A distribution-free method for change point detection in non-sparse high dimensional data

Reza Drikvandi[1] and Reza Modarres[2]

[1]*Department of Mathematical Sciences, Durham University, UK*

[2]*Department of Statistics, George Washington University, USA*

Abstract

We propose a distribution-free distance-based method for high dimensional change points that can address challenging situations when the sample size is very small compared to the dimension as in the so-called HDLSS data or when non-sparse changes may occur due to change in many variables but with small significant magnitudes. Our method can detect changes in mean or variance of high dimensional observations as well as other distributional changes. We present efficient algorithms that can detect single and multiple high dimensional change points. We use nonparametric metrics, including a new dissimilarity measure and some new distance and difference distance matrices, to develop a procedure to estimate change point locations. We also introduce a nonparametric test to determine the significance of estimated change points. We provide theoretical guaranties for our method and demonstrate its empirical performance in comparison with some of the recent methods for high dimensional change points. An R package called `HDDchangepoint` is developed to implement the proposed method.

Keywords: *Difference distance matrix; Dissimilarity measure; High dimensional change point; High dimensional data; Wild binary segmentation.*

1. Introduction

Change point analysis is frequently used in various fields such as economics, finance, engineering, genetics and medical research. The main objective is to detect significant changes in

the underlying distribution of a data sequence. The change point problem is well studied for low dimensional data in the literature, however change point analysis is challenging for high dimensional data which are growing in different domains. A change could happen in a small or large subset of the variables and with a small or large magnitude, but when the sample size is small compared to the number of variables it is difficult to distinguish a significant change point from just random variability.

In recent years there has been increasing interest in the so-called high-dimensional low-sample-size (HDLSS) data where the sample size n is very small while the dimension p can be very large. The asymptotics for HDLSS data is different than the usual high dimensional asymptotic setting (i.e., $p > n \to \infty$) in the sense that $p \to \infty$ but the sample size n can remain fixed (e.g., Hall et al., 2005; Jung and Marron, 2009; Li, 2020). Detecting change points is more challenging for HDLSS data with small samples. In this paper, we will see that recent methods for high dimensional change points struggle to have a good performance with HDLSS data where sample size is very small compared to dimension.

A common strategy in the literature of high dimensional change points is to simplify the problem by reducing the dimension of data so a simpler algorithm such as a univariate change point algorithm can be applied to the transformed data. This may be done using random projection (e.g., Wang and Samworth, 2018; Hahn et al., 2020), geometric mapping (e.g., Grundy et al., 2020) or any other relevant techniques such as principal components (e.g., Xiao et al., 2019), factor analysis (e.g., Barigozzi et al., 2018) and regularisation approaches (e.g., Lee et al., 2016; Safikhani and Shojaie, 2022). This strategy relies on sparsity, often a significant amount of sparsity, to maintain its oracle performance. Such techniques may not be suitable for non-sparse change point problems. By non-sparse change points, we mean the changes that can happen in many variables and with small significant magnitudes.

Another strategy is to search for change point locations through dissimilarity distances between pairs of observations. This may be done using appropriate dissimilarity measures

such as the interpoint distances (e.g., Li, 2020) or the divergence measures based on Euclidean distance (e.g., Matteson and James, 2014). Some authors including Garreau and Arlot (2018) and Chu and Chen (2019) indirectly use interpoint distances to develop methods based on counting the number of edges from a certain similarity graph constructed from interpoint distances. The performance, especially power, of such methods generally depends on distance measures and test statistics used, as also discussed in Li (2020). In this paper, we aim to develop a novel powerful method for detecting high dimensional change points based on dissimilarity measures, especially suitable for HDLSS data.

There has been a growing literature on change point analysis for high dimensional data in recent years. We review some of the recent methods focusing on offline change point detection. Chen and Zhang (2015), Garreau and Arlot (2018) and Chu and Chen (2019) developed change-point detection methods using kernels and similarity graphs. Wang and Samworth (2018) proposed a two-stage procedure called inspect for estimation of change points, which is based on random projection and sparsity. Their approach assumes the mean structure changes in a sparse subset of the variables and requires the normality assumption. Enikeeva and Harchaoui (2019) suggested a method for detecting a sparse change in mean in a sequence of high dimensional vectors. This method was studied for a single change point detection and was developed based on the normality assumption. Grundy et al. (2020) used geometric mapping to project the high dimensional data to two dimensions, namely distances and angles. Their approach requires the normality assumption. Liu et al. (2020) developed a general data-adaptive framework by extending the classical CUSUM statistic to construct a U-statistic-based CUSUM matrix. Their approach is based on the independence assumption for variables and is mainly studied for single change point detection. Li (2020) proposed an asymptotic distribution-free distance-based procedure for change point detection in high dimensional data. Using the random projection approach of Wang and Samworth (2018), Hahn et al. (2020) proposed a Bayesian algorithm to efficiently estimate the projection direction for

change point detection. Yu and Chen (2021) established a bootstrap CUSUM test for finite sample change point inference in high dimensions. Follain et al. (2022) extended the random projection approach to estimate change points with heterogeneous missingness. There have also been a few other methods, some of which are reviewed in Liu et al. (2022).

Below we highlight the importance of our work and summarise our main contributions.

i) Change point detection for HDLSS data, when the sample size is very small, is challenging and understudied. Also, many of the existing methods focused on sparse change points (e.g., Wang and Samworth, 2018; Enikeeva and Harchaoui, 2019; Follain et al., 2022). Our method can deal with non-sparse high dimensional situations. Unlike existing methods which mainly detect changes in mean of high dimensional observations, our method can detect changes in mean or variance of observations and other distributional changes.

ii) Unlike most of the recent methods for high dimensional change point detection which require the normality assumption as reviewed above, our approach does not require normality or any other distribution for variables. We use novel nonparametric tools to develop a method to detect change point locations.

iii) Many of the recent methods in the literature either used a bootstrap/permutation procedure or derived asymptotic distribution to test significance of change points (e.g., Wang and Samworth, 2018; Yu and Chen, 2021). We establish both asymptotic and permutation procedures for our method to handle both small and large sample size situations.

iv) Our strategy is to search for change point locations through an $n \times n$ matrix of distances instead of the $n \times p$ matrix of observations. Because the $n \times n$ matrix of distances has a smaller dimension which does not grow with p, it would be easier mathematically and computationally to investigate the distance matrix for a change point location.

We use the following notation throughout the paper. For a vector $\boldsymbol{u} \in \mathbb{R}^p$, we write the L_q-norm as $\|\boldsymbol{u}\|_q = \left(\sum_{j=1}^p |\boldsymbol{u}_j|^q\right)^{1/q}$. The infinity norm is defined as $\|\boldsymbol{u}\|_\infty = \max_j |\boldsymbol{u}_j|$. We

write $\surd$ as the short version of square root. For a real-value constant a, notation $|a|$ denotes the absolute value of a, while for a set C, $|C|$ denotes the cardinality of C. Also, $O(.)$ and $o(.)$ denote the usual big O and little o notation. We sometimes write $O^p(.)$ or $o^p(.)$ to emphasise them for $p \to \infty$. We write $O_P(.)$ and $o_P(.)$ to denote, respectively, the big O in probability and the little o in probability. We use $\xrightarrow{P}$ and $\xrightarrow{D}$ for convergence in probability and convergence in distribution, respectively. We also define some more specific notation as we develop the paper.

2. Methodology

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a sequence of independent p-dimensional random vectors with unknown probability distributions $F_1, F_2, \ldots, F_n$, respectively. In high dimensional settings we have $p \gg n$ and that the p variables in each observation $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \ldots, \boldsymbol{X}_{ip})$ are potentially correlated. We aim to develop a distribution-free approach, so we do not assume a parametric form for the distributions $F_1, F_2, \ldots, F_n$.

For the sake of clarity, we first present the proposed approach for detecting a single change point in high dimensions and then extend the idea for multiple change point detection in Section 3. The problem of detecting a single change point in general can be formulated as the following hypothesis testing problem

$$\begin{cases} H_0 : F_1 = F_2 = \ldots = F_n \\ H_1^s : F_1 = \ldots = F_{\tau-1} \neq F_\tau = \ldots = F_n, \end{cases} \tag{1}$$

where τ is a change point location, which is unknown too. When conducting this single change point problem, we first estimate the change point location τ and then test for significance of the estimated change point.

Let $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n]^T$ represent the entire data as an $n \times p$ matrix. As discussed in the introduction, the problem of change point detection in high dimensional situations is challenging, especially when n is very small compared to p. In our approach, instead of searching in the $n \times p$ space, we translate the problem into the lower dimensional space $n \times n$

without reducing the dimension of data. This is different than the dimensionality reduction techniques such as random projection (e.g., Wang and Samworth, 2018), geometric mapping (e.g., Grundy et al., 2020), principal components (e.g., Xiao et al., 2019) and factor analysis (e.g., Barigozzi et al., 2018) which reduce the dimension of data. We propose a distance-based method based on dissimilarity measures to find the change point location in the original data $\boldsymbol{X}$ through the distance matrix of $\boldsymbol{X}$. The dimension of the observed data $\boldsymbol{X}$ is $n \times p$ while the dimension of its distance matrix is $n \times n$, so it is easier mathematically and computationally to investigate the distance matrix for a change point location as n is small compared to p in high dimensions, especially in HDLSS data. We describe the idea in the sequel.

Let $d_{ij} := d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ be a dissimilarity distance between observations $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$, where $d(\cdot, \cdot)$ is a suitable dissimilarity distance function for high dimensional vectors. Since our proposal can be implemented with any suitable dissimilarity distance, we discuss the choice of dissimilarity distance measure after illustrating the main mechanism. We use the dissimilarity measure d_{ij} to obtain the $n \times n$ distance matrix between all pairs of the observations $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ as follows

$$D = \begin{bmatrix} d_{11} & d_{12} & \ldots & d_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \ldots & d_{nn} \end{bmatrix},$$

in which $d_{11} = \ldots = d_{nn} = 0$.

To find the location of change point τ, we propose a procedure that finds an estimate of τ by finding the maximum distance between observations in the distance matrix D. For this, we define the distance difference between each pair of the observations as

$$\Delta_{ij} := |d_{ij} - d_{i,j-1}|, \qquad i = 1, \ldots, n,\ j = 2, \ldots, n$$

$$\Delta_{i1} := 0, \qquad i = 1, \ldots, n.$$

This gives the following $n \times n$ matrix of distance differences, which we call difference distance

matrix,

$$\Delta = \begin{bmatrix} 0 & \Delta_{12} & \dots & \Delta_{1,n-1} & \Delta_{1n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \Delta_{n2} & \dots & \Delta_{n,n-1} & \Delta_{nn} \end{bmatrix}.$$

We then suggest an estimate of the change point location τ, based on the difference distance matrix Δ, as follows

$$\hat{\tau} = \underset{1 \le j \le n}{\arg\max} \Big\{ \frac{1}{n} \sum_{i=1}^{n} \Delta_{ij} \Big\}. \tag{2}$$

The change point estimate $\hat{\tau}$ is the location of maximum slope among the column sums of the difference distance matrix Δ (a simple illustrative example will be given later in Figure 1). If $\frac{1}{n}\sum_{i=1}^{n} \Delta_{ij} = 0$ for all $j = 1, \dots, n$ or if they all are equal, then we set $\hat{\tau} = \emptyset$, where the empty set $\emptyset$ means no change point is detected.

It is known that the Euclidean distance is not appropriate for high dimensional situations. A modified version of Euclidean distance can be obtained by dividing it by $\sqrt{p}$ for convergence guarantees in high dimensions, as suggested by Hall et al. (2005). One can use the modified Euclidean distance $d_{ij} = d(\boldsymbol{X}_i, \boldsymbol{X}_j) = p^{-1/2}\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2$ or other L_q-norm distances in our approach. However, a more general dissimilarity measure that takes into account the information of distances from all the $n-2$ other observations can be defined as

$$d_{ij} = d(\boldsymbol{X}_i, \boldsymbol{X}_j) = \frac{1}{n-2} \sum_{l \neq i,j} \big| \|\boldsymbol{X}_i - \boldsymbol{X}_l\|_D - \|\boldsymbol{X}_j - \boldsymbol{X}_l\|_D \big|, \tag{3}$$

where $\|\cdot - \cdot\|_D$ can be any appropriate distance function. We here suggest two simple options for this. A natural multivariate option is to use the modified Euclidean distance function $\|\boldsymbol{X}_u - \boldsymbol{X}_l\|_D = p^{-1/2}\|\boldsymbol{X}_u - \boldsymbol{X}_l\|_2$ or modified L_1-norm $\|\boldsymbol{X}_u - \boldsymbol{X}_l\|_D = p^{-1}\|\boldsymbol{X}_u - \boldsymbol{X}_l\|_1$. A univariate option is to use a distance function based on differences between the sample mean and variance of the observations. For this, let $\bar{\boldsymbol{X}}_i^p = p^{-1}\sum_{j=1}^{p} \boldsymbol{X}_{ij}$ and $S^p_{\boldsymbol{X}_i} = \sqrt{p^{-1}\sum_{j=1}^{p}\big(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i^p\big)^2}$ denote, respectively, the mean and standard deviation of the i-th observation $\boldsymbol{X}_i, i = 1, \dots, n$. We define $\|\boldsymbol{X}_u - \boldsymbol{X}_l\|_D = \sqrt{(\bar{\boldsymbol{X}}_u^p - \bar{\boldsymbol{X}}_l^p)^2 + (S^p_{\boldsymbol{X}_u} - S^p_{\boldsymbol{X}_l})^2}$, which quantifies the differences between the sample mean and variance of each pair of observations.

The following theorem shows the dissimilarity measure $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ in (3) is a pseudometric, that is $d(\boldsymbol{X}_i, \boldsymbol{X}_j) \geq 0$, $d(\boldsymbol{X}_i, \boldsymbol{X}_i) = 0$, $d(\boldsymbol{X}_i, \boldsymbol{X}_j) = d(\boldsymbol{X}_j, \boldsymbol{X}_i)$ and $d(\boldsymbol{X}_i, \boldsymbol{X}_j) \leq d(\boldsymbol{X}_i, \boldsymbol{X}_l) + d(\boldsymbol{X}_j, \boldsymbol{X}_l)$. The proof along with all other technical proofs are provided in Appendix A.

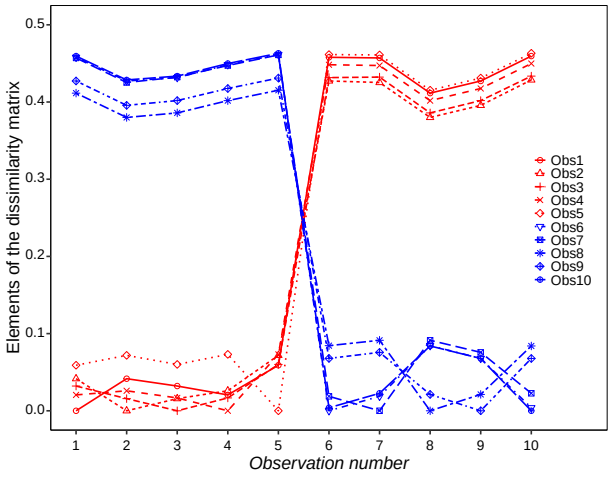
Theorem 1. *The dissimilarity measure $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ in (3) is a pseudometric on $\boldsymbol{X}$ for $n \geq 3$.*

The following theorem indicates that proximity in terms of the modified Euclidean distance implies proximity in terms of $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$, but the converse implication does not hold.

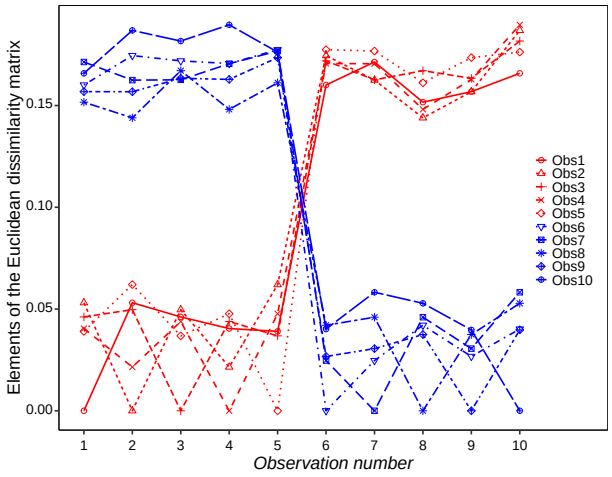
Theorem 2. *Consider the dissimilarity measure $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ in (3) and the Euclidean distance $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2$ for each pair of observations $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$. We have $d(\boldsymbol{X}_i, \boldsymbol{X}_j) \leq \frac{1}{\sqrt{p}}\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2$.*

In Section 4, we show that $d(\boldsymbol{X}_i, \boldsymbol{X}_j) \to 0$ if $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ have the same distribution. The computational cost of the dissimilarity measure $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ is of order $O(np)$. The cost of computing $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ for all $1 \leq i < j \leq n$ is $O(n^3 p)$ with n not being fixed. The cost is $O(p)$ with fixed n as in HDLSS data.

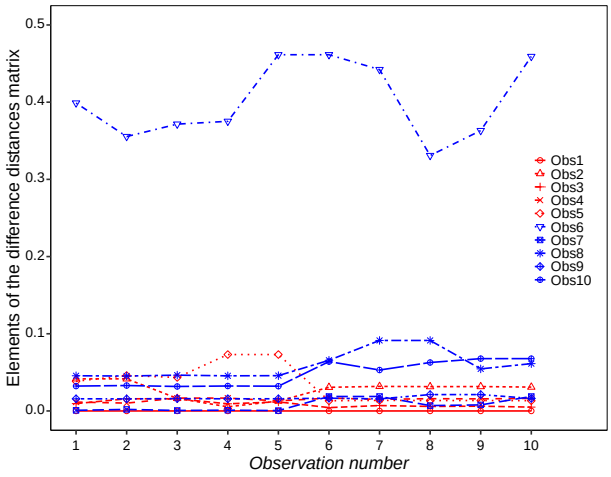
We here provide a simple illustrative example with one change point for which we generate 5 observations from a standard multivariate normal distribution with $p = 500$ and another 5 observations from the same distribution but with the mean being shifted by 0.5. Figures 1a and 1b visualise all elements of the distance matrix D obtained using the dissimilarity measure $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ in (3) with both the univariate distance function based on differences of the sample mean and variance of observations and the multivariate distance function based on the modified Euclidean distance. From both plots, it can be seen that the observations from the first distribution show a larger dissimilarity with the observations from the second distribution, and vice versa, so both distance functions capture the change point trend. More interestingly, Figure 1c visualises all elements of the matrix Δ based on the dissimilarity measure (3), which indicates that the change point location (i.e., location 6) has the largest values Δ_{ij}. Also, Figure 1d presents all the column sums of Δ, which reveals the change point location exactly as in the proposed estimate $\hat{\tau}$ in (2). In Section 4, we will prove that $\hat{\tau}$ is consistent for the true change point, denoted by τ_0, under some assumptions.



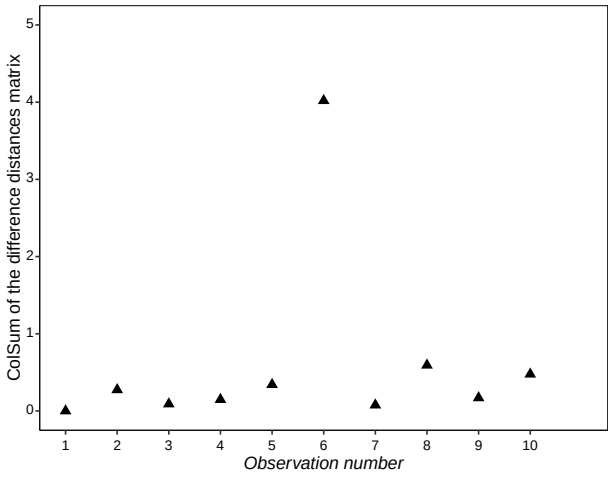
(a) For matrix D with the univariate distance function based on differences of the sample mean and variance



(b) For matrix D with the multivariate distance function based on the modified Euclidean distance



(c) For the difference distance matrix Δ



(d) For the difference distance matrix Δ

Figure 1: Illustrative example with one change point: plots for all elements of the distance matrix D obtained using the dissimilarity measure $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ in (3) with both the univariate and multivariate distance functions, as well as for all column sums of the difference distance matrix Δ.

We now construct a test statistic based on the change point estimate $\hat{\tau}$ to conduct the test whether or not the change point estimate $\hat{\tau}$ is statistically significant. Our proposed test statistic uses the information of d_{ij} and Δ_{ij} before and after the change point estimate $\hat{\tau}$. For this, considering the hypothesis test (1), we first define two sets of indices one for observations before and one for observations after the change point estimate $\hat{\tau}$ as follows

$$\nabla_{\hat{\tau}}^{-} := \{j : 1 \leq j \leq \hat{\tau} - 1\}, \;\; \nabla_{\hat{\tau}}^{+} := \{j : \hat{\tau} \leq j \leq n\}.$$

The two sets $\nabla_{\hat{\tau}}^{-}$ and $\nabla_{\hat{\tau}}^{+}$ are visualised in Figure 2, showing their connection with the change point candidate $\hat{\tau}$. We then propose the following test statistic for testing the significance of

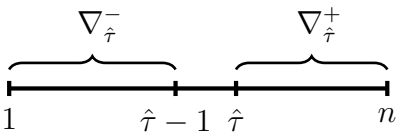


Figure 2: The sets $\nabla^-_{\hat\tau}$ and $\nabla^+_{\hat\tau}$ with the estimated change point location $\hat\tau$.

change point estimate $\hat\tau$

$$T(\hat\tau) = \frac{1}{n|\nabla^-_{\hat\tau}||\nabla^+_{\hat\tau}|}\sum_{i=1}^{n}\sum_{j\in\nabla^-_{\hat\tau}}\sum_{j'\in\nabla^+_{\hat\tau}}\left(d_{ij}-d_{ij'}\right)^2, \tag{4}$$

where $|\nabla^-_{\hat\tau}| = \hat\tau - 1$ and $|\nabla^+_{\hat\tau}| = n - \hat\tau + 1$ are the cardinalities of $\nabla^-_{\hat\tau}$ and $\nabla^+_{\hat\tau}$, respectively.

