proof of Theorem 12. Put $N = \kappa^{-2} \to \infty$ and

$$\xi_{L,n,j} = \operatorname{sgn}(\delta) Z_{k,\eta-j+1}, \qquad \xi_{R,n,j} = \operatorname{sgn}(\delta) Z_{k,\eta+j}, \qquad j \geq 1,$$

so $S_{\ell,n}(t) = N^{-1/2} \sum_{j=1}^{\lfloor Nt \rfloor} \xi_{\ell,n,j}$. For fixed $T > 0$, it suffices to prove joint finite-dimensional convergence to $(\omega_L B_{k,L}, \omega_R B_{k,R})$ and stochastic equicontinuity of both components on $[0,T]$.

Step 1: Moment and variance bounds. The extensions in Assumption 5 make each row $(\xi_{\ell,n,j})_{j \geq 1}$ centered, stationary, and α-mixing. With σ fixed, Theorem 15 gives $\sup_{\ell,n,j} \|\xi_{\ell,n,j}\|_{L^q} \leq C_q$ for every fixed $q \geq 1$. Davydov's inequality (Davydov, 1968) yields

$$|\operatorname{Cov}(\xi_{\ell,n,j}, \xi_{\ell,n,j+h})| \leq C \alpha_X(h)^{1/2} \leq C e^{-ch^\gamma}, \qquad h \geq 1, \quad \gamma < \infty. \tag{56}$$

This bound also holds across the change, with separation measured in original time. Under independence, all covariances at positive lags vanish. Moreover, Theorem 24 gives

$$\mathbb{E} \max_{0 \leq r \leq m} \left| \sum_{j=1}^{r} \xi_{\ell,n,a+j} \right|^4 \leq C m^2, \qquad a \geq 0, \quad m \geq 1. \tag{57}$$

For $\ell \in \{L, R\}$, write

$$c_{\ell,n}(h) = \mathrm{Cov}(\xi_{\ell,n,1}, \xi_{\ell,n,1+h}), \qquad h \geq 0, \qquad v^2_{\ell,n} = c_{\ell,n}(0) + 2\sum_{h=1}^{\infty} c_{\ell,n}(h).$$

The covariance bound gives $\sup_n \sum_{h\geq 1} h|c_{\ell,n}(h)| < \infty$, so stationarity implies

$$\left| \frac{1}{m} \mathrm{Var}\left(\sum_{j=1}^{m} \xi_{\ell,n,j} \right) - v^2_{\ell,n} \right| \leq \frac{2}{m} \sum_{h=1}^{m-1} h|c_{\ell,n}(h)| + 2\sum_{h=m}^{\infty} |c_{\ell,n}(h)| \leq \frac{C}{m}. \tag{58}$$

Since multiplication by $\mathrm{sgn}(\delta)$ preserves variance, Assumption 6 with $m = \lfloor N \rfloor$ and (58) give

$$v^2_{L,n} \longrightarrow \omega^2_L, \qquad v^2_{R,n} \longrightarrow \omega^2_R. \tag{59}$$

Step 2: Joint finite-dimensional convergence. Fix $0 = t_0 < t_1 < \cdots < t_J \leq T$ and real coefficients $a_{\ell,j}$, not all zero. Set

$$W_n = \sum_{\ell\in\{L,R\}} \sum_{j=1}^{J} a_{\ell,j}\{S_{\ell,n}(t_j) - S_{\ell,n}(t_{j-1})\}.$$

Two disjoint intervals in original time, possibly on opposite sides of the change, have at most h pairs of indices with separation h. The covariance of their sums is therefore bounded by $C\sum_{h\geq 1} he^{-ch} < \infty$. Together with (58) and (59), this gives

$$\mathrm{Var}(W_n) \longrightarrow \sum_{\ell\in\{L,R\}} \omega^2_\ell \sum_{j=1}^{J} a^2_{\ell,j}(t_j - t_{j-1}) > 0.$$

For $\lfloor Nt_{j-1} \rfloor < i \leq \lfloor Nt_j \rfloor$, define the weighted scores in original time order by

$$U_{n,1-i} = a_{L,j}\xi_{L,n,i}, \qquad U_{n,i} = a_{R,j}\xi_{R,n,i},$$
$$\sqrt{N}W_n = \sum_{r=1-\lfloor Nt_J \rfloor}^{\lfloor Nt_J \rfloor} U_{n,r}, \qquad \mathbb{E}U_{n,r} = 0.$$

The weights preserve the uniform α-mixing and covariance bounds in Step 1. Since J is fixed, (57) and Minkowski's inequality give

$$\left\| \sum_{r=a+1}^{a+b} U_{n,r} \right\|_{L^4} \leq C\sqrt{b}, \qquad \text{uniformly over admissible blocks.}$$

These bounds and the positive variance limit verify the α-mixing triangular-array CLT of Withers (1981, Corollary 2.1). Hence

$$\frac{W_n}{\sqrt{\mathrm{Var}(W_n)}} \xrightarrow{d} \mathcal{N}(0,1).$$

The variance limit and the Cramér–Wold device give the joint Gaussian finite-dimensional limit $(\omega_L B_{k,L}, \omega_R B_{k,R})$. Its left and right components satisfy

$$\mathrm{Cov}(\omega_L B_{k,L}(s), \omega_R B_{k,R}(t)) = 0, \qquad s, t \in [0, T].$$

Joint Gaussianity therefore implies that $B_{k,L}$ and $B_{k,R}$ are independent.

Step 3: Stochastic equicontinuity. For every $\varepsilon > 0$, we verify

$$\lim_{h\downarrow 0} \limsup_{n\to\infty} \mathbb{P}\left(\max_{\ell\in\{L,R\}} \sup_{\substack{s,t\in[0,T] \\ |t-s|\leq h}} |S_{\ell,n}(t) - S_{\ell,n}(s)| > \varepsilon \right) = 0. \tag{60}$$

Fix $0 < h \le T$. If $0 \le s \le t \le T$ and $t - s \le h$, then $j = \lfloor s/h \rfloor$ gives $s, t \in [jh, (j+2)h] \cap [0, T]$. Within this window, each increment is the difference of two prefix sums starting at index $\lfloor Njh \rfloor + 1$. Since $\lfloor Nt \rfloor - \lfloor Njh \rfloor \le \lceil 2Nh \rceil + 1$,

$$\sup_{s,t \in [jh,(j+2)h] \cap [0,T]} |S_{\ell,n}(t) - S_{\ell,n}(s)| \le \frac{2}{\sqrt{N}} \max_{0 \le r \le \lceil 2Nh \rceil + 1} \left| \sum_{i=1}^{r} \xi_{\ell,n,\lfloor Njh \rfloor + i} \right|.$$

A union bound over $\ell \in \{L, R\}$ and $j = 0, \ldots, \lfloor T/h \rfloor$, followed by Markov's inequality and (57), now gives

$$\begin{aligned}
&\mathbb{P}\left(\max_{\ell \in \{L,R\}} \sup_{\substack{s,t \in [0,T] \\ |t-s| \le h}} |S_{\ell,n}(t) - S_{\ell,n}(s)| > \varepsilon \right) \\
&\le \sum_{\ell \in \{L,R\}} \sum_{j=0}^{\lfloor T/h \rfloor} \mathbb{P}\left(\max_{0 \le r \le \lceil 2Nh \rceil + 1} \left| \sum_{i=1}^{r} \xi_{\ell,n,\lfloor Njh \rfloor + i} \right| > \frac{\varepsilon \sqrt{N}}{2} \right) \\
&\le \frac{C}{\varepsilon^4 N^2} \left(\left\lfloor \frac{T}{h} \right\rfloor + 1 \right) (\lceil 2Nh \rceil + 1)^2 \le C\varepsilon^{-4} (T/h + 1)(h + N^{-1})^2.
\end{aligned}$$

Taking $\limsup_{n \to \infty}$ and then $h \downarrow 0$ proves (60), since

$$\lim_{h \downarrow 0} C\varepsilon^{-4}(T/h + 1)h^2 = \lim_{h \downarrow 0} C\varepsilon^{-4}(Th + h^2) = 0.$$

For each n, the pair $(S_{L,n}, S_{R,n})$ is a continuous function of a finite random vector into its deterministic step-function subspace, hence is Borel measurable in $\ell^\infty([0,T])^2$. Because $S_{L,n}(0) = S_{R,n}(0) = 0$ and $[0,T]$ is compact, (60) gives asymptotic tightness, with every subsequential limit supported on $C([0,T])^2$. Step 2 identifies its finite-dimensional distributions as those of $(\omega_L B_{k,L}, \omega_R B_{k,R})$. Continuous functions are determined by their values at rational times, so every subsequential limit has this same law. This proves the claimed weak convergence in the supremum norm. □

Proof of Theorem 5. Since $\mathbb{P}(\mathcal{E}_{\text{unif}}^c) \to 0$, restriction to $\mathcal{E}_{\text{unif}}$ preserves weak limits. We work on this event.

Step 1: Convergence on compact intervals. Denote by $r_n(u)$ the rounded shift $\operatorname{sgn}(u)\lfloor |u|/\kappa^2 \rfloor$. The domain of G_n is $\{u \in \mathbb{R} : s < \eta + r_n(u) \le e - 1\}$, corresponding to the integer candidates in (17). On this set, define

$$G_n(u) = \widehat{Q}_k\{\eta + r_n(u)\} - \widehat{Q}_k(\eta).$$

With σ fixed, (52) gives $\kappa^{-2} = o(\Delta_{\min})$. Hence, for each fixed $T > 0$, (33) implies, eventually,

$$\sup_{|u| \le T} |r_n(u)| + 1 \le T/\kappa^2 + 1 < \frac{4}{5}\Delta_{\min} \le \min\{\eta - s, e - \eta\}.$$

Thus $s < \eta + r_n(u) \le e - 1$ for every $|u| \le T$, so G_n is defined throughout $[-T, T]$. For these candidates, (53) gives

$$Q_k^*\{\eta + r_n(u)\} - Q_k^*(\eta) = \kappa^2 |r_n(u)| + \begin{cases} 2S_{L,n}(-u), & u \le 0, \\ -2S_{R,n}(u), & u > 0. \end{cases}$$

Since $0 \le x - \lfloor x \rfloor < 1$ for $x \ge 0$,

$$0 \le |u| - \kappa^2 |r_n(u)| = \kappa^2 \left(\frac{|u|}{\kappa^2} - \left\lfloor \frac{|u|}{\kappa^2} \right\rfloor \right) < \kappa^2 \longrightarrow 0.$$

Thus the drift converges uniformly to $|u|$, and Theorem 12 gives the oracle limit from (23), namely

$$\mathcal{G}_k(u) = |u| + 2 \begin{cases} \omega_L B_{k,L}(-u), & u \le 0, \\ \omega_R B_{k,R}(u), & u > 0, \end{cases}$$

where $B_{k,L}$ and $B_{k,R}$ are independent standard Brownian motions. The right-hand minus sign in the oracle increment is absorbed into $B_{k,R}$ without changing this joint law. Choose a fixed $M \geq T/\sigma^{2p}$. Since σ is fixed, Theorem 11 gives

$$\sup_{|u|\leq T} \left|G_n(u) - \left[Q_k^*\{\eta + r_n(u)\} - Q_k^*(\eta)\right]\right| \leq \max_{\substack{r\in\mathbb{Z} \\ |r|\leq M\sigma^{2p}/\kappa^2}} \left|\left[\widehat{Q}_k(\eta + r) - \widehat{Q}_k(\eta)\right] - \left[Q_k^*(\eta + r) - Q_k^*(\eta)\right]\right| = o_{\mathbb{P}}(1).$$

