Received: 26 March 2019 | Revised: 21 January 2020 | Accepted: 26 February 2020

DOI: 10.1111/sjos.12460

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Multivariate analysis of variance and change points estimation for high-dimensional longitudinal data

Ping-Shou Zhong[1] | **Jun Li**[2] | **Piotr Kokoszka**[3]

[1]Mathematics, Statistics, and Computer Science, University of Illinois at Chicago

[2]Department of Mathematical Sciences, Kent State University

[3]Department of Statistics, Colorado State University

Correspondence
Ping-Shou Zhong, Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, IL 60607-7045.
Email: pszhong@uic.edu

Abstract

This article considers the problem of testing temporal homogeneity of p-dimensional population mean vectors from repeated measurements on n subjects over T times. To cope with the challenges brought about by high-dimensional longitudinal data, we propose methodology that takes into account not only the "large p, large T, and small n" situation but also the complex temporospatial dependence. We consider both the multivariate analysis of variance problem and the change point problem. The asymptotic distributions of the proposed test statistics are established under mild conditions. In the change point setting, when the null hypothesis of temporal homogeneity is rejected, we further propose a binary segmentation method and show that it is consistent with a rate that explicitly depends on p, T, and n. Simulation studies and an application to fMRI data are provided to demonstrate the performance and applicability of the proposed methods.

KEYWORDS

change points, fMRI data, high-dimensional means, longitudinal data, spatial dependence, temporal dependence

1 INTRODUCTION

High-dimensional longitudinal data are often observed in modern applications such as genomics studies and neuroimaging studies of brain function. Collected by repeatedly measuring a large number of components from a small number of subjects over many time points, the high-dimensional longitudinal data exhibit complex temporospatial dependence: the spatial dependence among the components of each high-dimensional measurement at a particular time point, and the temporal dependence among different high-dimensional measurements collected at different time points. For example, the functional magnetic resonance imaging (fMRI) data are collected by repeatedly measuring the p blood oxygen level-dependent (BOLD) responses from the brains over T times while a small number of subjects are given some task to perform (p, T, and n are typically of the order of 100,000, 100, and 10, respectively). The fMRI data are characterized by the spatial dependence between the BOLD responses in a large number of neighboring voxels at one time, and the temporal dependence among the BOLD responses of the same subject repeatedly measured at different time points (see Ashby, 2011).

This article aims to develop a data-driven and nonparametric method to detect and identify temporal changes in a course of high-dimensional time dependent data. Specifically, letting $X_{it} = (X_{it1}, \dots, X_{itp})'$ be a p-dimensional random vector observed for the ith subject ($i = 1, \dots, n$) at time t ($t = 1, \dots, T$), we are interested in testing

$$
\begin{aligned}
&H_0 : \mu_1 = \dots = \mu_T, \quad \text{vs.} \\
&H_1 : \mu_1 = \dots = \mu_{\tau_1} \neq \mu_{\tau_1+1} = \dots = \mu_{\tau_q} \neq \mu_{\tau_q+1} = \dots = \mu_T,
\end{aligned} \tag{1}
$$

where $\mu_t = E(X_{it})$ ($t = 1, \dots, T$) is a p-dimensional population mean vector and $1 \le \tau_1 < \dots < \tau_q < T$ are q ($q < \infty$) unknown locations of change points. If the null hypothesis is rejected, we will further estimate the locations of change points. The above hypotheses assume that all the individuals come from the same population with the same mean vectors and change points. In many applications, such as fMRI studies, it is more meaningful to allow the responding mechanism to be different across subjects. This motivates us to further generalize the above hypotheses to (14), where the whole population consists of G ($G > 1$) groups, and each group has its own unique means and change-points. A mixture model is proposed to accommodate such group effect (the details will be introduced in Section 2.5).

The classical multivariate analysis of variance (MANOVA) assumes independent normal populations with mean vectors $\mu_1, \dots, \mu_T$ and a common covariance. In the classical setting with $p < n$, the likelihood ratio test Wilks (1932) and Hotelling's T^2 test are commonly applied. When $p > n$, Dempster (1958), Dempster, 1960) first considered the MANOVA in the case of a two-sample problem. Since then, more methods have been developed. For instance, Bai and Saranadasa (1996) proposed a test by assuming p/n is a finite constant. Chen and Qin (2010) further improved the test in Bai and Saranadasa (1996) by proposing a test statistic formulated through the U-statistics (see also Schott, 2007; Srivastava & Kubokawa, 2013). Recently, Wang, Peng, and Li (2015) proposed a new multivariate test, which can accommodate heavy-tailed data. Readers are referred to Fujikoshi, Ulyanov, and Shimizu (2010) and Hu, Bai, Wang, and Wang (2017) for excellent reviews.

There exist several significant differences between the hypotheses (1) considered in this article and the classical MANOVA problem. First, the number of mean vectors T in (1) can be large, whereas the classical MANOVA considers the comparison of a small number of mean vectors. Second, the data considered in this article exhibit complex temporal and spatial dependence.

The MANOVA problem typically considers inference for independent samples without taking into account temporal dependence among $\{X_{it}\}_{t=1}^{T}$. Moreover, the classical MANOVA problem assumes the homogeneity among subjects, while this article also considers the mixture model to accommodate the group effect such that each group is allowed to have its own mean vectors and change points. Based on the above, none of the aforementioned MANOVA methods can be applied to test the hypotheses (1). What fundamentally distinguishes our work from the MANOVA research is that our work is closely related to research on change point detection; in contrast to MANOVA, the change points τ_j are unknown. There is a small but growing body of research on change point detection for high-dimensional data. Cho and Fryźlewicz (2015), Chen and Zhang (2015) and Jirak (2015) focus on change-point identification for high-dimensional time series or panel data with only one subject ($n = 1$). More recently, Wang and Samworth (2018) propose a sparse projection based method for high-dimensional change point estimation. Our approach takes into account both temporal and spatial dependence and imposes only weak moment conditions. The work of Aston and Kirch (2012a, 2012b) is also motivated by and applied to fMRI data very similar to that we consider in Section 5 (they focus on resting state fMRI). Their change-point detection methodology can only be applied to each subject separately. The essential innovation of our approach is that it is applicable to different data structures for which the existing approaches cannot be used. It should thus be seen as complementing and extending these approaches rather than competing with them.

The rest of the article is organized as follows. Section 2 introduces temporal homogeneity tests for the equality of high-dimensional mean vectors and studies their asymptotics, where Section 2.5 extends these methods to the mixture model. Section 3 proposes a change point identification estimator whose rate of convergence is derived. To further identify multiple change points, we consider a binary segmentation algorithm, which is shown to be consistent. Simulation experiments and a case study are conducted in Sections 4 and 5, respectively, to demonstrate the empirical performance of the proposed methods. A brief discussion is given in Section 6. All proofs are relegated to the Appendix. Some technical lemmas and additional simulation results are included in the supplemental material.

2 | TEMPORAL HOMOGENEITY TESTS

2.1 | Notation and data Model

We observe p-dimensional vectors X_{it} for ith individual at tth time point ($i = 1, \dots, n$ and $t = 1, \dots, T$). We assume that the observations are independent and identically distributed across individuals. This assumption is relaxed in Section 2.5. The mean and covariance of X_{it} are, respectively, μ_t and Σ_t. The covariance between X_{is} and X_{it} is defined as Ξ_{st}, which quantifies temporal correlation between X_{is} and X_{it} for the same individual measured at different time points s and t. The matrix Ξ_{st} becomes the covariance matrix Σ_t if $s = t$, and then describes the spatial dependence of X_{it} at time t. Define $X_i = (X'_{i1}, X'_{i2}, \dots, X'_{iT})'$ and $\mathrm{Var}(X_i) = \Sigma$. Then, Σ is a $(pT) \times (pT)$ matrix in which each main diagonal square matrix of size p represents the spatial dependence among the components of X_{it}, and each off diagonal square matrix represents the temporal dependence between X_{is} and X_{it} with $s \neq t$. Clearly, Σ becomes a block diagonal matrix if there is no temporal dependence.

We model X_{it} using a general factor model:

$$X_{it} = \mu_t + \Gamma_t Z_i \quad \text{for } i = 1, \dots, n \quad \text{and} \quad t = 1, \dots, T, \tag{2}$$

where Γ_t is a $p \times m$ matrix ($m \geq pT$) satisfying $\Gamma\Gamma' = \Sigma$ with $\Gamma = (\Gamma_1', \ldots, \Gamma_T')'$. The Z_i are m-variate i.i.d. random vectors satisfying $\mathrm{E}(Z_i) = 0$, $\mathrm{Var}(Z_i) = I_m$, the $m \times m$ identity matrix. If we write $Z_i = (z_{i1}, \ldots, z_{im})'$ and let Δ be a finite constant, we further assume that

$$\mathrm{E}(z_{ik}^4) = 3 + \Delta, \quad \text{and} \quad \mathrm{E}(z_{ik_1}^{l_1} z_{ik_2}^{l_2} \ldots z_{ik_h}^{l_h}) = \mathrm{E}(z_{ik_1}^{l_1})\mathrm{E}(z_{ik_2}^{l_2}) \ldots \mathrm{E}(z_{ik_h}^{l_h}), \tag{3}$$

where h is positive integer such that $\sum_{j=1}^{h} l_j \leq 8$ and $l_1 \neq l_2 \neq \ldots \neq l_h$. As in Chen and Qin (2010) and Bai and Saranadasa (1996), assumption (3) is a relaxation of Gaussianity.

We assume that the number of factors m is much larger than p. This includes the commonly used factor model as a special case, if we let the Γ_t be sparse matrices with many columns 0. Note that we do not need to estimate these factors in our detection and identification procedures. The above model facilitates our technical derivation and incorporates both spatial and temporal dependence of the data. Let $\delta_{ij} = 1$ if $i = j$, and 0 otherwise. From (2), it immediately follows that $\mathrm{Cov}(X_{is}, X_{jt}) = \delta_{ij}\Gamma_s\Gamma_t' \equiv \delta_{ij}\Xi_{st}$.

Throughout the article, $a \asymp b$ means that a and b are of the same asymptotic order.

2.2 | A measure of distance

To propose a test statistic for the hypotheses (1), for any $t \in \{1, \ldots, T-1\}$, we first quantify the difference between two sets of mean vectors $\{\mu_{s_1}\}_{s_1=1}^{t}$ and $\{\mu_{s_2}\}_{s_2=t+1}^{T}$ by defining a measure

$$M_t = h^{-1}(t) \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} (\mu_{s_1} - \mu_{s_2})'(\mu_{s_1} - \mu_{s_2}), \tag{4}$$

with the scale function $h(t) = t(T-t)$. We see that M_t is the average of $t(T-t)$ terms, each of which is the Euclidean distance between two population mean vectors chosen before and after a specific $t \in \{1, \ldots, T-1\}$.

Since $M_t = 0$ under H_0 and $M_t \neq 0$ under H_1, it can be used to distinguish the alternative from the null hypothesis. Another advantage of using M_t is that it attains its maximum at one of change-points $\{\tau_1, \ldots, \tau_q\}$ as shown in Lemma 1 in the supplemental material. Thus, it can also be used for identifying change-points when H_0 is rejected (details will be provided in Section 3). There is a connection between M_t and Schott (2007)'s test statistic based on the measure $S_{1T} = T\sum_{s=1}^{T}(\mu_s - \overline{\mu})'(\mu_s - \overline{\mu}) = \sum_{1 \leq s_1 < s_2 \leq T}(\mu_{s_1} - \mu_{s_2})'(\mu_{s_1} - \mu_{s_2})$, where $\overline{\mu} = \sum_{s=1}^{T} \mu_s / T$. It can be shown that $S_{1T} = h(t)M_t + S_{1t} + S_{(t+1)T}$. Note that S_{1t} measures distance among mean vectors before time t and $S_{(t+1)T}$ measures distance among mean vectors after time t. Neither S_{1t} nor $S_{(t+1)T}$ are informative for the differences between the mean vectors $\{\mu_{s_1}\}_{s_1=1}^{t}$ and $\{\mu_{s_2}\}_{s_2=t+1}^{T}$.

Given a random sample $\{X_{it}\}$, M_t can be estimated by

$$\widehat{M}_t = \frac{1}{h(t)n(n-1)} \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \left(\sum_{i \neq j}^{n} X_{is_1}' X_{js_1} + \sum_{i \neq j}^{n} X_{is_2}' X_{js_2} - 2 \sum_{i \neq j}^{n} X_{is_1}' X_{js_2} \right).$$

If the subjects are independent, elementary calculations show that $\mathrm{E}(\widehat{M}_t) = M_t$. Thus, $\widehat{M}_t$ is an unbiased estimator of M_t. If $T = 2$, the above statistic reduces to the two-sample U-statistic studied by Chen and Qin (2010) for testing the equality of two high-dimensional population means. However, the change points detection problem considered in this article is significantly different from the two sample mean testing problem considered in Chen and Qin (2010). The method proposed in Chen and Qin (2010) is not applicable to our change points detection problem.

We conclude this section by computing the variance of $\widehat{M}_t$. The expression we obtained will be used to formulate our test procedure. Define

$$A_{0t} = \sum_{r_1=1}^{t} \sum_{r_2=t+1}^{T} (\Gamma_{r_1} - \Gamma_{r_2})'(\Gamma_{r_1} - \Gamma_{r_2}) \text{ and } A_{1t} = \sum_{r_1=1}^{t} \sum_{r_2=t+1}^{T} (\mu_{r_1} - \mu_{r_2})'(\Gamma_{r_1} - \Gamma_{r_2}). \tag{5}$$

Proposition 1. *Under (2),*

$$\mathrm{Var}(\widehat{M}_t) \equiv \sigma_{nt}^2 = h^{-2}(t) \left\{ \frac{2}{n(n-1)} \mathrm{tr}(A_{0t}^2) + \frac{4}{n} \|A_{1t}\|^2 \right\}, \tag{6}$$

where A_{0t} and A_{1t} are specified in (5), and $\|\cdot\|$ denotes the vector l^2-norm.

Observe that A_{1t} becomes a $1 \times m$ vector of zeros under H_0 of (1). Proposition 1 implies that the variance of $\widehat{M}_t$ under H_0 is

$$\sigma_{nt,0}^2 = 2\mathrm{tr}(A_{0t}^2)/\{h^2(t)n(n-1)\}. \tag{7}$$

2.3 | Asymptotic properties of $\widehat{M}_t$

To establish the asymptotic normality of the statistic $\widehat{M}_t$ at any $t \in \{1, \dots, T-1\}$, we require the following condition.

(C1). As $n \to \infty$, $p \to \infty$ and $T \to \infty$, $\mathrm{tr}(A_{0t}^4) = o\{\mathrm{tr}^2(A_{0t}^2)\}$. In addition, under H_1, $A_{1t}A_{0t}^2A_{1t}' = o\{\mathrm{tr}(A_{0t}^2)\, \|A_{1t}\|^2\}$.

The first part of the condition (C1) is a generalization of condition (3.6) in Chen and Qin (2010) from a fixed T to the diverging T case, which is a mild condition. To appreciate this point, consider a scenario without temporal dependence. In this case,

$$\mathrm{tr}(A_{0t}^2) = \sum_{i=1}^{t} \sum_{j=t+1}^{T} \mathrm{tr}\{(\Sigma_i + \Sigma_j)^2\} + (T-t)(T-t-1) \sum_{i=1}^{t} \mathrm{tr}(\Sigma_i^2) + t(t-1) \sum_{j=t+1}^{T} \mathrm{tr}(\Sigma_j^2)$$

and

$$\mathrm{tr}(A_{0t}^4) = 2\left[\sum_{i=1}^{t} \sum_{j=t+1}^{T} \sum_{l=j+1}^{T} \mathrm{tr}\{((T-t)\Sigma_i^2 + \Sigma_i\Sigma_l + \Sigma_j\Sigma_i)^2\} + \sum_{i=1}^{t} \sum_{j=i+1}^{t} \sum_{k=t+1}^{T} \mathrm{tr}\{(t\Sigma_k^2 + \Sigma_i\Sigma_k + \Sigma_k\Sigma_j)^2\} \right] \times \{1 + o(1)\}.$$

If all the eigenvalues of Σ_t $(t = 1, \dots, T)$ are bounded, then $\mathrm{tr}(A_{0t}^4) \asymp T^5 p$ and $\mathrm{tr}^2(A_{0t}^2) \asymp T^6 p^2$. Thus, $\mathrm{tr}(A_{0t}^4) = o\{\mathrm{tr}^2(A_{0t}^2)\}$ as $p \to \infty$. If some of the eigenvalues of Σ_t diverges too fast such that $\mathrm{tr}(\Sigma_t^4) \asymp p^4$ and $\mathrm{tr}^2(\Sigma_t^2) \asymp p^4$ and the temporal dependence are very strong (e.g., both Σ_t and temporal correlation have the compound symmetric structure), then $\mathrm{tr}(A_{0t}^4) \asymp \mathrm{tr}^2(A_{0t}^2)$, which violates the condition. In this scenario, the asymptotic normality in Theorem 1 may not hold, and our proposed detection procedure needs some modification. A detailed discussion and more examples may be found in Zhong, Lan, Song, and Tsai (2017).