We will investigate the theoretical properties of the statistic $T(\hat\tau)$ in Section 4. Intuitively, the test statistic $T(\hat\tau)$ quantifies the differences between distances for the observations before change point and the observations after change point. If there is no change point then the differences will be small, and if there is a change point then $T(\hat\tau)$ will deviate from 0. In fact, we will prove that as $p \to \infty$

$$T(\hat\tau)\big|_{H_0} = o_P(1), \quad T(\hat\tau)\big|_{H_1^s} = \kappa^2_{\tau_0} + o_P(1),$$

where κ_{τ_0} is specified in Theorem 3 with τ_0 being the true change point. We can therefore carry out the hypothesis test for a single change point and reject the null hypothesis H_0 for large values of $T(\hat\tau)$. For large n, we conduct the test using the asymptotic distribution of $T(\hat\tau)$ under H_0, when $n, p \to \infty$. In Theorem 7, we will prove that the asymptotic distribution of $T(\hat\tau)$, denoted by $G_T(\cdot)$, is a normal distribution as $n, p \to \infty$ under some assumptions, so that

$$\frac{n|\nabla^-_{\hat\tau}||\nabla^+_{\hat\tau}|\left(T(\hat\tau) - \frac{1}{n|\nabla^-_{\hat\tau}||\nabla^+_{\hat\tau}|}\sum_{i=1}^{n}\sum_{j\in\nabla^-_{\hat\tau}}\sum_{j'\in\nabla^+_{\hat\tau}} m_{ijj'}\right)}{\sqrt{\sum_{i=1}^{n}\sum_{j\in\nabla^-_{\hat\tau}}\sum_{j'\in\nabla^+_{\hat\tau}}\sum_{k=1}^{n}\sum_{l\in\nabla^-_{\hat\tau}}\sum_{l'\in\nabla^+_{\hat\tau}} c_{ijj'kll'}}} \xrightarrow{D} N(0,1),$$

under H_0, where $m_{ijj'}$ and $c_{ijj'kll'}$ are specified in the statement of Theorem 7.

For small n, we use a permutation procedure based on $T(\hat\tau)$ to conduct the test. Permutation testing is useful here because all the observations have the same distribution under H_0 and hence are exchangeable or permutable. In each permutation step, we randomly permute the indices of observations before and after the change point estimate $\hat\tau$, while holding the

change point location, to get a random permutation sample. Based on R permutations, the approximate permutation distribution of the test statistic, denoted by $G_{T_R}(t)$, is defined as

$$G_{T_R}(t) = \frac{1}{R}\sum_{r=1}^{R} I\big(T_r(\hat{\tau}) \leq t\big), \tag{5}$$

where $I(\cdot)$ is the indicator function and $T_r(\hat{\tau})$ is the test statistic calculated for the r-th permutation sample. The p-value of the permutation test, denoted by p_{perm}, is given by

$$p_{\text{perm}} = 1 - G_{T_R}(T_{\text{obs}}(\hat{\tau})) = \frac{1}{R}\sum_{r=1}^{R} I\big(T_r(\hat{\tau}) > T_{\text{obs}}(\hat{\tau})\big), \tag{6}$$

where $T_{\text{obs}}(\hat{\tau})$ is the test statistic for the observed sample. Our theoretical results show that $P_{H_0}(p_{\text{perm}} \leq \alpha) \leq \alpha$ for all $0 \leq \alpha \leq 1$. We will show the asymptotic and permutation tests are equivalent asymptotically under H_0, that is $G_{T_R}(t) \xrightarrow{D} G_T(t)$ for all t as $n, p \to \infty$. Algorithm 1 summarises our proposed method for single change point detection in high dimensional data.

Algorithm 1: Single change point detection

Input : A data sequence or matrix of observations $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n]^T$.
Output: Change point estimate $\hat{\tau}$, or "NA" if there is no significant change point.

Step 1: Calculate the $n \times n$ distance matrix D for the $n \times p$ data matrix $\boldsymbol{X}$.
Step 2: Calculate the $n \times n$ difference distance matrix Δ.
Step 3: Calculate the change point estimate $\hat{\tau}$. If $\hat{\tau} = \emptyset$ stop the algorithm and return "NA" for no detected change point. Otherwise, go to the next step.
Step 4: Calculate the test statistic $T(\hat{\tau})$.
Step 5: Apply the permutation test based on $T(\hat{\tau})$ if n is small, or the asymptotic test if n is large, to test the significance of the change point estimate $\hat{\tau}$.
Step 6: If it is significant, return the change point estimate $\hat{\tau}$. Otherwise, return "NA" for no significant change point.

In addition to estimation and hypothesis test for change point τ, we can also construct confidence interval for change point location τ using the change point estimate $\hat{\tau}$. Finding the exact or asymptotic distribution of the change point estimate (2) is difficult due to the absolute value functions in both the definition of Δ_{ij} and the dissimilarity measure d_{ij} in (3). We instead obtain a permutation-based confidence interval for τ. Unlike the above permutation procedure which conducts under H_0, we here obtain a permutation sample by separately permuting observations before and after the change point location τ among themselves. The reason is

that observations before the change point τ have the same distribution and observations after the change point τ also have the same distribution. We calculate the change point estimate (2) for each permutation sample from R permutations and denote these permutation estimates by $\hat{\tau}_1^*, \ldots, \hat{\tau}_R^*$. Then, a $100(1-\alpha)\%$ confidence interval for change point location τ is

$$(2\hat{\tau} - \hat{\tau}^*_{(1-\alpha/2)}, 2\hat{\tau} - \hat{\tau}^*_{(\alpha/2)}) \tag{7}$$

in which $\hat{\tau}_{(\alpha/2)}$ and $\hat{\tau}_{(1-\alpha/2)}$ are, respectively, the $(\alpha/2)$-th and $(1-\alpha/2)$-th percentiles of the R ordered permutation estimates.

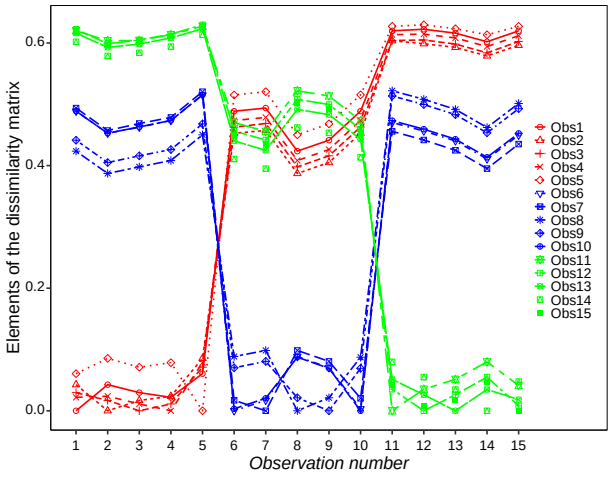
3. Multiple change point detection

In this section, we extend our approach to detect multiple change points in high dimensional data. The problem of detecting multiple change points in general can be formulated as

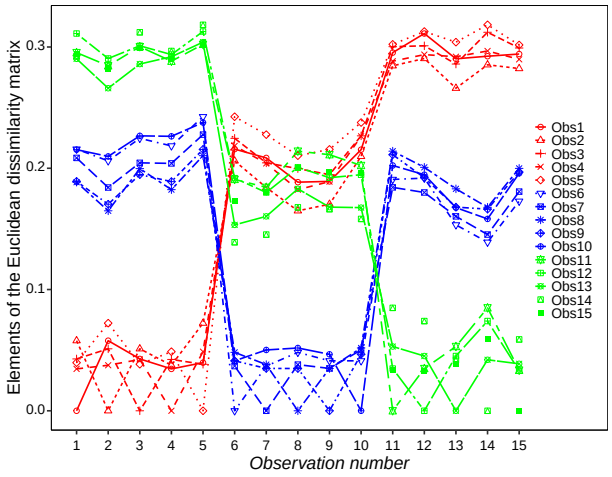
$$\begin{cases} H_0 : F_1 = F_2 = \ldots = F_n \\ H_1^m : F_1 = \ldots = F_{\tau_1 - 1} \neq F_{\tau_1} = \ldots = F_{\tau_2 - 1} \neq F_{\tau_2} = \ldots = F_{\tau_s - 1} \neq F_{\tau_s} = \ldots = F_n, \end{cases} \tag{8}$$

where $1 < \tau_1 < \tau_2 < \ldots < \tau_s < n$ are the unknown change point locations and s is the number of change points, which is also unknown. If the above null hypothesis is rejected, the main objective will be to find the s change point estimates $\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_s$.

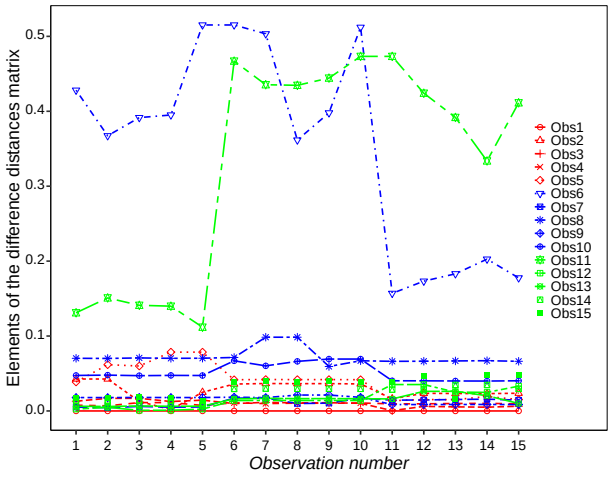
Similar to the illustrative example in Figure 1 for single change point, Figure 3 shows how the dissimilarity measure (3) and the proposed approach based on the difference distance matrix Δ can be helpful for finding multiple change points (here the same settings as previous example but with two true change points at locations 6 and 11). To carry out the problem of multiple change points, we use a recursive binary segmentation procedure on the basis of our method for single change point detection. We also demonstrate combining with the wild binary segmentation procedure of Fryzlewicz (2014) at the end of this section. We sequentially apply the proposed single change point method to uncover all significant change points in the data according to our asymptotic or permutation test. Suppose that s change points are detected



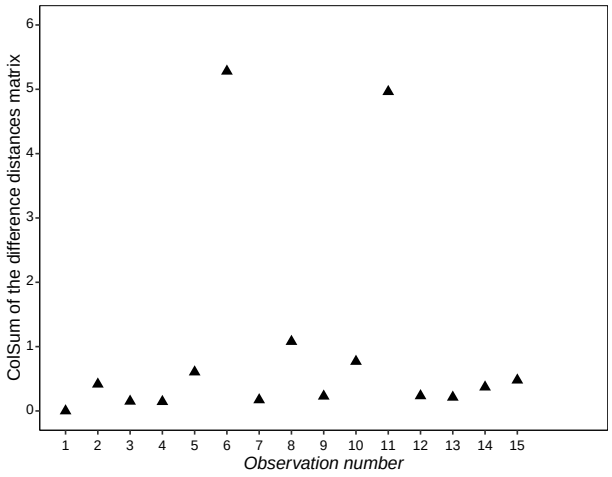
(a) For matrix D with the univariate distance function based on differences of the sample mean and variance



(b) For matrix D with the multivariate distance function based on the modified Euclidean distance



(c) For the difference distance matrix Δ



(d) For the difference distance matrix Δ

Figure 3: Illustrative example with two change points: plots for all elements of the distance matrix D obtained using the dissimilarity measure $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ in (3) with both the univariate and multivariate distance functions, as well as for all column sums of the difference distance matrix Δ.

sequentially and denoted by $\hat{\gamma}_k, k = 1, \ldots, s$. We use the notation $\hat{\gamma}_k$ because the detected change points using binary segmentation are not necessarily in increasing order, when there are more than one change point. Note that $(\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_s) = \text{sort}(\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_s)$.

The recursive binary segmentation starts with applying the single change point algorithm to the data sequence $[\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n]$, and if a change point is detected then the data sequence will be split to two segments (data sub-sequences) before and after the detected change point in order to continue the same process with each data sub-sequence separately for any further change points. Let b_k and e_k denote the beginning and ending indices of observations for a data

sub-sequence, say $[\boldsymbol{X}_{b_k}, \boldsymbol{X}_{b_k+1}, \ldots, \boldsymbol{X}_{e_k}]$, for finding a potential change point $\hat{\gamma}_k$. We apply the single change point algorithm to this data sub-sequence and if a change point location $\hat{\gamma}_k$ is detected, we split the data sequence $[\boldsymbol{X}_{b_k}, \boldsymbol{X}_{b_k+1}, \ldots, \boldsymbol{X}_{e_k}]$ to two sub-sequences before and after the change point $\hat{\gamma}_k$, say $[\boldsymbol{X}_{b_k}, \boldsymbol{X}_{b_k+1}, \ldots, \boldsymbol{X}_{\hat{\gamma}_k-1}]$ and $[\boldsymbol{X}_{\hat{\gamma}_k}, \boldsymbol{X}_{\hat{\gamma}_k+1}, \ldots, \boldsymbol{X}_{e_k}]$ respectively. We apply the single change point algorithm to each of these two sub-sequences to check for further change points. We continue the process until no further change points are detected.

For multiple change point detection, we extend the formula of single change point estimate (2) and write the estimate of the change point location γ_k as follows

$$\hat{\gamma}_k = \underset{b_k \le j \le e_k}{\arg\max} \Big\{ \frac{1}{e_k - b_k + 1} \sum_{i=b_k}^{e_k} \Delta_{ij} \Big\}, \tag{9}$$

where we note that $b_1 = 1$ and $e_1 = n$, implying $e_1 - b_1 + 1 = n$. When checking the significance of the change point estimate $\hat{\gamma}_k$, we need to update both ∇^- and ∇^+ according to the change point estimate $\hat{\gamma}_k$. For this, the two sets of indices for observations before and after the change point estimate $\hat{\gamma}_k$ are expressed as

$$\nabla^-_{\hat{\gamma}_k} = \{j : b_k \le j \le \hat{\gamma}_k - 1\}, \;\; \nabla^+_{\hat{\gamma}_k} = \{j : \hat{\gamma}_k \le j \le e_k\}.$$

To conduct the test for significance of $\hat{\gamma}_k$, the test statistic can be defined similarly as

$$T(\hat{\gamma}_k) = \frac{1}{(e_k - b_k + 1)|\nabla^-_{\hat{\gamma}_k}||\nabla^+_{\hat{\gamma}_k}|} \sum_{i=b_k}^{e_k} \sum_{j \in \nabla^-_{\hat{\gamma}_k}} \sum_{j' \in \nabla^+_{\hat{\gamma}_k}} \big(d_{ij} - d_{ij'}\big)^2,$$

which simplifies to the test statistic (4) when $k = 1$. Algorithm 2 summarises our method for multiple change point detection.

We also incorporate the wild binary segmentation procedure of Fryzlewicz (2014) in our proposal for multiple change points. For this, following the principle of wild binary segmentation, we first randomly draw a large number of pairs (b_k^*, e_k^*) from the whole domain $\{1, \ldots, n\}$ including the pair $(1, n)$, and find $\arg\max_{b_k^* \le j \le e_k^*} \big\{ \frac{1}{e_k^* - b_k^* + 1} \sum_{i=b_k^*}^{e_k^*} \Delta_{ij} \big\}$ for each draw. The candidate change point location is the one that has the largest $\big\{ \frac{1}{e_k^* - b_k^* + 1} \sum_{i=b_k^*}^{e_k^*} \Delta_{ij} \big\}$ among all the draws. We test the significance of the change point candidate using either the asymptotic or

Algorithm 2: Multiple change point detection

Input : A data sequence or matrix of observations $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n]^T$.
Output: A list of significant change point estimates $\{\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_s\}$, or "NA" if there is no significant change point.

Step 1: Apply the single change point Algorithm 1 to the data sequence $\boldsymbol{X}$. If there is no significant change point, end the process and return "NA". Otherwise, denote the detected change point by $\hat{\gamma}_1$ and go to next step.
Step 2: Split the data sequence to two sub-sequences before and after the detected change point. Apply Algorithm 1 to each of the two sub-sequences separately to check for further change points.
Step 3: Repeat Step 2 until no further sub-sequences have significant change points.
Step 4: Denote all the detected change points by $\{\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_s\}$. Return $(\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_s) = \text{sort}(\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_s)$ as a list of significant change points.

permutation test. If it is significant, the same procedure will be repeated to the left and to the right of it. The process is continued recursively until there is no further significant change point. The theory of wild binary segmentation shows it performs at least as good as the standard binary segmentation. Algorithm 2 then easily updates with the wild binary segmentation.

4. Asymptotic results

We study the asymptotic properties of our method for detecting single and multiple high dimensional change points. We establish the theory with both the multivariate modified Euclidean distance function and the univariate distance function based on differences between the sample mean and variance of observations, although one could see that the theory could similarly follow with other suitable distance measures including other L_q-norm distances. We study the situations when both n and p approach infinity, as well as when p approaches infinity but n remains finite as in HDLSS data.

To develop the asymptotic theory with $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ based on the multivariate modified Euclidean distance function, according to Hall et al. (2005) we make the following three assumptions on observations $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \ldots, \boldsymbol{X}_{ip})$, $i = 1, \ldots, n$.

(A1) Assume that $\max\limits_{1\le i\le n} \max\limits_{1\le j\le p} E(\boldsymbol{X}_{ij}^4) < \infty$.

(A2) Assume $tr(\boldsymbol{\Sigma}_i)/p \to \lambda_i$ and $p^{-1/2}\|E(\boldsymbol{X}_i) - E(\boldsymbol{X}_j)\|_2 \to \eta_{ij}$ for $1 \le i, j \le n$.

(A3) Assume $\sum_{j=1}^{p}\sum_{\substack{j'=1 \\ j\neq j'}}^{p}\mathrm{Cov}(\boldsymbol{X}_{ij}^2, \boldsymbol{X}_{ij'}^2) = o(p^2)$.

Alternatively, for using the univariate distance function based on differences between the sample mean and variance of observations, we make the following assumptions.

(B1) Assume that $\max\limits_{1\le i\le n}\max\limits_{1\le j\le p} E(\boldsymbol{X}_{ij}^2) < \infty$.

(B2) Let $\boldsymbol{S}_{ij}^2 = (\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i^p)^2$ and define $\sigma_{ij}^2 = E(\boldsymbol{S}_{ij}^2)$. Assume $\max\limits_{1\le i\le n}\max\limits_{1\le j\le p} \sigma_{ij}^2 < \infty$.

(B3) For all $\epsilon > 0$ and $i = 1, \ldots, n$, define

$$A_i = \sum_{j=1}^{p} E\big((\boldsymbol{X}_{ij} - \bar{\mu}_{ij})^2\big), \quad B_i = \sum_{j=1}^{p} E\big((\boldsymbol{X}_{ij} - \bar{\mu}_{ij})^2 I(|\boldsymbol{X}_{ij} - \bar{\mu}_{ij}| > \epsilon A_i)\big),$$

$$A_i^* = \sum_{j=1}^{p} E\big((\boldsymbol{S}_{ij}^2 - \sigma_{ij}^2)^2\big), \quad B_i^* = \sum_{j=1}^{p} E\big((\boldsymbol{S}_{ij}^2 - \sigma_{ij}^2)^2 I(|\boldsymbol{S}_{ij}^2 - \sigma_{ij}^2| > \epsilon A_i^*)\big),$$

and assume $\frac{B_i}{A_i} \to 0$ and $\frac{B_i^*}{A_i^*} \to 0$ as $p \to \infty$.

(B4) Assume $\sum_{j=1}^{p}\sum_{\substack{j'=1 \\ j\neq j'}}^{p}\mathrm{Cov}(\boldsymbol{X}_{ij}, \boldsymbol{X}_{ij'}) = o(p^2)$ and $\sum_{j=1}^{p}\sum_{\substack{j'=1 \\ j\neq j'}}^{p}\mathrm{Cov}(\boldsymbol{S}_{ij}^2, \boldsymbol{S}_{ij'}^2) = o(p^2)$ as $p \to \infty$.

(B5) Assume $\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j\neq k\neq l}}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathrm{Cov}\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2, \|\boldsymbol{X}_k - \boldsymbol{X}_l\|_D^2\big) = o(n^4)$ as $n \to \infty$.

Assumptions (A1)-(A3) are required for the convergence of the modified Euclidean distance used by Hall et al. (2005), Li (2020), and many others. Assumption (A3) implies weak dependence among variables and is only needed for the case of correlated variables, which is trivial if the variables are independent. Assumptions (B1) and (B2) ensure bounded second moments for convergence of the mean $\bar{\boldsymbol{X}}_i^p$ and standard deviation $S_{\boldsymbol{X}_i}^p$. Assumption (B3) is the usual Lindeberg condition which is a common assumption for applying central limit theorem (a weaker condition compared to Lyapunov condition). Assumptions (B4)-(B5) imply weak dependence among variables and are only required for the case of correlated variables. Assumptions (B4) guarantees bounded covariances for the case of correlated variables $\boldsymbol{X}_{ij}$, which is a typical assumption for dependent variables to satisfy the conditions for central limit theorem and weak law of large numbers. Assumption (B5) is a similar covariance condition for $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2$ which can be simplified since $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2 = (\bar{\boldsymbol{X}}_i^p - \bar{\boldsymbol{X}}_j^p)^2 + (S_{\boldsymbol{X}_i}^p - S_{\boldsymbol{X}_j}^p)^2$. Note that Assumptions (B1) and (B2) imply $\sum_{i=1}^{n}\sum_{j=1}^{n} E\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2\big) < \infty$. We also note that

Assumptions (A3) and (B4)-(B5) hold for variables with the ρ-mixing property (e.g., Utev, 1990) as well as with the spatial dependence (e.g., Wang and Samworth, 2018), so we here consider a more general dependence structure. For instance, under the spatial dependence $\mathrm{Cov}(\boldsymbol{X}_{ij}, \boldsymbol{X}_{ij'}) \propto \rho^{|j-j'|}$ with $|\rho| \leq 1$, it is easy to show $\sum_{j=1}^{p} \sum_{\substack{j'=1 \\ j \neq j'}}^{p} \mathrm{Cov}(\boldsymbol{X}_{ij}, \boldsymbol{X}_{ij'}) = o(p^2)$.

If observations $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ have the same distribution, we can write

$$\lambda_i = \lambda_j := \lambda_{\nabla_\tau^-}, \;\; \eta_{ij} := \eta_{\nabla_\tau^-} \qquad \forall i, j \in \nabla_\tau^-,$$

$$\lambda_i = \lambda_j := \lambda_{\nabla_\tau^+}, \;\; \eta_{ij} := \eta_{\nabla_\tau^+} \qquad \forall i, j \in \nabla_\tau^+,$$

and also

$$\bar{\mu}_i = \bar{\mu}_j := \bar{\mu}_{\nabla_\tau^-}, \;\; \sigma_i = \sigma_j := \sigma_{\nabla_\tau^-} \qquad \forall i, j \in \nabla_\tau^-,$$

$$\bar{\mu}_i = \bar{\mu}_j := \bar{\mu}_{\nabla_\tau^+}, \;\; \sigma_i = \sigma_j := \sigma_{\nabla_\tau^+} \qquad \forall i, j \in \nabla_\tau^+,$$

in which $\bar{\mu}_i = E(\bar{\boldsymbol{X}}_i^p)$ and $\sigma_i = \sqrt{E((S_{\boldsymbol{X}_i}^p)^2)}$ for $i = 1, \ldots, n$. Note that $\bar{\mu}_i = p^{-1} \sum_{j=1}^{p} \mu_{ij}$ and $\sigma_i^2 = p^{-1} \sum_{j=1}^{p} \sigma_{ij}^2$. The following theorem concerns the asymptotic behaviour of the dissimilarity measure $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ for high dimensional observations.