Slutsky's theorem therefore yields convergence in the supremum norm:

$$G_n \xrightarrow{d} \mathcal{G}_k \qquad \text{in } \ell^\infty([-T, T]).$$

Step 2: A unique finite limiting minimizer. For $\ell \in \{L, R\}$, $\varepsilon > 0$, and $j \geq 0$, Doob's maximal inequality applied to $B_{k,\ell}^2$ gives

$$\mathbb{P}\left(\sup_{2^j\leq t\leq 2^{j+1}} \frac{|B_{k,\ell}(t)|}{t} > \varepsilon\right) \leq \frac{\mathbb{E}B_{k,\ell}(2^{j+1})^2}{\varepsilon^2 2^{2j}} = \frac{2^{1-j}}{\varepsilon^2}.$$

The bound is summable in j. Borel–Cantelli, applied to both sides and all positive rational ε, therefore yields $B_{k,\ell}(t)/t \to 0$ almost surely as $t \to \infty$. Since $\omega_L, \omega_R < \infty$, it follows that

$$\frac{\mathcal{G}_k(u)}{|u|} \longrightarrow 1 \qquad \text{almost surely as } |u| \to \infty.$$

Thus $\mathcal{G}_k(u) \geq |u|/2$ for all sufficiently large $|u|$, so $\mathcal{G}_k(u) \to +\infty$. Because $\mathcal{G}_k(0) = 0$ and its paths are continuous, its global minimum is attained on a compact interval. For $u < v$,

$$\mathrm{Var}\{\mathcal{G}_k(v) - \mathcal{G}_k(u)\} = \begin{cases} 4\omega_L^2(v - u), & u < v \leq 0, \\ 4\omega_R^2(v - u), & 0 \leq u < v, \\ 4\{\omega_L^2|u| + \omega_R^2 v\}, & u < 0 < v, \end{cases}$$

which is strictly positive. Lemma 2.6 of Kim & Pollard (1990), applied to the continuous Gaussian process $-\mathcal{G}_k$ on $\mathbb{R}$, therefore gives almost sure uniqueness.

Step 3: Argmin convergence. The error $\kappa^2(\widehat{\eta}_k - \eta_k)$ exactly minimizes G_n over its domain, since $r_n(\kappa^2 r) = r$ for every integer r, and is tight by Theorem 2 with σ fixed. Extending G_n outside its domain by its minimum plus one preserves both the minimizers and the compact limits in Step 1. The limit paths are continuous, and Step 2 gives a unique finite, hence tight, minimizer. Applying the argmax theorem (van der Vaart & Wellner, 1996, Theorem 3.2.2) to the negative criterion gives

$$\kappa_k^2(\widehat{\eta}_k - \eta_k) \xrightarrow{d} \arg\min_{u\in\mathbb{R}} \mathcal{G}_k(u).$$

□

E.4 Variance estimation and confidence intervals

We first prove consistency of the Bartlett estimators (Newey & West, 1987) defined in (24) and (25). The proof controls both the selected interval and the estimated direction uniformly, because they are computed from the same observations as the variance estimates.

Lemma 13 (Consistency of the projected long-run variances)**.** *Under the conditions of Theorem 5,*

$$\widehat{\omega}_{k,L}^2 \xrightarrow{\mathbb{P}} \omega_{k,L}^2, \qquad \widehat{\omega}_{k,R}^2 \xrightarrow{\mathbb{P}} \omega_{k,R}^2,$$

where the estimators are defined in (24) *and* (25)*.*

Proof of Theorem 13. Fix k. We prove consistency of $\widehat{\omega}_{k,L}^2$; the proof for $\widehat{\omega}_{k,R}^2$ is identical. Relabel the left segment $(\eta_{k-1}, \eta_k]$ as $\{1, \ldots, N\}$, where $N = \Delta_{k-1}$. For deterministic $v \in \mathbb{R}^{D+1}$, write $z(v) = (z_t(v))_{t=1}^N$, where

$$z_t(v) = (v^\top Y_t)^p - \mathbb{E}\{(v^\top Y_t)^p\}, \qquad Z_k = z(\nu_k), \qquad \gamma_{n,h} = \mathrm{Cov}(Z_{k,0}, Z_{k,h}),$$

using the stationary regime law to define $\gamma_{n,h}$. Throughout, $I = (a, c] \subseteq \{1, \dots, N\}$ has $|I| \geq 7N/16$, and $\bar{z}_I(v) = |I|^{-1} \sum_{t \in I} z_t(v)$. On $\mathcal{E}_{\mathrm{unif}}$, the relabelled left fitting interval is

$$\widehat{I} = (a_{k-1} - \eta_{k-1},\, b_{k-1} - \eta_{k-1}] \subseteq \{1, \dots, N\}, \qquad |\widehat{I}| \geq 7N/16,$$

by (32). For $x \in \mathbb{R}^N$, extended by zero, define

$$H_N(x) = \frac{1}{N(\ell_n + 1)} \sum_{r=1}^{N+\ell_n} \left(\sum_{t=r-\ell_n}^{r} x_t \right)^2.$$

Then $H_N^{1/2}$ is a seminorm, and

$$\frac{N}{|I|} H_N(x \mathbf{1}_I) = \frac{1}{|I|} \sum_{\substack{s,t \in I \\ |s-t| \leq \ell_n}} \left(1 - \frac{|s-t|}{\ell_n + 1}\right) x_s x_t.$$

Since stationarity gives

$$z_t(v) - \bar{z}_I(v) = (v^\top Y_t)^p - |I|^{-1} \sum_{s \in I} (v^\top Y_s)^p,$$

the estimator in (24) is exactly

$$\widehat{\omega}_{k,L}^2 = \frac{N}{|\widehat{I}|} H_N\big([z(\widehat{\nu}_k) - \bar{z}_{\widehat{I}}(\widehat{\nu}_k)] \mathbf{1}_{\widehat{I}}\big) \qquad \text{on } \mathcal{E}_{\mathrm{unif}}.$$

Its nonnegative square root is $\widehat{\omega}_{k,L}$. Step 1 proves consistency with the population direction and centering; Step 2 bounds the effect of the fitted direction and empirical centering in this identity.

Step 1: Consistency for the true projected noise. The signal condition and $\kappa_{\min} \leq C\sigma^p$ give $\ell_n \to \infty$ and $\ell_n^2/N \to 0$. For the oracle form $H_N(Z_k \mathbf{1}_I)$, define the squared full-window sums

$$V_r = \frac{1}{\ell_n + 1} \left(\sum_{t=r-\ell_n}^{r} Z_{k,t} \right)^2, \qquad \ell_n + 1 \leq r \leq N.$$

Integrating (64) at order 16 gives $\sup_r \|V_r\|_{L^8} \leq C$. Within each residue class modulo $\ell_n + 1$, the windows are disjoint and their lag-h separation is at least h. Applying Theorem 24 to each class and summing gives

$$\left\| \max_{\ell_n + 1 \leq s \leq N} \left| \sum_{r=\ell_n+1}^{s} (V_r - \mathbb{E}V_r) \right| \right\|_{L^4} \leq C\sqrt{N(\ell_n + 1)}. \tag{61}$$

Only the at most $2\ell_n$ windows crossing the endpoints of I separate $H_N(Z_k \mathbf{1}_I)$ from its full-window sum. Hence Theorem 24 also gives

$$\begin{aligned} &\left\| \sup_I \left| H_N(Z_k \mathbf{1}_I) - \frac{1}{N} \sum_{r=a+\ell_n+1}^{c} V_r \right| \right\|_{L^2} \\ &\quad \leq \frac{2}{N} \left\| \max_{1 \leq r \leq N+\ell_n} \max_{\substack{J \subseteq [r-\ell_n, r] \cap [1,N] \\ J \text{ an interval}}} \left| \sum_{t \in J} Z_{k,t} \right|^2 \right\|_{L^2} \leq C \frac{\ell_n + 1}{\sqrt{N}}. \end{aligned} \tag{62}$$

Here an interval sum is a difference of two prefix sums, and $\| \max_r A_r \|_{L^2} \leq (\sum_r \|A_r\|_{L^2}^2)^{1/2}$. The covariance bound in (56) yields

$$\mathbb{E}V_r = \sum_{|h| \leq \ell_n} \left(1 - \frac{|h|}{\ell_n + 1}\right) \gamma_{n,h} = v_{L,n}^2 + O((\ell_n + 1)^{-1}), \qquad v_{L,n}^2 \longrightarrow \omega_{k,L}^2,$$

where (59) identifies the covariance sum. Since the full-window sum contains $|I| - \ell_n$ terms, (61)–(62) imply

$$\left\| \sup_I \left| \frac{N}{|I|} H_N(Z_k \mathbf{1}_I) - v_{L,n}^2 \right| \right\|_{L^2} \leq C \left\{ \frac{1}{\ell_n + 1} + \frac{\ell_n + 1}{\sqrt{N}} \right\} = o(1). \tag{63}$$

Markov's inequality proves uniform oracle consistency.

