DETECTION AND LOCALIZATION OF MULTI-TYPE PARAMETER CHANGES IN NON-STATIONARY DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper studies a common but complex change-point setting in time series, where a smoothly evolving mean trend may coexist with abrupt shifts in both the mean and variance. To address this joint problem in such complex non-stationary data, we propose a novel iterative estimation procedure that alternates between smoothing-spline estimation of the mean function and likelihood-ratio tests for mean and variance change-points.The resulting method jointly estimates the mean change-point and the jump size of the mean shift, the two smooth mean functions before and after the change-point, the variance change-point, and the variances before and after the change. We derive the asymptotic null distributions of the proposed test statistics and establish convergence rates for the estimators of the smooth mean functions, change-point locations, mean shift size, and variance parameters. Simulation studies and a battery-production case study demonstrate that the proposed procedure is stable across a range of settings and provides accurate change-point detection together with reliable trend estimation.The proposed method is computationally tractable and has potential applications in finance, meteorology, sensor monitoring, economics, and industrial process control.

1 INTRODUCTION

In the literature on parametric change-point analysis, it is often assumed that data are generated from a known distributional family and may exhibit abrupt changes in the mean, variance, or both. Most standard approaches rely on simplifying assumptions and target single-parameter change-point problems: mean change-point detection typically assumes constant variance (Chernoff & Zacks, 1964; Hinkley, 1970; Worsley, 1979; Zhang & Siegmund, 2007; Wang et al., 2020), while variance change-point detection generally assumes a constant mean (Hsu, 1977; Inclan & Tiao, 1994; Chen & Gupta, 1997; Li et al., 2015; Xu et al., 2016; Li et al., 2025; Cappello & Padilla, 2025). In practice, however, such single-parameter change scenarios are relatively uncommon, as multiple distributional characteristics may evolve simultaneously, which limits the applicability of these methods. This limitation motivates the present work, which addresses the more realistic and challenging setting where a smoothly evolving mean trend coexists with abrupt shifts in both the mean and variance.

Beyond these single-parameter frameworks, considerable attention has been devoted to change-point detection in more complex settings. In parametric settings, existing work has considered shifts in location and covariance-type parameters (Hawkins, 1992), normal mean–variance changes (Horvath, 1993), general parametric multiple change-point problems (Pan & Chen, 2006), exponential-family regression models (Frick et al., 2014), general joint mean–variance changes (Messer, 2022), and conditional mean–variance functions (Ghoudi & Laïb, 2025). For mean-function changes, existing methods include jump detection in smooth mean functions (Loader, 1996; Grégoire & Hamrouni, 2002), multiple piecewise-constant mean changes (Fryzlewicz, 2014; Kovács et al., 2023), and change-point-like or piecewise-polynomial mean structures (Baranowski et al., 2019; Fryzlewicz, 2024). Regression-model changes have been studied, ranging from multiple structural changes in linear regression models (Bai & Perron, 1998) and high-dimensional regression with a possible change point (Lee et al., 2016) to changes in high-dimensional regression coefficients and nonsparse regression structures (Rinaldo et al., 2021; Cho et al., 2025). Dependent-data settings cover mean shifts in dependent processes (Lavielle & Moulines, 2000), second-order structure changes (Cho & Fryzlewicz, 2015; Korkas & PryzlewiczV, 2017), distributional change-point inference under mixing conditions (Madrid Padilla et al., 2023), regression change-point inference under temporal dependence (Xu et al.,

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.

2024), and nonlinear serial-dependence changes (McGonigle & Cho, 2025). Functional data settings include functional mean changes and mean change-point localization for sparsely or densely sampled functional observations (Berkes et al., 2009; Aue et al., 2009; Madrid Padilla et al., 2022), as well as covariance-structure changes (Horváth et al., 2022). General distributional changes, where the changing parameter is not specified in advance, have been addressed mainly through nonparametric or distribution-free frameworks (Matteson & James, 2014; Chen & Zhang, 2015; Haynes et al., 2017; Padilla et al., 2021). The closest line of work to the present paper is Gao et al. (2019; 2020; 2026), which developed methods for variance change-point detection under a smoothly evolving mean function. However, the existing approaches reviewed above do not directly target the setting considered here, where a smooth mean trend, abrupt mean shifts, and variance change-points may coexist. Change-point detection in this simultaneous setting thus remains an open and challenging problem.

In this paper, we study change-point detection for data in which the mean and variance may both exhibit abrupt shifts while the underlying mean trend evolves smoothly over time. We propose an iterative approach that incorporates smoothing-spline-based mean estimation with likelihood-test-based mean and variance change-point detection. The main contributions of this paper are summarized as follows:

1. In terms of methodology, we develop a unified inferential framework for volatile data subject to complex parametric changes. This framework addresses a challenging setting not covered by existing methods, characterized by a smoothly evolving mean in which both mean and variance change-points coexist. This scope complements generic change scores or segmentation produced by broader machine-learning methods.
2. In terms of theory, we establish the asymptotic null distributions of the proposed test statistics for both mean and variance change-points, and derive convergence rates for the estimators of the smooth mean functions, change-points, shift size, and variance parameters.

The remainder of this paper is organized as follows. Section 2 introduces the model and notation. Section 3 presents the proposed parametric change detection algorithm, including the testing procedures for both mean and variance changes, as well as the estimation of the smooth mean function. Section 4 presents the main theoretical results, including asymptotic null distributions and convergence rates of the estimators. Section 5 evaluates the finite-sample performance and practical utility of the proposed method through simulation studies and an application to battery production data. Section 6 concludes the paper. All theoretical proofs are provided in the Appendix.

2 Model Setup

Suppose that $y_1, \ldots, y_n$ are independent observations generated from the following model:

$$y_i = \begin{cases} f_-(i/n) + \epsilon_i, & i = 1, \ldots, \tau_0, \\ f_+(i/n) + \epsilon_i, & i = \tau_0 + 1, \ldots, n, \end{cases}$$

where f_- and f_+ are unknown smooth functions, and τ_0 is the mean change-point. Let κ denote the quantity of the sudden shift of the mean functions, i.e., $\kappa = f_+((\tau_0+1)/n) - f_-(\tau_0/n)$. We assume $\epsilon_i \sim N\left(0, \sigma_i^2\right)$, with $\sigma_i = \sigma$ for $i \le \tau_1$ and $\sigma_i = \delta$ for $i > \tau_1$. Both σ^2 and δ^2 are unknown variances. When $\sigma = \delta$, there is no variance change-point; when $\sigma \neq \delta$, τ_1 is the variance change-point. Suppose that both f_- and f_+ belong to a reproducing kernel Hilbert space $\mathcal{H} = \{f \mid f : [0,1] \to \mathbb{R}, J(f) < \infty\}$, where J is a semi-norm on $\mathcal{H}$. In this paper, we consider $J(f) = \int_0^1 \left\{f^{(m)}(t)\right\}^2 dt$, where m is a positive integer. We aim to estimate $(f_-, f_+, \tau_0, \kappa, \tau_1, \sigma^2, \delta^2)$ through the minimization of the following penalized weighted least-squares criterion

$$\begin{cases} \frac{1}{n}(\mathbf{y}_- - \mathbf{f}_-)^T \Sigma^{-1}_{n,\tau_1,\sigma,\delta,\tau_0,-}(\mathbf{y}_- - \mathbf{f}_-) + \lambda_- J(f_-), & i \le \tau_0, \\ \frac{1}{n}(\mathbf{y}_+ - \mathbf{f}_+)^T \Sigma^{-1}_{n,\tau_1,\sigma,\delta,\tau_0,+}(\mathbf{y}_+ - \mathbf{f}_+) + \lambda_+ J(f_+), & i > \tau_0. \end{cases} \tag{1}$$

Here $\mathbf{y}_- = (y_1, \ldots, y_{\tau_0})^T$, $\mathbf{y}_+ = (y_{\tau_0+1}, \ldots, y_n)^T$, $\mathbf{f}_- = \left(f_-(1/n), \ldots, f_-(\tau_0/n)\right)^T$, and $\mathbf{f}_+ = \left(f_+((\tau_0+1)/n), \ldots, f_+(1)\right)^T$. The matrices $\Sigma_{n,\tau_1,\sigma,\delta,\tau_0,-}$ and $\Sigma_{n,\tau_1,\sigma,\delta,\tau_0,+}$ are the corresponding top-left and bottom-right submatrices of $\Sigma_{n,\tau_1,\sigma,\delta}$, where $\Sigma_{n,\tau_1,\sigma,\delta}$ has diagonal entries σ^2 before τ_1 and δ^2 after τ_1.

3 METHOD

The mean and variance change-points are detected using likelihood-based test statistics, following the ideas of Chen & Gupta (2000) and Gao et al. (2019). Given the variance parameters, the smooth mean functions are estimated by penalized weighted least squares, leading to smoothing-spline estimators; given the estimated mean functions, the variance change-point and variance levels are updated from the residual sequence. The criterion in equation 1 is used conditionally on the variance parameters rather than jointly minimized over all unknown quantities. Indeed, without the log-determinant term in the Gaussian likelihood, direct minimization over the variance parameters is degenerate, since the weighted residual term can approach zero as the variances diverge. We therefore adopt a block-wise iterative procedure that alternates between mean change-point detection and variance change-point detection, yielding estimates $(\widehat{f}_-, \widehat{f}_+, \widehat{\tau}_0, \widehat{\kappa}, \widehat{\tau}_1, \widehat{\sigma}^2, \widehat{\delta}^2)$. The procedure is summarized in Algorithm 1. The algorithm converges when the fitted residuals become stable, namely when $\max_{1\le i\le n} |\widehat{e}_i^{(\iota)} - \widehat{e}_i^{(\iota-1)}| < \epsilon$, where $\epsilon = 0.001$ in our numerical experiments.

Algorithm 1: Mean and variance shift detection

Input: Data $\{y_1, \ldots, y_n\}$

Output: $\widehat{f}_-, \widehat{f}_+, \widehat{\tau}_0, \widehat{\kappa}, \widehat{\tau}_1, \widehat{\sigma}^2, \widehat{\delta}^2$

Step 1: Initialize under constant variance

Assume $\sigma^2 = \delta^2$ and obtain the initial estimates $\widehat{f}_-^{(0)}, \widehat{f}_+^{(0)}, \widehat{\tau}_0^{(0)}, \widehat{\kappa}^{(0)}, \widehat{\mathbf{e}}^{(0)}$ via Algorithm 2;

Step 2: Iterative change detection at the ι-th iteration

2.1 Variance change detection:

Given $\widehat{f}_-^{(\iota-1)}, \widehat{f}_+^{(\iota-1)}, \widehat{\tau}_0^{(\iota-1)}$, estimate $\widehat{\tau}_1^{(\iota)}, \widehat{\sigma}^{2(\iota)}, \widehat{\delta}^{2(\iota)}$ by Algorithm 3;

2.2 Mean change detection:

Given $\widehat{\tau}_1^{(\iota)}, \widehat{\sigma}^{2(\iota)}, \widehat{\delta}^{2(\iota)}$, estimate $\widehat{f}_-^{(\iota)}, \widehat{f}_+^{(\iota)}, \widehat{\tau}_0^{(\iota)}, \widehat{\kappa}^{(\iota)}$ by Algorithm 2;

Step 3: Repeat Step 2 until convergence

Algorithm 1 alternates between mean change-point detection and variance change-point detection. The initialization is performed under a constant-variance working model, in which all observations receive equal weights. In our numerical experiments, the algorithm usually converges within a few iterations. The smoothing parameter λ is selected by minimizing the generalized cross-validation (GCV) score; see Gu (2013, Sec. 3.2).

3.1 MEAN FUNCTION ESTIMATION

Given τ_1, σ^2, and δ^2, the mean functions are estimated via penalized weighted least squares in a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$. A detailed theoretical background on the RKHS formulation, reproducing kernels, and spline examples is provided in Appendix A.

A key consequence of the RKHS formulation is that the infinite-dimensional optimization problem admits a finite-dimensional representation. Specifically, by the representer theorem, each estimated smooth mean function can be written as

$$\widehat{f}(t) = \sum_{\nu=1}^{m} d_\nu \phi_\nu(t) + \sum_{i=1}^{n} c_i R(t_i, t),$$

where $\{\phi_\nu\}_{\nu=1}^m$ spans the null space of the semi-norm $J(\cdot)$, $R(\cdot,\cdot)$ is the reproducing kernel associated with the penalized component of $\mathcal{H}$, and $\{d_\nu\}_{\nu=1}^m$ and $\{c_i\}_{i=1}^n$ are finite-dimensional coefficients to be estimated. Thus, although $\mathcal{H}$ is infinite-dimensional, the conditional minimizer of equation 1 can be obtained by solving a finite-dimensional penalized weighted least squares problem. This observation follows the classical spline theory of Wahba (1990).

3.2 MEAN CHANGE-POINT DETECTION GIVEN τ_1, σ^2, δ^2

Given the current estimates of the variance change-point and variances $(\widehat{\tau}_1, \widehat{\sigma}^2, \widehat{\delta}^2)$, we detect the mean change-point by traversing candidate locations. For each candidate k_0, we estimate the smooth

mean functions on the two sides of k_0, compute the corresponding residuals and the mean shift, and then test the significance of the calibrated mean shift.

Algorithm 2: Mean Change-Point Detection

Input: Observed data $\{y_1, \ldots, y_n\}$, variance parameters $(\tau_1, \sigma^2, \delta^2)$
Output: $\widehat{\tau}_0$, $\widehat{f}_-^{(\widehat{\tau}_0)}$, $\widehat{f}_+^{(\widehat{\tau}_0)}$, $\widehat{\kappa}^{(\widehat{\tau}_0)}$.
For each $k_0 \in [1, n]$, perform the following steps:
Step 1: Compute the mean function estimates $\widehat{f}_-^{(k_0)}$ and $\widehat{f}_+^{(k_0)}$, the residuals $\widehat{\mathbf{e}}^{(k_0)}$, and the mean shift quantity $\widehat{\kappa}^{(k_0)}$ by equation 2;
Step 2: Obtain the rescaled data $\{x_i^{(k_0)}\}$ by equation 3;
Step 3: Estimate the mean change-point $\widehat{\tau}_0$ through the hypothesis test procedure introduced in Section 3.2 on the rescaled data $\{x_i^{(k_0)}\}$.

The algorithm for mean function estimation and mean change-point detection is shown in Algorithm 2, which consists of three main steps. The first step is to estimate the mean function. To estimate the mean function, it is necessary to traverse every point in the data, calculate the residual sequence, and perform detection of mean change-points. Note that the mean of the residual sequence on both sides of each point is close to 0, making direct detection of mean change-points infeasible. For each potential mean change-point k_0, denote

$$\mathbf{y}_-^{(k_0)} = \{y_1, \ldots, y_{k_0}\} \quad \text{and} \quad \mathbf{y}_+^{(k_0)} = \{y_{k_0+1}, \ldots, y_n\}.$$

We then estimate the mean functions before and after the change-point, denoted as $\widehat{f}_-^{(k_0)}$ and $\widehat{f}_+^{(k_0)}$ respectively, and subsequently compute the residuals and mean shift quantities:

$$\begin{aligned} \widehat{f}_\pm^{(k_0)} &= \arg\min_f \left\{ \frac{1}{n} (\mathbf{y}_\pm^{(k_0)} - \mathbf{f})^\top \Sigma_{n,\tau_1,\sigma,\delta,k_0,\pm}^{-1} (\mathbf{y}_\pm^{(k_0)} - \mathbf{f}) + \lambda J(f) \right\}, \\ \widehat{e}_i^{(k_0)} &= y_i - \widehat{f}_\pm^{(k_0)}(i/n), \qquad \widehat{\kappa}^{(k_0)} = \widehat{f}_+^{(k_0)} \left(\frac{k_0+1}{n} \right) - \widehat{f}_-^{(k_0)} \left(\frac{k_0}{n} \right). \end{aligned} \tag{2}$$

where $\widehat{f}_\pm^{(k_0)}$ corresponds to $\widehat{f}_-^{(k_0)}$ and $\widehat{f}_+^{(k_0)}$ in the regions $i \le k_0$ and $i > k_0$, respectively. To test the significance of the estimated mean shift quantity $\widehat{\kappa}^{(k_0)}$, we eliminate the influence of the variance change from the residual sequence in the following way:

$$x_i^{(k_0)} = \tilde{e}_i^{(k_0)} + \widehat{\kappa}^{(k_0)} \mathbb{I}(i > k_0), \quad \text{where} \quad \tilde{e}_i^{(k_0)} = \begin{cases} \widehat{e}_i^{(k_0)}/\sigma, & i \le \widehat{\tau}_1 \\ \widehat{e}_i^{(k_0)}/\delta, & i > \widehat{\tau}_1 \end{cases}. \tag{3}$$

The normalization in equation 3 serves to remove the confounding effect of variance changes on the mean-change test. The third step is to perform mean change-point detection via hypothesis testing on the calibrated residual sequence $\{x_i^{(k_0)}\}$. Given known variances, we assess the significance of change-points using a simplified procedure adapted from Chen & Gupta (2000). To align with our iterative algorithm, we test only the current candidate position $t = k \in [2, n-2]$ rather than scanning the entire sequence. We want to test the hypothesis

$$\begin{aligned} &H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu \quad \text{versus} \\ &H_1 : \mu_1 = \cdots = \mu_{\tau_0} \neq \mu_{\tau_0+1} = \cdots = \mu_n, \end{aligned}$$

where τ_0 is a potential mean change-point. Define $RSS = \sum_{i=1}^n (x_i - \bar{x})^2$ and $RSS_{k_0} = \sum_{i=1}^{k_0} (x_i - \bar{x}_{k_0})^2 + \sum_{i=k_0+1}^n (x_i - \bar{x}_{n-k_0})^2$, where $\bar{x}$, $\bar{x}_{k_0}$, and $\bar{x}_{n-k_0}$ are the respective sample means. Then the hypothesis test statistic is

$$T_n^2 = \max_{2 \le k_0 \le n-2} (RSS - RSS_{k_0}).$$

Large values of T_n^2 lead us to reject the null hypothesis H_0. To make the decision statistically rigorous, we introduce a level-α critical value c_α. Detailed derivations can be found in Appendix B. We reject H_0 when $T_n^2 \ge c_\alpha$. The estimated mean change-point is the maximizer of the reduction in residual sum of squares,

$$\widehat{\tau}_0 = \operatorname*{argmax}_{2 \le k_0 \le n-2} (RSS - RSS_{k_0}).$$

3.3 VARIANCE CHANGE-POINT DETECTION GIVEN $\tau_0, f_-{}^{(\tau_0)}, f_+{}^{(\tau_0)}$

Given the estimates of the mean function and the mean change-point $\widehat{\tau}_0, \widehat{f}_-^{(\widehat{\tau}_0)}, \widehat{f}_+^{(\widehat{\tau}_0)}$, we propose the following test procedure for the variance change-point. We want to test the following hypothesis:

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 \quad \text{versus} \\ H_1 &: \sigma_1^2 = \cdots = \sigma_{\tau_1}^2 \neq \sigma_{\tau_1+1}^2 = \cdots = \sigma_n^2, \end{aligned} \tag{4}$$

for a potential variance change-point position τ_1.

Algorithm 3: Variance Change-Point Detection

Input: $\{y_1, \cdots, y_n\}, \widehat{\tau}_0, \widehat{f}_-^{(\widehat{\tau}_0)}, \widehat{f}_+^{(\widehat{\tau}_0)}$
Output: $\widehat{\tau}_1$, $\widehat{\sigma}^2$, $\widehat{\delta}^2$
Step 1: Compute the residuals $\widehat{\mathbf{e}}$ by equation 5;
Step 2: Detect the variance change-point $\widehat{\tau}_1$ through the hypothesis test procedure introduced in Section 3.3 on $\widehat{\mathbf{e}}$;
Step 3: Estimate variances $\widehat{\sigma}^2 = \widehat{\sigma}^2_{(\widehat{\tau}_1)}$ and $\widehat{\delta}^2 = \widehat{\delta}^2_{(\widehat{\tau}_1)}$.

