A General U-Statistic Framework for High-Dimensional Multiple Change-Point Analysis

Bin Liu* and Yufeng Liu†

Abstract

High-dimensional change-point analysis is essential in modern statistical inference. However, existing methods are often designed either for specific parameters (e.g., mean or variance) or for particular tasks (e.g., testing or estimation), making them difficult to generalize. Moreover, they typically rely on restrictive distributional assumptions, limiting their robustness to heavy-tailed data. We propose a unified framework for testing, estimating, and inferring multiple change points in high-dimensional data. Our approach leverages a two-sample U-statistic within a moving window, allowing flexible kernel function selection to accommodate structural changes in general parameters such as variance changes or robust statistics. For testing, we develop an ℓ_∞-norm-based statistic with a high-dimensional multiplier bootstrap procedure, achieving minimax-optimal power under sparse alternatives. For estimation, we construct an initial estimator for the change-point number and locations and refine it using the U-statistic Projection Refinement Algorithm (U-PRA), attaining minimax-optimal localization rates. We further derive the asymptotic distribution of refined estimators, enabling valid confidence interval construction. Extensive numerical experiments demonstrate the superior performance of our method

*Department of Statistics and Data Science, School of Management at Fudan University, China; E-mail:liubin0145@gmail.com;

†Department of Statistics and Operations Research, Department of Genetics, and Department of Biostatistics, Carolina Center for Genome Sciences, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, U.S.A; E-mail: yfliu@email.unc.edu

across various settings, including heavy-tailed distributions. Applications to genomic copy number variation and financial time series data highlight its practical utility.

Keywords: Bootstrap, Efficient computation, Heavy tail, Structure change

1 Introduction

Modern statistical applications frequently involve high-dimensional data sequences, where the dimensionality significantly exceeds the sample size. Due to the complex nature of data generation processes, heterogeneity is a prevalent characteristic in such datasets. These structural changes, occurring at unknown locations within the observed data sequence, are referred to as *change points*. Detecting and identifying these change points is crucial in many real-world applications, such as risk management in finance, disease diagnosis in biology, and neural activity analysis in neuroscience.

To formalize the problem, let $\boldsymbol{X} = (X_1, \cdots, X_d)^\top$ be a d-dimensional random vector, and let $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ be an independent sequence of observations drawn from $\boldsymbol{X}$, where $\boldsymbol{X}_t = (X_{t,1}, \cdots, X_{t,d})^\top$; $t = 1, \ldots, n$. Suppose $\boldsymbol{\theta} \in \mathbb{R}^d$ represents an underlying parameter of interest, and consider the following multiple change-point model:

$$\boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_{\gamma_1} \neq \boldsymbol{\theta}_{\gamma_1+1} = \cdots = \boldsymbol{\theta}_{\gamma_2} = \cdots = \boldsymbol{\theta}_{\gamma_{M_0}} \neq \boldsymbol{\theta}_{\gamma_{M_0+1}} = \cdots = \boldsymbol{\theta}_n, \tag{1.1}$$

where $0 < \gamma_1 < \cdots < \gamma_{M_0} < n$ are the unknown change-point locations such that $\boldsymbol{\theta}_t = \boldsymbol{\theta}^{(m)}$ for $\gamma_{m-1} < t \leq \gamma_m$ with $1 \leq m \leq M_0 + 1$. For notational convenience, we define $\gamma_0 = 0$ and $\gamma_{M_0+1} = n$. Given this model, high-dimensional change-point analysis typically involves three primary research questions: (1) detecting the presence of change points, (2) estimating multiple change points, and (3) conducting statistical inference for the estimated change-point locations.

The first problem, single change-point detection assuming $M_0 \leq 1$, has been extensively studied, particularly in the context of detecting mean shifts with $\boldsymbol{\theta}_t = \mathbb{E}[\boldsymbol{X}_t]$. In

high-dimensional settings, the change may occur in either a sparse subset of components or across a dense set of coordinates with small individual signals. Consequently, recent research has focused on strategies to efficiently aggregate information from affected components to enhance detection power. Existing detection methods are typically categorized into two classes: those that assume prior knowledge about the sparse or dense nature of the alternative hypothesis, and those that employ adaptive procedures to determine the alternative structure in a data-driven manner. For instance, sparse change-point detection using ℓ_∞-norm aggregation was explored in Jirak (2015); Yu and Chen (2021), while dense alternatives leveraging ℓ_2-norm aggregation were studied in Horváth and Hušková (2012); Wang et al. (2022). A more general adaptive framework combining both perspectives was developed in Cho (2016); Aston and Kirch (2018); Liu et al. (2020); Zhang et al. (2022); Wang and Feng (2023). Additionally, Enikeeva and Harchaoui (2019); Liu et al. (2021) investigated the minimax optimality of high-dimensional mean change detection. For a more comprehensive discussion, see the review by Liu et al. (2022). However, most of these methods assume that the error terms in the data are sub-Gaussian, sub-exponential or require stringent moment conditions. When these assumptions are violated, such as in the presence of heavy-tailed distributions or outliers, mean-based statistics can suffer from inflated false detection rates and reduced power, limiting their practical reliability.

Beyond detecting a single change point, multiple change-point estimation with $M_0 \geq 1$ aims to determine both the number and locations of structural breaks. Two dominant approaches for this problem include segmentation-based and penalization-based methods. The first category extends single change-point detection techniques by incorporating recursive partitioning algorithms, such as binary segmentation (Liu et al. (2020); Cho (2016)), wild binary segmentation (WBS) (Wang and Samworth (2018); Zhang et al. (2022)), and seed binary segmentation (Jiang et al. (2023)). While computationally efficient, these methods often lack theoretical guarantees or impose specific requirements on the strength

of the change-point signal and suffer from multiple testing issues, making it difficult to control false discovery rates. The second category treats multiple change points as unknown parameters and estimates them via penalized optimization techniques, such as the PELT algorithm (Killick et al. (2012)) or fused LASSO (Harchaoui and Lévy-Leduc (2010)). While these methods offer structured estimators, they generally assume that change points exist a priori and do not provide a formal hypothesis testing framework for their presence. Moreover, most of these methods are designed primarily for mean shifts and are not readily extendable to other parameters such as variance changes or robust statistics.

The third major challenge in high-dimensional change-point analysis is to conduct statistical inference for estimated change points. While numerous methods have been developed for detecting and estimating change points, relatively few offer confidence intervals or theoretical guarantees for their estimates. The difficulty in constructing confidence intervals lies in establishing optimality and characterizing the asymptotic distribution of change-point estimators. Existing results are largely limited to univariate settings, including recent studies by Verzelen et al. (2023), Eichinger and Kirch (2018), and Cho and Kirch (2022), as well as classical work by Bai and Perron (1998). Extending these methods to high-dimensional settings remains particularly challenging, and only recent progress has been made in this direction. Moreover, the construction of confidence intervals for change points has primarily focused on mean vector parameters, with most approaches relying on either an ℓ_2-loss-based framework (Kaul et al. (2021); Kaul and Michailidis (2025)) or mean-based CUSUM statistics (Chen et al. (2022)). These methodologies are inherently difficult to generalize to broader parameter settings and typically assume that the data follow a light-tailed distribution. Consequently, when applied to heavy-tailed data, these methods often yield unreliable confidence intervals.

Despite significant progress in high-dimensional change-point analysis, several challenges remain. First, most existing methods are designed for mean-based change detection

and rely on strong distributional assumptions, limiting their applicability when the underlying data deviate from Gaussianity. Second, while research on multiple change-point estimation has advanced, many existing techniques either lack theoretical guarantees or do not provide a formal hypothesis testing framework for determining the presence of change points. Moreover, these methods are often tailored to detecting mean shifts and struggle to generalize to other structural changes, such as variance shifts or robust statistics, particularly in the presence of heavy-tailed distributions or outliers. Beyond estimation, statistical inference for multiple change points remains an open challenge. Confidence interval construction has predominantly focused on mean vector parameters and often imposes restrictive assumptions on the error distribution. Many existing methods tend to break down under heavy-tailed data or in the presence of outliers, making them unreliable in broader applications. Furthermore, extending current inference frameworks to accommodate more general parameters remains a significant hurdle. This limitation highlights the need for a more general and statistically rigorous framework that extends beyond mean-based change point detection to capture broader forms of structural heterogeneity.

To overcome these challenges, we propose a novel methodology for detecting, estimating, and inferring multiple change points in high-dimensional data. Our approach leverages a two-sample U-statistic within a moving window framework, enabling flexible kernel function selection. By appropriately choosing the kernel, our method extends beyond traditional mean-based change detection to accommodate shifts in variance, robust mean, and robust variance. While existing methods have also investigated change-point detection and estimation for variance and covariance structures such as in Cho and Fryzlewicz (2015); Dette et al. (2022), they often focus on specific aspects under different assumptions. In contrast, our framework provides a unified and flexible approach that seamlessly integrates various forms of structural change, making it more general than many existing methods. Our contributions can be summarized as follows.

1. **Multiple Change-Point Testing**: We develop a multiple change-point testing framework based on a two-sample U-statistic within a moving window setting, incorporating an ℓ_∞-norm-based test statistic tailored for sparse alternatives. By appropriately selecting the kernel function, our method generalizes beyond mean shifts to detect changes in variance, robust mean, and robust variance. To approximate the limiting distribution of the testing statistic, we construct a high-dimensional multiplier bootstrap method that effectively captures the covariance structure among data components and mitigates dependencies induced by the moving window mechanism. Theoretically, we establish that our proposed test controls the significance level under the null hypothesis and achieves asymptotic power one under the alternative, even allowing the dimensionality d to grow exponentially with the sample size n. Notably, under sparse alternatives, our method attains minimax-optimal detection rates. Furthermore, our test imposes minimal moment conditions on the data by shifting regularity conditions onto the kernel function, enabling robust change-point detection in heavy-tailed distributions. By applying a robust kernel, such as the Wilcoxon-Mann-Whitney kernel, our method extends change-point testing to a robustness framework, a direction that, to our knowledge, remains unexplored in the existing literature.

2. **Optimal Estimation of Change-Point Locations and Numbers**: Inspired by Eichinger and Kirch (2018) in low-dimensional settings, we construct an initial estimator for the number and locations of change points in high-dimensional scenarios. We theoretically demonstrate that under mild model assumptions, this estimator correctly identifies the number of change points with probability tending to one, with location estimates achieving near-optimal convergence rates, up to a logarithmic factor $\log(d)$. To further refine the estimation, we introduce a U-statistic projection refinement algorithm (U-PRA). This method first identifies the subset of components at each change point with high probability and then projects the test statistic onto this subspace along the U-statistic's direction, reducing the

problem to a univariate change-point estimation. A local search then fine-tunes the final estimates. We rigorously establish that the refined estimators attain the minimax-optimal convergence rate. This framework provides optimal change-point estimation not only for mean vectors but also for variance and robust statistics, thereby solving a broad class of change-point estimation problems. Our simulations demonstrate that even in heavy-tailed settings with outliers, our method remains accurate and robust, significantly outperforming existing techniques.