Define $V_{1t}' = (\Gamma_1', \dots, \Gamma_1', \dots, \Gamma_t', \dots, \Gamma_t')$, where each Γ_l $(1 \le l \le t)$ is repeated for $T-t$ times, and $V_{2t}' = (\Gamma_{t+1}', \dots, \Gamma_T', \dots, \Gamma_{t+1}', \dots, \Gamma_T')$ where $(\Gamma_{t+1}', \dots, \Gamma_T')$ is repeated for t times. Then, we can write $A_{0t} = (V_{1t} - V_{2t})'(V_{1t} - V_{2t})$, which has the same nonzero eigenvalues as $A_{0t}^* = (V_{1t} - V_{2t})(V_{1t} - V_{2t})' = V_{1t}V_{1t}' + V_{2t}V_{2t}' - V_{1t}V_{2t}' - V_{2t}V_{1t}'$. Consider the multivariate linear process described in Equation (16) in Section 4.1, where the temporal dependence exists. If J is finite and the eigenvalues of Σ_t are bounded, then $\mathrm{tr}(A_{0t}^2) = \mathrm{tr}(A_{0t}^{*2}) = [\mathrm{tr}\{(V_{1t}V_{1t}')^2\} + \mathrm{tr}\{(V_{2t}V_{2t}')^2\}]\{1 + o(1)\} \asymp T^3 p$ and similarly $\mathrm{tr}(A_{0t}^4) \asymp T^5 p$. Therefore, $\mathrm{tr}(A_{0t}^4) = o\{\mathrm{tr}^2(A_{0t})^2\}$ holds for the multivariate linear process in Equation (16).

Note that the second part of condition (C1) is not needed for establishing the null distribution of our proposed test. Let λ_k be eigenvalues of A_{0t}. If the number of nonzero λ_ks diverges and all the nonzero λ_ks are bounded, the second part of condition (C1) is satisfied. Given that $A_{1t}A_{0t}^2A_{1t}' \le (\max_k \lambda_k^2)\|A_{1t}\|^2$, we have $A_{1t}A_{0t}^2A_{1t}' = o\{\mathrm{tr}(A_{0t}^2)\ \|A_{1t}\|^2\}$ if $\max_k \lambda_k^2 = o\{\mathrm{tr}(A_{0t}^2)\}$.

Theorem 1. *Under (2), (3), and condition (C1), as $n \to \infty$, $p \to \infty$ and $T \to \infty$,*

$$(\widehat{M}_t - M_t)/\sigma_{nt} \xrightarrow{d} N(0, 1),$$

where σ_{nt} is defined in (6).

In particular, under H_0, the variance of $\widehat{M}_t$ is (7) and $\widehat{M}_t/\sigma_{nt,0} \xrightarrow{d} N(0, 1)$. Since $\sigma_{nt,0}^2$ is unknown, to implement a testing procedure, we estimate $\sigma_{nt,0}^2$ by

$$\widehat{\sigma}_{nt,0}^2 = \frac{2}{h^2(t)n(n-1)} \sum_{r_1,s_1=1}^{t} \sum_{r_2,s_2=t+1}^{T} \sum_{a,b,c,d \in \{1,2\}} (-1)^{|a-b|+|c-d|} \mathrm{tr}\left(\widehat{\Gamma_{r_b}'\Gamma_{r_a}\Gamma_{s_c}'\Gamma_{s_d}}\right),$$

where, defining $P_n^4 = n(n-1)(n-2)(n-3)$ to be the permutation number,

$$\mathrm{tr}\left(\widehat{\Gamma_{r_b}'\Gamma_{r_a}\Gamma_{s_c}'\Gamma_{s_d}}\right) = \frac{1}{P_n^4} \sum_{i \ne j \ne k \ne l}^{n} \left(X_{ir_a}'X_{jr_b}X_{is_c}'X_{js_d} - X_{ir_a}'X_{jr_b}X_{is_c}'X_{ks_d} - X_{ir_a}'X_{jr_b}X_{ks_c}'X_{js_d} + X_{ir_a}'X_{jr_b}X_{ks_c}'X_{ls_d}\right). \tag{8}$$

Note that the computational cost of $\widehat{\sigma}_{nt,0}^2$ is not an issue. The main reason is twofold. First, some simple algebra can be applied to simplify the computation of the summations so that the computation complexity is at the order of $O(n^2T^2p)$. Second, the computational cost is mainly due to the size of n, T, not p, but n and T are typically not prohibitively large in fMRI and genomics applications.

The ratio consistency of $\widehat{\sigma}_{nt,0}^2$ is established by the following theorem.

Theorem 2. *Assume the same conditions in Theorem 1. As $n \to \infty$, $p \to \infty$, and $T \to \infty$,*

$$\widehat{\sigma}_{nt,0}^2/\sigma_{nt,0}^2 - 1 = O_p\left\{n^{-\frac{1}{2}}\mathrm{tr}^{-1}(A_{0t}^2)\mathrm{tr}^{\frac{1}{2}}(A_{0t}^4) + n^{-1}\right\} = o_p(1).$$

For a fixed t, Theorems 1 and 2 lead to a testing procedure that rejects H_0 if

$$\widehat{M}_t/\widehat{\sigma}_{nt,0} > z_\alpha, \tag{9}$$

where z_α is the upper α quantile of N(0, 1). A change point test must take into account all potential change points $t \in \{1, \dots, T-1\}$, this is what we do in the next section.

2.4 | Change point tests

To make the testing procedure for (1) free of tuning parameters, it is natural to consider the statistic

$$\widehat{\mathscr{M}} = \max_{0<t/T<1} \widehat{M}_t / \widehat{\sigma}_{nt,0}. \tag{10}$$

It formally resembles the maximally selected likelihood ratio statistic, see chapter 1 of Csörgő and Horváth (1997), so it may be hoped that it possesses some asymptotic optimality properties, but may also suffer from a slow convergence rate, as it might also converge to a Gumbel distribution. Theorem 3 shows that this is indeed the case. If T is finite, the asymptotic null distribution is not parameter-free. In this case, an adaptation of the method proposed in Peštová and Pešta (2015)) might be useful. In the case of $T \to \infty$, an extension of the self-normalized statistics proposed by Pešta and Wendler (2019) might offer an alternative approach.

To establish the asymptotic null distribution of $\widehat{\mathscr{M}}$, we need the following condition.

(C2). There exist $\phi(k) > 0$ satisfying $\sum_{k=1}^{\infty} \phi^{1/2}(k) < \infty$ such that for any $r, s \geq 1$, $\mathrm{tr}(\Xi_{rs}\Xi_{rs}') \asymp \phi(|r-s|)\mathrm{tr}(\Sigma_r\Sigma_s)$.

Condition (C2) imposes a mild weak dependence assumption on the time series $\{X_{it}\}_{t=1}^{T}$. To describe the limit of $\widehat{\mathscr{M}}$, we define the correlation coefficient

$$r_{nz,uv} = 2\mathrm{tr}(A_{0u}A_{0v})/\{n(n-1)h(u)h(v)\sigma_{nu,0}\sigma_{nv,0}\} \text{ and its limit } r_{z,uv} = \lim_{n\to\infty} r_{nz,uv},$$

Theorem 3. *Suppose (2), (3), (C1), (C2), and H_0 of (1) hold. As $n \to \infty$ and $p \to \infty$, (i) if T is finite, $\widehat{\mathscr{M}} \xrightarrow{d} \max_{0<t/T<1} W_t$, where W_t is the tth component of $W = (W_1, \dots, W_{T-1})' \sim \mathrm{N}(0, R_Z)$ with $R_Z = (r_{z,uv})$; (ii) if $T \to \infty$ and the maximum eigenvalue of R_Z is bounded, then $P(\widehat{\mathscr{M}} \leq \sqrt{2\log(T) - \log\log(T) + x}) \to \exp\{-(2\sqrt{\pi})^{-1}\exp(-x/2)\}$.*

To study the asymptotic power of the proposed test, we study the asymptotic behavior of the statistic $\widehat{\mathscr{M}}$ under local alternatives. For any fixed constants $1 > \eta > \nu > 0$, let $[T\nu]$ and $[T\eta]$ be largest integers no greater than $T\nu$ and $T\eta$, respectively. Define the following notations similar to (5), $A_{0,\nu\eta} = \sum_{r_1=1}^{[T\nu]}\sum_{r_2=[T\nu]+1}^{[T\eta]} (\Gamma_{r_1} - \Gamma_{r_2})'(\Gamma_{r_1} - \Gamma_{r_2})$,

$$A_{1,\nu\eta}^{(1)} = \sum_{r_1=1}^{[T\nu]}\sum_{r_2=[T\nu]+1}^{[T\eta]} (\mu_{r_1} - \mu_{r_2})'(\Gamma_{r_1} - \Gamma_{r_2}) \quad \text{and} \quad A_{1,\nu\eta}^{(2)} = \sum_{r_1=[T\nu]+1}^{[T\eta]}\sum_{r_2=[T\nu]}^{T} (\mu_{r_1} - \mu_{r_2})'(\Gamma_{r_1} - \Gamma_{r_2}).$$

Let $\sigma_{nuv} = [2\mathrm{tr}(A_{0u}A_{0v})/\{n(n-1)\} + 4A_{1u}A_{1v}^T/n]/\{h(u)h(v)\}$ be the covariance between $\widehat{M}_u$ and $\widehat{M}_v$ and define $r^*_{nz,uv} = \sigma_{nuv}/\sqrt{\sigma_{nuu}\sigma_{nvv}}$ as the corresponding correlation between $\widehat{M}_u$ and $\widehat{M}_v$. Let $r^*_{z,uv}$ be the limit of $r^*_{nz,uv}$. Consider the local alternatives H_{1n} that satisfy the following condition

$$\max\{\| \underset{1,\nu\eta}{\overset{(1)}{\mathrm{A}}} \|^2, \|A_{1,\nu\eta}^{(2)}\|^2\} = o\{\mathrm{tr}(A_{0,\nu\eta}^2/n)\}. \tag{11}$$

Theorem 4. *Suppose (2), (3), (C1), and (C2) hold. Under the local alternatives H_{1n} defined in (11), and assuming that $M_t/\sigma_{nt,0} \to M_t/\sigma_{t,0}$ and the sequence $M_t/\sigma_{t,0}$ is bounded. If $n \to \infty$ and $p \to \infty$, then $\widehat{\mathscr{M}}$ and $\max_{0<t<T}\left(W_t^* + M_t/\sigma_{t,0}\right)$ have the same limiting distribution for finite T or $T \to \infty$, where $W_t^* = (W_1^*, \dots, W_{T-1}^*)'$ is a Gaussian process with mean 0 and covariance R_Z^* with the (u, v) component $r^*_{z,uv}$.*

Due to the slow convergence suggested by Theorem 3, the empirical sizes based on $\widehat{\mathscr{M}}$ might not be accurate in finite samples. To address this issue, we propose a different test statistic by combining the building blocks of the $\widehat{M}_t$ in a different way, and define

$$\widehat{S}_n = \frac{2}{T(T-1)n(n-1)} \sum_{i \neq j}^{n} \sum_{s_1 < s_2}^{T} \left(X_{is_1}' X_{js_1} + X_{is_2}' X_{js_2} - 2X_{is_1}' X_{js_2} \right).$$

Theorem 5. *Suppose (2), (3), and (C1) hold.*
Let $S_n = 2\sum_{s_1<s_2}(\mu_{s_1} - \mu_{s_2})'(\mu_{s_1} - \mu_{s_2})/\{T(T-1)\}$. *As* $n \to \infty$, $p \to \infty$, *and* $T \to \infty$, $\sigma_n^{-1}(\widehat{S}_n - S_n) \xrightarrow{d} N(0,1)$, *where* $\sigma_n^2 = \left\{2\text{tr}(A_0^2)/\{n(n-1)\} + 4\|A_1\|^2/n\right\}/\{T(T-1)\}^2$. *Here* $A_0 = \sum_{r_1<r_2}^{T}(\Gamma_{r_1} - \Gamma_{r_2})'(\Gamma_{r_1} - \Gamma_{r_2})$ *and* $A_1 = \sum_{r_1<r_2}^{T}(\mu_{r_1} - \mu_{r_2})'(\Gamma_{r_1} - \Gamma_{r_2})$.

The convergence to the normal limit is due to replacing the maximum norm in $\widehat{\mathscr{M}}$ by a sum in $\widehat{S}_n$. Our proposed test statistic thus is

$$\widehat{\mathcal{S}}_n = \widehat{\sigma}_{n0}^{-1}\widehat{S}_n,$$

with

$$\widehat{\sigma}_{n0}^2 = \frac{2}{T^2(T-1)^2 n(n-1)} \sum_{r_1<r_2=1}^{T} \sum_{s_1<s_2=1}^{T} \sum_{a,b,c,d \in \{1,2\}} (-1)^{|a-b|+|c-d|} \text{tr}\left(\widehat{\Gamma_{r_b}'\Gamma_{r_a}\Gamma_{s_c}'\Gamma_{s_d}}\right),$$

and $\text{tr}(\widehat{\Gamma_{r_b}'\Gamma_{r_a}\Gamma_{s_c}'\Gamma_{s_d}})$ is defined in (8) in Section 2.3. Therefore, by Theorem 5, an asymptotic α-level test rejects null hypothesis if

$$\widehat{\mathcal{S}}_n > z_\alpha, \tag{12}$$

where z_α is the upper α quantile of the standard normal distribution.

2.5 | An extension to mixture models

Thus far we have focused on change point detection assuming that all subjects in the sample come from a population with the same potential change-points. In fMRI experiments, if different subjects choose different strategies to solve the same task, the patterns activated by stimuli will be different across subjects (see Ashby, 2011). Analytically, it is more attractive to consider that subjects show the same activation pattern within each group, but different patterns across groups.

In this subsection, we will generalize the approaches developed in the Sections 2.1–2.4 to accommodate such group effect. Instead of the model (2) considered in Section 2.1, we assume that data follow a mixture model

$$X_{it} = \sum_{g=1}^{G} \Lambda_{ig}\mu_{gt} + \Gamma_t Z_i, \tag{13}$$

where independent of $\{Z_i\}_{i=1}^n$, $(\Lambda_{i1}, \dots, \Lambda_{iG})$ follows a multinomial distribution with parameters 1 and $p = (p_1, \dots, p_G)$. This implies that $\sum_{g=1}^{G} \Lambda_{ig} = 1$ with $\Lambda_{ig} \in \{0,1\}$, and $\text{P}(\Lambda_{ig} = 1) = p_g$

satisfying $\sum_{g=1}^{G} p_g = 1$ with the number of groups $G \geq 1$. Note that the above model implies that ith subject only belongs to one of the G groups. The mixture model (13) allows each group to have its own population mean vectors $\{\mu_{gt}\}_{t=1}^{T}$ for $g = 1, \dots, G$. It reduces to (13) if there is only one group ($G = 1$).

In analogy to (1), we want to know whether there exist some change-points within some groups by testing

$$H_0^* : \mu_{g1} = \mu_{g2} = \dots = \mu_{gT} \quad \text{for all } \ 1 \leq g \leq G \quad \text{vs.}$$

$$H_1^* : \mu_{g1} = \dots = \mu_{g\tau_1^{(g)}} \neq \mu_{g(\tau_1^{(g)}+1)} = \dots = \mu_{g\tau_{q_g}^{(g)}} \neq \mu_{g(\tau_{q_g}^{(g)}+1)} = \dots = \mu_{gT}$$

$$\text{for some } \ g. \tag{14}$$

If H_0^* is rejected, we further identify $\{\tau_1^{(g)}, \tau_2^{(g)} \dots, \tau_{q_g}^{(g)}\}_{g=1}^{G}$, the collection of q ($q = \sum_{g=1}^{G} q_g$) change-points from G groups.

