Statistica Sinica Preprint No: SS-2025-0327	
Title	A Data-Adaptive Integrated Approach to Covariance Change Point Detection in High-dimensional Settings
Manuscript ID	SS-2025-0327
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0327
Complete List of Authors	Canhuang Xu, Lei Shu, Yu Chen and Qing Yang
Corresponding Authors	Yu Chen
E-mails	cyu@ustc.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

A Data-Adaptive Integrated Approach to Covariance Change Point Detection in High-Dimensional Settings

Canhuang Xu[1], Lei Shu[2], Yu Chen[3]* and Qing Yang[1]

[1]*School of Management, University of Science and Technology of China*

[2]*School of Big Data and Statistics, Anhui University*

[3]*School of Public Policy & Management, University of Science and Technology of China*

Abstract: This paper addresses the challenge of detecting the change point in the covariance matrix of a high-dimensional random vector sequence. A novel reweighted CUSUM-type statistic is introduced, incorporating a data-adaptive parameter selection method to optimize weight determination. Building on this statistic, we develop a comprehensive framework for change point detection. Additionally, a hypothesis testing procedure is proposed to assess the existence of the change point based on our methodology. The study provides rigorous theoretical foundations for the proposed method, demonstrating the validity of parameter selection and the consistency of change point estimation. The effectiveness of the method is substantiated through extensive simulation studies and real-world data analysis, confirming its practical applicability and statistical reliability.

Key words and phrases: Covariance change point detection; Data-adaptive; High-dimensional data; Reweighting.

*Corresponding author: cyu@ustc.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

1. Introduction

The change point problem is a prominent topic in statistics, originating in the domain of quality inspection and control, with its modeling and analysis pioneered by Page (1954). Nowadays, the change point problem extends far beyond its initial applications in industrial quality control and has been studied across various branches of science and engineering. It encompasses numerous application scenarios, including climate monitoring (Fischer et al., 2012), econometrics and finance (Lai and Xing, 2013), and biomedical time series (Lee et al., 2012), among others. In general, the study of change points involves detecting structural breaks in a sequence of observed data, where the locations of these breaks are termed change points.

Over recent decades, the change point problem has expanded far beyond industrial quality control to diverse applications in climate, finance, and genomics (Liu et al., 2022). The cumulative sum (CUSUM) statistic is a classical and widely used tool for detecting mean changes. In low-dimensional settings, CUSUM-based methods have been thoroughly investigated and successfully applied (Hawkins, 1977; Horváth et al., 1999). While high-dimensional mean change point detection has matured, with established L_∞-norm paradigms for sparse signals (Jirak, 2015; Yu and Chen, 2021), L_2-norm approaches for dense signals (Zhang et al., 2010; Wang et al., 2022), the double CUSUM method for reducing data to univariate processes (Cho, 2016), and projection-based techniques (Wang and Samworth, 2018). Beyond parametric approaches, nonparametric and data-driven methods have also emerged, for example, see Shu et al. (2022) and Li et al. (2024). Collectively, these contributions have led to a mature body of work on mean change point detection in high dimensions.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

Subsequent research has broadened the scope of change point analysis to diverse applications, including structural breaks in factor loadings (Ma and Su, 2018; Bai et al., 2020), regression coefficients in linear models (Lee et al., 2015; Cui et al., 2024), and network structures (Wang et al., 2021a; Chen et al., 2024). Among these, detecting changes in covariance matrices stands out as a particularly challenging and increasingly important problem in high-dimensional settings, owing to the intrinsic structural properties of covariances, and forms the primary focus of this paper. Early research on covariance change point detection focused primarily on fixed-dimensional settings or linear processes. For instance, Berkes et al. (2009) studied changes in the mean or covariance structure of linear processes using weighted CUSUM statistics. In fixed-dimensional settings, Aue et al. (2009) applied CUSUM-based methods to test for covariance changes in multivariate time series, and Kao et al. (2018) further proposed a normalized CUSUM-type statistic for testing the stability of covariance matrices and related eigensystems.

Recent advances have extended these efforts to high-dimensional regimes. For instance, Steland (2020) proposed maximally selected weighted CUSUM statistics for detecting changes in high-dimensional linear time series. Wang et al. (2021b) addressed the problem of multiple covariance change points in sequences of high-dimensional sub-Gaussian vectors, employing binary and wild binary segmentation procedures. A significant advance was made by Dette et al. (2022), who built upon the framework of Aue et al. (2009) by introducing a two-step nonparametric procedure that enhances estimation accuracy without imposing normality or sparsity assumptions. Similarly, Kaul et al. (2023) developed an approach for estimating a single covariance change point by minimizing an aggregated

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

squared loss. Additional contributions include the work of Li and Li (2023), who investigated online detection in high-dimensional M-dependent sequences; Pilliat et al. (2023), who proposed a general multi-scale bottom-up aggregation framework applicable to both mean and covariance change point problems; and Dörnemann and Dette (2024), who introduced a likelihood-ratio-based min-type statistic for covariance change point detection. Collectively, these contributions have substantially broadened the toolkit for high-dimensional covariance analysis.

However, a prevalent strategy in existing literature involves simply vectorizing the covariance matrices and applying vector-based change point methods. This transformation strategy tends to lose the key structural information of the covariance matrix, resulting in reduced performance; moreover, it also faces challenges in constructing appropriate statistics when the change pattern is unknown a priori. Motivated by these challenges and opportunities, we develop an integrated and fully data-adaptive framework for offline localization and inference of a single high-dimensional covariance change point. The key innovation lies in a reweighted CUSUM-type statistic adaptively tuning weights to the observed data, thereby avoiding naive vectorization and fixed-weight constructions.

Our main contributions are as follows. First, we propose a generalized reweighted statistic for precise change point localization. This statistic exploits a flexible decomposition that seamlessly accommodates both sparse and dense changes. Second, we develop fully data-driven procedures for parameter selection, including threshold-based scale tuning and rank-based sign alignment, that automatically focus statistical power on informative entries. Building on these advances, we introduce a companion hypothesis testing procedure to assess

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

the existence of a change point, resulting in a unified framework for both estimation and inference. Extensive simulation studies and real-data applications demonstrate the superior performance of our approach across a wide range of change patterns.

The article is structured as follows. Section 2 presents the model and methodology. Section 3 states the assumptions and main theoretical results. Sections 4 and 5 detail simulation studies and real data applications, respectively, including testing performance and comparisons. Section 6 concludes. All proofs are in the supplement.

Notations. For a symmetric matrix $\boldsymbol{H} = (H_{ab})_{a,b=1,\ldots,p} \in \mathbb{R}^{p\times p}$, $\mathrm{vech}(\boldsymbol{H})$ denotes the half-vectorization, a $p(p+1)/2$ vector formed by vectorizing the lower triangular part of $\boldsymbol{H}$. We denote by $\mathrm{diag}(\boldsymbol{A}, \boldsymbol{B})$ a block-diagonal matrix composed of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of appropriate dimension. For a vector $\boldsymbol{a} \in \mathbb{R}^p$, $\|\boldsymbol{a}\|_2 = \sqrt{\boldsymbol{a}^\top \boldsymbol{a}}$ represents the ℓ_2-norm of $\boldsymbol{a}$. The Hadamard product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is denoted by $\boldsymbol{a} \circ \boldsymbol{b}$, while $\mathrm{diag}(\boldsymbol{a})$ denotes the diagonal matrix with $\boldsymbol{a}$ as its diagonal entries. Additionally, $\sqrt{\boldsymbol{a}} = (\sqrt{a_1}, \ldots, \sqrt{a_p})^\top$ represents the element-wise square root of $\boldsymbol{a}$. For a real number c, $\lfloor c \rfloor$ denotes its integer part, and $(c)_2 = c(c-1)$. The $p \times p$ identity matrix is denoted by $\mathbf{I}_p$, while $\mathbf{0}_{p\times p}$, $\mathbf{0}_p$, and $\mathbf{1}_p$ denote the $p \times p$ zero matrix, p-dimensional zero vector, and p-dimensional all-ones vector, respectively.

2. Methodology

This paper focuses on detecting a single change point in the covariance structure of a sequence of high-dimensional random vectors, occurring at an unknown time point. We model the observations as independent p-dimensional random vectors with distinct covariance ma-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

trices before and after the change point. Specifically, let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{k_0}, \boldsymbol{X}_{k_0+1}, \ldots, \boldsymbol{X}_n$ be independent p-dimensional random vectors, where the i-th observation follows

$$\boldsymbol{X}_i = \begin{cases} \boldsymbol{\mu} + \boldsymbol{Z}_{1i}, & i = 1, \ldots, k_0 \\ \boldsymbol{\mu} + \boldsymbol{Z}_{2i}, & i = k_0 + 1, \ldots, n \end{cases}$$

with $\left(\boldsymbol{Z}_{1i}\right)_{i=1}^{k_0} \overset{i.i.d.}{\sim} \boldsymbol{Z}_1$ and $\left(\boldsymbol{Z}_{2i}\right)_{i=k_0+1}^{n} \overset{i.i.d.}{\sim} \boldsymbol{Z}_2$ satisfying $\mathbb{E}\boldsymbol{Z}_1 = \mathbb{E}\boldsymbol{Z}_2 = \boldsymbol{0}$,

$$\mathrm{Var}(\boldsymbol{Z}_1) = \boldsymbol{\Sigma}_1 = \left(\Sigma_{1,ab}\right)_{a,b=1,\ldots,p} \in \mathbb{R}^{p\times p}, \quad \mathrm{Var}(\boldsymbol{Z}_2) = \boldsymbol{\Sigma}_2 = \left(\Sigma_{2,ab}\right)_{a,b=1,\ldots,p} \in \mathbb{R}^{p\times p},$$

and $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$. Here, $\Sigma_{\nu,ab}$ denotes the (a,b)-entry of $\boldsymbol{\Sigma}_\nu$ for $\nu = 1, 2$. The unknown location k_0 is the true change point in the covariance matrices. Our primary objective is to estimate k_0 in high-dimensional settings. In the asymptotic analysis, we assume $k_0 = \lfloor r_0 n \rfloor$ for some fixed $r_0 \in (0,1)$, where $\lfloor \cdot \rfloor$ is the floor function. In the following, we propose a data-adaptive methodology for detecting the covariance change point k_0 based on the observed data.