Theorem 3. *Consider the data sequence $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ and dissimilarity measure $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ in (3). Suppose that τ is a change point so that $F_1 = \ldots = F_{\tau-1} \neq F_\tau = \ldots = F_n$. We have*

$$d(\boldsymbol{X}_i, \boldsymbol{X}_j) = o_P(1) \qquad \forall\, i \in \nabla_\tau^-,\, j \in \nabla_\tau^- \;\; or \;\; \forall\, i \in \nabla_\tau^+,\, j \in \nabla_\tau^+,$$

$$d(\boldsymbol{X}_i, \boldsymbol{X}_j) = \kappa_\tau + o_P(1) \qquad \forall\, i \in \nabla_\tau^-,\, j \in \nabla_\tau^+ \;\; or \;\; \forall\, i \in \nabla_\tau^+,\, j \in \nabla_\tau^-,$$

as $p \to \infty$, where for the modified Euclidean distance function, under Assumptions (A1)-(A3),

$$\kappa_\tau = \frac{1}{N-2} \Big\{ \big(|\nabla_\tau^-| - 1\big) \big| \sqrt{\lambda_{\nabla_\tau^-}^2 + \lambda_{\nabla_\tau^+}^2 + \eta_{\nabla_\tau^- \nabla_\tau^+}^2} - \sqrt{2\lambda_{\nabla_\tau^-}^2} \big| + \big(|\nabla_\tau^+| - 1\big) \big| \sqrt{\lambda_{\nabla_\tau^-}^2 + \lambda_{\nabla_\tau^+}^2 + \eta_{\nabla_\tau^- \nabla_\tau^+}^2} - \sqrt{2\lambda_{\nabla_\tau^+}^2} \big| \Big\},$$

while for the univariate distance function based on differences between the sample mean and variance, under Assumptions (B1)-(B4),

$$\kappa_\tau = \big| \sqrt{(\bar{\mu}_{\nabla_\tau^-} - \bar{\mu}_{\nabla_\tau^+})^2 + (\sigma_{\nabla_\tau^-} - \sigma_{\nabla_\tau^+})^2} \big|.$$

The next two theorems provide guaranties for consistency of the proposed single change point estimate (2) for high dimensional observations when $p \to \infty$ and n is fixed, as well as when $p \to \infty$ and n diverges too.

Theorem 4. *Suppose that there is a true single change point τ_0, $2 \le \tau_0 \le n$, so that $F_1 = \ldots = F_{\tau_0 - 1} \ne F_{\tau_0} = \ldots = F_n$. Under the assumptions of Theorem 3, (a) we have as $p \to \infty$, with any fixed n,*

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} \Delta_{ij} = o_P(1) & \text{if } j \ne \tau_0, \\ \\ \frac{1}{n}\sum_{i=1}^{n} \Delta_{ij} = \kappa_{\tau_0} + o_P(1) & \text{if } j = \tau_0, \end{cases}$$

where $\kappa_{\tau_0} = O(1)$ under Assumptions (A1)-(A2) or (B1)-(B2).

(b) it follows $\hat{\tau} - \tau_0 = o_P(1)$, hence the change point estimate $\hat{\tau} = \arg\max_{1 \le j \le n} \left\{ \frac{1}{n}\sum_{i=1}^{n} \Delta_{ij} \right\}$ is consistent for τ_0 as $p \to \infty$.

From this theorem, the consistency of the proposed change point estimate holds when $p \to \infty$ and n is fixed as in HDLSS data. In the following result, we show that the consistency also holds when $n \to \infty$ as well, as one expects.

Theorem 5. *Under the conditions in Theorem 4, we have as $n \to \infty$ and $p \to \infty$*

$$\max_{1 \le j \le n} \left\{ \frac{1}{n}\sum_{i=1}^{n} \Delta_{ij} - I(j = \tau_0)\kappa_{\tau_0} \right\} = o_P(1).$$

The next theorem studies the asymptotic limit of the proposed test statistic (4) under the null hypothesis H_0 as well as under the alternative hypothesis H_1^s.

Theorem 6. *Suppose that there is a true single change point τ_0, $2 \le \tau_0 \le n$, and consider the test statistic $T(\hat{\tau})$ in (4) based on the change point estimate $\hat{\tau} = \arg\max_{1 \le j \le n} \left\{ \frac{1}{n}\sum_{i=1}^{n} \Delta_{ij} \right\}$. Under the above assumptions, we have as $p \to \infty$*

$$T(\hat{\tau})\big|_{H_0} = o_P(1), \quad T(\hat{\tau})\big|_{H_1^s} = \kappa_{\tau_0}^2 + o_P(1).$$

The following theorem proves that the asymptotic distribution of the test statistic (4) is a normal distribution under above conditions when both n and p go to infinity.

Theorem 7. *Consider the test statistic $T(\hat{\tau})$ in (4) for testing a significant change point. Under the null hypothesis H_0 and the above assumptions, we have as $n \to \infty$ and $p \to \infty$*

$$T_{sd}(\hat{\tau}) := \frac{n|\nabla_{\hat{\tau}}^{-}||\nabla_{\hat{\tau}}^{+}|\Big(T(\hat{\tau}) - \frac{1}{n|\nabla_{\hat{\tau}}^{-}||\nabla_{\hat{\tau}}^{+}|}\sum_{i=1}^{n}\sum_{j\in\nabla_{\hat{\tau}}^{-}}\sum_{j'\in\nabla_{\hat{\tau}}^{+}} m_{ijj'}\Big)}{\sqrt{\sum_{i=1}^{n}\sum_{j\in\nabla_{\hat{\tau}}^{-}}\sum_{j'\in\nabla_{\hat{\tau}}^{+}}\sum_{k=1}^{n}\sum_{l\in\nabla_{\hat{\tau}}^{-}}\sum_{l'\in\nabla_{\hat{\tau}}^{+}} c_{ijj'kll'}}} \xrightarrow{D} N(0,1),$$

where $T_{sd}(\hat{\tau})$ is defined as standardised test statistic. For the univariate distance function we have $m_{ijj'} = v_{ij} + v^*_{ij} + v_{ij'} + v^*_{ij'} + o^p(1)$ and

$$\begin{aligned} c_{ijj'kll'} &= Cov\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2_D + \|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|^2_D, \|\boldsymbol{X}_k - \boldsymbol{X}_l\|^2_D + \|\boldsymbol{X}_k - \boldsymbol{X}_{l'}\|^2_D\big) \\ &= Cov\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2_D, \|\boldsymbol{X}_k - \boldsymbol{X}_l\|^2_D\big) + Cov\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2_D, \|\boldsymbol{X}_k - \boldsymbol{X}_{l'}\|^2_D\big) \\ &\quad + Cov\big(\|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|^2_D, \|\boldsymbol{X}_k - \boldsymbol{X}_l\|^2_D\big) + Cov\big(\|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|^2_D, \|\boldsymbol{X}_k - \boldsymbol{X}_{l'}\|^2_D\big), \end{aligned}$$

where

$$Cov\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2_D, \|\boldsymbol{X}_k - \boldsymbol{X}_l\|^2_D\big) = \begin{cases} 2(v^2_{ij} + v^{*2}_{ij}) + o^p(1) & \text{if } i \neq j, k \neq l, i = k, j = l, \\ (v^{*2}_{ii} + v^{*2}_{ii})/2 + o^p(1) & \text{if } i \neq j, k \neq l, i = k, j \neq l, \\ (v^{*2}_{jj} + v^{*2}_{jj})/2 + o^p(1) & \text{if } i \neq j, k \neq l, i \neq k, j = l, \\ o^p(1) & \text{otherwise,} \end{cases}$$

and $v_{ij} = Var(\bar{\boldsymbol{X}}^p_i) + Var(\bar{\boldsymbol{X}}^p_j)$ and $v^*_{ij} = Var(S^p_{\boldsymbol{X}_i}) + Var(S^p_{\boldsymbol{X}_j})$. For the multivariate modified Euclidean distance function we have $m_{ijj'} = E(p^{-1}\|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2_2) + E(p^{-1}\|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|^2_2)$ and

$$\begin{aligned} c_{ijj'kll'} &= Cov\big(p^{-1}\|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2_2 + p^{-1}\|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|^2_2, p^{-1}\|\boldsymbol{X}_k - \boldsymbol{X}_l\|^2_2 + p^{-1}\|\boldsymbol{X}_k - \boldsymbol{X}_{l'}\|^2_2\big) \\ &= Cov\big(p^{-1}\|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2_2, p^{-1}\|\boldsymbol{X}_k - \boldsymbol{X}_l\|^2_2\big) + Cov\big(p^{-1}\|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2_2, p^{-1}\|\boldsymbol{X}_k - \boldsymbol{X}_{l'}\|^2_2\big) \\ &\quad + Cov\big(p^{-1}\|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|^2_2, p^{-1}\|\boldsymbol{X}_k - \boldsymbol{X}_l\|^2_2\big) + Cov\big(p^{-1}\|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|^2_2, \|\boldsymbol{X}_k - \boldsymbol{X}_{l'}\|^2_2\big). \end{aligned}$$

Remark. We use estimates of the constant quantities v_{ij} and v^*_{ij} to calculate $m_{ijj'}$ and $c_{ijj'kll'}$. Following Slutsky's theorem, the result of Theorem 7 also holds when plugging in consistent estimates of the quantities v_{ij} and v^*_{ij}. For the univariate distance case, we get a consistent estimate of $\mathrm{Var}(\bar{\boldsymbol{X}}^p_i)$ in v_{ij} using (under Assumption (B4))

$$\mathrm{Var}(\bar{\boldsymbol{X}}^p_i) = \frac{1}{p^2}\sum_{l=1}^{p}\sum_{l'=1}^{p}\mathrm{Cov}(\boldsymbol{X}_{il}, \boldsymbol{X}_{il'}) = \frac{1}{p^2}\sum_{l=1}^{p}\mathrm{Var}(\boldsymbol{X}_{il}) + \frac{o(p^2)}{p^2}$$

and $p^{-1}\sum_{l=1}^{p}(\boldsymbol{X}_{il} - \bar{\boldsymbol{X}}^p_i)^2 - p^{-1}\sum_{l=1}^{p}\mathrm{Var}(\boldsymbol{X}_{il}) \xrightarrow{P} 0$. Similarly, we get a consistent estimate of

$\mathrm{Var}(S^p_{\boldsymbol{X}_i})$ in v^*_{ij} by first using (under Assumption (B4))

$$\begin{aligned}
\mathrm{Var}\big((S^p_{\boldsymbol{X}_i})^2\big) &= \frac{1}{p^2}\sum_{l=1}^{p}\sum_{l'=1}^{p}\mathrm{Cov}\big((\boldsymbol{X}_{il}-\bar{\boldsymbol{X}}^p_i)^2,(\boldsymbol{X}_{il'}-\bar{\boldsymbol{X}}^p_i)^2\big)\\
&= \frac{1}{p^2}\sum_{l=1}^{p}\mathrm{Var}\big((\boldsymbol{X}_{il}-\bar{\boldsymbol{X}}^p_i)^2\big)+\frac{o(p^2)}{p^2}\\
&= \frac{1}{p^2}\sum_{l=1}^{p}\mathrm{E}\big((\boldsymbol{X}_{il}-\bar{\boldsymbol{X}}^p_i)^4\big)-\frac{1}{p^2}\sum_{l=1}^{p}\big(\mathrm{Var}(\boldsymbol{X}_{il})\big)^2+o(1)\\
&= \frac{1}{p^2}\sum_{l=1}^{p}\mathrm{E}\big((\boldsymbol{X}_{il}-\bar{\boldsymbol{X}}^p_i)^4\big)-p\big(\mathrm{Var}(\bar{\boldsymbol{X}}^p_i)\big)^2+o(1)
\end{aligned} \tag{10}$$

and $p^{-1}\sum_{l=1}^{p}(\boldsymbol{X}_{il}-\bar{\boldsymbol{X}}^p_i)^4 - p^{-1}\sum_{l=1}^{p}\mathrm{E}\big((\boldsymbol{X}_{il}-\bar{\boldsymbol{X}}^p_i)^4\big) \xrightarrow{P} 0$, and then applying the delta method to obtain the required estimate as $\widehat{\mathrm{Var}}(S^p_{\boldsymbol{X}_i}) = \frac{1}{4(S^p_{\boldsymbol{X}_i})^2}\widehat{\mathrm{Var}}\big((S^p_{\boldsymbol{X}_i})^2\big)$. Note that the last equality in (10) is obtained by applying the Taylor expansion $\big(\mathrm{Var}(\boldsymbol{X}_{il})\big)^2 = \big(p\mathrm{Var}(\bar{\boldsymbol{X}}^p_i)\big)^2 + 2p\mathrm{Var}(\bar{\boldsymbol{X}}^p_i)\big(\mathrm{Var}(\boldsymbol{X}_{il}) - p\mathrm{Var}(\bar{\boldsymbol{X}}^p_i)\big) + o(p)$. Also for the modified Euclidean distance case, under Assumption (A3), we use the consistent estimates of $E(\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_2^2)$ and $\mathrm{Cov}\big(\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_2^2,\|\boldsymbol{X}_k-\boldsymbol{X}_l\|_2^2\big)$ as obtained in Li (2020).

We now study the optimality and detection power of the proposed method and obtain conditions for optimal detection rates and full power.

Theorem 8. *Consider the conditions in Theorem 3 and Theorem 7. We have, as $p\to\infty$,*

$$\begin{aligned}
&P_{H_0}(|T_{sd}(\hat{\tau})|>Z_{\alpha/2})\to\alpha,\\
&P_{H_1^s}(|T_{sd}(\hat{\tau})|>Z_{\alpha/2}) = 1-\Phi\big(Z_{\alpha/2}\frac{c}{c^*}-\frac{n^2\kappa^2_{\tau_0}}{c^*}\big)+\Phi\big(-Z_{\alpha/2}\frac{c}{c^*}-\frac{n^2\kappa^2_{\tau_0}}{c^*}\big)+o(1),
\end{aligned}$$

where Φ is the CDF of standard normal distribution, $Z_{\alpha/2}$ is the corresponding critical point of standard normal distribution, $c=\sqrt{\sum_{i=1}^{n}\sum_{j\in\nabla^-_{\hat{\tau}}}\sum_{j'\in\nabla^+_{\hat{\tau}}}\sum_{k=1}^{n}\sum_{l\in\nabla^-_{\hat{\tau}}}\sum_{l'\in\nabla^+_{\hat{\tau}}}c_{ijj'kll'}}$ with $c_{ijj'kll'}$ given in Theorem 7, and $c^*=\sqrt{\sum_{i=1}^{n}\sum_{j\in\nabla^-_{\hat{\tau}}}\sum_{j'\in\nabla^+_{\hat{\tau}}}\sum_{k=1}^{n}\sum_{l\in\nabla^-_{\hat{\tau}}}\sum_{l'\in\nabla^+_{\hat{\tau}}}c^*_{ijj'kll'}}$ with $c^*_{ijj'kll'}=c_{ijj'kll'}+b_{ijkl}+b_{ijkl'}+b_{ijj'kl}+b_{ij'kl'}$ where b_{ijkl} is specified in the proof of theorem.

Under Assumptions (B4)-(B5) or (A3) the covariances are bounded asymptotically, so c and c^* are finite. So, with $\kappa^2_{\tau_0}>0$, the test is consistent and has full power as $n\to\infty$, that is,

$$P_{H_1^s}(|T_{\mathrm{sd}}(\hat{\tau})|>Z_{\alpha/2})\to 1, \quad \text{as } n\to\infty \text{ and } p\to\infty.$$

Also, with fixed n as in HDLSS data, the test is consistent if $\kappa^2_{\tau_0}$ diverges. The following results demonstrate this for two cases when there is a change point in mean and when there is change in variance of observations considering the sparsity level and the change magnitude.

Corollary 1. *Consider a change point in mean of high dimensional observations, while variance remains unchanged, with the mean shift $\delta_k := \mu_{ik} - \mu_{jk}$, $k = 1, \ldots, p$, for all $i \in \nabla^-_\tau, j \in \nabla^+_\tau$. Let $S_0 = \{k : \delta_k \neq 0\}$ be the set of variables having a change point with $s_0 := |S_0|$ being its cardinality or the sparsity level. Then, $\kappa^2_{\tau_0} = \frac{1}{p^2}(\sum_{k\in S_0} \delta_k)^2$ and*

$$P_{H_1^s}(|T_{sd}(\hat{\tau})| > Z_{\alpha/2}) = 1 - \Phi\big(Z_{\alpha/2}\frac{c}{c^*} - \frac{n^2(\sum_{k\in S_0}\delta_k)^2/p^2}{c^*}\big) + \Phi\big(-Z_{\alpha/2}\frac{c}{c^*} - \frac{n^2(\sum_{k\in S_0}\delta_k)^2/p^2}{c^*}\big) + o(1).$$

Hence, with fixed n, the test is consistent if $(\sum_{k\in S_0}\delta_k)^2 > p^2/n^2$. In particular, when $\delta_k = \delta_{\min}$ for all $k \in S_0$ with $\delta_{\min} = \min_{k\in S_0}\delta_k$, the optimality is achieved if

$$s_0 > \frac{p}{n|\delta_{\min}|} \quad or \quad |\delta_{\min}| > \frac{p}{ns_0}.$$

Corollary 2. *Consider a change point in variance of high dimensional observations, while mean remains unchanged, with the variance shift $\omega_k := \sigma^2_{ik} - \sigma^2_{jk}$, $k = 1, \ldots, p$, for all $i \in \nabla^-_\tau, j \in \nabla^+_\tau$. Let $S_0 = \{k : \omega_k \neq 0\}$ be the set of variables having a change point in this case with $s_0 := |S_0|$ being the sparsity level. Then, $\kappa^2_{\tau_0} = \frac{(\sum_{k\in S_0}\omega_k)^2}{p^2(\sigma_{\nabla^-_{\tau_0}} + \sigma_{\nabla^+_{\tau_0}})^2}$. Hence, similar to Corollary 1, with fixed n, the test is consistent if $(\sum_{k\in S_0}\omega_k)^2 > p^2/n^2$, where the assumption of finite variance implies $\sigma_{\nabla^-_{\tau_0}} + \sigma_{\nabla^+_{\tau_0}} < \infty$. In particular, when $\omega_k = \omega_{\min}$ for all $k \in S_0$ with $\omega_{\min} = \min_{k\in S_0}\omega_k$, the optimality is achieved if*

$$s_0 > \frac{p}{n|\omega_{\min}|} \quad or \quad |\omega_{\min}| > \frac{p}{ns_0}.$$

We now demonstrate the consistency of multiple change point estimates from Algorithm 2 for multiple high dimensional change point detection.

Theorem 9. *Suppose there are s true change points $\tau^0_1, \tau^0_2, \ldots, \tau^0_s$, $2 \leq \tau^0_1 < \tau^0_2 < \ldots < \tau^0_s \leq n$, so that $F_1 = \ldots = F_{\tau^0_1 - 1} \neq F_{\tau^0_1} = \ldots = F_{\tau^0_2 - 1} \neq F_{\tau^0_2} = \ldots = F_{\tau^0_{s-1}} \neq F_{\tau^0_s} = \ldots = F_n$. Assume the minimum spacing between change points satisfies $\min_{1\leq i\leq s-1}|\tau^0_{i+1} - \tau^0_i| \geq Mn^\varepsilon$ for some $M > 0$*

and $\varepsilon \leq 1$. *Under the above assumptions, Algorithm 2 returns the change point estimates* $(\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_s) = sort(\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_s)$ *that satisfy* $\|(\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_s) - (\tau_1^0, \tau_2^0, \ldots, \tau_s^0)\|_\infty = o_P(1)$.

Note that the condition $\min_{1 \leq i \leq s-1} |\tau_{i+1}^0 - \tau_i^0| \geq Mn^\varepsilon$ ensures that there are not too many change points to detect, because it implies that $s \leq \left(\min_{1 \leq i \leq s-1} |\tau_{i+1}^0 - \tau_i^0|/M\right)^{1/\varepsilon}$ since $0 \leq s < n$. The following result shows that the asymptotic test and the permutation test are equivalent asymptotically and that the permutation test is also unbiased when $n \to \infty$ and $p \to \infty$.

Theorem 10. *Consider the asymptotic test with the distribution* $G_T(t)$ *obtained in Theorem 7 as well as the permutation test with the approximate permutation distribution* $G_{T_R}(t)$ *in (5) and with the permutation-based p-value* p_{perm} *in (6). Assume the above assumptions hold. Under the null hypothesis* H_0 *of no change points, we have as* $n \to \infty$ *and* $p \to \infty$

(a) $G_{T_R}(t) \xrightarrow{D} G_T(t) \quad \forall\, t,$

(b) $P_{H_0}(p_{perm} \leq \alpha) \leq \alpha \quad \forall\, 0 \leq \alpha \leq 1.$

5. Numerical results

In this section, we evaluate the performance of the proposed methods under various simulation scenarios and in comparison with some of the recent methods in the literature for high dimensional change points. We compare our methods with the nonparametric method proposed by Matteson and James (2014), called E-divisive, the method based on random projection developed by Wang and Samworth (2018), called Inspect, and the nonparametric method of Li et al. (2019) for high dimensional change points, called HDcpdetect. In the simulations, we set the tuning parameter selection of these methods as the recommended defaults in their R packages. We assess the performance of all the methods for detecting a single change point as well as multiple change points under different scenarios, especially for the challenging situation when n is very small compared to p as in HDLSS data. We investigate the performance of our asymptotic and permutation tests using both the univariate and multivariate distance functions. In

the simulations we find that the results of our method with the univariate and multivariate distance functions are quite similar, so for simplicity of presentation and comparisons with the other methods, we avoid presenting duplicate results unless when the results are different as in Section 5.4. Some simulation results are deferred to Appendix B due to space limitation.