Step 2: Estimated direction and sample centering. Apply Theorem 25 within each window and on the whole segment, respectively, to obtain

$$\mathbb{E} \sup_{I,\, \|h\|_2 \le \rho_{n,k}} H_N\big([z(\nu_k + h) - z(\nu_k)]\mathbf{1}_I\big) \le C \left\{ \sum_{j=1}^{p} (\rho_{n,k}\sqrt{D+1})^j \right\}^2 = o(1),$$

$$\left\| \sup_{I,\, \|h\|_2 \le \rho_{n,k}} |\bar{z}_I(\nu_k + h)| \right\|_{L^4} \le \frac{C}{\sqrt{N}} (1 + \rho_{n,k}\sqrt{D+1})^p.$$

Write $x_{I,h} = [z(\nu_k + h) - \bar{z}_I(\nu_k + h)]\mathbf{1}_I$; all suprema over (I, h) below use the intervals above and $\|h\|_2 \le \rho_{n,k}$. Expanding the squares in $H_N(\mathbf{1}_I)$, a pair $s, t \in I$ appears in $\ell_n + 1 - |s - t|$ windows when $|s - t| \le \ell_n$, and in none otherwise. Therefore,

$$\begin{aligned} H_N(\mathbf{1}_I) &= \frac{1}{N} \sum_{s \in I} \sum_{\substack{t \in I \\ |t-s| \le \ell_n}} \left(1 - \frac{|t-s|}{\ell_n + 1}\right) \\ &\le \frac{|I|}{N} \left\{ 1 + 2 \sum_{j=1}^{\ell_n} \left(1 - \frac{j}{\ell_n + 1}\right) \right\} = \frac{|I|}{N}(\ell_n + 1) \le \ell_n + 1. \end{aligned}$$

For each fixed s, there is one term with $t = s$ and at most two terms with $|t - s| = j \ge 1$; adding any missing nonnegative weights gives the first inequality. The equality uses $\sum_{j=1}^{\ell_n}(1 - j/(\ell_n + 1)) = \ell_n/2$, and the last inequality uses $|I| \le N$. The difference between the centered vectors is

$$x_{I,h} - Z_k\mathbf{1}_I = [z(\nu_k + h) - Z_k]\mathbf{1}_I - \bar{z}_I(\nu_k + h)\mathbf{1}_I.$$

Apply the triangle inequality for $H_N^{1/2}$ and $(a + b)^2 \le 2a^2 + 2b^2$. Using $\|\cdot\|_{L^2} \le \|\cdot\|_{L^4}$ for the sample-mean supremum, the two moment bounds above give

$$\begin{aligned} &\mathbb{E} \sup_{I,h} H_N(x_{I,h} - Z_k\mathbf{1}_I) \\ &\le 2\mathbb{E} \sup_{I,h} H_N([z(\nu_k + h) - Z_k]\mathbf{1}_I) + 2(\ell_n + 1) \left\| \sup_{I,h} |\bar{z}_I(\nu_k + h)| \right\|_{L^4}^2 \\ &\le C \left\{ \sum_{j=1}^{p} (\rho_{n,k}\sqrt{D+1})^j \right\}^2 + C \frac{\ell_n + 1}{N} (1 + \rho_{n,k}\sqrt{D+1})^{2p} = o(1). \end{aligned}$$

The last equality uses $\rho_{n,k}\sqrt{D+1} \to 0$ and $\ell_n/N \to 0$. Markov's inequality therefore yields, for every $\varepsilon > 0$,

$$\mathbb{P}\left\{ \sup_{I,h} H_N(x_{I,h} - Z_k\mathbf{1}_I)^{1/2} > \varepsilon \right\} \le \varepsilon^{-2}\mathbb{E} \sup_{I,h} H_N(x_{I,h} - Z_k\mathbf{1}_I) \longrightarrow 0.$$

By (63) and $v_{L,n}^2 \to \omega_{k,L}^2$, $\sup_I H_N(Z_k\mathbf{1}_I) = O_{\mathbb{P}}(1)$. The reverse triangle inequality and $|a^2 - b^2| = |a - b|\,|a + b|$ give, for any $x, y \in \mathbb{R}^N$,

$$|H_N(x) - H_N(y)| \le \{2H_N(y)^{1/2} + H_N(x - y)^{1/2}\} H_N(x - y)^{1/2}.$$

Taking $x = x_{I,h}$ and $y = Z_k\mathbf{1}_I$ in this bound gives

$$\begin{aligned} &\sup_{I,h} |H_N(x_{I,h}) - H_N(Z_k\mathbf{1}_I)| \\ &\quad \le \left\{ 2 \sup_I H_N(Z_k\mathbf{1}_I)^{1/2} + \sup_{I,h} H_N(x_{I,h} - Z_k\mathbf{1}_I)^{1/2} \right\} \\ &\qquad \times \sup_{I,h} H_N(x_{I,h} - Z_k\mathbf{1}_I)^{1/2} = \{O_{\mathbb{P}}(1) + o_{\mathbb{P}}(1)\}\, o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1). \end{aligned}$$

Finally, $N/|I| \le 16/7$ and (63) imply

$$\sup_{I,h}\left|\frac{N}{|I|}H_N(x_{I,h}) - v_{L,n}^2\right| \le \frac{16}{7}\sup_{I,h}|H_N(x_{I,h}) - H_N(Z_k\mathbf{1}_I)| + \sup_{I}\left|\frac{N}{|I|}H_N(Z_k\mathbf{1}_I) - v_{L,n}^2\right| = o_{\mathbb{P}}(1).$$

A sign change of $\widehat{\nu}_k$ leaves the Bartlett estimate unchanged. On $\mathcal{E}_{\mathrm{unif}}$, its sign-aligned version satisfies $\|\widehat{\nu}_k - \nu_k\|_2 \le \rho_{n,k}$ by Theorem 10. Evaluating the uniform bounds at $I = \widehat{I}$ and using the estimator identity above therefore gives

$$\widehat{\omega}_{k,L}^2 - v_{L,n}^2 \xrightarrow{\mathbb{P}} 0, \qquad v_{L,n}^2 \longrightarrow \omega_{k,L}^2,$$

since $\mathbb{P}(\mathcal{E}_{\mathrm{unif}}^c) \to 0$. □

Proof of Theorem 6. **Step 1: Consistent scales.** Fix k and write $\widehat{\boldsymbol{\vartheta}} = (\widehat{\omega}_{k,L}, \widehat{\omega}_{k,R})$. On $\mathcal{E}_{\mathrm{unif}}$, (16) gives, for all sufficiently large n,

$$|\widehat{\kappa}_k - \kappa_k| \le \|\widehat{\Theta}_k - \Theta_k\|_{\mathrm{op}} \le 2\epsilon_n = o(\kappa_k), \qquad \widehat{\kappa}_k \ge \kappa_k/2 > 0.$$

Together with Theorem 13 and $\mathbb{P}(\mathcal{E}_{\mathrm{unif}}^c) \le n^{-5}$, this yields

$$\widehat{\kappa}_k/\kappa_k \xrightarrow{\mathbb{P}} 1, \qquad \widehat{\boldsymbol{\vartheta}} \xrightarrow{\mathbb{P}} (\omega_{k,L}, \omega_{k,R}) \in (0,\infty)^2.$$

The corresponding failure probability satisfies

$$\mathbb{P}\big((\mathcal{E}_{\mathrm{unif}} \cap \{\widehat{\boldsymbol{\vartheta}} \in (0,\infty)^2\})^c\big) \le \mathbb{P}(\mathcal{E}_{\mathrm{unif}}^c) + \sum_{\ell\in\{L,R\}} \mathbb{P}\left(\mathcal{E}_{\mathrm{unif}} \cap \{|\widehat{\omega}_{k,\ell} - \omega_{k,\ell}| \ge \omega_{k,\ell}/2\}\right) = o(1).$$

For fitted quantities below, restrict to $\mathcal{E}_{\mathrm{unif}} \cap \{\widehat{\boldsymbol{\vartheta}} \in (0,\infty)^2\}$; its probability tends to one, so the weak limits are unchanged.

Step 2: Continuity of the Brownian quantiles. Use one pair of independent standard Brownian motions $B_{k,L}, B_{k,R}$ to define, for $\boldsymbol{\vartheta} = (\vartheta_L, \vartheta_R) \in (0,\infty)^2$,

$$\mathcal{H}_{\boldsymbol{\vartheta}}(-t) = t + 2\vartheta_L B_{k,L}(t), \qquad \mathcal{H}_{\boldsymbol{\vartheta}}(t) = t + 2\vartheta_R B_{k,R}(t), \qquad t \ge 0.$$

At the population scales this equals $\mathcal{G}_k$ in (23). For each fixed $\boldsymbol{\vartheta}$, define

$$U(\boldsymbol{\vartheta}) = \arg\min_{u\in\mathbb{R}} \mathcal{H}_{\boldsymbol{\vartheta}}(u).$$

This minimizer is almost surely finite and unique by the proof of Theorem 5. Its CDF and quantiles are

$$F_{\boldsymbol{\vartheta}}(x) = \mathbb{P}\{U(\boldsymbol{\vartheta}) \le x\}, \qquad q_a(\boldsymbol{\vartheta}) = \inf\{x : F_{\boldsymbol{\vartheta}}(x) \ge a\}, \qquad a \in (0,1).$$

The classical joint law of the depth and location of a drifted Brownian minimum (Abramson & Evans, 2014) gives independent minimum depths on the two sides, with exponential rates $(2\vartheta_L^2)^{-1}$ and $(2\vartheta_R^2)^{-1}$. Both depths are positive almost surely, so $\mathbb{P}\{U(\boldsymbol{\vartheta}) = 0\} = 0$. For $t > 0$, independence of the two halves yields the density

$$f_{\boldsymbol{\vartheta}}(t) = \int_0^\infty \frac{m\exp\{-(m+t)^2/(8\vartheta_R^2 t)\}}{4\vartheta_R^3\sqrt{2\pi t^3}}\left(1 - e^{-m/(2\vartheta_L^2)}\right) dm > 0.$$

Here the fraction is the joint density of the right minimum's depth m and location t; the second factor is the probability that the left minimum has smaller depth. Exchanging L and R gives $f_{\boldsymbol{\vartheta}}(-t)$. Thus $U(\boldsymbol{\vartheta})$ has no atoms, and

$$F_{\boldsymbol{\vartheta}}(y) - F_{\boldsymbol{\vartheta}}(x) = \int_x^y f_{\boldsymbol{\vartheta}}(t)\,dt > 0, \qquad x < y.$$

Hence $F_{\boldsymbol{\vartheta}}$ is continuous and strictly increasing.

For $\boldsymbol{\vartheta}_j \to \boldsymbol{\vartheta}$, the common Brownian paths give, almost surely, for every $T > 0$,

$$\sup_{|u|\le T} |\mathcal{H}_{\boldsymbol{\vartheta}_j}(u) - \mathcal{H}_{\boldsymbol{\vartheta}}(u)| \le 2\|\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}\|_\infty \max_{\ell\in\{L,R\}} \sup_{0\le t\le T} |B_{k,\ell}(t)| \longrightarrow 0.$$

The coefficient vectors are bounded, and $B_{k,\ell}(t)/t \to 0$ almost surely by the proof of Theorem 5. Thus, for some finite C and almost surely finite T_0,

$$\mathcal{H}_{\boldsymbol{d}}(\pm t) \ge t - 2C \max_{\ell\in\{L,R\}} |B_{k,\ell}(t)| \ge t/2, \quad t \ge T_0, \quad \boldsymbol{d} \in \{\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \ldots\}.$$

Since $\mathcal{H}_{\boldsymbol{d}}(0) = 0$, all minimizers lie in $[-T_0, T_0]$. Continuity and uniqueness give, for every $\varepsilon > 0$,

$$\inf_{\substack{|u|\le T_0 \\ |u - U(\boldsymbol{\vartheta})|\ge\varepsilon}} \{\mathcal{H}_{\boldsymbol{\vartheta}}(u) - \mathcal{H}_{\boldsymbol{\vartheta}}(U(\boldsymbol{\vartheta}))\} > 0.$$

Uniform convergence therefore implies $U(\boldsymbol{\vartheta}_j) \to U(\boldsymbol{\vartheta})$ almost surely and $F_{\boldsymbol{\vartheta}_j}(x) \to F_{\boldsymbol{\vartheta}}(x)$ for every x. For $a \in (0,1)$ and $\varepsilon > 0$, CDF convergence and strict increase give brackets proving continuity of q_a:

$$\begin{aligned} &F_{\boldsymbol{\vartheta}}(q_a(\boldsymbol{\vartheta}) - \varepsilon) < a < F_{\boldsymbol{\vartheta}}(q_a(\boldsymbol{\vartheta}) + \varepsilon), \\ &\quad q_a(\boldsymbol{\vartheta}) - \varepsilon < q_a(\boldsymbol{\vartheta}_j) \le q_a(\boldsymbol{\vartheta}) + \varepsilon \qquad \text{for all sufficiently large } j. \end{aligned}$$