3. **Confidence Interval Construction for Multiple Change Points**: Building on the refined estimation in the second step, we derive the asymptotic distribution of each change-point estimator within the two-sample U-statistic framework, establishing that it follows the arg max of a drifted weighted Brownian motion process. This result enables the construction of valid confidence intervals for individual change points as their signal strength approaches zero. Additionally, we prove that the limiting distributions of different change-point estimators are asymptotically independent, allowing us to construct joint confidence intervals for multiple change points. Importantly, since our framework is based on U-statistics, the limiting distribution exhibits a novel form involving weighted sums of three independent Brownian motions, a result established here for the first time in high-dimensional change-point analysis. By appropriately selecting the kernel function, our method extends confidence interval construction beyond mean changes to variance, robust mean, and robust variance. Our numerical experiments confirm that even in heavy-tailed settings, our constructed confidence intervals achieve coverage probabilities close to their nominal levels, making a significant advancement in high-dimensional change-point inference. Finally, we conduct extensive numerical studies validating both our theoretical results and methodological effectiveness. Our analysis on genomic copy number variation and S&P 500 datasets further highlights the practical utility of our proposed framework.

Our study is closely related to Liu et al. (2020), which employs U-statistics for change-

point detection. However, Liu et al. (2020) relies on one-sample U-statistics, whereas we adopt two-sample U-statistics. Moreover, Liu et al. (2020) focused exclusively on single change-point detection and primarily aimed to develop data-driven testing methods for both sparse and dense alternative structures. They did not address multiple change-point detection, estimation, or the construction of confidence intervals. Yu and Chen (2022) also used two-sample U-statistics for single change-point testing, achieving high power in single change-point global testing. However, their method cannot be directly extended to our general framework for multiple change-point detection, necessitating the use of a moving window approach. Our approach incorporates the moving window technique, making it more relevant to Eichinger and Kirch (2018) and Chen et al. (2022). Eichinger and Kirch (2018) first established the theoretical foundation for multiple change-point inference in mean sequences using moving windows in univariate time series, while Chen et al. (2022) extended this approach to high-dimensional mean sequences. Unlike these studies, we construct a multiple change-point detection statistic in high-dimensional settings using two-sample U-statistics. When an appropriate kernel function is chosen, our method includes the mean model as a special case. Furthermore, we develop a specialized multiplier bootstrap method to approximate the limiting distribution of our test statistic. The high-dimensional setting, the nonlinearity of U-statistics, and the dependence structure induced by the moving window approach introduce significant theoretical challenges that were not encountered in the aforementioned studies.

The rest of this paper is organized as follows. In Section 2, we introduce our methodology for multiple change point testing and introduce a multiplier bootstrap based procedure for approximating the limiting null distribution in Section 3. The estimation and inference of multiple change points are given in Section 4. In Section 5, extensive numerical studies are investigated. Section 6 presents the application of the proposed method to S&P 500 real data analysis. Discussions are given in Section 7. The supplementary materials provide

proofs of the theoretical results and additional numerical results.

We end this section with some notations. Let $\mathcal{X} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$. We define the ℓ_p norm as $\|\boldsymbol{v}\|_p = (\sum_{j=1}^d |v_j|^p)^{1/p}$ for $\boldsymbol{v} = (v_1, v_2, \cdots, v_d)^\top \in \mathbb{R}^d$. For $p = \infty$, $\|\boldsymbol{v}\|_\infty = \max_{1\le j\le d} |v_j|$. For $p = 0$, $\|\boldsymbol{v}\|_0 := \#\{j : v_j \neq 0\}$, where $\#\{\mathcal{S}\}$ denotes the cardinality of a set $\mathcal{S}$. For two real numbered sequences a_n and b_n, we set $a_n = O(b_n)$ if there exits a constant C such that $|a_n| \le C|b_n|$ for a sufficiently large n; $a_n = o(b_n)$ if $a_n/b_n \to 0$ as $n \to \infty$; $a_n \asymp b_n$ if there exists constants c and C such that $c|b_n| \le |a_n| \le C|b_n|$ for a sufficiently large n. For a sequence of random variables (r.v.s) $\{\xi_1, \xi_2, \cdots\}$, we set $\xi_n \to \xi$ if ξ_n converges to ξ in probability as $n \to \infty$. We also denote $\xi_n = o_p(1)$ if $\xi_n \to 0$.

2 Testing existence of multiple change points

In this section, we present our methodology for the multiple change point detection. Before presenting our methodology, we introduce some model settings. Let $h(x, y) \in \mathbb{R}^1$ be some user specified two-sample kernel which is antisymmetric in the sense that $h(y, x) = -h(x, y)$. In the following sections, for $\boldsymbol{x} = (x_1, \ldots, x_d)$ and $\boldsymbol{y} = (y_1, \ldots, y_d)$, we set $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}) := \big(h(x_1, y_1), \ldots, h(x_d, y_d)\big) \in \mathbb{R}^d$. Let $\theta_j^{(m)} := \mathbb{E}h(X_{\gamma_m,j}, X_{\gamma_m+1,j})$ be the signal jump for $m = 1, \ldots, M_0$ and $j = 1, \ldots, d$ and define $\boldsymbol{\theta}^{(m)} := \mathbb{E}\boldsymbol{h}(\boldsymbol{X}_{\gamma_m}, \boldsymbol{X}_{\gamma_m+1})$ as the $\mathbb{R}^d$-dimensional signal jump at the γ_m-th change point. It is worth noting that in many ultrahigh-dimensional change-point detection problems, the associated kernel functions are often antisymmetric. Moreover, by appropriately selecting some kernel functions, our framework can recover several existing high-dimensional change-point detection methods for specific parameters. For instance, setting $h(x, y) = y - x$ leads to $\theta_j^{(m)} := \mathbb{E}(X_{\gamma_m+1,j} - X_{\gamma_m,j})$ representing the mean shift between segments. In subsequent sections, we will provide further examples to illustrate the versatility and broad applicability of our approach. We set $\Pi_m = \{j \in \{1, \cdots, d\} : \theta_j^{(m)} \neq 0\}$ as the set of coordinates with a change point at the change point γ_m and $s_m = |\Pi_m|$ as the total number of coordinates.

Our first goal is to present a new test to determine the presence of multiple breaks. To formally establish the multiple change-point detection framework, we define the signal jump at time t as:

$$\boldsymbol{\theta}_t := \mathbb{E}\boldsymbol{h}(\boldsymbol{X}_t, \boldsymbol{X}_{t+1}), \quad \text{for} \quad t = 1, \ldots, n-1, \tag{2.1}$$

where $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y})$ is a user-specified kernel function that quantifies the structural difference between two consecutive observations. Based on this definition, we formulate the hypothesis testing problem as follows:

$$\begin{aligned} &\mathbf{H}_0 : \boldsymbol{\theta}_t = 0, \quad \forall t = 1, \ldots, n-1, \\ &\mathbf{H}_1 : \exists\ \gamma_1 < \gamma_2 < \cdots < \gamma_{M_0}, \text{ such that } \boldsymbol{\theta}_{\gamma_m} \neq 0, \quad \forall m = 1, \ldots, M_0. \end{aligned} \tag{2.2}$$

Under the null hypothesis $\mathbf{H}_0$, the entire sequence remains stable, meaning that no significant structural changes occur. In contrast, under the alternative hypothesis $\mathbf{H}_1$, there exist at least one or more change-points γ_m where the underlying process exhibits abrupt shifts.

Note that our testing method allows the number of change points M_0 scale with the sample observations n which covers the single change point testing as a special case when $M_0 = 1$. Our method is based on the moving window based procedure proposed in Eichinger and Kirch (2018). Specifically, let $k \in \{1, \ldots, n\}$ be a candidate search location and $1 \leq G \leq n$ be some bandwidth. For each search window $[s_k, e_k] := \{k - G + 1, k - G + 2, \ldots, k, k + 1, \ldots, k + G\}$ and for each coordinate $j \in \{1, \ldots, d\}$, we define the following two-sample U-statistic based testing statistics:

$$T_j(k) := \frac{1}{G^{3/2}} \sum_{t_1 = k-G+1}^{k} \sum_{t_2 = k+1}^{k+G} h(X_{t_1,j}, X_{t_2,j}), \quad \text{for} \quad j = 1, \ldots, d. \tag{2.3}$$

The bandwidth parameter G satisfies $G = o(n)$ and $G \to \infty$. Intuitively, the bandwidth G should be sufficiently large to ensure that each search interval contains an adequate number of observations. However, G should not be so large as to allow more than one change point within any given scanning interval. Details on this condition will be provided in the theoretical analysis section.

Next, we first explain why the aforementioned test statistic can effectively detect multi-

ple change points in the data. It can be proven that by selecting an appropriate bandwidth G, the sliding window $[s_k, e_k]$ contains at most one change point, and these sliding windows with a single change point determine the testing and estimation of the change point. Suppose the bandwidth G is chosen properly such that there exists only one change point, say γ_m, in $[s_k, e_k]$. To further elucidate this issue, we first introduce Hoeffding's decomposition (Hoeffding (1948)) for the kernel $h(x, y)$. In particular, for each coordinate $j = 1, \ldots, d$ and change point γ_m, by Hoeffding's decomposition, we can decompose $h(x, y)$ as follows:

$$h(X_{t_1,j}, X_{t_2,j}) = \theta_j^{(m)} + h_{1,j}(X_{t_1,j}) + h_{2,j}(X_{t_2,j}) + g_j(X_{t_1,j}, X_{t_2,j}), \tag{2.4}$$

where $\theta_j^{(m)} = \mathbb{E}h(X_{\gamma_m,j}, X_{\gamma_m+1,j})$, and for any $\gamma_{m-1} < t_1 \leq \gamma_m, \gamma_m + 1 \leq t_2 \leq \gamma_{m+1}$, the three terms $h_{1,j}(X_{t_1,j})$, $h_{2,j}(X_{t_2,j})$ and $g_j(X_{t_1,j}, X_{t_2,j})$ are defined as

$$\begin{aligned} &h_{1,j}(X_{t_1,j}) = \mathbb{E}[h(X_{t_1,j}, X_{t_2,j})|X_{t_1,j}] - \theta_j^{(m)}, \quad h_{2,j}(X_{t_2,j}) = \mathbb{E}[h(X_{t_1,j}, X_{t_2,j})|X_{t_2,j}] - \theta_j^{(m)}, \\ &\text{and} \quad g_j(X_{t_1,j}, X_{t_2,j}) = h(X_{t_1,j}, X_{t_2,j}) - \theta_j^{(m)} - h_{1,j}(X_{t_1,j}) - h_{2,j}(X_{t_2,j}). \end{aligned} \tag{2.5}$$