5.1. Simulations for a single change in mean

We first start with the case of a single change point in mean of high dimensional observations, where we generate data with n random observations $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ from a p-variate normal distribution, where the first $3n/5$ observations are generated from $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and the other $2n/5$ observations are generated from $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$. We consider different high dimensional settings with $n \in \{45, 90\}$ and $p \in \{500, 1000, 1500\}$, as well as with $\boldsymbol{\mu}_1 = \mathbf{0}_p$ and $\boldsymbol{\mu}_2 \in \{\mathbf{0}_p, (0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4}), (0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})\}$ where $\mathbf{0}_p$ and $\mathbf{1}_p$ denote p-dimensional vectors of zeros and ones, respectively. We here set $\boldsymbol{\Sigma} = \sigma_p^2 \boldsymbol{V}_p$ where $\sigma_p^2 \in \{0.5, 1\}$ and $\boldsymbol{V}_p$ represents the covariance structure of data. In the simulations, we consider two covariance structures: the uncorrelated structure $\boldsymbol{V}_p = \boldsymbol{I}_p$ where $\boldsymbol{I}_p$ is the identity matrix of size p, and the correlated autoregressive structure $\boldsymbol{V}_p = \left[\boldsymbol{V}_{ij}\right]_{i,j=1}^{p} = \left[0.5^{|i-j|}\right]_{i,j=1}^{p}$. Note that the true change point location here is $\tau_1 = 3n/5 + 1$ for all scenarios, except for the scenario when $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}_p$ as it implies there is no change point. We consider 250 replications for each simulation scenario and use $R = 200$ random permutations for the permutation test. We apply each of the change point methods to the generated data sets and record, in addition to the change point estimates, the frequency and average number of detected change points over 250 replications for each method. The results on frequency and average number of the change points detected are presented in Table 1. From the table, it can be seen that all the methods perform well when there is no change point, but for the cases with a true change point our proposed method based on both the asymptotic and permutation tests performs better than all the other methods E-divisive, Inspect and HDcpdetect. On average across the 250 replications, the proposed method detects the change

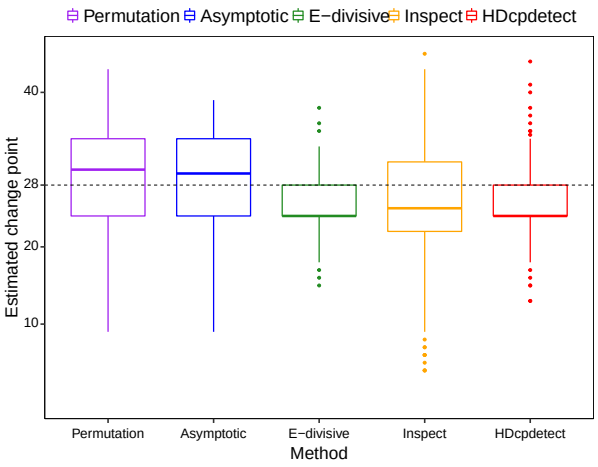
Table 1: Frequency and average number of the detected change points over 250 replications by each of the methods when there is one true change point, also including the case with no change point.

$\boldsymbol{\mu}_2$	n	p	Number of true change points	Method	Frequency of the detected change points		Average number of detected change points
					0	1	
	45	500	0	Permutation	0.98	0.02	0.02
	45	500	0	Asymptotic	0.98	0.02	0.02
$\mathbf{0}_p$	45	500	0	E-divisive	0.97	0.03	0.03
	45	500	0	Inspect	0.97	0.03	0.03
	45	500	0	HDcpdetect	0.98	0.02	0.02
	45	1000	0	Permutation	0.97	0.03	0.03
	45	1000	0	Asymptotic	0.96	0.04	0.04
$\mathbf{0}_p$	45	1000	0	E-divisive	0.95	0.05	0.05
	45	1000	0	Inspect	0.94	0.06	0.06
	45	1000	0	HDcpdetect	0.96	0.04	0.04
	45	1500	0	Permutation	0.95	0.05	0.05
	45	1500	0	Asymptotic	0.95	0.05	0.05
$\mathbf{0}_p$	45	1500	0	E-divisive	0.96	0.04	0.04
	45	1500	0	Inspect	0.94	0.06	0.06
	45	1500	0	HDcpdetect	0.96	0.04	0.04
	45	500	1	Permutation	0.80	0.20	0.20
	45	500	1	Asymptotic	0.73	0.27	0.27
$(0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	45	500	1	E-divisive	0.83	0.17	0.17
	45	500	1	Inspect	0.91	0.09	0.09
	45	500	1	HDcpdetect	0.87	0.13	0.13
	45	1000	1	Permutation	0.42	0.58	0.58
	45	1000	1	Asymptotic	0.33	0.67	0.67
$(0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	45	1000	1	E-divisive	0.67	0.33	0.33
	45	1000	1	Inspect	0.93	0.07	0.07
	45	1000	1	HDcpdetect	0.71	0.29	0.29
	45	1500	1	Permutation	0.16	0.84	0.84
	45	1500	1	Asymptotic	0.13	0.87	0.87
$(0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	45	1500	1	E-divisive	0.38	0.62	0.62
	45	1500	1	Inspect	0.93	0.07	0.07
	45	1500	1	HDcpdetect	0.49	0.51	0.51
	45	500	1	Permutation	0.12	0.88	0.88
	45	500	1	Asymptotic	0.10	0.90	0.90
$(0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	45	500	1	E-divisive	0.14	0.86	0.86
	45	500	1	Inspect	0.77	0.23	0.23
	45	500	1	HDcpdetect	0.15	0.85	0.85
	45	1000	1	Permutation	0.00	1.00	1.00
	45	1000	1	Asymptotic	0.00	1.00	1.00
$(0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	45	1000	1	E-divisive	0.03	0.97	0.97
	45	1000	1	Inspect	0.72	0.28	0.28
	45	1000	1	HDcpdetect	0.05	0.95	0.95
	45	1500	1	Permutation	0.00	1.00	1.00
	45	1500	1	Asymptotic	0.00	1.00	1.00
$(0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	45	1500	1	E-divisive	0.02	0.98	0.98
	45	1500	1	Inspect	0.63	0.37	0.37
	45	1500	1	HDcpdetect	0.02	0.98	0.98

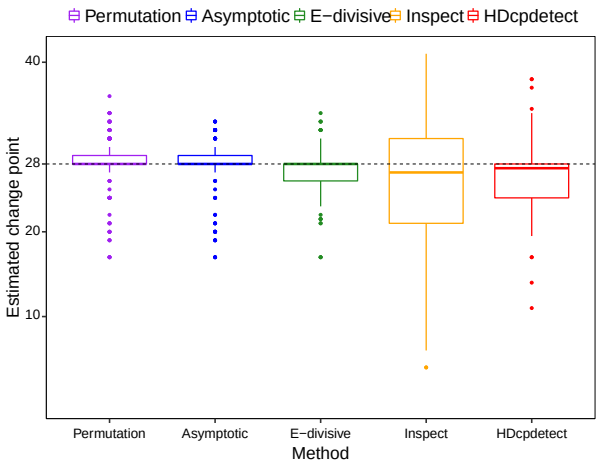
point in a higher frequency compared to the other methods under all scenarios considered. We note that the method Inspect of Wang and Samworth (2018) is not very competitive under such non-sparse high dimensional scenarios because it requires sparsity and performs better with much larger sample sizes n when p is very large (see Wang and Samworth, 2018; Hahn et al., 2020; Follain et al., 2022). Also, Figure 4 presents the box plots of the estimated change point from each of the methods over 250 replications. The box plots show that the proposed method also produces a more accurate change point estimate compared to the other methods. The performance of our method for the correlated autoregressive structure is similar, so we skip the similar results because of space limitation.

5.2. Simulations for multiple changes in mean

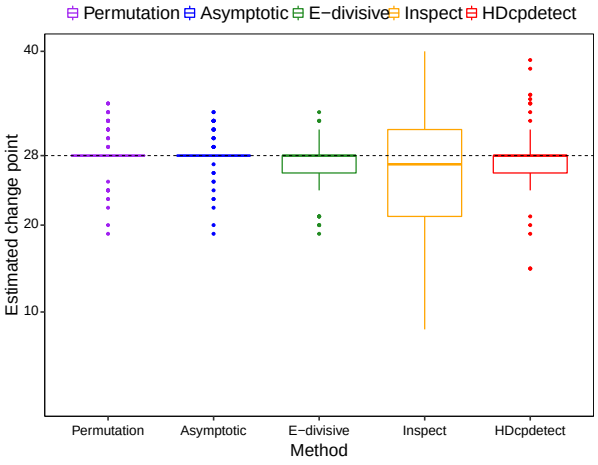
We next consider the case of multiple change points in mean of high dimensional observations, where we use the same simulation settings as before but here with three true change points in the simulated data. For this, we generate data with n random observations $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ from a p-variate normal distribution, where the first $3n/10$ observations are generated from $N(\boldsymbol{\mu}_1 = \mathbf{0}_p, \boldsymbol{\Sigma})$, the next $n/5$ observations are generated from $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, the next $3n/10$ observations are generated from $N(2\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ and the last $n/5$ observations are generated from $N(3\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$. So the three true change point locations here are $\tau_1 = 3n/10 + 1$, $\tau_2 = n/2 + 1$ and $\tau_3 = 4n/5 + 1$. We again use 250 replications for each simulation scenario and apply the proposed method and the other methods to the generated data sets. We here set the minimum segment length to 5. For each method we calculate the frequency and average number of the correctly detected change points over 250 replications. We also calculate the total number of change points detected (correct or incorrect detection). Table 2 shows the results on frequency and average number of the correctly detected change points for each method. From the results in Table 2, we can see that all the methods tend to perform well in the cases when there is no change point. For the cases with three true change points, our method based on both the



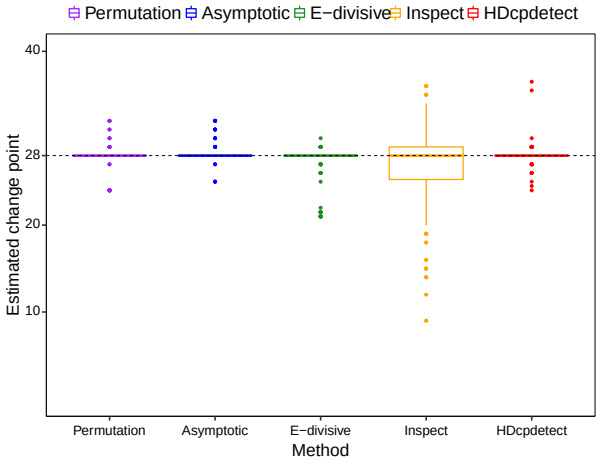
(a) $p = 500$ and $\boldsymbol{\mu}_2 = (0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$



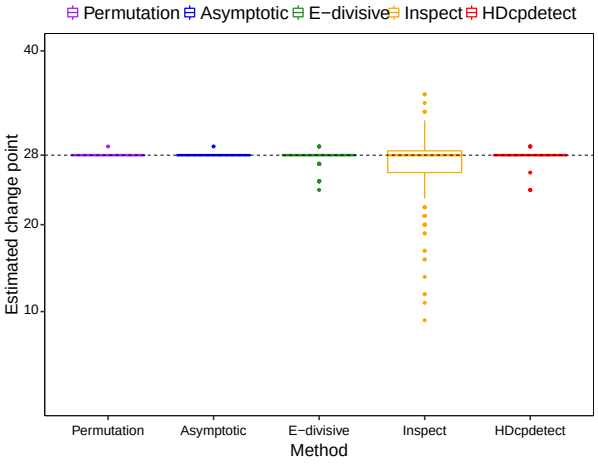
(b) $p = 1000$ and $\boldsymbol{\mu}_2 = (0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$



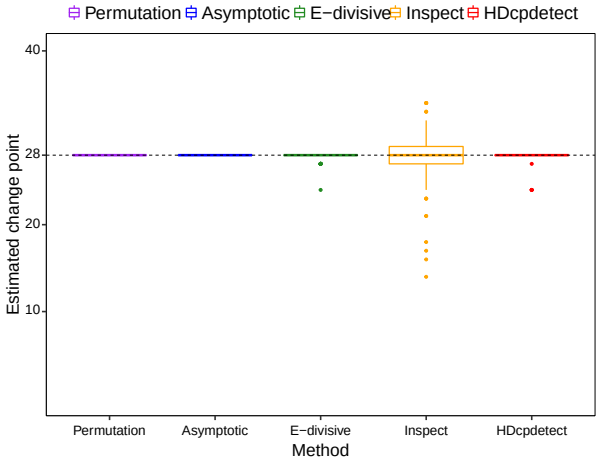
(c) $p = 1500$ and $\boldsymbol{\mu}_2 = (0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$



(d) $p = 500$ and $\boldsymbol{\mu}_2 = (0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$



(e) $p = 1000$ and $\boldsymbol{\mu}_2 = (0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$

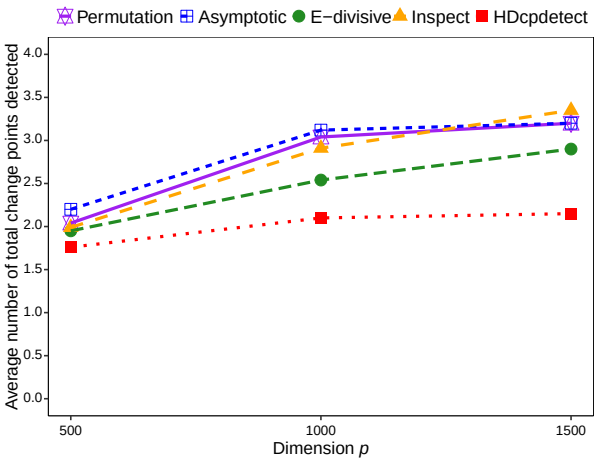


(f) $p = 1500$ and $\boldsymbol{\mu}_2 = (0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$

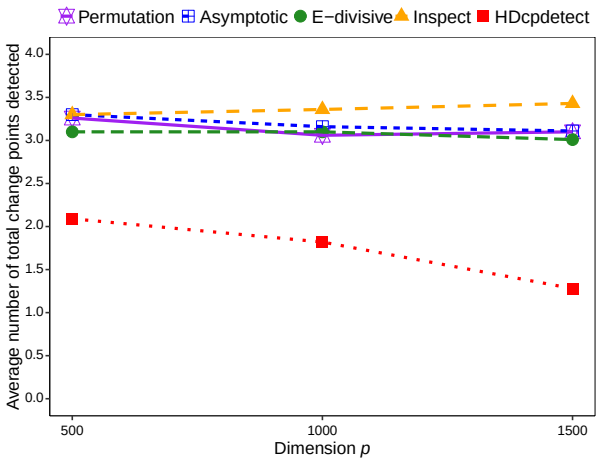
Figure 4: The estimate of change point from each of the methods over 250 replications with a true single change point at location $3n/5 + 1 = 28$ when $n = 45$.

Table 2: Frequency and average number of the correctly detected change points over 250 replications by each of the methods when there are three true change points, also including the case with no change points.

$\boldsymbol{\mu}_2$	n	p	Number of true change points	Method	Frequency of the detected change points				Average number of detected change points
					0	1	2	3	
	90	500	0	Permutation	0.97	0.03	0.00	0.00	0.03
	90	500	0	Asymptotic	0.97	0.03	0.00	0.00	0.03
$\mathbf{0}_p$	90	500	0	E-divisive	0.96	0.04	0.00	0.00	0.04
	90	500	0	Inspect	0.95	0.05	0.00	0.00	0.05
	90	500	0	HDcpdetect	0.96	0.04	0.00	0.00	0.04
	90	1000	0	Permutation	0.97	0.03	0.00	0.00	0.03
	90	1000	0	Asymptotic	0.96	0.04	0.00	0.00	0.04
$\mathbf{0}_p$	90	1000	0	E-divisive	0.95	0.05	0.00	0.00	0.05
	90	1000	0	Inspect	0.96	0.04	0.00	0.00	0.04
	90	1000	0	HDcpdetect	0.95	0.05	0.00	0.00	0.05
	90	1500	0	Permutation	0.96	0.04	0.00	0.00	0.04
	90	1500	0	Asymptotic	0.96	0.04	0.00	0.00	0.04
$\mathbf{0}_p$	90	1500	0	E-divisive	0.96	0.04	0.00	0.00	0.04
	90	1500	0	Inspect	0.94	0.06	0.00	0.00	0.06
	90	1500	0	HDcpdetect	0.94	0.06	0.00	0.00	0.06
	90	500	3	Permutation	0.32	0.46	0.20	0.02	0.92
	90	500	3	Asymptotic	0.24	0.48	0.26	0.02	1.06
$(0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	90	500	3	E-divisive	0.30	0.56	0.12	0.02	0.86
	90	500	3	Inspect	0.56	0.38	0.06	0.00	0.50
	90	500	3	HDcpdetect	0.32	0.62	0.06	0.00	0.74
	90	1000	3	Permutation	0.09	0.24	0.56	0.11	1.69
	90	1000	3	Asymptotic	0.06	0.27	0.50	0.17	1.77
$(0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	90	1000	3	E-divisive	0.07	0.60	0.23	0.10	1.36
	90	1000	3	Inspect	0.46	0.45	0.09	0.00	0.63
	90	1000	3	HDcpdetect	0.20	0.63	0.14	0.03	1.00
	90	1500	3	Permutation	0.00	0.05	0.19	0.76	2.71
	90	1500	3	Asymptotic	0.00	0.05	0.15	0.80	2.75
$(0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	90	1500	3	E-divisive	0.05	0.25	0.46	0.24	1.89
	90	1500	3	Inspect	0.22	0.58	0.19	0.01	0.99
	90	1500	3	HDcpdetect	0.08	0.44	0.38	0.10	1.50
	90	500	3	Permutation	0.00	0.02	0.18	0.80	2.78
	90	500	3	Asymptotic	0.00	0.02	0.16	0.82	2.80
$(0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	90	500	3	E-divisive	0.00	0.02	0.24	0.74	2.72
	90	500	3	Inspect	0.06	0.30	0.36	0.28	1.86
	90	500	3	HDcpdetect	0.01	0.45	0.14	0.40	1.93
	90	1000	3	Permutation	0.00	0.00	0.00	1.00	3.00
	90	1000	3	Asymptotic	0.00	0.00	0.00	1.00	3.00
$(0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	90	1000	3	E-divisive	0.00	0.00	0.14	0.86	2.86
	90	1000	3	Inspect	0.00	0.13	0.57	0.30	2.16
	90	1000	3	HDcpdetect	0.00	0.53	0.19	0.28	1.75
	90	1500	3	Permutation	0.00	0.00	0.00	1.00	3.00
	90	1500	3	Asymptotic	0.00	0.00	0.00	1.00	3.00
$(0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	90	1500	3	E-divisive	0.00	0.00	0.02	0.98	2.98
	90	1500	3	Inspect	0.00	0.08	0.35	0.57	2.49
	90	1500	3	HDcpdetect	0.00	0.85	0.04	0.11	1.26



(a) The cases of three true change points in mean with $\boldsymbol{\mu}_2 = (0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$



(b) The cases of three true change points in mean with $\boldsymbol{\mu}_2 = (0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$

Figure 5: Average number of the total change points detected (correct or incorrect detection) over 250 replications by each of the methods for multiple change point detection.

asymptotic and permutation tests outperforms all the other methods in all scenarios considered. The frequency of the correctly detected change points, reported in the table, shows that the proposed method detects the true change points more accurately compared to the other methods. The performance of our method and the E-divisive by Matteson and James (2014) improves when the dimension p increases, but this is not the case for the Inspect by Wang and Samworth (2018) as it relies on sparsity and tends to improve with the sample size n (see Wang and Samworth, 2018; Hahn et al., 2020; Follain et al., 2022). The average number of correctly detected change points over the 250 replications is much closer to the actual number of true change points for our method in the case of multiple change points too. The results on average number of the total change points detected (correct or incorrect) are reported in Figures 5a and 5b, which suggest that our method does not over-detect or under-detect change points.

5.3. Simulations for a change in variance

We then investigate the performance of the methods for detecting a change in variance of observations while mean remains unchanged. We consider two scenarios for this when the variance of observations is increased by 0.1 and 0.2, that is $\boldsymbol{\Sigma}_1 = 0.5\boldsymbol{V}_p$ and $\boldsymbol{\Sigma}_2 \in \{0.6\boldsymbol{V}_p, 0.7\boldsymbol{V}_p\}$,

Table 3: Frequency and average number of the detected change points over 250 replications by each of the methods when there is a true change in variance of observations.

$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	n	p	Method	Frequency of the detected change points		Average number of detected change points
					0	1	
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	500	Permutation	0.87	0.13	0.13
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	500	Asymptotic	0.85	0.15	0.15
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	500	E-divisive	0.98	0.02	0.02
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	500	Inspect	0.98	0.02	0.02
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	500	HDcpdetect	0.99	0.01	0.01
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	1000	Permutation	0.32	0.68	0.68
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	1000	Asymptotic	0.29	0.71	0.71
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	1000	E-divisive	0.96	0.04	0.04
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	1000	Inspect	0.96	0.04	0.04
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	1000	HDcpdetect	0.98	0.02	0.02
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	1500	Permutation	0.09	0.91	0.91
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	1500	Asymptotic	0.08	0.92	0.92
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	1500	E-divisive	0.96	0.04	0.04
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	1500	Inspect	0.97	0.03	0.03
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.6\boldsymbol{V}_p)$	45	1500	HDcpdetect	0.97	0.03	0.03
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	500	Permutation	0.09	0.91	0.91
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	500	Asymptotic	0.09	0.91	0.91
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	500	E-divisive	0.86	0.14	0.14
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	500	Inspect	0.97	0.03	0.03
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	500	HDcpdetect	0.98	0.02	0.02
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	1000	Permutation	0.00	1.00	1.00
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	1000	Asymptotic	0.00	1.00	1.00
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	1000	E-divisive	0.72	0.28	0.28
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	1000	Inspect	0.94	0.06	0.06
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	1000	HDcpdetect	0.97	0.03	0.03
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	1500	Permutation	0.00	1.00	1.00
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	1500	Asymptotic	0.00	1.00	1.00
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	1500	E-divisive	0.49	0.51	0.51
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	1500	Inspect	0.95	0.05	0.05
$(\mathbf{0}_p, 0.5\boldsymbol{V}_p)$	$(\mathbf{0}_p, 0.7\boldsymbol{V}_p)$	45	1500	HDcpdetect	0.96	0.04	0.04

while the mean of observations is the same, that is $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. The simulation results for all the methods considered are reported in Table 3. From the results, one can see that our proposed method based on both the asymptotic and permutation tests perform well in detecting such a change in variance, but the other methods do not have power for detecting the change point due to the variance of observations.