Step 3: Coverage. Take $\boldsymbol{\vartheta} = (\omega_{k,L}, \omega_{k,R})$. The Brownian motions in (26) are independent of the observations, so Steps 1–2 give $\widehat{q}_{k,a} = q_a(\widehat{\boldsymbol{\vartheta}}) \xrightarrow{\mathbb{P}} q_a(\boldsymbol{\vartheta})$ for $a \in \{\alpha/2, 1-\alpha/2\}$. Together with Theorem 5, Slutsky's theorem yields

$$\big(\widehat{\kappa}_k^2(\widehat{\eta}_k - \eta_k), \widehat{q}_{k,\alpha/2}, \widehat{q}_{k,1-\alpha/2}\big) \xrightarrow{d} \big(U(\boldsymbol{\vartheta}), q_{\alpha/2}(\boldsymbol{\vartheta}), q_{1-\alpha/2}(\boldsymbol{\vartheta})\big).$$

On the restriction in Step 1, $\widehat{\kappa}_k > 0$, so (27) gives

$$\eta_k \in \mathcal{J}_{k,\alpha} \quad \Longleftrightarrow \quad \widehat{q}_{k,\alpha/2} \le \widehat{\kappa}_k^2(\widehat{\eta}_k - \eta_k) \le \widehat{q}_{k,1-\alpha/2}.$$

Since $\mathbb{P}\{U(\boldsymbol{\vartheta}) = q_a(\boldsymbol{\vartheta})\} = 0$ and $F_{\boldsymbol{\vartheta}}(q_a(\boldsymbol{\vartheta})) = a$, the limiting event has boundary probability zero. Therefore,

$$\begin{aligned} &\mathbb{P}\{\mathcal{E}_{\text{unif}},\ \widehat{\boldsymbol{\vartheta}} \in (0,\infty)^2,\ \eta_k \in \mathcal{J}_{k,\alpha}\} \\ &\qquad \longrightarrow F_{\boldsymbol{\vartheta}}(q_{1-\alpha/2}(\boldsymbol{\vartheta})) - F_{\boldsymbol{\vartheta}}(q_{\alpha/2}(\boldsymbol{\vartheta})) = 1 - \alpha. \end{aligned}$$

The excluded probability is $o(1)$ by Step 1, proving the claim. □

F Auxiliary results

F.1 Uniform empirical-moment and tensor CUSUM bounds

The following maximal deviation bound is a direct consequence of Theorem 1 and Remarks 1–3 of Merlevède et al. (2011).

Proposition 14 (Sub-Weibull deviation under generalized-exponential α-mixing)**.** *Let $N \ge 1$, and let $\{V_i\}_{i=1}^N$ be centered random variables. Suppose the α-mixing coefficients $\alpha_V(\ell)$ of $\{V_i\}_{i=1}^N$, defined as in* (19)*, satisfy*

$$\alpha_V(\ell) \le b_0 \exp(-b_1 \ell^{\gamma_1}), \qquad 1 \le \ell < N,$$

for some $b_0, b_1 > 0$ and $\gamma_1 \in (0, \infty]$. When $\gamma_1 = \infty$, this bound means that the variables are independent. Let $\gamma_2 \in (0, \infty]$ and define the uniform tail envelope by

$$\sigma_{V,\gamma_2} = \begin{cases} \max\limits_{1\le t\le N} \|V_t\|_{\psi_{\gamma_2}}, & \gamma_2 < \infty, \\ \max\limits_{1\le t\le N} \|V_t\|_\infty, & \gamma_2 = \infty, \end{cases} < \infty.$$

Assume that

$$\bar{\gamma} = \left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2}\right)^{-1} < 1,$$

where $1/\infty = 0$*; the boundary case* $(\gamma_1, \gamma_2) = (\infty, 1)$ *is also allowed. Then, for every* $x \geq 1$*, with probability at least* $1 - e^{-x}$*,*

$$\max_{1 \leq m \leq N} \left| \sum_{t=1}^{m} V_t \right| \leq C\sigma_{V,\gamma_2} \left\{ \sqrt{Nx} + (x + \log N)^{1/\gamma_1 + 1/\gamma_2} \right\}. \tag{64}$$

Here $C > 0$ *depends only on* $\gamma_1, \gamma_2, b_0, b_1$*; when* $\gamma_2 = \infty$*, the dependence on* γ_2 *is omitted, and when* $\gamma_1 = \infty$*, the dependence on* γ_1, b_0, b_1 *is omitted.*

Proof of Theorem 14. Theorem 1 of Merlevède et al. (2011) is stated for $N \geq 4$. To cover the remaining cases, note that, for $1 \leq N \leq 3$, the tail envelope σ_{V,γ_2} and a union bound over at most three variables give the stated inequality. Hence assume $N \geq 4$ below.

If $\gamma_1 = \infty$, V_t are independent and $\gamma_2 \leq 1$. The independent Bernstein bounds in equations (1.2)–(1.3) of Merlevède et al. (2011), integrated over the tail and combined with Doob's maximal inequality for the partial-sum martingale, give for every $q \geq 2$

$$\left\| \max_{1 \leq m \leq N} \left| \sum_{t=1}^{m} V_t \right| \right\|_{L^q} \leq C\sigma_{V,\gamma_2} \left\{ \sqrt{Nq} + (q + \log N)^{1/\gamma_2} \right\}.$$

Taking $q = 2 \vee x$ and applying Markov's inequality proves (64), with $1/\gamma_1 = 0$. Hence assume $\gamma_1 < \infty$ below.

If $\sigma_{V,\gamma_2} = 0$, all V_t vanish almost surely and the conclusion is immediate. Otherwise, normalize by σ_{V,γ_2} and extend the sequence by zero outside $\{1, \ldots, N\}$. For $\gamma_2 < \infty$, the moment definition in Additional notation in the appendix gives

$$\max_{1 \leq t \leq N} \|V_t/\sigma_{V,\gamma_2}\|_{L^q} \leq q^{1/\gamma_2}, \qquad q \geq 1.$$

For $z \geq 1$, Markov's inequality with $q = z^{\gamma_2}$ yields

$$\max_{1 \leq t \leq N} \mathbb{P}\big(|V_t| > e\,\sigma_{V,\gamma_2} z\big) \leq \left(\frac{q^{1/\gamma_2}}{ez}\right)^q = e^{-z^{\gamma_2}}.$$

For $0 \leq z < 1$, use the trivial bound by one. Thus, for all $z \geq 0$,

$$\max_{1 \leq t \leq N} \mathbb{P}\big(|V_t| > e\,\sigma_{V,\gamma_2} z\big) \leq \exp(1 - z^{\gamma_2}),$$

which is the required uniform generalized exponential tail bound after a fixed rescaling. If $\gamma_2 = \infty$, then $|V_t|/\sigma_{V,\infty} \leq 1$ almost surely, so the normalized tail probability is zero for $z \geq 1$.

Theorem 1 of Merlevède et al. (2011) is stated under τ-mixing. As explained in their Remark 1, its dependence condition also follows from α-mixing: the comparison between the two coefficients, together with the uniformly bounded second moments of the normalized variables, gives $\tau(\ell) \leq C\alpha_V(\ell)^{1/2} \leq C\exp(-c\ell^{\gamma_1})$. Thus the required τ-mixing decay holds with the same exponent γ_1, after adjusting constants. The zero extension preserves these bounds and has $\alpha_V(\ell) = 0$ for $\ell \geq N$.

We next bound the variance proxy in Theorem 1 of Merlevède et al. (2011). Let $Q(v)$ be the envelope of the upper quantile functions of the absolute normalized variables. The tail condition gives $Q(v) \leq C\{\log(e/v)\}^{1/\gamma_2}$, with $1/\infty = 0$. By Remark 3 of Merlevède et al. (2011), the α-mixing bound therefore yields

$$V \leq C\left\{1 + \sum_{\ell \geq 1} \int_0^{2\alpha_V(\ell)} Q(v)^2\,dv\right\} \leq C\left\{1 + \sum_{\ell \geq 1} \alpha_V(\ell)^{1/2}\right\} \leq C'.$$

Thus Theorem 1 of the cited paper applies with $\bar{\gamma} = (1/\gamma_1 + 1/\gamma_2)^{-1}$.

For $x \geq 1$, substitute

$$u = C_* \left\{ \sqrt{Nx} + (x + \log N)^{1/\bar{\gamma}} \right\}$$

into that theorem. For a sufficiently large C_*, its first probability term is at most $e^{-x}/3$ because $u^{\bar{\gamma}} \geq C(x + \log N)$; the second and third are each at most $e^{-x}/3$ because $V \leq C'$ and $u^2/N \geq Cx$. Rescaling by σ_{V,γ_2} and using $1/\bar{\gamma} = 1/\gamma_1 + 1/\gamma_2$ proves (64). □

The concentration arguments below use Proposition 14. We record the tensor-product tail and product-net reductions needed in the later proofs.

Lemma 15 (Tensor-product tails)**.** *Under Assumption 1, fix $u_1, \ldots, u_p \in \mathbb{S}^D$. The variables*

$$V_t = \prod_{a=1}^{p} u_a^\top Y_t - \mathbb{E} \prod_{a=1}^{p} u_a^\top Y_t$$

are centered and satisfy $\sup_t \|V_t\|_{\psi_{2/p}} \leq C_p \sigma^p$. Their α-mixing coefficients are at most $b_0 \exp(-b_1 \ell^\gamma)$ for $1 \leq \ell < n$.

Proof of Theorem 15. Write $u_a = (c_a, w_a^\top)^\top \in \mathbb{S}^D$, with $w_a \in \mathbb{R}^D$. Assumption 1, the triangle inequality, and Cauchy–Schwarz give

$$\left\| u_a^\top Y_t \right\|_{\psi_2} \leq |c_a| + \sigma \|w_a\|_2 \leq \sqrt{1 + \sigma^2} \leq \sqrt{2}\, \sigma,$$

where $c_a^2 + \|w_a\|_2^2 = 1$ and $\sigma \geq 1$. The moment definition of the ψ_2 norm in Additional notation in the appendix therefore gives, for $u \geq 1$,

$$\left\| u_a^\top Y_t \right\|_{L^{pu}} \leq \sqrt{pu}\, \left\| u_a^\top Y_t \right\|_{\psi_2} \leq \sigma \sqrt{2pu}.$$

See also Vershynin (2018, Proposition 2.5.2(ii)) for this moment characterization of sub-Gaussian variables. Hölder's inequality now yields

$$\left\| \prod_{a=1}^{p} u_a^\top Y_t \right\|_{L^u} \leq (\sigma \sqrt{2pu})^p.$$

Centering costs at most a factor two. The mixing assertion follows because measurable transformations cannot increase α-mixing coefficients. □

For $I = (s, e]$, recall the empirical moment tensor $\overline{\mathcal{M}}_I^{(p)}$ and the CUSUM statistic $\mathcal{T}_I(q)$ defined in Section 2.1.