First, we obtain the residual vector $\widehat{\mathbf{e}}$ based on the known $\widehat{\tau}_0$ and $\widehat{f}_-^{(\widehat{\tau}_0)}$. $\widehat{\mathbf{e}}$ is a zero-mean sequence with variance change-points. The i-th component of $\widehat{\mathbf{e}}$, denoted by $\widehat{e}_i$, is computed as follows:

$$\widehat{e}_i = \begin{cases} y_i - \widehat{f}_-^{(\widehat{\tau}_0)}(i/n) & 1 \leq i \leq \widehat{\tau}_0, \\ y_i - \widehat{f}_+^{(\widehat{\tau}_0)}(i/n) & \widehat{\tau}_0 < i \leq n, \end{cases} \tag{5}$$

Proceeding from this, the variance change-point is identified from $\widehat{\mathbf{e}}$. In the second step, we use the likelihood ratio to conduct a hypothesis test. The scaled negative log-likelihood under H_1 is $\ell(k_1) = k_1 \log \widehat{\sigma}^2_{(k_1)} + (n-k_1)\log \widehat{\delta}^2_{(k_1)}$, and under H_0 it is $\ell(n) = n \log \widehat{\sigma}^2_{(n)}$, where $\widehat{\sigma}^2_{(k_1)} = \frac{1}{k_1}\sum_{i=1}^{k_1} \widehat{e}_i^2$, and $\widehat{\delta}^2_{(k_1)} = \frac{1}{n-k_1}\sum_{i=k_1+1}^{n} \widehat{e}_i^2$. We define the test statistic for the variance change-point as

$$\lambda_n^2 = \max_{2 \leq k_1 \leq n-2} [\ell(n) - \ell(k_1)].$$

By the principle of the likelihood ratio test, we reject H_0 if $\ell(n) > \ell(k_1)$ for some k_1. To make the decision statistically rigorous, we use a critical value c_α. The detailed derivations are provided in Appendix B. We reject H_0 when $\lambda_n^2 > \log n + c_\alpha$. The estimated variance change-point is

$$\widehat{\tau}_1 = \operatorname*{argmin}_{2 \leq k_1 \leq n-2} \ell(k_1).$$

4 THEORETICAL RESULTS

In this paper, we establish three main theoretical results: the consistency of the parameter estimators, the asymptotic null distribution of the mean change-point test statistic, and the asymptotic null distribution of the variance change-point test statistic. Before presenting the theorems, we impose several regularity assumptions on the change-points, mean functions, and error distributions. For brevity, the detailed statements of these assumptions (Assumptions C.1 - C.5) and their corresponding remarks are provided in Appendix C.1. The proposed theory relies on independence, Gaussian errors, and mean functions smoothness assumptions. Although our theoretical framework is developed under Gaussian errors for simplicity, the proposed method remains robust under heavy-tailed noise and non-Gaussian distributions (Appendices D.4).

Theorem 4.1 (Consistency of Parameter Estimates)**.** *Under Assumptions C.1 - C.4, the estimates* $\{\widehat{f}_-, \widehat{f}_+, \widehat{\tau}_0, \widehat{\kappa}, \widehat{\tau}_1, \widehat{\delta}^2, \widehat{\sigma}^2\}$ *in Algorithm 1 are consistent and satisfy the following convergence rates:*

$$\|\widehat{f}_- - f_-\|_n^2 = O_P\left(\lambda + (nh)^{-1} + h^{-1}\widetilde{r}_n^2\right), \quad \|\widehat{f}_+ - f_+\|_n^2 = O_P\left(\lambda + (nh)^{-1} + h^{-1}\widetilde{r}_n^2\right),$$
$$|\widehat{\tau}_0 - \tau_0| = O_P((\log n)^4(\log\log n)^2), \quad |\widehat{\tau}_1 - \tau_1| = O_P((\log n)^4(\log\log n)^2),$$
$$|\widehat{\kappa} - \kappa| = O_P\big(r_n + (\log n)^4(\log\log n)^2/n\big), \quad |\widehat{\delta}^2 - \delta_0^2| = O_P(\widetilde{r}_n),$$

where $h = \lambda^{1/(2m)}$, $r_n = \sqrt{\log n/(nh)} + h^{m-1/2}$, $\widetilde{r}_n = r_n^2 + (nh)^{-3/4} + (\log n)^4(\log\log n)^2/n + n^{-1/2}$, *and* $\|f\|_n = \sqrt{\sum_{i=1}^n f(i/n)^2/n}$ *is an empirical norm of function* f.

Corollary 4.1. *When the regularization parameter satisfies $\lambda \asymp n^{-2m/(2m+1)}$, we have:*

$$\|\widehat{f}_- - f_-\|_n^2 = O_P\left(n^{-\frac{2m}{2m+1}}\right), \quad \|\widehat{f}_+ - f_+\|_n^2 = O_P\left(n^{-\frac{2m}{2m+1}}\right),$$
$$|\widehat{\kappa} - \kappa| = O_P\left(n^{-\frac{2m-1}{4m+2}}\right), \quad |\widehat{\sigma}^2 - \sigma_0^2| = O_P(n^{-1/2}), \quad |\widehat{\delta}^2 - \delta_0^2| = O_P(n^{-1/2}).$$

Corollary 4.1 shows that both $\widehat{f}_-$ and $\widehat{f}_+$ attain the stated rate for spline function estimation. Under the fixed-order Sobolev class and squared empirical L_2 loss, this rate is the corresponding minimax benchmark; the statement does not claim universal optimality over other function classes or losses. In addition, this indicates that $\widehat{\sigma}^2$ and $\widehat{\delta}^2$ are $\sqrt{n}$-consistent. Notice that when $m = 1$, the results of $|\widehat{\sigma}^2 - \sigma_0^2| = O_P(n^{-1/2})$ and $|\widehat{\delta}^2 - \delta_0^2| = O_P(n^{-1/2})$ are not straightforward in Corollary 4.1. The detailed proof is provided in the supplement. Next, we introduce the asymptotic distributions of the mean and variance change-point test statistics under H_0.

Theorem 4.2 (Asymptotic Null Distribution of the Mean Change-Point Test Statistic)**.** *Suppose $n \to \infty$, $h = \lambda^{1/(2m)} \to 0$, $r_n^2 \log n \to 0$, where $r_n = \sqrt{\log n/(nh)} + h^{m-1/2}$. Under H_0 and Assumptions C.1 - C.3, for all real numbers x, T_n in Algorithm 2 has the following asymptotic property:*

$$P\left\{a_n^{-1}\left(T_n - b_n\right) \le x\right\} \to \exp\left(-2\pi^{1/2}e^{-x}\right),$$

where $a_n = (2\log\log n)^{-1/2}$, $b_n = a_n^{-1} + \frac{1}{2}a_n \log\log\log n$.

Let $\lambda_n^2 = \max_{2\le k\le n-2}[\ell(n) - \ell(k)]$. The asymptotic distribution of λ_n under the null hypothesis H_0 is described in Theorem 4.3.

Theorem 4.3 (Asymptotic Null Distribution of the Variance Change-Point Test Statistic)**.** *Suppose $n \to \infty$, $h = \lambda^{1/(2m)} \to 0$, $r_n^2 \log n \to 0$, where $r_n = \sqrt{\log n/(nh)} + h^{m-1/2}$. Under H_0 and Assumptions C.1 - C.3 and C.5, for all real numbers x,*

$$P\left\{a_n(\log n)^{-1/2}\lambda_n - b_n \log n \le x\right\} \to \exp\left(-2e^{-x}\right),$$

where $a_n = (2\log\log n)^{1/2}/\log n$, $b_n = \{2\log\log n + \frac{1}{2}\log\log\log n - \log\Gamma\left(\frac{1}{2}\right)\}/\log n$.

Theorems 4.2 and 4.3 state that the normalized test statistics T_n and λ_n converge to Gumbel-type extreme-value distributions under the null hypothesis.

5 Simulation and Application

5.1 Simulations

We compare the proposed iterative algorithm with two existing methods: the parametric CG method of Chen & Gupta (1997) and the nonparametric MJ method of Matteson & James (2014). Their moving-window variants based on Niu & Zhang (2012) are also included, with window sizes {10, 20, 50, 100}, denoted by MW10, ..., MW100. We consider two sample sizes, $n \in \{150, 450\}$, with the proposed method denoted by New and New450, respectively. For each setting, 1,000 replicates are generated, and all change-point estimates are rescaled by n for easier comparison. Four change-point structures are considered: simultaneous mean and variance changes, co-located joint mean-variance change, mean change only, and variance change only. Estimation performance is evaluated for $(f_-, f_+, \tau_0, \kappa, \tau_1, \sigma^2, \delta^2)$, with empirical power reported in Appendices D.2 and D.3. Mean function estimation is measured by mean squared error (MSE), and variance estimation is reported for the proposed method using the log ratio of the estimated to true variance. The simulation experiments in this section are based on Gaussian errors. In addition, we conduct simulation experiments based on the Student's t distribution in Appendices D.4 to verify the robustness of the proposed method under heavy-tailed noise and non-Gaussian distributions.

The two baseline trend functions used in the simulations are

$$f_1(x) = 20 + 12x(1-x), \qquad x \in [0.15, 0.8],$$
$$f_2(x) = \sin(x) + x^5 - 8x^3 + 10x + 6, \quad x \in [-1.5, 2.5].$$

Here $f_1(x) = 20 + 2b_{2,2}(x)$, where $b_{2,2}(x) = 6x(1-x)$ is the Beta$(2,2)$ density. For a sample of size n, the baseline mean at observation i is evaluated at $x_{1,i} = 0.15 + 0.65(i-1)/(n-1)$ for f_1 and $x_{2,i} = -1.5 + 4(i-1)/(n-1)$ for f_2, for $i = 1, \ldots, n$.

Due to space limitations, we report only the simultaneous-change case in the main text and relegate the remaining cases to Appendix D. We consider the following four settings:
Scenario 1: $n = 150,\ f = f_1,\ \tau_0 = 50,\ \kappa = 0.75,\ \tau_1 = 100,\ \sigma^2 = 0.24,\ \delta^2 = 0.06$;
Scenario 2: $n = 450,\ f = f_1,\ \tau_0 = 150,\ \kappa = 0.75,\ \tau_1 = 300,\ \sigma^2 = 0.24,\ \delta^2 = 0.06$;
Scenario 3: $n = 150,\ f = f_2,\ \tau_0 = 50,\ \kappa = 3,\ \tau_1 = 100,\ \sigma^2 = 8,\ \delta^2 = 2$;
Scenario 4: $n = 450,\ f = f_2,\ \tau_0 = 150,\ \kappa = 3,\ \tau_1 = 300,\ \sigma^2 = 8,\ \delta^2 = 2$.
Figure 1 illustrates four representative realizations generated under the above settings.

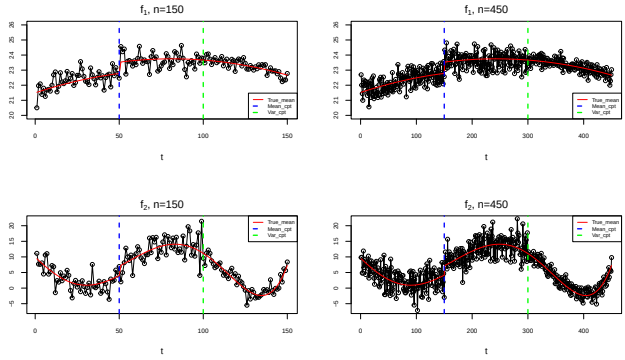

Figure 1: Simulated data profiles under Scenarios 1–4. Red curves show the true mean functions; blue and green dashed lines mark the true mean change-point ($\tau_0/n = 1/3$) and variance change-point ($\tau_1/n = 2/3$), respectively.

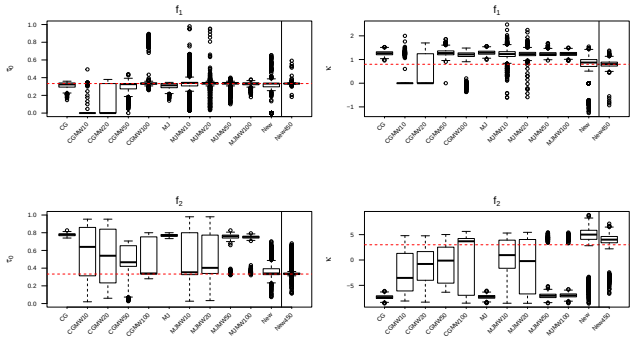

Figure 2: Boxplots of mean change-point estimates and mean-shift estimates under Scenarios 1–4. *Left column:* mean change-point estimates; *right column:* mean-shift estimates. Red dashed lines indicate the true values, with $\tau_0/n = 1/3$ for the mean change-point and the corresponding true κ for the mean shift. *Top row:* $f = f_1$; *bottom row:* $f = f_2$.

Figure 2 summarizes the mean change-point and mean-shift estimation results under simultaneous mean and variance changes. The proposed method(New) gives the most accurate and stable estimates across Scenarios 1–4. Under the simpler trend f_1, CG and MJ are reasonably accurate, and their moving-window variants improve as the window size increases. Under the more complex trend f_2, CG, MJ, and their moving-window variants show clear biases. The estimation of κ also improves as n increases from 150 to 450.

Figure 3 reports variance change-point estimates for all methods and variance-level estimates for the proposed method under simultaneous mean and variance changes. New accurately estimates both the variance change-point and variance levels under f_1 and f_2, while CG and MJ show biases in change-point localization due to the unknown smooth mean and mean shift. For New, the pre-/post-change variance log-ratios center around zero and become less variable at $n = 450$.

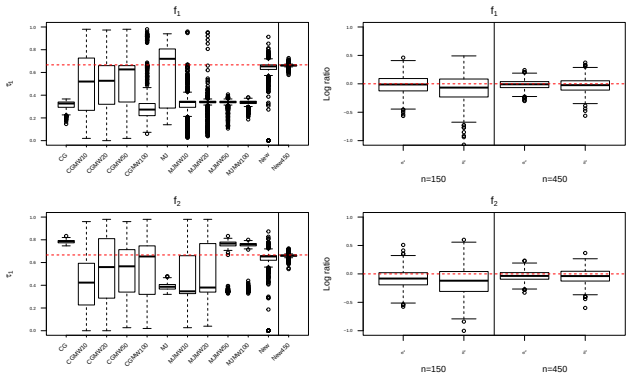

Figure 3: Boxplots of variance change-point estimates for all methods and variance-level estimates for the proposed method under Scenarios 1–4. *Left column:* variance change-point estimates; *right column:* variance-level estimates. Red dashed lines mark the true variance change-point, $\tau_1/n = 2/3$. For the proposed method, variance estimates are reported as log ratios of the estimated to true values for σ^2 and δ^2. *Top row:* $f = f_1$; *bottom row:* $f = f_2$.

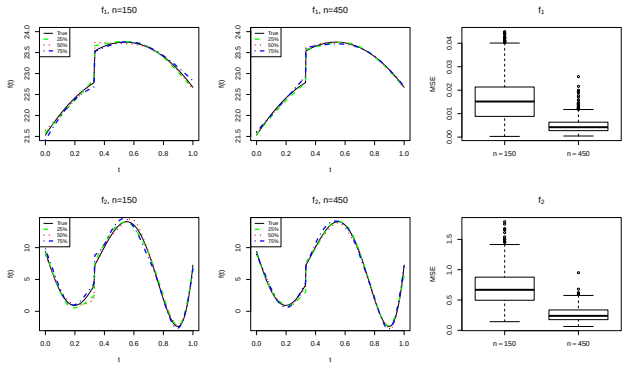

Figure 4: Mean function estimation performance of the proposed method. *Left and middle columns:* true mean functions and estimated curves corresponding to the 25th, 50th, and 75th percentiles of the MSE over 1,000 replications for $n = 150$ and $n = 450$, respectively. *Right column:* boxplots of the MSE distributions for the two sample sizes. *Top row:* $f = f_1$; *bottom row:* $f = f_2$.

Figure 4 evaluates the mean-function estimation performance of the proposed method when both mean and variance change-points coexist. Across all four scenarios, the estimated curves track the true mean functions closely. At $n = 150$, the curves corresponding to larger MSE values show slight deviations near the change-points, reflecting the expected variability for smaller samples. In contrast, the estimates for $n = 450$ conform more closely to the true trends. The MSE distributions also shrink markedly as n increases, indicating improved mean-function estimation accuracy.

5.2 Application

In this section, we apply the proposed method to evaluate the quality control process in electric vehicle battery manufacturing. For confidentiality reasons, the data have been appropriately desensitized. To identify potential defects and maintain stable equipment performance, more than one hundred critical dimensions are monitored in real time. Significant changes in the data of each dimension reflect structural shifts in the manufacturing process. Both the number and nature of such changes vary across different dimensions. For each dimension, a time-ordered data profile comprising 691 consecutive battery measurements is collected. In our study, the first $n = 200$ observations from each dimension are used for change detection. We categorize the detection results into the following four representative scenarios: (1) simultaneous changes in mean and variance at the same location;

(2) changes in mean and variance at different locations; (3) change in mean only; and (4) change in variance only.

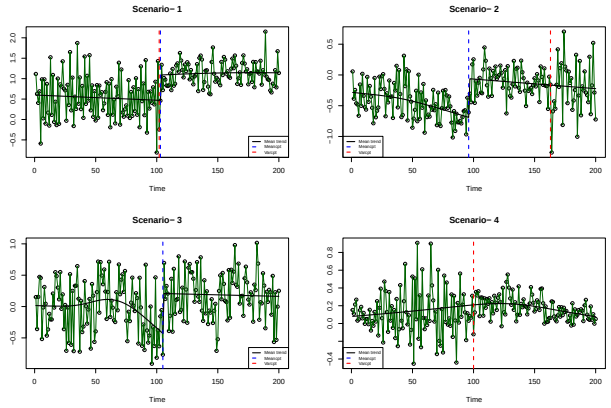

Figure 5: Change-point detection results of Scenarios (1)-(4). Each panel displays a randomly selected data profile from the corresponding scenario. Solid black curves are the estimated mean functions. Vertical blue and red dashed lines are the estimated mean and variance change-points, respectively.

Four representative data profiles are randomly selected for each scenario, and the corresponding results are presented in Figure 5. For Scenario (i), there is a co-occurring mean and variance change with estimates $\widehat{\tau}_0 = 103$ and $\widehat{\tau}_1 = 102$. Scenario-ii exhibits changes in both mean and variance at distinct locations with $\widehat{\tau}_0 = 96$ and $\widehat{\tau}_1 = 163$. Scenario-iii involves a shift exclusively in the mean, identified by $\widehat{\tau}_0 = 105$. Scenario-iv presents a change only in the variance, indicated by $\widehat{\tau}_1 = 100$. Across all scenarios, the estimated change-points accurately capture the structural changes visually apparent in the data, while the fitted mean functions closely follow the underlying trends.

All experiments were run on a standard CPU workstation: Intel(R) Core(TM) i5-8500 CPU at 3.00GHz and 8 GB RAM. No large-scale cluster or cloud computing was used.

6 Conclusion and discussion

In practice, complex non-stationary data often conceal various parametric changes that are difficult to detect jointly. To address this challenge, this paper develops an iterative algorithm for detecting multiple parametric changes within volatile data. Unlike traditional methods that assume a piecewise-constant mean, our approach accommodates a smoothly evolving mean trend alongside abrupt shifts in either the mean or variance. Operating within a penalized least squares framework, the algorithm iteratively estimates the mean functions, the location and magnitude of mean shifts, the variance change-point, and the pre- and post-change variances. Simulation comparisons with state-of-the-art methods and their moving-window variants under Gaussian and heavy-tailed error distributions show that our approach identifies change-points more accurately across diverse scenarios, demonstrating its strong practical value.

The proposed method has two limitations. First, being based on hypothesis testing, it inherently risks Type I errors (false alarms) and Type II errors (missed detections). Second, small mean shifts may be partially absorbed by the smooth mean trend and obscured by data variability, particularly around regions with a steep local slope. Using a less conservative significance level (e.g., $\alpha = 0.10$ or 0.15) can improve sensitivity but may increase false positives, motivating future work on more adaptive testing strategies.

Disclosure statement

The authors declare no competing interests.

AI USE STATEMENT

The authors used generative AI tools solely for grammar checking. No other use of generative AI was involved in this work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

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. ISSN 0047-259X.

Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78, 1998. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/2998540.

Rafal 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.

István Berkes, Robertas Gabrys, Lajos Horváth, and Piotr Kokoszka. Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(5):927–946, 11 2009.

Lorenzo Cappello and Oscar Hernan Madrid Padilla. Bayesian variance change point detection with credible sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Hao Chen and Nancy Zhang. Graph-based change-point detection. *The Annals of Statistics*, 43(1):139 – 176, 2015. doi: 10.1214/14-AOS1269. URL https://doi.org/10.1214/14-AOS1269.