Hence, for any $k \in [s_k, e_k]$ with only one change point $\gamma_m \in [s_k, e_k]$, $T_j(k)$ can have the following explicit form:

$$T_j(k) = \begin{cases} \delta_j^{(m)}(k) + \underline{R}_{1,j}(k) + \underline{R}_{2,j}(k) + R_{3,j}(k), & \gamma_{m-1} < k \leq \gamma_m, \\ \delta_j^{(m)}(k) + \overline{R}_{1,j}(k) + \overline{R}_{2,j}(k) + R_{3,j}(k), & \gamma_m < k \leq \gamma_{m+1}, \end{cases} \tag{2.6}$$

where

$$\delta_j^{(m)}(k) = \frac{G(k + G - \gamma_m)}{G^{3/2}} \theta_j^{(m)} \mathbf{1}\{\gamma_{m-1} < k \leq \gamma_m\} + \frac{G(\gamma_m - k + G)}{G^{3/2}} \theta_j^{(m)} \mathbf{1}\{\gamma_m < k \leq \gamma_{m+1}\}$$

is the called signal jump function and

$$\begin{aligned} \underline{R}_{1,j}(k) &:= \frac{k + G - \gamma_m}{G^{3/2}} \sum_{t_1=k-G+1}^{k} h_{1,j}(X_{t_1,j}) + \frac{G}{G^{3/2}} \sum_{t_2=\gamma_m+1}^{k+G} h_{2,j}(X_{t_2,j}), \\ \underline{R}_{2,j}(k) &:= \frac{1}{G^{3/2}} \sum_{t_1=k-G+1}^{k} \sum_{t_2=k+1}^{\gamma_m} h(X_{t_1,j}, X_{t_2,j}), \quad \text{for} \quad \gamma_{m-1} < k \leq \gamma_m; \\ \overline{R}_{1,j}(k) &:= \frac{G}{G^{3/2}} \sum_{t_1=k-G+1}^{\gamma_m} h_{1,j}(X_{t_1,j}) + \frac{\gamma_m - k + G}{G^{3/2}} \sum_{t_2=k+1}^{k+G} h_{2,j}(X_{t_2,j}), \\ \overline{R}_{2,j}(k) &:= \frac{1}{G^{3/2}} \sum_{t_1=\gamma_m+1}^{k} \sum_{t_2=k+1}^{k+G} h(X_{t_1,j}, X_{t_2,j}), \quad \text{for} \quad \gamma_m < k \leq \gamma_{m+1}; \\ R_{3,j}(k) &:= \frac{1}{G^{3/2}} \sum_{t_1=k-G+1}^{k} \sum_{t_2=k+1}^{k+G} g_j(X_{t_1,j}, X_{t_2,j}), \quad \text{for} \quad \gamma_{m-1} < k \leq \gamma_{m+1}. \end{aligned}$$

Note that $\underline{R}_{1,j}(k)$ and $\overline{R}_{1,j}(k)$ are the leading term (random noise) of the Hoeffding's decomposition, $R_{3,j}(k)$ is the negligible term (random bias), and $\underline{R}_{2,j}(k)$ and $\overline{R}_{2,j}(k)$ involve data with no change point. Based on the above decomposition, we see that $|\delta_j^{(m)}(k)|$ obtains its maximum value at the true change point γ_m. Moreover, under some regular conditions, we can prove that

$$\max_{1\le j\le d}\max_{G\le k\le n-G}\Big(|\underline{R}_{1,j}(k)|+|\underline{R}_{2,j}(k)|+|\overline{R}_{1,j}(k)|+|\overline{R}_{2,j}(k)|+|R_{3,j}(k)|\Big)=O_p(\sqrt{\log(nd)}).$$

The above observation suggests that by imposing certain signal-to-noise ratio requirements on the change-point signals $\theta_j^{(m)}$, we can ensure that the signal function $\delta_j^{(m)}(k)$ dominates the random noise and biases, thereby enabling accurate detection and identification of change points.

To further illustrate the novelty of our proposed methodology, we provide several specific examples of kernel functions $h(x,y)$ used in the construction of the two-sample U-statistics.

Example 1. When the kernel function is defined as $h(x,y)=y-x$, the resulting U-statistic focuses on detecting changes in the mean of the distribution. Specifically, in this case, we have $\theta_j^{(m)}=\mathbb{E}(X_{\gamma_m+1}-X_{\gamma_m})$ as the mean shift and $T_j(k)$ reduces to

$$T_j(k)=\frac{1}{\sqrt{G}}\Big(\sum_{t_1=k-G+1}^{k}X_{t_1,j}-\sum_{t_2=k+1}^{k+G}X_{t_2,j}\Big),\quad \text{for}\quad j=1,\ldots,d.$$

Thus, our method effectively tackles the problem of detecting multiple change points in ultra-high-dimensional mean parameters, which has been recently extensively studied in Eichinger and Kirch (2018); Chen et al. (2022).

Example 2. To handle cases where the data may contain outliers or heavy-tailed distributions, we consider a robust kernel function: $h(x,y)=\text{sign}(y-x)$, where $\text{sign}(x)=\mathbf{1}\{x\ge 0\}-\mathbf{1}\{x<0\}$ is the indicator function. This kernel function leads to a **rank-based U-statistic**, making the test resistant to the influence of outliers when detecting data with mean shifts. To see this, $X_{\gamma_m,j}=\mu_j^{(m)}+\epsilon_j$ and $X_{\gamma_m+1,j}=\mu_j^{(m+1)}+\epsilon_j'$ and ϵ_j, ϵ_j' are the random erros with mean zeros. Let $\Delta_j^{(m)}:=\mu_j^{(m+1)}-\mu_j^{(m)}$ be the signal jump at the m-th

change point for the j-th coordidate and $\xi_j = \epsilon'_j - \epsilon_j$. By the defintion of $\theta_j^{(m)}$, we have

$$\begin{aligned}\theta_j^{(m)} &= \mathbb{E}h(X_{\gamma_m,j}, X_{\gamma_m+1,j}) \\ &=_{(1)} \mathbb{E}[\mathbf{1}\{\mu_j^{(m+1)} - \mu_j^{(m)} + \epsilon'_j - \epsilon_j \geq 0\}] - \mathbb{E}[\mathbf{1}\{\mu_j^{(m+1)} - \mu_j^{(m)} + \epsilon'_j - \epsilon_j < 0\}] \\ &=_{(2)} \mathbb{E}\mathbf{1}\{\xi_j \geq -\Delta_j^{(m)}\} - \mathbb{E}\mathbf{1}\{\xi_j < -\Delta_j^{(m)}\} =_{(3)} -2(F_j(-\Delta_j^{(m)}) - F(0)) \approx_{(4)} 2f_j(0)\Delta_j^{(m)},\end{aligned} \tag{2.7}$$

where $F_j(x)$ and $f_j(x)$ are the CDF and pdf of ξ_j respectively, Equation (3) is derived from the symmetry of ξ_j, and (4) follows from a Taylor expansion. Thus, $\theta_j^{(m)}$ with vanishing signal jumps encapsulates information about the signal jump and shares equivalent magnitudes with $\Delta_j^{(m)}$, indicating that any distortion of the signal through the sign kernel is limited to a multiplicative constant.

Example 3. Consider a sequence of observations $\{X_{t,j}\}_{t=1}^n$ with $j = 1, \ldots, d$, where the variance changes at γ_m:

$$X_{t,j} = \begin{cases} \sigma_j^{(m)} \times \epsilon_{t,j} \sim (0, (\sigma_j^{(m)})^2), & \gamma_{m-1} < t \leq \gamma_m, \\ \sigma_j^{(m+1)} \times \epsilon'_{t,j} \sim (0, (\sigma_j^{(m+1)})^2), & \gamma_m < t \leq \gamma_{m+1}, \end{cases}$$

where $(\sigma_j^{(m+1)})^2 = \lambda_j^{(m)}(\sigma_j^{(m)})^2$ with $\lambda_j^{(m)}$ represents the relative change in variance and $\epsilon_{t,j}$ or $\epsilon'_{t,j}$ are i.i.d random variables with mean zero and variance one. We can use the kernel $h(x, y) = y^2 - x^2$. In this case, $\theta_j^{(m)} = (\lambda_j^{(m)} - 1)\sigma_{1,j}^2$ denotes the value of the variance shift. If we use $h(x, y) = \text{sign}(y^2 - x^2)$, we can construct a robust variance change-point detection method that is more effective for heavy-tailed data. Specifically, similar to Example 2, we can prove that $\theta_j^{(m)}$ is proportional to the relative changes of variances with

$$\theta_j^{(m)} = \mathbb{E}h(X_{\gamma_m,j}, X_{\gamma_m+1,j}) \approx 2f_{\epsilon_{t,j}^2/\epsilon_{t,j}'^2}(1)(\lambda_j^{(m)} - 1),$$

where $f_{\epsilon_{t,j}^2/\epsilon_{t,j}'^2}(x)$ is the pdf of $\epsilon_{t,j}^2/\epsilon_{t,j}'^2$.

Let $\boldsymbol{T}(k) = (T_1(k), \ldots, T_d(k))^\top$. Note that we aim to propose a powerful testing procedure that is tailored to sparse alternatives. Hence, we consider the ℓ_∞-norm aggregation of $\boldsymbol{T}(k)$ which has been used in Jirak (2015); Yu and Chen (2021, 2022). Specifically, our final testing statistic is defined as follows:

$$W = \max_{G \le k \le n-G} \max_{1 \le j \le d} |T_j(k)| = \max_{G \le k \le n-G} \|\boldsymbol{T}(k)\|_\infty. \tag{2.8}$$

Since W is constructed by adopting the ℓ_∞ norm, it is more sensitive to sparse signals with strong perturbations on a small number of coordinates. More importantly, we do not employ the CUSUM statistic; instead, we utilize information of two-sample U statistics that constructs the testing statistic. This ensures that the test may not use the data points directly and may maintain high power while also being robust to heavy-tailed data and outliers.

Given the significance level α, we reject the null hypothesis once W exceeds a critical value, say $c_{\alpha,W}$. In the following section, we propose a technique to obtain the critical values $c_{\alpha,W}$ which is based on the multiplier bootstrap such that the testing can control the type I error while being consistent under sparse alternatives.

Remark 2.1 (**Computational cost**). For any search interval $[s, e] \subset [1, T]$ and the bandwith G, define $F_j([s_1, e_1], [s_2, e_2]) = \sum_{t_1=s_1}^{e_1} \sum_{t_2=s_2}^{e_2} h(X_{t_1,j}, X_{t_2,j})$. By the definition of $T_j(k)$ in (2.3), we can have the following iterative formula:

$$\begin{aligned} T_j(k+1) = T_j(k) - \frac{1}{G^{3/2}}\big(&F_j([k-b, k-b], [k+1, k+1]) + F_j([k-b, k-b], [k+2, k+b]) \\ &- F_j([k-b+1, k], [k+1, k+1])\big) + \frac{1}{G^{3/2}}\big(F_j([k-b+1, k], [k+b+1, k+b+1]) \\ &+ F_j([k+1, k+1], [k+2, k+b]) + F_j([k+1, k+1], [k+b+1, k+b+1]\big). \end{aligned}$$

Hence, we can calculate the above statistic $\{T_j(k), k = G, \ldots, n-G\}$ with $1 \le j \le d$ in linear time over $G \le k \le n-G$. Considering that our dataset is d-dimensional, the total computational time required to compute the test statistic has a complexity of $O(d(n-2G))$.

3 Testing using bootstrap quantiles

In the previous section, we introduce the ℓ_∞-norm based testing statistic W in (2.8). It is difficult to obtain its limiting distribution directly as n and d go to infinity. Under some extremely stringent model assumptions, we may derive that the theoretical asymptotic

distribution of this statistic is the Gumbel distribution. However, this typically imposes very strong requirements on the covariance structure of the data, which are often difficult to satisfy in practical applications. More importantly, the method relying on approximating the statistic's distribution with a Gumbel distribution is known to have slow convergence rates using such asymptotic results. This was also pointed out in Jirak (2015); Gösmann et al. (2022). Consequently, for small sample sizes, the critical values derived may result in somewhat imprecise approximations in practical settings. To overcome this problem, in this section, we introduce the multiplier bootstrap method for approximating the limiting distribution of W.