Lemma 16 (Fixed interval-direction concentration)**.** *Let $I = (s, e] \subseteq (0, n]$ be a fixed nonempty interval, and let $p \geq 2$ be fixed. Under Assumption 1, fix $u_1, \ldots, u_p \in \mathbb{S}^D$. Then, for every $x \geq 1$, with probability at least $1 - e^{-x}$,*

$$\left| \left\langle \overline{\mathcal{M}}_I^{(p)} - \mathbb{E}\overline{\mathcal{M}}_I^{(p)}, u_1 \otimes \cdots \otimes u_p \right\rangle_{\mathrm{F}} \right| \leq C\sigma^p \left\{ \sqrt{\frac{x}{|I|}} + \frac{(x + \log |I|)^{p/2 + 1/\gamma}}{|I|} \right\}. \tag{65}$$

Proof of Lemma 16. Set

$$V_t = \prod_{j=1}^{p} \langle Y_t, u_j \rangle - \mathbb{E} \prod_{j=1}^{p} \langle Y_t, u_j \rangle .$$

Lemma 15 gives $\max_{t \in I} \|V_t\|_{\psi_{2/p}} \leq C_p \sigma^p$ and bounds the α-mixing coefficients of $\{V_t\}_{t \in I}$ by the bound in Assumption 1. Moreover,

$$\left\langle \overline{\mathcal{M}}_I^{(p)} - \mathbb{E}\overline{\mathcal{M}}_I^{(p)}, u_1 \otimes \cdots \otimes u_p \right\rangle_{\mathrm{F}} = \frac{1}{|I|} \sum_{t \in I} V_t. \tag{66}$$

After reindexing I, apply Proposition 14 with

$$\gamma_1 = \gamma, \qquad \gamma_2 = 2/p, \qquad N = |I|, \qquad \sigma_{V,\gamma_2} = \max_{t \in I} \|V_t\|_{\psi_{\gamma_2}} \leq C_p \sigma^p.$$

Since $p \geq 2$ and $\gamma \in [1, \infty)$, $1/\gamma_1 + 1/\gamma_2 = p/2 + 1/\gamma > 1$. Note that the temporally independent case, i.e. $\gamma = \infty$, is also included in the proposition. Therefore the case $m = N$ of Proposition 14 and (66), after division by N, give (65). □

Lemma 17 (Approximating the tensor operator norm by a product net)**.** *Let $\mathcal{M} \in (\mathbb{R}^{D+1})^{\otimes p}$, let $\mathcal{N}_\varepsilon \subset \mathbb{S}^D$ be a finite Euclidean ε-net, and suppose $p\varepsilon < 1$. Then*

$$\|\mathcal{M}\|_{\mathrm{op}} \le \frac{1}{1-p\varepsilon} \max_{v_1,\dots,v_p \in \mathcal{N}_\varepsilon} \left|\langle \mathcal{M}, v_1 \otimes \cdots \otimes v_p \rangle_{\mathrm{F}}\right|. \tag{67}$$

Moreover, such a net exists with cardinality at most $(1+2/\varepsilon)^{D+1}$.

Proof of Lemma 17. Fix $u_1, \dots, u_p \in \mathbb{S}^D$. For each j, choose $v_j \in \mathcal{N}_\varepsilon$ satisfying

$$\|u_j - v_j\|_2 \le \varepsilon. \tag{68}$$

A telescoping expansion gives

$$\begin{aligned}&\langle \mathcal{M}, u_1 \otimes \cdots \otimes u_p \rangle_{\mathrm{F}} - \langle \mathcal{M}, v_1 \otimes \cdots \otimes v_p \rangle_{\mathrm{F}} \\ &\quad = \sum_{j=1}^{p} \langle \mathcal{M}, v_1 \otimes \cdots \otimes v_{j-1} \otimes (u_j - v_j) \otimes u_{j+1} \otimes \cdots \otimes u_p \rangle_{\mathrm{F}}. \end{aligned} \tag{69}$$

By multilinearity and the definition of $\|\mathcal{M}\|_{\mathrm{op}}$, $\|u_i\|_2 = \|v_i\|_2 = 1$, and (68), every summand in (69) has absolute value at most

$$\|\mathcal{M}\|_{\mathrm{op}} \prod_{i<j} \|v_i\|_2 \, \|u_j - v_j\|_2 \prod_{i>j} \|u_i\|_2 \le \varepsilon \, \|\mathcal{M}\|_{\mathrm{op}}.$$

Hence

$$\left|\langle \mathcal{M}, u_1 \otimes \cdots \otimes u_p \rangle_{\mathrm{F}}\right| \le \left|\langle \mathcal{M}, v_1 \otimes \cdots \otimes v_p \rangle_{\mathrm{F}}\right| + p\varepsilon \, \|\mathcal{M}\|_{\mathrm{op}}. \tag{70}$$

Taking the supremum over $u_1, \dots, u_p \in \mathbb{S}^D$ in (70) and rearranging proves (67). The cardinality assertion follows from the standard covering number bound (see e.g. Vershynin, 2018, Corollary 4.2.13). □

Theorem 18 (Uniform empirical-moment and tensor CUSUM bounds)**.** *Let $n \ge 2$. Suppose Assumption 1 holds and $p \ge 2$ is fixed. There exist constants $C_1, C_2 > 0$, depending only on p, γ, b_0, b_1, such that the following holds. With $I = (s, e] \subseteq (0, n]$ ranging over all nonempty index intervals, define events*

$$\mathcal{E}_{\mathrm{unif}} = \left\{ \sup_{|I| \ge 1} \frac{\left\| \overline{\mathcal{M}}_I^{(p)} - \mathbb{E}\overline{\mathcal{M}}_I^{(p)} \right\|_{\mathrm{op}}}{\sqrt{(D+\log n)/|I|} + (D+\log n)^{p/2+1/\gamma}/|I|} \le C_1 \sigma^p \right\}, \tag{71}$$

$$\mathcal{E}_{\mathrm{cusum}} = \left\{ \sup_{\substack{I=(s,e] \\ 1 \le h \le |I|/2}} \max_{s+h \le q \le e-h} \frac{\|\mathcal{T}_I(q) - \mathbb{E}\mathcal{T}_I(q)\|_{\mathrm{op}}}{\sqrt{D+\log n} + (D+\log n)^{p/2+1/\gamma}/\sqrt{h}} \le C_2 \sigma^p \right\}. \tag{72}$$

Then $\mathcal{E}_{\mathrm{unif}} \subseteq \mathcal{E}_{\mathrm{cusum}}$ and

$$\mathbb{P}(\mathcal{E}_{\mathrm{unif}} \cap \mathcal{E}_{\mathrm{cusum}}) = \mathbb{P}(\mathcal{E}_{\mathrm{unif}}) \ge 1 - n^{-5}.$$

Proof of Theorem 18. Set $\nu = p/2 + 1/\gamma$ and $\varepsilon = (2p)^{-1}$, and choose a Euclidean ε-net $\mathcal{N}_\varepsilon \subset \mathbb{S}^D$ with $|\mathcal{N}_\varepsilon| \le (1+4p)^{D+1}$. Apply Lemma 16 with

$$y_0 = 7\log n + p(D+1)\log(1+4p)$$

to every nonempty integer interval and every element of $\mathcal{N}_\varepsilon^p$. There are fewer than $n^2(1+4p)^{p(D+1)}$ interval–direction pairs, so a union bound makes their total failure probability at most n^{-5}. On the complementary event, Lemma 17 gives, simultaneously for all nonempty intervals,

$$\begin{aligned}\left\| \overline{\mathcal{M}}_I^{(p)} - \mathbb{E}\overline{\mathcal{M}}_I^{(p)} \right\|_{\mathrm{op}} &\le C\sigma^p \left\{ \sqrt{\frac{y_0}{|I|}} + \frac{(y_0 + \log|I|)^\nu}{|I|} \right\} \\ &\le C\sigma^p \left\{ \sqrt{\frac{D+\log n}{|I|}} + \frac{(D+\log n)^\nu}{|I|} \right\}.\end{aligned}$$

The second inequality uses $\log|I| \le \log n$, $D+1 \le 2D$, and fixed p. Thus, after enlarging C_1 if necessary, $\mathbb{P}(\mathcal{E}_{\mathrm{unif}}) \ge 1 - n^{-5}$.

Work on $\mathcal{E}_{\mathrm{unif}}$ and set $L = D + \log n$. Fix an admissible interval $I = (s, e]$, an integer h, and $q \in \{s+h, \dots, e-h\}$. Put

$$\ell = q - s, \qquad r = e - q.$$

The CUSUM definition gives the exact identity

$$\mathcal{T}_I(q) - \mathbb{E}\mathcal{T}_I(q) = \sqrt{\frac{\ell r}{\ell + r}} \left\{ (\overline{\mathcal{M}}^{(p)}_{(s,q]} - \mathbb{E}\overline{\mathcal{M}}^{(p)}_{(s,q]}) - (\overline{\mathcal{M}}^{(p)}_{(q,e]} - \mathbb{E}\overline{\mathcal{M}}^{(p)}_{(q,e]}) \right\}.$$

Because $\ell, r \ge h \ge 1$, both child intervals are nonempty, so the defining bound for $\mathcal{E}_{\mathrm{unif}}$ applies to each. The triangle inequality therefore gives

$$\|\mathcal{T}_I(q) - \mathbb{E}\mathcal{T}_I(q)\|_{\mathrm{op}} \le C\sigma^p \sqrt{\frac{\ell r}{\ell + r}} \left\{ \sqrt{\frac{L}{\ell}} + \sqrt{\frac{L}{r}} + \frac{L^\nu}{\ell} + \frac{L^\nu}{r} \right\}.$$

The Gaussian coefficients satisfy

$$\sqrt{\frac{\ell r}{\ell + r}} \left(\frac{1}{\sqrt{\ell}} + \frac{1}{\sqrt{r}} \right) = \sqrt{\frac{r}{\ell + r}} + \sqrt{\frac{\ell}{\ell + r}} \le \sqrt{2},$$

and the large-deviation coefficients satisfy

$$\sqrt{\frac{\ell r}{\ell + r}} \left(\frac{1}{\ell} + \frac{1}{r} \right) \le \frac{1}{\sqrt{\ell}} + \frac{1}{\sqrt{r}} \le \frac{2}{\sqrt{h}}.$$

Since I, h, and q were arbitrary, enlarging C_2 if necessary gives $\mathcal{E}_{\mathrm{unif}} \subseteq \mathcal{E}_{\mathrm{cusum}}$. Combining this inclusion with the probability bound for $\mathcal{E}_{\mathrm{unif}}$ completes the proof. □

F.2 Results on seeded intervals

Lemma 19 (Size of the seeded interval collection)**.** *The seeded interval collection $\mathcal{I}$ in* (8) *satisfies*

$$|\mathcal{I}| \le \frac{n}{4} \quad \textit{and} \quad \sum_{I \in \mathcal{I}} |I| \le Cn \log n. \tag{73}$$