5.4. Simulations for change in distribution while mean and variance remain unchanged

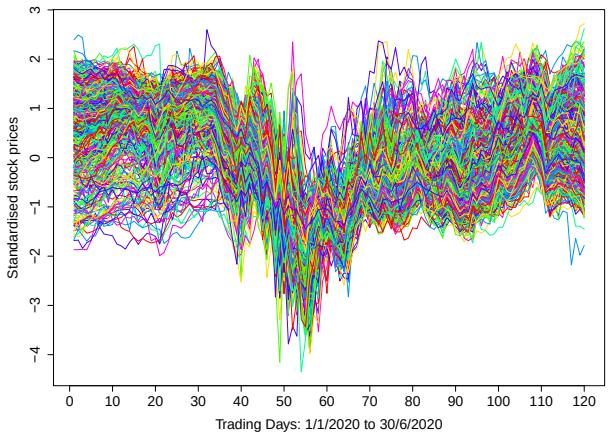
It is a challenging problem for many methods in the literature to detect a change in distribution while the mean and variance of observations remain unchanged. We consider two simulation scenarios for this with $n = 45$ and a change in distribution at location 28 when the variables for observations are generated from $N(0, 5/3)$ and $t(5)$ respectively, both having the same means and variances, as well as from $N(1, 1)$ and Exp(1). Considering the asymptotic limit of the univariate distance function and the modified Euclidean distance, they might not distinguish between distributions when their mean and variance are the same, so we here also include the modified L_1 norm distance to see how it performs in this situation. Thus, we try our method with these three distance functions all using the permutation test for a fair comparison. The simulation results over 250 replications for our methods and the other methods are reported in Table 4. The results show that all the methods perform quite poorly in this case as expected, except our method with the modified L_1 norm distance function which performs reasonably good in this challenging situation. This is because the asymptotic limit of the L_1 norm distance does not simplify to expressions in terms of the mean and variance of observations.

6. Real data application

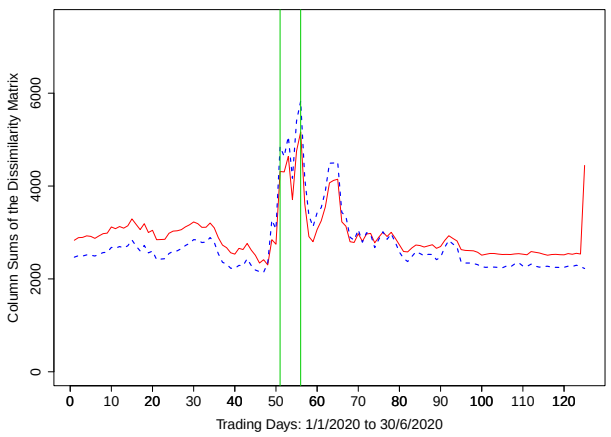
We apply the proposed method to a real data application from the US stock return data. The data set is available at https://www.finance.yahoo.com and can be obtained using the R package `BatchGetSymbols` for different time periods. The data we use here holds the daily closing prices of stocks from the S&P 500 index during the first year of COVID-19 pandemic between 1st January 2020 and 30th June 2020, which results in $n = 125$ time points and $p = 496$ stocks. This specific time period is chosen because based on the experts analysis reported in Statista Research Department (2022), the S&P 500 index showed much volatility

Table 4: Frequency and average number of the detected change points over 250 replications by each of the methods when there is a change in distribution of observations while their mean and variance remain the same.

Two different distributions	n	p	Number of true change points	Method	Frequency of the detected change points		Average number of detected change points
					0	1	
$N(0, 5/3)$ & $t(5)$	45	500	1	Univariate	0.86	0.14	0.14
	45	500	1	L_2 norm	0.84	0.16	0.16
	45	500	1	L_1 norm	0.74	0.26	0.26
	45	500	1	E-divisive	0.55	0.45	0.45
	45	500	1	Inspect	0.99	0.01	0.01
	45	500	1	HDcpdetect	0.98	0.02	0.02
$N(0, 5/3)$ & $t(5)$	45	1000	1	Univariate	0.83	0.17	0.17
	45	1000	1	L_2 norm	0.80	0.20	0.20
	45	1000	1	L_1 norm	0.69	0.31	0.31
	45	1000	1	E-divisive	0.51	0.49	0.49
	45	1000	1	Inspect	0.97	0.03	0.03
	45	1000	1	HDcpdetect	0.96	0.04	0.04
$N(0, 5/3)$ & $t(5)$	45	1500	1	Univariate	0.80	0.20	0.20
	45	1500	1	L_2 norm	0.77	0.23	0.23
	45	1500	1	L_1 norm	0.66	0.34	0.34
	45	1500	1	E-divisive	0.50	0.50	0.50
	45	1500	1	Inspect	0.94	0.06	0.06
	45	1500	1	HDcpdetect	0.94	0.06	0.06
$N(1, 1)$ & Exp(1)	45	500	1	Univariate	0.99	0.01	0.01
	45	500	1	L_2 norm	0.98	0.02	0.02
	45	500	1	L_1 norm	0.58	0.42	0.42
	45	500	1	E-divisive	0.99	0.01	0.01
	45	500	1	Inspect	0.96	0.04	0.04
	45	500	1	HDcpdetect	0.99	0.01	0.01
$N(1, 1)$ & Exp(1)	45	1000	1	Univariate	0.98	0.02	0.02
	45	1000	1	L_2 norm	0.95	0.05	0.05
	45	1000	1	L_1 norm	0.29	0.71	0.71
	45	1000	1	E-divisive	0.98	0.02	0.02
	45	1000	1	Inspect	0.94	0.06	0.06
	45	1000	1	HDcpdetect	0.98	0.02	0.02
$N(1, 1)$ & Exp(1)	45	1500	1	Univariate	0.96	0.04	0.04
	45	1500	1	L_2 norm	0.90	0.10	0.10
	45	1500	1	L_1 norm	0.09	0.91	0.91
	45	1500	1	E-divisive	0.92	0.08	0.08
	45	1500	1	Inspect	0.93	0.07	0.07
	45	1500	1	HDcpdetect	0.97	0.03	0.03



(a) The prices for all the 496 stocks in the S&P 500 data over the time period 1/1/2020 and 30/6/2020. Note that the prices are standardised here for the purpose of visualisation.



(b) Dissimilarity visualisation of S&P 500 data using dissimilarity measure $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ with the modified Euclidean distance function (red solid line) and univariate distance function (blue dashed line).

Figure 6: Change point plots for the S&P 500 data over the time period 1/1/2020 and 30/6/2020 during the COVID-19 pandemic. In plot (b) the two trading days 13th and 16th March 2020 show the largest total dissimilarity from all the other days and are marked by vertical bars (in green).

and dropped by about 20% in early March 2020 entering into a bear market. While the drop was the steepest one-day fall since 1987, S&P 500 index began to recover at the start of April 2020. Stock markets fell in the wake of the COVID-19 pandemic, with investors fearing its spread could destroy economic growth. Figure 6a shows a rough display of the price changes for all the stocks over this time period, where all the stock prices are standardised for visualisation purpose. One can see the very steep drop around early March 2020, as explained. The drop seems to be happened for a majority of stocks with some different magnitudes, suggesting a non-sparse high dimensional change point problem here.

Figure 6b displays the dissimilarity visualisation of the S&P 500 data using the dissimilarity measure $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ with both the modified Euclidean distance function and the univariate distance function, which show a similar trend. Note that the column sums of the two dissimilarity indices are drawn for those trading days in the first half of 2020. The figure suggests that the trading days 13th and 16th March 2020 show the largest total dissimilarity from all the other observations and are marked by vertical bars.

We first implement our distribution-free method, using our R package `HDDchangepoint`,

to find significant change points in this high dimensional data set. Our method with the asymptotic test, using both distance functions, returns six change point locations, namely 51, 57, 66, 95, 106, 112. The method with the permutation test returns four change point locations, namely 51, 57, 66, 95. While the two tests produce four common change points, the permutation test is a bit conservative considering the dimension of data, which is in line with our numerical results. The asymptotic p-values for these four significant change points are 5.64e−09, 5.06e−04, 1.27e−05 and 2.50e−07, respectively. Also, the same four change points are obtained when we use 10 or 15 as the minimum segment length for the binary segmentation. We then apply the E-divisive method of Matteson and James (2014) with the minimum segment length being 15, which detects seven significant change points, namely 15, 37, 48, 68, 80, 95, 106. Also, the random projection method of Wang and Samworth (2018) finds nine change point locations, namely 18, 35, 44, 50, 65, 78, 99, 111, 124. The HDcpdetect method of Li et al. (2019) only finds two change point locations 48, 80. As in our numerical results, HDcpdetect tends to detect fewer change points when there are multiple change points (see Figures 5a and 5b), and the random projection method shows a lower accuracy in such high dimensional data with a small sample size (see Table 2). Considering our simulation results especially those in Table 2 for multiple change points, we believe our estimates of change point locations are more accurate, especially the detected locations 51 and 57 which coincide with the steep drop in the stock prices in early March 2020 due to the COVID-19 impact on the market. Some further results on our analysis of this data set is reported in Appendix C.

7. Concluding remarks

We have proposed a distance-based method for detecting single and multiple change points in non-sparse high dimensional data. The proposed approach is based on new dissimilarity measures and some proposed distance and difference distance matrices, which does not require normality or any other specific distribution for the observations. However, we note that our

asymptotic tests require up to four finite moments of the distribution. Our method is especially useful for change point detection in HDLSS data when the sample size is very small compared to the dimension of data. This is an understudied problem in the literature of high dimensional change points. Our method can handle non-sparse high dimensional situations where changes may happen in many variables and with small significant magnitudes. We have introduced a novel test statistic to formally test the significance of estimated change points and established its asymptotic and permutation distributions to address both small and large sample size situations. We have shown that our proposed estimates of change point locations are consistent for the true unknown change points under some standard conditions and that our proposed tests are consistent asymptotically. Our simulation results showed that both asymptotic and permutation tests perform well compared to some of the recent methods for high dimensional change points. Our R package `HDDchangepoint` for implementation of the proposed method, including both the recursive binary segmentation and the wild binary segmentation as well as the real data application, can be obtained from "`https://github.com/statistics11/HDDchangepoint`". The R package returns significant change point estimates and their corresponding p-values, and it can also be applied with any other dissimilarity measure specified by the user.

References

Barigozzi, M., Cho, H. and Fryzlewicz, P. (2018), 'Simultaneous multiple change-point and factor analysis for high-dimensional time series', *Journal of Econometrics* **206**, 187–225.

Chen, H. and Zhang, N. (2015), 'Graph-based change-point detection', *The Annals of Statistics* **43**, 139–176.

Chu, L. and Chen, H. (2019), 'Asymptotic distribution-free change-point detection for multivariate and non-euclidean data', *The Annals of Statistics* **47**, 382–414.

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

Follain, B., Wang, T. and Samworth, R. J. (2022), 'High-dimensional changepoint estimation with heterogeneous missingness', *Journal of the Royal Statistical Society Series B* **84**, 1023–1055.

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

Garreau, D. and Arlot, S. (2018), 'Consistent change-point detection with kernels', *Electronic Journal of Statistics* **12**, 4440–4486.

Grundy, T., Killick, R. and Mihaylov, G. (2020), 'High-dimensional changepoint detection via a geometrically inspired mapping', *Statistics and Computing* **30**, 1155–1166.

Hahn, G., Fearnhead, P. and Eckley, I. A. (2020), 'BayesProject: Fast computation of a projection direction for multivariate changepoint detection', *Statistics and Computing* **30**, 1691–1705.

Hall, P., Marron, J. S. and Neeman, A. (2005), 'Geometric representation of high dimension, low sample size data', *Journal of the Royal Statistical Society: Series B* **67**, 427–444.

Jung, S. and Marron, J. (2009), 'PCA consistency in high dimension, low sample size context', *The Annals of Statistics* **37**, 4104–4130.

Lee, S., Seo, M. H. and Shin, Y. (2016), 'The lasso for high dimensional regression with a possible change point', *Journal of the Royal Statistical Society: Series B* **78**, 193–210.

Li, J. (2020), 'Asymptotic distribution-free change-point detection based on interpoint distances for high-dimensional data', *Journal of Nonparametric Statistics* **32**, 157–184.

Li, J., Xu, M., Zhong, P.-S. and Li, L. (2019), 'Change point detection in the mean of high-dimensional time series data under dependence'. arXiv:1903.07006.

Liu, B., Zhang, X. and Liu, Y. (2022), 'High dimensional change point inference: Recent developments and extensions', *Journal of Multivariate Analysis* **188**, 1–19.

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

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

Safikhani, A. and Shojaie, A. (2022), 'Joint structural break detection and parameter estimation in high-dimensional nonstationary VAR models', *Journal of the American Statistical Association* **117**, 251–264.

Statista Research Department (2022), 'Weekly development of the S&P 500 index from January 2020 to September 2022', `https://www.statista.com/statistics/1104270/weekly-sandp-500-index-performance/`.

Utev, S. A. (1990), 'Central limit theorem for dependent random variables', *Probability Theory and Mathematical Statistics* **2**, 519–528.

Wang, T. and Samworth, R. J. (2018), 'High dimensional change point estimation via sparse projection', *Journal of the Royal Statistical Society: Series B* **80**, 57–83.

Xiao, W., Huang, X., He, F., Silva, J., Emrani, S. and Chaudhuri, A. (2019), 'Online robust principal component analysis with change point detection', *IEEE Transactions on Multimedia* **22**, 59–68.

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* **83**, 247–270.

Supplementary materials for "A distribution-free method for change point detection in non-sparse high dimensional data"

Appendix A: Technical proofs

Appendix A.1: Proof of Theorem 1

Considering the definition of $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$, we can easily observe $d(\boldsymbol{X}_i, \boldsymbol{X}_j) \geq 0$, $d(\boldsymbol{X}_i, \boldsymbol{X}_i) = 0$ and $d(\boldsymbol{X}_i, \boldsymbol{X}_j) = d(\boldsymbol{X}_j, \boldsymbol{X}_i)$. To prove the triangle inequality $d(\boldsymbol{X}_i, \boldsymbol{X}_j) \leq d(\boldsymbol{X}_i, \boldsymbol{X}_l) + d(\boldsymbol{X}_j, \boldsymbol{X}_l)$, we first note that if $n = 3$,

$$\begin{aligned} d(\boldsymbol{X}_i, \boldsymbol{X}_j) &= \big|\|\boldsymbol{X}_i - \boldsymbol{X}_l\|_D - \|\boldsymbol{X}_j - \boldsymbol{X}_l\|_D\big| \\ &= \big|\|\boldsymbol{X}_i - \boldsymbol{X}_l\|_D - \|\boldsymbol{X}_j - \boldsymbol{X}_l\|_D + \|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D - \|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D\big| \\ &= \big|(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D - \|\boldsymbol{X}_j - \boldsymbol{X}_l\|_D) - (\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D - \|\boldsymbol{X}_i - \boldsymbol{X}_l\|_D)\big| \\ &= \big|(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D - \|\boldsymbol{X}_l - \boldsymbol{X}_j\|_D) - (\|\boldsymbol{X}_j - \boldsymbol{X}_i\|_D - \|\boldsymbol{X}_l - \boldsymbol{X}_i\|_D)\big| \\ &\leq \big|\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D - \|\boldsymbol{X}_l - \boldsymbol{X}_j\|_D\big| + \big|\|\boldsymbol{X}_j - \boldsymbol{X}_i\|_D - \|\boldsymbol{X}_l - \boldsymbol{X}_i\|_D\big| \\ &\leq d(\boldsymbol{X}_i, \boldsymbol{X}_l) + d(\boldsymbol{X}_j, \boldsymbol{X}_l). \end{aligned} \tag{A.1}$$

So $d(\boldsymbol{X}_i, \boldsymbol{X}_j) \leq d(\boldsymbol{X}_i, \boldsymbol{X}_l) + d(\boldsymbol{X}_j, \boldsymbol{X}_l)$ holds for $n = 3$. For $n \geq 4$ and any $\boldsymbol{X}_k$ with $k \geq 4$,

$$\begin{aligned} \big|\|\boldsymbol{X}_i - \boldsymbol{X}_k\|_D - \|\boldsymbol{X}_j - \boldsymbol{X}_k\|_D\big| &= \big|\|\boldsymbol{X}_i - \boldsymbol{X}_k\|_D + \|\boldsymbol{X}_k - \boldsymbol{X}_l\|_D - \|\boldsymbol{X}_k - \boldsymbol{X}_l\|_D - \|\boldsymbol{X}_j - \boldsymbol{X}_k\|_D\big| \\ &\leq \big|\|\boldsymbol{X}_i - \boldsymbol{X}_k\|_D - \|\boldsymbol{X}_l - \boldsymbol{X}_k\|_D\big| + \big|\|\boldsymbol{X}_j - \boldsymbol{X}_k\|_D - \|\boldsymbol{X}_l - \boldsymbol{X}_k\|_D\big|. \end{aligned} \tag{A.2}$$

Using (A.1) and (A.2), we obtain

$$\begin{aligned} \sum_{k \neq i,j} \big|\|\boldsymbol{X}_i - \boldsymbol{X}_k\|_D - \|\boldsymbol{X}_j - \boldsymbol{X}_k\|_D\big| \leq &\sum_{k \neq i,l} \big|\|\boldsymbol{X}_i - \boldsymbol{X}_k\|_D - \|\boldsymbol{X}_l - \boldsymbol{X}_k\|_D\big| \\ &+ \sum_{k \neq j,l} \big|\|\boldsymbol{X}_j - \boldsymbol{X}_k\|_D - \|\boldsymbol{X}_l - \boldsymbol{X}_k\|_D\big|. \end{aligned}$$

Hence, $d(\boldsymbol{X}_i, \boldsymbol{X}_j) \leq d(\boldsymbol{X}_i, \boldsymbol{X}_l) + d(\boldsymbol{X}_j, \boldsymbol{X}_l)$ for $n \geq 3$.

Appendix A.2: Proof of Theorem 2

Using the triangle inequality for $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ proved in Theorem 1, we find

$$\begin{aligned}
d(\boldsymbol{X}_i, \boldsymbol{X}_j) &= \frac{1}{n-2} \sum_{l \neq i,j} \big| \|\boldsymbol{X}_i - \boldsymbol{X}_l\|_D - \|\boldsymbol{X}_j - \boldsymbol{X}_l\|_D \big| \\
&\leq \frac{1}{n-2} \sum_{l \neq i,j} \big| \|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D + \|\boldsymbol{X}_j - \boldsymbol{X}_l\|_D - \|\boldsymbol{X}_j - \boldsymbol{X}_l\|_D \big| \\
&= \frac{1}{n-2} \sum_{l \neq i,j} \|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D \\
&= \|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D.
\end{aligned}$$

Hence, the theorem immediately follows if the multivariate modified Euclidean distance function is used for distance function $\| \cdot - \cdot \|_D$. If the univariate distance function based on differences between the sample mean and variance of the observations is used, we can write

$$\begin{aligned}
\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2^2 &= \sum_{k=1}^{p} \big(\boldsymbol{X}_{ik} - \boldsymbol{X}_{jk}\big)^2 \\
&= \sum_{k=1}^{p} \big(\boldsymbol{X}_{ik} - \bar{\boldsymbol{X}}_i^p + \bar{\boldsymbol{X}}_i^p - \boldsymbol{X}_{jk} - \bar{\boldsymbol{X}}_j^p + \bar{\boldsymbol{X}}_j^p\big)^2 \\
&= \sum_{k=1}^{p} \Big((\boldsymbol{X}_{ik} - \bar{\boldsymbol{X}}_i^p) - (\boldsymbol{X}_{jk} - \bar{\boldsymbol{X}}_j^p) + (\bar{\boldsymbol{X}}_i^p - \bar{\boldsymbol{X}}_j^p)\Big)^2 \\
&= p(S_{\boldsymbol{X}_i}^p)^2 + p(S_{\boldsymbol{X}_j}^p)^2 + p(\bar{\boldsymbol{X}}_i^p - \bar{\boldsymbol{X}}_j^p)^2 - 2\sum_{k=1}^{p} (\boldsymbol{X}_{ik} - \bar{\boldsymbol{X}}_i^p)(\boldsymbol{X}_{jk} - \bar{\boldsymbol{X}}_j^p) \\
&= p(S_{\boldsymbol{X}_i}^p - S_{\boldsymbol{X}_j}^p)^2 + p(\bar{\boldsymbol{X}}_i^p - \bar{\boldsymbol{X}}_j^p)^2 - 2\sum_{k=1}^{p} (\boldsymbol{X}_{ik} - \bar{\boldsymbol{X}}_i^p)(\boldsymbol{X}_{jk} - \bar{\boldsymbol{X}}_j^p) + 2pS_{\boldsymbol{X}_i}^p S_{\boldsymbol{X}_j}^p \\
&\geq p(S_{\boldsymbol{X}_i}^p - S_{\boldsymbol{X}_j}^p)^2 + p(\bar{\boldsymbol{X}}_i^p - \bar{\boldsymbol{X}}_j^p)^2 \\
&= p\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2,
\end{aligned}$$

where the inequality is obtained using $\sum_{k=1}^{p} (\boldsymbol{X}_{ik} - \bar{\boldsymbol{X}}_i^p)(\boldsymbol{X}_{jk} - \bar{\boldsymbol{X}}_j^p) \leq pS_{\boldsymbol{X}_i}^p S_{\boldsymbol{X}_j}^p$ which holds by the Cauchy–Schwarz inequality. Hence, $d(\boldsymbol{X}_i, \boldsymbol{X}_j) \leq \frac{1}{\sqrt{p}} \|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2$.

Appendix A.3: Proof of Theorem 3

First, we give the proof for the multivariate distance function based on the modified Euclidean distance. For this, using the weak law of large numbers and under Assumptions (A1)-(A3), it

is straightforward to show (see also Hall et al., 2005)

$$
\begin{aligned}
&p^{-1/2}\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2 \xrightarrow{P} \sqrt{2\lambda^2_{\nabla_\tau^-}} && \forall\, i \in \nabla_\tau^-,\, j \in \nabla_\tau^- \\
&p^{-1/2}\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2 \xrightarrow{P} \sqrt{2\lambda^2_{\nabla_\tau^+}} && \forall\, i \in \nabla_\tau^+,\, j \in \nabla_\tau^+ \\
&p^{-1/2}\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2 \xrightarrow{P} \sqrt{\lambda^2_{\nabla_\tau^-} + \lambda^2_{\nabla_\tau^+} + \eta^2_{\nabla_\tau^-\nabla_\tau^+}} && \forall\, i \in \nabla_\tau^-,\, j \in \nabla_\tau^+ \\
&p^{-1/2}\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2 \xrightarrow{P} \sqrt{\lambda^2_{\nabla_\tau^-} + \lambda^2_{\nabla_\tau^+} + \eta^2_{\nabla_\tau^-\nabla_\tau^+}} && \forall\, i \in \nabla_\tau^+,\, j \in \nabla_\tau^-.
\end{aligned}
\tag{A.3}
$$

Recall $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D = p^{-1/2}\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2$ in this case. We then write

$$
\begin{aligned}
d(\boldsymbol{X}_i, \boldsymbol{X}_j) &= \frac{1}{n-2}\sum_{l\neq i,j} \big|\|\boldsymbol{X}_i - \boldsymbol{X}_l\|_D - \|\boldsymbol{X}_j - \boldsymbol{X}_l\|_D\big| \\
&= \frac{1}{n-2}\Bigg\{ \sum_{l\in\nabla_\tau^- - \{i,j\}} \big|\|\boldsymbol{X}_i - \boldsymbol{X}_l\|_D - \|\boldsymbol{X}_j - \boldsymbol{X}_l\|_D\big| \\
&\qquad + \sum_{l\in\nabla_\tau^+ - \{i,j\}} \big|\|\boldsymbol{X}_i - \boldsymbol{X}_l\|_D - \|\boldsymbol{X}_j - \boldsymbol{X}_l\|_D\big| \Bigg\}.
\end{aligned}
\tag{A.4}
$$

Inserting (A.3) into (A.4) leads to the result stated in the theorem for $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ for each i and j belonging to ∇_τ^- or ∇_τ^+ accordingly.