Jie Chen and Arjun K Gupta. Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association*, 92(438):739–747, 1997.

Jie Chen and Arjun K. Gupta. *Parametric statistical change point analysis*, volume 192. Springer, 2000.

Herman Chernoff and Shelemyahu Zacks. Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics*, 35(3):999–1018, 1964.

Haeran Cho and Piotr Fryzlewicz. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 77(2):475–507, 03 2015.

Haeran Cho, Tobias Kley, and Housen Li. Detection and inference of changes in high-dimensional linear regression with nonsparse structures. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(5):1528–1552, 2025.

Paulus Petrus Bernardus Eggermont and Vincent N LaRiccia. *Maximum penalized likelihood estimation*, volume 1. Springer, 2001.

Klaus Frick, Axel Munk, and Hannes Sieling. Multiscale change point inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(3):495–580, 2014.

Piotr Fryzlewicz. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243 – 2281, 2014. doi: 10.1214/14-AOS1245. URL https://doi.org/10.1214/14-AOS1245.

Piotr Fryzlewicz. Narrowest significance pursuit: inference for multiple change-points in linear models. *Journal of the American Statistical Association*, 119(546):1633–1646, 2024.

Zhenguo Gao, Zuofeng Shang, Pang Du, and John L. Robertson. Variance change point detection under a smoothly-changing mean trend with application to liver procurement. *Journal of the American Statistical Association*, 114(526):773–781, 2019. doi: 10.1080/01621459.2018.1442341.

Zhenguo Gao, Pang Du, Ran Jin, and John L. Robertson. Surface temperature monitoring in liver procurement via functional variance change-point analysis. *The Annals of Applied Statistics*, 14(1):143 – 159, 2020. doi: 10.1214/19-AOAS1297. URL https://doi.org/10.1214/19-AOAS1297.

Zhenguo Gao, Yan Li, Longqin Lai, and Rui Miao. An effective variance change detection method for data with constantly changing mean function. *Stat*, 15(1):e70114, 2026. doi: https://doi.org/10.1002/sta4.70114. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sta4.70114. e70114 STAT-24-0263.R6.

K. Ghoudi and N. Laïb. On joint testing of changes in conditional mean and variance functions of stationary and ergodic time series. *Statistical Papers (Berlin, Germany)*, 66(5):115, 2025.

Gérard Grégoire and Zouhir Hamrouni. Change point estimation by local linear smoothing. *Journal of Multivariate Analysis*, 83(1):56–83, 2002.

Chong Gu. *Smoothing Spline ANOVA Models*. Springer, New York, 2 edition, 2013. doi: 10.1007/978-1-4614-5369-7.

D. L. Hawkins. Detecting shifts in functions of multivariate location and covariance parameters. *Journal of Statistical Planning and Inference*, 33(2):233–244, 1992.

Kaylea Haynes, Paul Fearnhead, and Idris A Eckley. A computationally efficient nonparametric approach for changepoint detection. *Statistics and Computing*, 27:1293–1305, 2017.

David V Hinkley. Inference about the change-point in a sequence of random variables. *Biometrika*, 57(1):1–17, 1970.

Lajos Horvath. The maximum likelihood method for testing changes in the parameters of normal observations. *The Annals of Statistics*, 21(2):671–680, 1993. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/2242253.

Lajos Horváth, Gregory Rice, and Yuqian Zhao. Change point analysis of covariance functions: A weighted cumulative sum approach. *Journal of Multivariate Analysis*, 189:104877, 2022. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2021.104877. URL https://www.sciencedirect.com/science/article/pii/S0047259X2100155X.

Der-Ann Hsu. Tests for variance shift at an unknown time point. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 26(3):279–284, 1977.

Carla Inclan and George C Tiao. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 89(427):913–923, 1994.

Karolos K. Korkas and Piotr PryzlewiczV. Multiple change-point detection for non-stationary time series using wild binary segmentation. *Statistica Sinica*, 27(1):287–311, 2017. ISSN 10170405, 19968507.

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.

Marc Lavielle and Eric Moulines. Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, 21(1):33–59, 2000.

Sokbae Lee, Myung Hwan Seo, and Youngki Shin. The lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(1):193–210, 01 2016.

Fuxiao Li, Zheng Tian, Yanting Xiao, and Zhanshou Chen. Variance change-point detection in panel data models. *Economics Letters*, 126:140–143, 2015.

Yang Li, Qijing Yan, Mixia Wu, and Aiyi Liu. Bootstrap-based inference for multiple variance changepoint models. *Journal of Applied Statistics*, 52(14):2636–2671, 2025.

Clive R Loader. Change point estimation using nonparametric regression. *The Annals of Statistics*, 24(4):1667–1678, 1996.

Carlos Misael Madrid Padilla, Daren Wang, Zifeng Zhao, and Yi Yu. Change-point detection for sparse and dense functional data in general dimensions. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 37121–37133. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/f0add74c2f1ac58197173a38c01b2210-Paper-Conference.pdf.

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 A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 21081–21134. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/42a0de6b8a1809ceba8fdad1661be06c-Paper-Conference.pdf.

David S Matteson and Nicholas A James. A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505):334–345, 2014.

Euan T McGonigle and Haeran Cho. Nonparametric data segmentation in multivariate time series via joint characteristic functions. *Biometrika*, 112(2):asaf024, 2025.

Michael Messer. Bivariate change point detection: joint detection of changes in expectation and variance. *Scandinavian Journal of Statistics*, 49(2):886–916, 2022.

Yue S Niu and Heping Zhang. The screening and ranking algorithm to detect dna copy number variations. *The Annals of Applied Statistics*, 6(3):1306, 2012.

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, 2021.

Jianmin Pan and Jiahua Chen. Application of modified information criterion to multiple change point problems. *Journal of Multivariate Analysis*, 97(10):2221–2241, 2006.

Alessandro Rinaldo, Daren Wang, Qin Wen, Rebecca Willett, and Yi Yu. Localizing changes in high-dimensional regression models. In Arindam Banerjee and Kenji Fukumizu (eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 2089–2097. PMLR, 13–15 Apr 2021.

Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

Grace Wahba. *Spline models for observational data*. SIAM, 1990.

Grace Wahba and Yuedong Wang. Representer theorem. *Wiley StatsRef: Statistics Reference Online*, pp. 1–11, 2019.

Daren Wang, Yi Yu, and Alessandro Rinaldo. Univariate mean change point detection: Penalization, CUSUM and optimality. *Electronic Journal of Statistics*, 14(1):1917 – 1961, 2020. doi: 10.1214/20-EJS1710. URL https://doi.org/10.1214/20-EJS1710.

Keith J Worsley. On the likelihood ratio test for a shift in location of normal populations. *Journal of the American Statistical Association*, 74(366a):365–367, 1979.

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.

Minya Xu, Ping-Shou Zhong, and Wei Wang. Detecting variance change-points for blocked time series and dependent panel data. *Journal of Business & Economic Statistics*, 34(2):213–226, 2016.

Nancy R Zhang and David O Siegmund. A modified bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, 63(1):22–32, 2007.

A THEORETICAL BACKGROUND OF MEAN FUNCTION ESTIMATION

At each $t \in \mathcal{T}$, the evaluation functional $[t] : \mathcal{H} \to \mathbb{R}, f \mapsto f(t)$ is continuous on the Hilbert space $\mathcal{H}$. In that case, $\mathcal{H}$ is referred to as a Reproducing Kernel Hilbert Space (RKHS). By the Riesz representation theorem, for all $t \in \mathcal{T}$, there exists a unique $R_t \in \mathcal{H}$ satisfying the reproducing property, i.e., $\langle R_t, f\rangle = [t](f) = f(t)$, where $\langle \cdot, \cdot \rangle$ is the inner product on $\mathcal{H}$. A reproducing kernel $R : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ on $\mathcal{H}$ satisfies $R(s,t) = \langle R_s, R_t\rangle$. It can be proved that there is a one-to-one correspondence between the reproducing kernels and RKHSs. The penalty term J in equation 1 is a semi-norm on $\mathcal{H}$. From the null space of J, denoted as $\mathcal{N}_J = \{f : J(f) = 0\}$, we can obtain a direct sum decomposition of $\mathcal{H}$ as $\mathcal{H} = \mathcal{N}_J \oplus \mathcal{H}_J$, where $\mathcal{H}_J$ is the complement space of $\mathcal{N}_J$ in $\mathcal{H}$. This leads to a decomposition of the reproducing kernel as $R = R_0 + R_J$, where R_0 and R_J are the reproducing kernels on the subspaces $\mathcal{N}_J$ and $\mathcal{H}_J$ respectively.

Example. Here we illustrate these concepts using an example of a cubic smoothing spline. Assume $\mathcal{T} = [0,1]$, and define $J(f) = \int_0^1 \left(f''\right)^2 dt$. If the inner product on the null space $\mathcal{N}_J$ is $\left(\int_0^1 f dt\right)\left(\int_0^1 g dt\right) + \left(\int_0^1 f' dt\right)\left(\int_0^1 g' dt\right)$, then we have $\mathcal{H}_J = \mathcal{H} \ominus \mathcal{N}_J = \{f : \int_0^1 f dt = \int_0^1 f' dt = 0, J(f) < \infty\}$, and the corresponding reproducing kernel is $R_J(s,t) = k_2(s)k_2(t) - k_4(|s-t|)$, where $k_v(t) = B_v(t)/\nu!$ is the scaled Bernoulli polynomial over $t \in [0,1]$. The null space $\mathcal{N}_J$ has a basis $\{1, k_1(t)\}$, where $k_1(t) = t - 0.5, t \in [0,1]$.

The Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$ is infinite dimensional. Therefore, minimizing equation 1 on $\mathcal{H}$ is not feasible. However, since we are only interested in the values of f at a finite number of points, Wahba (1990) has shown that the minimizer of equation 1 can be achieved in a finite dimensional subspace of $\mathcal{H}$.

B THEORETICAL DERIVATIONS OF CHANGE-POINTS DETECTION

B.1 MEAN CHANGE-POINT

Let $x_i \sim N(\mu_i, \sigma_i)$ be independent observations, where $i = 1, \ldots, n$, and σ_i is known. We want to test the hypothesis $H_0 : \mu_1 = \cdots = \mu_n = \mu$ versus $H_1 : \mu_1 = \cdots = \mu_{\tau_0} \neq \mu_{\tau_0+1} = \cdots = \mu_n$, where τ_0 is a potential mean change-point. Under H_0, assuming $\sigma_i^2 = 1$ for simplicity, the likelihood function is $L_0(\mu) = \frac{1}{(\sqrt{2\pi})^n} e^{-\sum_{i=1}^n (x_i-\mu)^2/2}$. Under H_1, the likelihood function is $L_1(\mu_1, \mu_2) = \frac{1}{(\sqrt{2\pi})^n} e^{-\left(\sum_{i=1}^k (x_i-\mu_1)^2 + \sum_{i=k+1}^n (x_i-\mu_2)^2\right)/2}$, where $\widehat{\mu_1} = \bar{x}_k = \frac{1}{k}\sum_{i=1}^k x_i$ and $\widehat{\mu_2} = \bar{x}_{n-k} = \frac{1}{n-k}\sum_{i=k+1}^n x_i$.

According to Theorem 4.2, the critical value c_α satisfies:

$$
\begin{aligned}
1-\alpha =& P\left[T_n^2 < c_\alpha \mid H_0\right] \\
=& P\left[-a_n^{-1} b_n < a_n^{-1}\left(T_n - b_n\right) < a_n^{-1}\left(c_\alpha^{1/2} - b_n\right) \mid H_0\right] \\
\cong& \exp\left\{-2\pi^{1/2}\exp\left[-a_n^{-1}\left(c_\alpha^{1/2} - b_n\right)\right]\right\} - \exp[-2\pi^{1/2}\exp(a_n^{-1} b_n)].
\end{aligned}
$$

Therefore, we have:

$$
c_\alpha \cong \left[-a_n \log\left\{-\frac{1}{2\pi^{1/2}}\log\left[1-\alpha+\exp\left(-2\pi^{1/2} e^{a_n^{-1} b_n}\right)\right]\right\} + b_n\right]^2,
$$

where $a_n = (2\log\log n)^{-1/2}, b_n = a_n^{-1} + \frac{1}{2} a_n \log\log\log n$.

B.2 VARIANCE CHANGE-POINT

The corresponding likelihood ratio limits and critical values are derived similarly based on the likelihood framework. Let $\ell(n) = -2L_0(\widehat{\sigma}^2) - n - n\log 2\pi$ and $\ell(k_1) = -2L_1(\widehat{\sigma}^2, \widehat{\delta}^2) - n - n\log 2\pi$, where L_0 and L_1 are the likelihood functions under the null and the alternative hypotheses of variance shifts.

According to Theorem 4.3, the critical value c_α for the test statistic $\lambda_n^2 = \max_{2\le k_1\le n-2}[\ell(n)-\ell(k_1)]$ satisfies

$$\begin{aligned}1-\alpha =& P\left[\lambda_n^2 < \log n + c_\alpha \mid H_0\right]\\ =& P\left[-b_n\log n < a_n(\log n)\lambda_n - b_n\log n < a_n\log n(\log n + c_\alpha)^{1/2} - b_n\log n \mid H_0\right]\\ \cong& \exp\left\{-2\exp\left[-a_n\log n(\log n + c_\alpha)^{1/2} + b_n\log n\right]\right\} - \exp[-2\exp(b_n\log n)].\end{aligned}$$

Therefore, we have

$$c_\alpha \cong \left\{-\frac{1}{a_n\log n}\log\log\left[1-\alpha+\exp\left(-2e^{b_n\log n}\right)\right]^{-1/2} + \frac{b_n\log n}{a_n\log n}\right\}^2 - \log n,$$

where $a_n = (2\log\log n)^{1/2}/\log n$, $b_n = \{2\log\log n + \frac{1}{2}\log\log\log n - \log\Gamma\left(\frac{1}{2}\right)\}/\log n$.

C Proofs of the Main Theorems

C.1 Regularity Assumptions

Assumption C.1. *When there exists a mean change point, the true mean change point* $\tau_0 \in \left[\frac{cn}{\log n}, n - \frac{cn}{\log n}\right]$ *for some* $c > 0$*, and* $\tau_0/n \overset{n\to\infty}{\longrightarrow} q_0 \in (0,1)$.

Assumption C.2. *When there exists a variance change point, the true variance change point* $\tau_1 \in \left[\frac{cn}{\log n}, n - \frac{cn}{\log n}\right]$ *for some* $c > 0$*, and* $\tau_1/n \overset{n\to\infty}{\longrightarrow} q_1 \in (0,1)$.

Assumption C.3. *The true mean functions* $f_-, f_+ \in S^m$*, i.e., the* m*-th order periodic Sobolev space defined on* $[0,1]$ *with period* 1.

Assumption C.4. *The random errors* ϵ_i*,* $i = 1, 2, \ldots, n$ *are independent normal random variables satisfying* $\mathbb{E}(\epsilon_i) = 0$ *and* $\mathrm{Var}(\epsilon_i) = \sigma_i^2$*, where* $\sigma_i = \sigma_0$ *for* $i \le \tau_1$*, and* $\sigma_i = \delta_0$ *for* $i > \tau_1$.

Assumption C.5. *Under the null hypothesis of no variance change-point, the random errors* $\epsilon_i, i = 1, 2, \ldots, n$ *are mutually independent and identically distributed normal random variables satisfying* $\mathbb{E}(\epsilon_i) = 0$ *and* $\mathrm{Var}(\epsilon_i) = \sigma_0^2$.

Remark C.1. Assumptions C.1 and C.2 are common in the literature on change-point analysis and essentially guarantee that the change-points are bounded away from the sequence boundaries. Assumption C.3 restricts our theory to the periodic reproducing kernel Hilbert space $\mathcal{H} = S^m$ to avoid complex matrix computations, though more general spaces could be considered. Assumptions C.4 and C.5 specify the error distributions under the alternative and null hypotheses of variance change, respectively.

The Gaussian assumption is part of the stated theorem scope; extensions to sub-Gaussian or heavy-tailed observations require separate arguments and calibration.

C.2 Technical Lemmas

In simultaneous estimation, direct proofs can be challenging. We therefore decompose the problem into several subproblems and facilitate subsequent iterations of the joint proof by establishing lemmas regarding the estimation properties of initial values. The main ideas and proof procedures for the three theorems are outlined below.

Before proving the theorem, we first present and prove two lemmas. These lemmas establish the convergence of initial values and the corresponding convergence of errors and residuals, laying the groundwork for the iterative proof. The detailed process is as follows.

First, we define some parameters under the alternative hypothesis H_1:

$$\begin{aligned}\delta_i &= \widehat{f}_-^{(0)}(i/n) - E\{\widehat{f}_-^{(0)}\}(i/n),\\ \delta_i^0 &= E\{\widehat{f}_-^{(0)}\}(i/n) - f_-(i/n),\\ \varphi_i &= \widehat{f}_+^{(0)}(i/n) - E\{\widehat{f}_+^{(0)}\}(i/n),\\ \varphi_i^0 &= E\{\widehat{f}_+^{(0)}\}(i/n) - f_+(i/n).\end{aligned}$$

The above four equations define the centered terms and estimation errors for the mean function estimates at both ends. Let $h = \lambda^{1/(2m)}$, $r_n = \sqrt{\log n/(nh)} + h^{m-1/2}$, $\widetilde{r}_n = r_n^2 + (nh)^{-3/4} + (\log n)^5(\log\log n)^2/n + n^{-1/2}$.

Lemma C.1. *There exist constants c_m and c'_m depending only on m such that*

$$\|E\{\widehat{f}_-^{(0)}\} - f_-\|_{\sup} \le c_m\sqrt{J(f_-)}\left(h^{m-1/2} + \left(\frac{h\log n}{n}\right)^{-1/2}\right), \tag{6}$$

$$\|E\{\widehat{f}_+^{(0)}\} - f_+\|_{\sup} \le c'_m\sqrt{J(f_+)}\left(h^{m-1/2} + \left(\frac{h\log n}{n}\right)^{-1/2}\right). \tag{7}$$

Lemma C.2. *Assume H_1 holds, and denote the true mean change point as τ_0 and variance change point as τ_1. Then the following hold:*

$$\max_{1\le k_1<k_2\le n}(k_2-k_1)^{-1/2}\left|\sum_{i=k_1+1}^{k_2}[\epsilon_i^2 - E(\epsilon_i^2)]\right| = O_p(\log n), \tag{8}$$

$$\max_{1\le k\le\tau_0}\frac{1}{\sqrt{k}}\left|\sum_{i=1}^{k}\epsilon_i(\widehat{f}_-^{(0)}(i/n) - f_-(i/n))\right| = O_p(n^{-1/4}h^{-3/4}), \tag{9}$$

$$\max_{1\le k\le\tau_0}\frac{1}{\sqrt{\tau_0-k}}\left|\sum_{i=k+1}^{\tau_0}\epsilon_i(\widehat{f}_-^{(0)}(i/n) - f_-(i/n))\right| = O_p(n^{-1/4}h^{-3/4}), \tag{10}$$

$$\max_{\tau_0\le k\le n}\frac{1}{\sqrt{k-\tau_0}}\left|\sum_{i=\tau_0}^{k}\epsilon_i(\widehat{f}_+^{(0)}(i/n) - f_+(i/n))\right| = O_p(n^{-1/4}h^{-3/4}), \tag{11}$$

$$\max_{\tau_0\le k\le n}\frac{1}{\sqrt{n-k}}\left|\sum_{i=k+1}^{n}\epsilon_i(\widehat{f}_+^{(0)}(i/n) - f_+(i/n))\right| = O_p(n^{-1/4}h^{-3/4}), \tag{12}$$

$$\max_{1\le i\le\tau_0}|\widehat{f}_-^{(0)}(i/n) - f_-(i/n)| = O_p(\sqrt{\log n/(nh)} + h^{m-1/2}), \tag{13}$$

$$\max_{\tau_0\le i\le n}|\widehat{f}_+^{(0)}(i/n) - f_+(i/n)| = O_p(\sqrt{\log n/(nh)} + h^{m-1/2}). \tag{14}$$

The results in equation 9, equation 10, equation 11, and equation 12 also hold under the null hypothesis H_0. When H_0 holds, $f_- = f_+$, and τ_0 represents an arbitrary integer in $[1, n]$. That is, in the absence of a mean change point, these expressions convey identical meanings and remain valid.

Below we provide proofs for lemmas C.1 and C.2.