Bootstrap methods for U-statistics have been developed in low dimensions (Janssen, 1994; Bücher and Kojadinovic, 2016) and extended to high dimensions (Chernozhukov et al., 2013, 2017; Chen, 2018; Zhou et al., 2018; Jirak, 2015; Yu and Chen, 2021; Liu et al., 2020). Unlike existing literature, we propose a novel multiplier bootstrap procedure using two-sample U-statistics with moving windows for multiple change point detection. Let $e_1^b, \cdots, e_n^b$ be i.i.d $N(0,1)$ random variables with $b = 1, \cdots, B$. Then the b-th bootstrap version of the coordinate based testing statistic $T_j(k)$ can be defined as:

$$T_j^b(k) = \frac{1}{G^{3/2}} \sum_{t_1=k-G+1}^{k} \sum_{t_2=k+1}^{k+G} (e_{t_1} + e_{t_2}) h(X_{t_1,j}, X_{t_2,j}), \quad \text{for} \quad j = 1, \ldots, d. \tag{3.1}$$

Accordingly, the b-th bootstrap version of the final testing statistic is defined as:

$$W^b = \max_{G \le k \le n-G} \max_{1 \le j \le d} |T_j^b(k)| = \max_{G \le k \le n-G} \|\boldsymbol{T}^b(k)\|_\infty.$$

Given the significance level α, let $c_{\alpha,W} := \inf\{t \in \mathbb{R} : \mathbb{P}(W \le t) \ge 1-\alpha\}$ be the theoretical critical value for W. Based on the bootstrap samples $\{W^1, \cdots, W^B\}$, we estimate $c_{\alpha,W}$ by

$$\widehat{c}_{\alpha,W} = \inf\Big\{t : \frac{\sum_{b=1}^{B} \mathbb{1}\{W^b \le t\}}{B} \ge 1-\alpha\Big\}. \tag{3.2}$$

Hence, based on the estimated critical values $\widehat{c}_{\alpha,W}$, we define our test as follows

$$\Psi_{\alpha,W} = \mathbb{1}\big\{W \ge \widehat{c}_{\alpha,W}\big\}. \tag{3.3}$$

For W in (2.8), we reject $\mathbf{H}_0$ if and only if $\Psi_{\alpha,W} = 1$.

The multiplier bootstrap procedure captures the underlying covariance structure of the data in an automatic way. More importantly, as a result of the moving window procedure, $\{\boldsymbol{T}(k), k = G, \ldots, n-G\}$ possesses a complex correlation structure across different scanning positions (see Equation 3.8). We prove that the bootstrap-based test statistic can approximate this correlation structure in a fully data-driven manner, thereby circumventing the necessity of first estimating the covariance matrix and then generating the corresponding Gaussian random variables for approximation as used in Chen et al. (2022). Moreover, the bootstrap method is straightforward to implement, requiring no hyperparameter tuning or additional estimation of high-dimensional parameters. Furthermore, when estimating multiple change points, we can utilize the bootstrap-based critical values as thresholds for identifying change-point intervals, thereby avoiding the need to manually specify a threshold as used in Wang and Samworth (2018). Additionally, the proposed bootstrap procedure is not limited to mean parameters; it is also well-suited for detecting variance change points and robust mean shifts, highlighting its high flexibility and broad applicability.

Through simulations, we demonstrate that the bootstrap approach not only achieves high detection power in multiple change-point testing but also mitigates the issue of over- or under-estimating change points caused by manually chosen critical values. As a result, our method exhibits superior accuracy and stability compared to existing approaches.

Next, we provide theoretical results for the bootstrap based procedure under $\mathbf{H}_0$. Before that, we give some model assumptions. Let $\boldsymbol{X} = (X_1, \ldots, X_d)^\top \sim F(\boldsymbol{x})$ under $\mathbf{H}_0$ and $\boldsymbol{X}' = (X_1', \ldots, X_d')^\top$ be the independent copy of $\boldsymbol{X}$. Under $\mathbf{H}_0$, for the kernel $h(x, y)$, consider the following Hoeffding's decomposition

$$h(X_j, X_j') = 0 + h_{1,j}(X_j) + h_{2,j}(X_j') + g_j(X_j, X_j'), \ j = 1, \ldots, d, \tag{3.4}$$

where $h_{1,j}(X_j) := \mathbb{E}[h(X_j, X_j')|X_j]$, $h_{2,j}(X_j') := \mathbb{E}[h(X_j, X_j')|X_j']$, and $g_j(X_j, X_j') = h(X_j, X_j') - h_{1,j}(X_j) - h_{2,j}(X_j')$. Note that due to the antisymmetry property of $h(x, y)$, we have

$h_{2,j}(x) = -h_{1,j}(x)$. The following Assumptions (A.1) – (A.3) impose some moment conditions on the kernel. Specifically,

Assumption (A.1): There exists some constant b such that $\mathbb{E}(h_{1,j}(X_j))^2 \geq b$ for all $j = 1, \ldots, d$.

Assumption (A.2): There exists some constant D such that $\|h_{1,j}(X_j)\|_{\psi_1} \leq D$ for all $j = 1, \ldots, d$, where for any random variable X, $\|X\|_{\psi_1}$ is defined as $\|X\|_{\psi_1} := \inf\{C > 0 : \mathbb{E}\psi_1(|X|/C)|) \leq 1\}$ with $\psi_1(x) := \exp(x) - 1$.

Assumption (A.3): There exists some constant D such that $\mathbb{E}|h_{1,j}(X_j)|^{2+\ell} \leq D^\ell$ with $\ell = 1, 2$ for all $j = 1, \ldots, d$.

Assumption (A.4) For the bandwidth parameter G, we require

$$\frac{\log^7((n - 2G + 1)d)}{G} \to 0, \text{ as } (n, d) \to \infty.$$

Assumptions (A.1) – (A.4) are crucial for proving the results of Gaussian approximations for the ℓ_∞-norm-based testing statistic. Assumption (A.1) requires that the two sample kernel based U-statistics are non-degenerate. Assumption (A.2) requires that $h_{1,j}(X_j)$ follows sub-exponential distributions. Many bounded kernels such as Wilcoxon can satisfy this condition. Assumption (A.3) requires that $h_{1,j}(X_j)$ has bounded third and forth moments. Lastly, Assumption (A.4) characterizes the scaling relationship between (G, n, d) which allows the data dimension d can grow exponentially with n as long as $G \gg \log^7((n - 2G + 1)d)$ holds. Bases on the above assumptions, the following Theorem 3.1 justifies the validity of our proposed bootstrap procedure.

Theorem 3.1. Suppose Assumptions (**A.1**) – (**A.4**) hold. Under $\mathbf{H}_0$, we have

$$\sup_{z \in (0,\infty)} \left|\mathbb{P}(W \leq z) - \mathbb{P}(W^b \leq z|\mathcal{X})\right| = o_p(1), \text{ as } n, d \to \infty. \tag{3.5}$$

Theorem 3.1 shows that we can uniformly approximate the distribution of W using that of W^b. As a corollary, the following Corollary 3.1 shows that our proposed new test can control the Type I error asymptotically for any given pre-specified significant level α.

Corollary 3.1. Assume Assumptions (**A.1**)–(**A.4**) hold. Under $\mathbf{H}_0$, we have $\mathbb{P}(\Psi_{\alpha,W} = 1) \to \alpha$, as $n, d, B \to \infty$.

The proof of Theorem 3.1 relies on the essential modifications of high dimensional Gaussian approximation theorey of Chernozhukov et al. (2017) to the two-sample U-statistic-based multiple change point detection with moving window. The proof mainly proceeds in three steps. We only sketch the proof here. More detailed proof is presented in Appendix B.1 of supplementary materials. For any $1 \le t_1 \ne t_2 \le n$, define the $\mathbb{R}^d$ dimensional kernel and the corresponding leading and residual terms as:

$$\boldsymbol{h}(\boldsymbol{X}_{t_1}, \boldsymbol{X}_{t_2}) := \big(h(X_{t_1,1}, X_{t_2,1}), \ldots, h(X_{t_1,d}, X_{t_2,d})\big)^\top, \boldsymbol{h}_1(\boldsymbol{X}_t) := \big(h_{1,1}(X_{t,1}), \ldots, h_{1,d}(X_{t,d})\big)^\top,$$
$$\text{and} \quad \boldsymbol{g}(\boldsymbol{X}_{t_1}, \boldsymbol{X}_{t_2}) := \big(g_1(X_{t_1,1}, X_{t_2,1}), \ldots, g_d(X_{t_1,d}, X_{t_2,d})\big)^\top,$$

where $h_{1,j}(x)$ and $g_j(x, y)$ are defiend in (3.4). Let $\boldsymbol{\Gamma} = (\gamma_{i,j}) \in \mathbb{R}^{d\times d} = \mathrm{Cov}(\boldsymbol{h}_1(\boldsymbol{X}_1))$.

Step (i). In this step, we approximate the distribution of W under the null hypothesis. Note that under the null hypothesis, we have $\mathbb{E}h(X_j, X_j') = 0$. For each fixed $G \le k \le n-G$ and $t = 1, \ldots, n$, let

$$a_t(k) = \frac{\sqrt{n}}{\sqrt{G}}\big(\mathbf{1}\{k - G + 1 \le t \le k\} - \mathbf{1}\{k + 1 \le t \le k + G\}\big).$$

Then, we can rewrite $\boldsymbol{T}(k)$ as the following form:

$$\boldsymbol{T}(k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_t(k)\boldsymbol{h}_1(\boldsymbol{X}_t) + \frac{1}{G^{3/2}} \sum_{t_1=k-G+1}^{k} \sum_{t_2=k+1}^{k+G} \boldsymbol{g}(\boldsymbol{X}_{t_1}, \boldsymbol{X}_{t_2}). \tag{3.6}$$

Our theorem shows that the residual term of the Hoeffding's decomposition can be uniformly negligible over k and $1 \le j \le d$ in the sense that

$$\max_{G\le k\le n-G} \big\|\boldsymbol{T}(k) - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_t(k)\boldsymbol{h}_1(\boldsymbol{X}_t)\big\|_\infty = O_p(\log(nd)/\sqrt{G}).$$

Hence, by the Hoeffding's decomposition, we can approximate W by $\max_{G\le k\le n-G} \|\boldsymbol{T}^{(1)}(k)\|_\infty$, where $\boldsymbol{T}^{(1)}(k) := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_t(k)\boldsymbol{h}_1(\boldsymbol{X}_t)$.