Next, we provide the proof for the univariate distance function based on differences between the sample mean and variance of the observations. For this, using the weak law of large numbers and under Assumptions (B1)-(B4), we have $\bar{\boldsymbol{X}}_i^p - \bar{\mu}_{\nabla_\tau^-} \xrightarrow{P} 0\ \forall\, i \in \nabla_\tau^-$ and $\bar{\boldsymbol{X}}_i^p - \bar{\mu}_{\nabla_\tau^+} \xrightarrow{P} 0\ \forall\, i \in \nabla_\tau^+$ as $p \to \infty$. Similarly, $S^p_{\boldsymbol{X}_i} - \sigma_{\nabla_\tau^-} \xrightarrow{P} 0\ \forall\, i \in \nabla_\tau^-$ and $S^p_{\boldsymbol{X}_i} - \sigma_{\nabla_\tau^+} \xrightarrow{P} 0\ \forall\, i \in \nabla_\tau^+$ as $p \to \infty$. Hence, recalling $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D = \sqrt{(\bar{\boldsymbol{X}}_i^p - \bar{\boldsymbol{X}}_j^p)^2 + (S^p_{\boldsymbol{X}_i} - S^p_{\boldsymbol{X}_j})^2}$ in this case, we get

$$
\begin{aligned}
&\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D \xrightarrow{P} 0 && \forall\, i \in \nabla_\tau^-,\, j \in \nabla_\tau^- \\
&\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D \xrightarrow{P} 0 && \forall\, i \in \nabla_\tau^+,\, j \in \nabla_\tau^+ \\
&\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D - \big|\sqrt{(\bar{\mu}_{\nabla_\tau^-} - \bar{\mu}_{\nabla_\tau^+})^2 + (\sigma_{\nabla_\tau^-} - \sigma_{\nabla_\tau^+})^2}\big| \xrightarrow{P} 0 && \forall\, i \in \nabla_\tau^-,\, j \in \nabla_\tau^+ \\
&\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D - \big|\sqrt{(\bar{\mu}_{\nabla_\tau^-} - \bar{\mu}_{\nabla_\tau^+})^2 + (\sigma_{\nabla_\tau^-} - \sigma_{\nabla_\tau^+})^2}\big| \xrightarrow{P} 0 && \forall\, i \in \nabla_\tau^+,\, j \in \nabla_\tau^-.
\end{aligned}
$$

Again inserting the above into (A.4) will give the result stated in the theorem for $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$ for each i and j belonging to ∇_τ^- or ∇_τ^+ accordingly.

Appendix A.4: Proof of Theorem 4

(a) Since τ_0 is the true change point, using the result of Theorem 3 it is straightforward to find that as $p \to \infty$

$$\Delta_{ij} = o_P(1) \qquad \forall\, i = 1, \ldots, n,\ j \neq \tau_0$$

$$\Delta_{ij} = \kappa_{\tau_0} + o_P(1) \qquad \forall\, i = 1, \ldots, n,\ j = \tau_0.$$

Therefore, if $j \neq \tau_0$, we have $\frac{1}{n}\sum_{i=1}^{n} \Delta_{ij} = o_P(1)$. If $j = \tau_0$, we have as $p \to \infty$

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} \Delta_{ij} &= \frac{1}{n}\sum_{i=1}^{n} \Delta_{i,\tau_0} \\
&= \frac{1}{n}\sum_{i \in \nabla_{\tau_0}^{-}} \Delta_{i,\tau_0} + \frac{1}{n}\sum_{i \in \nabla_{\tau_0}^{+}} \Delta_{i,\tau_0} \\
&= \frac{1}{n}|\nabla_{\tau_0}^{-}|\kappa_{\tau_0} + o_P(1) + \frac{1}{n}|\nabla_{\tau_0}^{+}|\kappa_{\tau_0} + o_P(1) \\
&= \kappa_{\tau_0} + o_P(1),
\end{aligned}$$

where in the last step we used the fact that $|\nabla_{\tau_0}^{-}| + |\nabla_{\tau_0}^{+}| = n$.

(b) Using the result of part (a) and the fact that arg max is a continuous function, we find that as $p \to \infty$

$$\begin{aligned}
\hat{\tau} &= \underset{1 \leq j \leq n}{\arg\max} \Big\{ \frac{1}{n}\sum_{i=1}^{n} \Delta_{ij} \Big\} \\
&= \underset{1 \leq j \leq n}{\arg\max} \Big\{ \frac{1}{n}\sum_{i=1}^{n} \Delta_{ij} I(j \neq \tau_0) + \frac{1}{n}\sum_{i=1}^{n} \Delta_{ij} I(j = \tau_0) \Big\} \\
&= \underset{1 \leq j \leq n}{\arg\max} \Big\{ \frac{1}{n}\sum_{i=1}^{n} o_P(1) I(j \neq \tau_0) + \frac{1}{n}\sum_{i=1}^{n} (\kappa_{\tau_0} + o_P(1)) I(j = \tau_0) \Big\} \\
&\xrightarrow{P} \underset{1 \leq j \leq n}{\arg\max} \Big\{ \frac{1}{n}\sum_{i=1}^{n} \kappa_{\tau_0} I(j = \tau_0) \Big\} \\
&= \underset{1 \leq j \leq n}{\arg\max} \big\{ \kappa_{\tau_0} I(j = \tau_0) \big\} \\
&= \tau_0,
\end{aligned}$$

and hence $\hat{\tau} - \tau_0 = o_P(1)$, where $o_P(1)$ follows from the third equality above. This completes the proof of theorem.

Appendix A.5: Proof of Theorem 5

First, using the result of Theorem 4, we have for any fixed n

$$\frac{1}{n}\sum_{i=1}^{n}\Delta_{ij} - I(j=\tau_0)\kappa_{\tau_0} = o_P(1) \quad \text{as } p\to\infty.$$

Furthermore, when $p\to\infty$ and with any fixed n, Δ_{ij} has a degenerate distribution at κ_{τ_0} for all $i=1,\ldots,n$ as

$$P(\Delta_{ij}=\kappa_{\tau_0}) \to \begin{cases} 1 & \text{if } j=\tau_0 \\ 0 & \text{if } j\neq\tau_0. \end{cases}$$

Thus, $E(\Delta_{ij})\to I(j=\tau_0)\kappa_{\tau_0}<\infty$ and $E(\Delta_{ij}^2)\to I(j=\tau_0)\kappa_{\tau_0}^2<\infty$ as $p\to\infty$. Then, using the weak law of large numbers when $n\to\infty$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\Delta_{ij} - \frac{1}{n}\sum_{i=1}^{n}E(\Delta_{ij}) \xrightarrow{P} 0 \quad \text{as } n\to\infty. \tag{A.5}$$

Also, for any fixed n, we have as $p\to\infty$

$$\frac{1}{n}\sum_{i=1}^{n}E(\Delta_{ij}) \to \frac{1}{n}\sum_{i=1}^{n}I(j=\tau_0)\kappa_{\tau_0} = I(j=\tau_0)\kappa_{\tau_0}. \tag{A.6}$$

From (A.5) and (A.6), we get

$$\frac{1}{n}\sum_{i=1}^{n}\Delta_{ij} - I(j=\tau_0)\kappa_{\tau_0} = o_P(1) \quad \text{as } n,p\to\infty.$$

Hence, we find as $n\to\infty$ and $p\to\infty$

$$\max_{1\le j\le n}\left\{\frac{1}{n}\sum_{i=1}^{n}\Delta_{ij} - I(j=\tau_0)\kappa_{\tau_0}\right\} = \max_{1\le j\le n}\left\{\left\{I(j=\tau_0)\kappa_{\tau_0}+o_P(1)\right\} - I(j=\tau_0)\kappa_{\tau_0}\right\} = o_P(1),$$

which completes the proof.

Appendix A.6: Proof of Theorem 6

Using Theorem 3, under the null hypothesis of no change points, we have as $p\to\infty$

$$d_{ij} = d(\boldsymbol{X}_i,\boldsymbol{X}_j) = o_P(1) \qquad \forall\, i,j=1,\ldots,n.$$

This implies $T(\hat{\tau}) = o_P(1)$ under the null hypothesis H_0.

Using Theorem 3, under the alternative hypothesis H_1^s, we get as $p \to \infty$

$$\begin{aligned}
d_{ij} &= o_P(1) && \forall\, i \in \nabla_{\hat{\tau}}^-,\, j \in \nabla_{\hat{\tau}}^- \\
d_{ij} &= o_P(1) && \forall\, i \in \nabla_{\hat{\tau}}^+,\, j \in \nabla_{\hat{\tau}}^+ \\
d_{ij} &= \kappa_{\tau_0} + o_P(1) && \forall\, i \in \nabla_{\hat{\tau}}^-,\, j \in \nabla_{\hat{\tau}}^+ \\
d_{ij} &= \kappa_{\tau_0} + o_P(1) && \forall\, i \in \nabla_{\hat{\tau}}^+,\, j \in \nabla_{\hat{\tau}}^-.
\end{aligned}$$

We note that this result initially holds for the two sets $\nabla_{\tau_0}^-$ and $\nabla_{\tau_0}^+$, and consequently for $\nabla_{\hat{\tau}}^-$ and $\nabla_{\hat{\tau}}^+$ since $\hat{\tau}$ is a consistent estimate of τ_0 as shown in Theorem 4.

Using the above result, we find as $p \to \infty$

$$\begin{aligned}
T(\hat{\tau}) &= \frac{1}{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|} \sum_{i=1}^{n} \sum_{j \in \nabla_{\hat{\tau}}^-} \sum_{j' \in \nabla_{\hat{\tau}}^+} \left(d_{ij} - d_{ij'}\right)^2 \\
&= \frac{1}{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|} \sum_{i=1}^{n} \sum_{j \in \nabla_{\hat{\tau}}^-} \sum_{j' \in \nabla_{\hat{\tau}}^+} d_{ij}^2 + \frac{1}{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|} \sum_{i=1}^{n} \sum_{j \in \nabla_{\hat{\tau}}^-} \sum_{j' \in \nabla_{\hat{\tau}}^+} d_{ij'}^2 \\
&\quad - \frac{2}{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|} \sum_{i=1}^{n} \sum_{j \in \nabla_{\hat{\tau}}^-} \sum_{j' \in \nabla_{\hat{\tau}}^+} d_{ij} d_{ij'} \\
&= \frac{1}{n|\nabla_{\hat{\tau}}^-|} \sum_{i=1}^{n} \sum_{j \in \nabla_{\hat{\tau}}^-} d_{ij}^2 + \frac{1}{n|\nabla_{\hat{\tau}}^+|} \sum_{i=1}^{n} \sum_{j' \in \nabla_{\hat{\tau}}^+} d_{ij'}^2 - \frac{2}{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|} \sum_{i=1}^{n} \sum_{j \in \nabla_{\hat{\tau}}^-} \sum_{j' \in \nabla_{\hat{\tau}}^+} d_{ij} d_{ij'} \\
&= \frac{1}{n|\nabla_{\hat{\tau}}^-|} \sum_{i \in \nabla_{\hat{\tau}}^-} \sum_{j \in \nabla_{\hat{\tau}}^-} d_{ij}^2 + \frac{1}{n|\nabla_{\hat{\tau}}^-|} \sum_{i \in \nabla_{\hat{\tau}}^+} \sum_{j \in \nabla_{\hat{\tau}}^-} d_{ij}^2 + \frac{1}{n|\nabla_{\hat{\tau}}^+|} \sum_{i \in \nabla_{\hat{\tau}}^-} \sum_{j' \in \nabla_{\hat{\tau}}^+} d_{ij'}^2 + \frac{1}{n|\nabla_{\hat{\tau}}^+|} \sum_{i \in \nabla_{\hat{\tau}}^+} \sum_{j' \in \nabla_{\hat{\tau}}^+} d_{ij'}^2 \\
&\quad - \frac{2}{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|} \sum_{i=1}^{n} \sum_{j \in \nabla_{\hat{\tau}}^-} \sum_{j' \in \nabla_{\hat{\tau}}^+} d_{ij} d_{ij'} \\
&= o_P(1) + \left\{ \frac{|\nabla_{\tau_0}^+|}{n} \kappa_{\tau_0}^2 + o_P(1) \right\} + \left\{ \frac{|\nabla_{\tau_0}^-|}{n} \kappa_{\tau_0}^2 + o_P(1) \right\} + o_P(1) - o_P(1) \\
&= \frac{|\nabla_{\tau_0}^-| + |\nabla_{\tau_0}^+|}{n} \kappa_{\tau_0}^2 + o_P(1) \\
&= \kappa_{\tau_0}^2 + o_P(1),
\end{aligned}$$

which holds under the alternative hypothesis H_1^s.

Appendix A.7: Proof of Theorem 7

We first prove the asymptotic distribution when using the univariate distance function $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D = \sqrt{(\bar{\boldsymbol{X}}_i^p - \bar{\boldsymbol{X}}_j^p)^2 + (S^p_{\boldsymbol{X}_i} - S^p_{\boldsymbol{X}_j})^2}$ in (3). To obtain the asymptotic distribution of the test statistic $T(\hat{\tau})$ in this case, we know from the result of Theorem 6 that, under H_1^s,

$$T(\hat{\tau}) = \kappa^2_{\tau_0} + o_P(1),$$

and furthermore, under H_0, $T(\hat{\tau}) = o_P(1)$ as $p \to \infty$. Let us define

$$W(\hat{\tau}) := \frac{1}{n|\nabla^-_{\hat{\tau}}||\nabla^+_{\hat{\tau}}|} \sum_{i=1}^{n} \sum_{j \in \nabla^-_{\hat{\tau}}} \sum_{j' \in \nabla^+_{\hat{\tau}}} \left\{ \|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2_D + \|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|^2_D \right\},$$

which can be rewritten as

$$\begin{aligned} W(\hat{\tau}) =& \frac{1}{n|\nabla^-_{\hat{\tau}}||\nabla^+_{\hat{\tau}}|} \sum_{i \in \nabla^-_{\hat{\tau}}} \sum_{j \in \nabla^-_{\hat{\tau}}} \sum_{j' \in \nabla^+_{\hat{\tau}}} \left\{ \|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2_D + \|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|^2_D \right\} \\ &+ \frac{1}{n|\nabla^-_{\hat{\tau}}||\nabla^+_{\hat{\tau}}|} \sum_{i \in \nabla^+_{\hat{\tau}}} \sum_{j \in \nabla^-_{\hat{\tau}}} \sum_{j' \in \nabla^+_{\hat{\tau}}} \left\{ \|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2_D + \|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|^2_D \right\}. \end{aligned}$$

Using the asymptotic results for $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D$ given in the proof of Theorem 3, one can find that as $p \to \infty$

$$W(\hat{\tau}) = \begin{cases} \kappa^2_{\tau_0} + o_P(1) & \text{under } H_1^s \\ o_P(1) & \text{under } H_0, \end{cases}$$

and furthermore

$$\|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2_D + \|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|^2_D = \begin{cases} \kappa^2_{\tau_0} + o_P(1) & i \in \nabla^-_{\hat{\tau}}, j \in \nabla^-_{\hat{\tau}}, j' \in \nabla^+_{\hat{\tau}} \\ \kappa^2_{\tau_0} + o_P(1) & i \in \nabla^+_{\hat{\tau}}, j \in \nabla^-_{\hat{\tau}}, j' \in \nabla^+_{\hat{\tau}}, \end{cases}$$

and

$$\left(d_{ij} - d_{ij'}\right)^2 = \begin{cases} \kappa^2_{\tau_0} + o_P(1) & i \in \nabla^-_{\hat{\tau}}, j \in \nabla^-_{\hat{\tau}}, j' \in \nabla^+_{\hat{\tau}} \\ \kappa^2_{\tau_0} + o_P(1) & i \in \nabla^+_{\hat{\tau}}, j \in \nabla^-_{\hat{\tau}}, j' \in \nabla^+_{\hat{\tau}}. \end{cases}$$

Therefore, the asymptotic distribution of $T(\hat{\tau})$ is the same as the asymptotic distribution of $W(\hat{\tau})$ when both p and n go to infinity. It thus suffices to derive the asymptotic distribution

of $W(\hat{\tau})$ as it is easier to calculate. For this, we first note that using CLT as $p \to \infty$ and under Assumptions (B1)-(B4), we have

$$\begin{aligned} v_{ij}^{-1/2}\big((\bar{\boldsymbol{X}}_i^p - \bar{\boldsymbol{X}}_j^p) - (\bar{\mu}_i - \bar{\mu}_j)\big) &\xrightarrow{D} N(0,1), \\ v_{ij}^{*-1/2}\big((S_{\boldsymbol{X}_i}^p - S_{\boldsymbol{X}_j}^p) - (\sigma_i - \sigma_j)\big) &\xrightarrow{D} N(0,1), \end{aligned} \tag{A.7}$$

in which

$$\begin{aligned} v_{ij} &= \mathrm{Var}(\bar{\boldsymbol{X}}_i^p) + \mathrm{Var}(\bar{\boldsymbol{X}}_j^p) = \frac{1}{p^2}\sum_{l=1}^{p}\sum_{l'=1}^{p}\mathrm{Cov}(\boldsymbol{X}_{il}, \boldsymbol{X}_{il'}) + \frac{1}{p^2}\sum_{l=1}^{p}\sum_{l'=1}^{p}\mathrm{Cov}(\boldsymbol{X}_{jl}, \boldsymbol{X}_{jl'}), \\ v_{ij}^{*} &= \mathrm{Var}(S_{\boldsymbol{X}_i}^p) + \mathrm{Var}(S_{\boldsymbol{X}_j}^p) = \frac{1}{4\sigma_i^2 p^2}\sum_{l=1}^{p}\sum_{l'=1}^{p}\mathrm{Cov}\big((\boldsymbol{X}_{il} - \bar{\boldsymbol{X}}_i^p)^2, (\boldsymbol{X}_{il'} - \bar{\boldsymbol{X}}_i^p)^2\big) \\ &\quad + \frac{1}{4\sigma_j^2 p^2}\sum_{l=1}^{p}\sum_{l'=1}^{p}\mathrm{Cov}\big((\boldsymbol{X}_{jl} - \bar{\boldsymbol{X}}_j^p)^2, (\boldsymbol{X}_{jl'} - \bar{\boldsymbol{X}}_j^p)^2\big). \end{aligned}$$

Note that we used the delta method to obtain the second equation in (A.7) following the asymptotic normality of $(S_{\boldsymbol{X}_i}^p)^2$ and $(S_{\boldsymbol{X}_j}^p)^2$. From (A.7), we get as $p \to \infty$

$$\begin{aligned} v_{ij}^{-1}(\bar{\boldsymbol{X}}_i^p - \bar{\boldsymbol{X}}_j^p)^2 - \chi_1^2(v_{ij}^{-1}(\bar{\mu}_i - \bar{\mu}_j)^2) &\xrightarrow{D} 0, \\ v_{ij}^{*-1}(S_{\boldsymbol{X}_i}^p - S_{\boldsymbol{X}_j}^p)^2 - \chi_1^2(v_{ij}^{*-1}(\sigma_i - \sigma_j)^2) &\xrightarrow{D} 0, \end{aligned} \tag{A.8}$$

where $\chi_a^2(b)$ denotes a noncentral chi-squared variable whose asymptotic distribution is non-central chi-squared with the degrees of freedom a and the noncentrality parameter b. Since in this case $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2 = (\bar{\boldsymbol{X}}_i^p - \bar{\boldsymbol{X}}_j^p)^2 + (S_{\boldsymbol{X}_i}^p - S_{\boldsymbol{X}_j}^p)^2$, using (A.8) we obtain the expectation and variance of $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2$ when $p \to \infty$ as follows

$$\begin{aligned} E\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2\big) &= \big[(v_{ij} + v_{ij}^{*}) + (\bar{\mu}_i - \bar{\mu}_j)^2 + (\sigma_i - \sigma_j)^2\big] + o^p(1) \\ \mathrm{Var}\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2\big) &= \big[2(v_{ij}^2 + v_{ij}^{*2}) + 4v_{ij}(\bar{\mu}_i - \bar{\mu}_j)^2 + 4v_{ij}^{*}(\sigma_i - \sigma_j)^2\big] + o^p(1). \end{aligned} \tag{A.9}$$

Furthermore, it is straightforward to see that

$$\mathrm{Cov}\big(\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_D^2,\|\boldsymbol{X}_k-\boldsymbol{X}_l\|_D^2\big)=$$

$$\begin{cases}
\mathrm{Var}\big(\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_D^2\big) & \text{if } i\neq j,k\neq l,i=k,j=l,\\
\mathrm{Cov}\big((\bar{\boldsymbol{X}}_i^p-\bar{\boldsymbol{X}}_j^p)^2,(\bar{\boldsymbol{X}}_i^p-\bar{\boldsymbol{X}}_l^p)^2\big)+\mathrm{Cov}\big((S^p_{\boldsymbol{X}_i}-S^p_{\boldsymbol{X}_j})^2,(S^p_{\boldsymbol{X}_i}-S^p_{\boldsymbol{X}_l})^2\big) & \text{if } i\neq j,k\neq l,i=k,j\neq l,\\
\mathrm{Cov}\big((\bar{\boldsymbol{X}}_i^p-\bar{\boldsymbol{X}}_j^p)^2,(\bar{\boldsymbol{X}}_k^p-\bar{\boldsymbol{X}}_j^p)^2\big)+\mathrm{Cov}\big((S^p_{\boldsymbol{X}_i}-S^p_{\boldsymbol{X}_j})^2,(S^p_{\boldsymbol{X}_k}-S^p_{\boldsymbol{X}_j})^2\big) & \text{if } i\neq j,k\neq l,i\neq k,j=l,\\
\mathrm{Var}\big(\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_D^2\big) & \text{if } i\neq j,k\neq l,i=l,j=k,\\
\mathrm{Cov}\big((\bar{\boldsymbol{X}}_i^p-\bar{\boldsymbol{X}}_j^p)^2,(\bar{\boldsymbol{X}}_k^p-\bar{\boldsymbol{X}}_i^p)^2\big)+\mathrm{Cov}\big((S^p_{\boldsymbol{X}_i}-S^p_{\boldsymbol{X}_j})^2,(S^p_{\boldsymbol{X}_k}-S^p_{\boldsymbol{X}_i})^2\big) & \text{if } i\neq j,k\neq l,i=l,j\neq k,\\
\mathrm{Cov}\big((\bar{\boldsymbol{X}}_i^p-\bar{\boldsymbol{X}}_j^p)^2,(\bar{\boldsymbol{X}}_j^p-\bar{\boldsymbol{X}}_l^p)^2\big)+\mathrm{Cov}\big((S^p_{\boldsymbol{X}_i}-S^p_{\boldsymbol{X}_j})^2,(S^p_{\boldsymbol{X}_j}-S^p_{\boldsymbol{X}_l})^2\big) & \text{if } i\neq j,k\neq l,i\neq l,j=k,\\
0 & \text{otherwise,}
\end{cases}\tag{A.10}$$

To calculate the covariances in (A.10), we know from the theory of normal distribution that if $(Z_1,Z_2)\sim N(\mu_{Z_1},\mu_{Z_2},\sigma^2_{Z_1},\sigma^2_{Z_2},\rho_{Z_1,Z_2})$ then $\mathrm{Cov}(Z_1^2,Z_2^2)=4\mu_{Z_1}\mu_{Z_2}\mathrm{Cov}(Z_1,Z_2)+2\mathrm{Cov}^2(Z_1,Z_2)$, so we get

$$\mathrm{Cov}\big((\bar{\boldsymbol{X}}_i^p-\bar{\boldsymbol{X}}_j^p)^2,(\bar{\boldsymbol{X}}_i^p-\bar{\boldsymbol{X}}_l^p)^2\big)-\big[2(\bar{\mu}_i-\bar{\mu}_j)(\bar{\mu}_i-\bar{\mu}_l)v_{ii}+v_{ii}^2/2\big]\to 0$$

$$\mathrm{Cov}\big((S^p_{\boldsymbol{X}_i}-S^p_{\boldsymbol{X}_j})^2,(S^p_{\boldsymbol{X}_i}-S^p_{\boldsymbol{X}_l})^2\big)-\big[2(\sigma_i-\sigma_j)(\sigma_i-\sigma_l)v^*_{ii}+v^{*2}_{ii}/2\big]\to 0,$$

where we used the fact that $\mathrm{Cov}\big((\bar{\boldsymbol{X}}_i^p-\bar{\boldsymbol{X}}_j^p),(\bar{\boldsymbol{X}}_i^p-\bar{\boldsymbol{X}}_l^p)\big)=\mathrm{Var}(\bar{\boldsymbol{X}}_i^p)=v_{ii}/2$ and $\mathrm{Cov}\big((S^p_{\boldsymbol{X}_i}-S^p_{\boldsymbol{X}_j}),(S^p_{\boldsymbol{X}_i}-S^p_{\boldsymbol{X}_l})\big)=\mathrm{Var}(S^p_{\boldsymbol{X}_i})=v^*_{ii}/2$. We calculate the other covariance terms in (A.10) similarly.