Toward this end, we first evaluate the mean and variance of the statistic $\widehat{M}_t$ under the mixture model (13). Similar to Proposition 1, the mean is $E(\widehat{M}_t) = \tilde{M}(t) = h^{-1}(t) \sum_{r_1=1}^{t} \sum_{r_2=t+1}^{T} (\tilde{\mu}_{r_1} - \tilde{\mu}_{r_2})'(\tilde{\mu}_{r_1} - \tilde{\mu}_{r_2})$ with $\tilde{\mu}_{r_i} = \sum_{g=1}^{G} p_g \mu_{gr_i}$ for $i = 1, 2$. The variance of $\widehat{M}_t$ is

$$\text{Var}(\widehat{M}_t) \equiv \tilde{\sigma}_{nt}^2 = \frac{2}{n(n-1)h^2(t)} \{\text{tr}(A_{0t}^2) + \tilde{A}_{3t}\} + \frac{4}{nh^2(t)} \{\|\tilde{A}_{1t}\|^2 + \tilde{A}_{2t}\}, \tag{15}$$

where A_{0t} is defined in (5), $\tilde{A}_{1t} = \sum_{r_1=1}^{t} \sum_{r_2=t+1}^{T} (\tilde{\mu}_{r_1} - \tilde{\mu}_{r_2})'(\Gamma_{r_1} - \Gamma_{r_2})$. In addition, with $\delta_{g_1g_2r_i} = \mu_{g_1r_i} - \mu_{g_2r_i}$ for $i = 1, 2$,

$$\tilde{A}_{2t} = \sum_{g_1<g_2}^{G} p_{g_1} p_{g_2} \left\{ \sum_{r_1=1}^{t} \sum_{r_2=t+1}^{T} (\delta_{g_1g_2r_1} - \delta_{g_1g_2r_2})'(\tilde{\mu}_{r_1} - \tilde{\mu}_{r_2}) \right\}^2 \quad \text{and}$$

$$\tilde{A}_{3t} = \sum_{g_1<g_2, g_3<g_4}^{G} p_{g_1} p_{g_2} p_{g_3} p_{g_4} \left\{ \sum_{r_1=1}^{t} \sum_{r_2=t+1}^{T} (\delta_{g_1g_2r_1} - \delta_{g_1g_2r_2})'(\delta_{g_3g_4r_1} - \delta_{g_3g_4r_2}) \right\}^2.$$

It is worth discussing some special cases of (15). First, if there is only one group ($G = 1$), it can be shown that $\tilde{A}_{2t} = \tilde{A}_{3t} = 0$, and $\tilde{A}_{1t} = A_{1t}$ defined in (5). Therefore, the variance formulated in Proposition 1 is a special case of the variance (15) under the mixture model. Second, under H_0^* of (14), $\tilde{\sigma}_{nt,0}^2 \equiv \text{Var}(\widehat{M}_t) = 2\text{tr}(A_{0t}^2)/\{n(n-1)h^2(t)\}$ because $\tilde{A}_{1t} = \tilde{A}_{2t} = \tilde{A}_{3t} = 0$. The unknown $\tilde{\sigma}_{nt,0}^2$ can be estimated by

$$\hat{\tilde{\sigma}}_{nt,0}^2 = \frac{2}{h^2(t)n^2(n-1)^2} \sum_{i\neq j}^{n} \left\{ \sum_{r_1=1}^{t} \sum_{r_2=t+1}^{T} \sum_{a,b\in\{1,2\}} (-1)^{|a-b|} X_{ir_a}' X_{jr_b} \right\}^2.$$

Asymptotic results of Section 2 can be extended to the mixture model (13) under some regularity conditions. We do not state these notationally complex results, but demonstrate the empirical performance under the mixture model through simulation studies.

3 | CHANGE POINTS IDENTIFICATION

When H_0 of (1) is rejected, it is often useful to identify the change points. We first consider the case of a single change point $\tau \in \{1, \dots, T-1\}$. It can be shown that M_t attains its maximum at τ, which motivates us to identify the change point τ by the following estimator

$$\widehat{\tau} = \arg \max_{0<t/T<1} \widehat{M}_t.$$

Let

$$v_{\max} = \max_{1 \le t \le T-1} \max \left\{ \sqrt{\mathrm{tr}(\Sigma_t^2)}, \sqrt{n(\mu_1 - \mu_T)' \Sigma_t (\mu_1 - \mu_T)} \right\}$$

and $\delta^2 = (\mu_1 - \mu_T)'(\mu_1 - \mu_T)$. The following theorem establishes the rate of convergence for the change point estimator $\widehat{\tau}$.

Theorem 6. *Assume that a change-point $\tau = \tau_T$ satisfies $\lim_{T \to \infty} \tau/T = \kappa$ with $0 < \kappa < 1$. Assume that $(\mu_1 - \mu_T)' \Xi_{rs} (\mu_1 - \mu_T) \asymp \phi(|r-s|)(\mu_1 - \mu_T)' \Sigma_r (\mu_1 - \mu_T)$, where $\phi(\cdot)$ is defined in condition (C2). Under (2), (3), (C1) and (C2), as $n \to \infty$,*

$$\widehat{\tau} - \tau = O_p \left\{ \sqrt{T \log(T)} \; v_{\max}/(n \, \delta^2) \right\}.$$

Theorem 6 shows that $\widehat{\tau}$ is consistent if $n\delta^2/\{v_{\max}\sqrt{T\log(T)}\} \to \infty$, where $n\delta^2$ is a measure of signal and $v_{\max}$ is associated with noise. It explicitly demonstrates the contributions of the dimension p, series length T, and sample size n to the rate of convergence. First, if both p and T are fixed, $\widehat{\tau} - \tau = O_p(n^{-1/2})$ as $n \to \infty$. Second, if p is fixed but T diverges as n increases, $\widehat{\tau} - \tau = O_p(\sqrt{T\log(T)/n})$. Finally, if both p and T diverge as n increases, the convergence rate can be faster than $O_p(\sqrt{T\log(T)/n})$. To appreciate this, we consider a special setting where X_{it} in (2) has the identity covariance $\Sigma_t = I_p$, the nonzero components of δ^2 are equal and fixed, and the number of nonzero components is $p^{1-\beta}$ for $\beta \in (0, 1)$. Under such setting,

$$\widehat{\tau} - \tau = O_p \left(\frac{\{T \log(T)\}^{1/2}}{\min\{np^{1/2-\beta}, n^{1/2} \, p^{(1-\beta)/2}\}} \right),$$

which is faster than the rate $O_p\{\sqrt{T\log(T)/n}\}$ if $n^{1/2}p^{1/2-\beta} \to \infty$.

Next, we consider the case of more than one change-point. To identify these change-points, we first introduce some notation. Let $\mathbb{S} = \{1 \le \tau_1 < \dots < \tau_q < T\}$ be the set containing all q ($q \ge 1$) change points. For any $t_1, t_2 \in \{1, \dots, T\}$ satisfying $t_1 < t_2$, let $\widehat{\mathcal{S}}_n[t_1, t_2]$ denote the test statistic in Section 2.4 computed using data within $[t_1, t_2]$. Lemma 1 in the supplemental material shows that M_t in (4) always attains its maximum at one of the change-points, which motivates us to identify all change points by the following binary segmentation algorithm (Venkatraman, 1992; Vostrikova, 1981).

1 Check if $\widehat{\mathcal{S}}_n[1, T] \le z_{\alpha_n}$. If yes, then no change point is identified and stop. Otherwise, a change point $\widehat{\tau}_{(1)}$ is selected by $\widehat{\tau}_{(1)} = \arg\max_{1 \le t \le T-1} \widehat{M}_t$ and included into $\widehat{\mathbb{S}} = \{\widehat{\tau}_{(1)}\}$;

2 Treat $\{1, \hat{\tau}_{(1)}, T\}$ as new ending points and first check if $\hat{\mathcal{S}}_n[1, \hat{\tau}_{(1)}] \le z_{\alpha_n}[1, \hat{\tau}_{(1)}]$. If yes, no change-point is selected from time 1 to $\hat{\tau}_{(1)}$. Otherwise, one change point is selected by $\hat{\tau}^1_{(2)} = \arg\max_{1\le t\le \hat{\tau}_{(1)}-1} \hat{M}_t$ and updated $\hat{\mathbb{S}}$ by adding $\hat{\tau}^1_{(2)}$. Next check if $\hat{\mathcal{S}}_n[\hat{\tau}_{(1)}+1, T] \le z_{\alpha_n}$. If yes, no time point is selected from time $\hat{\tau}_{(1)}+1$ to T. Otherwise, one change point is selected by $\hat{\tau}^2_{(2)} = \arg\max_{\hat{\tau}_{(1)}+1\le t\le T-1} \hat{M}_t$, and $\hat{\mathbb{S}}$ is updated by including $\hat{\tau}^2_{(2)}$. If no any change point has been identified from both $[1, \hat{\tau}_{(1)}]$ and $[\hat{\tau}_{(1)}+1, T]$, then stop. Otherwise, rearrange $\hat{\mathbb{S}}$ by sorting its elements from smallest to largest and update ending points by $\{1, \hat{\mathbb{S}}, T\}$;
3 Repeat Step 2 until no more change point is identified from each time segment, and obtain the final set $\hat{\mathbb{S}}$ as an estimate of the set $\mathbb{S}$.

Let $S_\mu = \sum_{s_1=1}^T \sum_{s_2\ne s_1} (\mu_{s_1} - \mu_{s_2})'(\mu_{s_1} - \mu_{s_2})/\{T(T-1)\}$. Define $\tau_0 = 1$ and $\tau_{q+1} = T$. Consider intervals $I_{l,l^*} = [\tau_l + 1, \tau_{l^*}]$ with $l + 1 < l^*$. Define the smallest maximum signal-to-noise ratio to be

$$\mathscr{R}^* = \min_{l+1<l^*} \max_{\tau_i \in I_t} S_\mu \left[I_{l,l^*}\right] / \sigma_n \left[I_{l,l^*}\right],$$

where $S_\mu \left[I_{l,l^*}\right]$ and $\sigma_n \left[I_{l,l^*}\right]$ are defined over I_{l,l^*}. To establish the consistency of $\hat{\mathbb{S}}$ obtained from the above binary segmentation algorithm, we need the following condition.

(C3). As $T \to \infty$, τ_i/T converges to κ_i, $0 < \kappa_1 < \ldots < \kappa_q < 1$ ($q \ge 1$ is fixed).

Theorem 7. *Assume (2), (3), (C1)–(C3). Suppose $\mathscr{R}^*$ diverges at a rate such that the upper α_n-quantile of the standard normal distribution $z_{\alpha_n} = o(\mathscr{R}^*)$, as $\alpha_n \to 0$. Furthermore, assume that $v_{\max}[I_{l,l^*}] = o\{n\delta^2[I_{l,l^*}]/\sqrt{T\log(T)}\}$. Then, $\hat{\mathbb{S}} \xrightarrow{p} \mathbb{S}$, as $n \to \infty$ and $T \to \infty$.*

4 | SIMULATION STUDIES

In this section, we evaluate finite sample performance of our methods.

4.1 | Change point detection

We first evaluate the performance of the test (12). To make a comparison, we consider the classical likelihood ratio test (LRT) and a high-dimensional test for MANOVA proposed by Schott (2007). It is well known that the classical likelihood ratio test is applicable only if the dimension p is fixed and $p \le n(T-1)$ in the notation in this article. The test of Schott extends the likelihood ratio test to the high-dimensional setting by allowing $p > n(T-1)$ and $p\{n(T-1)\}^{-1} \to \gamma \in (0, \infty)$. However, both the likelihood ratio and Schott's tests assume temporal independence. As we will demonstrate in the following, their performance is severely affected if the temporal dependence does exist in data; our test is robust to temporal dependence.

The data $\{X_{it}\}$, $i = 1, \ldots, n$ and $t = 1, \ldots, T$, were generated from the following multivariate linear process

$$X_{it} = \mu_t + \sum_{l=0}^{J} Q_{lt}\, \epsilon_{i(t-l)}, \tag{16}$$

where μ_t is the p-dimensional population mean vector at time t, Q_{lt} is a $p \times p$ matrix, and ϵ_{it} is p-variate normally distributed with mean 0 and identity covariance I_p. The model generates both the temporal dependence of X_{it} and X_{is} at $t \neq s$ and the spatial dependence among the p components of X_{it}. Specifically, it can be seen that $\mathrm{Cov}(X_{it}, X_{is}) = \sum_{l=t-s}^{J} Q_{lt} Q_{(l-t+s)s}$ if $t - s \leq J$ and $\mathrm{Cov}(X_{it}, X_{is}) = 0$ otherwise. The maximum lag J controls the extent of temporal dependence; if $J = 0$, data are temporally independent.

We use $J = 0, 2$ and $Q_{lt} = \{0.5^{|i-j|} \mathrm{I}(|i - j| < p/2)/(J - l + 1)\}$ for $i, j = 1, \dots, p$ and $0 \leq l \leq J$. To evaluate the empirical size of all three tests, we set $\mu_t = 0$ for all t. Under H_1, we considered one change point located at κT such that $\mu_t = 0$ for $t = 1, \dots, \kappa T$ and $\mu_t = \mu$ for $t = \kappa T + 1, \dots, T$. Two κ values 0.1 and 0.4 were used in our simulation. The nonzero mean vector μ had $[p^{0.7}]$ nonzero components, which were uniformly and randomly drawn from p coordinates $\{1, \dots, p\}$. The magnitude of nonzero entry of μ was controlled by a constant δ multiplied by a random sign. The effect of sample size, dimensionality, and length of time series on the performance of the proposed testing procedure was demonstrated by different combinations of $n \in \{30, 60, 90\}$, $p \in \{50, 200, 600, 1,000\}$, and $T \in \{50, 100, 150\}$. The nominal significance level is .05. All simulation results were obtained based on 1,000 replications.

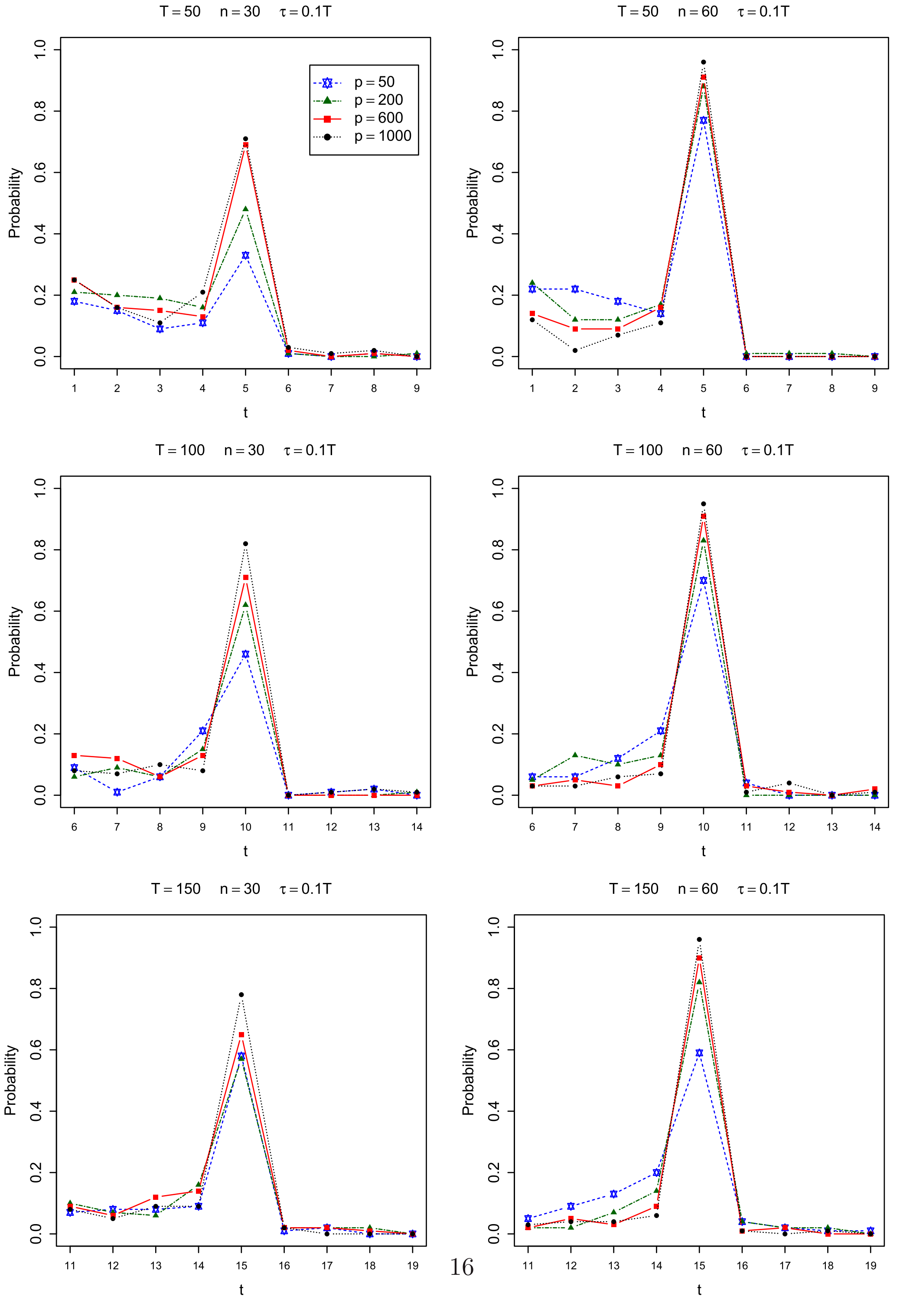
Table 1 summarizes the empirical sizes of the above three tests. The sizes of the LRT could not be computed in some cases with $p = 600$ and 1,000 due to the aforementioned upper bound on p. Under temporally independence, $J = 0$, the LRT is optimal for $p = 50$, but it overrejects or cannot be applied for larger values of p. For those values of p, our test and the test of Schott give comparable results. Under temporally dependence, $J = 2$, only our test is reliable, and the test of Schott is practically unusable. We emphasize that Schott's test was developed for temporally independent data, so the above evaluation is not its criticism, but rather stresses the need for a new test.

