Journal of Machine Learning Research 24 (2023) 1-44 Submitted 10/20; Revised 1/23; Published 3/23

Online Change-Point Detection in High-Dimensional Covariance Structure with Application to Dynamic Networks

Lingjun Li lli13@kent.edu
Jun Li jli49@kent.edu
Department of Mathematical Sciences
Kent State University
Kent, Ohio 44242, USA

Editor: Zaid Harchaoui

Abstract

In this paper, we develop an online change-point detection procedure in the covariance structure of high-dimensional data. A new stopping rule is proposed to terminate the process as early as possible when a change in covariance structure occurs. The stopping rule allows spatial and temporal dependence and can be applied to non-Gaussian data. An explicit expression for the average run length is derived, so that the level of threshold in the stopping rule can be easily obtained with no need to run time-consuming Monte Carlo simulations. We also establish an upper bound for the expected detection delay, the expression of which demonstrates the impact of data dependence and magnitude of change in the covariance structure. Simulation studies are provided to confirm accuracy of the theoretical results. The practical usefulness of the proposed procedure is illustrated by detecting the change of brain's covariance network in a resting-state fMRI data set. The implementation of the methodology is provided in the R package OnlineCOV.

Keywords: change-point detection, high-dimensional data, spatial and temporal dependence

1. Introduction

Online change-point detection or sequential change-point detection originally arises from the problem of quality control. The product quality is monitored based on the observations continually arriving during an industrial process and a stopping rule is chosen to terminate and reset the process as early as possible when an anomaly or a change occurs. There are two errors to be controlled by a stopping rule. One is false alarm, which is the time when the stopping rule terminates a monitored process without undergoing any change. Another is a detection delay which is the number of additional observations the stopping rule collects in order to detect a change point. The goal is to construct a stopping rule which can make the expected detection delay as small as possible, subject to the constraint that the average run length to false alarm is controlled at a pre-specified level. In modern applications, there has been a resurgence of interest in detecting abrupt change from streaming data with a large number of measurements. Examples include real-time monitoring for sensor networks and threat detection from surveillance videos. More can be found in studying dynamic connectivity of resting state functional magnetic resonance imaging, and in detecting threat of fake news from the group of fake accounts in social networks (Bara et al., 2015).

Extensive research has been done for online change-point detection of univariate data; see, for example, Page (1954), Shiryayev (1963), Lorden (1971), Wald (1973), Siegmund (1985) and Siegmund and Venkatraman (1995). The proposed stopping rules are based on the CUSUM test or the quasi-Bayesian test which assume the distributions of data before and after the change point to be known, or its variants proposed to relax the restrictive assumption of known distributions. There also exist many developments in online change-point detection of multivariate data. For example, Tartakovsky and Veeravalli (2008) propose the stopping rule for the common change point detection from all dimensions based on the assumption that the distributions of data before and after the change point are known. By relaxing the common change to the change of only subset of data, Xie and Siegmund (2013) study the stopping rule for the multivariate normally distributed data with the identity covariance matrix.

In this paper, we consider online change-point detection in the covariance structure of high-dimensional data. More precisely, letting $\{X_1, X_2, \cdots\}$ be a sequence of continually arriving p-dimensional random vectors, each of which has its own covariance matrix Σ_i, we consider the hypotheses

$$\begin{aligned} H_0 &: \Sigma_1 = \Sigma_2 = \cdots \qquad \text{against} \\ H_1 &: \Sigma_1 = \cdots = \Sigma_\tau \neq \Sigma_{\tau+1} = \cdots , \end{aligned} \tag{1}$$

where τ is an unknown change point. The motivation behind the considered problem stems from the applications of detecting covariance network changes, such as dynamic changes in brain's functional connectivity, where the network can be quantified by the covariance or precision (inverse covariance) matrix (Varoquaux et al., 2010). Despite the practical usefulness of the considered problem, the aforementioned methods cannot be applied to the hypotheses (1), either because they work specifically for change-point detection in the mean or because they require fixed dimensionality. There are some recent developments for change-point detection in the high-dimensional covariance structure (Aue et al., 2009). Nevertheless, they cannot be directly applied to the online change-point detection problem (1), because those methods are designed for the offline change-point detection problem, where the entire data need to be collected before any statistical analysis is carried out.

Our contribution in this paper is to propose a new stopping rule for the above high-dimensional online change-point detection problem. We study the asymptotic behavior of the stopping time of the stopping rule under both null and alternative hypotheses. More specifically, we derive an explicit expression for the average run length of the stopping time under the null hypothesis, so that the level of threshold in the stopping rule can be easily obtained with no need to run time-consuming Monte Carlo simulations. Under the alternative hypothesis, we establish an upper bound for the expected detection delay, which demonstrates the impact of data dependence and magnitude of change in the covariance structure. The proposed stopping rule has some advantages. First, unlike vast majority of online change-point detection procedures such as Avanesov and Buzun (2018) and Avanesov (2019) that assume temporal independence, it allows temporal dependence among different high-dimensional measurements at different time points. More precisely, we incorporate spatial and temporal dependence through a nonparametric factor model in Section 2.1 that does not assume Gaussian distribution. Our method is essentially nonparametric and thus different from the sequential detection procedures in Lai (1995), where dependence is incorporated by replacing parametric probability densities with conditional densities in likelihood ratios. There also exist some offline change-point detection methods that allow both spatial and temporal dependence; see, for example, Aue et al. (2009) and Jirak (2015). One option is to implement these offline methods every time when a new observation arrives. But this will raise challenge to control the false alarm especially when there is a highly correlated multiple testing procedure. Second, we estimate the temporal dependence consistently through a data-driven procedure, and establish the distribution of the stopping time with the correctly specified dependence. Consequently, the average run length of the proposed stopping rule can be well controlled even in the presence of temporal dependence. Third, the stopping rule is implementable when the dimension p diverges and thus suitable for monitoring modern networks whose size varies enormously from thousands to millions. Finally, we identify the key factors and establish their impact on the expected detection delay through an explicitly derived upper bound. In particular, we reveal that the expected detection delay based on the L_2-norm statistic increases as the strength of temporal dependence increases, but decreases as the magnitude of change $||\Sigma_{\tau+1} - \Sigma_\tau||_F/||\Sigma_\tau||_F$ increases. Here $||\cdot||_F$ represents the matrix Frobenius norm. The implementation of the proposed stopping rule is provided in the R package OnlineCOV (Li and Li, 2020).