Proof of Lemma C.1. Since $\widehat{f}_-^{(0)}$ and $\widehat{f}_+^{(0)}$ are minimizers of the optimization problem, under H_1, $\widehat{f}_-^{(0)}$ optimizes the problem on $[1, \tau_0]$ while $\widehat{f}_+^{(0)}$ optimizes on $[\tau_0+1, n]$. Considering $\widehat{f}_-^{(0)}$ first, we have

$$-\frac{1}{\tau_0}\sum_{i=1}^{\tau_0}\left(Y_i - \widehat{f}_-^{(0)}(i/\tau_0)\right)R_J(i/\tau_0,\cdot) + \lambda\widehat{f}_-^{(0)} = 0, \tag{15}$$

where R_J is the reproducing kernel associated with the space S^m (and J), and $R_J(x,\cdot)$ denotes the univariate function obtained by fixing the first argument at x. Taking expectations in equation 15 yields

$$\frac{1}{\tau_0}\sum_{i=1}^{\tau_0}\left(\bar{f}_-(i/\tau_0) - f_-(i/\tau_0)\right)R_J(i/\tau_0,\cdot) + \lambda\bar{f}_- = 0, \tag{16}$$

where $\bar{f}_- = E\{\widehat{f}_-^{(0)}\}$. Thus, $\bar{f}_-$ minimizes the following functional:

$$\ell_0(f) = \frac{1}{\tau_0}\sum_{i=1}^{\tau_0}\left(f(i/\tau_0) - f_-(i/\tau_0)\right)^2 + \lambda J(f).$$

From $\ell_0(\bar{f}_-) \le \ell_0(f_-)$, we obtain

$$\frac{1}{\tau_0}\sum_{i=1}^{\tau_0}\left(\bar{f}_-(i/\tau_0) - f_-(i/\tau_0)\right)^2 + \lambda J(\bar{f}_-) \le \lambda J(f_-),$$

implying $J(\bar{f}_-) \le J(f_-)$. Define $g(t) = (\bar{f}_-(t) - f_-(t))^2$. By Lemma 2.24 in Eggermont & LaRiccia (2001) we have

$$\begin{aligned}
\left|\frac{1}{\tau_0}\sum_{i=1}^{\tau_0} g(i/\tau_0) - \int_0^1 g(t)dt\right| &\le \frac{1}{\tau_0}\int_0^1 |g'(t)|dt \\
&= \frac{2}{\tau_0}\int_0^1 |\bar{f}_-'(t) - f_-'(t)| \cdot |\bar{f}_-(t) - f_-(t)|dt \\
&\le \frac{2}{\tau_0}\|\bar{f}_-' - f_-'\|_{L^2[0,1]}\|\bar{f}_- - f_-\|_{L^2[0,1]} \\
&\le \frac{2}{\tau_0}\|\bar{f}_-^{(m)} - f_-^{(m)}\|_{L^2[0,1]}\|\bar{f}_- - f_-\|_{L^2[0,1]} \\
&\le \frac{8}{\tau_0}J(f_-).
\end{aligned}$$

From equation 16, we have

$$\begin{aligned}
\|\bar{f}_- - f_-\|^2 \equiv \|\bar{f}_- - f_-\|_{L^2}^2 + \lambda J(\bar{f}_- - f_-) \\
&= -\lambda J(f_-, \bar{f}_- - f_-) + \int_0^1 g(t)dt - \frac{1}{\tau_0}\sum_{i=1}^{\tau_0} g(i/\tau_0) \\
&\le 2\lambda J(f_-) + \frac{8}{\tau_0}J(f_-) \\
&= 2J(f_-)\left(\lambda + \frac{4}{\tau_0}\right).
\end{aligned}$$

By Eggermont & LaRiccia (2001), we further obtain

$$\|\bar{f}_- - f_-\|_{\sup} \le d_m h^{-1/2}\|\bar{f}_- - f_-\| \le c_m\sqrt{J(f_-)}\left(h^{m-1/2} + (\tau_0 h)^{-1/2}\right). \tag{17}$$

By Assumption C.1, $\tau_0 \ge \frac{cn}{\log n}$. Without loss of generality, let $c = 1$, we have $\tau_0 \ge \frac{n}{\log n}$. Thus, equation 17 simplifies to

$$\|\bar{f}_- - f_-\|_{\sup} \le c_m\sqrt{J(f_-)}\left(h^{m-1/2} + \left(\frac{h\log n}{n}\right)^{-1/2}\right),$$

where c_m, d_m are positive constants depending only on m. Therefore, equation 6 holds.

Similarly, considering $\widehat{f}_+^{(0)}$ on $[\tau_0 + 1, n]$, we have

$$\|\bar{f}_+ - f_+\|_{\sup} \le d_m' h^{-1/2}\|\bar{f}_+ - f_+\| \le c_m'\sqrt{J(f_+)}\left(h^{m-1/2} + \{(n-\tau_0)h\}^{-1/2}\right). \tag{18}$$

By Assumption C.1, $\tau_0 \le n - \frac{cn}{\log n}$. Letting $c = 1$, we have $n - \tau_0 \ge \frac{n}{\log n}$. Therefore equation 18 becomes

$$\|\bar{f}_+ - f_+\|_{\sup} \le c_m'\sqrt{J(f_+)}\left(h^{m-1/2} + \left(\frac{h\log n}{n}\right)^{-1/2}\right),$$

where c_m', d_m' are positive constants depending only on m. Thus, equation 7 holds. □

Proof of Lemma 2. Let $\sigma_i^2 = E(\epsilon_i^2)$. Without loss of generality, assume $\sigma^2 = \sigma_1^2 = \cdots = \sigma_{k_0}^2 < \sigma_{k_0+1}^2 = \cdots = \sigma_n^2 = \delta^2$. Since $\epsilon_i^2 - \sigma_i^2$ are independent centered sub-exponential random variables,

by Vershynin (2010), there exist constants $c, d > 0$ such that for any $1 \le k_1 < k_2 \le n$,

$$\begin{aligned}
&P\left(\left|\sum_{i=k_1+1}^{k_2}[\epsilon_i^2 - \sigma_i^2]\right| \ge C\sqrt{k_2-k_1}\log n\right) \\
&\le 2\exp\left(-c\min\left\{\frac{C^2(k_2-k_1)(\log n)^2}{d^2(k_2-k_1)}, \frac{C\sqrt{k_2-k_1}\log n}{d}\right\}\right) \\
&\le 2\exp\left(-c\min\left\{\frac{C^2}{d^2}, \frac{C}{d}\right\}\log n\right) \\
&\le 2\exp(-3\log n) \\
&= \frac{2}{n^3},
\end{aligned}$$

where $C = \max\left\{\sqrt{\frac{3d^2}{c}}, \frac{3d}{c}\right\} > 0$. Therefore, as $n \to \infty$,

$$P\left(\max_{1\le k_1<k_2\le n}(k_2-k_1)^{-1/2}\left|\sum_{i=k_1+1}^{k_2}[\epsilon_i^2-\sigma_i^2]\right| \ge C\log n\right) \le \frac{2}{n} \to 0.$$

This establishes equation 8.

Next we prove the remaining results. We focus on the case under H_1, as the proof for H_0 is similar. Define $\Omega = (\Omega_1^T, \ldots, \Omega_n^T)^T$, where $\Omega_i = (R_J(1/n, i/n), \ldots, R_J(n/n, i/n))/n$. By the representer theorem (Wahba & Wang (2019)), we have

$$(\widehat{f}^{(0)}(1/n), \ldots, \widehat{f}^{(0)}(n/n))^T = \Omega(\Omega + \lambda I_n)^{-1}\mathbf{y}.$$

$$R_J(x,y) = \sum_{\nu=1}^{\infty}\frac{\varphi_\nu(x)\varphi_\nu(y)}{\gamma_\nu} = 2\sum_{k=1}^{\infty}\frac{\cos(2\pi k(x-y))}{(2\pi k)^{2m}}, \quad x, y \in \mathbb{I}.$$

For $0 \le l \le n-1$, define $c_l = \frac{2}{n}\sum_{k=1}^{\infty}\frac{\cos(2\pi kl/n)}{(2\pi k)^{2m}}$. Since $c_l = c_{n-l}$ for $l = 1, 2, \ldots, n-1$, Ω is a symmetric circulant matrix of order n.

Let $\zeta = \exp(2\pi\sqrt{-1}/n)$. The normalized eigenvectors of Ω are:

$$\mathbf{x}_k = \frac{1}{\sqrt{n}}\left(1, \zeta^k, \zeta^{2k}, \ldots, \zeta^{(n-1)k}\right)^T, \quad k = 0, 1, \ldots, n-1.$$

Let $M = (\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{n-1})$, and denote M^* as the conjugate transpose of M. Clearly, $MM^* = I_n$, and Ω can be decomposed as $\Omega = M\Lambda M^*$, where $\Lambda = \mathrm{diag}(\lambda_0, \lambda_1, \ldots, \lambda_{n-1})$ with

$$\lambda_l = c_0 + c_1\zeta^l + \cdots + c_{n-1}\zeta^{(n-1)l}.$$

Exact computation yields:

$$\lambda_l = \begin{cases} 2\sum_{k=1}^{\infty}\frac{1}{(2\pi kn)^{2m}}, & l = 0, \\ \sum_{k=1}^{\infty}\frac{1}{[2\pi(kn-l)]^{2m}} + \sum_{k=0}^{\infty}\frac{1}{[2\pi(kn+l)]^{2m}}, & 1 \le l \le n-1. \end{cases}$$

It can be verified that for $l = 0$, $\lambda_0 = 2\bar{c}_m(2\pi n)^{-2m}$ where $\bar{c}_m := \sum_{k=1}^{\infty}k^{-2m}$; while for $1 \le l \le n-1$,

$$\lambda_l = \frac{1}{[2\pi(n-l)]^{2m}} + \frac{1}{(2\pi l)^{2m}} + \sum_{k=2}^{\infty}\frac{1}{[2\pi(kn-l)]^{2m}} + \sum_{k=1}^{\infty}\frac{1}{[2\pi(kn+l)]^{2m}}. \tag{19}$$

Let $\underline{c}_m = \sum_{k=2}^{\infty}k^{-2m}$. Then:

$$\underline{c}_m(2\pi n)^{-2m} \le \sum_{k=2}^{\infty}\frac{1}{[2\pi(kn-l)]^{2m}} \le \bar{c}_m(2\pi n)^{-2m},$$

$$\underline{c}_m(2\pi n)^{-2m} \le \sum_{k=1}^{\infty} \frac{1}{[2\pi(kn+l)]^{2m}} \le \bar{c}_m(2\pi n)^{-2m}.$$

Assume $\epsilon \sim N(0, I_n)$ is a vector of independent standard normal random variables, which can be written as $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T = D_0\epsilon$, where D_0 is the square root of the true covariance matrix Σ_0 of ϵ, i.e., D_0 is diagonal with first τ_0 diagonal elements as σ_0 and the rest as δ_0. Define $\delta_{(k)} = (\delta_1, \ldots, \delta_k)^T$ and $\Omega_{(k)} = (\Omega_1^T, \ldots, \Omega_k^T)^T$. Define vectors:

$$\boldsymbol{\epsilon}_{(k)} = (\epsilon_1, \ldots, \epsilon_k)^T, \quad \boldsymbol{\epsilon}_{*(n-k)} = (\epsilon_{k+1}, \ldots, \epsilon_n)^T,$$
$$\boldsymbol{e}_{(k)} = (e_1, \ldots, e_k)^T, \quad \boldsymbol{e}_{*(n-k)} = (e_{k+1}, \ldots, e_n)^T.$$

Then:

$$\begin{aligned}
\delta_{(k)} &= \Omega_{(k)}(\Omega + \lambda I_n)^{-1}\boldsymbol{\epsilon} \\
&= (\mathbf{x}_0, \ldots, \mathbf{x}_{k-1})^T \Lambda M^* M(\Lambda + \lambda I_n)^{-1} M^* \boldsymbol{\epsilon} \\
&= (\mathbf{x}_0, \ldots, \mathbf{x}_{k-1})^T \Lambda(\Lambda + \lambda I_n)^{-1} M^* \boldsymbol{\epsilon} \\
&= (\mathbf{x}_0, \ldots, \mathbf{x}_{k-1})^T \Lambda(\Lambda + \lambda I_n)^{-1}(\bar{\mathbf{x}}_0, \ldots, \bar{\mathbf{x}}_{k-1})\boldsymbol{\epsilon}_{(k)} \\
&\quad + (\mathbf{x}_0, \ldots, \mathbf{x}_{k-1})^T \Lambda(\Lambda + \lambda I_n)^{-1}(\bar{\mathbf{x}}_k, \ldots, \bar{\mathbf{x}}_{n-1})\boldsymbol{\epsilon}_{*(n-k)},
\end{aligned}$$

where $\bar{\mathbf{x}}_k$ denotes the conjugate of $\mathbf{x}_k$.

When $k \le \tau_1$, define

$$D_k = \mathrm{diag}(\underbrace{\sigma_0, \ldots, \sigma_0}_{k \text{ elements}}).$$

When $k > \tau_1$, define

$$D_k = \mathrm{diag}(\underbrace{\sigma_0, \ldots, \sigma_0}_{\tau_1 \text{ elements}}, \underbrace{\delta_0, \ldots, \delta_0}_{k-\tau_1 \text{ elements}}).$$

Similarly, when $k \le n - \tau_1$, define

$$D_{*k} = \mathrm{diag}(\overbrace{\delta_0, \ldots, \delta_0}^{k \text{ elements}}).$$

When $k > n - \tau_1$, define

$$D_{*k} = \mathrm{diag}(\overbrace{\sigma_0, \ldots, \sigma_0}^{k-n+\tau_1 \text{ elements}}, \overbrace{\delta_0, \ldots, \delta_0}^{n-\tau_1 \text{ elements}}).$$

Clearly, $\boldsymbol{\epsilon}_{(k)} = D_k \boldsymbol{e}_{(k)}$ and $\boldsymbol{\epsilon}_{*(k)} = D_{*k}\boldsymbol{e}_{*(k)}$. Let

$$\widetilde{A}_k = (\mathbf{x}_0, \ldots, \mathbf{x}_{k-1})^T \Lambda(\Lambda + \lambda I_n)^{-1}(\bar{\mathbf{x}}_0, \ldots, \bar{\mathbf{x}}_{k-1}),$$
$$\widetilde{B}_k = (\mathbf{x}_0, \ldots, \mathbf{x}_{k-1})^T \Lambda(\Lambda + \lambda I_n)^{-1}(\bar{\mathbf{x}}_k, \ldots, \bar{\mathbf{x}}_{n-1}),$$

and define $A_k = D_k \widetilde{A}_k D_k$ and $B_k = D_k \widetilde{B}_k D_{*n-k}$. Then

$$\boldsymbol{\epsilon}_{(k)}^T \delta_{(k)} = \boldsymbol{e}_{(k)}^T A_k \boldsymbol{e}_{(k)} + \boldsymbol{e}_{(k)}^T B_k \boldsymbol{e}_{*(n-k)}.$$

By the Hanson-Wright inequality, for any $k = 1, \ldots, n$,

$$\begin{aligned}
&P\left(|\boldsymbol{e}_{(k)}^T A_k \boldsymbol{e}_{(k)} - E\{\boldsymbol{e}_{(k)}^T A_k \boldsymbol{e}_{(k)}\}| \ge C_n\sqrt{k/(nh)}\right) \\
&\le 2\exp\left(-\min\left\{\frac{c^2 C_n^2 k/(nh)}{\|A_k\|_F^2}, \frac{cC_n k/(nh)}{\|A_k\|_{\mathrm{op}}}\right\}\right),
\end{aligned} \tag{20}$$

where $\|\cdot\|_F$ and $\|\cdot\|_{\mathrm{op}}$ denote the Frobenius norm and spectral norm respectively, $C_n > 0$ is a constant depending only on n, and $c > 0$ is another constant.
Let $\Gamma = \Lambda(\Lambda + \lambda I_n)^{-1}$. Since Γ is diagonal, it can be expressed as $\Gamma = \mathrm{diag}(\gamma_0, \ldots, \gamma_{n-1})$.

Let $M_{(k)} = (\mathbf{x}_0, \ldots, \mathbf{x}_{k-1})^T$. Since $M^*M = I_n$, we have $M^*_{(k)}M_{(k)} \leq I_n$. Set $a_0 = \max\{\sigma^2, \delta^2\}$. Then we can show:

$$\begin{aligned}\|A_k\|_F^2 &\leq a_0^2 \operatorname{Tr}(\widetilde{A}_k^* \widetilde{A}_k) = a_0^2 \operatorname{Tr}(M_{(k)}\Gamma M^*_{(k)} M_{(k)}\Gamma M^*_{(k)}) \\ &\leq a_0^2 \operatorname{Tr}(M_{(k)}\Gamma^2 M^*_{(k)}) = a_0^2 \sum_{l=0}^{k-1} \mathbf{x}_l^T \operatorname{diag}(\gamma_0^2, \ldots, \gamma_{n-1}^2)\bar{\mathbf{x}}_l \\ &= \frac{a_0^2 k}{n} \sum_{r=0}^{n-1} \gamma_r^2 = O\left(\frac{k}{nh}\right), \quad \text{uniformly in } k.\end{aligned}$$

This implies $\|A_k\|_{\text{op}} \leq \|A_k\|_F = O\left(\sqrt{k/(nh)}\right)$ uniformly in k. Therefore, for $C_n > 1$, equation 20 simplifies to:

$$P\left(|\boldsymbol{e}_{(k)}^T A_k \boldsymbol{e}_{(k)} - E\{\boldsymbol{e}_{(k)}^T A_k \boldsymbol{e}_{(k)}\}| \geq C_n\sqrt{k/(nh)}\right) \leq 2\exp(-cC_n).$$

Furthermore:

$$P\left(\max_{1\leq k\leq n} \frac{|\boldsymbol{e}_{(k)}^T A_k \boldsymbol{e}_{(k)} - E\{\boldsymbol{e}_{(k)}^T A_k \boldsymbol{e}_{(k)}\}|}{\sqrt{k/(nh)}} \geq C_n\right) \leq 2n\exp(-cC_n).$$

Setting $C_n = (2/c)\log n$, we obtain

$$\max_{1\leq k\leq n} \frac{|\boldsymbol{e}_{(k)}^T A_k \boldsymbol{e}_{(k)} - E\{\boldsymbol{e}_{(k)}^T A_k \boldsymbol{e}_{(k)}\}|}{\sqrt{k/(nh)}} = O_P(\log n).$$

Additionally, by probability inequality:

$$P\left(\max_{1\leq k\leq n} \frac{|\sum_{i=1}^k \delta_i^0 \epsilon_i|}{\sqrt{\sum_{i=1}^k (\delta_i^0\sigma_i)^2}} \geq C\right) \leq nP(|Z| > C) = O(n\exp(-C^2/2)),$$

which implies

$$\max_{1\leq k\leq n} \frac{|\sum_{i=1}^k \delta_i^0 \epsilon_i|}{\sqrt{\sum_{i=1}^k (\delta_i^0\sigma_i)^2}} = O_P(\sqrt{\log n}).$$

By Lemma C.1 and conditions $h(\log n)^2 = o(1)$ and $nh^{4m+1}(\log n)^2 = o(1)$, we have

$$\begin{aligned}&\frac{1}{\sqrt{k}} \sum_{i=1}^k \epsilon_i(\delta_i + \delta_i^0) \\ &= \frac{1}{\sqrt{k}} \epsilon_{(k)}^T \delta_{(k)} + \frac{1}{\sqrt{k}} \sum_{i=1}^k \epsilon_i \delta_i^0 \\ &= \frac{1}{\sqrt{k}} \left(\boldsymbol{e}_{(k)}^T A_k \boldsymbol{e}_{(k)} + \boldsymbol{e}_{(k)}^T B_k \boldsymbol{e}_{*(n-k)}\right) + \frac{1}{\sqrt{k}} \sum_{i=1}^k \epsilon_i \delta_i^0 \\ &= O_P\Bigg(\frac{\sqrt{k}}{nh} + \frac{\log n}{\sqrt{nh}} \\ &\quad + \sqrt{\left(\frac{1}{nh} + \sqrt{\frac{1}{knh}}\right)\left(\frac{n-k}{nh} + \sqrt{\frac{n-k}{nh}}\right)} \\ &\quad + \sqrt{\log n\left(h^{m-1/2} + (nh)^{-1/2}\right)}\Bigg)\end{aligned}$$

$$
\begin{aligned}
&= O_P\Bigg(\frac{1}{\sqrt{nh}} + \frac{\log n}{\sqrt{nh}} \\
&\quad + \sqrt{\left(\frac{1}{nh} + \frac{1}{(nh)^{1/4}}\right)\left(\frac{1}{h} + \frac{1}{h^{1/2}}\right)} \\
&\quad + \sqrt{\log n\left(h^{m-1/2} + (nh)^{-1/2}\right)}\Bigg) \\
&= O_P\left(\frac{1}{\sqrt{nh}} + \frac{\log n}{\sqrt{nh}} + \left(\frac{1}{nh^3}\right)^{1/4} + \sqrt{\log n\left(h^{m-1/2} + (nh)^{-1/2}\right)}\right) \\
&= O_P\left(n^{-1/4}h^{-3/4}\right),
\end{aligned}
$$

where the O_P term holds uniformly for $1 \le k \le n$, thus establishing equation 9. Similarly, we can show that equation 10, equation 11, and equation 12 hold.