Table 2 displays the empirical power of our test for $J = 2$ for two change points at $\tau = 0.1T$ and $0.4T$. The power increases as the dimension p, the sample size n, and the series length T increase. The results also demonstrate the effect of the change point location on the power of the test; it is easier to detect a change if the two samples are of comparable length.

4.2 | Change point identification

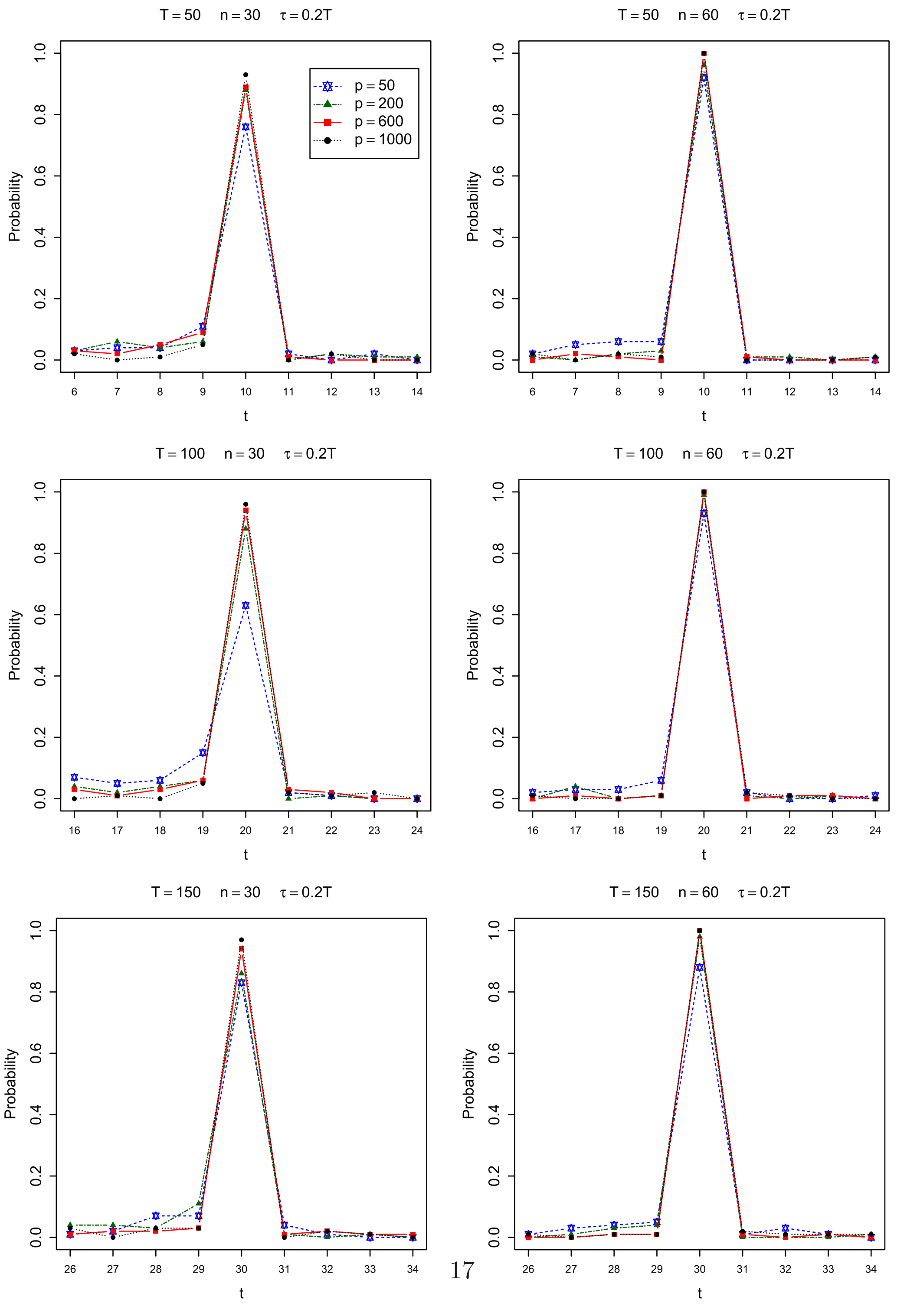
We now evaluate finite sample properties of the change point identification procedure of Section 3. We generated data using a similar setup as in the previous subsection, namely, we considered one change-point at κT with $\kappa = 0.1$ and 0.2, respectively. The power and location identification improved as κ approaches 1/2. We set $\mu_t = 0$ for $t = 1, \dots, \kappa T$ $\mu_t = \mu$ for $t = \kappa T + 1, \dots, T$. Again, the nonzero mean vector μ had $[p^{0.7}]$ nonzero components, which were uniformly and randomly drawn from $\{1, \dots, p\}$. The nonzero entry of μ was $\delta = 0.6$, multiplied by a random sign. The nominal significance level was chosen to be $\alpha = .05$.

Rather than using standard tables, we display graphs that show the empirical probability (based on 100 simulation replications) of identifying a change point at any specific t in the range where these probabilities are positive. This is done in Figure 1 for $\tau = 0.1T$ and Figure 2 for $\tau = 0.2T$. For each chosen T and n, the probability of identifying the change point increased as the dimension p increased. The probability of detecting the correct change point also increased with the series length T and the sample size n increase. It is easier to correctly detect and identify a change point at $\tau = 0.2T$ than at $\tau = 0.1T$.



16

FIGURE 1 The probability of identifying a change point at $\tau = 0.1T$ subject to different combination of T, n, and p [Colour figure can be viewed at wileyonlinelibrary.com]



17

FIGURE 2 The probability of identifying a change point at $\tau = 0.2T$ subject to different combination of T, n, and p [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Empirical sizes of the likelihood ratio test (LRT), Schott's (Sch), and the proposed test (New) for several combinations of n, p, and T

		$J = 0$											
		$T = 50$				$T = 100$				$T = 150$			
Method	n/p	**50**	**200**	**600**	**1000**	**50**	**200**	**600**	**1000**	**50**	**200**	**600**	**1000**
LRT		0.056	0.079	—	—	0.054	0.053	0.177	—	0.043	0.049	0.099	0.458
Sch	30	0.052	0.058	0.051	0.054	0.055	0.056	0.057	0.064	0.044	0.049	0.041	0.046
New		0.052	0.064	0.052	0.059	0.055	0.056	0.056	0.062	0.048	0.050	0.043	0.050
LRT		0.053	0.061	0.127	——	0.054	0.061	0.064	0.147	0.056	0.053	0.062	0.090
Sch	60	0.041	0.050	0.053	0.043	0.056	0.049	0.054	0.057	0.057	0.052	0.045	0.049
New		0.042	0.050	0.054	0.044	0.059	0.049	0.055	0.059	0.055	0.052	0.046	0.050
LRT		0.047	0.059	0.073	0.187	0.051	0.055	0.043	0.086	0.042	0.056	0.046	0.060
Sch	90	0.060	0.055	0.048	0.038	0.055	0.046	0.057	0.058	0.051	0.045	0.059	0.051
New		0.058	0.053	0.048	0.038	0.055	0.046	0.056	0.056	0.052	0.046	0.060	0.049
		$J=2$											
		$T=50$				$T=100$				$T=150$			
Method	n/p	**50**	**200**	**600**	**1000**	**50**	**200**	**600**	**1000**	**50**	**200**	**600**	**1000**
LRT		0.021	0.267	—	—	0.038	0.208	—	—	0.077	0.184	—	—
Sch	30	0.044	0.011	0.001	0.002	0.062	0.021	0.010	0.003	0.056	0.036	0.011	0.006
New		0.056	0.050	0.057	0.048	0.061	0.039	0.068	0.050	0.050	0.054	0.039	0.057
LRT		0.020	0.034	0.985	—	0.044	0.045	0.843	——	0.037	0.062	0.666	—
Sch	60	0.040	0.016	0.001	0.001	0.050	0.021	0.006	0.003	0.052	0.035	0.017	0.009
New		0.050	0.067	0.059	0.036	0.047	0.053	0.046	0.044	0.049	0.046	0.048	0.058
LRT		0.007	0.011	0.480	—	0.041	0.033	0.310	0.989	0.050	0.036	0.247	0.942
Sch	90	0.034	0.013	0.000	0.000	0.059	0.030	0.010	0.004	0.051	0.039	0.011	0.013
New		0.052	0.050	0.049	0.051	0.062	0.059	0.039	0.055	0.043	0.062	0.035	0.063

There are two types of errors for change point identification: the false positive (FP) and the false negative (FN). The FP means that a time point without changing the mean is wrongly identified as a change point, and the FN refers that a change point is wrongly treated as a time point without changing the mean. The accuracy of the proposed change point identification was measured by the sum of FP and FN. Figure 3 demonstrates the FP+FN associated with the change-point identification procedure for $\tau = 0.1T$ and $0.2T$, respectively, under different combinations of T, n, and p. The average FP+FN decreased as p increased. From left to right, the average FP+FN decreased as n increased. And from up to down, the average FP+FN decreased as the change point got closer to the center of the time interval $[1, T]$.

We also conducted simulation studies for the proposed change point detection and identification methods for non-Gaussian data and mixture models. Due to the space limitation, these results are reported in Section 2 of the supplementary material.

TABLE 2 Empirical power of the proposed test for $J = 2$, under several combinations of n, p, and T and two change point locations

		$\tau = 0.1T$				$\tau = 0.4T$			
T	n/p	50	200	600	1000	50	200	600	1000
50	30	0.086	0.093	0.100	0.129	0.166	0.211	0.285	0.302
	60	0.113	0.160	0.211	0.259	0.355	0.517	0.647	0.741
	90	0.171	0.246	0.316	0.353	0.610	0.781	0.918	0.962
100	30	0.101	0.104	0.141	0.165	0.209	0.310	0.393	0.463
	60	0.157	0.213	0.269	0.320	0.495	0.744	0.894	0.929
	90	0.256	0.358	0.466	0.571	0.817	0.958	0.999	0.998
150	30	0.103	0.133	0.178	0.185	0.290	0.405	0.517	0.580
	60	0.194	0.298	0.381	0.412	0.678	0.881	0.963	0.986
	90	0.329	0.463	0.623	0.695	0.922	0.992	1.000	1.000

5 | REAL DATA ANALYSIS

Recent studies suggest that the parahippocampal region of the brain activates more significantly to images with spatial structures than others without such structures (Epstein & Kanwisher, 1998; Henderson, Larson, & Zhu, 2007). An experiment was conducted to investigate the function of this region in scene processing. During the experiment, 14 students at Michigan State University were presented alternatively with six sets of scene images and six sets of object images. The order of presenting the images follows "sosososososoos" where "s" and "o" represent a set of scene images and object images, respectively. The fMRI data were acquired by placing each brain into a 3T GE Sigma EXCITE scanner. After the data were preprocessed by shifting time difference, correcting rigid-body motion and removing trends (more detail can be found in Henderson, Zhu, & Larson, 2011), the resulting dataset consists of BOLD measurements of $p = 33,866$ voxels from $n = 14$ subjects and at $T = 192$ time points.

Let X_{it} be a p-dim ($p = 33,866$) random vector representing the fMRI image data for the ith subject measured at time point t ($i = 1, \dots, 14$ and $t = 1, \dots, 192$). We first applied the testing procedure described in Section 2.4 to the dataset for testing the homogeneity of mean vectors, namely, the hypothesis (14). The test statistic $\widehat{\mathscr{M}} = 9.117$ with p-value less than 10^{-6}, which indicates existence of change-points. After further implementing the proposed binary segmentation approach, we identified 59 change-points, which is not surprising because the large number of change-points arise from the time-altered scene and object images stimuli. To crosscheck the credibility of the identified change-points, we compared them with the predicted BOLD responses obtained from the convolution of the boxcar function with a gamma HRF function (see Ashby, 2011). In Figure 4, the green solid and the green dot dash curves following the order of presenting the images are predicted BOLD responses to the scene images and object images, respectively. The x-values and y-values of the red stars marked on the curves are the identified change-points and the corresponding BOLD responses. Based on the predicted BOLD response function, we found that 58 out of 59 identified change-points were expected to have signal changes. Keeping in mind that the proposed change-point detection and identification approach is nonparametric with no

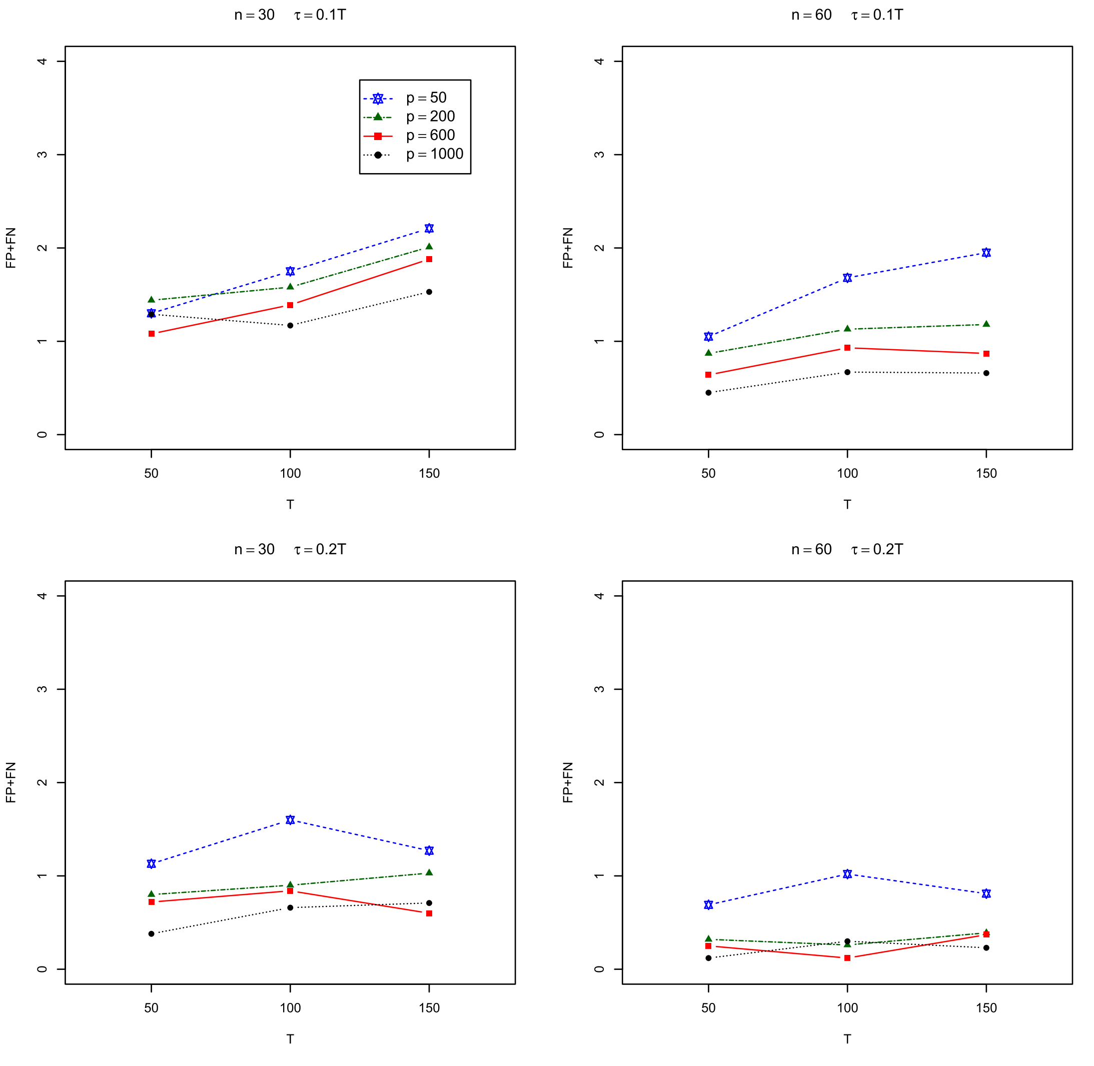


FIGURE 3 The average FP+FN subject to different combination of T, n, and p. Upper panel: The change point is located at $\tau = 0.1T$. Lower panel: The change point is located at $\tau = 0.2T$ [Colour figure can be viewed at wileyonlinelibrary.com]

attempt to model neural activation, we have demonstrated that it has satisfactory performance for the fMRI data analysis.