Proof of Lemma 19. Recall the definitions given in Section 2.1 including the dyadic interval scale set $\mathcal{L}_n$ and the seeded interval $\mathcal{I}_\ell$ at scale ℓ. At a dyadic interval scale $\ell \in \mathcal{L}_n$, we have

$$|\mathcal{I}_\ell| \le \left\lfloor \frac{2(n-\ell)}{\ell} \right\rfloor + 2 \le \frac{2n}{\ell}.$$

Therefore

$$|\mathcal{I}| \le \sum_{\ell \in \mathcal{L}_n} \frac{2n}{\ell} \le 2n \sum_{r=4}^{\infty} 2^{-r} = \frac{n}{4},$$

and

$$\sum_{I \in \mathcal{I}} |I| \le \sum_{\ell \in \mathcal{L}_n} \ell\, |\mathcal{I}_\ell| \le 2n\, |\mathcal{L}_n| \le Cn \log n.$$

Removing duplicate intervals can only decrease both quantities. □

Lemma 20 (Isolating seeded interval)**.** *If $\Delta_{\min} \ge 256$, then, for every $k = 1, \dots, K$, there exists $I_k^* = (a_k^*, b_k^*] \in \mathcal{I}$ that contains η_k and no other change point and satisfies*

$$\frac{\Delta_{\min}}{32} < |I_k^*| \le \frac{\Delta_{\min}}{16}, \tag{74}$$

and

$$\min\{\eta_k - a_k^*, b_k^* - \eta_k\} \ge \frac{|I_k^*|}{4}. \tag{75}$$

Proof of Lemma 20. **Step 1: Choice of scale.** Let ℓ be the largest dyadic length satisfying $\ell \le \Delta_{\min}/16$. Because $256 \le \Delta_{\min} \le n$, we have $16 \le \ell \le n$, so $\ell \in \mathcal{L}_n$. The next larger dyadic length satisfies $2\ell > \Delta_{\min}/16$, and therefore

$$\frac{\Delta_{\min}}{32} < \ell \le \frac{\Delta_{\min}}{16}.$$

This proves (74) once an interval of length ℓ is selected.

Step 2: Choice of interval. For $j \in \mathbb{Z}$, the central half of the grid interval $(j\ell/2, j\ell/2 + \ell]$ is

$$H_j = (j\ell/2 + \ell/4,\ j\ell/2 + 3\ell/4]\,.$$

The intervals $\{H_j : j \in \mathbb{Z}\}$ cover the real line, so choose j such that $\eta_k \in H_j$, and set $a_k^* = j\ell/2$, $b_k^* = a_k^* + \ell$, and $I_k^* = (a_k^*, b_k^*]$. It remains to check that this grid interval belongs to the finite collection $\mathcal{I}_\ell$. Since $\eta_k \ge \Delta_{k-1} \ge \Delta_{\min}$ and $n - \eta_k \ge \Delta_k \ge \Delta_{\min}$, while $\Delta_{\min} \ge 16\ell$, central-half membership gives

$$a_k^* \ge \eta_k - 3\ell/4 \ge 0, \quad \text{and} \quad b_k^* \le \eta_k + 3\ell/4 \le n.$$

Thus $j \ge 0$ and $j \le \lfloor 2(n-\ell)/\ell \rfloor$, which proves $I_k^* \in \mathcal{I}_\ell \subseteq \mathcal{I}$.

Step 3: Margins and isolation. The inclusion $\eta_k \in H_j$ yields

$$\eta_k - a_k^* \ge \ell/4, \quad \text{and} \quad b_k^* - \eta_k \ge \ell/4.$$

This proves (75). Finally, $\ell < \Delta_{\min}$ and I_k^* contains η_k, whereas the adjacent changes are at distances $\Delta_{k-1} \ge \Delta_{\min}$ and $\Delta_k \ge \Delta_{\min}$ from η_k. Hence I_k^* lies strictly between η_{k-1} and η_{k+1} and contains no change other than η_k. □

F.3 Estimating the most significant projection direction

We use the distance $d_\pm$ defined in Additional notation in the appendix. Throughout this subsection, we consider order-p moment tensors and omit their superscript (p).

We impose the following eigengap type of assumption such that the leading projection direction is identifiable.

Assumption 7. *There exist fixed constants $c_0, r_0 > 0$ such that, for every $k = 1, \dots, K$ and $0 < \rho \le r_0$,*

$$\kappa_k - \sup_{\substack{u \in \mathbb{S}^D \\ d_\pm(u, \nu_k) \ge \rho}} \left|\left\langle \Theta_k, u^{\otimes p} \right\rangle_{\mathrm{F}}\right| \ge c_0 \kappa_k \rho^2.$$

Similar assumptions have been considered in the study of tensor eigenvectors and best rank-one approximations by Kolda & Mayo (2011), Qi et al. (2016), and Jaffe et al. (2018). Theorem 21 shows that Assumption 7 holds for tensors admitting an orthogonal symmetric CP decomposition with a suitable eigengap assumption.

Lemma 21. *Suppose $2 \le R \le D + 1$ and the symmetric tensor $\mathcal{T} \in (\mathbb{R}^{D+1})^{\otimes p}$ admits the orthogonal symmetric CP decomposition*

$$\mathcal{T} = \sum_{j=1}^{R} \lambda_j \phi_j^{\otimes p},$$

where $\phi_1, \dots, \phi_R \in \mathbb{R}^{D+1}$ are orthonormal. Define the strictly positive weight gap by

$$\Delta = |\lambda_1| - \max_{2 \le j \le R} |\lambda_j| > 0. \tag{76}$$

Then ϕ_1 is the unique maximizer, up to sign, of

$$u \longmapsto \left|\left\langle \mathcal{T}, u^{\otimes p} \right\rangle_{\mathrm{F}}\right|, \qquad u \in \mathbb{S}^D.$$

Moreover, Assumption 7 holds with $r_0 = 1$ and $c_0 = \Delta/(2\,|\lambda_1|)$.

Proof of Theorem 21. For $v \in \mathbb{S}^D$, orthonormality and $p \geq 2$ give

$$\sum_{j=1}^{R} \left|\phi_j^\top v\right|^p \leq \sum_{j=1}^{R} \left|\phi_j^\top v\right|^2 \leq 1.$$

Consequently, generalized Hölder's inequality gives, for unit vectors $v_1, \ldots, v_p$,

$$\left|\langle \mathcal{T}, v_1 \otimes \cdots \otimes v_p \rangle_{\mathrm{F}}\right| \leq |\lambda_1| \prod_{\ell=1}^{p} \left(\sum_{j=1}^{R} \left|\phi_j^\top v_\ell\right|^p \right)^{1/p} \leq |\lambda_1|.$$

Equality is attained at $v_1 = \cdots = v_p = \phi_1$, so $\|\mathcal{T}\|_{\mathrm{op}} = |\lambda_1|$. For any $v \in \mathbb{S}^D$, the weight gap in (76) yields

$$\begin{aligned}\left|\langle \mathcal{T}, v^{\otimes p} \rangle_{\mathrm{F}}\right| &\leq |\lambda_1| \left|\phi_1^\top v\right|^p + (|\lambda_1| - \Delta) \sum_{j=2}^{R} \left|\phi_j^\top v\right|^p \\ &\leq |\lambda_1| - \Delta(1 - \left|\phi_1^\top v\right|^2).\end{aligned}$$

Since $d_\pm(v, \phi_1)^2 = 2(1 - \left|\phi_1^\top v\right|)$, it follows that

$$\|\mathcal{T}\|_{\mathrm{op}} - \left|\langle \mathcal{T}, v^{\otimes p} \rangle_{\mathrm{F}}\right| \geq \frac{\Delta}{2} d_\pm(v, \phi_1)^2.$$

The right side is positive unless $v \in \{\phi_1, -\phi_1\}$. Taking the infimum over $d_\pm(v, \phi_1) \geq \rho$ proves the separation condition with $r_0 = 1$ and $c_0 = \Delta/(2\,|\lambda_1|)$. □

The following deterministic perturbation bound applies to symmetric matrices and to symmetric tensors of higher order.

Lemma 22 (Perturbation of a maximizing direction)**.** *Let $p \geq 2$ and $D \geq 1$, and let $\mathcal{A}, \widehat{\mathcal{A}} \in (\mathbb{R}^{D+1})^{\otimes p}$ be real symmetric tensors. Let*

$$\kappa = \|\mathcal{A}\|_{\mathrm{op}} > 0 \quad \textit{and} \quad \nu \in \arg\max_{u \in \mathbb{S}^D} \left|\langle \mathcal{A}, u^{\otimes p} \rangle_{\mathrm{F}}\right|.$$

Suppose that $c_0, r_0 > 0$ satisfy, for every $0 < \rho \leq r_0$,

$$\kappa - \sup_{\substack{u \in \mathbb{S}^D \\ d_\pm(u,\nu) \geq \rho}} \left|\langle \mathcal{A}, u^{\otimes p} \rangle_{\mathrm{F}}\right| \geq c_0 \kappa \rho^2. \tag{77}$$

Let $\widehat{\nu}$ be any exact maximizer of $\left|\left\langle \widehat{\mathcal{A}}, u^{\otimes p} \right\rangle_{\mathrm{F}}\right|$ over $u \in \mathbb{S}^D$, and define $\varepsilon = \left\|\widehat{\mathcal{A}} - \mathcal{A}\right\|_{\mathrm{op}}$. If

$$\frac{\varepsilon}{\kappa} \leq \min\left\{\frac{1}{4}, \frac{c_0 r_0^2}{4}, \frac{c_0}{16p^2}\right\}, \tag{78}$$

then

$$d_\pm(\widehat{\nu}, \nu) \leq \frac{p\varepsilon}{c_0 \kappa}.$$

Proof of Theorem 22. For a real symmetric tensor, the operator norm equals the maximum absolute diagonal contraction by Banach's identity (Friedland & Lim, 2018, Eq. (5.1)). For every $v \in \mathbb{S}^D$, the definition of ε gives

$$\left|\left\langle \widehat{\mathcal{A}} - \mathcal{A}, v^{\otimes p} \right\rangle_{\mathrm{F}}\right| \leq \varepsilon. \tag{79}$$

We write $d = d_\pm(\widehat{\nu}, \nu)$.