We now prove the uniform bound for $\|\widehat{f}_i^{(0)} - \bar{f}_i\|_{\sup}$. By the representer theorem, we have

$$
\widehat{f}_i^{(0)} - \bar{f}_i = \left(R_J(1/n,\cdot),\ldots,R_J(n/n,\cdot)\right)(\Omega + \lambda I_n)^{-1}\boldsymbol{\epsilon}/n.
$$

where, Ω_i is the i-th row of Ω, and for $1 \le i \le \tau_0$,

$$
\delta_i = \Omega_i(\Omega + \lambda I_n)^{-1}\boldsymbol{\epsilon} \sim N\left(0, \Omega_i(\Omega + \lambda I_n)^{-2}\Omega_i^T\sigma_i^2\right),
$$

while for $\tau_0 + 1 \le i \le n$,

$$
\varphi_i = \Omega_i(\Omega + \lambda I_n)^{-1}\boldsymbol{\epsilon} \sim N\left(0, \Omega_i(\Omega + \lambda I_n)^{-2}\Omega_i^T\delta_i^2\right).
$$

Note that

$$
\Omega_i(\Omega + \lambda I_n)^{-2}\Omega_i^T = \mathbf{x}_{i-1}^T\Lambda(\Lambda + \lambda I_n)^{-2}\Lambda\bar{\mathbf{x}}_{i-1} = \frac{1}{n}\sum_{r=0}^{n-1}\frac{\lambda_r^2}{(\lambda + \lambda_r)^2} \approx \frac{1}{nh}.
$$

Therefore, as $n \to \infty$, for $1 \le i \le \tau_0$,

$$
P\left(\max_{1\le i\le\tau_0}|\delta_i| \ge C\frac{\sqrt{\log n}}{\sqrt{nh}}\right) \le nP\left(|\delta_i| \ge C\frac{\sqrt{\log n}}{\sqrt{nh}}\right) \le n\exp(-C\log n) \to 0.
$$

This shows $\max_{1\le i\le\tau_0}|\delta_i| = O_P\left(\sqrt{\frac{\log n}{nh}}\right)$. Combined with Lemma 1, this establishes (3.8). Similarly, as $n \to \infty$, for $\tau_0 + 1 \le i \le n$,

$$
P\left(\max_{\tau_0+1\le i\le n}|\varphi_i| \ge C\frac{\sqrt{\log n}}{\sqrt{nh}}\right) \le nP\left(|\varphi_i| \ge C\frac{\sqrt{\log n}}{\sqrt{nh}}\right) \le n\exp(-C\log n) \to 0.
$$

showing $\max_{\tau_0+1\le i\le n}|\varphi_i| = O_P\left(\sqrt{\frac{\log n}{nh}}\right)$. Combined with Lemma C.1, this establishes equation 14. □

With Lemmas C.1 and C.2 prepared as our main tools, we now proceed to prove the three theorems sequentially. For readability and writing convenience, we will state additional lemmas as needed within the proof process.

C.3 Proof of Theorem 4.1

The consistency of parameter estimation aims to verify the convergence of parameter estimates, demonstrating the accuracy and general applicability of our algorithmic procedure. We decompose the proof using conclusions from the lemmas to derive results step by step. Below is the detailed proof process and reasoning for parameter consistency estimation.

Proof of Theorem 4.1. The consistency results for estimated parameters can be decomposed into the following five parts:

I. Consistency of the initialized mean function estimates;

II. Consistency of the mean and variance change point estimates given the consistent estimates of the mean functions;

III. Consistency of the mean shift quantity estimate given the consistent estimates of the mean functions;

IV. Consistency of the variance estimates given the consistent estimates of the mean functions;

V. Consistency of the mean function estimates given the consistent estimates of the mean/variance change points, quantity of the mean shift, and variances.

Specifically, we will prove the following objectives:

$$\max_{1\le i\le \tau_0} |\widehat{f}_-^{(0)}(i/n) - f_-(i/n)| = O_P(r_n), \tag{21}$$

$$\max_{\tau_0+1\le i\le n} |\widehat{f}_+^{(0)}(i/n) - f_+(i/n)| = O_P(r_n). \tag{22}$$

At the ι-th iteration, given $\max_{1\le i\le \tau_0} |\widehat{f}_-^{(\iota-1)}(i/n) - f_-(i/n)| = O_P(r_n)$ and $\max_{\tau_0+1\le i\le n} |\widehat{f}_+^{(\iota-1)}(i/n) - f_+(i/n)| = O_P(r_n)$, we have

$$\begin{aligned}
\left|\widehat{\tau}_0^{(\iota)} - \tau_0\right| &= O_P\left((\log n)^4(\log\log n)^2\right), \\
\left|\widehat{\tau}_1^{(\iota)} - \tau_1\right| &= O_P\left((\log n)^4(\log\log n)^2\right), \\
\left|\widehat{\kappa}^{(\iota)} - \kappa\right| &= O_P\left(r_n + \frac{(\log n)^4(\log\log n)^2}{n}\right), \\
\left[\widehat{\sigma}^2\right]^{(\iota)} &= \sigma_0^2 + O_P(\widetilde{r}_n), \\
\left[\widehat{\delta}^2\right]^{(\iota)} &= \delta_0^2 + O_P(\widetilde{r}_n).
\end{aligned}$$

Given the convergence above, we have

$$\begin{aligned}
\max_{1\le i\le \tau_0} |\widehat{f}_-^{(\iota)}(i/n) - f_-(i/n)| &= O_P(r_n), \\
\max_{\tau_0+1\le i\le n} |\widehat{f}_+^{(\iota)}(i/n) - f_+(i/n)| &= O_P(r_n).
\end{aligned}$$

These results collectively ensure the consistency of the estimated parameters $(\widehat{f}_-, \widehat{f}_+, \widehat{\tau}_0, \widehat{\tau}_1, \widehat{\kappa}, \widehat{\sigma}^2, \widehat{\delta}^2)$. When combining iterative convergence results, a practical concern often arises: since all results are stated in probability, after multiple iterations the "high probability" might weaken in practice. However, our algorithm typically converges in few iterations, making this issue negligible. To simplify notation, we omit superscripts $(\iota-1)$ and (ι) in this proof.

I. Proof of the consistency of the initial mean function estimates:

By Lemma C.2, objectives equation 21 and equation 22 follow directly from equation 13 and equation 14.

II. Proof of the consistency of the mean and variance change point estimates:

For simplicity and without loss of generality, assume $\sigma_0^2 < \delta_0^2$ and $\tau_0 \ge \tau_1$. For the mean change point, our approach is to show that for $k \in [cn/\log n, n - cn/\log n]$ satisfying $|k - \tau_0| \ge (\log n)^4(\log\log n)^2$, we have $S_k > S_{\tau_0}$ uniformly. Similarly, for the variance change point, we show that for $k \in [cn/\log n, n - cn/\log n]$ satisfying $|k - \tau_1| \ge (\log n)^4(\log\log n)^2$, we have $\ell(k) > \ell(\tau_1)$ uniformly.

We only need to consider $cn/\log n \le k < \tau_i - (\log n)^4(\log\log n)^2$ for $i = 0, 1$ (the case $n - cn/\log n \ge k > \tau_i + (\log n)^4(\log\log n)^2$ is similar).

Define $\eta_i = \widehat{f}_-(i/n) - f_-(i/n)$ for $1 \le i \le \tau_0$ and $\gamma_i = \widehat{f}_+(i/n) - f_+(i/n)$ for $\tau_0 + 1 \le i \le n$. Let

$$\widehat{\sigma}_0^2 = \frac{1}{\tau_1}\sum_{i=1}^{\tau_1}(\eta_i + \epsilon_i)^2, \quad \widehat{\sigma}_n^2 = \frac{1}{n-\tau_1}\left[\sum_{i=\tau_1+1}^{\tau_0}(\eta_i+\epsilon_i)^2 + \sum_{i=\tau_0+1}^{n}(\gamma_i+\epsilon_i)^2\right],$$

$$\widehat{\sigma}_k^2 = \frac{1}{\tau_1 - k}\sum_{i=k+1}^{\tau_1}(\eta_i+\epsilon_i)^2.$$

By Lemmas C.1 and C.2 and the scaled sequence variance being 1, we have uniformly

$$\frac{1}{\tau_1 - k}\sum_{i=k+1}^{\tau_1}\eta_i\epsilon_i = O_P(r_n\sqrt{\log n})$$

for $k \le \tau_1 - (\log n)^4(\log\log n)^2$. Thus we obtain

$$\begin{aligned}\widehat{\sigma}_k^2 =& \frac{1}{\tau_1-k}\sum_{i=k+1}^{\tau_1}\eta_i^2 + \frac{2}{\tau_1-k}\sum_{i=k+1}^{\tau_1}\eta_i\epsilon_i + \frac{1}{\tau_1-k}\sum_{i=k+1}^{\tau_1}\epsilon_i^2\\ =&\sigma^2 + O_P\left(r_n^2 + r_n\sqrt{\log n} + (\log n\log\log n)^{-1}\right)\\ =&\sigma^2 + O_P(r_{1n}^2),\end{aligned}$$

where $r_{1n}^2 = r_n^2 + r_n\sqrt{\log n} + (\log n\log\log n)^{-1}$, which is $o(1)$ by assumption. Similarly, we get

$$\widehat{\sigma}_0^2 = \sigma^2 + O_P(r_{1n}^2), \quad \widehat{\sigma}_n^2 = \delta^2 + O_P(r_{1n}^2).$$

Therefore, for $k \le \tau_1 - (\log n)^4(\log\log n)^2$, the following hold uniformly with probability approaching 1:

$$|\widehat{\sigma}_k^2/\widehat{\sigma}_n^2 - \sigma^2/\delta^2| = O_P(r_{1n}^2), \quad |\widehat{\sigma}_0^2/\widehat{\sigma}_n^2 - \sigma^2/\delta^2| = O_P(r_{1n}^2),$$

$$\left|1 - \frac{\widehat{\sigma}_k^2}{\widehat{\sigma}_0^2}\right| = \left|\frac{\widehat{\sigma}_0^2 - \widehat{\sigma}_k^2}{\widehat{\sigma}_0^2}\right| = O_P(r_{1n}^2).$$

Observe that

$$\frac{1}{k}\sum_{i=1}^{k}(y_i - \widehat{f}_-(i/n))^2 = \frac{1}{k}\sum_{i=1}^{k}(\delta_i+\epsilon_i)^2 = \widehat{\sigma}_0^2 + \frac{\tau_1-k}{k}(\widehat{\sigma}_0^2 - \widehat{\sigma}_k^2),$$

$$\begin{aligned}&\frac{1}{n-k}\left[\sum_{i=k+1}^{\tau_0}(y_i - \widehat{f}_-(i/n))^2 + \sum_{i=\tau_0+1}^{n}(y_i - \widehat{f}_+(i/n))^2\right]\\ =&\frac{1}{n-k}\left[\sum_{i=k+1}^{\tau_0}(\delta_i+\epsilon_i)^2 + \sum_{i=\tau_0+1}^{n}(\varphi_i+\epsilon_i)^2\right]\\ =&\widehat{\sigma}_n^2 + \frac{\tau_1-k}{n-k}(\widehat{\sigma}_k^2 - \widehat{\sigma}_n^2).\end{aligned}$$

Therefore,

$$\begin{aligned}&\ell(k)\\ =&\frac{k}{n}\log\left(\widehat{\sigma}_0^2 + \frac{\tau_1-k}{k}(\widehat{\sigma}_0^2 - \widehat{\sigma}_k^2)\right) + \frac{n-k}{n}\log\left(\widehat{\sigma}_n^2 + \frac{\tau_1-k}{n-k}(\widehat{\sigma}_k^2 - \widehat{\sigma}_n^2)\right)\\ =&\ell(\tau_1) - \frac{\tau_1-k}{n}\log\left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2}\right) + \frac{n-k}{n}\log\left(1 + \frac{\tau_1-k}{n-k}\left(\frac{\widehat{\sigma}_k^2}{\widehat{\sigma}_n^2} - 1\right)\right)\\ &+ \frac{k}{n}\log\left(1 + \frac{\tau_1-k}{k}\left(1 - \frac{\widehat{\sigma}_k^2}{\widehat{\sigma}_0^2}\right)\right).\end{aligned}$$

Note that

$$\begin{aligned}
&\frac{n-k}{n}\log\left(1+\frac{\tau_1-k}{n-k}\left(\frac{\widehat{\sigma}_k^2}{\widehat{\sigma}_n^2}-1\right)\right)\\
=&\frac{n-k}{n}\log\left(1+\frac{\tau_1-k}{n-k}\left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2}-1+\frac{\widehat{\sigma}_k^2-\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2}\right)\right)\\
=&\frac{n-k}{n}\log\left(1+\frac{\tau_1-k}{n-k}\left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2}-1\right)\right)+\frac{\tau_1-k}{n}O_P(r_{1n}^2).
\end{aligned}$$

Thus, for $k<\tau_1-(\log n)^4(\log\log n)^2$, we have uniformly

$$\begin{aligned}
&\ell(k)-\ell(\tau_1)\\
=&-\frac{\tau_1-k}{n}\log\left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2}\right)+\frac{n-k}{n}\log\left(1+\frac{\tau_1-k}{n-k}\left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2}-1\right)\right)+\frac{\tau_1-k}{n}O_P(r_{1n}^2)\\
=&\frac{\tau_1-k}{n}\sum_{l=1}^{\infty}\frac{1}{l}\left(1-\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2}\right)^l\left[1-\left(\frac{\tau_1-k}{n-k}\right)^{l-1}\right]+\frac{\tau_1-k}{n}O_P(r_{1n}^2)\\
\geq&\frac{\tau_1-k}{2n}\left(1-\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_n^2}\right)^2\frac{n-\tau_1}{n-k}+\frac{\tau_1-k}{n}O_P(r_{1n}^2)\\
\geq&\frac{\tau_1-k}{2n}\left[(1-\sigma^2/\delta^2)^2(1-q_0)+O_P(r_{1n}^2)\right]>0,
\end{aligned}$$

where the last inequality holds because $r_{1n}^2=o(1)$. This shows that with probability approaching 1, $\widehat{\tau}_1\geq\tau_1-(\log n)^4(\log\log n)^2$. Similarly, we can show that with probability approaching 1, for $k>\tau_1+(\log n)^4(\log\log n)^2$, we have $\ell(k)-\ell(\tau_1)>0$ uniformly, implying $\widehat{\tau}_1\leq\tau_1+(\log n)^4(\log\log n)^2$. Therefore, we have

$$|\widehat{\tau}_1-\tau_1|=O_P((\log n)^4(\log\log n)^2). \tag{23}$$

Similarly, for the mean change point τ_0, we have

$$S_k=\sum_{i=1}^{k}\left(x_i-\bar{x}_k\right)^2+\sum_{i=k+1}^{n}\left(x_i-\bar{x}_{n-k}\right)^2.$$

By Lemmas C.1-C.2, we obtain

$$\begin{aligned}
\sum_{i=1}^{k}\left(x_i-\bar{x}_k\right)^2=&\sum_{i=1}^{k}\left(x_i-E(x_i)+E(x_i)-\bar{x}_k\right)^2\\
=&\sum_{i=1}^{k}\left[(x_i-E(x_i))^2+(E(x_i)-\bar{x}_k)^2+2(x_i-E(x_i))(E(x_i)-\bar{x}_k)\right]\\
=&k+O_P\left(r_n^2+r_n\sqrt{\log n}+(\log n\log\log n)^{-1}\right)\\
=&k+O_P(r_{1n}^2).
\end{aligned}$$

Similarly, we get

$$\sum_{i=k+1}^{n}\left(x_i-\bar{x}_{n-k}\right)^2=n-k+O_P(r_{1n}^2),$$

$$\sum_{i=1}^{\tau_0}\left(x_i-\bar{x}_{\tau_0}\right)^2=\tau_0+O_P(r_{1n}^2),$$

$$\sum_{i=\tau_0+1}^{n}\left(x_i-\bar{x}_{n-\tau_0}\right)^2=n-\tau_0+O_P(r_{1n}^2).$$

Therefore,

$$\begin{aligned} S_k - S_{\tau_0} =& \sum_{i=1}^{k} (x_i - \bar{x}_k)^2 + \sum_{i=k+1}^{n} (x_i - \bar{x}_{n-k})^2 - \sum_{i=1}^{\tau_0} (x_i - \bar{x}_{\tau_0})^2 - \sum_{i=\tau_0+1}^{n} (x_i - \bar{x}_{n-\tau_0})^2 \\ =& k + (n-k) - \tau_0 - (n-\tau_0) + O_P(r_{1n}^2) \\ =& O_P(r_{1n}^2). \end{aligned}$$

This shows that with probability approaching 1, $\widehat{\tau}_0 \geq \tau_0 - (\log n)^4(\log\log n)^2$. Similarly, we can show that with probability approaching 1, for $k > \tau_0 + (\log n)^4(\log\log n)^2$, we have $S_k - S_{\tau_0} = O_P(r_{1n}^2)$ uniformly, implying $\widehat{\tau}_0 \leq \tau_0 + (\log n)^4(\log\log n)^2$. Therefore,

$$|\widehat{\tau}_0 - \tau_0| = O_P((\log n)^4(\log\log n)^2).$$

III. Proof of the consistency of the estimate of the mean shift quantity:

Recall that $\kappa = f_+(t_0) - f_-(t_0)$ with $t_0 = \tau_0/n$, and the estimator is $\widehat{\kappa} = \widehat{f}_+^{(\widehat{\tau}_0)}(\widehat{t}) - \widehat{f}_-^{(\widehat{\tau}_0)}(\widehat{t})$ with $\widehat{t} = \widehat{\tau}_0/n$. Write

$$\begin{aligned} |\widehat{\kappa} - \kappa| =& \Big| \big[\widehat{f}_+^{(\widehat{\tau}_0)}(\widehat{t}) - f_+(\widehat{t})\big] - \big[\widehat{f}_-^{(\widehat{\tau}_0)}(\widehat{t}) - f_-(\widehat{t})\big] \\ &\qquad + \big(f_+(\widehat{t}) - f_+(t_0)\big) - \big(f_-(\widehat{t}) - f_-(t_0)\big) \Big| \\ \leq& \underbrace{\big|\widehat{f}_+^{(\widehat{\tau}_0)}(\widehat{t}) - f_+(\widehat{t})\big| + \big|\widehat{f}_-^{(\widehat{\tau}_0)}(\widehat{t}) - f_-(\widehat{t})\big|}_{(I)} + \underbrace{\big|f_+(\widehat{t}) - f_+(t_0) - \{f_-(\widehat{t}) - f_-(t_0)\}\big|}_{(II)}. \end{aligned}$$

For term (I): by the same spline/RKHS arguments equation 13 and equation 14 in Lemma C.2, we have the following uniform error bounds over the observation grid:

$$\max_{1\leq i\leq \tau_0} \big|\widehat{f}_-^{(k_0)}(i/n) - f_-(i/n)\big| = O_P(r_n), \qquad \max_{\tau_0 \leq i\leq n} \big|\widehat{f}_+^{(k_0)}(i/n) - f_+(i/n)\big| = O_P(r_n).$$

In particular, evaluating at $\widehat{t}_- = \widehat{\tau}_0/n$ and $\widehat{t}_+ = (\widehat{\tau}_0 + 1)/n$ yields

$$\big|\widehat{f}_-^{(k_0)}(\widehat{t}_-) - f_-(\widehat{t}_-)\big| = O_P(r_n), \qquad \big|\widehat{f}_+^{(k_0)}(\widehat{t}_+) - f_+(\widehat{t}_+)\big| = O_P(r_n).$$

Hence, $(I) = O_P(r_n)$.