Step (ii). Note that we don't know the exact distribution of $\max_{G\le k\le n-G} \|\boldsymbol{T}^{(1)}(k)\|_\infty$. As $\boldsymbol{T}^{(1)}(k)$ is a sum of independent random vectors with zero mean, this motivates us to use a Gaussian random vector to further approximate $\max_{G\le k\le n-G} \|\boldsymbol{T}^{(1)}(k)\|_\infty$. Specifically,

let $\boldsymbol{G}_1, \ldots, \boldsymbol{G}_n$ be $i.i.d$ Gaussian random vectors with $\boldsymbol{G}_t = (G_{t,1}, \ldots, G_{t,d}) \in \mathbb{R}^d$ and $\boldsymbol{G}_t \sim N(0, \boldsymbol{\Gamma})$. We define the Gaussian random vector-based testing statistic $\boldsymbol{T}^{\boldsymbol{G}}(k) = (\boldsymbol{T}_1^{\boldsymbol{G}}(k), \ldots, T_d^{\boldsymbol{G}}(k))^\top$ with

$$\boldsymbol{T}^{\boldsymbol{G}}(k) \;\; = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_t(k) \boldsymbol{G}_t, \quad \text{with} \;\; G \leq k \leq n - G. \tag{3.7}$$

Our theorem shows that we can approximate $\boldsymbol{T}^{(1)}(k)$ using $\boldsymbol{T}^{\boldsymbol{G}}(k)$ in the sense that

$$\sup_{z \in (0,\infty)} |\mathbb{P}(\max_{G \leq k \leq n-G} \|\boldsymbol{T}^{(1)}(k)\|_\infty \leq z) - \mathbb{P}(\max_{G \leq k \leq n-G} \|\boldsymbol{T}^{G}(k)\|_\infty \leq z)| \leq C_1 \Big(\frac{\log^7((n-2G+1)d)}{G} \Big)^{1/6}.$$

Hence, we can use the distribution of $\max_{G \leq k \leq n-G} \|\boldsymbol{T}^{G}(k)\|_\infty$ to approximate that of $\max_{G \leq k \leq n-G} \|\boldsymbol{T}^{(1)}(k)\|_\infty$.

Step (iii). Since the covariance structure $\boldsymbol{\Gamma}$ is typically unknown, one way is to obtain some estimator $\widehat{\boldsymbol{\Gamma}}$ and use the corresponding Gaussian random vectors to approximate the testing statistic. However, the estimation of the covariance matrix imposes certain requirements, which can be particularly challenging in ultra-high dimensional problems. Instead, we can use the multiplier bootstrap-based testing statistic in (3.1) to approximate $\boldsymbol{T}^{\boldsymbol{G}}(k)$ in a data-driven way. Specifically, for the Gaussian random vector-based testing statistic $\boldsymbol{T}^{\boldsymbol{G}}(k)$, it has very complicated covariance structure in terms of the search location:

$$\mathrm{Cov}(\boldsymbol{T}^{\boldsymbol{G}}(k_1), \boldsymbol{T}^{\boldsymbol{G}}(k_2)) = \begin{cases} 0, & \text{if } |k_2 - k_1| > 2G - 1; \\ -\dfrac{2G - |k_2 - k_1|}{G} \boldsymbol{\Gamma}, & \text{if } G - 1 < |k_2 - k_1| \leq 2G - 1; \\ \dfrac{2G - 3|k_2 - k_1|}{G} \boldsymbol{\Gamma}, & \text{if } 0 < |k_2 - k_1| \leq G - 1; \\ 2\boldsymbol{\Gamma}, & \text{if } |k_2 - k_1| = 0. \end{cases} \tag{3.8}$$

Our theorem shows that the multiplier bootstrap-based testing statistic can capture the above covariance structure accurately in the sense that

$$\max_{G \leq k_1, k_2 \leq n-G} \|\mathrm{Cov}(\boldsymbol{T}^{\boldsymbol{G}}(k_1), \boldsymbol{T}^{\boldsymbol{G}}(k_2)) - \mathrm{Cov}(\boldsymbol{T}^{b}(k_1), \boldsymbol{T}^{b}(k_2))\|_\infty = O_p(\epsilon_n),$$

where

$$\epsilon_n = C_2 D^2 \max\Big(\sqrt{\frac{\log(nd)}{G}}, \frac{\log^2(nd)\log^2(Gd)}{G} \Big).$$

The above results allow us to use $\|\boldsymbol{T}^{b}(k)\|_\infty$ to approximate $\max_{G \leq k \leq n-G} \|\boldsymbol{T}^{G}(k)\|_\infty$. Lastly, we analyze the estimation error and combine results from **Steps (i) - (iii)** to finish

the proof of Theorem 3.1. The detailed proof is in Appendix B.1 of the supplementary materials.

After analyzing the size, we next discuss the power properties. To this end, we need some additional notations. Recall $\Pi_m := \{j \in \{1, \ldots, d\} : \theta_j^{(m)} \neq 0\}$ as the set of coordinates having a change point at γ_m. For each $m = 1, \ldots, M_0$, define the signal jump vector $\boldsymbol{\theta}^{(m)} := (\theta_1^{(m)}, \ldots, \theta_d^{(m)})^\top$. Define $\boldsymbol{\theta}^\diamondsuit = \max_{1 \leq m \leq M_0} \|\boldsymbol{\theta}^{(m)}\|_\infty$ be the maximum signal jump among all change points and let $\Delta := \min_{1 \leq m \leq M_0}(\gamma_{m+1} - \gamma_m)$ be the minimum segmentation length between change points with $\gamma_{M_0+1} = n$. Next, we present some conditions for $\mathbf{H}_1$.

Assumption (B.1): For the bandwidth parameter G, we require

$$\Delta \geq 2G, \quad \text{and} \quad \frac{\log^2(nd)\log^2(Gd)}{G} \to 0 \text{ as } (n, d) \to \infty.$$

Additionally, suppose there exist some constants such that $\max_{1 \leq m \leq M_0} \|\boldsymbol{\theta}^{(m)}\|_\infty \leq C^*$.

Assumption (B.2): There exists some constants D^ℓ such that $\mathbb{E}|h(X_{\gamma_m, j}, X_{\gamma_m+1, j}) - \theta_j^{(m)}|^{2+\ell} \leq D^\ell$ with $\ell = 1, 2$ for all $j = 1, \ldots, d$ and $m = 1, \ldots, M_0$.

Assumption (B.3): There exists some constant D such that $\|h(X_{\gamma_m, j}, X_{\gamma_m+1, j}) - \theta_j^{(m)}\|_{\psi_1} \leq D$ for all $j = 1, \ldots, d$ and $m = 1, \ldots, M_0$.

Note that Assumption (B.1) requires the some common conditions on the segmentation length Δ and the bandwidth parameter in the literature. It requires that the bandwidth should not be too large in the sense that $G \leq \Delta/2$ such that there is at most one change point in any candidate moving window $[k - G + 1, k + G]$. Moreover, the bandwidth is required to have an order $G \gg \log^2(nd)\log^2(Gd)$ so that the large sample property holds. See also Eichinger and Kirch (2018). Assumptions (B.2) and (B.3) make some similar moment conditions on the kernel to Assumptions (A.2) and (A.3). Notably, our moment condition assumptions are primarily imposed on the kernel rather than the data itself. Therefore, for many heavy-tailed datasets, the theoretical framework of the model remains applicable as long as an appropriate kernel function is chosen.

With the above notations and some regularity conditions, the following Theorem 3.2

shows that we can reject the null hypothesis of no change point with overwhelming probability.

Theorem 3.2. Suppose Assumptions (**B.1**) – (**B.3**) hold. If $\boldsymbol{\theta}^{\diamondsuit}$ satisfies

$$\sqrt{G} \times \boldsymbol{\theta}^{\diamondsuit} \geq C_0\Big(\sqrt{2\log(dn)} + \sqrt{2\log(\alpha^{-1})}\Big), \tag{3.9}$$

under $\mathbf{H}_1$, we have $\mathbb{P}(\Phi_{\alpha,W} = 1) \to 1$, as $n, d, B \to \infty$, where C_0 is a large enough universal positive constant not depending on n or d.

Theorem 3.2 demonstrates that with probability tending to one, our proposed new test can detect the existence of multiple change points as long as the corresponding signal to noise ratio satisfies (3.9). Specifically, considering Assumption (B.1) with $\Delta \geq 2G$, it mainly requires that

$$\max_{1\leq m\leq M_0} \|\boldsymbol{\theta}^{(m)}\|_\infty \geq \sqrt{\frac{\log(nd)}{\Delta}}. \tag{3.10}$$

Hence, with a larger signal jump and larger segmentation length, it is more likely to trigger a rejection of the null hypothesis. From the above conditions, we can also observe that to detect multiple change points, it is sufficient for the signal-to-noise ratio condition of just one of the change points to meet the requirements. This is fundamentally different from the conditions for change point estimation discussed later.

Remark 3.3. The rate in (3.10) is rate-optimal for high-dimensional change point detection. Specifically, Enikeeva and Harchaoui (2019) studied the minimax rate optimality for detecting a singe mean shift in independent Gaussian vectors with an identity covariance matrix. Defining $\gamma = d^{1-\beta}$ for $\beta \in (0, 1)$ as the proportion of coordinates having a change point, assume a common single change point occurs at $\gamma_1 \in (0, 1)$, with common signal strength $|\theta_j^{(2)} - \theta_j^{(1)}| = a_{n,d}, \quad \forall j \in \Pi_1$. For high sparsity with $\beta \in (1/2, 1)$, they derived the minimax separation rate:

$$a_{n,d} \approx \Big(\frac{\log(d)}{\gamma_1 \frac{(n-\gamma_1)}{n}}\Big)^{1/2}. \tag{3.11}$$

This result implies that no α-level test can reject $\mathbf{H}_0$ uniformly when $a_{n,d} = C_*\Big(\frac{\log(d)}{\gamma_1 \frac{(n-\gamma_1)}{n}}\Big)^{1/2}$

for sufficiently small C_*. Considering (3.10) and $\gamma_1 \frac{(n-\gamma_1)}{n} \approx \min(\gamma_1, n-\gamma_1) = \Delta$, our proposed tests achieve rate-optimality for sparse alternatives.

4 Estimation and inference of break points

After introducing our test statistic for Problem 2.2, we next present our methodology for identifying the unknown change point number, locations, as well as constructing their valid confidence intervals. Our estimation algorithms proceed into two steps. In the first step, we modify the ideas in Eichinger and Kirch (2018) to the high dimensional settings based on the two sample kernel $h(x, y)$ with ℓ_∞-norm aggregations.

Step 1: Obtain initial estimator. Recall $\boldsymbol{T}(k) = (T_1(k), \ldots, T_d(k))^\top$ with $T_j(k)$ defined in (2.3). Consider all pairs of indices $(v_j, w_j) \subset \{G, \ldots, n-G\}$ such that

$$\begin{aligned} &\|\boldsymbol{T}(k)\|_\infty \geq c_{\alpha,W}, \text{ for } k = v_j, \ldots, w_j, \\ &\|\boldsymbol{T}(k)\|_\infty < c_{\alpha,W}, \text{ for } k = v_j - 1, w_j + 1, \\ &w_j - v_j \geq \eta G, \text{for some fixed but arbitary } \ 0 < \eta < 1/2, \end{aligned} \tag{4.1}$$

where $c_{\alpha,W}$ is the α quantile level obtained from the bootstrap procedure in Section 3. Condition 4.1 is necessary to avoid overestimation by spurious local maxima exceeding the critical value on the boundary between significant and insignificant areas. Based on (4.1), our initial estimation for change point number and locations are defined as:

$$\widehat{M}_0 = \text{number of pairs } (v_j, w_j), \quad \text{and } \widehat{\gamma}_m = \underset{v_m \leq k \leq w_m}{\arg\max} \|\boldsymbol{T}(k)\|_\infty \text{ for } m = 1, \ldots, M_0. \tag{4.2}$$

Note that the principle in (4.1) was proposed in Eichinger and Kirch (2018) for one-dimensional sequences and recently modified by Chen et al. (2022). Different from Eichinger and Kirch (2018), we consider the high dimensional setting using the ℓ_∞-norm aggregations for the two-sample U-statistic-based moving window process. In the following, we provide consistency results of the initial estimated change point number and locations. To that end, we need the following conditions on the signal strengh $\boldsymbol{\theta}_\diamondsuit = \min_{1 \leq m \leq M_0} \|\boldsymbol{\theta}^{(m)}\|_\infty$ with

$\boldsymbol{\theta}^{(m)} := \mathbb{E}\boldsymbol{h}(\boldsymbol{X}_{\gamma_m}, \boldsymbol{X}_{\gamma_m+1})$.