Now, when $n\to\infty$ we find that the asymptotic distribution of $W(\tau_0)$ converges to a normal distribution using CLT and under Assumptions (B1)-(B5), as follows

$$\frac{n|\nabla^-_{\tau_0}||\nabla^+_{\tau_0}|\Big(W(\tau_0)-\frac{1}{n|\nabla^-_{\tau_0}||\nabla^+_{\tau_0}|}\sum_{i=1}^n\sum_{j\in\nabla^-_{\tau_0}}\sum_{j'\in\nabla^+_{\tau_0}}m_{ijj'}\Big)}{\sqrt{\sum_{i=1}^n\sum_{j\in\nabla^-_{\tau_0}}\sum_{j'\in\nabla^+_{\tau_0}}\sum_{k=1}^n\sum_{l\in\nabla^-_{\tau_0}}\sum_{l'\in\nabla^+_{\tau_0}}c_{ijj'kll'}}}\xrightarrow{D}N(0,1),\tag{A.11}$$

where

$$m_{ijj'}=E\big(\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_D^2\big)+E\big(\|\boldsymbol{X}_i-\boldsymbol{X}_{j'}\|_D^2\big)$$

and

$$\begin{aligned} c_{ijj'kll'} &= \mathrm{Cov}\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2 + \|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|_D^2, \|\boldsymbol{X}_k - \boldsymbol{X}_l\|_D^2 + \|\boldsymbol{X}_k - \boldsymbol{X}_{l'}\|_D^2\big) \\ &= \mathrm{Cov}\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2, \|\boldsymbol{X}_k - \boldsymbol{X}_l\|_D^2\big) + \mathrm{Cov}\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2, \|\boldsymbol{X}_k - \boldsymbol{X}_{l'}\|_D^2\big) \\ &\quad + \mathrm{Cov}\big(\|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|_D^2, \|\boldsymbol{X}_k - \boldsymbol{X}_l\|_D^2\big) + \mathrm{Cov}\big(\|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|_D^2, \|\boldsymbol{X}_k - \boldsymbol{X}_{l'}\|_D^2\big). \end{aligned}$$

Note that for this result, in addition to Assumption (A5), we require $E\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2\big) < \infty$ which is directly followed from Assumptions (A1) and (A2).

Letting $p \to \infty$, we can calculate both $m_{ijj'}$ and $c_{ijj'kll'}$ using the results in (A.9) and (A.10) obtained as $p \to \infty$. In particular,

$$\begin{aligned} m_{ijj'} &= E\big(\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_D^2\big) + E\big(\|\boldsymbol{X}_i - \boldsymbol{X}_{j'}\|_D^2\big) \\ &= (v_{ij} + v_{ij}^* + v_{ij'} + v_{ij'}^*) + (\bar{\mu}_i - \bar{\mu}_j)^2 + (\sigma_i - \sigma_j)^2 + (\bar{\mu}_i - \bar{\mu}_{j'})^2 + (\sigma_i - \sigma_{j'})^2 + o^p(1). \end{aligned}$$

The asymptotic normality in (A.11) also holds if we replace τ_0 with $\hat{\tau}$ as it is consistent for τ_0. Therefore, we find as $n \to \infty$ and $p \to \infty$

$$\frac{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|\Big(W(\hat{\tau}) - \frac{1}{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|}\sum_{i=1}^n \sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+} m_{ijj'}\Big)}{\sqrt{\sum_{i=1}^n \sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+}\sum_{k=1}^n \sum_{l\in\nabla_{\hat{\tau}}^-}\sum_{l'\in\nabla_{\hat{\tau}}^+} c_{ijj'kll'}}} \xrightarrow{D} N(0,1).$$

Under the null hypothesis H_0, the asymptotic distribution of $W(\hat{\tau})$ when $n \to \infty$ and $p \to \infty$ simplifies since $\bar{\mu}_i = \bar{\mu}_j$ and $\sigma_i = \sigma_j$ for all i and j under H_0, resulting in the following simpler formulae

$$m_{ijj'} = v_{ij} + v_{ij}^* + v_{ij'} + v_{ij'}^* + o^p(1)$$

and

$$\mathrm{Cov}\big(\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_D^2,\|\boldsymbol{X}_k-\boldsymbol{X}_l\|_D^2\big)=\begin{cases}2(v_{ij}^2+v_{ij}^{*2})+o^p(1) & \text{if } i\neq j,k\neq l,i=k,j=l,\\ (v_{ii}^{*2}+v_{ii}^{*2})/2+o^p(1) & \text{if } i\neq j,k\neq l,i=k,j\neq l,\\ (v_{jj}^{*2}+v_{jj}^{*2})/2+o^p(1) & \text{if } i\neq j,k\neq l,i\neq k,j=l,\\ 2(v_{ij}^2+v_{ij}^{*2})+o^p(1) & \text{if } i\neq j,k\neq l,i=l,j=k,\\ (v_{ii}^{*2}+v_{ii}^{*2})/2+o^p(1) & \text{if } i\neq j,k\neq l,i=l,j\neq k,\\ (v_{jj}^{*2}+v_{jj}^{*2})/2+o^p(1) & \text{if } i\neq j,k\neq l,i\neq l,j=k,\\ o^p(1) & \text{otherwise.}\end{cases}$$

For the multivariate modified Euclidean distance function $\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_D=p^{-1/2}\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_2$, we similarly define

$$W^*(\hat{\tau}):=\frac{1}{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|}\sum_{i=1}^{n}\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+}\big\{p^{-1}\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_2^2+p^{-1}\|\boldsymbol{X}_i-\boldsymbol{X}_{j'}\|_2^2\big\},$$

which has the same asymptotic limit as $T(\hat{\tau})$ when both p and n go to infinity under H_0. Again using CLT as above, we obtain in this case

$$m_{ijj'}=E(p^{-1}\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_2^2)+E(p^{-1}\|\boldsymbol{X}_i-\boldsymbol{X}_{j'}\|_2^2),$$

and

$$\begin{aligned}c_{ijj'kll'}&=\mathrm{Cov}\big(p^{-1}\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_2^2+p^{-1}\|\boldsymbol{X}_i-\boldsymbol{X}_{j'}\|_2^2,p^{-1}\|\boldsymbol{X}_k-\boldsymbol{X}_l\|_2^2+p^{-1}\|\boldsymbol{X}_k-\boldsymbol{X}_{l'}\|_2^2\big)\\&=\mathrm{Cov}\big(p^{-1}\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_2^2,p^{-1}\|\boldsymbol{X}_k-\boldsymbol{X}_l\|_2^2\big)+\mathrm{Cov}\big(p^{-1}\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_2^2,p^{-1}\|\boldsymbol{X}_k-\boldsymbol{X}_{l'}\|_2^2\big)\\&\quad+\mathrm{Cov}\big(p^{-1}\|\boldsymbol{X}_i-\boldsymbol{X}_{j'}\|_2^2,p^{-1}\|\boldsymbol{X}_k-\boldsymbol{X}_l\|_2^2\big)+\mathrm{Cov}\big(p^{-1}\|\boldsymbol{X}_i-\boldsymbol{X}_{j'}\|_2^2,\|\boldsymbol{X}_k-\boldsymbol{X}_{l'}\|_2^2\big).\end{aligned}$$

Appendix A.8: Proof of Theorem 8

As we obtained in Theorem 7, the asymptotic distribution of the test statistic $T(\hat{\tau})$ is a normal distribution in general under H_0 or under H_1^s. So, the proof is straightforward for $P_{H_0}(|T_{\mathrm{sd}}(\hat{\tau})|>Z_{\alpha/2})\to\alpha$ using the standard normal distribution of $T_{\mathrm{sd}}(\hat{\tau})$ under H_0. To prove the second

result, we work with the distribution of $T_{\mathrm{sd}}(\hat{\tau})$ under H_1^s by writing

$$
\begin{aligned}
&P_{H_1^s}(T_{\mathrm{sd}}(\hat{\tau}) > Z_{\alpha/2}) = P_{H_1^s}\Big(\frac{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|\Big(T(\hat{\tau}) - \frac{1}{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|}\sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+} m_{ijj'}\Big)}{\sqrt{\sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+}\sum_{k=1}^n\sum_{l\in\nabla_{\hat{\tau}}^-}\sum_{l'\in\nabla_{\hat{\tau}}^+} c_{ijj'kll'}}} > Z_{\alpha/2}\Big) \\
&= P_{H_1^s}\Big(\frac{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|\Big(T(\hat{\tau}) - \frac{1}{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|}\sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+}(m_{ijj'} + a_{ijj'})\Big)}{\sqrt{\sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+}\sum_{k=1}^n\sum_{l\in\nabla_{\hat{\tau}}^-}\sum_{l'\in\nabla_{\hat{\tau}}^+} c_{ijj'kll'}}} > \\
&\qquad Z_{\alpha/2} - \frac{\sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+} a_{ijj'}}{\sqrt{\sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+}\sum_{k=1}^n\sum_{l\in\nabla_{\hat{\tau}}^-}\sum_{l'\in\nabla_{\hat{\tau}}^+} c_{ijj'kll'}}}\Big) \\
&= P_{H_1^s}\Big(\frac{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|\Big(T(\hat{\tau}) - \frac{1}{n|\nabla_{\hat{\tau}}^-||\nabla_{\hat{\tau}}^+|}\sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+}(m_{ijj'} + a_{ijj'})\Big)}{\sqrt{\sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+}\sum_{k=1}^n\sum_{l\in\nabla_{\hat{\tau}}^-}\sum_{l'\in\nabla_{\hat{\tau}}^+} c^*_{ijj'kll'}}} > \\
&\qquad \frac{Z_{\alpha/2}\sqrt{\sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+}\sum_{k=1}^n\sum_{l\in\nabla_{\hat{\tau}}^-}\sum_{l'\in\nabla_{\hat{\tau}}^+} c_{ijj'kll'}} - \sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+} a_{ijj'}}{\sqrt{\sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+}\sum_{k=1}^n\sum_{l\in\nabla_{\hat{\tau}}^-}\sum_{l'\in\nabla_{\hat{\tau}}^+} c^*_{ijj'kll'}}}\Big) \\
&= P_{H_1^s}\Big(Z > \frac{Z_{\alpha/2}\sqrt{\sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+}\sum_{k=1}^n\sum_{l\in\nabla_{\hat{\tau}}^-}\sum_{l'\in\nabla_{\hat{\tau}}^+} c_{ijj'kll'}} - \sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+} a_{ijj'}}{\sqrt{\sum_{i=1}^n\sum_{j\in\nabla_{\hat{\tau}}^-}\sum_{j'\in\nabla_{\hat{\tau}}^+}\sum_{k=1}^n\sum_{l\in\nabla_{\hat{\tau}}^-}\sum_{l'\in\nabla_{\hat{\tau}}^+} c^*_{ijj'kll'}}}\Big) \\
&\quad + o(1),
\end{aligned}
$$

where $a_{ijj'} = (\bar{\mu}_i - \bar{\mu}_j)^2 + (\sigma_i - \sigma_j)^2 + (\bar{\mu}_i - \bar{\mu}_{j'})^2 + (\sigma_i - \sigma_{j'})^2$ and $c^*_{ijj'kll'} = c_{ijj'kll'} + b_{ijkl} + b_{ijkl'} + b_{ij'kl} + b_{ij'kl'}$ in which

$$
b_{ijkl} = \begin{cases}
4v_{ij}(\bar{\mu}_i - \bar{\mu}_j)^2 + 4v^*_{ij}(\sigma_i - \sigma_j)^2 & \text{if } i \neq j, k \neq l, i = k, j = l, \\
2(\bar{\mu}_i - \bar{\mu}_j)(\bar{\mu}_i - \bar{\mu}_l)v_{ii} + 2(\sigma_i - \sigma_j)(\sigma_i - \sigma_l)v^*_{ii} & \text{if } i \neq j, k \neq l, i = k, j \neq l, \\
2(\bar{\mu}_i - \bar{\mu}_j)(\bar{\mu}_k - \bar{\mu}_j)v_{jj} + 2(\sigma_i - \sigma_j)(\sigma_k - \sigma_j)v^*_{jj} & \text{if } i \neq j, k \neq l, i \neq k, j = l, \\
4v_{ij}(\bar{\mu}_i - \bar{\mu}_j)^2 + 4v^*_{ij}(\sigma_i - \sigma_j)^2 & \text{if } i \neq j, k \neq l, i = l, j = k, \\
2(\bar{\mu}_i - \bar{\mu}_j)(\bar{\mu}_k - \bar{\mu}_i)v_{ii} + 2(\sigma_i - \sigma_j)(\sigma_k - \sigma_i)v^*_{ii} & \text{if } i \neq j, k \neq l, i = l, j \neq k, \\
2(\bar{\mu}_i - \bar{\mu}_j)(\bar{\mu}_j - \bar{\mu}_l)v_{jj} + 2(\sigma_i - \sigma_j)(\sigma_j - \sigma_l)v^*_{jj} & \text{if } i \neq j, k \neq l, i \neq l, j = k, \\
0 & \text{otherwise.}
\end{cases}
$$

Then, since

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j\in\nabla_{\hat\tau}^-}\sum_{j'\in\nabla_{\hat\tau}^+} a_{ijj'} &= \sum_{i=1}^{n}\sum_{j\in\nabla_{\hat\tau}^-}\sum_{j'\in\nabla_{\hat\tau}^+} (\bar\mu_i-\bar\mu_j)^2+(\sigma_i-\sigma_j)^2+(\bar\mu_i-\bar\mu_{j'})^2+(\sigma_i-\sigma_{j'})^2\\
&= \sum_{i\in\nabla_{\hat\tau}^-}\sum_{j\in\nabla_{\hat\tau}^-}\sum_{j'\in\nabla_{\hat\tau}^+} (\bar\mu_i-\bar\mu_j)^2+(\sigma_i-\sigma_j)^2+(\bar\mu_i-\bar\mu_{j'})^2+(\sigma_i-\sigma_{j'})^2\\
&\quad+\sum_{i\in\nabla_{\hat\tau}^+}\sum_{j\in\nabla_{\hat\tau}^-}\sum_{j'\in\nabla_{\hat\tau}^+} (\bar\mu_i-\bar\mu_j)^2+(\sigma_i-\sigma_j)^2+(\bar\mu_i-\bar\mu_{j'})^2+(\sigma_i-\sigma_{j'})^2\\
&= \sum_{i\in\nabla_{\hat\tau}^-}\sum_{j\in\nabla_{\hat\tau}^-}\sum_{j'\in\nabla_{\hat\tau}^+} (\bar\mu_i-\bar\mu_{j'})^2+(\sigma_i-\sigma_{j'})^2\\
&\quad+\sum_{i\in\nabla_{\hat\tau}^+}\sum_{j\in\nabla_{\hat\tau}^-}\sum_{j'\in\nabla_{\hat\tau}^+} (\bar\mu_i-\bar\mu_j)^2+(\sigma_i-\sigma_j)^2\\
&= |\nabla_{\hat\tau}^-|\sum_{i\in\nabla_{\hat\tau}^-}\sum_{j'\in\nabla_{\hat\tau}^+} (\bar\mu_i-\bar\mu_{j'})^2+(\sigma_i-\sigma_{j'})^2\\
&\quad+|\nabla_{\hat\tau}^+|\sum_{i\in\nabla_{\hat\tau}^-}\sum_{j\in\nabla_{\hat\tau}^-} (\bar\mu_i-\bar\mu_j)^2+(\sigma_i-\sigma_j)^2\\
&= |\nabla_{\hat\tau}^-||\nabla_{\hat\tau}^-||\nabla_{\hat\tau}^+|\big((\bar\mu_{\nabla_\tau^-}-\bar\mu_{\nabla_\tau^+})^2+(\sigma_{\nabla_\tau^-}-\sigma_{\nabla_\tau^+})^2\big)\\
&\quad+|\nabla_{\hat\tau}^+||\nabla_{\hat\tau}^-||\nabla_{\hat\tau}^+|\big((\bar\mu_{\nabla_\tau^-}-\bar\mu_{\nabla_\tau^+})^2+(\sigma_{\nabla_\tau^-}-\sigma_{\nabla_\tau^+})^2\big)\\
&= (|\nabla_{\hat\tau}^-|+|\nabla_{\hat\tau}^+|)|\nabla_{\hat\tau}^-||\nabla_{\hat\tau}^+|\kappa_{\tau_0}^2\\
&= n|\nabla_{\hat\tau}^-||\nabla_{\hat\tau}^+|\kappa_{\tau_0}^2
\end{aligned}$$

and $|\nabla_{\hat\tau}^-||\nabla_{\hat\tau}^+| = O(n)$, we find

$$P_{H_1^s}(T_{\mathrm{sd}}(\hat\tau) > Z_{\alpha/2}) = P_{H_1^s}\big(Z > Z_{\alpha/2}\frac{c}{c^*} - \frac{n^2\kappa_{\tau_0}^2}{c^*}\big) + o(1) = 1-\Phi\big(Z_{\alpha/2}\frac{c}{c^*} - \frac{n^2\kappa_{\tau_0}^2}{c^*}\big) + o(1), \tag{A.12}$$

where we used the short notation $c = \sqrt{\sum_{i=1}^n \sum_{j\in\nabla_{\hat\tau}^-}\sum_{j'\in\nabla_{\hat\tau}^+}\sum_{k=1}^n\sum_{l\in\nabla_{\hat\tau}^-}\sum_{l'\in\nabla_{\hat\tau}^+} c_{ijj'kll'}}$ and $c^* = \sqrt{\sum_{i=1}^n \sum_{j\in\nabla_{\hat\tau}^-}\sum_{j'\in\nabla_{\hat\tau}^+}\sum_{k=1}^n\sum_{l\in\nabla_{\hat\tau}^-}\sum_{l'\in\nabla_{\hat\tau}^+} c^*_{ijj'kll'}}$ as stated in the theorem.

It is also clear to obtain similarly

$$P_{H_1^s}(T_{\mathrm{sd}}(\hat\tau) < -Z_{\alpha/2}) = P_{H_1^s}\big(Z < -Z_{\alpha/2}\frac{c}{c^*} - \frac{n^2\kappa_{\tau_0}^2}{c^*}\big) + o(1) = \Phi\big(-Z_{\alpha/2}\frac{c}{c^*} - \frac{n^2\kappa_{\tau_0}^2}{c^*}\big) + o(1). \tag{A.13}$$

From (A.12) and (A.13), we get

$$P_{H_1^s}(|T_{\mathrm{sd}}(\hat\tau)| > Z_{\alpha/2}) = 1-\Phi\big(Z_{\alpha/2}\frac{c}{c^*} - \frac{n^2\kappa_{\tau_0}^2}{c^*}\big) + \Phi\big(-Z_{\alpha/2}\frac{c}{c^*} - \frac{n^2\kappa_{\tau_0}^2}{c^*}\big) + o(1).$$

Appendix A.9: Proof of Corollary 1

The proof of this corollary is straightforward using the result of Theorem 8. See also the proof of next corollary.