For term (II): by the smoothness assumptions on f_-, f_+ (Assumptions C.1 and C.2 imply bounded first derivatives on $[0,1]$), the mean value theorem gives

$$(II) \;\leq\; \|f_+'\|_\infty\, |\widehat{t} - t_0| + \|f_-'\|_\infty\, |\widehat{t} - t_0| \;\leq\; C\, \frac{|\widehat{\tau}_0 - \tau_0|}{n}.$$

By Part 2 proved above, $|\widehat{\tau}_0 - \tau_0| = O_P\big((\log n)^4(\log\log n)^2\big)$, so

$$(II) = O_P\left(\frac{(\log n)^4(\log\log n)^2}{n}\right).$$

Combining (I) and (II) yields

$$|\widehat{\kappa} - \kappa| = O_P\left(r_n + \frac{(\log n)^4(\log\log n)^2}{n}\right).$$

Since $\lambda \asymp n^{-2m/(2m+1)}$ implies $h \asymp \lambda^{1/(2m)} \asymp n^{-1/(2m+1)}$, we have $r_n = \sqrt{\log n/(nh)} + h^{m-1/2} \asymp n^{-\frac{2m-1}{4m+2}}$ (up to a $\sqrt{\log n}$ factor), which dominates $(\log n)^4(\log\log n)^2/n$ for any $m \geq 1$. Hence $|\widehat{\kappa} - \kappa| = O_P(r_n)$.

IV. Proof of the consistency of the variance estimates:

To prove the consistency of $\widehat{\sigma}^2$ and $\widehat{\delta}^2$, by Lemma C.2 we have

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{1}{\widehat{\tau}_1}\sum_{i=1}^{\widehat{\tau}_1}\eta_i^2 + \frac{2}{\widehat{\tau}_1}\sum_{i=1}^{\widehat{\tau}_1}\epsilon_i\eta_i + \frac{1}{\widehat{\tau}_1}\sum_{i=1}^{\widehat{\tau}_1}\epsilon_i^2 \\ &= O_P\left(\frac{\log n}{nh} + h^{2m-1} + (nh)^{-3/4}\right) + \frac{1}{\widehat{\tau}_1}\sum_{i=1}^{\widehat{\tau}_1}\epsilon_i^2 \\ &= O_P\left(r_n^2 + (nh)^{-3/4}\right) + \frac{1}{\widehat{\tau}_1}\left(\sum_{i=1}^{\widehat{\tau}_1}\epsilon_i^2 - \sum_{i=1}^{\tau_1}\epsilon_i^2\right) + \frac{1}{\widehat{\tau}_1}\sum_{i=1}^{\tau_1}\epsilon_i^2.\end{aligned}$$

Combining Lemma C.2 and equation 23, we have

$$\begin{aligned}\frac{1}{\widehat{\tau}_1}\left|\sum_{i=1}^{\widehat{\tau}_1}\epsilon_i^2 - \sum_{i=1}^{\tau_1}\epsilon_i^2\right| &\le \frac{1}{\widehat{\tau}_1}\sum_{i=\tau_1+1-|\Delta|}^{\tau_1+1+|\Delta|}\epsilon_i^2 \\ &= \frac{2|\widehat{\tau}_1-\tau_1|+1}{\widehat{\tau}_1} \times \frac{1}{2|\widehat{\tau}_1-\tau_1|+1}\sum_{i=\tau_1+1-|\widehat{\tau}_1-\tau_1|}^{\tau_1+1+|\widehat{\tau}_1-\tau_1|}\epsilon_i^2 \\ &= \frac{2|\widehat{\tau}_1-\tau_1|+1}{\widehat{\tau}_1} \times O_P\left(1 + \frac{\log n}{\sqrt{2|\widehat{\tau}_1-\tau_1|+1}}\right) \\ &= \frac{2|\widehat{\tau}_1-\tau_1|+1}{\widehat{\tau}_1} \times O_P\left(1 + \frac{\log n}{(\log n)^2\log\log n}\right) \\ &= O_P\left(\frac{(\log n)^4(\log\log n)^2}{n}\right)\end{aligned}$$

and

$$\frac{1}{\widehat{\tau}_1}\sum_{i=1}^{\tau_1}\epsilon_i^2 = \frac{\tau_1}{\widehat{\tau}_1}\frac{1}{\tau_1}\sum_{i=1}^{\tau_1}\epsilon_i^2 = \sigma_0^2 + O_P(n^{-1/2}). \tag{24}$$

In summary, we obtain

$$\widehat{\sigma}^2 = \sigma_0^2 + O_P\left(r_n^2 + (nh)^{-3/4} + (\log n)^4(\log\log n)^2/n + n^{-1/2}\right) = \sigma_0^2 + O_P(\widetilde{r}_n).$$

Similarly, we can prove $\widehat{\delta}^2 = \delta_0^2 + O_P(\widetilde{r}_n)$.

V. Proof of the consistency of the mean function estimates:

For the consistency of mean function estimates given consistent estimators $\widehat{\tau}_0$, $\widehat{\tau}_1$, $\widehat{\sigma}_0^2$, and $\widehat{\delta}_0^2$, recall the proof of Lemma C.2. We write $\epsilon = D_0\mathbf{e}$, where D_0 is the square root of the true covariance matrix of ϵ, and $\mathbf{e}$ is an n-dimensional vector of independent standard normal random variables. By the representer theorem, the estimators $\widehat{f}_-^{(0)}$ and $\widehat{f}_-$ have explicit expressions:

$$\widehat{f}_-^{(0)} = \Omega_{(\tau_0)}(\Omega + \lambda I)^{-1}f_- + \Omega_{(\tau_0)}(\Omega + \lambda I)^{-1}D_0\mathbf{e},$$

$$\widehat{f}_- = \Omega_{(\tau_0)}(\Omega + \lambda_{\text{new}}\widehat{\Sigma})^{-1}f_- + \Omega_{(\tau_0)}(\Omega + \lambda_{\text{new}}\widehat{\Sigma})^{-1}D_0\mathbf{e}.$$

Similarly, the estimators $\widehat{f}_+^{(0)}$ and $\widehat{f}_+$ have explicit expressions:

$$\widehat{f}_+^{(0)} = \Omega_{(*(n-\tau_0))}(\Omega + \lambda I)^{-1}f_+ + \Omega_{(*(n-\tau_0))}(\Omega + \lambda I)^{-1}D_0\mathbf{e},$$

$$\widehat{f}_+ = \Omega_{(*(n-\tau_0))}(\Omega + \lambda_{\text{new}}\widehat{\Sigma})^{-1}f_+ + \Omega_{(*(n-\tau_0))}(\Omega + \lambda_{\text{new}}\widehat{\Sigma})^{-1}D_0\mathbf{e}.$$

Without loss of generality, assume $\sigma^2 < \delta^2$, and $\kappa > 0$. Let $\widehat{c} = \widehat{\delta}^2/\widehat{\sigma}^2$ and $c_0 = \delta^2/\sigma^2$. By the consistency of $\widehat{\tau}_0$, $\widehat{\tau}_1$, $\widehat{\kappa}$, $\widehat{\sigma}^2$, and $\widehat{\delta}^2$, the set $\mathcal{E}_n \equiv \{\widehat{\sigma}^2 < \widehat{\delta}^2, |\widehat{c} - c_0| \le C\widetilde{r}_n\}$ holds with probability

approaching 1, where $\varepsilon > 0$ can be arbitrarily small. Let 1_k denote a k-dimensional vector of all ones, and 0_k a k-dimensional vector of all zeros. On $\mathcal{E}_n$:

$$\begin{aligned}\lambda_{\text{new}}\widehat{\Sigma} &= \lambda \ \mathrm{diag}(\widehat{c}1_{\widehat{\tau}_1}, 1_{n-\widehat{\tau}_1}) \\ &= \lambda \ \mathrm{diag}(c_0 1_{\widehat{\tau}_1}, 1_{n-\widehat{\tau}_1}) + \lambda(\widehat{c} - c_0) \ \mathrm{diag}(1_{\widehat{\tau}_1}, 0_{n-\widehat{\tau}_1}) \\ &\equiv \lambda\Gamma + \lambda(\widehat{c} - c_0)E,\end{aligned}$$

where $\Gamma = \mathrm{diag}(c_0 1_{\widehat{\tau}_1}, 1_{n-\widehat{\tau}_1})$ and $E = \mathrm{diag}(1_{\widehat{\tau}_1}, 0_{n-\widehat{\tau}_1})$.

By the Sherman-Morrison formula,

$$\begin{aligned}&(\Omega + \lambda_{\text{new}}\widehat{\Sigma})^{-1} - (\Omega + \lambda\Gamma)^{-1} \\ &= -(\widehat{c} - c_0)\lambda(\Omega + \lambda\Gamma)^{-1}E\left[I + (\widehat{c} - c_0)\lambda E(\Omega + \lambda\Gamma)^{-1}E\right]^{-1}E(\Omega + \lambda\Gamma)^{-1} \\ &\equiv -(\widehat{c} - c_0)\Delta,\end{aligned}$$

where $\Delta = \lambda(\Omega + \lambda\Gamma)^{-1}E\left[I + (\widehat{c} - c_0)\lambda E(\Omega + \lambda\Gamma)^{-1}E\right]^{-1}E(\Omega + \lambda\Gamma)^{-1}$.

Therefore, on $\mathcal{E}_n$, we have:

$$0 \le \Delta \le \frac{\lambda}{1-\varepsilon}(\Omega + \lambda\Gamma)^{-2} \le \frac{1}{1-\varepsilon}(\Omega + \lambda I)^{-1}.$$

Note that $\lambda I \le \lambda_{\text{new}}\widehat{\Sigma} \le \lambda\widehat{c}I$. Thus:

$$\begin{aligned}&(\widehat{f}_- - f_-)'(\widehat{f}_- - f_-) \\ &\le 2f_-'(\Omega + \lambda_{\text{new}}\widehat{\Sigma})^{-1}(\lambda_{\text{new}}\widehat{\Sigma})^2(\Omega + \lambda_{\text{new}}\widehat{\Sigma})^{-1}f_- \\ &\quad + 2\mathbf{e}^T D_0(\Omega + \lambda_{\text{new}}\widehat{\Sigma})^{-1}\Omega_{(\tau_0)}^2(\Omega + \lambda_{\text{new}}\widehat{\Sigma})^{-1}D_0\mathbf{e} \\ &\le 2\widehat{c}^2\lambda^2 f_-'(\Omega + \lambda I)^{-2}f_- \\ &\quad + 4\mathbf{e}^T D_0(\Omega + \lambda\Gamma)^{-1}\Omega_{(\tau_0)}^2(\Omega + \lambda\Gamma)^{-1}D_0\mathbf{e} + 4(\widehat{c} - c_0)^2\mathbf{e}^T D_0\Delta\Omega_{(\tau_0)}^2\Delta D_0\mathbf{e}.\end{aligned}$$

We will bound these three terms separately. For the first term, by the results of Wahba & Wang (2019), its bound is:

$$2\widehat{c}^2\left(E\{\widehat{f}^{(0)}\} - f_-\right)'\left(E\{\widehat{f}^{(0)}\} - f_-\right) = O_P(n\lambda).$$

For the second term, note that:

$$\begin{aligned}&E\left\{\mathbf{e}^T D_0(\Omega + \lambda\Gamma)^{-1}\Omega_{(\tau_0)}^2(\Omega + \lambda\Gamma)^{-1}D_0\mathbf{e}\right\} \\ &= \mathrm{Tr}\left(D_0(\Omega + \lambda\Gamma)^{-1}\Omega_{(\tau_0)}^2(\Omega + \lambda\Gamma)^{-1}D_0\right) \\ &\le \delta^4 \ \mathrm{Tr}\left(\Omega_{(\tau_0)}(\Omega + \lambda\Gamma)^{-2}\Omega_{(\tau_0)}\right) \\ &\le \delta^4 \ \mathrm{Tr}\left(\Omega_{(\tau_0)}(\Omega + \lambda I)^{-2}\Omega_{(\tau_0)}\right) = O(h^{-1}),\end{aligned}$$

so the second term is $O_P(h^{-1})$. For the third term, observe:

$$\begin{aligned}\mathbf{e}^T D_0\Delta\Omega_{(\tau_0)}^2\Delta D_0\mathbf{e} &\le \mathbf{e}^T\mathbf{e} \ \mathrm{Tr}\left(D_0\Delta\Omega_{(\tau_0)}^2\Delta D_0\right) \\ &\le \delta^4\mathbf{e}^T\mathbf{e} \times \frac{1}{(1-\varepsilon)^2} \ \mathrm{Tr}\left(\Omega_{(\tau_0)}(\Omega + \lambda I)^{-2}\Omega_{(\tau_0)}\right) = O_P(nh^{-1}),\end{aligned}$$

so the third term is $O_P(nh^{-1}\widetilde{r}_n^2)$. In summary, we obtain:

$$\|\widehat{f}_- - f_-\|_n^2 = O_P\left(\lambda + (nh)^{-1} + h^{-1}\widetilde{r}_n^2\right).$$

Similarly, we have

$$\begin{aligned}\|\widehat{f}_+ - f_+\|_n^2 &\le 2\widehat{c}^2\lambda^2 f_+'(\Omega + \lambda I)^{-2}f_+ + 4\mathbf{e}^T D_0(\Omega + \lambda\Gamma)^{-1}\Omega_{(*(n-\tau_0))}^2(\Omega + \lambda\Gamma)^{-1}D_0\mathbf{e} \\ &\quad + 4(\widehat{c} - c_0)^2\mathbf{e}^T D_0\Delta\Omega_{(*(n-\tau_0))}^2\Delta D_0\mathbf{e} \\ &= O_P\left(\lambda + (nh)^{-1} + h^{-1}\widetilde{r}_n^2\right).\end{aligned}$$

This completes the proof of Theorem 4.1. □

C.4 Proof of Corollary 4.1

Proof of Corollary 4.1. When $m \geq 2$, the results of Corollary 4.1 are obvious based on Theorem 4.1. When $m = 1$, from Theorem 4.1. we have

$$\frac{1}{\widehat{\tau}_1}\sum_{i=1}^{\widehat{\tau}_1}\eta_i^2 \leq \|\widehat{f}_- - f_-\|_n^2 = O_P(n^{-2/3}).$$

By Lemma C.2, we have

$$\frac{2}{\widehat{\tau}_1}\sum_{i=1}^{\widehat{\tau}_1}\epsilon_i\eta_i = O_p(n^{-1/4}h^{-3/4}/n) = O_p(n^{-1/2}).$$

equation 24 shows that

$$\frac{1}{\widehat{\tau}_1}\sum_{i=1}^{\tau_1}\epsilon_i^2 = \sigma_0^2 + O_P(n^{-1/2}).$$

Therefore, we have

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{1}{\widehat{\tau}_1}\sum_{i=1}^{\widehat{\tau}_1}\eta_i^2 + \frac{2}{\widehat{\tau}_1}\sum_{i=1}^{\widehat{\tau}_1}\epsilon_i\eta_i + \frac{1}{\widehat{\tau}_1}\sum_{i=1}^{\widehat{\tau}_1}\epsilon_i^2 \\ &= \sigma_0^2 + O_P(n^{-2/3} + n^{-1/2} + n^{-1/2}) \\ &= \sigma_0^2 + O_P(n^{-1/2}),\end{aligned}$$

Similarly, we can prove $\widehat{\delta}^2 = \delta_0^2 + O_P(n^{-1/2})$. □

C.5 Proof of Theorem 4.2

The key feature of our approach is the interaction between mean and variance change points. Thus, conventional proofs for single mean change points cannot be directly applied. We must address the new scenario: how to obtain the statistical properties of the mean change point test statistic under the influence of variance change points. We observe that we can rescale the detrended data and compensate for the shift at the variance change point to construct our dataset. Below is the detailed proof.

Proof of Theorem 4.2. When detecting the mean change point, we always use the variance change point and variance estimates from the previous iteration. That is, mean change point estimation is performed under known variance conditions. Assume a variance change point τ_1 exists such that $Var(\epsilon_i) = \sigma^2$ for $1 \leq i \leq \tau_1$ and $Var(\epsilon_i) = \delta^2$ for $\tau_1 + 1 \leq i \leq n$.

Although the presence of a variance change point prevents direct application of constant variance methods, we note that scaling the data preserves the mean change structure. Therefore, we rescale the dataset y_i in two segments: $\{x_i\}_{1\leq i\leq\tau_1}$ and $\{x_i\}_{\tau_1+1\leq i\leq n}$

$$x_i = \begin{cases}\frac{\widehat{e}_i}{\widehat{\sigma}}, & 1 \leq i \leq \tau_1, \\ \frac{\widehat{e}_i}{\widehat{\delta}} + \widehat{\kappa}, & \tau_1 + 1 \leq i \leq n.\end{cases}$$

This transformation yields $Var(x_i) = 1$ for all $1 \leq i \leq n$, converting our problem to a mean change point detection with known constant variance.
Under H_0, the likelihood is

$$L_0(\mu) = \frac{1}{(\sqrt{2\pi})^n}\exp\left(-\frac{\sum_{i=1}^n(x_i-\mu)^2}{2}\right),$$

with MLE

$$\widehat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^n x_i.$$

Under H_1,

$$L_1(\mu_1, \mu_2) = \frac{1}{(\sqrt{2\pi})^n} \exp\left(- \left[\frac{\sum_{i=1}^{k}(x_i - \mu_1)^2}{2} + \frac{\sum_{i=k+1}^{n}(x_i - \mu_2)^2}{2} \right] \right),$$

with MLEs

$$\widehat{\mu_1} = \bar{x}_k = \frac{1}{k}\sum_{i=1}^{k} x_i, \quad \widehat{\mu_2} = \bar{x}_{n-k} = \frac{1}{n-k}\sum_{i=k+1}^{n} x_i.$$

Define $S_n = \sum_{i=1}^{n}(x_i - \bar{x})^2$ and $S_k = \sum_{i=1}^{k}(x_i - \bar{x}_k)^2 + \sum_{i=k+1}^{n}(x_i - \bar{x}_{n-k})^2$. The test statistic is

$$T_n^2 = \max_{1\le k\le n-1} T_k^2 = \max_{1\le k\le n-1} [S_n - S_k].$$

Let $W_k = \sum_{i=1}^{k} x_i = k\bar{x}_k$ for $1 \le k \le n$. Through algebraic manipulation, we have

$$T_n = \max_{1\le k\le n-1} |T_k| = \max_{1\le k\le n-1} \left| \frac{W_k}{\sqrt{n}} - \frac{k}{n}\frac{W_n}{\sqrt{n}} \right| \Big/ \left[\frac{k}{n}\left(1 - \frac{k}{n}\right) \right]^{1/2}.$$

Let $\{B(t); 0 \le t < \infty\}$ be a standard Brownian motion. Under H_0, by properties of normal random variables, we have

$$\left\{ (W_k - k\mu)/\sqrt{n}; 1 \le k \le n \right\} \overset{D}{=} \left\{ B\left(\frac{k}{n}\right); 1 \le k \le n \right\},$$

where "$\overset{D}{=}$" denotes equality in distribution. Further

$$\begin{aligned}
T_n &= \max_{1\le k\le n-1} \left| \frac{W_k}{\sqrt{n}} - \frac{k}{n}\frac{W_n}{\sqrt{n}} \right| \Big/ \left[\frac{k}{n}\left(1 - \frac{k}{n}\right) \right]^{1/2} \\
&= \max_{t=k/n} \left| \frac{W_k}{\sqrt{n}} - t\frac{W_n}{\sqrt{n}} \right| \Big/ [t(1-t)]^{1/2} \\
&= \max_{t=k/n} \left| \frac{W_k - k\mu}{\sqrt{n}} - t\left(\frac{W_n - n\mu}{\sqrt{n}}\right) \right| \Big/ [t(1-t)]^{1/2} \\
&\overset{D}{=} \max_{t=k/n} |B(t) - tB(1)| \big/ [t(1-t)]^{1/2} \\
&= \max_{t=k/n} |B_0(t)| \big/ [t(1-t)]^{1/2},
\end{aligned}$$

where $t = k/n$ and $B_0(t) = B(t) - tB(1)$ is a Brownian bridge.
Before proving the theorem, we require the following lemmas.