2.1 Change point detection

Building on the model introduced above, we now describe our data-adaptive methodology for detecting the covariance change point k_0. To isolate the covariance structure, we center the observations using the global sample mean $\dot{\boldsymbol{X}}_i = \boldsymbol{X}_i - \bar{\boldsymbol{X}}$, where $\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$. For each observation, define the vectorized centered second-moment estimator $\dot{\boldsymbol{\sigma}}_{(i)} = \mathrm{vech}(\dot{\boldsymbol{X}}_i \dot{\boldsymbol{X}}_i^\top)$, where $\mathrm{vech}(\dot{\boldsymbol{X}}_i \dot{\boldsymbol{X}}_i^\top)$ denotes the half-vectorization that stacks the lower triangular elements (including the diagonal) of a symmetric matrix into a vector of length $p(p+1)/2$. The indexing mapping $h = (b-1)(p - b/2) + a$ establishes a bijection between pairs (a,b) with $1 \leq b \leq a \leq p$ and indices $h = 1, \ldots, p(p+1)/2$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

To estimate the change point k_0, we propose the reweighted CUSUM-type statistic

$$T_{\boldsymbol{W}}(k) = \frac{1}{(k)_2(n-k)_2}\sum_{(i\neq t)=1}^{k}\sum^{k}\sum_{(j\neq l)=k+1}^{n}\sum^{n}\big(\dot{\boldsymbol{\sigma}}_{(i)}-\dot{\boldsymbol{\sigma}}_{(j)}\big)^{\top}\boldsymbol{W}\big(\dot{\boldsymbol{\sigma}}_{(t)}-\dot{\boldsymbol{\sigma}}_{(l)}\big), \tag{2.1}$$

where $(k)_2 = k(k-1)$. The weight matrix $\boldsymbol{W} \in \mathbb{R}^{p(p+1)/2\times p(p+1)/2}$ assigns different importance to individual covariance entries. The exclusion of diagonal terms ($i = t$ and $j = l$) in the summation removes the bias introduced by the squared noise terms, ensuring that the statistic centers on the signal difference. Intuitively, placing larger weights on entries where the pre- and post-change covariances $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ differ most substantially increases detection sensitivity. In what follows, we develop a data-adaptive procedure for selecting $\boldsymbol{W}$ to optimize performance. This enables $T_{\boldsymbol{W}}(k)$ to effectively capture structural changes, allowing for precise identification of the change point in the covariance matrices.

The effectiveness of $T_{\boldsymbol{W}}(k)$ hinges on the choice of $\boldsymbol{W}$. In high-dimensional settings, a common modeling strategy represents structured matrices as the sum of a low-rank component and a diagonal component (Johnstone, 2001; Cai et al., 2015; Wu et al., 2020). Following this idea, we parameterize

$$\boldsymbol{W} = \big(\boldsymbol{W}^{1/2}\big)^{\top}\boldsymbol{W}^{1/2} = \boldsymbol{\Gamma} + \boldsymbol{\delta}\boldsymbol{\delta}^{\top} \in \mathbb{R}^{\frac{p(p+1)}{2}\times\frac{p(p+1)}{2}},$$

where $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$ is a nonnegative diagonal matrix with entries $\boldsymbol{\gamma} = (\gamma_h)_{1\le h\le \frac{p(p+1)}{2}}$ and $\boldsymbol{\delta} = (\delta_h)_{1\le h\le \frac{p(p+1)}{2}}$ is a $\frac{p(p+1)}{2}$-dimensional vector. This decomposition allows the statistic to be expressed as the sum of two interpretable components:

$$\begin{aligned} T_{\boldsymbol{W}}(k) = T_1(k) + T_2(k) =& \frac{1}{(k)_2(n-k)_2}\sum_{(i\neq t)=1}^{k}\sum^{k}\sum_{(j\neq l)=k+1}^{n}\sum^{n}\big(\dot{\boldsymbol{\sigma}}_{(i)}-\dot{\boldsymbol{\sigma}}_{(j)}\big)^{\top}\boldsymbol{\Gamma}\big(\dot{\boldsymbol{\sigma}}_{(t)}-\dot{\boldsymbol{\sigma}}_{(l)}\big) \\ &+ \frac{1}{(k)_2(n-k)_2}\sum_{(i\neq t)=1}^{k}\sum^{k}\sum_{(j\neq l)=k+1}^{n}\sum^{n}\big\{\boldsymbol{\delta}^{\top}\big(\dot{\boldsymbol{\sigma}}_{(i)}-\dot{\boldsymbol{\sigma}}_{(j)}\big)\big\}\big\{\boldsymbol{\delta}^{\top}\big(\dot{\boldsymbol{\sigma}}_{(t)}-\dot{\boldsymbol{\sigma}}_{(l)}\big)\big\}. \end{aligned} \tag{2.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

We observe that the statistic $T_1(k)$ is a diagonally weighted CUSUM statistic commonly used in change point analysis (Aue et al., 2009), and the statistic $T_2(k)$ resembles a burden-type statistic frequently employed in sparse settings (Lee et al., 2014).

The statistic $T_{\boldsymbol{W}}(k)$ quantifies the weighted discrepancy between the empirical covariance structures of the observations before and after the candidate location k. The change point estimation problem thus reduces to finding the position k that maximizes this discrepancy. With the true change point at $k_0 = \lfloor r_0 n \rfloor$ for some fixed $r_0 \in (0, 1)$, we define the estimator as

$$\hat{k}_0 = \underset{c_1 n \le k \le c_2 n}{\arg\max}\ T_{\boldsymbol{W}}(k). \tag{2.3}$$

The boundary trimming constants c_1 and c_2 are fixed prespecified values satisfying $0 < c_1 < c_2 < 1$, chosen to ensure sufficient observations in both pre- and post-candidate segments (in our simulations and real data analysis, we set $c_1 = 0.1$ and $c_2 = 0.9$). As established in Theorem 1, under suitable regularity conditions and appropriate parameter choices, $\hat{k}_0$ is consistent for k_0 as the sample size n and dimension p tend to infinity.

Remark 1. The statistic $T_{\boldsymbol{W}}(k)$ shares conceptual similarities with those proposed by Shu et al. (2022) and Jiang et al. (2024), which use random integration for two-sample mean testing and mean change point detection, respectively. Prior works suggest various fixed choices for $\boldsymbol{\delta}$ and $\boldsymbol{\Gamma}$ when constructing the weight matrix $\boldsymbol{W}$ for $T_{\boldsymbol{W}}(k)$ (e.g., $\boldsymbol{\delta} = \mathbf{0}$ with $\boldsymbol{\Gamma}$ a positive definite diagonal matrix), but they do not provide explicit data-adaptive selection procedures. In the next section, we introduce such a scheme and establish its theoretical properties.

Remark 2. Boundary trimming is a standard practice in the change point literature to

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

ensure adequate sample sizes on both sides of each candidate location, thereby mitigating edge effects and potential boundary bias. Choosing (c_1, c_2) entails trading off maximal data utilization against robustness near the sample endpoints. For example, Wang and Feng (2023) set $(c_1, c_2) = (0.2, 0.8)$. In a related high-dimensional setting, Dörnemann and Dette (2024) employs a symmetric search region with $c_2 = 1 - c_1$, recommending $c_1 > \max\{p/n + 0.05, 0.2\}$ for hypothesis testing to ensure stable finite-sample behavior, while choosing c_1 closer to the critical boundary p/n when the primary goal is change point localization. We adopt the setting $(c_1, c_2) = (0.1, 0.9)$, a design that imposes no additional constraints on n and p and preserves more observations while effectively controlling boundary interference.

2.2 Data-adaptive weight selection

The performance of the statistic $T_{\boldsymbol{W}}(k)$ in (2.2) depends critically on the choice of parameters $\boldsymbol{\Gamma}$ and $\boldsymbol{\delta}$. Appropriate selection of these weights enhances detection power by emphasizing informative covariance entries while suppressing noise and weak signals.

Since $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, there exist entries where $\Sigma_{1,ab} \neq \Sigma_{2,ab}$. To formalize this, define the index sets for the lower triangular elements (including the diagonal): $\mathcal{N} = \big\{(a, b) : 1 \leq b \leq a \leq p;\ \Sigma_{1,ab} = \Sigma_{2,ab}\big\}$ as the positions with no change, and $\mathcal{G} = \big\{(a, b) : 1 \leq b \leq a \leq p;\ \Sigma_{1,ab} \neq \Sigma_{2,ab}\big\}$, as those with a change. Additionally, let

$$\mathcal{P} = \big\{(a, b) : 1 \leq b \leq a \leq p;\ \big|\Sigma_{1,ab} - \Sigma_{2,ab}\big| \geq c\sqrt{\tau/n}\big\}$$

denote the subset of $\mathcal{G}$ where the change magnitude exceeds $c\sqrt{\tau/n}$ for some constant $c > 0$, with τ a threshold that separates detectable signals from noise (specified in Section 2.2.3;

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

for context, thresholds of order $\sqrt{(\log p)/n}$ are common in high-dimensional covariance estimation, as discussed in Bickel and Levina (2008)). Here, τ depends on the sample size n and dimension p, as the signal magnitude varies with both.