Assumption (**C.1**) Assume $G \times \boldsymbol{\theta}_{\diamondsuit}^2 \gg \log(nd)$.

Assumption (C.1) is the so called signal to noise ratio condition for identifying multiple change points. Considering Assumption (B.1) with $\Delta \geq 2G$, it can be equivalently stated as $\Delta \times \boldsymbol{\theta}_{\diamondsuit}^2 \gg \log(nd)$. Hence, with a larger segmentation length $\boldsymbol{\Delta}$ and larger signal strength $\boldsymbol{\theta}_{\diamondsuit}$, Assumption (C.1) is more likely to be satisfied.

Theorem 4.1. (**Consistency of initial estimators**) Suppose Assumptions **(B.1)-(B.3)** and **(C.1)** hold. Then we have

(1) $\mathbb{P}\{\widehat{M}_0 = M_0\} \to 1$, as $n, d \to \infty$;

(2) $|\widehat{\gamma}_m - \gamma_m| = O_P\Big(\dfrac{\log(nd)}{\|\boldsymbol{\theta}^{(m)}\|^2}\Big)$, for $m = 1, \ldots, M_0$.

Theorem 4.1 (1) shows that the initial procedure can correctly identify the change point number M_0 using the ℓ_∞-norm aggregation of U-statistic-based moving window process. This is important for the refining procedure in the following step. More importantly, Theorem 4.1 (2) demonstrates that the initial estimator $\widehat{\gamma}_m$ has an estimation error bound of $O_p(\log(nd)/\|\boldsymbol{\theta}^{(m)}\|^2)$. Moreover, for d-dimensional Gaussian random vector with a single change point at γ_1, Proposition 3 in Wang and Samworth (2018) shows that the estimation lower bound has an error of $O_p(\dfrac{1}{\|\boldsymbol{\theta}^{(m)}\|^2})$. Hence, our theoretical result is minimax rate optimal up to a $\log(nd)$ factor for detecting change points with sparse alternatives.

The problem of high-dimensional change point estimation has been extensively studied, leading to the development of several algorithms. In contrast to previous studies, our approach not only facilitates multiple change point estimation but also extends to more general parameters, such as variance or robust mean estimation. Under this broader modeling framework, our algorithm achieves a multiple change point convergence rate comparable to the near-optimal level of Wang and Samworth (2018), while imposing fewer restrictions on the signal-to-noise ratio as imposed in Yu and Chen (2021); Cho (2016). Our simulations in Section 5 demonstrate that the proposed method performs competitively compared with

existing techniques and exhibits superior robustness in estimating both the number and locations of change points, particularly when dealing with heavy-tailed distributions or outliers.

Lastly, we would like to point out that the requirements for identifying and detecting multiple change points are different. More specifically, from Theorem 4.1, to correctly identify the locations of change point with desirable accuracy, the signal strength should at least satisfy $\boldsymbol{\theta}_{\diamondsuit} \gg \sqrt{\log(nd)/\Delta}$. In contrast, Theorem 3.2 shows that it is sufficient to detect a change point if $\boldsymbol{\theta}^{\diamondsuit} \geq \sqrt{\log(nd)/\Delta}$. holds. Hence, we need more stringent conditions for locating change points than detecting its existence.

Step 2: Obtain refined estimator. Next, we introduce more refined theory of change point estimation. To that end, we propose a U-statistic projection refinement algorithm (U-PRA). This approach initially detects the subset of components at each change point with high probability and subsequently projects the test statistic onto this subspace in the direction of the U-statistic. This transformation simplifies the problem to univariate change-point estimation. Finally, a local search is applied to refine the estimates further. Our objectives are twofold. First, we aim to further optimize the theoretical results of change point estimation in Theorem 4.1 (2) from $O_P(\log(nd)/\|\boldsymbol{\theta}^{(m)}\|^2)$ to $O_P(1/\|\boldsymbol{\theta}^{(m)}\|^2)$, which is known to be theoretical optimality (Verzelen et al. (2023)). Building on this, we construct valid confidence intervals for the positions of multiple change points.

To accomplish the above goal, we need to construct U-statistic-based projection directions. Recall Π_m as the set of coordinates having a change point at γ_m. We next construct estimation for Π_m. For each change point γ_m, we first estimate its signal jump size for each coordinate by

$$\widehat{\theta}_j^m = \frac{1}{G^2} \sum_{t_1=\widehat{\gamma}_m-3G/2}^{\widehat{\gamma}_m-G/2} \sum_{t_2=\widehat{\gamma}_m+G/2}^{\widehat{\gamma}_m+3G/2} h(X_{t_1 j,}, X_{t_2,j}), \text{for } j = 1, \ldots, d. \tag{4.3}$$

Note that by Theorem 4.1, we have proved that with probability tending to one, $|\widehat{\gamma}_m - \gamma_m| = o(G)$. Hence, this gurantees that the samples that used to obtain $\widehat{\theta}_j^m$ are not mixed. Then,

the final estimation for Π_m is defined as

$$\widehat{\Pi}_m = \{1 \le j \le d : |\widehat{\theta}_j^{(m)}| \ge w^+\}. \tag{4.4}$$

In practice, w^+ is some threshold parameter which can be chosen to be large enough to ensure that we can detect all the jumps with probability tending to one as in Theorem 4.2.

Theorem 4.2. Suppose Assumptions (**B**.1)–(**B**.3) and (**C**.1) hold. Furthermore, for the threshold parameter w^+, we require

$$w^+ \le \min_{1\le m\le M_0} \min_{1\le j\le d} |\theta_j^{(m)}| - C_2 D\Big(\sqrt{\frac{\log(nd)}{G}} + \frac{\log(Gd)\log(nd)}{G}\Big)$$
$$\text{and} \quad w^+ \ge C_2 D\Big(\sqrt{\frac{\log(nd)}{G}} + \frac{\log(Gd)\log(nd)}{G}\Big),$$

for some $C_2 > 0$. Then we have

$$\mathbb{P}(\widehat{\Pi}_m = \Pi_m) \to 1.$$

Using $\widehat{\Pi}_m$, instead of aggregating the coordinate with the maximum value, we aggregate coordinates that have a signal jump larger than w^+. Specifically, for each estimated change point location $\widehat{\gamma}_m$ with $|k - \widehat{\gamma}_m| \le G/4$, we define

$$\widetilde{T}_j(k) := \frac{1}{G} \sum_{t_1=k-G+1}^{k} \sum_{t_2=k+1}^{k+G} \widehat{\theta}_j^{(m)} h(X_{t_1,j}, X_{t_2,j}), \quad \text{for} \quad j = 1, \ldots, d. \tag{4.5}$$

Note that $\widetilde{T}_j(k)$ can be understood as a projection of the high-dimensional observations $h(X_{t_1,j}, X_{t_2,j})$ according to the direction of $\widehat{\boldsymbol{\theta}}^{(m)} = (\widehat{\theta}_1^m, \ldots, \widehat{\theta}_d^m)^\top$. The idea is simiar to Wang and Samworth (2018). In addition, in high-dimensional settings, there are many projection-based methods for constructing change point tests, including those by Aston and Kirch (2018); Chen et al. (2022) and others. However, these methods primarily rely on sample means to determine projection directions. Unlike existing approaches, we have developed U-statistic based projection directions using the general kernel $h(x, y)$ to obtain robust support sets for change points. Based on the aggregated series, the refined change-point locations can be detected as follows:

$$\widetilde{\gamma}_m = \underset{|k-\widehat{\gamma}_m|\le G/4}{\arg\max} \sum_{j\in\widehat{\Pi}_m} \widetilde{T}_j(k). \tag{4.6}$$

The following theorem shows that after refinemet, our proposed U-PRA can have an optimal convergence rate. Thanks to this result, we are able to construct valid confidence intervals for the positions of change points. To that end, we need more notations. Recall $\Pi_m = \{j \in \{1, \cdots, d\} : \theta_j^{(m)} \neq 0\}$ as the coordinates having a change point at γ_m and $s_m = |\Pi_m|$ is the total number. Without loss of generality, we assume $\Pi_m = \{1, \ldots, s_m\}$. For each $m = 1, \ldots, M_0$ and $j = 1, \ldots, d$, define the centralized kernel $\widetilde{h}(x, y) = h(x, y) - \theta_j^{(m)}$ and let

$$h_{1,j}(X_{\gamma_m,j}) := \mathbb{E}[\widetilde{h}(X_{\gamma_m,j}, X_{\gamma_m+1,j}|X_{\gamma_m,j})], \quad \underline{h}_{1,j}(X_{\gamma_m,j}) := \mathbb{E}[h(X_{\gamma_m-1,j}, X_{\gamma_m,j}|X_{\gamma_m,j})],$$
$$h_{2,j}(X_{\gamma_m+1,j}) := \mathbb{E}[\widetilde{h}(X_{\gamma_m,j}, X_{\gamma_m+1,j}|X_{\gamma_m+1,j})], \ \overline{h}_{1,j}(X_{\gamma_m+1,j}) := \mathbb{E}[h(X_{\gamma_m+1,j}, X_{\gamma_m+2,j}|X_{\gamma_m+1,j})],$$

with $\theta_j^{(m)} := \mathbb{E}[h(X_{\gamma_m,j}, X_{\gamma_m+1,j}]$ being the signal jump of the j-th coordinate at γ_m. Note that by above definitions, $h_{1,j}(X_{\gamma_m,j})$ and $h_{2,j}(X_{\gamma_m+1,j})$ are the Hoeffding's projections for the two-sample kernel $h(x, y)$ with heterogeneous data and $\underline{h}_{1,j}(X_{\gamma_m,j})$, $\overline{h}_{1,j}(X_{\gamma_m+1,j})$ are the Hoeffding's projections for the two-sample kernel before and after the change point with homogeneous data. Moreover, define the four $\mathbb{R}^{s_m}$-dimensional vectorized version of Hoeffding's projections as follows

$$\boldsymbol{h}_1(\boldsymbol{X}_{\gamma_m}) = (h_{1,1}(X_{\gamma_m,1}), \ldots, h_{1,s_m}(X_{\gamma_m,s_m}))^\top, \ \boldsymbol{h}_2(\boldsymbol{X}_{\gamma_m+1}) = (h_{2,1}(X_{\gamma_m+1,1}), \ldots, h_{2,s_m}(X_{\gamma_m+1,s_m}))^\top,$$
$$\overline{\boldsymbol{h}}_1(\boldsymbol{X}_{\gamma_m+1}) = (\overline{h}_{1,1}(X_{\gamma_m+1,1}), \ldots, \overline{h}_{1,s_m}(X_{\gamma_m+1,s_m}))^\top, \ \underline{\boldsymbol{h}}_1(\boldsymbol{X}_{\gamma_m}) = (\underline{h}_{1,1}(X_{\gamma_m,1}), \ldots, \underline{h}_{1,s_m}(X_{\gamma_m,s_m}))^\top,$$

and define the following four $\mathbb{R}^{s_m \times s_m}$-dimensional covariance matrices as

$$\boldsymbol{\Sigma}_1^{(m)} = \mathrm{Cov}(\boldsymbol{h}_1(\boldsymbol{X}_{\gamma_m})), \quad \boldsymbol{\Sigma}_2^{(m)} = \mathrm{Cov}(\boldsymbol{h}_2(\boldsymbol{X}_{\gamma_m+1})),$$
$$\boldsymbol{\Sigma}_3^{(m)} = \mathrm{Cov}(\overline{\boldsymbol{h}}_1(\boldsymbol{X}_{\gamma_m+1}) - \boldsymbol{h}_2(\boldsymbol{X}_{\gamma_m+1})), \quad \boldsymbol{\Sigma}_4^{(m)} = \mathrm{Cov}(\underline{\boldsymbol{h}}_1(\boldsymbol{X}_{\gamma_m}) + \boldsymbol{h}_1(\boldsymbol{X}_{\gamma_m})).$$

To present the change point inference results, we need the following additional assumption.