To confirm that the parahippocampal region is selectively activated by the scenes over the objects, we compared the brain region activated by the scene images and with that activated by the object images. To do this, we let $X_{i\tau j}$ be the jth component (voxel) of the random vector $X_{i\tau}$ for ith subject at the change-point τ where $i = 1, \dots, 14$, $\tau = 1, \dots, 59$, and $j = 1, \dots, 33,866$. Similarly, let $X_{i\tau+1j}$ be the jth component of the random vector $X_{i\tau+1}$ after the change-point τ. For each voxel ($j = 1, \dots, 33,866$), we computed the difference between two sample means $\overline{X}_{\tau j}$ and $\overline{X}_{\tau+1j}$ and then conducted paired t-test for the significance of the mean difference before and after the change-point. Based on obtained p-values, we allocated the activated brain regions composed of all significant voxels after controlling the false discovery rate at 0.01 (see Storey, 2003). The

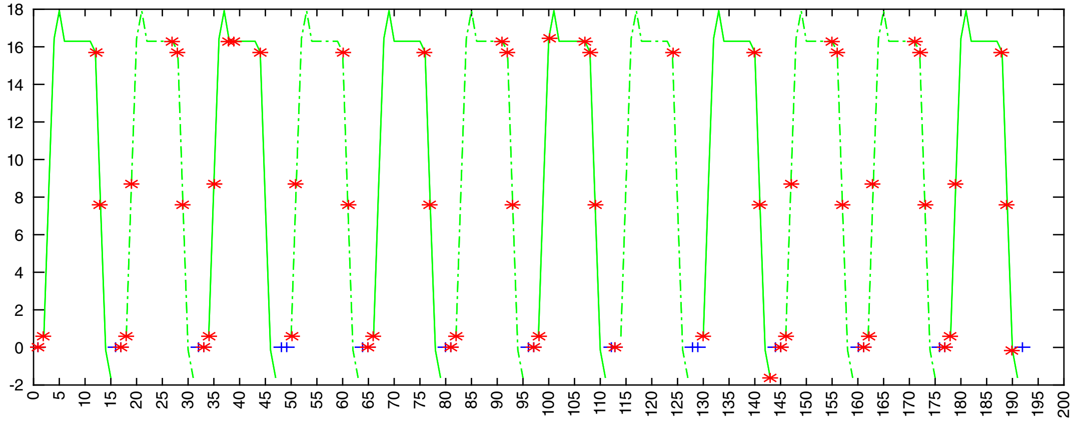


FIGURE 4 The illustration of change-points identified by the proposed method. The green solid and dash curves, respectively, represent the expected blood oxygen level-dependent (BOLD) responses to the scene and objective images. The x-values and y-values of the red stars marked on the curves, are the identified change-points and the corresponding BOLD responses. The blue plus signs represent the locations where subjects rest such that the BOLD responses are zero. Out of the 59 identified change-points, 58 are expected to have signal changes. [Colour figure can be viewed at wileyonlinelibrary.com]

results showed that the activated brain regions were quite similar across the same type of images, but significantly different between scene and object images. More specifically, the brain region activated by the scene images was located at both the visual cortex area and the parahippocampal area, whereas the region activated by the object images was only located at the visual cortex area. Our findings are consistent with the results in Henderson et al. (2011). For illustration purpose, we only included pictures at two change-points in Figure 5.

6 | DISCUSSION

Motivated by applications such as the fMRI studies, we consider the problem of testing the homogeneity of high-dimensional mean vectors. The data structure we consider is characterized dimension p which is large, the series length T which is moderate or large, and the sample size n which is small or moderate. The main contribution of our article is to develop a complete change point detection and identification procedure for such data. The existing procedures consider only the case of $n = 1$. The second contribution is to develop a MANOVA test, which is applicable to temporally dependent data. The existing procedures for testing the equality of high-dimensional means assume temporal independence. In both cases, we propose new test statistics and establish their asymptotic distributions under mild conditions. In the change point problem, when the null hypothesis is rejected, we further propose a procedure that identifies the change-points with probability converging to one. The rate of consistency of the change-point estimator is also established. The rate explicitly displays the interplay of the three crucial sizes, p, T, and n. The proposed methods have also been generalized to a mixture model to allow heterogeneity among subjects. Although the current article is motivated by fMRI data analysis, our methods can be also applied to other high-dimensional longitudinal data with the characteristics formulated above.

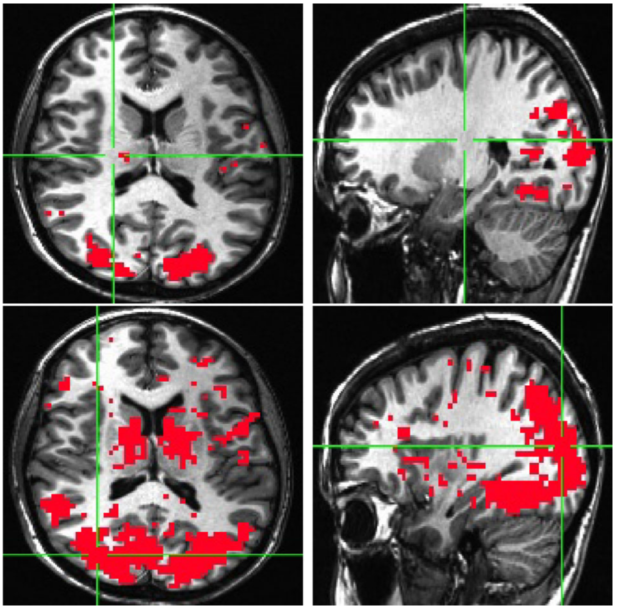


FIGURE 5 Upper panels: the activated brain regions at the fifth identified change-point (17th time point), where the object images were presented. Most of the significant changes (red areas) occurred at visual cortex areas. Lower panels: the activated brain regions at the 57th change-point (188th time point), where the scene images were presented. Most of the significant changes (red areas) occurred at both visual cortex and parahippocampal areas [Colour figure can be viewed at wileyonlinelibrary.com]

ACKNOWLEDGEMENTS

The authors thank the Editor, Professor Håkon K. Gjessing, an Associate Editor, and two referees for their comments, which helped to improve the article. The research of Zhong was partially supported by NSF grant FRG-1462156, of Li by NSF grant DMS-1916239, and of Kokoszka by NSF grant DMS-1914882.

ORCID

Piotr Kokoszka https://orcid.org/0000-0001-9979-6536

REFERENCES

Ashby, F. G. (2011). *Statistical analysis of fMRI Data*. Cambridge MA: MIT press.

Aston, J. A. D., & Kirch, C. (2012a). Detecting and estimating epidemic changes in dependent functional data. *Journal of Multivariate Analysis*, *109*, 204–220.

Aston, J. A. D., & Kirch, C. (2012b). Evaluating stationarity via change–point alternatives with applications to fMRI data. *The Annals of Applied Statistics*, *6*, 1906–1948.

Bai, Z., & Saranadasa, H. (1996). Effect of high dimension: By an example of two sample problem. *Statistica Sinica*, *6*, 311–329.

Billingsley, P. (1999). *Convergence of probability measures*. New York, NY: Wiley.

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

Chen, S. X., & Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, *38*, 808–835.

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

Csörgő, M., & Horváth, L. (1997). *Limit theorems in change-point analysis*. Hoboken, NJ: Wiley.

Dempster, A. P. (1958). A high dimensional two sample significance test. *The Annals of Mathematical Statistics*, *29*, 995–1010.

Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics*, *16*, 41–50.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.

Fujikoshi, Y., Ulyanov, V. V., & Shimizu, R. (2010). *Multivariate statistics: High-dimensional and large-sample approximations*. New York, NY: Wiley.

Hall, P., & Heyde, C. (1980). *Martingale limit theory and applications*. New York, NY: Academic Press.

Henderson, J. M., Larson, C. L., & Zhu, D. C. (2007). Cortical activation to indoor versus outdoor scenes: An fMRI study. *Experimental Brain Research*, *179*, 75–84.

Henderson, J. M., Zhu, D. C., & Larson, C. L. (2011). Functions of parahippocampal place area and retrosplenial cortex in real-world scene analysis: An fMRI study. *Visual Cognition*, *19*, 910–927.

Hu, J., Bai, Z., Wang, C., & Wang, W. (2017). On testing the equality of high dimensional mean vectors with unequal covariance matrices. *Annals of the Institute of Statistical Mathematics*, *69*, 365–387.

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

Pešta, M., & Wendler, M. (2019). Nuisance-parameter-free changepoint detection in non-stationary series. *Test*, *2019*, 1–30.

Peštová, B., & Pešta, M. (2015). Testing structural changes in panel data with small fixed panel size and bootstrap. *Metrika*, *78*, 665–689.

Schott, J. R. (2007). Some High-dimensional Tests for a one-way MANOVA. *Journal of Multivariate Analysis*, *98*, 1825–1839.

Srivastava, M. S., & Kubokawa, T. (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *Journal of Multivariate Analysis*, *115*, 204–216.

Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *The Annals of Statistics*, *31*, 2013–2035.

Venkatraman, E. (1992) Consistency results in multiple change-points problems. *Stanford University Technical Report*, *24*.

Vostrikova, L. J. (1981). Detecting "disorder" in multidimensional random processes. *Soviet Mathematics: Doklady*, *24*, 55–59.

Wang, L., Peng, B., & Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association*, *110*, 1658–1669.

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

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, *24*, 471–494.

Zhong, P.-S., Lan, W., Song, P. X. K., & Tsai, C.-L. (2017). Tests for covariance structures with high-dimensional repeated measurements. *The Annals of Statistics*, *45*, 1185–1213.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhong P-S, Li J, Kokoszka P. Multivariate analysis of variance and change points estimation for high-dimensional longitudinal data. *Scand J Statist*. 2020;1–31. https://doi.org/10.1111/sjos.12460

APPENDIX PROOFS OF THE THEOREMS OF PREVIOUS SECTIONS

In this Appendix, we provide proofs to the theorems and propositions in the article. Assume $\mu_t = 0$ in (2) and (3). For any squared $m \times m$ matrix A and B, the following results commonly used in the Appendix can be derived: $\mathrm{E}(X_{is}'AX_{it}) = \mathrm{tr}(\Gamma_s'A\Gamma_t)$, and

$$\mathrm{E}(X_{is}'AX_{it}X_{is^*}'BX_{it^*}) = \mathrm{tr}(\Gamma_s'A\Gamma_t)\mathrm{tr}(\Gamma_{s^*}'B\Gamma_{t^*}) + \mathrm{tr}(\Gamma_s'A\Gamma_t\Gamma_{s^*}'B\Gamma_{t^*}) + \mathrm{tr}(\Gamma_s'A\Gamma_t\Gamma_{t^*}'B'\Gamma_{s^*}) + (3+\Delta)\mathrm{tr}(\Gamma_s'A\Gamma_t \circ \Gamma_{s^*}'B\Gamma_{t^*}), \tag{A1}$$

where $A \circ B$ is the Hadamard product of A and B.

Proof of Theorem 1. Theorem 1 can be established by the martingale central limit theorem. Toward this end, we first construct a martingale difference sequence. If we define $Y_{is_a} = X_{is_a} - \mu_{s_a}$, then $\widehat{M}_t - M_t = \sum_{i=1}^n M_{ti}$, where

$$M_{ti} = \frac{2}{n(n-1)h(t)} \sum_{j=1}^{i-1} \left\{ \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \sum_{a,b\in\{1,2\}} (-1)^{|a-b|} Y_{is_a}' Y_{js_b} \right\} + \frac{2}{nh(t)} \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \sum_{a,b\in\{1,2\}} (-1)^{|a-b|} \mu_{s_a}' Y_{is_b}.$$

Let $\{\mathcal{F}_i, 1 \le i \le n\}$ be σ-fields generated by $\sigma\{\mathbb{Y}_1, \dots, \mathbb{Y}_i\}$ where $\mathbb{Y}_i = \{Y_{i1}, \dots, Y_{iT}\}'$. Then it can be shown that $\mathrm{E}(M_{tk}|\mathcal{F}_{k-1}) = 0$ for $k = 1, \dots, n$. Therefore, $\{M_{ti}, 1 \le i \le n\}$ is a martingale difference sequence with respect to σ-fields $\{\mathcal{F}_i, 1 \le i \le n\}$.

Based on Lemmas 1 and 2 proven in the supplementary material, Theorem 1 can be proven using the martingale central limit theorem (see Hall & Heyde, 1980). ■

Proof of Theorem 2. Note that the estimator $\mathrm{tr}(\widehat{\Xi_{r_as_c}\Xi_{r_bs_d}'})$ in (8) is invariant by transforming X_{it} to $X_{it} - \mu_t$ where $t = 1, \dots, \tau$. With loss of generality, we assume that $\mu_1 = \mu_2 = \ldots = \mu_T = 0$. First,

$$\mathrm{E}\left\{ \mathrm{tr}(\widehat{\Xi_{r_as_c}\Xi_{r_bs_d}'}) \right\} = \mathrm{E}(X_{ir_a}'X_{jr_b}X_{is_c}'X_{js_d}) - \mathrm{E}(X_{ir_a}'X_{jr_b}X_{is_c}'X_{ks_d}) - \mathrm{E}(X_{ir_a}'X_{jr_b}X_{ks_c}'X_{js_d}) + \mathrm{E}(X_{ir_a}'X_{jr_b}X_{ks_c}'X_{ls_d}) = \mathrm{tr}(\Xi_{r_as_c}\Xi_{r_bs_d}').$$

This shows that $\mathrm{E}(\widehat{\sigma}_{nt,0}^2) = \sigma_{nt,0}^2$. Therefore, to prove Theorem 2, we only need to show that $\mathrm{Var}(\widehat{\sigma}_{nt,0}^2)/\sigma_{nt,0}^4 \to 0$.