Focusing on entries in $\mathcal{P}$ is advantageous for two reasons. First, positions in $\mathcal{N}$ contribute no signal and only increase effective dimensionality, reducing power and raising computational burden. Second, changes in $\mathcal{G}$ but not in $\mathcal{P}$ are too weak to be reliably distinguished from sampling variability, potentially degrading detection. Even within $\mathcal{P}$, signal strengths vary, so differential weighting can further improve performance by prioritizing greater changes. To measure entry-wise change magnitudes, we introduce the statistic

$$\begin{aligned}\boldsymbol{v}_k =& \big(v_{k,h}\big)_{1\le h\le \frac{p(p+1)}{2}} \\ =& \frac{1}{(k)_2}\sum_{i\neq j\le k}\sum \dot{\boldsymbol{\sigma}}_{(i)}\circ\dot{\boldsymbol{\sigma}}_{(j)} + \frac{1}{(n-k)_2}\sum_{i\neq j>k}\sum \dot{\boldsymbol{\sigma}}_{(i)}\circ\dot{\boldsymbol{\sigma}}_{(j)} - \frac{2}{k\big(n-k\big)}\sum_{i\le k}\sum_{j>k}\dot{\boldsymbol{\sigma}}_{(i)}\circ\dot{\boldsymbol{\sigma}}_{(j)}.\end{aligned} \tag{2.4}$$

This is a simplified version of a related quantity $\tilde{\boldsymbol{v}}_k$ discussed in Remark 3, omitting higher-order terms to reduce estimation bias.

Remark 3. The statistic $\boldsymbol{v}_k$ is deliberately constructed by excluding the diagonal terms (i.e., cases where $i = j$) from a natural alternative formulation, thereby avoiding bias arising from higher-order moment contributions. To illustrate this choice, consider the alternative statistic

$$\begin{aligned}\widetilde{\boldsymbol{v}}_k &= \Big(\frac{1}{k}\sum_{i=1}^{k}\dot{\boldsymbol{\sigma}}_{(i)} - \frac{1}{n-k}\sum_{i=k+1}^{n}\dot{\boldsymbol{\sigma}}_{(i)}\Big)\circ\Big(\frac{1}{k}\sum_{i=1}^{k}\dot{\boldsymbol{\sigma}}_{(i)} - \frac{1}{n-k}\sum_{i=k+1}^{n}\dot{\boldsymbol{\sigma}}_{(i)}\Big) \\ &= \frac{1}{k^2}\sum_{i,j=1}^{k}\dot{\boldsymbol{\sigma}}_{(i)}\circ\dot{\boldsymbol{\sigma}}_{(j)} + \frac{1}{(n-k)^2}\sum_{i,j=k+1}^{n}\dot{\boldsymbol{\sigma}}_{(i)}\circ\dot{\boldsymbol{\sigma}}_{(j)} - \frac{2}{k(n-k)}\sum_{i=1}^{k}\sum_{j=k+1}^{n}\dot{\boldsymbol{\sigma}}_{(i)}\circ\dot{\boldsymbol{\sigma}}_{(j)}.\end{aligned}$$

This $\widetilde{\boldsymbol{v}}_k$ represents the element-wise squared difference between the pre-k and post-k sample second-moment estimators. However, the inclusion of $i = j$ terms renders $\widetilde{v}_{k,h}$ a biased

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

estimator of a quantity proportional to $(\Sigma_{1,ab} - \Sigma_{2,ab})^2$, with the bias originating from the variance of these diagonal terms. By omitting the $i = j$ contributions, $\boldsymbol{v}_k$ eliminates this higher-order bias. Furthermore, the expectation of $v_{k,h}$ is

$$\mathbb{E}v_{k,h} = \begin{cases} (1 - \frac{2}{n})^2(\Sigma_{1,ab} - \Sigma_{2,ab})^2, & k = k_0, \\ \frac{(k_0)_2}{(k)_2}(1 - \frac{2}{n})^2(\Sigma_{1,ab} - \Sigma_{2,ab})^2, & k > k_0, \\ \frac{(n-k_0)_2}{(n-k)_2}(1 - \frac{2}{n})^2(\Sigma_{1,ab} - \Sigma_{2,ab})^2, & k < k_0. \end{cases}$$

Thus, for any fixed $k \neq k_0$, larger absolute differences $|\Sigma_{1,ab} - \Sigma_{2,ab}|$ yield larger expected values of $v_{k,h}$.

Since the true k_0 is unknown, we aggregate over candidate locations k:

$$\boldsymbol{d} = \big(d_h\big)_{1 \le h \le \frac{p(p+1)}{2}} = \frac{1}{n-3}\sum_{k=2}^{n-2} \frac{k\big(n-k\big)}{n}\boldsymbol{v}_k. \tag{2.5}$$

The weights $k(n-k)/n$ are introduced to mitigate the effect of large variances near the boundaries, thereby enhancing the robustness of the statistic.

2.2.1 Selection of $\boldsymbol{\Gamma}$ and $\boldsymbol{\delta}$

Each entry d_h of $\boldsymbol{d}$ provides an aggregated estimate of the squared change magnitude in the corresponding covariance element. Larger values of d_h thus indicate stronger differences between $\Sigma_{1,ab}$ and $\Sigma_{2,ab}$. The vector $\boldsymbol{d}$ therefore captures essential information about the magnitude of entry-wise changes between the pre- and post-change covariances. Under sufficient signal strength in the set $\mathcal{G}$, a suitably chosen threshold τ can distinguish the strongly changed entries in $\mathcal{P}$ (as formalized in Section 3). This thresholding strategy for dimensionality reduction aligns with ideas in Dette et al. (2022).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

We accordingly define the weights as

$$\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma}), \quad \gamma_h = \mathbb{I}(d_h > \tau) \quad \text{and} \quad \delta_h = s_h \cdot \frac{\sqrt{d_h} \cdot \mathbb{I}(d_h > \tau)}{\|\sqrt{\boldsymbol{d}} \circ \mathbb{I}(\boldsymbol{d} > \tau)\|_2}, \tag{2.6}$$

where s_h is the sign of the estimated change in the h-th entry (detailed below). Note that the threshold τ is chosen to be positive (as described in Section 2.2.3), ensuring the validity of the square root operation. The diagonal form of $\boldsymbol{\Gamma}$ is standard for capturing dense changes across many entries. For $\boldsymbol{\delta}$, the element-wise magnitudes of $\mathrm{vech}(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)$ are naturally estimated by $\sqrt{\boldsymbol{d}}$ (inspired by Wang and Samworth (2018)). Normalization ensures $\|\boldsymbol{\delta}\|_2 = 1$, rendering $\hat{k}_0$ scale-invariant and preventing one component from dominating the other. Covariance changes may increase or decrease individual entries, leading to positive or negative differences $\Sigma_{1,ab} - \Sigma_{2,ab}$ across positions. Ignoring signs when constructing $\boldsymbol{\delta}$ risks cancellation in terms like $\boldsymbol{\delta}^\top(\dot{\boldsymbol{\sigma}}_{(i)} - \dot{\boldsymbol{\sigma}}_{(j)})$. The factor s_h therefore aligns δ_h with the dominant directional trend of the change, as determined by the rank-based procedure described next.

2.2.2 Rank-based sign selection

To capture the directional trend of covariance changes, i.e., whether individual entries increase or decrease after the change point, while enhancing practical robustness to outliers, we propose a rank-based statistic. For each entry $h = 1, \ldots, p(p+1)/2$, define

$$s_h^* = \sum_{k=1}^{n-1} \sqrt{\frac{k(n-k)}{n}} \Big\{ \frac{1}{k} \sum_{i=1}^{k} \mathrm{Rank}(\dot{\sigma}_{(i),h}) - \frac{1}{n-k} \sum_{i=k+1}^{n} \mathrm{Rank}(\dot{\sigma}_{(i),h}) \Big\},$$

where $\mathrm{Rank}(\dot{\sigma}_{(i),h})$ assigns ranks to the values $\{\dot{\sigma}_{(1),h}, \ldots, \dot{\sigma}_{(n),h}\}$ separately for each h: the smallest value receives rank 1, the next rank 2, and the largest rank n. In the rare event of ties, we assign the average rank to all tied values, following standard statistical practice.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

The weights $\sqrt{k(n-k)/n}$ stabilize contributions from candidates k near the boundaries, where fewer observations increase variability.

The final sign vector is $\boldsymbol{s} = (s_h)_{1 \le h \le \frac{p(p+1)}{2}}$, where

$$s_h = sgn(s_h^*) = \begin{cases} 1 & \text{if } s_h^* > 0 \\ 0 & \text{if } s_h^* = 0 \\ -1, & \text{if } s_h^* < 0 \end{cases}.$$

This rank-based approach focuses on relative ordering rather than absolute magnitudes, making sign estimation less sensitive to outliers or extreme observations. By applying signs only to entries that receive non-zero weights (those in the strongly changed set $\mathcal{P}$), the procedure further mitigates noise from weak or negligible changes. Overall, this strategy reduces effective dimensionality, improves estimation accuracy, and enhances computational efficiency.

2.2.3 Threshold selection

To determine a data-driven threshold τ in (2.6) that effectively separates unchanged entries ($\mathcal{N}$) from those with substantial changes ($\mathcal{P}$), we employ a nonparametric bootstrap procedure. The threshold is chosen based on the distribution of the statistic $\boldsymbol{d}$ in (2.5) under a null-like scenario where no systematic covariance change is present. For an entry h corresponding to a position in $\mathcal{N}$, the expected value of $v_{k,h}$ (and thus d_h) is small across candidates k, reflecting only random variation.

We generate bootstrap resamples by sampling the centered observations $\dot{\boldsymbol{X}}_i = \boldsymbol{X}_i - \bar{\boldsymbol{X}}$ (for $i = 1, \ldots, n$) with replacement. Let $\dot{\boldsymbol{X}}^{(q)} = (\dot{\boldsymbol{X}}_1^{(q)}, \ldots, \dot{\boldsymbol{X}}_n^{(q)})$ denote the q-th resample,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

for $q = 1, \ldots, Q$ where Q is sufficiently large (e.g., $Q = 500$ in our implementations). This resampling destroys the original change point structure, producing a "mixed" sample akin to the null hypothesis of no change. Under thorough mixing, the pre- and post-k segments for any candidate k exhibit no systematic covariance difference, resulting in small values of $d_h^{(q)}$ across all entries h.