Appendix A.10: Proof of Corollary 2

Using the result of Theorem 8, the proof of this corollary is also straightforward and we just prove $\kappa^2_{\tau_0} = \frac{(\sum_{k\in S_0}\omega_k)^2}{p^2(\sigma_{\nabla^-_{\tau_0}}+\sigma_{\nabla^+_{\tau_0}})^2}$ in this case. For this, first recall $\sigma_i^2 = p^{-1}\sum_{k=1}^p \sigma_{ik}^2$, $i = 1, \ldots, n$. Since $\sigma_i^2 - \sigma_j^2 = (\sigma_i - \sigma_j)(\sigma_i + \sigma_j)$, we can write for $i \in \nabla^-_\tau, j \in \nabla^+_\tau$

$$(\sigma_i - \sigma_j)^2 = \frac{(\sigma_i^2 - \sigma_j^2)^2}{(\sigma_i + \sigma_j)^2} = \frac{\left(p^{-1}\sum_{k=1}^p \sigma_{ik}^2 - p^{-1}\sum_{k=1}^p \sigma_{jk}^2\right)^2}{(\sigma_i + \sigma_j)^2} = \frac{p^{-2}(\sum_{k=1}^p \omega_k)^2}{(\sigma_i + \sigma_j)^2}. \tag{A.14}$$

Considering that the mean of observations is unchanged in this case, we get from (A.14) that

$$\kappa^2_{\tau_0} = \frac{p^{-2}(\sum_{k=1}^p \omega_k)^2}{(\sigma_{\nabla^-_{\tau_0}} + \sigma_{\nabla^+_{\tau_0}})^2} = \frac{(\sum_{k\in S_0}\omega_k)^2}{p^2(\sigma_{\nabla^-_{\tau_0}} + \sigma_{\nabla^+_{\tau_0}})^2}.$$

Appendix A.11: Proof of Theorem 9

Let us define the notation

$$\bar{\Delta}_{j,b_k,e_k} := \frac{1}{e_k - b_k + 1}\sum_{i=b_k}^{e_k} \Delta_{ij}, \qquad j = 1, \ldots, n.$$

We can then rewrite the change point estimate (9) as

$$\hat{\gamma}_k = \underset{b_k \le j \le e_k}{\arg\max} \left\{\bar{\Delta}_{j,b_k,e_k}\right\}.$$

For clarity of the proof, first consider the case when there are two true change points τ_1^0, τ_2^0, $2 \le \tau_1^0 < \tau_2^0 \le n$. Similar to the proof of Theorem 4, it is straightforward to show as $p \to \infty$ that

$$\begin{aligned}
\bar{\Delta}_{j,b_k,e_k} &= o_P(1) && j \neq \tau_1^0, \tau_2^0 \\
\bar{\Delta}_{j,b_k,e_k} &= \kappa_{\tau_1^0} + o_P(1) && j = \tau_1^0 \\
\bar{\Delta}_{j,b_k,e_k} &= \kappa_{\tau_2^0} + o_P(1) && j = \tau_2^0,
\end{aligned} \tag{A.15}$$

where $\kappa_{\tau_i^0}$, $i = 1, 2$, are obtained by plugging in τ_i^0 in the formula of κ_τ given in Theorem 3. Note that the above result holds with any b_k and e_k. The binary segmentation starts with $k = 1$, $b_k = 1$ and $e_k = n$, so

$$\hat{\gamma}_1 = \underset{1\le j\le n}{\arg\max} \left\{\bar{\Delta}_{j,1,n}\right\},$$

and using (A.15) we find as $p \to \infty$

$$\begin{aligned}
\hat{\gamma}_1 &= \underset{1\le j\le n}{\arg\max} \left\{\bar{\Delta}_{j,1,n}\right\} \\
&= \underset{1\le j\le n}{\arg\max} \left\{\bar{\Delta}_{j,1,n} I(j \neq \tau_1^0, \tau_2^0) + \bar{\Delta}_{j,1,n} I(j = \tau_1^0) + \bar{\Delta}_{j,1,n} I(j = \tau_2^0)\right\} \\
&= \underset{1\le j\le n}{\arg\max} \left\{o_P(1) I(j \neq \tau_1^0, \tau_2^0) + (\kappa_{\tau_1^0} + o_P(1)) I(j = \tau_1^0) + (\kappa_{\tau_2^0} + o_P(1)) I(j = \tau_2^0)\right\} \\
&\xrightarrow{P} \underset{1\le j\le n}{\arg\max} \left\{\kappa_{\tau_1^0} I(j = \tau_1^0) + \kappa_{\tau_2^0} I(j = \tau_2^0)\right\}.
\end{aligned}$$

Hence, we get either $\hat{\gamma}_1 \xrightarrow{P} \tau_1^0$ or $\hat{\gamma}_1 \xrightarrow{P} \tau_2^0$, depending on which of $\kappa_{\tau_1^0}$ and $\kappa_{\tau_2^0}$ is larger. First, if $\kappa_{\tau_1^0} > \kappa_{\tau_2^0}$, we get $\hat{\gamma}_1 \xrightarrow{P} \tau_1^0$ and then split the data sequence to two sub-sequences before and after $\hat{\gamma}_1$: one with $b_2 = 1$ and $e_2 = \hat{\gamma}_1 - 1$, and one with $b_2 = \hat{\gamma}_1$ and $e_2 = n$. Then, since $\hat{\gamma}_1 \xrightarrow{P} \tau_1^0 < \tau_2^0$, we have either

$$\hat{\gamma}_2 = \underset{1\le j\le \hat{\gamma}_1 - 1}{\arg\max} \left\{\bar{\Delta}_{j,1,\hat{\gamma}_1 - 1}\right\} = \underset{1\le j\le \hat{\gamma}_1 - 1}{\arg\max} \left\{o_P(1)\right\} \xrightarrow{P} \underset{1\le j\le \hat{\gamma}_1 - 1}{\arg\max} \left\{0\right\} = \emptyset \tag{A.16}$$

where the empty set $\emptyset$ means no change point returned as defined in the paper, or

$$\begin{aligned}
\hat{\gamma}_2 &= \underset{\hat{\gamma}_1\le j\le \hat{n}}{\arg\max} \left\{\bar{\Delta}_{j,\hat{\gamma}_1,n}\right\} \\
&= \underset{\hat{\gamma}_1\le j\le \hat{n}}{\arg\max} \left\{\bar{\Delta}_{j,\hat{\gamma}_1,n} I(j \neq \tau_2^0) + \bar{\Delta}_{j,\hat{\gamma}_1,n} I(j = \tau_2^0)\right\} \\
&= \underset{\hat{\gamma}_1\le j\le \hat{n}}{\arg\max} \left\{o_P(1) I(j \neq \tau_2^0) + (\kappa_{\tau_2^0} + o_P(1)) I(j = \tau_2^0)\right\} \\
&\xrightarrow{P} \underset{\hat{\gamma}_1\le j\le n}{\arg\max} \left\{\kappa_{\tau_2^0} I(j = \tau_2^0)\right\} \\
&= \tau_2^0.
\end{aligned}$$

Thus, we must have $\hat{\gamma}_2 \xrightarrow{P} \tau_2^0$. Second, if $\kappa_{\tau_1^0} < \kappa_{\tau_2^0}$, then we get $\hat{\gamma}_1 \xrightarrow{P} \tau_2^0$ and we can similarly as above show that $\hat{\gamma}_2 \xrightarrow{P} \tau_1^0$. If we further split the data sub-sequences before and after the second change point estimate $\hat{\gamma}_2$, no more $\hat{\gamma}_k$ will converge by considering (A.15) and a similar

calculation as in (A.16). So, denoting $(\hat{\tau}_1, \hat{\tau}_2) = \text{sort}(\hat{\gamma}_1, \hat{\gamma}_2)$, we have $\|(\hat{\tau}_1, \hat{\tau}_2) - (\tau_1^0, \tau_2^0)\|_\infty = o_P(1)$ for this case of two true change points.

Now suppose there are s true change points $\tau_1^0, \tau_2^0, \ldots, \tau_s^0$, $2 \leq \tau_1^0 < \tau_2^0 < \ldots < \tau_s^0 \leq n$. Similar to the above, we obtain as $p \to \infty$

$$\begin{aligned}
\bar{\Delta}_{j,b_k,e_k} &= o_P(1) && j \neq \tau_1^0, \tau_2^0, \ldots, \tau_s^0 \\
\bar{\Delta}_{j,b_k,e_k} &= \kappa_{\tau_1^0} + o_P(1) && j = \tau_1^0 \\
\bar{\Delta}_{j,b_k,e_k} &= \kappa_{\tau_2^0} + o_P(1) && j = \tau_2^0 \\
&\vdots \\
\bar{\Delta}_{j,b_k,e_k} &= \kappa_{\tau_s^0} + o_P(1) && j = \tau_s^0.
\end{aligned} \tag{A.17}$$

The binary segmentation starts with $k = 1$, $b_k = 1$ and $e_k = n$, so

$$\hat{\gamma}_1 = \underset{1 \leq j \leq n}{\arg\max} \left\{ \bar{\Delta}_{j,1,n} \right\},$$

and using (A.17) we find as $p \to \infty$

$$\begin{aligned}
\hat{\gamma}_1 &= \underset{1 \leq j \leq n}{\arg\max} \left\{ \bar{\Delta}_{j,1,n} \right\} \\
&= \underset{1 \leq j \leq n}{\arg\max} \left\{ \bar{\Delta}_{j,1,n} I(j \neq \tau_1^0, \tau_2^0, \ldots, \tau_s^0) + \sum_{i=1}^{s} \bar{\Delta}_{j,1,n} I(j = \tau_i^0) \right\} \\
&= \underset{1 \leq j \leq n}{\arg\max} \left\{ o_P(1) I(j \neq \tau_1^0, \tau_2^0, \ldots, \tau_s^0) + \sum_{i=1}^{s} (\kappa_{\tau_i^0} + o_P(1)) I(j = \tau_i^0) \right\} \\
&\xrightarrow{P} \underset{1 \leq j \leq n}{\arg\max} \left\{ \sum_{i=1}^{s} \kappa_{\tau_i^0} I(j = \tau_i^0) \right\}.
\end{aligned}$$

Hence, we have $\hat{\gamma}_1 \xrightarrow{P} \tau_{k_1}^0$ for a $\tau_{k_1}^0$, $k_1 \in \{1, \ldots, s\}$, for which $\kappa_{\tau_{k_1}^0} = \max\{\kappa_{\tau_1^0}, \kappa_{\tau_2^0}, \ldots, \kappa_{\tau_s^0}\}$. Then, we split the data sequence to two sub-sequences before and after $\hat{\gamma}_1$: one with $b_2 = 1$ and $e_2 = \hat{\gamma}_1 - 1$, and one with $b_2 = \hat{\gamma}_1$ and $e_2 = n$. Applying the same process as above for the two true change points case to each of these two sub-sequences, we find for the next change point estimates $\hat{\gamma}_2$ and $\hat{\gamma}_3$ that $\hat{\gamma}_2 \xrightarrow{P} \tau_{k_2}^0$ and $\hat{\gamma}_3 \xrightarrow{P} \tau_{k_3}^0$ where we have either $\tau_{k_2}^0 < \tau_{k_1}^0$ and $\tau_{k_3}^0 > \tau_{k_1}^0$ or $\tau_{k_2}^0 > \tau_{k_1}^0$ and $\tau_{k_3}^0 < \tau_{k_1}^0$. Continuing this process, we find $\|(\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_s) - (\tau_1^0, \tau_2^0, \ldots, \tau_s^0)\|_\infty = o_P(1)$, as $p \to \infty$. We note that for the last two sub-sequences no additional change points will

converge using the same argument as above for the case with two change points according to a similar calculation as in (A.16).

Appendix A.12: Proof of Theorem 10

(a) First, because under the null hypothesis H_0 all the variables $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ have the same distribution, $T_r(\hat{\tau})$ must have the same distribution as $T(\hat{\tau})$ under H_0 for all random permutations, and hence

$$E(G_{T_R}(t)) = \frac{1}{R}\sum_{r=1}^{R} P\big(T_r(\hat{\tau}) \leq t\big) = P\big(T(\hat{\tau}) \leq t\big).$$

This then implies $E(G_{T_R}(t)) \to G_T(t)$ for all t as $n \to \infty$ and $p \to \infty$ under H_0, since the asymptotic distribution of $T(\hat{\tau})$ is $G_T(t)$ using Theorem 7. To show that $G_{T_R}(t) \xrightarrow{D} G_T(t)$, it suffices to prove that $\mathrm{Var}(G_{T_R}(t)) \to 0$ as $n \to \infty$ and $p \to \infty$. This holds because

$$E(G^2_{T_r}(t)) = \frac{1}{R^2}\sum_{r=1}^{R}\sum_{r'=1}^{R} P\big(T_r(\hat{\tau}) \leq t, T_{r'}(\hat{\tau}) \leq t\big) = P\big(T(\hat{\tau}) \leq t, T(\hat{\tau}) \leq t\big) \to G^2_T(t).$$

It should be noted that this holds only under H_0 because the variables would not have the same distribution under the alternative hypothesis.

(b) Using part (a) above, as $n \to \infty$ and $p \to \infty$, we can write for $0 \leq \alpha \leq 1$

$$\begin{aligned} P_{H_0}(p_{\mathrm{perm}} \leq \alpha) &= P_{H_0}(1 - G_{T_R}(T_{\mathrm{obs}}(\hat{\tau})) \leq \alpha) \to P_{H_0}(1 - G_T(T_{\mathrm{obs}}(\hat{\tau})) \leq \alpha) \\ &= P_{H_0}(G_T(T_{\mathrm{obs}}(\hat{\tau})) \geq 1 - \alpha) = P(U \geq 1 - \alpha) = \alpha, \end{aligned}$$

where U denotes a random variable having uniform distribution on the interval (0,1).

Appendix B: Additional simulation results

We here provide some further simulation results on our proposed method and in comparison with the other methods, as well as on the computations of our method using the wild binary segmentation versus the recursive binary segmentation.

Appendix B1: Simulations for change points with non-normal data

We also evaluate the performance of our methods in detecting high dimensional change points when the distribution of observations is not normal. For this, we generate data from the asymmetric chi-squared distribution with one degree of freedom and scale the simulated observations to have the same mean and variance as in the previous cases of normal distribution. Table A.1 presents the results on frequency and average number of the detected change points for this scenario. We can see that our method based on both the asymptotic and permutation tests performs better than the other methods. Unlike our method as well as E-divisive and HDcpdetect which are distribution-free, the performance of the Inspect is more affected compared to the previous results as this method requires the normality of observations.

Appendix B2: Computation time and wild binary segmentation

We here report the computation time of our method for multiple change point detection and also evaluate the use of wild binary segmentation in our approach as described at the end of Section 3. We are interested to see how the wild binary segmentation may improve the results. For this, we consider the same change point problem with three true change points as before, focusing on the cases with $n = 90$, $p \in \{500, 1000, 1500\}$ and $\boldsymbol{\mu}_2 = (0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$. The three true change point locations are thus $\{28, 46, 73\}$. We report the percentage of times the first detected change point is a true change point (this is important for any binary segmentation), the average number of the true change points detected and the average computation time, on a desktop PC (2.50 GHz+64 GB RAM), for our multiple change point algorithm once with the recursive binary segmentation and once with the wild binary segmentation using 1000 random draws, across 250 replications. The results, which are presented in Table A.2, show that the wild binary segmentation slightly improves the results compared to the recursive binary segmentation. We note that our single change point algorithm is consistent so it tends to produce a true change point, thus the two binary segmentation procedures perform similarly well. As the results

Table A.1: Frequency and average number of the detected change points over 250 replications by each of the methods when data are generated from chi-squared distribution with one degree of freedom and when there is one true change point, also including the case with no change point.

$\boldsymbol{\mu}_2$	n	p	Number of true change points	Method	Frequency of the detected change points		Average number of detected change points
					0	1	
	45	500	0	Permutation	0.98	0.02	0.02
	45	500	0	Asymptotic	0.97	0.03	0.03
$\mathbf{0}_p$	45	500	0	E-divisive	0.97	0.03	0.03
	45	500	0	Inspect	0.98	0.02	0.02
	45	500	0	HDcpdetect	0.97	0.03	0.03
	45	1000	0	Permutation	0.97	0.03	0.03
	45	1000	0	Asymptotic	0.97	0.03	0.03
$\mathbf{0}_p$	45	1000	0	E-divisive	0.96	0.04	0.04
	45	1000	0	Inspect	0.97	0.03	0.03
	45	1000	0	HDcpdetect	0.97	0.03	0.03
	45	1500	0	Permutation	0.96	0.04	0.04
	45	1500	0	Asymptotic	0.96	0.04	0.04
$\mathbf{0}_p$	45	1500	0	E-divisive	0.96	0.04	0.04
	45	1500	0	Inspect	0.97	0.03	0.03
	45	1500	0	HDcpdetect	0.95	0.05	0.05
	45	500	1	Permutation	0.85	0.15	0.15
	45	500	1	Asymptotic	0.84	0.16	0.16
$(0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	45	500	1	E-divisive	0.86	0.14	0.14
	45	500	1	Inspect	0.97	0.03	0.03
	45	500	1	HDcpdetect	0.88	0.12	0.12
	45	1000	1	Permutation	0.54	0.46	0.46
	45	1000	1	Asymptotic	0.47	0.53	0.53
$(0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	45	1000	1	E-divisive	0.60	0.40	0.40
	45	1000	1	Inspect	0.96	0.04	0.04
	45	1000	1	HDcpdetect	0.62	0.38	0.38
	45	1500	1	Permutation	0.22	0.78	0.78
	45	1500	1	Asymptotic	0.18	0.82	0.82
$(0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	45	1500	1	E-divisive	0.39	0.61	0.61
	45	1500	1	Inspect	0.94	0.06	0.06
	45	1500	1	HDcpdetect	0.44	0.56	0.56
	45	500	1	Permutation	0.20	0.80	0.80
	45	500	1	Asymptotic	0.17	0.83	0.83
$(0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	45	500	1	E-divisive	0.21	0.79	0.79
	45	500	1	Inspect	0.89	0.11	0.11
	45	500	1	HDcpdetect	0.23	0.77	0.77
	45	1000	1	Permutation	0.04	0.96	0.96
	45	1000	1	Asymptotic	0.04	0.96	0.96
$(0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	45	1000	1	E-divisive	0.06	0.94	0.94
	45	1000	1	Inspect	0.88	0.12	0.12
	45	1000	1	HDcpdetect	0.08	0.92	0.92
	45	1500	1	Permutation	0.00	1.00	1.00
	45	1500	1	Asymptotic	0.00	1.00	1.00
$(0.2 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$	45	1500	1	E-divisive	0.02	0.98	0.98
	45	1500	1	Inspect	0.86	0.14	0.14
	45	1500	1	HDcpdetect	0.04	0.96	0.96

Table A.2: Computation time when there are three true change points and results on using the wild binary segmentation (WBS) versus the recursive binary segmentation (RBS) in our method over 250 replications, when $\boldsymbol{\mu}_2 = (0.1 \times \mathbf{1}_{3p/4}, 0 \times \mathbf{1}_{p/4})$.

Test	n	p	True change point locations	Segmentation method	Percentage of times the first detected change point is a true change point	Average number of detected change points	Computation time (in minute)
Asymptotic	90	500	$\{28, 46, 73\}$	RBS	0.38	1.04	1.52
	90	500	$\{28, 46, 73\}$	WBS	0.43	1.15	3.11
	90	1000	$\{28, 46, 73\}$	RBS	0.60	1.73	1.64
	90	1000	$\{28, 46, 73\}$	WBS	0.72	1.80	5.30
	90	1500	$\{28, 46, 73\}$	RBS	0.91	2.72	1.69
	90	1500	$\{28, 46, 73\}$	WBS	0.91	2.72	7.28
Permutation	90	500	$\{28, 46, 73\}$	RBS	0.38	0.93	26.44
	90	500	$\{28, 46, 73\}$	WBS	0.42	1.03	43.51
	90	1000	$\{28, 46, 73\}$	RBS	0.58	1.68	37.09
	90	1000	$\{28, 46, 73\}$	WBS	0.69	1.74	67.12
	90	1500	$\{28, 46, 73\}$	RBS	0.91	2.70	57.17
	90	1500	$\{28, 46, 73\}$	WBS	0.91	2.71	96.91

indicate, the use of wild binary segmentation requires a fairly more computation time.

Appendix C: Additional results on our data application

For the S&P 500 data, we also calculate confidence intervals for the detected change point locations using our method. For this, we use the confidence interval formula (7) obtained at the end of Section 2 in the paper. Figure A.1 shows the 95% confidence intervals for the detected change point locations.

Finally, a rather naive approach in high dimensional settings which entirely ignores the high dimensional nature of the data is to aggregate the data for example calculate the means of observations and then apply a univariate change point method to the means of observations. To test this ad hoc approach on the S&P 500 data, we compute the means of 125 observations and apply the wild binary segmentation of Fryzlewicz (2014) using CUSUM as a univariate technique to the means of 125 observations. It returns sixteen significant change points 28, 37, 46, 49, 51, 55, 57, 58, 63, 66, 68, 80, 95, 99, 106, 112, whose significance are confirmed using the `changepoints` function as in the `wbs` package. The number of significant change points detected is too large compared to all the other methods considered in the paper, which could

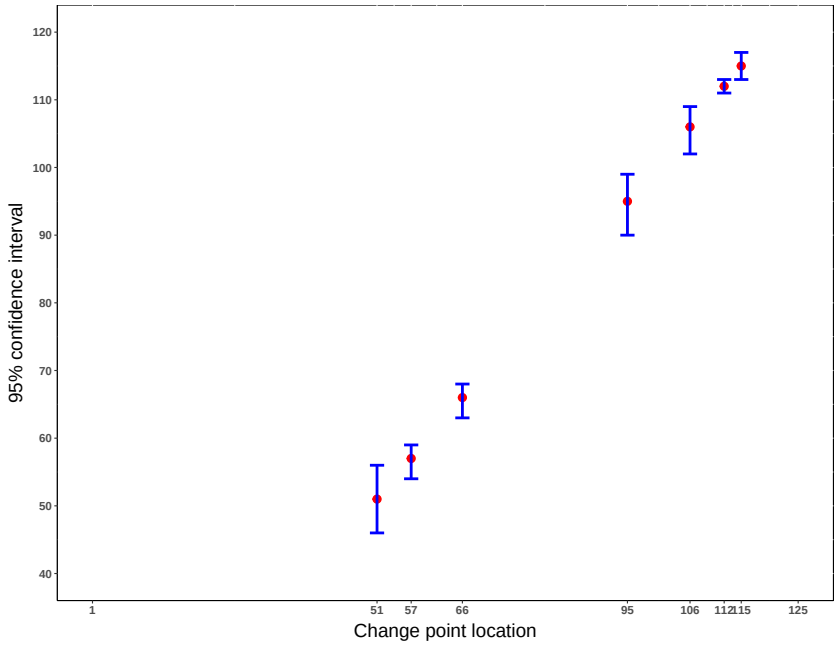


Figure A.1: S&P 500 data: 95% confidence intervals for the detected change points using the proposed method. The estimated change points are also highlighted using red dots.

be misleading especially the detection of several successive change points.

References

Hall, P., Marron, J. S. and Neeman, A. (2005), 'Geometric representation of high dimension, low sample size data', *Journal of the Royal Statistical Society: Series B* **67**, 427–444.