For convenience, we denote the summation $\sum_{r_1=1}^t \sum_{r_2=t+1}^T \sum_{s_1=1}^t \sum_{s_2=t+1}^T$ by $\sum_{r_1,r_2,s_1,s_2}$. Define the right-hand side of "=" in (8) as $B_1 + B_2 + B_3 + B_4$, and accordingly,

$$\widehat{\sigma}_{nt,0}^2 = \frac{2}{h^2(t)n(n-1)} \sum_{r_1,r_2,s_1,s_2} \sum_{a,b,c,d\in\{1,2\}} (-1)^{|a-b|+|c-d|}(B_1 + B_2 + B_3 + B_4) \equiv \widehat{\sigma}_{nt,0}^{2(1)} + \widehat{\sigma}_{nt,0}^{2(2)} + \widehat{\sigma}_{nt,0}^{2(3)} + \widehat{\sigma}_{nt,0}^{2(4)}.$$

Therefore, we only need to show that $\mathrm{Var}(\widehat{\sigma}_{nt,0}^{2(i)})/\sigma_{nt,0}^4 \to 0$ for $i = 1, 2, 3$, and 4 respectively. Toward this end, we first show that $\mathrm{Var}(\widehat{\sigma}_{nt,0}^{2(1)})/\sigma_{nt,0}^4 \to 0$ as follows.

$$\mathrm{Var}(\widehat{\sigma}_{nt,0}^{2(1)}) = \frac{4}{h^4(t)n^4(n-1)^4} \mathrm{Var}\left\{ \sum_{r_1,r_2,s_1,s_2} \sum_{a,b,c,d\in\{1,2\}} (-1)^{|a-b|+|c-d|} \sum_{i\neq j}^{n} X_{ir_a}'X_{jr_b}X_{is_c}'X_{js_d} \right\}$$

$$= \frac{4}{h^4(t)n^4(n-1)^4} \sum \left\{ \sum_{i\neq j,k\neq l}^{n} \mathrm{E}(X'_{ir_a} X_{jr_b} X'_{is_c} X_{js_d} X'_{kr^*_{a^*}} X_{lr^*_{b^*}} X'_{ks^*_{c^*}} X_{ls^*_{d^*}}) \right.$$
$$\left. - n^2(n-1)^2 \mathrm{tr}(\Gamma'_{r_a}\Gamma_{r_b}\Gamma'_{s_c}\Gamma_{s_d})\mathrm{tr}(\Gamma'_{r^*_{a^*}}\Gamma_{r^*_{b^*}}\Gamma'_{s^*_{c^*}}\Gamma_{s^*_{d^*}}) \right\}, \tag{A2}$$

where $\sum$ represents $\sum_{r_1,r_2,s_1,s_2}\sum_{a,b,c,d\in\{1,2\}}\sum_{r_1^*,r_2^*,s_1^*,s_2}\sum_{a^*,b^*,c^*,d^*\in\{1,2\}}$.

Now we evaluate $\sum_{i\neq j,k\neq l}^{n} \mathrm{E}(X'_{ir_a} X_{jr_b} X'_{is_c} X_{js_d} X'_{kr^*_{a^*}} X_{lr^*_{b^*}} X'_{ks^*_{c^*}} X_{ls^*_{d^*}})$ with respect to different cases in the following. First, if all indices are distinct, that is, $i\neq j\neq k\neq l$. Using (A1), we have

$$\sum_{i\neq j,k\neq l}^{n} \mathrm{E}(X'_{ir_a} X_{jr_b} X'_{is_c} X_{js_d} X'_{kr^*_{a^*}} X_{lr^*_{b^*}} X'_{ks^*_{c^*}} X_{ls^*_{d^*}}) \asymp n^4 \mathrm{tr}(\Gamma'_{r_a}\Gamma_{r_b}\Gamma'_{s_d}\Gamma_{s_c})\mathrm{tr}(\Gamma'_{r^*_{a^*}}\Gamma_{r^*_{b^*}}\Gamma'_{s^*_{d^*}}\Gamma_{s^*_{c^*}}).$$

Next, if $(i=k)\neq j\neq l$, then by (A1),

$$\sum_{i\neq j,k\neq l}^{n} \mathrm{E}(X'_{ir_a} X_{jr_b} X'_{is_c} X_{js_d} X'_{kr^*_{a^*}} X_{lr^*_{b^*}} X'_{ks^*_{c^*}} X_{ls^*_{d^*}})$$
$$\asymp n^3 \left\{ (3+\Delta)\mathrm{tr}(\Gamma'_{r_a}\Gamma_{r_b}\Gamma'_{s_d}\Gamma_{s_c}\circ\Gamma'_{r^*_{a^*}}\Gamma_{r^*_{b^*}}\Gamma'_{s^*_{d^*}}\Gamma_{s^*_{c^*}}) + \mathrm{tr}(\Gamma'_{r_a}\Gamma_{r_b}\Gamma'_{s_d}\Gamma_{s_c})\mathrm{tr}(\Gamma'_{r^*_{a^*}}\Gamma_{r^*_{b^*}}\Gamma'_{s^*_{d^*}}\Gamma_{s^*_{c^*}}) \right.$$
$$\left. + \mathrm{tr}(\Gamma'_{r_a}\Gamma_{r_b}\Gamma'_{s_d}\Gamma_{s_c}\Gamma'_{r^*_{a^*}}\Gamma_{r^*_{b^*}}\Gamma'_{s^*_{d^*}}\Gamma_{s^*_{c^*}}) + \mathrm{tr}(\Gamma'_{r_a}\Gamma_{r_b}\Gamma'_{s_d}\Gamma_{s_c}\Gamma'_{s^*_{c^*}}\Gamma_{s^*_{d^*}}\Gamma'_{r^*_{b^*}}\Gamma_{r^*_{a^*}}) \right\},$$

which is equal to other cases $(j=k)\neq i\neq l$, $(i=l)\neq j\neq k$ and $(j=l)\neq i\neq k$. Finally, we consider the cases $(i=k)\neq(j=l)$ and $(i=l)\neq(j=k)$. For the case $(i=k)\neq(j=l)$,

$$\sum_{i\neq j,k\neq l}^{n} \mathrm{E}(X'_{ir_a} X_{jr_b} X'_{is_c} X_{js_d} X'_{kr^*_{a^*}} X_{lr^*_{b^*}} X'_{ks^*_{c^*}} X_{ls^*_{d^*}})$$
$$\asymp n^2 \left\{ 3\mathrm{tr}(\Gamma'_{r_a}\Gamma_{r_b}\Gamma'_{s_d}\Gamma_{s_c})\mathrm{tr}(\Gamma'_{r^*_{a^*}}\Gamma_{r^*_{b^*}}\Gamma'_{s^*_{d^*}}\Gamma_{s^*_{c^*}}) + 3Q_1 + (3+\Delta)Q_2 \right.$$
$$+ 3(3+\Delta)\mathrm{tr}(\Gamma'_{s_d}\Gamma_{s_c}\Gamma'_{r_a}\Gamma_{r_b}\circ\Gamma'_{s^*_{d^*}}\Gamma_{s^*_{c^*}}\Gamma'_{r^*_{a^*}}\Gamma_{r^*_{b^*}})$$
$$\left. + (3+\Delta)^2 \sum_{\alpha\beta} (\Gamma'_{r_a}\Gamma_{r_b})_{\alpha\beta}(\Gamma'_{s_d}\Gamma_{s_c})_{\beta\alpha}(\Gamma'_{r^*_{a^*}}\Gamma_{r^*_{b^*}})_{\alpha\beta}(\Gamma'_{s_{d^*}}\Gamma_{s_{c^*}})_{\beta\alpha} \right\},$$

where $Q_1 = \mathrm{tr}(\Gamma'_{s_d}\Gamma_{s_c}\Gamma'_{r_a}\Gamma_{r_b}\Gamma'_{s^*_{d^*}}\Gamma_{s^*_{c^*}}\Gamma'_{r^*_{a^*}}\Gamma_{r^*_{b^*}}) + \mathrm{tr}(\Gamma'_{s_d}\Gamma_{s_c}\Gamma'_{r_a}\Gamma_{r_b}\Gamma'_{r^*_{b^*}}\Gamma_{r^*_{a^*}}\Gamma'_{s^*_{c^*}}\Gamma_{s^*_{d^*}})$ and $Q_2 = \mathrm{tr}(\Gamma'_{r_a}\Gamma_{r_b}\Gamma'_{s_d}\Gamma_{s_c}\circ\Gamma'_{r^*_{a^*}}\Gamma_{r^*_{b^*}}\Gamma'_{s^*_{d^*}}\Gamma_{s^*_{c^*}}) + \mathrm{tr}(\Gamma'_{r_a}\Gamma_{r_b}\Gamma'_{r_{b^*}}\Gamma_{r_{a^*}}\circ\Gamma'_{s_d}\Gamma_{s_c}\Gamma'_{s^*_{c^*}}\Gamma_{s^*_{d^*}}) + \mathrm{tr}(\Gamma'_{r_a}\Gamma_{r_b}\Gamma'_{s_{d^*}}\Gamma_{s_{c^*}}\circ\Gamma'_{r_{a^*}}\Gamma_{r_{b^*}}\Gamma'_{s_d}\Gamma_{s_c})$. It can be shown that the case $(j=l)\neq i\neq k$ is the as the case $(i=k)\neq(j=l)$.

Plugging all the above results into (A2), we have

$$\mathrm{Var}(\hat{\sigma}^{2(1)}_{nt,0}) \asymp h^{-4}(t)n^{-5}\sum \mathrm{tr}(\Gamma'_{r_b}\Gamma_{r_a}\Gamma'_{s_c}\Gamma_{s_d}\Gamma'_{s^*_{d^*}}\Gamma_{s^*_{c^*}}\Gamma'_{r^*_{a^*}}\Gamma_{r^*_{b^*}}) + h^{-4}(t)n^{-6}\mathrm{tr}(A^2_{0t}).$$

Following the same procedure, it can be also shown that $\mathrm{Var}(\hat{\sigma}^{2(j)}_{nt,0}) = o\{\mathrm{Var}(\hat{\sigma}^{2(1)}_{nt,0})\}$ for $j=2,3$, and 4. Then, using condition (C1), we have $\mathrm{Var}(\hat{\sigma}^{2(j)}_{nt,0})/\sigma^4_{nt,0}\to 0$ for $j=1,2,3$, and 4. This completes the proof of Theorem 2. ∎

Proof of Theorem 3. First, we derive $\mathrm{Cov}(\widehat{M}_u, \widehat{M}_v)$ for $u, v \in \{1, \dots, T-1\}$ under H_0 of (1). Without loss of generality, we assume that $\mu_1 = \mu_2 = \dots = \mu_T = 0$. Recall that

$$\widehat{M}_u = \frac{1}{h(u)n(n-1)} \sum_{s_1=1}^{u} \sum_{s_2=u+1}^{T} \left\{ \sum_{i \neq j}^{n} X'_{is_1} X_{js_1} + \sum_{i \neq j}^{n} X'_{is_2} X_{js_2} - 2 \sum_{i \neq j}^{n} X'_{is_1} X_{js_2} \right\},$$
$$\widehat{M}_v = \frac{1}{h(v)n(n-1)} \sum_{s_1=1}^{v} \sum_{s_2=v+1}^{T} \left\{ \sum_{i \neq j}^{n} X'_{is_1} X_{js_1} + \sum_{i \neq j}^{n} X'_{is_2} X_{js_2} - 2 \sum_{i \neq j}^{n} X'_{is_1} X_{js_2} \right\}.$$

Following similar derivations for the variance of $\widehat{M}_t$ in the proof of Proposition 1 in the supplementary material, we can derive that

$$\mathrm{Cov}(\widehat{M}_u, \widehat{M}_v) = \frac{2}{h(u)h(v)n(n-1)} \sum_{r_1=1}^{u} \sum_{r_2=u+1}^{T} \sum_{s_1=1}^{v} \sum_{s_2=v+1}^{T} \times \sum_{a,b,c,d \in \{1,2\}} (-1)^{|a-b|+|c-d|} \mathrm{tr}(\Xi_{r_a s_c} \Xi'_{r_b s_d}).$$

Next, we show that $\{\widehat{M}_t\}_{t=1}^{T-1}$ follow a joint multivariate normal distribution when T is fixed. According to the Cramer-word device, we only need to show that for any nonzero constant vector $a = (a_1, \dots, a_{T-1})'$, $\sum_{t=1}^{T-1} a_t \widehat{M}_t$ is asymptotically normal under H_0 of (1). Toward this end, we note that $\mathrm{Var}(\sum_{t=1}^{T-1} a_t \widehat{M}_t) = \sum_{u=1}^{T-1} \sum_{v=1}^{T-1} a_u a_v \mathrm{Cov}(\widehat{M}_u, \widehat{M}_v)$. Then we only need to show that $\sum_{t=1}^{T-1} a_t \widehat{M}_t / \sqrt{\mathrm{Var}(\sum_{t=1}^{T-1} a_t \widehat{M}_t)} \xrightarrow{d} N(0,1)$, which can be proved by the martingale central limit theorem. Since the proof is very similar to that of Theorem 1, we omit it. With the joint normality of $\{\widehat{M}_t\}_{t=1}^{T-1}$, the distribution of $\widehat{\mathscr{M}} \to \max_{1 \le t \le T-1} Z_t$ can be established by the continuous mapping theorem.

To establish the asymptotic distribution of $\widehat{\mathscr{M}}$ for T diverging case, we need to show that under H_0, $\max_{1 \le t \le T-1} \sigma_{nt}^{-1} \widehat{M}_t$ converges to $\max_{1 \le t \le T-1} Z_t$, where Z_t is a Gaussian process with mean 0 and covariance Σ_Z. To this end, we need to show (i) the joint asymptotic normality of $(\sigma_{nt_1}^{-1} \widehat{M}_{t_1}, \dots, \sigma_{nt_d}^{-1} \widehat{M}_{t_d})'$ for $t_1 < t_2 < \dots < t_d$. (ii) the tightness of $\max_{1 \le t \le T-1} \sigma_{nt}^{-1} \widehat{M}_t$. The proof of (i) is the similar to the proof of the joint asymptotic normality under finite T case. We need to prove (ii).

To prove (ii), let $W_n(s_1, s_2) = \sum_{a,b \in \{1,2\}} (-1)^{|a-b|} \{n(n-1)\}^{-1} \sum_{i \neq j} X'_{is_a} X_{js_b}$ and the first-order projection as $W_{n1}(s_1) = \{n(n-1)\}^{-1} \sum_{i \neq j} X'_{is_1} X_{js_1}$. Then we have the following Hoeffding-type decomposition for $\widehat{M}_t$,

$$\widehat{M}_t = \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} g_n(s_1, s_2) + \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \{W_{n1}(s_1) + W_{n2}(s_2)\} := \widehat{M}_t^{(1)} + \widehat{M}_t^{(2)},$$

where $g_n(s_1, s_2) = W_n(s_1, s_2) - W_{n1}(s_1) - W_{n2}(s_2)$. The covariance between $\widehat{M}_t^{(1)}$ and $\widehat{M}_t^{(2)}$ is 0. First, we compute the variances of $\widehat{M}_t^{(2)}$ under the the null hypothesis H_0. We first write $\widehat{M}_t^{(2)} = (T-t) \sum_{s_1=1}^{t} W_{n1}(s_1) + t \sum_{s_2=t+1}^{T} W_{n2}(s_2) := \widehat{M}_t^{(21)} + \widehat{M}_t^{(22)}$. Then we have

$$\mathrm{Var}(\widehat{M}_t^{(21)}) = \frac{2(T-t)^2}{n(n-1)} \sum_{s_1=1}^{t} \sum_{r_1=1}^{t} \mathrm{tr}(\Xi_{s_1 r_1} \Xi'_{s_1 r_1})$$

Similarly, we have

$$\mathrm{Var}(\widehat{M}_t^{(22)}) = \frac{2t^2}{n(n-1)} \sum_{s_2=t+1}^{T} \sum_{r_2=t+1}^{T} \mathrm{tr}(\Xi_{s_2 r_2} \Xi'_{s_2 r_2}).$$

In addition, the covariance between $\widehat{M}_t^{(21)}$ and $\widehat{M}_t^{(22)}$ is

$$\mathrm{Cov}(\widehat{M}_t^{(21)}, \widehat{M}_t^{(22)}) = \frac{2t(T-t)}{n(n-1)} \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \mathrm{tr}(\Xi_{s_1 s_2} \Xi'_{s_1 s_2}).$$

In summary, the variance for $\widehat{M}_t^{(2)}$ is

$$\mathrm{Var}(\widehat{M}_t^{(2)}) = \frac{2}{n(n-1)} \sum_{s_1, r_1=1}^{t} \sum_{s_2, r_2=t+1}^{T} \left\{ \mathrm{tr}(\Xi_{s_1 r_1} \Xi'_{s_1 r_1}) + \mathrm{tr}(\Xi_{s_2 r_2} \Xi'_{s_2 r_2}) + 2\mathrm{tr}(\Xi_{s_1 s_2} \Xi'_{s_1 s_2}) \right\}.$$

Moreover, we have

$$\begin{aligned} \mathrm{Var}(\widehat{M}_t^{(1)}) = & \frac{4}{n(n-1)} \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \left\{ \mathrm{tr}(\Sigma_{s_1} \Sigma_{s_2}) + \mathrm{tr}(\Xi_{s_2 s_1} \Xi_{s_2 s_1}) \right\} \\ & + \frac{4}{n(n-1)} \sum_{s_1 \neq r_1=1}^{t} \sum_{s_2 \neq r_2=t+1}^{T} \left\{ \mathrm{tr}(\Xi_{s_1 r_1} \Xi'_{s_2 r_2}) + \mathrm{tr}(\Xi_{s_2 r_1} \Xi'_{s_1 r_2}) \right\}. \end{aligned}$$

According to the condition (C2), $\mathrm{tr}(\Xi_{s_1 r_1} \Xi'_{s_1 r_1}) \asymp \phi(|s_1 - r_1|)\mathrm{tr}(\Sigma_{s_1} \Sigma_{r_1})$ and $\sum_{k=1}^{T} \phi^{1/2}(k) < \infty$. Under the null hypothesis H_0, we have

$$\begin{aligned} \mathrm{Var}(\widehat{M}_t^{(2)}) & \asymp \frac{2\mathrm{tr}(\Sigma^2)}{n(n-1)} \sum_{s_1, r_1=1}^{t} \sum_{s_2, r_2=t+1}^{T} \{\phi(|s_1 - r_1|) + \phi(|s_2 - r_2|) + 2\phi(|s_1 - s_2|)\} \\ & \asymp \frac{2\mathrm{tr}(\Sigma^2)}{n(n-1)} \{(T-t)^2 t + t^2 (T-t)\}. \end{aligned}$$

On the other hand, we notice that the first term of $\mathrm{Var}(\widehat{M}_t^{(1)})$ has the same order as $t(T-t)\mathrm{tr}(\Sigma^2)/\{n(n-1)\}$. Using the Cauchy–Schwarz inequality and under H_0, we have

$$\mathrm{tr}^2(\Xi_{s_1 r_1} \Xi'_{s_2 r_2}) \leq \mathrm{tr}(\Xi_{s_1 r_1} \Xi'_{s_1 r_1}) \mathrm{tr}(\Xi_{s_2 r_2} \Xi'_{s_2 r_2}) \asymp \phi(|s_1 - r_1|)\phi(|s_2 - r_2|)\mathrm{tr}^2(\Sigma^2).$$

Therefore, using the condition $\sum_{k=1}^{T} \phi^{1/2}(k) < \infty$, the second term in $\mathrm{Var}(\widehat{M}_t^{(1)})$ is also of order $t(T-t)\mathrm{tr}(\Sigma^2)/\{n(n-1)\}$. In summary, $\widehat{M}_t^{(1)}$ is a small order of $\widehat{M}_t^{(2)}$. This also implies that $\sigma_{nt}^2 = \mathrm{Var}(\widehat{M}_t^{(2)})\{1 + o(1)\}$.