The procedure is as follows:

1. For each resample q, compute the bootstrap analogs $\boldsymbol{v}_k^{(q)}$ and $\boldsymbol{d}^{(q)}$ using the definitions in Eqs. (2.4) and (2.5).

2. Set $\tau^{(q)} = \max_h d_h^{(q)}$, the maximum entry in the q-th bootstrap $\boldsymbol{d}$.

3. Define the threshold as $\tau = \min_{q=1,\ldots,Q} \tau^{(q)}$.

This conservative choice (the minimum over bootstrap maxima) ensures that τ captures the upper tail of the null-like distribution of $\max_h d_h$, providing a data-driven cutoff to identify strongly changed entries.

2.3 Existence of the change point

In addition to estimating the covariance change point location, it is often of interest to test whether a covariance change is present at all:

$$H_0 : \text{no change point} \quad \text{vs.} \quad H_1 : \text{a change point exists.}$$

This hypothesis testing problem has been widely studied (e.g., Aue and Horváth (2013); Kao et al. (2018)).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

Recall that the weight construction in (2.6) retains only entries of $\boldsymbol{d}$ exceeding the threshold τ, setting the others to zero. Let m denote the number of non-zero elements in the resulting vector $\boldsymbol{\delta}$. Under H_0 (no change), all true entry-wise differences are zero, so $m = 0$ is expected. Under H_1, at least some differences are non-zero, leading to $m > 0$ with high probability when the signal is sufficiently strong. A natural test therefore rejects H_0 if $m > 0$. The threshold $\tau = \min_{q=1,\ldots,Q} \tau^{(q)}$ from Section 2.2.3 (the minimum over bootstrap maxima) tends to be conservative, often yielding excessively small values that may fail to detect moderate changes. To achieve a better size-power balance, we adjust the threshold to the α-level quantile of the bootstrap maxima.

Specifically, for a nominal level α (e.g., $\alpha = 0.05$), sort the bootstrap maxima $\{\tau^{(1)}, \ldots, \tau^{(Q)}\}$ in increasing order and set $\tau_\alpha = \tau^{(q)}$, where q satisfies

$$\#\{i : \tau^{(i)} > \tau^{(q)}\} = \lfloor \alpha \cdot Q \rfloor.$$

The value of τ_α is used to calculate m in the testing procedure. Under H_0, the resampling-induced mixing ensures that the bootstrap maxima approximate the null distribution of $\max_h d_h$, so τ_α controls the probability of falsely identifying non-zero entries (yielding empirical size close to α). Under H_1, the systematic change in the original data produces larger values of d_h for changed entries, exceeding τ_α and leading to $m > 0$.

Remark 4. Methods such as Yu and Chen (2021) and Zhou et al. (2025) employ Gaussian multiplier bootstraps to approximate the null distribution of their test statistics and calibrate critical values, ensuring asymptotic size control. In contrast, our nonparametric resampling bootstrap serves dual purposes: (i) identifying informative covariance entries for data-adaptive weighting (Section 2.2.3), and (ii) providing an empirical quantile-based

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

threshold for testing (this section). Although not a formal multiplier calibration, extensive simulations show excellent empirical size control (typically at or below the nominal α) alongside high power. Note that we employ a 'min-max' criterion for weight selection to include potentially informative signals, whereas a stricter quantile-based threshold is used in Section 2.3 for formal hypothesis testing to control the type I error rate.

3. Asymptotic Properties

To analyze the theoretical properties of our proposed approach, we first outline the necessary assumptions regarding the distribution of the random vectors $\boldsymbol{X}_i$ and the structure of the covariance matrices. These assumptions are critical to ensure the validity and consistency of our method.

Assumption 1. (a) *Sub-Gaussianity and Bounded Covariance: Let* $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^\top, i = 1, \ldots, n$. *For any* $1 \leq j \leq p$, X_{ij} *is a sub-Gaussian random variable, meaning there exist positive constants* C_1, C_2 *(independent of indices* i *and* j*) such that, for any* $t > 0$,

$$\mathbb{P}\big\{|X_{ij}| > t\big\} \leq C_1 e^{-C_2 t^2}.$$

Additionally, the covariance matrices before and after the change point satisfy $\max_{1 \leq a,b \leq p} \big|\Sigma_{\nu,ab}\big| \leq M, \nu = 1, 2$ *for some positive constant* M.

(b) *Magnitude of Covariance Matrix Difference: The smallest nonzero element of the matrix* $\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2$ *satisfies* $|\Sigma_{1,ab} - \Sigma_{2,ab}| \geq c\sqrt{\tau/n}$ *for some positive constant* c.

Assumption 1(a) imposes sub-Gaussian tails, which is crucial for deriving concentration inequalities and ensuring stability in high-dimensional settings. It is standard in high-dimensional statistics to guarantee finite-sample performance, as seen in works like Vershynin

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

(2018). This condition also bounds individual covariance entries, stabilizing the estimation process. Assumption 1(b) imposes a minimal signal strength on the difference between $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, ensuring that the change in covariance structure is sufficiently pronounced to be detectable in high-dimensional settings, thus enabling reliable identification of the change point. Under these assumptions, we establish the following Proposition 1 and Theorem 1.

Proposition 1. (a) *No Covariance Change: For any position* (a, b) *where* $\Sigma_{1,ab} = \Sigma_{2,ab}$*, the following conclusion holds under Assumption 1(a):*

$$\mathbb{P}\big\{d_h > \tau\big\} \leq c_1 n\big(e^{-c_2\sqrt{\tau}} + e^{-c_3\sqrt{n\tau}} + e^{-c_4 n\sqrt{\tau}}\big),$$

where $c_i (i = 1, \ldots, 4)$ *are positive constants. In particular, if* $p^2 n = o(e^{c\sqrt{\tau}})$ *for some constant* c*, then*

$$p^2 \cdot \mathbb{P}\{d_h > \tau\} \to 0. \tag{3.1}$$

(b) *Significant Covariance Change: For any position* (a, b) *where* $\big|\Sigma_{1,ab} - \Sigma_{2,ab}\big| \geq c\sqrt{\tau/n}$ *for some constant* c*, the following holds under Assumption 1(a):*

$$\mathbb{P}\big\{d_h \leq \tau\big\} \leq c_1 n\big(e^{-c_2\tau} + e^{-c_3\sqrt{n\tau}} + ne^{-c_4\sqrt{\tau}} + ne^{-c_5 n\sqrt{\tau}}\big),$$

where $c_i (i = 1, \ldots, 5)$ *are positive constants. In particular, if* $p^2 n^2 = o(e^{c\sqrt{\tau}})$ *for some constant* c*, then*

$$p^2 \cdot \mathbb{P}\{d_h \leq \tau\} \to 0. \tag{3.2}$$

Proposition 1(a) implies that, when the true covariances are equal at position (a, b), the probability that d_h exceeds the threshold is negligible. Consequently, such entries contribute little to the aggregated statistic $T_{\boldsymbol{W}}$, thereby reducing the influence of random fluctuations.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

In contrast, Proposition 1(b) shows that, when a genuine covariance difference exists at position (a, b), the probability that d_h fails to exceed the threshold is negligible. This ensures that truly informative entries are reliably retained in $T_{\boldsymbol{W}}$. Note that the condition $p^2n^2 = o(e^{c\sqrt{\tau}})$ implies $p^2n = o(e^{c\sqrt{\tau}})$ suggesting that these asymptotic results (3.1) and (3.2) hold under the condition $p^2n^2 = o(e^{c\sqrt{\tau}})$.

Theorem 1. *Suppose Assumption 1 holds, and the statistic* (2.1) *satisfies* $\gamma_h > 0$ *for a nonempty subset* $J_1 \subseteq \mathcal{P}$ *and* $\gamma_h = 0$ *otherwise, while* δ_h *matches the sign of* $\Sigma_{1,ab} - \Sigma_{2,ab}$ *for a nonempty subset* $J_2 \subseteq \mathcal{P}$ *and* $\delta_h = 0$ *otherwise, where* $\mathcal{P} = \big\{(a, b) : 1 \le b \le a \le p;\ \big|\Sigma_{1,ab} - \Sigma_{2,ab}\big| \ge c\sqrt{\tau/n}\big\}$. *Then, the estimator* (2.3) *satisfies*

$$\mathbb{P}\Big\{\Big|\frac{\hat{k}_0}{k_0} - 1\Big| \ge \epsilon\Big\} \le c_1 p^2 n\big(e^{-c_2\tau} + e^{-c_3\sqrt{n\tau}} + ne^{-c_4\sqrt{\tau}} + ne^{-c_5 n\sqrt{\tau}}\big),$$

where $c_i (i = 1, \ldots, 5)$ *are some constants. In particular, if* $p^2n^2 = o(e^{c\sqrt{\tau}})$ *for some constant* c, *then*

$$\frac{\hat{k}_0}{k_0} \xrightarrow{p} 1.$$

Theorem 1 establishes the consistency of change point estimation under stringent regularity conditions imposed on the weight matrix $\boldsymbol{W} = \boldsymbol{\Gamma} + \boldsymbol{\delta}\boldsymbol{\delta}^\top$. Notably, Proposition 1 demonstrates that the equality $J_1 = J_2 = \mathcal{P}$ is statistically guaranteed in the asymptotic setting. The data-adaptive parameter selection framework introduced in Section 2.2 inherently satisfies these conditions, ensuring that the consistency guarantees are not only theoretical but also practically achievable.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