Lemma C.3 (Chen & Gupta (2000) Lemma 2.6)**.**

$$\max_{1\le k\le [n/\log n]} \frac{|B_0(k/n)|}{\sqrt{(k/n)(1-k/n)}} - \max_{1\le k\le [n/\log n]} \frac{|B(k/n)|}{\sqrt{k/n}} = o_p(a_n).$$

Lemma C.4 (Chen & Gupta (2000) Lemma 2.7)**.**

$$\max_{[n/\log n]\le k\le [n/2]} \frac{|B_0(k/n)|}{\sqrt{(k/n)(1-k/n)}} = O_p((\log\log\log n)^{1/2}).$$

Lemma C.5 (Chen & Gupta (2000) Lemma 2.8)**.** *For any* $-\infty < x < \infty$,

$$\lim_{n\to\infty} P\left[a_n^{-1}\left(\max_{1\le k\le [n/\log n]} \frac{|B(k/n)|}{\sqrt{k/n}} - b_n \right) \le x \right] = \exp\left\{ -\pi^{1/2} e^{-x} \right\}.$$

With these preparations, the asymptotic distribution of our test statistic is

$$
\begin{aligned}
&P\left[a_n^{-1}(T_n - b_n) \le x \mid H_0\right] \\
&= P\left[T_n \le a_n x + b_n \mid H_0\right] \\
&= P\left[\max_{1\le k\le n-1} \frac{|B_0(k/n)|}{\sqrt{(k/n)(1-k/n)}} \le a_n x + b_n\right] \\
&= P\left[\max_{1\le k\le [n/2]} \frac{|B_0(k/n)|}{\sqrt{(k/n)(1-k/n)}} \le a_n x + b_n, \max_{1\le k\le [n/2]} \frac{|B_0(1-k/n)|}{\sqrt{(k/n)(1-k/n)}} \le a_n x + b_n\right] \\
&= P\left[\max_{1\le k\le [n/\log n]} \frac{|B(k/n)|}{\sqrt{k/n}} \le a_n x + b_n, \max_{1\le k\le [n/\log n]} \frac{|B(1) - B(1-k/n)|}{\sqrt{k/n}} \le a_n x + b_n\right] \\
&\quad + o_p(1) \\
&= P\left[\max_{1\le k\le [n/\log n]} \frac{|B(k/n)|}{\sqrt{k/n}} \le a_n x + b_n\right] \\
&\quad \cdot P\left[\max_{1\le k\le [n/\log n]} \frac{|B(1) - B(1-k/n)|}{\sqrt{k/n}} \le a_n x + b_n\right] + o_p(1) \\
&\xrightarrow{n\to\infty} \exp\left(-\pi^{-1/2} e^{-x}\right) \cdot \exp\left(-\pi^{-1/2} e^{-x}\right) = \exp\left(-2\pi^{-1/2} e^{-x}\right).
\end{aligned}
$$

Thus, we conclude that the asymptotic null distribution of T_n is a Gumbel distribution. □

C.6 Proof of Theorem 4.3

Similar to the mean change point scenario, the joint effect of mean and variance change points prevents direct application of traditional variance change point test statistics. This paper leverages the piecewise nature of the mean function, examining the data segments separately and expressing residuals at the mean change point to develop the proof strategy. Below is the detailed proof process.

Proof of Theorem 4.3. For the variance change point, without loss of generality, assume the existence of a mean change point τ_0. If no mean change point exists, the case reduces to that of Gao et al. (2019), where the variance of ϵ_i is normalized to 1 . Under the alternative hypothesis H_1 for the variance change point, the mean function estimators are $\widehat{f}_-^{(0)}$ and $\widehat{f}_+^{(0)}$.

Recall

$$
\begin{aligned}
\delta_i &= \widehat{f}_-^{(0)}(i/n) - E\{\widehat{f}_-^{(0)}\}(i/n), & \delta_i^0 &= E\{\widehat{f}_-^{(0)}\}(i/n) - f_-(i/n), \\
\varphi_i &= \widehat{f}_+^{(0)}(i/n) - E\{\widehat{f}_+^{(0)}\}(i/n), & \varphi_i^0 &= E\{\widehat{f}_+^{(0)}\}(i/n) - f_+(i/n).
\end{aligned}
$$

Define

$$
\begin{aligned}
\eta_i &= \delta_i + \delta_i^0 = \widehat{f}_-^{(0)}(i/n) - f_-(i/n), & 1 \le i \le \tau_0; \\
\gamma_i &= \varphi_i + \varphi_i^0 = \widehat{f}_+^{(0)}(i/n) - f_+(i/n), & \tau_0 + 1 \le i \le n.
\end{aligned}
$$

For any $1 < k \le \tau_0$, the variance estimates are

$$
\begin{aligned}
\widehat{\sigma}_n^2 &= \frac{1}{n}\left[\sum_{i=1}^{\tau_0}(\eta_i + \epsilon_i)^2 + \sum_{i=\tau_0+1}^{n}(\gamma_i + \epsilon_i)^2\right], \\
\widehat{\sigma}_k^2 &= \frac{1}{k}\sum_{i=1}^{k}(\eta_i + \epsilon_i)^2, \\
\widehat{\sigma}_{n-k}^2 &= \frac{1}{n-k}\left[\sum_{i=k+1}^{\tau_0}(\eta_i + \epsilon_i)^2 + \sum_{i=\tau_0+1}^{n}(\gamma_i + \epsilon_i)^2\right].
\end{aligned}
$$

Using Taylor expansion and the results of Lemmas C.1 and C.2, we derive

$$
\begin{aligned}
&\ell(k) - \ell(n) \\
&= n \log \widehat{\sigma}_n^2 - k \log \widehat{\sigma}_k^2 - (n-k) \log \widehat{\sigma}_{n-k}^2 \\
&= n \log \left(1 + \frac{\sum_{i=1}^n (\epsilon_i^2 - 1) + 2\sum_{i=1}^{\tau_0} \eta_i \epsilon_i + 2\sum_{i=\tau_0+1}^n \gamma_i \epsilon_i + \sum_{i=1}^{\tau_0} \eta_i^2 + \sum_{i=\tau_0+1}^n \gamma_i^2}{n} \right) \\
&\quad - k \log \left(1 + \frac{\sum_{i=1}^k (\epsilon_i^2 - 1) + 2\sum_{i=1}^k \eta_i \epsilon_i + \sum_{i=1}^k \eta_i^2}{k} \right) \\
&\quad - (n-k) \log \left(1 + \frac{\sum_{i=k+1}^n (\epsilon_i^2 - 1) + 2\sum_{i=k+1}^{\tau_0} \eta_i \epsilon_i + 2\sum_{i=\tau_0+1}^n \gamma_i \epsilon_i + \sum_{i=k+1}^{\tau_0} \eta_i^2 + \sum_{i=\tau_0+1}^n \gamma_i^2}{n-k} \right) \\
&= -\frac{1}{2n} \left(\sum_{i=1}^n (\epsilon_i^2 - 1) + 2\sum_{i=1}^{\tau_0} \eta_i \epsilon_i + 2\sum_{i=\tau_0+1}^n \gamma_i \epsilon_i + \sum_{i=1}^{\tau_0} \eta_i^2 + \sum_{i=\tau_0+1}^n \gamma_i^2 \right)^2 \\
&\quad + \frac{1}{2k} \left(\sum_{i=1}^k (\epsilon_i^2 - 1) + 2\sum_{i=1}^k \eta_i \epsilon_i + \sum_{i=1}^k \eta_i^2 \right)^2 \\
&\quad + \frac{1}{2(n-k)} \left(\sum_{i=k+1}^n (\epsilon_i^2 - 1) + 2\sum_{i=k+1}^{\tau_0} \eta_i \epsilon_i + 2\sum_{i=\tau_0+1}^n \gamma_i \epsilon_i + \sum_{i=k+1}^{\tau_0} \eta_i^2 + \sum_{i=\tau_0+1}^n \gamma_i^2 \right)^2 \\
&\quad + O_P \left(n \left[\frac{\sum_{i=1}^n (\epsilon_i^2 - 1) + \cdots}{n} \right]^3 + k \left[\frac{\sum_{i=1}^k (\epsilon_i^2 - 1) + \cdots}{k} \right]^3 + (n-k) \left[\frac{\sum_{i=k+1}^n (\epsilon_i^2 - 1) + \cdots}{n-k} \right]^3 \right) \\
&= -n \log \left(\frac{\sum_{i=1}^n \epsilon_i^2}{n} \right) + k \log \left(\frac{\sum_{i=1}^k \epsilon_i^2}{k} \right) + (n-k) \log \left(\frac{\sum_{i=k+1}^n \epsilon_i^2}{n-k} \right) \\
&\quad + O_P \left(\log n \left(n^{-1/4} h^{-3/4} + n^{-1/2} h^{-1} \log n + n^{1/2} h^{2m-1} \right) \right) \\
&= \mathrm{SIC}(k) - \mathrm{SIC}(n) + O_P(r_n),
\end{aligned}
$$

where the O_P term holds uniformly in k, and $r_n = \log n (n^{-1/4} h^{-3/4} + n^{-1/2} h^{-1} \log n + n^{1/2} h^{2m-1})$.

Thus,

$$\max_{1 \le k \le \tau_0} [\ell(k) - \ell(n)] = \max_{1 \le k \le \tau_0} [\mathrm{SIC}(k) - \mathrm{SIC}(n)] + O_P(r_n).$$

Similarly, for any $\tau_0 + 1 \le k \le n$,

$$\max_{\tau_0+1 \le k \le n} [\ell(k) - \ell(n)] = \max_{\tau_0+1 \le k \le n} [\mathrm{SIC}(k) - \mathrm{SIC}(n)] + O_P(r_n).$$

Combining both results, we have

$$\max_{1 \le k \le n} [\ell(k) - \ell(n)] = \max_{1 \le k \le n} [\mathrm{SIC}(k) - \mathrm{SIC}(n)] + O_P(r_n).$$

By the result of Chen & Gupta (1997), for any $x \in \mathbb{R}$,

$$\Pr \left(a_n (\log n)^{1/2} \lambda_{*n} - b_n \log n \le x \right) \to \exp\left(-2\exp(-x)\right),$$

where $a_n = (2 \log\log n)^{1/2} / \log n$ and

$$b_n = \left\{ 2\log\log n + \frac{1}{2} \log\log\log n - \log \Gamma\left(\frac{1}{2}\right) \right\} / \log n.$$

Since r_n satisfies $r_n \log^2 n = o(1)$, it follows that $a_n \log n (\lambda_n - \lambda_{*n}) = o_P(1)$. Therefore,

$$\Pr\left\{ a(\log n) \lambda_n - b(\log n) \le x \right\} \to \exp\left(-2e^{-x}\right),$$

where

$$a(\log n) = \frac{(2\log\log n)^{1/2}}{(\log n)^{1/2}}, \qquad b(\log n) = 2\log\log n + \frac{1}{2}\log\log\log n - \log\Gamma\left(\frac{1}{2}\right).$$

□

D Additional Data Simulations

In this appendix, we present the simulation of the other three specialized change structures: data with a co-located joint mean-variance Change-point, data with only a mean change-point (Section D.2) and data with only a variance change-point (Section D.3). The data generation mechanism, comparison methods, and evaluation metrics remain identical to those described in Section 5.1 of the main text.

D.1 Co-located Joint Mean-Variance Changes

We examined co-located mean and variance changes with the original two-sided RKHS implementation, using the same Gaussian configurations as the other supplementary runs: $\kappa = 0.75$ for the f_1 configurations and the original f_2 with $\kappa = 6$ for the f_2 configurations. Each configuration is based on 1,000 replications. Figure 6 shows one representative replication of each configuration. Because the two changes are co-located, $\tau_0 = \tau_1 = n/2$, a single dashed line marks both, and the mean shift and the drop of the error variance occur at that same point.

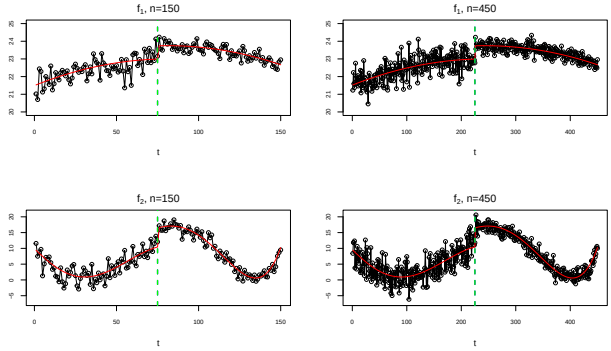

Figure 6: Representative data profiles with co-located mean and variance change-points, one panel per configuration. Red curves show the true mean functions and the dashed green lines mark the co-located change-points, $\tau_0 = \tau_1 = n/2$.

Figure 7 summarizes the localization of the two change-points. Conditional on detection, both are located essentially exactly: the median of $\widehat{\tau}_0/n$ is 0.50 in all four configurations, and the median of $\widehat{\tau}_1/n$ lies between 0.49 and 0.50, the estimates at $n = 450$ being slightly tighter than those at $n = 150$.

Figure 8 reports the estimated mean shift and the two variance levels. The shift is recovered with medians of 0.75 and 0.86 in the two f_1 configurations and of 6.0 and 6.1 in the two f_2 configurations, against the true 0.75 and 6; as in the simultaneous-change case of Section 5.1, the estimate is slightly upward biased under the smoother trend f_1 at $n = 150$. The log ratios of the estimated to true variances centre near zero for both the pre- and the post-change level, the largest deviations occurring at f_2, $n = 150$ (-0.11 and -0.13).

Figure 9 evaluates the mean function estimation. The MSE is far larger under f_2 than under f_1 (medians of 0.46 and 0.14 at $n = 150$ and 450, against 0.016 and 0.0033), which reflects the much larger variance levels of the f_2 configurations ($\sigma^2 = 8$, $\delta^2 = 2$) rather than a difference in the estimation of the trend itself: in all four configurations the fitted curves at the three MSE percentiles follow the true mean closely, separating only in the immediate neighbourhood of the change-point.

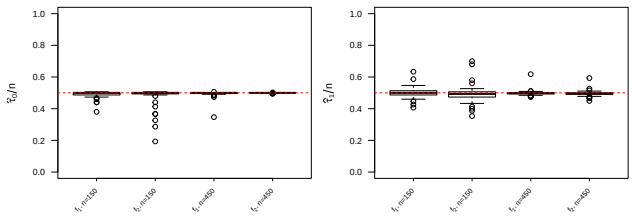

Figure 7: Boxplots of the co-located mean and variance change-point estimates, normalised by n, so that the true co-located change-point sits at 0.5. Dashed red lines mark the true locations. Replications in which the corresponding change-point is not detected are excluded. Each configuration is based on 1,000 replications.

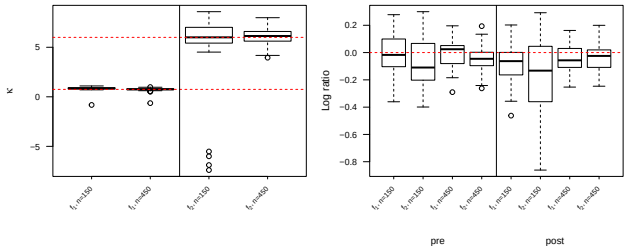

Figure 8: Boxplots of the estimated mean shift $\widehat{\kappa}$ (*left*) and of the log ratios of the estimated to true pre- and post-change variances (*right*; within it the pre-change group is on the left and the post-change group on the right). Dashed red lines mark the true values, κ for the mean shift and 0 for both log ratios. Replications in which the corresponding change-point is not detected are excluded. Each configuration is based on 1,000 replications.

D.2 Mean Change-point Only

When only the mean shifts and the variance remains constant, we assess the estimation of the mean change-point, the mean function, and the mean shift quantity. We consider the following four scenarios:

- Scenario 5: $n = 150,\ f = f_1,\ \tau_0 = 50,\ \kappa = 0.75,\ \sigma^2 = 1$;
- Scenario 6: $n = 450,\ f = f_1,\ \tau_0 = 150,\ \kappa = 0.75,\ \sigma^2 = 1$;
- Scenario 7: $n = 150,\ f = f_2,\ \tau_0 = 50,\ \kappa = 3,\ \sigma^2 = 1$;
- Scenario 8: $n = 450,\ f = f_2,\ \tau_0 = 150,\ \kappa = 3,\ \sigma^2 = 1$.

Figure 10 shows four typical realizations under the settings above.

Figure 11 summarizes the mean change-point and mean-shift estimation results under mean-only changes. The results are consistent with those in Figure 2. The proposed method provides accurate and stable estimates across Scenarios 5–8, whereas CG, MJ, and their moving-window variants are less reliable, especially under the more complex trend f_2. The estimation of κ also improves as n increases from 150 to 450. Our method yields excellent estimation results in Scenario 1-4. CG and MJ and their moving-window variants provide unreliable estimates.

Figure 12 evaluates the mean-function estimation performance of the proposed method under mean-only changes. Across Scenarios 5–8, the estimated curves track the true mean functions closely. At $n = 150$, the curves corresponding to larger MSE values show slight deviations near the change-points, reflecting the expected variability for smaller samples. In contrast, the estimates for $n = 450$ conform more closely to the true trends. The MSE distributions also shrink markedly as n increases, indicating improved mean-function estimation accuracy.

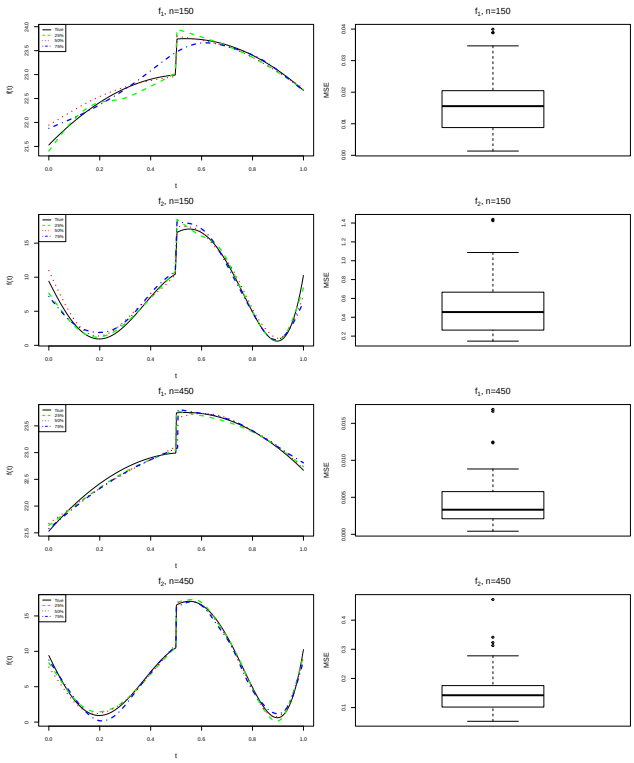

Figure 9: Mean function estimation for the co-located configurations. For each configuration the left panel shows the true mean function and the estimated curves corresponding to the 25th, 50th and 75th percentiles of the MSE, and the right panel the corresponding MSE distribution. Each configuration is based on 1,000 replications.

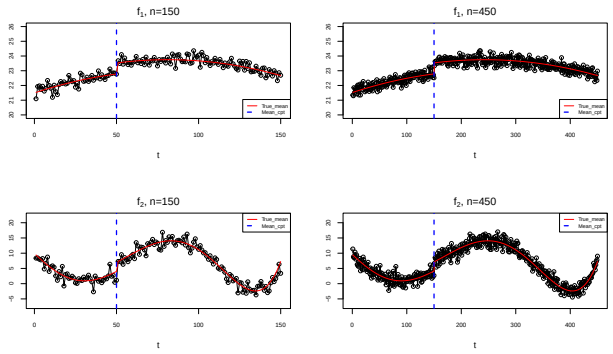

Figure 10: Simulated data profiles under Scenarios 5–8 (mean change only). Red curves show the true mean functions, and blue dashed lines mark the true mean change-point, $\tau_0/n = 1/3$.

Let the signal-to-noise ratio (SNR) κ/σ vary in the range $[-1.5, 1.5]$ with step size 0.1. Figure 13 presents the empirical power of the mean change-point detection procedure against SNR. The detection power rises monotonically with $|\kappa|/\sigma$ and is symmetric about 0, indicating that upward and downward shifts of the same magnitude are equally detectable. For any fixed nonzero SNR, the $n = 450$ panel achieves uniformly higher power than $n = 150$, and the curves approach one more rapidly as $|\kappa|/\sigma$ increases.