Consider $t = [Tv]$ for $v = j/T \in (0,1)$ with $j = 1, \ldots, T-1$. Based on the above results, to show the tightness of $\max_{1 \leq t \leq T-1} \sigma_{nt}^{-1} \widehat{M}_t$ is equivalent to show the tightness of $G_n(v)$ where

$$G_n(v) = T^{-3/2} n^{-1} \mathrm{tr}^{-1/2}(\Sigma^2)(\widehat{M}_{[Tv]}^{(1)} + \widehat{M}_{[Tv]}^{(2)}) := G_n^{(1)}(v) + G_n^{(2)}(v).$$

We first show the tightness of $G_n^{(1)}(v)$. To this end, we first note that, for $1 > \eta > v > 0$,

$$E\left\{|G_n^{(1)}(v) - G_n^{(1)}(\eta)|^2\right\} = \frac{1}{T^3 n^2 \mathrm{tr}(\Sigma^2)} E\left\{\left|\sum_{s_1=1}^{[Tv]} \sum_{s_2=[Tv]+1}^{[T\eta]} g_n(s_1, s_2) - \sum_{s_1=[Tv]+1}^{[T\eta]} \sum_{s_2=[T\eta]+1}^{T} g_n(s_1, s_2)\right|^2\right\}$$
$$\leq CT^{-3}\{[Tv]([T\eta] - [Tv]) + (T - [T\eta])([T\eta] - [Tv])\} \leq C(\eta - v)/T.$$

Applying the above inequality with $v = k/T$ and $\eta = m/T$ for $0 \leq k \leq m < T$ for integers k, m, and T and using Chebyshev's inequality, we have, for any $\epsilon > 0$,

$$P\left(\left|G_n^{(1)}(k/T) - G_n^{(1)}(m/T)\right| \geq \epsilon\right) \leq E\left\{|G_n^{(1)}(k/T) - G_n^{(1)}(m/T)|^2\right\}/\epsilon^2$$
$$\leq C(m-k)/(\epsilon T)^2 \leq (C/\epsilon^2)(m-k)^{1+\alpha}/T^{2-\alpha},$$

where $0 < \alpha < 1/2$. Now if we define $\xi_i = G_n^{(1)}(i/T) - G_n^{(1)}((i-1)/T)$ for $i = 1, \ldots, T-1$. Then $G_n^{(1)}(i/T)$ is equal to the partial sum of ξ_i, namely $S_i = \xi_1 + \ldots + \xi_i = G_n^{(1)}(i/T)$. Here $S_0 = 0$. Then we have

$$P(|S_m - S_k| \geq \epsilon) \leq (1/\epsilon^2)\{C^{1/(1+\alpha)}(m-k)/T^{(2-\alpha)/(1+\alpha)}\}^{1+\alpha}.$$

Then using theorem 10.2 in Billingsley (1999), we conclude the following

$$P\left(\max_{1\leq i\leq T} |S_i| \geq \epsilon\right) \leq (KC/\epsilon^2)\{T/T^{(2-\alpha)/(1+\alpha)}\}^{1+\alpha} \leq (KC/\epsilon^2)T^{-1+2\alpha}.$$

The right-hand side of the above inequality goes to 0 as $T \to \infty$ because $\alpha < 1/2$. Based on the relationship between S_i and $G_n^{(1)}(i/T)$, we have shown the tightness of $G_n^{(1)}(v)$.

Next, we consider the tightness of $G_n^{(2)}(v)$. Recall that

$$G_n^{(2)}(v) = T^{-3/2} n^{-1} \mathrm{tr}^{-1/2}(\Sigma^2) \sum_{s_1=1}^{[Tv]} \sum_{s_2=[Tv]+1}^{T} \{W_{n1}(s_1) + W_{n2}(s_2)\}$$
$$= T^{-3/2} n^{-1} \mathrm{tr}^{-1/2}(\Sigma^2)(T - [Tv]) \sum_{s_1=1}^{[Tv]} W_{n1}(s_1)$$
$$+ T^{-3/2} n^{-1} \mathrm{tr}^{-1/2}(\Sigma^2)[Tv] \sum_{s_2=[Tv]+1}^{T} W_{n2}(s_2) := G_n^{(21)}(v) + G_n^{(22)}(v).$$

It is enough to show the tightness of $G_n^{(21)}(v)$, since the tightness of $G_n^{(22)}(v)$ is similar. Let $h(i,j) = T^{-1/2} \sum_{s_1=[Tv]+1}^{[T\eta]} (X_{is_1} - \mu)'(X_{js_1} - \mu)$. Then, we have the following

$$G_n^{(21)}(\eta) - G_n^{(21)}(v) = T^{-1/2} n^{-1} \mathrm{tr}^{-1/2}(\Sigma^2) \sum_{s_1=[Tv]+1}^{[T\eta]} \frac{1}{n(n-1)} \sum_{i\neq j} X_{is_1}' X_{js_1}$$
$$= \frac{1}{\sqrt{n(n-1)}\mathrm{tr}(\Sigma^2)} \sum_{i\neq j} h(i,j).$$

First, note that

$$\{G_n^{(21)}(\eta) - G_n^{(21)}(\nu)\}^2 = \frac{2}{n(n-1)\mathrm{tr}(\Sigma^2)} \sum_{i \neq j} h^2(i,j) + \frac{4}{n(n-1)\mathrm{tr}(\Sigma^2)} \sum_{i \neq j \neq k} h(i,j)h(i,k) + \frac{1}{n(n-1)\mathrm{tr}(\Sigma^2)} \sum_{i \neq j \neq k \neq l} h(i,j)h(k,l).$$

Then, we have the following

$$\begin{aligned} E[\{G_n^{(21)}(\eta) - G_n^{(21)}(\nu)\}^4] \leq\; & E\left[\frac{8}{n^2(n-1)^2\mathrm{tr}^2(\Sigma^2)} \left\{\sum_{i \neq j} h^2(i,j)\right\}^2\right] \\ & + E\left[\frac{32}{n^2(n-1)^2\mathrm{tr}^2(\Sigma^2)} \left\{\sum_{i \neq j \neq k} h(i,j)h(i,k)\right\}^2\right] \\ & + E\left[\frac{2}{n^2(n-1)^2\mathrm{tr}^2(\Sigma^2)} \left\{\sum_{i \neq j \neq k \neq l} h(i,j)h(k,l)\right\}^2\right] \\ := \; & I_1 + I_2 + I_3. \end{aligned}$$

First, we consider I_1 in the above expression.

$$\begin{aligned} I_1 &= E\left[\frac{8}{n^2(n-1)^2\mathrm{tr}^2(\Sigma^2)} \sum_{i \neq j} \sum_{i_i \neq j_1} h^2(i,j)h^2(i_1,j_1)\right] \\ &= E\left[\frac{16}{n^2(n-1)^2\mathrm{tr}^2(\Sigma^2)} \sum_{i \neq j} h^4(i,j)\right] \\ &\quad + E\left[\frac{32}{n^2(n-1)^2\mathrm{tr}^2(\Sigma^2)} \sum_{i \neq j \neq k} h^2(i,j)h^2(i,k)\right] \\ &\quad + E\left[\frac{8}{n^2(n-1)^2\mathrm{tr}^2(\Sigma^2)} \sum_{i \neq j \neq i_i \neq j_1} h^2(i,j)h^2(i_1,j_1)\right] := I_{11} + I_{12} + I_{13}. \end{aligned}$$

We see that

$$I_{13} \asymp \frac{C}{T^2\mathrm{tr}^2(\Sigma^2)} \left\{\sum_{s_1=[T\nu]+1}^{[T\eta]} \sum_{r_1=[T\nu]+1}^{[T\eta]} \mathrm{tr}(\Xi_{s_1 r_1}\Xi'_{s_1 r_1})\right\}^2 \asymp \frac{C}{T^2}\{[T\eta]-[T\nu]\}^2.$$

After some calculation, we obtain that

$$\begin{aligned} I_{11} = \frac{C}{n(n-1)T^2\mathrm{tr}^2(\Sigma^2)} \Bigg[& \left\{\sum_{s_1=[T\nu]+1}^{[T\eta]} \sum_{r_1=[T\nu]+1}^{[T\eta]} \mathrm{tr}(\Xi_{s_1 r_1}\Xi'_{s_1 r_1})\right\}^2 \\ & + \sum_{s_1=[T\nu]+1}^{[T\eta]} \sum_{r_1=[T\nu]+1}^{[T\eta]} \sum_{u_1=[T\nu]+1}^{[T\eta]} \sum_{v_1=[T\nu]+1}^{[T\eta]} \mathrm{tr}(\Xi_{r_1 s_1}\Xi_{s_1 v_1}\Xi_{v_1 u_1}\Xi_{u_1 r_1})\Bigg] = o(I_{13}). \end{aligned}$$

Similarly, it can be shown that $I_{12} = o(I_{13})$. In summary, $I_1 \le C\{[T\eta] - [T\nu]\}^2/T^2$.

Now, we check I_2. We have the following

$$I_2 = E\left[\frac{64}{n^2(n-1)^2\mathrm{tr}^2(\Sigma^2)}\sum_{i\neq i_1\neq j\neq k} h(i,j)h(i,k)h(i_1,j)h(i_1,k)\right] + E\left[\frac{64}{n^2(n-1)^2\mathrm{tr}^2(\Sigma^2)}\sum_{i\neq j\neq k} h(i,j)h(i,k)h(i,j)h(i,k)\right] := I_{21} + I_{22}.$$

It can be seen that

$$I_{21} \le \frac{C}{\mathrm{tr}^2(\Sigma^2)}E\left[h(i,j)h(i,k)h(i_1,j)h(i_1,k)\right] = \frac{C}{T^2\mathrm{tr}^2(\Sigma^2)}\sum_{s_1,r_1,u_1,v_1}\mathrm{tr}(\Xi_{s_1r_1}\Xi_{r_1v_1}\Xi_{v_1u_1}\Xi_{u_1s_1}),$$

which is a smaller order of I_{13}. For I_{22}, we have

$$I_{22} = \frac{C}{n\mathrm{tr}^2(\Sigma^2)}E\left[h(i,j)h(i,k)h(i,j)h(i,k)\right] = \frac{C}{nT^2\mathrm{tr}^2(\Sigma^2)}\sum_{s_1,r_1,u_1,v_1}\left\{\mathrm{tr}(\Xi_{s_1u_1}\Xi'_{s_1u_1})\mathrm{tr}(\Xi_{r_1v_1}\Xi'_{r_1v_1}) + \mathrm{tr}(\Xi_{s_1u_1}\Xi_{u_1r_1}\Xi_{r_1v_1}\Xi_{v_1s_1})\right\}.$$

Therefore, I_{22} is also a smaller order of I_{13}. In summary, I_1 is a smaller order of I_{13}.

Finally, let us consider I_3. After some calculation, we have the following

$$I_3 \asymp E\left[\frac{C}{\mathrm{tr}^2(\Sigma^2)}\{h^2(i,j)h^2(k,l) + h(i,j)h(k,l)h(i,k)h(j,l)\}\right] = \frac{C}{T^2\mathrm{tr}^2(\Sigma^2)}\left\{\sum_{s_1=[T\nu]+1}^{[T\eta]}\sum_{r_1=[T\nu]+1}^{[T\eta]}\mathrm{tr}(\Xi_{s_1r_1}\Xi'_{s_1r_1})\right\}^2 + \frac{C}{T^2\mathrm{tr}^2(\Sigma^2)}\sum_{s_1,r_1,u_1,v_1}\mathrm{tr}(\Xi_{s_1r_1}\Xi_{r_1v_1}\Xi_{v_1u_1}\Xi_{u_1s_1}).$$

Now it is clear that the first term in I_3 is of the same order as I_{13} and the second term is of the same order as I_{21}. Therefore, $I_3 \le C\{[T\eta] - [T\nu]\}^2/T^2$.

Let $\nu = k/T$ and $\eta = m/T$ for $0 \le k \le m < T$ for integers k, m, and T and using the above bounds for the fourth moment of $|G_n^{(21)}(\eta) - G_n^{(21)}(\nu)|$, we have, for any $L > 0$,

$$P\left(\left|G_n^{(21)}(k/T) - G_n^{(21)}(m/T)\right| \ge L\right) \le E\left\{|G_n^{(21)}(k/T) - G_n^{(21)}(m/T)|^4\right\}/L^4 \le (C/L^4)\{(m-k)/T\}^2.$$

Applying theorem 10.2 in Billingsley (1999) again, we have

$$P(\max_{1\le i\le T}|G_n^{(21)}(i/T)| \ge L) \le KC/L^4.$$

If L is large enough, the above probability could be smaller than any $\epsilon > 0$. Therefore, $\max_{1\le i\le T}|G_n^{(21)}(i/T)|$ is tight. Similarly, we can show the tightness of $\max_{1\le i\le T}|G_n^{(22)}(i/T)|$. In summary, we have shown the tightness of $G_n^{(1)}(v)$ and $G_n^{(2)}(v)$. Hence, $G_n(v)$ is also tight. Combining (i) and (ii) together, we know that $\sigma_{nt}^{-1}\widehat{M}_t$ converges to a Gaussian process with mean 0 and covariance Σ_Z.