4. Simulations

In this section, we present simulation studies to evaluate the numerical performance of the proposed estimator in finite samples. We begin by defining the notation. Unless otherwise specified, all simulation results are based on 1000 independent replications per scenario. For each replication $i = 1, \ldots, B$ (with $B = 1000$), we obtain an estimate $\hat{k}_i$ of the true change point k_0 and compute the absolute error $\Delta_i = |\hat{k}_i - k_0|$. To evaluate change point localization performance in the subsequent analysis, we report the mean absolute error (MAE) and standard deviation (SD) of the $\{\Delta_i\}_{i=1}^{B}$ as follows:

$$\text{MAE} = \frac{1}{B}\sum_{i=1}^{B}\Delta_i, \quad \text{SD} = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\Delta_i - \text{MAE})^2}.$$

4.1 Data generation settings

For each Monte Carlo replication, we first generate a mean vector $\mu = (\mu_1, \ldots, \mu_p)$, where each component μ_j is drawn independently from Unif$(1, 2)$. Conditional on this fixed mean vector μ, we then generate n observations from a multivariate normal distribution with mean μ and the specified covariance matrix. Specifically, we consider four distinct matrices $\mathbf{\Sigma}_1$ as follows:

1) $\mathbf{\Sigma}_1^{(1)} = \mathbf{I}_p$;

2) $\mathbf{\Sigma}_1^{(2)} = \left(\Sigma_{ij}^{(2)}\right)_{i,j=1,\ldots,p}$, where $\Sigma_{ii}^{(2)} = 1, \Sigma_{ij}^{(2)} = 0.4$ for $5(k-1)+1 \le i \ne j \le 5k$, $k = 1, \ldots, \lfloor p/5 \rfloor$, and $\Sigma_{ij}^{(2)} = 0$ otherwise;

3) $\mathbf{\Sigma}_1^{(3)} = (\Sigma_{ij}^{(3)})_{i,j=1,\ldots,p}$, where $\Sigma_{ii}^{(3)} = 1$ and $\Sigma_{ij}^{(3)} = (-1)^{\min\{i,j\}+1}0.4^{|i-j|}$;

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

4) $\boldsymbol{\Sigma}_1^{(4)} = \boldsymbol{Q}\boldsymbol{D}_1\boldsymbol{Q}^\top$, where $\boldsymbol{Q}$ is an orthogonal matrix and $\boldsymbol{D}_1 = \text{diag}(\boldsymbol{u}_1)$ with $\boldsymbol{u}_1$ being a p-dimensional vector whose elements are i.i.d. from Unif(0.5, 1.5).

In these settings, $\boldsymbol{\Sigma}_1^{(1)}$ is a diagonal matrix, while $\boldsymbol{\Sigma}_1^{(2)}$, $\boldsymbol{\Sigma}_1^{(3)}$ and $\boldsymbol{\Sigma}_1^{(4)}$ introduce weak correlations. Specifically, $\boldsymbol{\Sigma}_1^{(2)}$ is a block diagonal matrix with 5×5 blocks, as used in Cai et al. (2013); $\boldsymbol{\Sigma}_1^{(3)}$ is Toeplitz-type matrix incorporating positive and negative correlations via $(-1)^{\min\{i,j\}+1}$; and $\boldsymbol{\Sigma}_1^{(4)}$ represents a general covariance matrix, as employed in Ding et al. (2025). The structure of each $\boldsymbol{\Sigma}_1$ is illustrated in Figure 1 with $p = 20$. To examine variations in covariance, we consider different relationships between p and n, which determine the signal strength. For $n = 200$ and $p \in \{50, 100\}$, given a specific $\boldsymbol{\Sigma}_1$, we study three distinct covariance matrices $\boldsymbol{\Sigma}_2$ as follows:

1) $\boldsymbol{\Sigma}_2^{(1)} = \boldsymbol{\Sigma}_1 + \boldsymbol{\Delta}_1$, where $\boldsymbol{\Delta}_1 = \text{diag}\big((\boldsymbol{u}_{1,0.3p}^\top, \boldsymbol{u}_{2,0.7p}^\top)^\top\big)$, with $\boldsymbol{u}_{1,0.3p}$ and $\boldsymbol{u}_{2,0.7p}$ being $0.3p$- and $0.7p$-dimensional vectors whose elements are i.i.d. from Unif(2, 3) and Unif(0, 1), respectively;

2) $\boldsymbol{\Sigma}_2^{(2)} = \boldsymbol{\Sigma}_1 + \boldsymbol{\Delta}_2$, where $\boldsymbol{\Delta}_2 = (\delta_{ij})_{i,j=1,\ldots,p}$, with $\delta_{ii} = 2$, $\delta_{ij} = (-1)^{\min\{i,j\}+1}$ for $|i - j| = 1$ and $\delta_{ij} = 0$ otherwise;

3) $\boldsymbol{\Sigma}_2^{(3)} = \boldsymbol{D}^{\frac{1}{2}}\boldsymbol{\Sigma}_1\boldsymbol{D}^{\frac{1}{2}}$, where $\boldsymbol{D} = \text{diag}(\boldsymbol{u})$ with $\boldsymbol{u}$ being a p-dimensional vector whose elements are i.i.d. from Unif(2, 4).

In these settings for $\boldsymbol{\Sigma}_2$, $\boldsymbol{\Sigma}_2^{(1)}$ introduces changes on the diagonal with varying magnitudes; $\boldsymbol{\Sigma}_2^{(2)}$ indicates a band structure change, introducing increasing and decreasing variations through $(-1)^{\min\{i,j\}+1}$; and $\boldsymbol{\Sigma}_2^{(3)}$ represents a covariance matrix where each element undergoes multiplicative changes of varying magnitudes, a configuration similarly explored

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

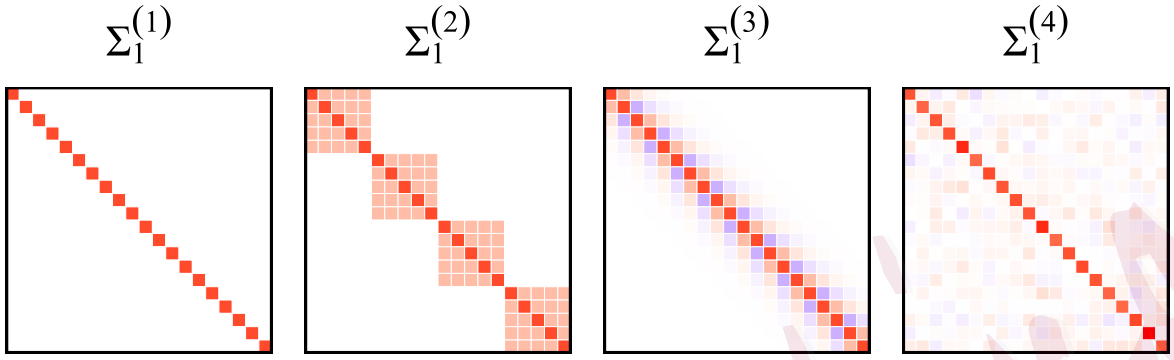

Figure 1: Four distinct structures of $\mathbf{\Sigma}_1$ examined in the simulation with $p = 20$.

in Cai et al. (2013). The primary objective of this simulation setup is to compare various covariance matrix structures and assess how differences in these matrices influence the performance of change point detection methods. For $n = 200$ and $p \in \{200, 300\}$, preliminary simulations indicate that a strong difference signal, as described above, is not necessary to achieve satisfactory change point detection results. Therefore, we weaken the signal by considering, for a given $\mathbf{\Sigma}_1$, three new covariance matrices $\mathbf{\Sigma}_2$ as follows:

1) $\mathbf{\Sigma}_2^{(1*)} = \mathbf{\Sigma}_1 + \mathbf{\Delta}_1^*$, where $\mathbf{\Delta}_1^* = \mathrm{diag}\big((\boldsymbol{u}_{1,20}^\top, \boldsymbol{u}_{2,30}^\top, \mathbf{0}_{p-50}^\top)^\top\big)$, with $\boldsymbol{u}_{1,20}$ and $\boldsymbol{u}_{2,30}$ being 20- and 30-dimensional vectors whose elements are i.i.d. from Unif(1.5, 3.5) and Unif(0, 1);

2) $\mathbf{\Sigma}_2^{(2*)} = \mathbf{\Sigma}_1 + \mathbf{\Delta}_2^*$, where $\mathbf{\Delta}_2^* = \mathrm{diag}(\mathbf{\Delta}_2, \mathbf{0}_{p-50\times p-50})$, with $\mathbf{\Delta}_2 = (\delta_{ij})_{50\times 50}$, $\delta_{ii} = 2$, $\delta_{ij} = (-1)^{\min\{i,j\}+1}$ for $|i-j| = 1$ and $\delta_{ij} = 0$ otherwise;

3) $\mathbf{\Sigma}_2^{(3*)} = \boldsymbol{D}^{*\frac{1}{2}} \mathbf{\Sigma}_1 \boldsymbol{D}^{*\frac{1}{2}}$, where $\boldsymbol{D}^* = \mathrm{diag}(\boldsymbol{u}^*)$ with $\boldsymbol{u}^*$ being a p-dimensional vector whose elements are i.i.d. from Unif(1, 3).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

4.2 Testing for the existence of a change point

To evaluate the finite-sample performance of our proposed test for the existence of a change point, we conduct a simulation study. Under the null hypothesis H_0, we consider four distinct covariance matrices $\mathbf{\Sigma}_1$, as described in Section 4.1, with $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_1$. Under the alternative hypothesis H_1, for each $\mathbf{\Sigma}_1$, we pair it with three corresponding post-change covariance matrices $\mathbf{\Sigma}_2^{(1)}, \mathbf{\Sigma}_2^{(2)}, \mathbf{\Sigma}_2^{(3)}$, as specified in Section 4.1. We set the sample size to $n = 200$, the change point to $k_0 = \lfloor 0.3n \rfloor$, the number of resampling iterations to $Q = 500$, and vary the dimension p from 50 to 200. For each configuration, we report the empirical frequency of rejecting H_0. The simulation results under H_0 and H_1 are summarized in Table 1. The proportions of rejecting H_0 when $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ are consistently close to 0, while the proportions of rejecting H_0 when $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$ are consistently close to 1. These results demonstrate the effectiveness of our method in detecting the existence of a change point.