Assumption (C.2): We require there exists constants $\kappa_1 > 0$ and $\kappa_2 > 0$ such that

$$\begin{aligned} &\kappa_1 \leq \min(\lambda_{\min}(\boldsymbol{\Sigma}_1^{(m)}), \lambda_{\min}(\boldsymbol{\Sigma}_2^{(m)}), \lambda_{\min}(\boldsymbol{\Sigma}_3^{(m)}), \lambda_{\min}(\boldsymbol{\Sigma}_4^{(m)}), \\ &\max(\lambda_{\max}(\boldsymbol{\Sigma}_1^{(m)}), \lambda_{\max}(\boldsymbol{\Sigma}_2^{(m)}), \lambda_{\max}(\boldsymbol{\Sigma}_3^{(m)}), \lambda_{\max}(\boldsymbol{\Sigma}_4^{(m)}) \leq \kappa_2, \text{ for } m = 1, \ldots, M_0. \end{aligned} \tag{4.7}$$

Remark 4.3. Assumption (C.2) makes some regular assumptions on the covariance structures of the data and is very mild to satisfy. For example, under the mean shift model, if $h(x, y) = y - x$, it reduces to

$$\kappa_1 \leq \min(\lambda_{\min}(\mathrm{Cov}(\boldsymbol{X}_{\gamma_m})), \lambda_{\min}(\mathrm{Cov}(\boldsymbol{X}_{\gamma_m+1}))) \leq \max(\lambda_{\max}(\mathrm{Cov}(\boldsymbol{X}_{\gamma_m})), \lambda_{\max}(\mathrm{Cov}(\boldsymbol{X}_{\gamma_m+1}))) \leq \kappa_2.$$

With the above notations, we are ready to give the change point inference results.

Theorem 4.4. Suppose Assumptions (**B.1**)–(**B.3**) and (**C.1**) – (**C.2**) hold. Then we have

$$\begin{aligned} &(1)|\widetilde{\gamma}_m - \gamma_m| = O_P\big(\frac{1}{\|\boldsymbol{\theta}^{(m)}\|^2}\big), \text{for } m = 1, \ldots, M_0; \\ &(2)\|\boldsymbol{\theta}^{(m)}\|^2(\widetilde{\gamma}_m - \gamma_m) \Rightarrow \arg\max_{s\in\mathbb{R}} Z^{(m)}(s), \text{for } m = 1, \ldots, M_0, \end{aligned} \tag{4.8}$$

with

$$Z^m(s) = \begin{cases} -s + \sigma_{1,*}^{(m)} W_1^{(m)}(s) + \sigma_{3,*}^{(m)} W_2^{(2)}(s) + \sigma_{2,*}^{(m)} W_3^{(m)}(s), & s > 0, \\ 0, & s = 0, \\ s + \sigma_{1,*}^{(m)} W_1^{(m)}(s) + \sigma_{4,*}^{(m)} W_2^{(m)}(s) + \sigma_{2,*}^{(m)} W_3^{(m)}(s), & s < 0, \end{cases}$$

where $\{W_1^{(m)}(s)\}$, $\{W_2^{(m)}(s)\}$, $\{W_3^{(m)}(s)\}$ are standard Brownian motion processes on $(-\infty, \infty)$ which are independent with each other and $\sigma_{j,*}^{(m)}; j = 1, 2, 3, 4$ are defined as

$$\begin{aligned} &\sigma_{1,*}^{(m)} = \frac{((\boldsymbol{\theta}^{(m)})^\top \boldsymbol{\Sigma}_1^{(m)} \boldsymbol{\theta}^{(m)})^{1/2}}{\|\boldsymbol{\theta}^{(m)}\|}, \sigma_{2,*}^{(m)} = \frac{((\boldsymbol{\theta}^{(m)})^\top \boldsymbol{\Sigma}_2^{(m)} \boldsymbol{\theta}^{(m)})^{1/2}}{\|\boldsymbol{\theta}^{(m)}\|}, \\ &\sigma_{3,*}^{(m)} = \frac{((\boldsymbol{\theta}^{(m)})^\top \boldsymbol{\Sigma}_3^{(m)} \boldsymbol{\theta}^{(m)})^{1/2}}{\|\boldsymbol{\theta}^{(m)}\|}, \sigma_{4,*}^{(m)} = \frac{((\boldsymbol{\theta}^{(m)})^\top \boldsymbol{\Sigma}_4^{(m)} \boldsymbol{\theta}^{(m)})^{1/2}}{\|\boldsymbol{\theta}^{(m)}\|}. \end{aligned} \tag{4.9}$$

Moreover, we can prove that $\|\boldsymbol{\theta}^{(m)}\|^2(\widetilde{\gamma}_m - \gamma_m)$'s are asymptotically independent.

Theorem 4.4 (1) demonstrates that the U-statistic projection refinement algorithm U-PRA has an estimation error bound of $O_P(\frac{1}{\|\boldsymbol{\theta}^{(m)}\|^2})$, which has been improved compared to the initial estimators obtained in Theorem 4.1. To our best knowledge, the convergence rate is optimal. From the estimation point of view, the optimal rate $O_P(\frac{1}{\|\boldsymbol{\theta}^{(m)}\|^2})$ of γ_m may not appear to be a substantial enhancement compared to the initial estimator in Theorem 4.1. Nevertheless, this refinement is crucial from an inferential standpoint, as the key factor is the existence of an $O_P(\frac{1}{\|\boldsymbol{\theta}^{(m)}\|^2})$ rate, which ensures the presence of a limiting distribution. Moreover, the refined change point estimator itself achieves a convergence rate of $O_P(\frac{1}{\|\boldsymbol{\theta}^{(m)}\|^2})$, which remains independent of the dimensional parameters d, the sampling period n, and holds even in high-dimensional settings or when the jump size is potentially

diminishing.

Theorem 4.4 (2) shows that the centered and scaled change point estimator, after refining, has an limiting distribution which follows the arg max of a drifted weighted Brownian motion process. Hence, we can make valid confidence intervals for the detected change points. Let $q_{\alpha/2}^{(m)}$ and $q_{1-\alpha/2}^{(m)}$ be the quantiles of the $\arg\max_{s\in\mathbb{R}} Z^{(m)}(s)$. Using Theorem 4.4, we can construct the theoretical confidence interval for each change point γ_m as

$$\Big[\widetilde{l}_\alpha(m), \widetilde{u}_\alpha(m)\Big] := \Big[\widetilde{\gamma}_m - \frac{q_{1-\alpha/2}^{(m)}}{\|\boldsymbol{\theta}^{(m)}\|^2}, \widetilde{\gamma}_m - \frac{q_{\alpha/2}^{(m)}}{\|\boldsymbol{\theta}^{(m)}\|^2}\Big], \text{ for } m = 1, \ldots, M_0.$$

Moreover, Theorem 4.4 allows us to make simultaneous confidence intervals of the detected change points. Specifically, for any $1 \le J \le M_0$ given change points $\gamma_{m_1}, \ldots, \gamma_{m_J} \subset \{\gamma_1, \cdots, \gamma_{M_0}\}$, let $[\widetilde{l}_{\alpha/J}(m_1), \widetilde{u}_{\alpha/J}(m_1)]$,..., $[\widetilde{l}_{\alpha/J}(m_J), \widetilde{u}_{\alpha/J}(m_J)]$ be the corresponding confidence intervals. Then, as $(n, d) \to \infty$, by Theorem 4.4, we have

$$\mathbb{P}\Big(\gamma_{m_1} \in [\widetilde{l}_{\alpha/J}(m_1), \widetilde{u}_{\alpha/J}(m_1)], \ldots, \gamma_{m_J} \in [\widetilde{l}_{\alpha/J}(m_J), \widetilde{u}_{\alpha/J}(m_J)]\Big) \to 1 - \alpha.$$

Remark 4.5. From Theorem 4.4, we see that Assumption C.2 mainly requires $\sigma_{1,*}^{(m)} \ldots \sigma_{4,*}^{(m)}$ have moderate values. Specifically, it suggests that the stochastic influence of mixtures of three Brownian motion processes is neither negligible nor predominant which balances the signal and noise. In Appendix B, we introduce how to obtain estimators for $\sigma_{1,*}^{(m)} \ldots \sigma_{4,*}^{(m)}$ and construct the estimators for $[\widetilde{l}_\alpha(m), \widetilde{u}_\alpha(m)]$.

5 Numerical performance

In this section, we examine the numerical performance of our proposed method in terms of change point detection, estimation, as well as inference for the mean shift and variance change point models. We also compare our method with several existing state-of-art techniques. To save space, we move the detailed model settings and results to Appendix C of the supplementary materials. The simulation results can be summarized below.

For mean change point models, we consider two types of kernels: $h_1(x, y) = y - x$ and

$h_2(x, y) = \text{sign}(y - x)$. The main findings are as follows:

1. **Size control**: When the error distribution is Gaussian, both methods effectively control the empirical size. However, under heavy-tailed errors or outlier contamination, the method based on $h_1(x, y)$ becomes overly conservative, while the h_2-based method still maintains proper size control.

2. **Multiple change point testing**: Under Gaussian errors, our method performs comparably to existing approaches. In contrast, when data are heavy-tailed or contaminated with outliers, existing methods either become too conservative or fail to control the size. In these cases, the h_2-based method achieves the best performance.

3. **Multiple change point estimation**: When the data follow a Gaussian distribution, the h_1-based method outperforms both the h_2-based method and other mainstream approaches. However, under heavy-tailed errors, both the h_1-based method and existing methods suffer from large estimation errors, while the h_2-based method remains robust and accurately identifies both the number and locations of change points. Moreover, for both h_1 and h_2, the refined change point estimators significantly improve over the initial estimators.