Finally, applying Lemma 4 in the supplementary material, we can show that the asymptotic distribution of $\max_{1\le t\le T-1}\sigma_{nt,0}^{-1}\widehat{M}_t$ is the desired Gumbel distribution. This completes the proof of Theorem 3. ∎

Proof of Theorem 4. We first obtain the covariance between $\widehat{M}_u$ and $\widehat{M}_v$ under alternatives. Let $L(s_a, s_b) = \sum_{i\neq j}^n X'_{is_a}X_{js_b}$ for $a, b \in \{1, 2\}$. Following the derivation of Proposition ‘ in the supplementary material, we note that

$$\begin{aligned}\sigma_{nuv} &= \frac{1}{n^2(n-1)^2h(u)h(v)}\sum_{s_1,s_2,r_1,r_2}\sum_{\substack{a,b,\\c,d\in\{1,2\}}}(-1)^{|a-b|+|c-d|}\mathrm{Cov}\{L(s_a,s_b),L(r_c,r_d)\}\\ &= \frac{1}{n^2(n-1)^2h(u)h(v)}\sum_{s_1,s_2,r_1,r_2}\sum_{\substack{a,b,\\c,d\in\{1,2\}}}(-1)^{|a-b|+|c-d|}\Big[n(n-1)\{\mathrm{tr}(\Xi_{s_ar_c}\Xi'_{s_br_d})\\ &\quad+\mathrm{tr}(\Xi_{s_ar_d}\Xi'_{s_br_c})\}+n(n-1)^2\{\mu'_{s_a}\Xi_{s_br_d}\mu_{r_c}+\mu'_{s_a}\Xi_{s_br_c}\mu_{r_d}+\mu'_{s_b}\Xi_{s_ar_c}\mu_{r_d}\\ &\quad+\mu'_{s_b}\Xi_{s_ar_d}\mu_{r_c}\}\Big]\\ &= \frac{2}{n(n-1)h(u)h(v)}\mathrm{tr}(A_{0u}A_{0v})+\frac{4}{nh(u)h(v)}A_{1u}A'_{1v}.\end{aligned}$$

Following the proof of Theorems 1 and 3, if T is a finite number, we can see that

$$\max_{0<u<T}\frac{\widehat{M}_u - M_u}{\sqrt{\sigma_{nuu}}}\xrightarrow{d}\max_{0<t<T}W_t^*,$$

where W_t^* is a Gaussian random vector defined in Theorem 4. Under the condition (11), we have $\sigma_{nuu} = \sigma_{nu,0}^2\{1 + o(1)\}$ and thus,

$$\max_{0<u<T}\frac{\widehat{M}_u}{\sigma_{nu,0}}\xrightarrow{d}\max_{0<t<T}\left(W_t^*+\frac{M_t}{\sigma_{nu,0}}\right).$$

If $T\to\infty$, we need to show the tightness of $\widehat{\mathscr{M}} = \max_{0<u<T}\widehat{M}_u/\sigma_{nu,0}$. To this end, we note that

$$\widehat{M}_u = \widehat{M}_{u,0} + \widehat{M}_{u,1} + M_u$$

where

$$\begin{aligned}\widehat{M}_{u,0} &= \frac{1}{h(t)}\sum_{s_1=1}^{u}\sum_{s_2=u+1}^{T}\frac{1}{n(n-1)}\sum_{i\neq j}\Big\{(X_{is_1}-\mu_{s_1})'(X_{js_1}-\mu_{s_1})\\ &\quad+(X_{is_2}-\mu_{s_2})'(X_{js_2}-\mu_{s_2})-2(X_{is_1}-\mu_{s_1})'(X_{js_2}-\mu_{s_2})\Big\};\\ \widehat{M}_{u,1} &= \frac{1}{h(t)}\sum_{s_1=1}^{u}\sum_{s_2=u+1}^{T}\frac{2}{n}(\mu_{s_1}-\mu_{s_2})'\{(X_{is_1}-X_{is_2})-(\mu_{s_1}-\mu_{s_2})\}.\end{aligned}$$

Note that $\widehat{M}_{u,0}/\sigma_{nu,0}$ is asymptotically the same as the $\widehat{M}_u/\sigma_{nu,0}$ under the null hypothesis, which has been shown to be tight in the proof of Theorem 3. In addition, $M_u/\sigma_{nu,0}$ is a sequence of nonrandom numbers, which is a bounded sequence by assumption. Therefore, to show the tightness of $\widehat{\mathscr{M}}$, we only need to show the tightness of $\widehat{M}_{u,1}/\sigma_{nu,0}$.

Using the results in the proof of Theorem 3, we note that the asymptotic order of $\sigma^2_{nu,0}$ is $n^{-2}T^3\text{tr}(\Sigma^2)$. Define

$$G_{n1}(\nu) = T^{-3/2}\text{tr}^{-1/2}(\Sigma^2)\sum_{s_1=1}^{[T\nu]}\sum_{s_2=[T\nu]+1}^{T}\sum_{i=1}^{n}(\mu_{s_1}-\mu_{s_2})'\{(X_{is_1}-X_{is_2})-(\mu_{s_1}-\mu_{s_2})\}.$$

It is then enough to show the tightness of $G_{n1}(\nu)$. Following the similar method in the proof of Theorem 3, for $1 > \eta > \nu > 0$,

$$E\{|G_{n1}(\nu)-G_{n1}(\eta)|^2\} \le nT^{-3}\text{tr}^{-1}(\Sigma^2)\left\|\sum_{s_1=1}^{[T\nu]}\sum_{s_2=[T\nu]+1}^{[T\eta]}(\mu_{s_1}-\mu_{s_2})'(\Gamma_{s_1}-\Gamma_{s_2})\right\|^2 + nT^{-3}\text{tr}^{-1}(\Sigma^2)\left\|\sum_{s_1=[T\nu]+1}^{[T\eta]}\sum_{s_2=[T\eta]+1}^{T}(\mu_{s_1}-\mu_{s_2})'(\Gamma_{s_1}-\Gamma_{s_2})\right\|^2.$$

Under the alternatives defined in (11), we have $E\{|G_{n1}(\nu)-G_{n1}(\eta)|^2\} = o\{|\eta-\nu|^2\}$. Thus, following the same steps in the proof of Theorem 3, we can show the tightness of $G_{n1}(\nu)$. This completes the proof of Theorem 4. ∎

Proof of Theorem 5. Similar to Theorem 1, Theorem 5 can be established by the martingale central limit theorem. To construct a martingale difference sequence, we define $Y_{is_a} = X_{is_a} - \mu_{s_a}$, then $\widehat{S}_n - S_n = \sum_{i=1}^n S_{ni}$, where

$$S_{ni} = \frac{4}{n(n-1)h(T)}\sum_{j=1}^{i-1}\left\{\sum_{s_1=1}^{T}\sum_{s_2=s_1+1}^{T}\sum_{a,b\in\{1,2\}}(-1)^{|a-b|}Y'_{is_a}Y_{js_b}\right\} + \frac{4}{nh(T)}\sum_{s_1=1}^{T}\sum_{s_2=s_1+1}^{T}\sum_{a,b\in\{1,2\}}(-1)^{|a-b|}\mu'_{s_a}Y_{is_b}.$$

Let $\{\mathscr{F}_i, 1 \le i \le n\}$ be σ-fields generated by $\sigma\{\mathbb{Y}_1, \dots, \mathbb{Y}_i\}$, where $\mathbb{Y}_i = \{Y_{i1}, \dots, Y_{iT}\}'$. Then it can be shown that $\text{E}(M_{tk}|\mathscr{F}_{k-1}) = 0$ for $k = 1, \dots, n$. Therefore, $\{M_{ti}, 1 \le i \le n\}$ is a martingale difference sequence with respect to σ-fields $\{\mathscr{F}_i, 1 \le i \le n\}$. By modifying Lemmas 1 and 2 in the supplementary material via changing the definition of the summation $\sum$ to

$$\sum \equiv \sum_{r_1<r_2}^{T}\sum_{s_1<s_2}^{T}\sum_{a,b,c,d\in\{1,2\}}\sum_{r_1^*<r_2^*}^{T}\sum_{s_1^*<s_2^*}^{T}\sum_{a^*,b^*,c^*,d^*\in\{1,2\}}(-1)^{|a-b|+|c-d|+|a^*-b^*|+|c^*-d^*|}.$$

Theorem 5 can be proved similarly to the proof of Theorem 1. ∎

Proof of Theorem 6. Recall that $\sigma_{\max} = \max_{0<t/T<1} \max\{\sqrt{\mathrm{tr}(A_{0t}^2)/h^2(t)}, \sqrt{n\|A_{1t}\|^2/h^2(t)}\}$ and $\delta = \|\mu_1 - \mu_T\|^2$. Given a constant C, we define a set

$$K(C) = \{t : |t-\tau| > CT\log^{1/2}T\sigma_{\max}/(n\delta),\ 1 \le t \le T-1\}.$$

To show Theorem 6, we first show that for any $\epsilon > 0$, there exists a constant C such that

$$\mathrm{P}\left\{|\hat{\tau} - \tau| > CT\log^{1/2}T\sigma_{\max}/(n\delta)\right\} < \epsilon. \tag{A3}$$

Since the event $\{\hat{\tau} \in K(C)\}$ implies the event $\{\max_{t\in K(C)}\hat{M}_t > \hat{M}_\tau\}$, then it is enough to show that

$$\mathrm{P}\left(\max_{t\in K(C)} \hat{M}_t > \hat{M}_\tau\right) < \epsilon.$$

Toward this end, we first derive the result based on the definition of M_t:

$$M_t = \left\{\frac{T-\tau}{T-t} I(1 \le t \le \tau) + \frac{\tau}{t} I(\tau < t \le T)\right\}\delta,$$

where $\delta = (\mu_1 - \mu_T)'(\mu_1 - \mu_T)$. Specially, M_t attains its maximum δ at $t = \tau$ since $1/(T-t)$ is an increasing function and $1/t$ is a decreasing function. As a result, by union sum inequality and letting $A(t, \tau|1, T) = 1/(T-t)I(1 \le t \le \tau) + 1/tI(\tau < t \le T)$, we have

$$\begin{aligned}
\mathrm{P}(\max_{t\in K(C)} \hat{M}_t > \hat{M}_\tau) &\le \sum_{t\in K(C)} \mathrm{P}(\hat{M}_t - M_t + M_t - M_\tau > \hat{M}_\tau - M_\tau) \\
&\le \sum_{t\in K(C)} \mathrm{P}\left\{\left|\frac{\hat{M}_t - M_t}{\sigma_{nt}}\right| > \frac{A(t,\tau|1,T)}{2}\frac{\delta}{\sigma_{\max}}|\tau - t|\right\} \\
&\quad + \sum_{t\in K(C)} \mathrm{P}\left\{\left|\frac{\hat{M}_\tau - M_\tau}{\sigma_{n\tau}}\right| > \frac{A(t,\tau|1,T)}{2}\frac{\delta}{\sigma_{\max}}|\tau - t|\right\} \\
&\le \sum_{t\in K(C)} \mathrm{P}\left\{\left|\frac{\hat{M}_t - M_t}{\sigma_{nt}}\right| > \sqrt{C\log T}\right\} + \sum_{t\in K(C)} \mathrm{P}\left\{\left|\frac{\hat{M}_\tau - M_\tau}{\sigma_{n\tau}}\right| > \sqrt{C\log T}\right\},
\end{aligned}$$

where the result of $A(t, \tau|1, T) = O(1/T)$ has been used.

Since $(\hat{M}_t - M_t)/\sigma_{nt} \sim \mathrm{N}(0, 1)$, for a large C,

$$\sum_{t\in K(C)} \mathrm{P}\left\{\left|\frac{\hat{M}_t - M_t}{\sigma_{nt}}\right| > \sqrt{C\log T}\right\} = \sum_{t\in K(C)} C(\log T)^{-1/2}T^{-C} \le \epsilon.$$

Similarly, we can show that

$$\sum_{t\in K(C)} \mathrm{P}\left\{\left|\frac{\hat{M}_\tau - M_\tau}{\sigma_{n\tau}}\right| > \sqrt{C\log T}\right\} \le \epsilon.$$

Hence, (A3) is true, which implies that $\hat{\tau} - \tau = O_p\left\{T\log^{1/2}T\sigma_{\max}/(n\delta)\right\}$.

Recall that $\sigma_{\max} = \max_{0<t/T<1}\max\{\sqrt{\mathrm{tr}(A_{0t}^2)/h^2(t)}, \sqrt{n\|A_{1t}\|^2/h^2(t)}\}$ and the assumption $\mathrm{tr}(\Xi_{s_1r_1}\Xi_{s_1r_1}') \asymp \phi(|s_1-r_1|)\mathrm{tr}(\Sigma_{s_1}\Sigma_{r_1})$ and $\sum_{k=1}^{T}\phi^{1/2}(k) < \infty$, following the proofs in Theorem 3, we have $\mathrm{tr}(A_{0t}^2) \asymp T^3\mathrm{tr}(\Sigma^2)$. Thus, we have $\mathrm{tr}(A_{0t}^2)/h^2(t) \asymp \mathrm{tr}(\Sigma^2)/T$.

For the second part in $\sigma_{\max}$, if $1 \le t \le \tau$, we have

$$\|A_{1t}\|^2 = (\mu_1-\mu_T)'\sum_{r_1,s_1=1}^{t}\sum_{r_2,s_2=t+1}^{T}(\Gamma_{r_1}-\Gamma_{r_2})(\Gamma_{s_1}-\Gamma_{s_2})'(\mu_1-\mu_T).$$

Using the assumption that $(\mu_1-\mu_T)'\Xi_{r_1s_1}(\mu_1-\mu_T) \asymp \phi(|r_1-s_1|)(\mu_1-\mu_T)'\Sigma(\mu_1-\mu_T)$, it can be checked that $\|A_{1t}\|^2 \asymp T^3(\mu_1-\mu_T)'\Sigma(\mu_1-\mu_T)$. In summary, we have

$$\sigma_{\max} = \max\{\sqrt{\mathrm{tr}(\Sigma^2)}, \sqrt{n(\mu_1-\mu_T)'\Sigma(\mu_1-\mu_T)}\}/\sqrt{T} = v_{\max}/\sqrt{T}.$$

This completes the proof of Theorem 6. ■

Proof of Theorem 7. To prove Theorem 7, we need the following Lemma 1, whose proof is presented in the supplementary material. It asserts that the maximum of M_t given by (4) is attained at one of the change-points $1 \le \tau_1 < \ldots < \tau_q < T$.

Lemma 1. *Let* $1 \le \tau_1 < \ldots < \tau_q < T$ *be* $q \ge 1$ *change-points such that* $\mu_1 = \ldots = \mu_{\tau_1} \ne \mu_{\tau_1+1} = \ldots = \mu_{\tau_q} \ne \mu_{\tau_q+1} = \ldots = \mu_T$. *Then,* M_t *defined by (4) attains its maximum at one of the change-points.* ■

We now prove Theorem 7. Recall that within the time interval $[1, T]$, there are q change-points. First, we will show that the proposed binary segmentation algorithm detects the existence of change-points with probability one. To show this, according to Theorem 3, we only need to show that $\mathrm{P}(\widehat{\mathcal{S}}_n[1,T] > z_{\alpha_n}) = 1$, where z_{α_n} is the upper α_n quantile of the standard normal distribution. This can be shown because for any $1 \le t \le T-1$,

$$\mathrm{P}(\widehat{\mathcal{S}}_n[1,T] > z_{\alpha_n}) = \mathrm{P}\left(\frac{\widehat{S}_n[1,T]}{\sigma_{n,0}[1,T]} > z_{\alpha_n}\right) = 1-\Phi\left(\frac{\sigma_{n,0}[1,T]}{\sigma_n[1,T]}z_{\alpha_n} - \frac{S_\mu[1,T]}{\sigma_n[1,T]}\right),$$

which converges to 1 because $\sigma_{n,0}[1,T] \le \sigma_n[1,T]$, $S_\mu[1,T]/\sigma_n[1,T] \to \infty$, and $z_{\alpha_n} = o(S_n[1,T]/\sigma_n[1,T])$.

Once the existence of change-points is detected, the proposed binary segmentation algorithm will continue to identify change-points. Since $v_{\max} = o\{n\delta/(T\sqrt{\log T})\}$, one change-point $\tau_{(1)} \in \{\tau_1, \ldots, \tau_q\}$ can be identified correctly with probability 1 based on similar derivations given in the proof of Theorem 6, and the fact that M_t achieves its maximum at one of change-points as shown in Lemma 3.

Since each subsequence satisfies the condition that $z_{\alpha_n} = o(\mathscr{R}^*)$, the detection continues. Suppose that there are less than q change-points identified successfully, then there exists a segment I_t contains a change-point. Since $z_{\alpha_n} = o(\mathscr{R}^*)$ and $v_{\max}[I_t] = o\{n\delta[I_t]/(T\sqrt{\log T})\}$, the change-point will be detected and identified by the proposed binary segmentation method. Once all q change-points have been identified consistently, each of all the subsequent segments has two end points chosen from $1, \tau_1, \ldots, \tau_q, T$. Then the proposed binary segmentation algorithm will not wrongly detect any change-point from any segment I_t that contains no change-point, $\mathrm{P}(\widehat{\mathcal{S}}_n[I_t] > z_{\alpha_n}[1,T]) = \alpha_n \to 0$, which implies that no change-point will be identified further. This completes the proof of Theorem 7. ■