Table 1: Rejection frequencies for the existence of a change point over 1000 simulation replications: rows labeled H_0 report rejections when $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_1$, while rows labeled H_1 report the average rejections across three distinct $\mathbf{\Sigma}_2$ configurations.

	$\mathbf{\Sigma}_1^{(1)}$	$\mathbf{\Sigma}_1^{(2)}$	$\mathbf{\Sigma}_1^{(3)}$	$\mathbf{\Sigma}_1^{(4)}$	$\mathbf{\Sigma}_1^{(1)}$	$\mathbf{\Sigma}_1^{(2)}$	$\mathbf{\Sigma}_1^{(3)}$	$\mathbf{\Sigma}_1^{(4)}$
	$p = 50$				$p = 100$			
H_0	33	45	38	32	27	39	25	35
H_1	999.7	1000	999.3	1000	1000	1000	1000	1000
	$p = 150$				$p = 200$			
H_0	26	19	26	21	21	22	20	20
H_1	1000	1000	1000	1000	1000	1000	1000	1000

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

4.3 Accuracy of change point localization

Before presenting the simulation setup, we introduce two alternative statistics corresponding to different choices of $\boldsymbol{\delta}$. The first statistic, $T_{\boldsymbol{W}_1}$, sets $\boldsymbol{\delta}_1 = \mathbf{0}$. The second statistic, $T_{\boldsymbol{W}_2}$, employs

$$\delta_{2,h} = s_h \cdot \mathbb{I}(d_h > \tau)/\|\mathbb{I}(\boldsymbol{d} > \tau)\|_2,$$

which assigns equal normalized weights to the selected entries. For the data-generating process, we consider multiple configurations of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ as described in Section 4.1. The sample size is fixed at $n = 200$, with parameters $c_1 = 0.1$ and $c_2 = 0.9$; the dimension p varies from 50 to 300, and the change point is set at $k_0 = \lfloor 0.3n \rfloor$. For each configuration, both the method introduced by Dette et al. (2022) and our method were used to estimate the change point k_0, performing 1000 repetitions to compute the mean absolute error (MAE) and standard deviation (SD) of the estimates. The numerical results are summarized in Table 2 for strong signal scenarios and Table 3 for weak signal scenarios.

By comparing the estimators, we identify several key insights that highlight the advantages of our approach. First, across all scenarios, the first three estimates introduced in this paper exhibit an MAE closer to zero compared to those of Dette et al. (2022), along with a smaller SD in most cases, indicating more effective covariance change point detection. Second, while the first three estimators perform similarly in most cases, they exhibit notable differences in specific scenarios. For instance, when $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_2^{(1)}$ with $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1^{(1)}, \boldsymbol{\Sigma}_1^{(2)}, \boldsymbol{\Sigma}_1^{(3)}$ for $p = 50$, $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_2^{(1*)}$ with $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1^{(1)}$ for $p = 200$, and $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_2^{(1*)}$ with $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1^{(2)}$ for $p = 300$, the SD of $T_{\boldsymbol{W}_1}$ is significantly larger. This is attributed to the emergence of bounded estimates, whereas $T_{\boldsymbol{W}}$ performs robustly, suggesting that incorporating $T_2(k)$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

Table 2: MAE (SD) of estimators for different cases with $p = 50$ and 100.

	$\boldsymbol{\Sigma}_2^{(1)}$	$\boldsymbol{\Sigma}_2^{(2)}$	$\boldsymbol{\Sigma}_2^{(3)}$	$\boldsymbol{\Sigma}_2^{(1)}$	$\boldsymbol{\Sigma}_2^{(2)}$	$\boldsymbol{\Sigma}_2^{(3)}$
$p = 50$	$\boldsymbol{\Sigma}_1^{(1)}$			$\boldsymbol{\Sigma}_1^{(2)}$		
$T_{\boldsymbol{W}}$	0.157(0.457)	0.023(0.150)	0.025(0.156)	0.216(0.544)	0.025(0.156)	0.103(0.350)
$T_{\boldsymbol{W_1}}$	0.273(3.754)	0.021(0.143)	0.026(0.159)	0.345(3.828)	0.025(0.156)	0.098(0.347)
$T_{\boldsymbol{W_2}}$	0.220(0.569)	0.034(0.187)	0.028(0.182)	3.197(17.99)	0.026(0.159)	0.636(7.551)
H.Dette	1.608(2.845)	0.794(1.493)	0.677(1.378)	1.786(3.029)	0.891(1.722)	1.469(2.643)
	$\boldsymbol{\Sigma}_1^{(3)}$			$\boldsymbol{\Sigma}_1^{(4)}$		
$T_{\boldsymbol{W}}$	0.196(0.512)	0.055(0.249)	0.045(0.226)	0.157(0.457)	0.030(0.176)	0.015(0.122)
$T_{\boldsymbol{W_1}}$	0.321(3.822)	0.054(0.247)	0.045(0.230)	0.137(0.393)	0.028(0.171)	0.017(0.129)
$T_{\boldsymbol{W_2}}$	0.615(6.494)	0.069(0.280)	0.058(0.250)	0.321(3.823)	0.041(0.203)	0.019(0.144)
H.Dette	1.664(2.932)	1.077(1.963)	1.002(1.940)	1.611(2.850)	0.939(1.741)	0.753(1.388)
$p = 100$	$\boldsymbol{\Sigma}_1^{(1)}$			$\boldsymbol{\Sigma}_1^{(2)}$		
$T_{\boldsymbol{W}}$	0.035(0.189)	0.001(0.032)	0.001(0.032)	0.059(0.252)	0.002(0.045)	0.014(0.118)
$T_{\boldsymbol{W_1}}$	0.031(0.179)	0.001(0.032)	0.001(0.032)	0.063(0.259)	0.002(0.045)	0.012(0.109)
$T_{\boldsymbol{W_2}}$	0.033(0.190)	0.001(0.032)	0.002(0.045)	0.065(0.266)	0.003(0.055)	0.027(0.168)
H.Dette	0.695(1.471)	0.371(0.881)	0.357(0.844)	0.705(1.493)	0.387(0.901)	0.621(1.271)
	$\boldsymbol{\Sigma}_1^{(3)}$			$\boldsymbol{\Sigma}_1^{(4)}$		
$T_{\boldsymbol{W}}$	0.040(0.201)	0.006(0.077)	0.003(0.055)	0.037(0.199)	0.001(0.032)	0.002(0.045)
$T_{\boldsymbol{W_1}}$	0.039(0.199)	0.005(0.071)	0.003(0.055)	0.038(0.201)	0.002(0.045)	0.002(0.045)
$T_{\boldsymbol{W_2}}$	0.051(0.246)	0.006(0.077)	0.004(0.063)	0.052(0.239)	0.002(0.045)	0.002(0.045)
H.Dette	0.728(1.505)	0.530(1.141)	0.541(1.109)	0.750(1.514)	0.375(0.904)	0.377(0.917)

enhances the original statistic $T_1(k)$. However, this improvement relies on a well-chosen $\boldsymbol{\delta}$. Similarly, in certain cases, particularly those involving $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_2^{(1)}$ with $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1^{(2)}, \boldsymbol{\Sigma}_1^{(3)}, \boldsymbol{\Sigma}_1^{(4)}$ for $p = 50$ and $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_2^{(1*)}$ with $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1^{(1)}, \boldsymbol{\Sigma}_1^{(2)}, \boldsymbol{\Sigma}_1^{(3)}, \boldsymbol{\Sigma}_1^{(4)}$ for $p = 200, 300$, the SD of

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

Table 3: MAE (SD) of estimators for different cases with $p = 200$ and 300.