4. **Multiple change point inference**: Under Gaussian errors, both h_1 and h_2-based methods yield confidence intervals with empirical coverage probabilities close to the nominal level (95%). However, when data are heavy-tailed (e.g., t_3 or t_2 distributions) or contaminated with outliers, the coverage of the h_1-based method drops substantially below the nominal level, while the h_2-based method remains valid.

Finally, for variance change point models, we also compare two kernels: $h_3(x, y) = y^2 - x^2$ and $h_4(x, y) = \text{sign}(y^2 - x^2)$, and observe similar patterns in the results. In summary, the simulation results highlight the effectiveness and competitive performance of our proposed methods compared to existing approaches.

6 Application to real-world data

In this section, we aim to detect variance change points in the S&P 500 dataset to identify shifts in market volatility over two critical periods: January 1, 2007 to December 31, 2011 and January 1, 2019 to December 31, 2022. The first time period (2007-2011) encompasses the 2008 Global Financial Crisis and the second period (2019-2022) includes the onset of the COVID-19 pandemic, which triggered sharp declines in financial markets in early 2020. The dataset is obtained from Yahoo! Finance (https://finance.yahoo.com/) consisting of daily returns for 234 ($d = 234$) continuously listed stocks within the S&P 500 index. It is well known that volatility change points often indicate transitions between stable and turbulent periods, which can be driven by significant events and is crucial for understanding shifts in market dynamics and managing financial risks. Note that the financial data are typically heavy-tailed and we have no prior-knowledge about the tail structure of the data. Hence, for this real data analysis, it seems very suitable to use the kernel $\boldsymbol{h}_4(\boldsymbol{x}, \boldsymbol{y}) = \text{sign}(\boldsymbol{y}^2 - \boldsymbol{x}^2)$. For this data analysis, we set the bandwith parameter $G = 80$ and use the testing and estimation procedure in Sections 2-4 to detect and identify change points. The data are scaled to have mean zeros and variance ones before the change point detection.

Table 1: Multiple change poins detection for the S&P 500 dataset.

initial change point	refined change point	confidence interval	Events
2007/7/19	2007/7/19	2007/7/12–2007/7/27	TED Spread
2008/9/3	2008/9/5	2008/9/2-2008/9/10	Collapse of Lehman Brothers
2009/6/5	2009/5/19	2009/5/14-2009/5/21	End of recession
2011/8/1	2011/8/1	2011/7/27-2011/8/5	Global stock markets fell due to fears contagion of the European sovereign debt crisis
2020/2/21	2020/2/21	2020/2/19-2020/2/25	The early outbreak of COVID-19
2020/6/17	2020/6/17	2020/6/10-2020/6/23	Partial stabilization of the pandemic
2021/10/12	2021/11/18	2021/11/9-2021/11/29	The surge in U.S. inflation

We first analyze the period from 2007-2011. The first change point occurs on July 19, 2007, which marks the onset of the 2008 Global Financial Crisis, reflecting the potential instability in the market during the second half of 2007. To further substantiate the mean-

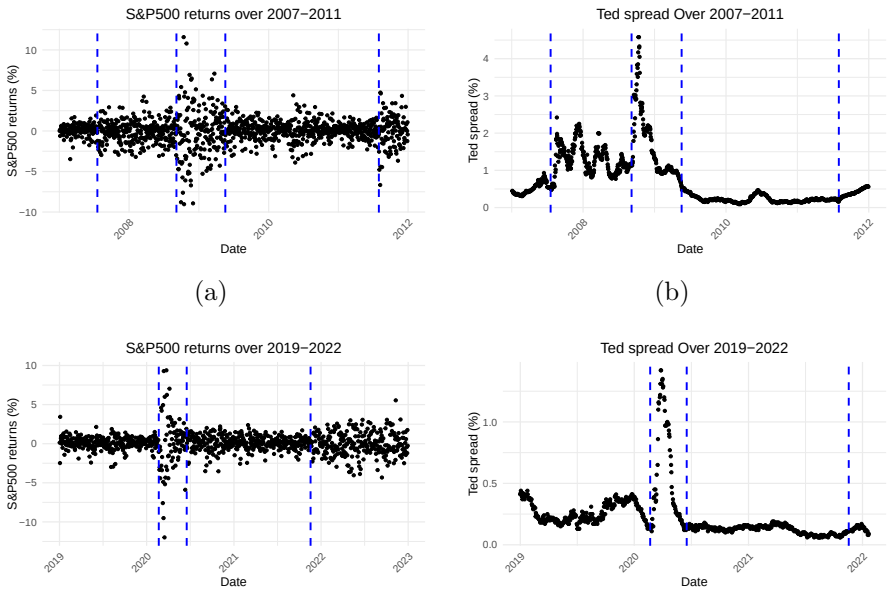

Figure 1: Change point estimation results for the S&P500 data set.

ingful findings of our proposed methods, we refer to the T-bills and TED (TED) spread, which is the difference between the 3-month London Interbank Offered Rate (LIBOR) and the 3-month U.S. government debt (T-bills). It is well known that the TED spread is an indicator of perceived risk in the economy, and an increase in the TED spread during the financial crisis reflects an uptick in credit risk. We observe from Figure 1 that after the first change point, the TED spread experiences dramatic fluctuations, and the estimated change points successfully capture this trend. The second change point occurs on September 5, 2008. On September 15, 2008, Lehman Brothers filed for bankruptcy protection, a critical event in the financial crisis. The change point we identified reflects the sharp fluctuations in the financial markets just before Lehman Brothers' bankruptcy. In fact, after this point, the TED spread reached its peak during the 2008 financial crisis. The third change point occurs on May 19, 2009. According to the National Bureau of Economic Research (NBER), the U.S. economy officially exited the recession in June 2009. The identified change point can capture market changes related to the end of the U.S. recession. The final change point occurs on August 1, 2011. In August 2011, market concerns about the European

debt crisis peaked, particularly related to the fiscal situations of Spain and Italy, which raised concerns about the economic stability of the entire Eurozone.

We next analyze the second period from 2019-2022. The first change point occurs on February 21, 2020. This change point corresponds to the early outbreak of the COVID-19 pandemic, particularly as the virus began to spread to Europe. From February 21 to 23, 2020, Italy reported the first local cases of COVID-19, becoming the first European country to be severely affected. As shown in Figure 1, the TED spread started to fluctuate significantly at this time. During the week of February 24, 2020, global financial markets began to react strongly, with the S&P 500 index dropping by 12%, marking the largest weekly decline since the 2008 financial crisis. The second change point occurs on June 17, 2020. After the initial wave of the pandemic, the situation was under partial control, and the market gradually regained confidence in economic recovery, with financial markets showing signs of stabilization. The change point we identified accurately captures this moment. The third change point occurs on November 18, 2021. In November 2021, the annualized Consumer Price Index (CPI) in the United States reached 6.8%, the highest level in nearly 40 years, reflecting a sharp rise in inflation in the market. This change point may reflect the market's fear of sustained long-term inflation.

7 Summary

This paper introduces a general framework for high-dimensional change-point detection, estimation, and inference based on U-statistics within a moving window. By enabling flexible kernel function selection, our method extends beyond traditional mean-based change-point analysis to accommodate a broader range of structural changes, including variance and robust mean shifts. We develop a minimax-optimal test statistic tailored for sparse alternatives and propose a high-dimensional multiplier bootstrap procedure that ensures valid significance control while maintaining high power. For estimation, we construct an

initial estimator for the number and locations of change points and refine it using the U-statistic Projection Refinement Algorithm (U-PRA), achieving minimax-optimal localization rates. Furthermore, we derive the asymptotic distribution of the refined estimators, facilitating the construction of valid confidence intervals. The effectiveness of our method is demonstrated through both theoretical analysis and empirical validation on synthetic and real-world datasets. Interesting future directions include extending the change-point analysis framework to other statistical settings, such as graphical models.

References

Aston, J. A. D. and Kirch, C. (2018). High dimensional efficiency with applications to change point tests. *Electronic Journal of Statistics* **12** 1901–1947.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78.

Bücher, A. and Kojadinovic, I. (2016). Dependent multiplier bootstraps for non-degenerate U-statistics under mixing conditions with applications. *Journal of Statistical Planning and Inference* **170** 83–105.

Chen, L., Wang, W. and Wu, W. B. (2022). Inference of breakpoints in high-dimensional time series. *Journal of the American Statistical Association* **117** 1951–1963.

Chen, X. (2018). Gaussian and bootstrap approximations for high-dimensional u-statistics and their applications. *The Annals of Statistics* **46** 642–678.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics* **41** 2786–2819.

Chernozhukov, V., Chetverikov, D., Kato, K. et al. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability* **45** 2309–2352.

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

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

Cho, H. and Kirch, C. (2022). Bootstrap confidence intervals for multiple change points based on moving sum procedures. *Computational Statistics & Data Analysis* **175** 107552.

Dette, H., Pan, G. and Yang, Q. (2022). Estimating a Change Point in a Sequence of Very High-Dimensional Covariance Matrices. *Journal of the American Statistical Association* **117** 444–454.

Eichinger, B. and Kirch, C. (2018). A mosum procedure for the estimation of multiple random change points. *Bernoulli* **24** 526–564.

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

Gösmann, J., Stoehr, C., Heiny, J. and Dette, H. (2022). Sequential change point detection in high dimensional time series. *Electronic Journal of Statistics* **16** 3608–3671.

Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association* **105** 1480–1493.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19** 293–325.

Horváth, L. and Hušková, M. (2012). Change-point detection in panel data. *Journal of Time Series Analysis* **33** 631–648.

Janssen, P. (1994). Weighted bootstrapping of U-statistics. *Journal of Statistical Planning and Inference* **38** 31–41.

Jiang, F., Wang, R. and Shao, X. (2023). Robust inference for change points in high dimension. *Journal of Multivariate Analysis* **193** 105114.

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

Kaul, A., Fotopoulos, S. B., Jandhyala, V. K. and Safikhani, A. (2021). Inference on the change point under a high dimensional sparse mean shift. *Electronic Journal of Statistics* **15** 71–134.

Kaul, A. and Michailidis, G. (2025). Inference for Change Points in High Dimensional Mean Shift Models. *Statistica Sinica* **35** 1–25.

Killick, R., Fearnhead, P. and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* **107** 1590–1598.

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

Liu, B., Zhou, C., Zhang, X.-S. and Liu, Y. (2020). A unified data-adaptive framework for high dimensional change point detection. *Journal of Royal Statistical Society, Series B* **82** 933–963.

Liu, H., Gao, C. and Samworth, R. J. (2021). Minimax rates in sparse, high-dimensional change point detection. *The Annals of Statistics* **49** 1081–1112.

Verzelen, N., Fromont, M., Lerasle, M. and Reynaud-Bouret, P. (2023). Optimal change-point detection and localization. *The Annals of Statistics* **51** 1586–1610.

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

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

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

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

Yu, M. and Chen, X. (2022). A robust bootstrap change point test for high-dimensional location parameter. *Electronic Journal of Statistics* **16** 1096–1152.

Zhang, Y., Wang, R. and Shao, X. (2022). Adaptive inference for change points in high-dimensional data. *Journal of the American Statistical Association* **117** 1751–1762.

Zhou, C., Zhou, W.-X., Zhang, X.-S. and Liu, H. (2018). A unified framework for testing high dimensional parameters: a data-adaptive approach. *Preprint arXiv:1808.02648* .