Step 1: localization. Comparison of the empirical objective at ν and $\widehat{\nu}$ yields

$$\kappa - \left|\langle \mathcal{A}, \widehat{\nu}^{\otimes p} \rangle_{\mathrm{F}}\right| \leq \left|\left\langle \widehat{\mathcal{A}}, \nu^{\otimes p} \right\rangle_{\mathrm{F}}\right| - \left|\left\langle \widehat{\mathcal{A}}, \widehat{\nu}^{\otimes p} \right\rangle_{\mathrm{F}}\right| + 2\varepsilon \leq 2\varepsilon.$$

If $d > 0$, separation at $\rho = \min\{d, r_0\}$ gives $c_0\kappa \min\{d, r_0\}^2 \le 2\varepsilon$. Condition (78) therefore forces $d < r_0$ and $d^2 \le 2\varepsilon/(c_0\kappa) \le 1/(8p^2)$. The case $d = 0$ is immediate. Consequently,

$$d < \min\{r_0, 1/(2p)\}. \tag{80}$$

Step 2: sign alignment. Choose $\widetilde{\nu} \in \{\widehat{\nu}, -\widehat{\nu}\}$ so that $\|\widetilde{\nu} - \nu\|_2 = d$. Its absolute population and empirical objectives agree with those of $\widehat{\nu}$ by homogeneity. For any order-p tensor $\mathcal{B}$ and unit vectors u, v, telescoping gives

$$\left|\left\langle \mathcal{B}, u^{\otimes p} - v^{\otimes p}\right\rangle_{\mathrm{F}}\right| = \left|\sum_{a=1}^{p} \left\langle \mathcal{B}, u^{\otimes(a-1)} \otimes (u - v) \otimes v^{\otimes(p-a)}\right\rangle_{\mathrm{F}}\right| \le p\,\|\mathcal{B}\|_{\mathrm{op}}\,\|u - v\|_2\,. \tag{81}$$

It follows from (80) that

$$\left|\left\langle \mathcal{A}, \widetilde{\nu}^{\otimes p} - \nu^{\otimes p}\right\rangle_{\mathrm{F}}\right| \le p\kappa d < \kappa/2.$$

Let $s = \mathrm{sign}\{\langle \mathcal{A}, \nu^{\otimes p}\rangle_{\mathrm{F}}\}$. Then $s \in \{-1, 1\}$ and

$$s\left\langle \mathcal{A}, \widetilde{\nu}^{\otimes p}\right\rangle_{\mathrm{F}} > \kappa/2, \qquad\qquad s\left\langle \mathcal{A}, \nu^{\otimes p}\right\rangle_{\mathrm{F}} = \kappa.$$

By (79) and $\varepsilon \le \kappa/4$, the corresponding empirical contractions satisfy

$$s\left\langle \widehat{\mathcal{A}}, \widetilde{\nu}^{\otimes p}\right\rangle_{\mathrm{F}} > \kappa/4, \qquad\qquad s\left\langle \widehat{\mathcal{A}}, \nu^{\otimes p}\right\rangle_{\mathrm{F}} \ge 3\kappa/4.$$

All four contractions therefore have sign s.

Step 3: the refined objective comparison. Exact maximization and the common empirical sign imply

$$s\left\langle \widehat{\mathcal{A}}, \widetilde{\nu}^{\otimes p}\right\rangle_{\mathrm{F}} \ge s\left\langle \widehat{\mathcal{A}}, \nu^{\otimes p}\right\rangle_{\mathrm{F}}.$$

Using the common population sign, we obtain

$$\begin{aligned}\kappa - \left|\left\langle \mathcal{A}, \widetilde{\nu}^{\otimes p}\right\rangle_{\mathrm{F}}\right| &= s\left\langle \mathcal{A}, \nu^{\otimes p} - \widetilde{\nu}^{\otimes p}\right\rangle_{\mathrm{F}} \\ &\le s\left\langle \widehat{\mathcal{A}} - \mathcal{A}, \widetilde{\nu}^{\otimes p} - \nu^{\otimes p}\right\rangle_{\mathrm{F}} \\ &\le p\varepsilon d,\end{aligned}$$

where the last inequality uses (81).

Step 4: the direction bound. If $d = 0$, there is nothing to prove. Otherwise $0 < d < r_0$, and applying (77) at $\rho = d$ to $\widetilde{\nu}$ gives

$$c_0\kappa d^2 \le \kappa - \left|\left\langle \mathcal{A}, \widetilde{\nu}^{\otimes p}\right\rangle_{\mathrm{F}}\right| \le p\varepsilon d.$$

Dividing by $c_0\kappa d$ proves $d \le p\varepsilon/(c_0\kappa)$. □

Theorem 23 (Accuracy of the estimated direction)**.** *Let $p \ge 2$ and $D \ge 1$ be integers. Let $\mathcal{M}_L, \mathcal{M}_R$ be real symmetric order-p tensors on $\mathbb{R}^{D+1}$, with symmetric estimators $\widehat{\mathcal{M}}_L, \widehat{\mathcal{M}}_R$. Define*

$$\Theta = \mathcal{M}_L - \mathcal{M}_R, \qquad \widehat{\Theta} = \widehat{\mathcal{M}}_L - \widehat{\mathcal{M}}_R, \qquad \kappa = \|\Theta\|_{\mathrm{op}} > 0,$$

and choose exact maximizing directions on the unit sphere $\mathbb{S}^D$

$$\nu \in \arg\max_{u \in \mathbb{S}^D} \left|\left\langle \Theta, u^{\otimes p}\right\rangle_{\mathrm{F}}\right|, \qquad \widehat{\nu} \in \arg\max_{u \in \mathbb{S}^D} \left|\left\langle \widehat{\Theta}, u^{\otimes p}\right\rangle_{\mathrm{F}}\right|.$$

For $u, v \in \mathbb{S}^D$, let $d_{\pm}(u, v) = \min\{\|u - v\|_2, \|u + v\|_2\}$. Suppose $c_0, r_0 > 0$ satisfy, for every $0 < \rho \le r_0$,

$$\kappa - \sup_{\substack{u \in \mathbb{S}^D \\ d_{\pm}(u,\nu) \ge \rho}} \left|\left\langle \Theta, u^{\otimes p}\right\rangle_{\mathrm{F}}\right| \ge c_0\kappa\rho^2.$$

For a deterministic $\epsilon_n > 0$, define

$$\mathcal{E}_{\mathrm{mom}} = \left\{ \max_{\ell \in \{L,R\}} \left\| \widehat{\mathcal{M}}_\ell - \mathcal{M}_\ell \right\|_{\mathrm{op}} \le \epsilon_n \right\}. \tag{82}$$

If

$$\frac{\epsilon_n}{\kappa} \le \min\left\{ \frac{1}{8}, \frac{c_0 r_0^2}{8}, \frac{c_0}{32p^2} \right\}, \tag{83}$$

then, on $\mathcal{E}_{\mathrm{mom}}$, every such $\widehat{\nu}$ satisfies

$$d_{\pm}(\widehat{\nu}, \nu) \le \frac{2p\epsilon_n}{c_0 \kappa}. \tag{84}$$

Proof of Theorem 23. On $\mathcal{E}_{\mathrm{mom}}$,

$$\left\| \widehat{\Theta} - \Theta \right\|_{\mathrm{op}} \le \sum_{\ell \in \{L,R\}} \left\| \widehat{\mathcal{M}}_\ell - \mathcal{M}_\ell \right\|_{\mathrm{op}} \le 2\epsilon_n.$$

Apply Theorem 22 with $\mathcal{A} = \Theta$, $\widehat{\mathcal{A}} = \widehat{\Theta}$, and $\varepsilon \le 2\epsilon_n$. Condition (83) implies (78), giving (84). □

F.4 Maximal bounds for projected moment sums

Lemma 24 (Fourth-moment maximal bound)**.** *Let $V_1, \dots, V_m$ be centered real random variables with $\max_t \|V_t\|_{L^8} \le B$. Suppose their α-mixing coefficients satisfy the bound in Assumption 1, including independence when $\gamma = \infty$. Then*

$$\left\| \max_{0 \le r \le m} \left| \sum_{t=1}^{r} V_t \right| \right\|_{L^4} \le CB\sqrt{m}, \tag{85}$$

where $C > 0$ depends only on the fixed mixing parameters.

Proof of Theorem 24. **Step 1: Fourth moments of interval sums.** For ordered indices $t_1 \le t_2 \le t_3 \le t_4$, let $d = \max\{t_2 - t_1, t_4 - t_3\}$. If $d = t_2 - t_1 \ge 1$, centering, Davydov's covariance inequality (Davydov, 1968) with exponents 8 and $8/3$, and Hölder's inequality give

$$|\mathbb{E}(V_{t_1} V_{t_2} V_{t_3} V_{t_4})| = |\mathrm{Cov}(V_{t_1}, V_{t_2} V_{t_3} V_{t_4})| \le CB^4 \alpha_V(d)^{1/2}.$$

If $d = t_4 - t_3 \ge 1$, isolate V_{t_4} instead. For $d = 0$, Hölder's inequality bounds the product moment by B^4. Hence, for finite γ,

$$|\mathbb{E}(V_{t_1} V_{t_2} V_{t_3} V_{t_4})| \le CB^4 \exp(-cd^\gamma).$$

For a consecutive interval I of ℓ indices,

$$\mathbb{E}\left| \sum_{t \in I} V_t \right|^4 \le 24 \sum_{\substack{t_1, \dots, t_4 \in I \\ t_1 \le t_2 \le t_3 \le t_4}} |\mathbb{E}(V_{t_1} V_{t_2} V_{t_3} V_{t_4})| \le CB^4 \ell^2 \sum_{d=0}^{\infty} (2d+1) e^{-cd^\gamma} \le CB^4 \ell^2. \tag{86}$$

The first inequality uses the fourth-power expansion, the triangle inequality, and at most $4!$ permutations per ordered quadruple. The second uses the preceding moment bound, at most ℓ^2 choices of t_1 and $t_3 - t_2$ for each outer-gap pair $(a, b) = (t_2 - t_1, t_4 - t_3)$, and $2d+1$ pairs with $\max\{a, b\} = d$. The last inequality follows from convergence of the series for finite $\gamma \ge 1$. Under independence, a product expectation can be nonzero only if $t_1 = t_2$ and $t_3 = t_4$. There are at most ℓ^2 such ordered quadruples, each with absolute product expectation at most B^4, so the same bound holds.

Step 2: Dyadic decomposition. Choose $J \ge 0$ with $m \le 2^J < 2m$, and set $V_t = 0$ for $m < t \le 2^J$. For every interval $I \subseteq \{1, \dots, 2^J\}$,

$$\sum_{t \in I} V_t = \sum_{t \in I \cap \{1, \dots, m\}} V_t.$$

The intersection is empty or an interval of length at most $|I|$; hence (86) also holds for the extended sequence.

For $q = 0, \dots, J$ and $b = 0, \dots, 2^q - 1$, define

$$I_{q,b} = \{b2^{J-q} + 1, \dots, (b+1)2^{J-q}\}.$$

Every prefix $\{1, \dots, r\}$ is a disjoint union of at most one such interval at each level q. Thus

$$\max_{0 \le r \le m} \left| \sum_{t=1}^{r} V_t \right| \le \sum_{q=0}^{J} \max_{0 \le b < 2^q} \left| \sum_{t \in I_{q,b}} V_t \right|.$$

Minkowski's inequality, followed by bounding a maximum of fourth powers by their sum, gives

$$\begin{aligned} \left\| \max_{0 \le r \le m} \left| \sum_{t=1}^{r} V_t \right| \right\|_{L^4} &\le \sum_{q=0}^{J} \left\{ \sum_{b=0}^{2^q - 1} \mathbb{E} \left| \sum_{t \in I_{q,b}} V_t \right|^4 \right\}^{1/4} \\ &\le CB\, 2^{J/2} \sum_{q=0}^{J} 2^{-q/4} \le CB\sqrt{m}. \end{aligned}$$

The second inequality uses (86) and $|I_{q,b}| = 2^{J-q}$; the last uses $2^J < 2m$ and $\sum_{q=0}^{\infty} 2^{-q/4} < \infty$. □

The next lemma controls projected moment sums uniformly over directions near the population direction, using forward and reverse sums for candidate locations to the right and left of a true change, respectively.