	$\boldsymbol{\Sigma}_2^{(1*)}$	$\boldsymbol{\Sigma}_2^{(2*)}$	$\boldsymbol{\Sigma}_2^{(3*)}$	$\boldsymbol{\Sigma}_2^{(1*)}$	$\boldsymbol{\Sigma}_2^{(2*)}$	$\boldsymbol{\Sigma}_2^{(3*)}$
$p = 200$	$\boldsymbol{\Sigma}_1^{(1)}$			$\boldsymbol{\Sigma}_1^{(2)}$		
$T_{\boldsymbol{W}}$	0.082(0.312)	0.024(0.160)	0.017(0.129)	0.133(0.462)	0.036(0.197)	0.047(0.221)
$T_{\boldsymbol{W_1}}$	0.202(3.774)	0.027(0.168)	0.017(0.129)	0.141(0.479)	0.037(0.194)	0.041(0.203)
$T_{\boldsymbol{W_2}}$	0.820(8.948)	0.027(0.168)	0.020(0.140)	1.579(12.86)	0.041(0.203)	0.052(0.235)
H.Dette	1.323(2.602)	0.789(1.511)	0.738(1.390)	1.277(2.372)	0.830(1.619)	1.287(2.178)
	$\boldsymbol{\Sigma}_1^{(3)}$			$\boldsymbol{\Sigma}_1^{(4)}$		
$T_{\boldsymbol{W}}$	0.125(0.440)	0.066(0.282)	0.032(0.176)	0.093(0.329)	0.021(0.150)	0.031(0.173)
$T_{\boldsymbol{W_1}}$	0.122(0.389)	0.061(0.267)	0.032(0.176)	0.089(0.327)	0.022(0.153)	0.032(0.176)
$T_{\boldsymbol{W_2}}$	0.262(3.788)	0.195(3.804)	0.040(0.196)	1.070(10.39)	0.023(0.157)	0.035(0.189)
H.Dette	1.305(2.313)	1.048(2.045)	1.155(1.996)	1.215(2.150)	0.729(1.493)	1.085(1.963)
$p = 300$	$\boldsymbol{\Sigma}_1^{(1)}$			$\boldsymbol{\Sigma}_1^{(2)}$		
$T_{\boldsymbol{W}}$	0.090(0.319)	0.029(0.168)	0.018(0.133)	0.163(0.520)	0.041(0.208)	0.019(0.137)
$T_{\boldsymbol{W_1}}$	0.086(0.308)	0.028(0.171)	0.018(0.133)	0.278(3.764)	0.038(0.201)	0.018(0.133)
$T_{\boldsymbol{W_2}}$	0.468(6.371)	0.030(0.171)	0.021(0.143)	0.317(3.832)	0.033(0.184)	0.018(0.133)
H.Dette	1.234(2.333)	0.844(1.670)	0.887(1.780)	1.193(2.184)	0.826(1.616)	1.034(1.829)
	$\boldsymbol{\Sigma}_1^{(3)}$			$\boldsymbol{\Sigma}_1^{(4)}$		
$T_{\boldsymbol{W}}$	0.140(0.430)	0.070(0.309)	0.014(0.118)	0.111(0.378)	0.022(0.153)	0.007(0.095)
$T_{\boldsymbol{W_1}}$	0.143(0.437)	0.065(0.277)	0.016(0.148)	0.096(0.321)	0.024(0.160)	0.008(0.100)
$T_{\boldsymbol{W_2}}$	0.144(0.438)	0.069(0.304)	0.020(0.140)	0.742(8.371)	0.034(0.207)	0.008(0.100)
H.Dette	1.201(2.224)	1.012(1.882)	1.059(1.816)	1.245(2.178)	0.857(1.707)	0.693(1.379)

$T_{\boldsymbol{W_2}}$ is notably larger, reinforcing the appropriateness of our $\boldsymbol{\delta}$ selection. Third, comparing the data generation process settings and results in Tables 2 and 3, we observe that as the dimension p increases from less than n to greater than n, our proposed statistic does not

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

require the strength of the change to scale with p. Instead, a fixed-scale signal strength, independent of p, is sufficient to achieve superior detection results, highlighting the effectiveness of the reweighting method in high-dimensional settings. Finally, our estimator consistently delivers robust performance across all scenarios, demonstrating its capability to accurately estimate change point locations under diverse covariance structures and varying degrees of covariance changes.

In summary, the simulations demonstrate substantial improvements from our proposed estimator, especially when the true change point lies away from the sample center. To demonstrate the computational efficiency of our method, we measured the execution time of each step. Specifically, using MATLAB (R2018a) on an 11th Gen Intel® Core™ i5-11320H 3.20 GHz processor with $p = 100$ and $n = 200$, detecting the existence of the change point and computing the threshold took approximately 7.8 seconds, and estimating the change point took only 0.17 seconds. The robustness and accuracy of our method across a wide range of covariance structures and change scenarios underscore its potential as a reliable tool for change point detection in diverse applications.

5. Empirical Evidence

In this section, we apply our method to a dataset of handwritten numeral features ('0'–'9') extracted from a collection of Dutch utility maps (`https://archive.ics.uci.edu/dataset/72/multiple+features`) and compare its performance with other estimators described in Section 4.3. The dataset comprises 2,000 handwritten numerals ('0'–'9'), with 200 patterns per class, digitized into binary images. These digits are represented using six different fea-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

ture sets, and we selected the set of 76 Fourier coefficients (i.e., $p = 76$). We focus on the change point problem involving the numeral '0' relative to other numerals.

To investigate the existence of a change point, we construct concatenated datasets by randomly selecting 100 patterns of the numeral '0' and 100 patterns of another numeral, resulting in a dataset with $n = 200$ and $k_0 = 100$. We then apply our proposed method, as detailed in Section 2.3, to test for the existence of a change point in these newly formed datasets. For every pair of numerals, we repeat this experiment 1000 times and report the empirical rejection proportion of the null hypothesis H_0. Furthermore, given the rotation-invariant nature of Fourier feature sets, we do not anticipate a change point between the numerals '6' and '9' (van Breukelen et al., 1998). To validate this, we similarly combine the datasets for numerals '6' and '9' and conduct the change point detection test. The results, summarized in Table 4, demonstrate that our method effectively identifies the existence of the change point. Specifically, a change point is consistently detected between the numeral '0' and each of the other numerals, while the results for numerals '6' and '9' align with the theoretical expectation, confirming the absence of a change point between them.

Table 4: Proportion of rejecting H_0 for various pairs of numbers.

	'0' and '1'	'0' and '2'	'0' and '3'	'0' and '4'	'0' and '5'
Prop	1.000	1.000	1.000	1.000	1.000
	'0' and '6'	'0' and '7'	'0' and '8'	'0' and '9'	'6' and '9'
Prop	1.000	1.000	0.962	1.000	0.022

Next, we address the problem of estimating the change point. Building on the above

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

design, we concatenate the dataset for '0' with each other numeral in turn by randomly sampling 120 patterns of '0' and 180 patterns of the other numeral, resulting in $n = 300$ with the true change point $k_0 = 0.4n$. We then estimate the change point using the boundary trimming parameters $c_1 = 0.1$ and $c_2 = 0.9$, repeating the procedure 1000 times for each combination. We visualize the estimation results by plotting boxplots of estimated relative change point locations $\hat{k}_i/n$ $(i = 1, \ldots, B)$ for each combination in Figure 2.

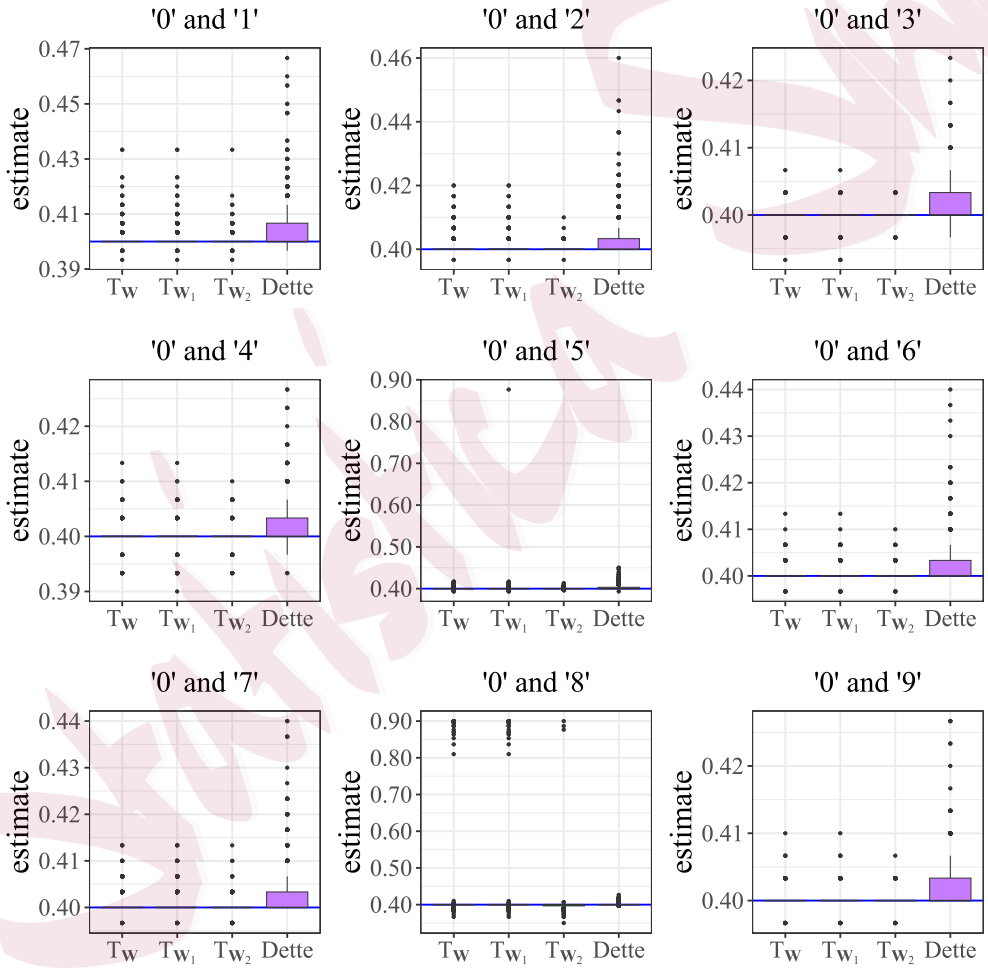

Figure 2: Boxplots of the change point estimates for each combination.

As shown in Figure 2, $T_{\boldsymbol{W}}, T_{\boldsymbol{W}_1}, T_{\boldsymbol{W}_2}$ are generally more accurate than the benchmark estimator of Dette et al. (2022) across most pairings. However, for the combination of '0' and

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

'5', $T_{\boldsymbol{W}_1}$ produces the boundary estimate, while $T_{\boldsymbol{W}_2}$ and $T_{\boldsymbol{W}}$ perform well, which is consistent with our simulation results. For the combination of '0' and '8', the latter three estimators all produce the boundary estimate, and multiple contributing factors influence this outcome. First, the difference between the covariance matrices of '0' and '8' is relatively small, as seen in the results in Table 4, which affects the performance of estimators. Second, similar to the CUSUM statistic, a general statistic can be expressed as $T'_{\boldsymbol{W}}(k) = \{\frac{(k)_2(n-k)_2}{n^2}\}^{\eta} \cdot T_{\boldsymbol{W}}(k)$, where $\eta \in [0, 1]$ is a tuning parameter. Hariz et al. (2007) considered the CUSUM statistic with different tuning parameter settings. The tuning parameters $\eta = 1$ and $\eta = 0.5$ correspond to the statistics used in Fryzlewicz (2014) and Enikeeva and Harchaoui (2019), respectively. Our boundary removal procedure is reasonable given our $k_0 = \lfloor r_0 n \rfloor$ setting, and following Jiang et al. (2024), we use the statistic with $\eta = 0$. With boundary removal, the estimate with $\eta = 0$ can be more accurate, especially when the relative location of the change point r_0 is not near 0.5. However, the estimate is inevitably susceptible to boundary effects. Better results may be obtained by estimating the change point with a $\eta \neq 0$ for the '0' and '8' combination. Furthermore, the boundary removal procedure may not be strictly necessary with a sufficiently large η. The selection of η will be explored in future work.