The conservative nature of our method is evident from Table 1, where the empirical rejection rates under H_0 are below the nominal level. This behavior stems from the slow finite-sample convergence

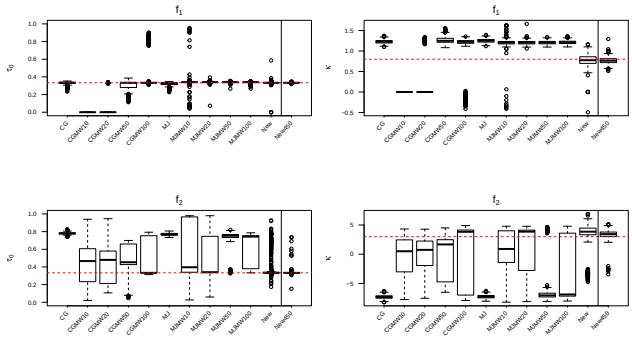

Figure 11: Boxplots of mean change-point estimates and mean-shift estimates under Scenarios 5–8 (mean change only). *Left column:* mean change-point estimates; *right column:* mean-shift estimates. Red dashed lines indicate the true values, with $\tau_0/n = 1/3$ for the mean change-point and the corresponding true κ for the mean shift. *Top row:* $f = f_1$; *bottom row:* $f = f_2$.

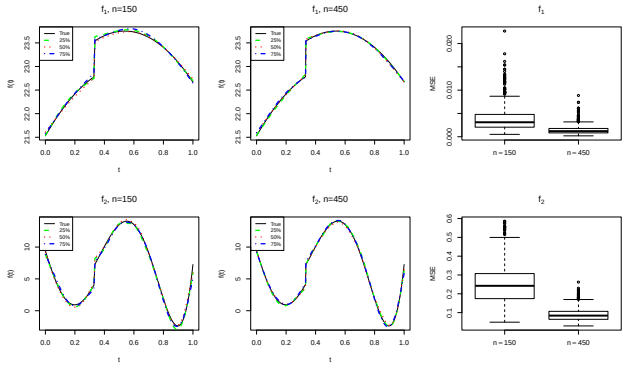

Figure 12: Mean function estimation performance of the proposed method under Scenarios 5–8 (mean change only). *Left and middle columns:* true mean functions and estimated curves corresponding to the 25th, 50th, and 75th percentiles of the MSE over 1,000 replications for $n = 150$ and $n = 450$, respectively. *Right column:* boxplots of the MSE distributions for the two sample sizes. *Top row:* $f = f_1$; *bottom row:* $f = f_2$.

due to the heavy-tailed asymptotic distribution. As the sample size n grows, the empirical size moves toward the nominal level, indicating the method's calibration improves with larger samples.

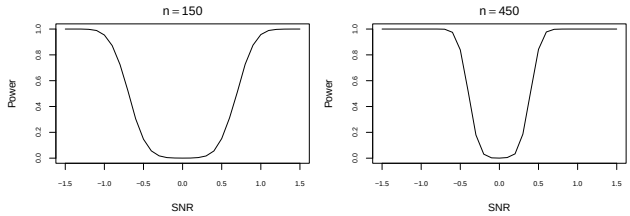

Figure 13: Detection power against signal-to-noise ratio (mean change only). *Left*: $n = 150$. *Right*: $n = 450$.

Table 1: Type I error validation in testing the mean change-point. The true mean function is f_1.

Test Level α	Sample Size n			
	150	450	2000	10000
0.05	0.0005	0.007	0.0010	0.0017
0.10	0.0023	0.0041	0.0058	0.0062

D.3 Variance Change-point Only

When only the variance shifts and the mean remains constant, we assess the estimation of the variance change-point, the mean function, and the variances before and after the change-point. We consider the following four scenarios:

- Scenario 9: $n = 150,\ f = f_1,\ \tau_1 = 100,\ \sigma^2 = 0.24,\ \delta^2 = 0.06$;
- Scenario 10: $n = 450,\ f = f_1,\ \tau_1 = 300,\ \sigma^2 = 0.24,\ \delta^2 = 0.06$;
- Scenario 11: $n = 150,\ f = f_2,\ \tau_1 = 100,\ \sigma^2 = 8,\ \delta^2 = 2$;
- Scenario 12: $n = 450,\ f = f_2,\ \tau_1 = 300,\ \sigma^2 = 8,\ \delta^2 = 2$.

Figure 14 shows four typical realizations under the settings above.

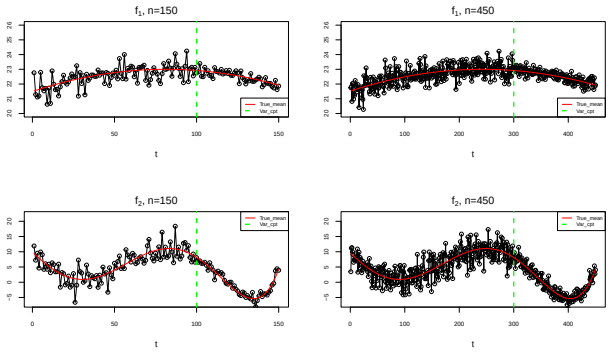

Figure 14: Simulated data profiles under Scenarios 9–12 (variance change only). Red curves show the true mean functions, and green dashed lines mark the true variance change-point, $\tau_1/n = 2/3$.

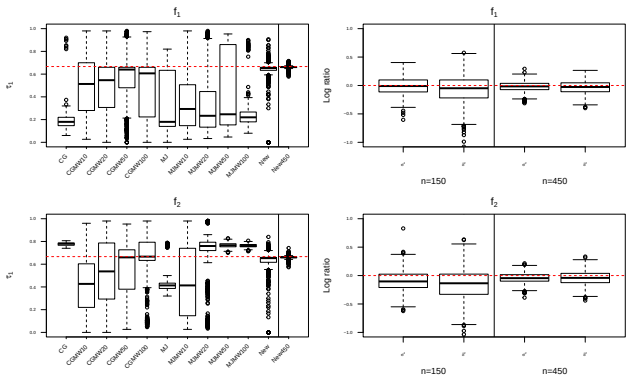

Figure 15: Boxplots of variance change-point estimates for all methods and variance-level estimates for the proposed method under Scenarios 9–12 (variance change only). *Left column:* variance change-point estimates; *right column:* variance-level estimates. Red dashed lines mark the true variance change-point, $\tau_1/n = 2/3$. For the proposed method, variance estimates are reported as log ratios of the estimated to true values for σ^2 and δ^2. *Top row:* $f = f_1$; *bottom row:* $f = f_2$.

Figure 15 summarizes the variance change-point and variance-level estimation results under variance-only changes. The proposed method provides accurate and stable estimates under both f_1 and f_2. In contrast, CG, MJ, and their moving-window variants are affected by the unknown smooth mean trend, resulting in noticeable bias in change-point localization. For the proposed method, the pre-/post-change variance log-ratios cluster around zero with relatively small spread, indicating accurate variance estimation. The results further improve as the sample size increases from 150 to 450.

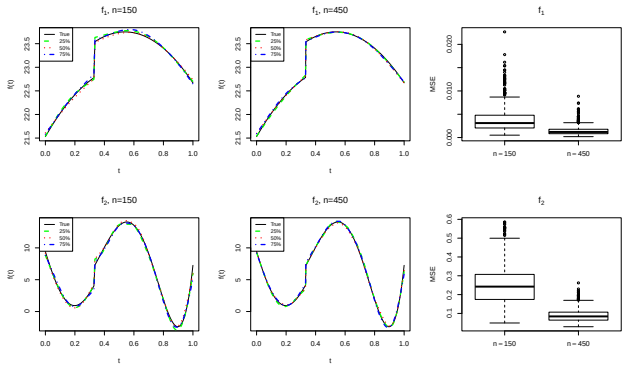

Figure 16: Mean function estimation performance of the proposed method under Scenarios 9–12 (variance change only). *Left and middle columns:* true mean functions and estimated curves corresponding to the 25th, 50th, and 75th percentiles of the MSE over 1,000 replications for $n = 150$ and $n = 450$, respectively. *Right column:* boxplots of the MSE distributions for the two sample sizes. *Top row:* $f = f_1$; *bottom row:* $f = f_2$.

Figure 16 evaluates the mean-function estimation performance of the proposed method under variance-only changes. Across Scenarios 9–12, the estimated curves track the true mean functions closely. At $n = 150$, the curves corresponding to larger MSE values show slight deviations near the change-points, reflecting the expected variability for smaller samples. In contrast, the estimates for $n = 450$ conform more closely to the true trends. The MSE distributions also shrink markedly as n increases, indicating improved mean-function estimation accuracy.

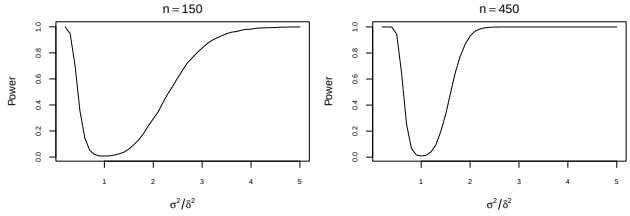

Figure 17: Detection power versus ratio of variances θ (variance change only). *Left*: $n = 150$. *Right*: $n = 450$.

Let $\theta = \sigma^2/\delta^2$ be the ratio of the variances in the range of $[0.1, 5]$ with step size 0.1. Figure 17 displays the empirical power of the variance change-point detection procedure against θ. Power increases as θ moves away from 1 , yielding a U -shaped pattern with a troughm near $\theta = 1$ (no change). Across the range, the $n = 450$ curves lie above the $n = 150$ curves, reaching high power at less extreme ratios.

Table 2 summarizes the empirical sizes of the test when no variance change is present ($\theta = \sigma^2/\delta^2 = 1$). Similar to the mean-change test, the empirical size lies below the nominal level, indicating that our method is conservative. This slow convergence of the finite-sample distribution is due to the heavy-tailed, extreme-value nature of the asymptotic distribution under H_0. As the sample size n increases, the empirical size approaches the nominal level.

Table 2: Type I error validation in testing the variance change-point.

Significance Level α	Sample Size n			
	150	450	2000	10000
0.05	0.0072	0.0097	0.0118	0.0137
0.10	0.0242	0.0305	0.0384	0.0435

D.4 Heavy-Tailed Errors Scenario

To assess the robustness of the proposed procedure when the Gaussian error assumption is violated, we repeat the separated change-point setting of Scenario 1 with heavy-tailed errors. The errors are drawn from a scaled Student-t distribution, $\epsilon_i = \sqrt{(\nu-2)/\nu}\, t_{\nu,i}$ with $t_{\nu,i}$ i.i.d. Student-t with ν degrees of freedom, so that $\mathrm{Var}(\epsilon_i) = 1$ for every ν considered; the observations are generated with $n = 150$, $f = f_1$, $\tau_0 = 50$, $\kappa = 0.75$, $\tau_1 = 100$, $\sigma^2 = 0.24$ and $\delta^2 = 0.06$. We consider $\nu \in \{5, 10, 20, 50\}$, where smaller values correspond to heavier tails and $\nu \to \infty$ recovers the Gaussian case. Each setting is repeated 1,000 times.

Figure 18 shows one representative realization for each ν. Even at $\nu = 5$, where the noise exhibits pronounced spikes, the mean change-point and the variance change-point remain separated at $\tau_0/n = 1/3$ and $\tau_1/n = 2/3$, so that the two types of change can still be attributed to different locations.

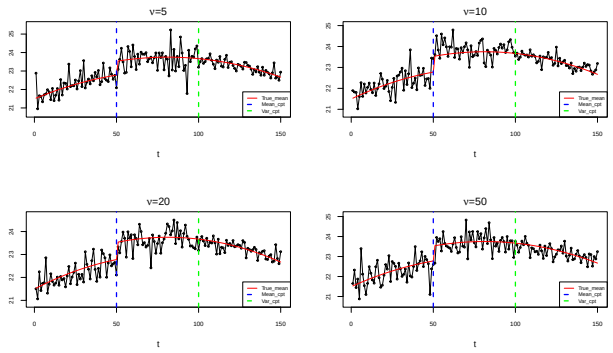

Figure 18: Simulated data profiles under heavy-tailed noise, one panel per degree of freedom ν ($n = 150$, separated mean and variance change-points). Red curves show the true mean function f_1; blue and green dashed lines mark the true mean change-point ($\tau_0/n = 1/3$) and variance change-point ($\tau_1/n = 2/3$), respectively.

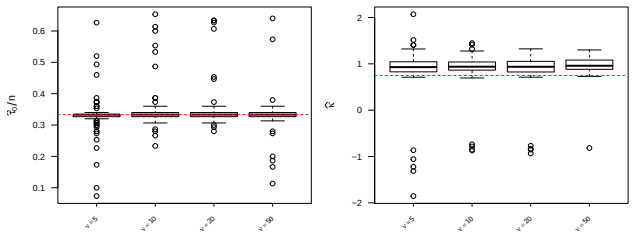

Figure 19: Boxplots of the mean change-point fraction $\widehat{\tau}_0/n$ (*left*) and the mean-shift estimate $\widehat{\kappa}$ (*right*) under heavy-tailed noise, one box per degree of freedom ν with $n = 150$. Red dashed lines mark the true values, $1/3$ for the mean change-point and 0.75 for the mean shift. Replications in which no mean change-point is detected are excluded.

Figure 19 summarizes the mean change-point and mean-shift estimates for the four tail indices; replications in which no mean change-point is declared are excluded, so the panels show the localization conditional on detection. When a change-point is detected it is localized essentially exactly at every tail index: the median of $\widehat{\tau}_0/n$ equals $1/3$ throughout and its interquartile range stays between 0.007 and 0.013. The mean-shift estimate behaves differently: $\widehat{\kappa}$ is biased upward at every ν, with medians between 0.93 ($\nu = 5$) and 0.96 ($\nu = 50$) against the true 0.75, and the bias is of the same order at $\nu = 50$ as at $\nu = 5$, so it is not induced by the tail index. The detection rate itself varies between 0.75 ($\nu = 20$) and 0.89 ($\nu = 50$) with no monotone trend in ν, a spread consistent with the Monte Carlo error of 100 replications.

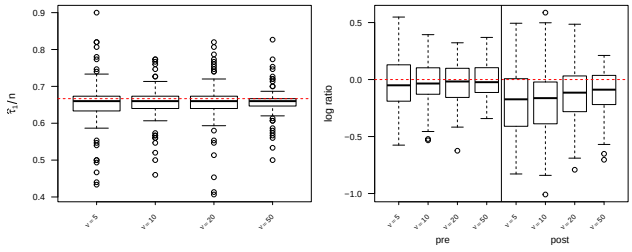

Figure 20: Boxplots of the variance change-point fraction $\widehat{\tau}_1/n$ (*left*) and of the log ratios of the estimated to true pre- and post-change variances (*right*; within it the pre-change group is on the left and the post-change group on the right) under heavy-tailed noise, one box per degree of freedom ν with $n = 150$. Red dashed lines mark the true values, $2/3$ for the variance change-point and 0 for both log ratios. Replications in which no variance change-point is detected are excluded.

Figure 20 reports the variance change-point and variance-level estimates, again restricted to the replications in which the variance change-point is detected. The change-point fraction is stable, with medians of 0.66 against the true $2/3$ and interquartile ranges between 0.020 and 0.040. Of the two variance levels, the pre-change one stays centred, the median of $\log(\widehat{\sigma}^2/\sigma^2)$ lying between -0.05 and -0.02 at every ν. The post-change level is shifted downwards at every ν: the median of $\log(\widehat{\delta}^2/\delta^2)$ lies between -0.17 ($\nu = 5$) and -0.09 ($\nu = 50$), and its interquartile range widens from 0.25 at $\nu = 50$ to 0.42 at $\nu = 5$. In the Gaussian setting of Section 5.1 both log ratios are centred around zero at $n = 150$, so this downwards shift of $\widehat{\delta}^2$ is specific to the heavy-tailed errors rather than a general small-sample property of the estimator. The variance change-point is detected in 0.89–0.98 of the replications.

Figure 21 evaluates the mean function estimation under heavy-tailed noise. Its accuracy is almost invariant to the tail index: the MSE quartiles are 0.0099, 0.0150 and 0.0223 at $\nu = 5$ against 0.0094, 0.0141 and 0.0215 at $\nu = 50$, a shift of less than 6% in the median, as expected because the Student-t errors are scaled to unit variance. What changes with ν is the shape of the individual fits: at $\nu = 5$ the fitted curves at the 25th, 50th and 75th MSE percentiles separate visibly in the neighbourhood of the mean change-point, whereas for $\nu \geq 20$ they follow the true mean almost indistinguishably.

The estimation study above keeps the change-point structure of Scenario 1 and varies only the error distribution. To separate the effect of the tail index on the two tests from its effect on the estimators, we complement it with a calibration experiment in which no change point is present: the mean and the variance are both constant, the errors follow the scaled Student-t law introduced above, and only ν varies. For each $\nu \in \{5, 10, 20, 50\}$ and each $n \in \{150, 450, 2000\}$, 1,000 null series are generated and the two tests are applied at the nominal level $\alpha = 0.05$.

Table 3 reports the empirical rejection rates of the two tests when no change is present. The mean change-point test is conservative at every tail index and sample size, with rejection rates between 0.000 and 0.002 against the nominal 0.05, so heavier tails do not make it anti-conservative. The variance change-point test behaves differently. Its rejection rate is at or below the nominal level for $\nu \geq 20$ (0.010–0.031), but it exceeds the nominal level once the tails are heavy: 0.139 at $\nu = 5$ and $n = 150$, rising to 0.436 at $\nu = 5$ and $n = 2000$, so that here the discrepancy grows with the sample size instead of shrinking towards the nominal level. The ordering of the rejection rates follows the kurtosis of the scaled t errors, which is 9.00, 4.00, 3.38 and 3.13 for $\nu = 5, 10, 20, 50$ against 3 for

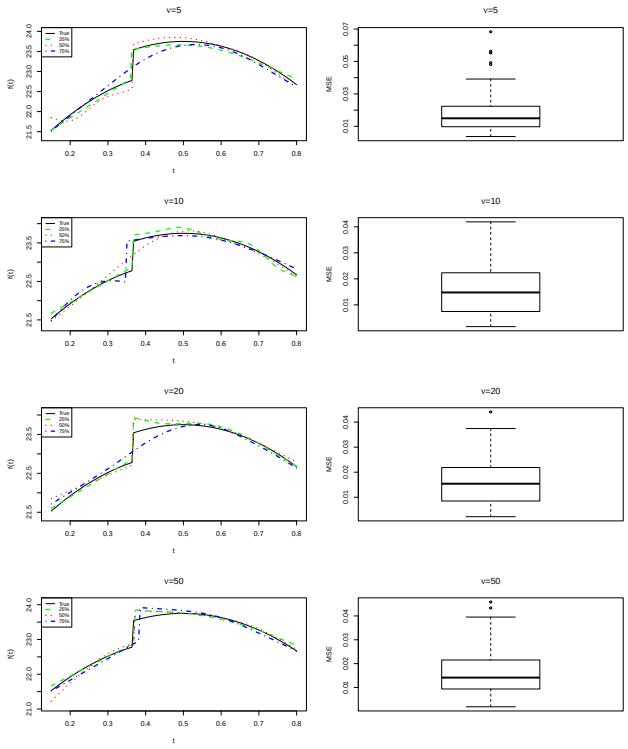

Figure 21: Mean function estimation performance of the proposed method under heavy-tailed noise. For each ν, the left panel shows the true mean function and the estimated curves corresponding to the 25th, 50th, and 75th percentiles of the MSE over 1,000 replications, and the right panel shows the corresponding MSE distribution ($n = 150$).

Table 3: Type I error validation under heavy-tailed noise. No change point is present: the mean is constant and the noise variance is constant, with errors $\sqrt{(\nu-2)/\nu}\, t_\nu$. The nominal level is $\alpha = 0.05$, and each cell is based on 1,000 replications.

ν	Mean change-point test			Variance change-point test		
	$n = 150$	$n = 450$	$n = 2000$	$n = 150$	$n = 450$	$n = 2000$
5	0.000	0.002	0.002	0.139	0.256	0.436
10	0.001	0.001	0.000	0.038	0.065	0.075
20	0.001	0.002	0.002	0.012	0.016	0.031
50	0.000	0.002	0.000	0.010	0.016	0.011

Gaussian errors; the variance test statistic is built from squared residuals, whose sampling variability is governed by the fourth moment, so the degradation is a property of the heavy-tailed errors rather than of the sample size.