Lemma 25 (Maximal bounds near a fixed direction)**.** *Suppose Assumption 1 holds and $p \ge 2$ is fixed. Fix deterministic integers $0 \le a < b \le n$ and take either the forward list $(Y_{a+1}, \dots, Y_b)$ or the reverse list $(Y_b, \dots, Y_{a+1})$. Within this lemma, relabel the chosen list as $Y_1, \dots, Y_{b-a}$ and define*

$$S_r = \sum_{t=1}^{r} \{Y_t^{\otimes p} - \mathbb{E} Y_t^{\otimes p}\}, \qquad 1 \le r \le m.$$

Zero centered tensor summands may be included at either end, with $m \ge 1$ denoting the total number of terms. For every fixed $\nu \in \mathbb{S}^D$ and $\rho \ge 0$,

$$\left\| \max_{1 \le r \le m} \left| \langle S_r, \nu^{\otimes p} \rangle_{\mathrm{F}} \right| \right\|_{L^4} \le C\sigma^p \sqrt{m}, \tag{87}$$

$$\left\| \sup_{\|h\|_2 \le \rho} \max_{1 \le r \le m} \left| \langle S_r, (\nu + h)^{\otimes p} - \nu^{\otimes p} \rangle_{\mathrm{F}} \right| \right\|_{L^4} \le C\sigma^p \sqrt{m} \sum_{j=1}^{p} \{\rho\sqrt{D+1}\}^j. \tag{88}$$

Here $C > 0$ depends only on p and the fixed mixing parameters.

Proof of Theorem 25. **Step 1: bounds for contracted tensor sums.** For $j = 1, \dots, p$, define the order-j tensor $A_{r,j} \in (\mathbb{R}^{D+1})^{\otimes j}$ by

$$A_{r,j} = \sum_{t=1}^{r} \left\{ (\nu^\top Y_t)^{p-j} Y_t^{\otimes j} - \mathbb{E}[(\nu^\top Y_t)^{p-j} Y_t^{\otimes j}] \right\},$$

with added summands set to zero, and let $A_{r,0} = \langle S_r, \nu^{\otimes p} \rangle_{\mathrm{F}}$. Each coordinate summand is a centered product of p unit-vector projections, so its L^8 norm is at most $C_p \sigma^p$ by Theorem 15. The mixing bound is inherited under coordinate maps and restriction to a block. Reversal uses $\alpha(\mathcal{A}, \mathcal{B}) = \alpha(\mathcal{B}, \mathcal{A})$; added zeros generate trivial sigma-fields. Thus Theorem 24 yields

$$\left\| \max_{1 \le r \le m} |(A_{r,j})_{\boldsymbol{i}}| \right\|_{L^4} \le C\sigma^p \sqrt{m}$$

for every coordinate $\boldsymbol{i}$. For $j = 0$, the coordinate index is omitted, giving (87). For $j \geq 1$,

$$\begin{aligned}\left\| \max_{1\leq r\leq m} \|A_{r,j}\|_{\mathrm{F}} \right\|_{L^4}^2 &\leq \left\| \sum_{\boldsymbol{i}\in\{1,\ldots,D+1\}^j} \max_{1\leq r\leq m} |(A_{r,j})_{\boldsymbol{i}}|^2 \right\|_{L^2} \\ &\leq \sum_{\boldsymbol{i}\in\{1,\ldots,D+1\}^j} \left\| \max_{1\leq r\leq m} |(A_{r,j})_{\boldsymbol{i}}| \right\|_{L^4}^2 \\ &\leq Cm\sigma^{2p}(D+1)^j. \end{aligned} \tag{89}$$

The first inequality takes the maximum coordinatewise, the second is Minkowski's inequality in L^2, and the last sums the coordinate bounds over $(D+1)^j$ entries.

Step 2: expansion around the fixed direction. Symmetry gives, for every $h \in \mathbb{R}^{D+1}$,

$$\left\langle S_r, (\nu+h)^{\otimes p} - \nu^{\otimes p} \right\rangle_{\mathrm{F}} = \sum_{j=1}^{p} \binom{p}{j} \left\langle A_{r,j}, h^{\otimes j} \right\rangle_{\mathrm{F}}. \tag{90}$$

Since $\left\|h^{\otimes j}\right\|_{\mathrm{F}} = \|h\|_2^j$, Cauchy–Schwarz implies

$$\sup_{\|h\|_2\leq\rho} \max_{1\leq r\leq m} \left|\left\langle S_r, (\nu+h)^{\otimes p} - \nu^{\otimes p} \right\rangle_{\mathrm{F}}\right| \leq \sum_{j=1}^{p} \binom{p}{j} \rho^j \max_{1\leq r\leq m} \|A_{r,j}\|_{\mathrm{F}}.$$

Taking L^4 norms gives

$$\begin{aligned}\left\| \sup_{\|h\|_2\leq\rho} \max_{1\leq r\leq m} \left|\left\langle S_r, (\nu+h)^{\otimes p} - \nu^{\otimes p} \right\rangle_{\mathrm{F}}\right| \right\|_{L^4} &\leq \sum_{j=1}^{p} \binom{p}{j} \rho^j \left\| \max_{1\leq r\leq m} \|A_{r,j}\|_{\mathrm{F}} \right\|_{L^4} \\ &\leq C\sigma^p\sqrt{m} \sum_{j=1}^{p} \binom{p}{j} \{\rho\sqrt{D+1}\}^j \\ &\leq 2^p C\sigma^p\sqrt{m} \sum_{j=1}^{p} \{\rho\sqrt{D+1}\}^j. \end{aligned}$$

The first inequality is Minkowski's inequality, the second follows from (89), and the third uses $\binom{p}{j} \leq 2^p$. Since p is fixed, absorbing 2^p into C proves (88). □

G Additional experimental details

G.1 Tuning parameter selection

For our method, we set $\gamma = 1$ in Algorithm 1 and jointly select the moment order $p \in \{2, 3\}$ and the threshold parameter $\tau \in \{1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 3, 4, 6, 8, 12, 16\}$. For each candidate pair (p, τ), we split the time series into its odd- and even-indexed subsequences, denoted by $X_t^{\mathrm{odd}} = X_{2t-1}$ and $X_t^{\mathrm{even}} = X_{2t}$. We apply Algorithm 1, with the threshold in (10), to the odd-indexed time series and subsequently apply the local refinement procedure in Section 2.2. The resulting change-point estimates produce the estimated segments $\widehat{\mathcal{I}}_1, \ldots, \widehat{\mathcal{I}}_{\widehat{K}}$. To select the candidate pair in a data-driven manner, we evaluate these segments on the even-indexed observations using an empirical kernel score (Steinwart & Ziegel, 2021). Let $\mathcal{H}$ be the reproducing kernel Hilbert space associated with a bounded Gaussian kernel, and let ϕ be its induced feature map. The score is

$$\mathcal{S}(p,\tau) = \sum_{k=1}^{\widehat{K}} \sum_{t\in\widehat{\mathcal{I}}_k} \left\| \phi\left(X_t^{\mathrm{even}}\right) - \frac{1}{|\widehat{\mathcal{I}}_k|} \sum_{s\in\widehat{\mathcal{I}}_k} \phi\left(X_s^{\mathrm{odd}}\right) \right\|_{\mathcal{H}}^2.$$

We select the tuning parameters as $(\widehat{p}, \widehat{\tau}) = \arg\min_{p,\tau} \mathcal{S}(p,\tau)$ and then apply Algorithm 1 to the complete time series using the selected parameters.

G.2 HAUSDORFF DISTANCE

For nonempty change-point sets $A, B \subseteq \{1, \ldots, n-1\}$, define

$$d_H(A, B) = \max\left\{\max_{a \in A} \min_{b \in B} |a - b|, \max_{b \in B} \min_{a \in A} |b - a|\right\}.$$

We apply this distance to the true set $A = \{\eta_1, \ldots, \eta_K\}$ and estimated set $B = \{\widehat{\eta}_1, \ldots, \widehat{\eta}_{\widehat{K}}\}$. It compares both sets without requiring $\widehat{K} = K$ and is reported in observation-index units, without division by n. For empty outputs, the simulation implementation uses the finite convention

$$d_H(A, \varnothing) = d_H(\varnothing, A) = \max_{a \in A} \min\{a, n - a\}, \qquad A \neq \varnothing,$$
$$d_H(\varnothing, \varnothing) = 0.$$

This empty-set convention extends the simulation score; the first display is the usual Hausdorff distance for nonempty sets.

G.3 CONFIDENCE INTERVAL COVERAGE

We report empirical coverage and illustrate the simulation-calibrated 95% intervals for settings R4 and R5.

Table 3: Empirical coverage of simulation-calibrated 95% Tensor intervals, conditional on $\widehat{K} = K = 2$. Entries give covered/eligible repetitions.

Setting	First change point	Second change point
R4	95/100	95/100
R5	96/100	96/100

Figures 2 and 3 show the intervals from the first ten repetitions.

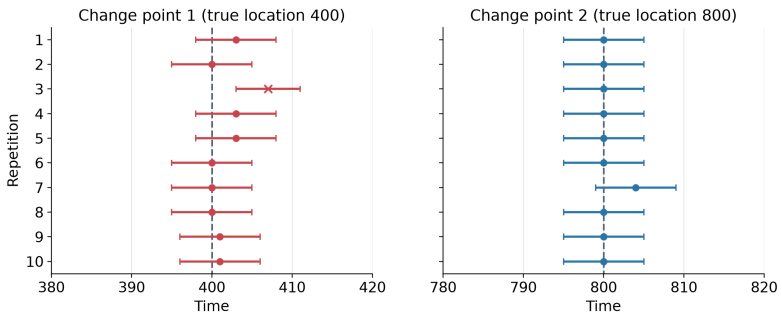


Figure 2: Simulation-calibrated 95% confidence intervals from the Tensor method for the first ten repetitions in R4.

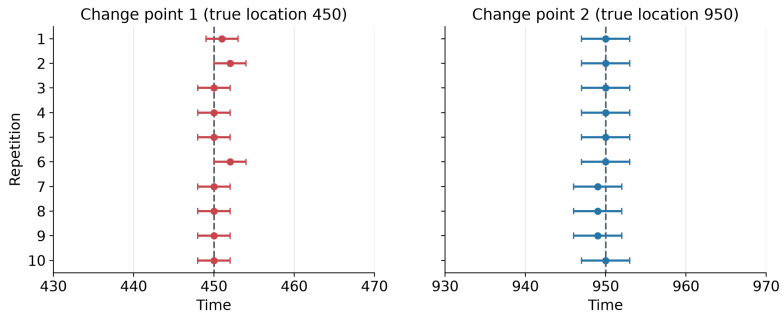


Figure 3: Simulation-calibrated 95% confidence intervals from the Tensor method for the first ten repetitions in R5.