6. Conclusions

This article addresses the problem of detecting the change point in covariance matrices. First, we propose a reweighted CUSUM-type statistic for estimating the change point. Next, we introduce a data-adaptive parameter selection method. Additionally, building on this change point estimation approach, we develop a hypothesis testing procedure to assess the existence

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

of a change point. Extensive simulation studies and real-world data analyses validate the accuracy of our change point estimation, the precision of our parameter selection, and the effectiveness of our method for testing the existence of the change point. Finally, our theoretical results establish the consistency of the proposed estimators. Future research could further investigate the construction of theoretically optimal estimators, relax the tail assumptions, and extend the proposed framework to broader settings, including online change point detection and multiple change point scenarios.

Disclosure Statement

The authors report there are no competing interests to declare.

Supplementary Material

The online Supplementary Material contains some lemmas that are essential for our proofs, as well as the proofs of the theorem and proposition introduced in Section 3.

Acknowledgments

The authors are grateful to the editor, associate editor, and reviewers for their insightful comments and suggestions, which have improved the article significantly. The work is supported by the National Natural Science Foundation of China (Nos. 12371279, 12501391,12571297, 12231017), National Key R&D Program of China-2022YFA1008000, and the CAS Talent Introduction Program (Category B). Canhuang Xu and Lei Shu are co-first authors.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

References

Aue, A., S. Hörmann, L. Horváth, and M. Reimherr (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics 37*(6B), 4046–4087.

Aue, A. and L. Horváth (2013). Structural breaks in time series. *Journal of Time Series Analysis 34*(1), 1–16.

Bai, J., X. Han, and Y. Shi (2020). Estimation and inference of change points in high-dimensional factor models. *Journal of Econometrics 219*(1), 66–100.

Berkes, I., E. Gombay, and L. Horváth (2009). Testing for changes in the covariance structure of linear processes. *Journal of Statistical Planning and Inference 139*(6), 2044–2063.

Bickel, P. J. and E. Levina (2008). Covariance regularization by thresholding. *The Annals of Statistics 36*(6), 2577–2604.

Cai, T., W. Liu, and Y. Xia (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association 108*(501), 265–277.

Cai, T., Z. Ma, and Y. Wu (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probability Theory and Related Fields 161*(3), 781–815.

Chen, C. Y. H., Y. Okhrin, and T. Wang (2024). Monitoring network changes in social media. *Journal of Business & Economic Statistics 42*(2), 391–406.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

Cho, H. (2016). Change-point detection in panel data via double cusum statistic. *Electronic Journal of Statistics 10*, 2000–2038.

Cui, X., H. Geng, Z. Wang, and C. Zou (2024). Robust estimation of high-dimensional linear regression with changepoints. *IEEE Transactions on Information Theory 70*(10), 7297–7319.

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

Ding, X., Y. Hu, and Z. Wang (2025). Two sample test for covariance matrices in ultra-high dimension. *Journal of the American Statistical Association 120*(552), 2210–2221.

Dörnemann, N. and H. Dette (2024). Detecting change points of covariance matrices in high dimensions. *arXiv*, 10.48550/arXiv.2409.15588.

Enikeeva, F. and Z. Harchaoui (2019). High-dimensional change-point detection under sparse alternatives. *The Annals of Statistics 47*(4), 2051–2079.

Fischer, T., M. Gemmer, L. Liu, and B. Su (2012). Change-points in climate extremes in the Zhujiang River Basin, South China, 1961–2007. *Climatic Change 110*(3), 783–799.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics 42*(6), 2243–2281.

Hariz, S. B., J. J. Wylie, and Q. Zhang (2007). Optimal rate of convergence for nonparametric

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

change-point estimators for nonstationary sequences. *The Annals of Statistics 35*(4), 1802–1826.

Hawkins, D. M. (1977). Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association 72*(357), 180–186.

Horváth, L., P. Kokoszka, and J. G. Steinebach (1999). Testing for changes in multivariate dependent observations with an application to temperature changes. *Journal of Multivariate Analysis 68*, 96–119.

Jiang, Y., X. Wang, C. Wen, Y. Jiang, and H. Zhang (2024). Nonparametric two-sample tests of high dimensional mean vectors via random integration. *Journal of the American Statistical Association 119*(545), 701–714.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics 43*(6), 2451–2483.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics 29*(2), 295–327.

Kao, C., L. Trapani, and G. Urga (2018). Testing for instability in covariance structures. *Bernoulli 24*(1), 740–771.

Kaul, A., H. Zhang, K. Tsampourakis, and G. Michailidis (2023). Inference on the change point under a high dimensional covariance shift. *Journal of Machine Learning Research 24*(168), 1–68.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

Lai, T. L. and H. Xing (2013). Stochastic change-point ARX-GARCH models and their applications to econometric time series. *Statistica Sinica 23*, 1573–1594.

Lee, J., S. Nemati, I. Silva, B. A. Edwards, J. P. Butler, and A. Malhotra (2012). Transfer entropy estimation and directional coupling change detection in biomedical time series. *Biomedical Engineering Online 11*, 1–17.

Lee, S., G. R. Abecasis, M. Boehnke, and X. Lin (2014). Rare-variant association analysis: study designs and statistical tests. *The American Journal of Human Genetics 95*(1), 5–23.

Lee, S., M. H. Seo, and Y. Shin (2015, 02). 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.

Li, J., P. Fearnhead, P. Fryzlewicz, and T. Wang (2024). Automatic change-point detection in time series via deep learning. *Journal of the Royal Statistical Society Series B: Statistical Methodology 86*(2), 273–285.

Li, L. and J. Li (2023). Online change-point detection in high-dimensional covariance structure with application to dynamic networks. *Journal of Machine Learning Research 24*(51), 1–44.

Liu, B., X. Zhang, and Y. Liu (2022). High dimensional change point inference: Recent developments and extensions. *Journal of Multivariate Analysis 188*, 104833.

Ma, S. and L. Su (2018). Estimation of large dimensional factor models with an unknown number of breaks. *Journal of Econometrics 207*(1), 1–29.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

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

Pilliat, E., A. Carpentier, and N. Verzelen (2023). Optimal multiple change-point detection for high-dimensional data. *Electronic Journal of Statistics 17*(1), 1240–1315.

Shu, L., Y. Chen, W. Zhang, and X. Wang (2022). Spatial rank-based high-dimensional change point detection via random integration. *Journal of Multivariate Analysis 189*, 104942.

Steland, A. (2020). Testing and estimating change-points in the covariance matrix of a high-dimensional time series. *Journal of Multivariate Analysis 177*, 104582.

van Breukelen, M., R. P. Duin, D. M. Tax, and J. Den Hartog (1998). Handwritten digit recognition by combined classifiers. *Kybernetika 34*(4), 381–386.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, Volume 47. Cambridge University Press.

Wang, D., Y. Yu, and A. Rinaldo (2021a). Optimal change point detection and localization in sparse dynamic networks. *The Annals of Statistics 49*(1), 203–232.

Wang, D., Y. Yu, and A. Rinaldo (2021b). Optimal covariance change point localization in high dimensions. *Bernoulli 27*(1), 554–575.

Wang, G. and L. Feng (2023). Computationally efficient and data-adaptive changepoint inference in high dimension. *Journal of the Royal Statistical Society Series B: Statistical Methodology 85*(3), 936–958.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

Wang, R., C. Zhu, S. Volgushev, and X. Shao (2022). Inference for change points in high-dimensional data via selfnormalization. *The Annals of Statistics 50*(2), 781–806.

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

Wu, Y., Y. Qin, and M. Zhu (2020). High-dimensional covariance matrix estimation using a low-rank and diagonal decomposition. *Canadian Journal of Statistics 48*(2), 308–337.

Yu, M. and X. Chen (2021, 12). Finite sample change point inference and identification for high-dimensional mean vectors. *Journal of the Royal Statistical Society Series B: Statistical Methodology 83*(2), 247–270.

Zhang, N. R., D. O. Siegmund, H. Ji, and J. Z. Li (2010, 06). Detecting simultaneous changepoints in multiple sequences. *Biometrika 97*(3), 631–645.

Zhou, H., H. Zhu, and X. Wang (2025). High-dimensional data analysis: Change point detection via bootstrap MOSUM. *Journal of Multivariate Analysis 209*, 105449.

Canhuang Xu, Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei, Anhui, P.R.China.

ORCID: 0009-0005-5089-9901

E-mail: xch02@mail.ustc.edu.cn

Lei Shu, School of Big Data and Statistics, Anhui University, Hefei, Anhui, P.R.China.

ORCID: 0000-0003-1901-2143

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0327

E-mail: leis@ahu.edu.cn

Yu Chen, School of Public Policy & Management, University of Science and Technology of China, Hefei, Anhui, P.R.China.

ORCID: 0000-0002-2438-3451

E-mail: cyu@ustc.edu.cn

Qing Yang, School of Management, University of Science and Technology of China, Hefei, Anhui, P.R.China.

ORCID: 0009-0002-8015-1978

E-mail: yangq@ustc.edu.cn

Statistica Sinica