It is worth mentioning that Avanesov and Buzun (2018) and Avanesov (2019) also consider online change-point detection in the precision matrix and in the covariance matrix, respectively. There are some differences between the approaches in Avanesov and Buzun (2018) and Avanesov (2019) and the current approach. The first difference is that the approaches in Avanesov and Buzun (2018) and Avanesov (2019) are multiple testing procedures with critical values to control the probability to raise false alarms. The critical values are chosen by a bootstrap calibration scheme. The current approach is based on a proposed stopping rule, which terminates or continues a monitored process through the comparison of a proposed test statistic with a threshold. The threshold controls the average run length to false alarm, and can be obtained by solving a equation involving an explicit expression of the average run length as a function of the threshold. Another difference is that the test statistics in Avanesov and Buzun (2018) and Avanesov (2019) employ the matrix sup-norm and the test statistic in current work considers the Frobenius norm. Since the sup-norm is the largest element of a matrix in absolute value, the approaches in Avanesov and Buzun (2018) and Avanesov (2019) would be advantageous if changes of the precision matrix or the covariance matrix happen in a sparse number of elements. On the other hand, since the Frobenius norm sums the squared elements of a matrix, our proposed stopping rule would be advantageous if changes of the covariance matrix happen in a large number of elements.

The rest of the paper is organized as follows. Section 2 introduces the proposed stopping rule. Section 3 presents its asymptotic properties. Simulation studies and real data analysis are given in Sections 4 and 5, respectively. We conclude the paper with some discussions in Section 6. All the proofs are delegated to the Appendix.

2. Methodology

In Sections 2.1-2.2, we first discuss a model (2), conditions (C1)-(C2) and a test statistic (3) based on offline data, in the sense that a sequence of p-dimensional random vectors $\{X_i, 1 \leq i \leq n\}$ have been already observed. In Section 2.3, we extend them to online framework, where observations are continually arriving and a stopping rule is proposed to determine whether a process needs to be terminated.

2.1 Modeling Spatial and Temporal Dependence

Since our main focus is change point detection of covariance structure, we assume $\mathrm{E}(X_i) = \mu$ to facilitate our analysis. We model the sequence by

$$X_i = \mu + \Gamma_i Z \qquad \text{for} \quad i = 1, \cdots, n, \tag{2}$$

where Γ_i is a $p \times m$ matrix with $m \geq n \cdot p$, and $Z = (z_1, \cdots, z_m)^T$ such that $\{z_i\}_{i=1}^m$ are mutually independent and satisfy $\mathrm{E}(z_i) = 0$, $\mathrm{var}(z_i) = 1$ and $\mathrm{E}(z_i^4) = 3 + \beta$ for some constant β.

Admittedly, Bai and Saranadasa (1996), Chen and Qin (2010) and Li and Chen (2012) employ a similar factor model for the one-sample or two-sample offline hypothesis testing problems, where the observations $\{X_i, 1 \leq i \leq n\}$ are assumed to be independent. Our model (2) is more general as it can incorporate both spatial and temporal dependence of the sequence $\{X_i, 1 \leq i \leq n\}$. To appreciate this, we let $X = (X_1^T, \cdots, X_n^T)^T$ and $\Gamma = (\Gamma_1^T, \cdots, \Gamma_n^T)^T$. From (2), the covariance matrix of X is $\Gamma\Gamma^T$, in which each $p \times p$ block diagonal sub-matrix $\Gamma_i\Gamma_i^T \equiv \Sigma_i$ represents the spatial dependence of each X_i and each $p \times p$ block off-diagonal sub-matrix $\Gamma_i\Gamma_j^T \equiv C(j-i)$ describes the spatial and temporal dependence between X_i and X_j at $i \neq j$. Here we require $m \geq n \times p$ to ensure the positive definiteness of $\Gamma\Gamma^T$.

Based on (2), we accommodate the spatial and temporal dependence by the following two conditions.

(C1). The sequence is M-dependent, such that for some integer $M \geq 0$, $C(j-i) \neq 0$ if and only if $|j-i| \leq M$. Moreover, under H_0 of (1), $C(j-i) = C(h)$ for all i and j satisfying $j - i = h$ with $h \in \{0, \pm 1, \cdots, \pm M\}$.

Under the null hypothesis, we assume that the sequence is M-dependent, and the spatial and temporal dependence is stationary. Under the alternative hypothesis, the covariance structure changes and consequently, the stationarity of the spatial and temporal dependence cannot hold. We thus only assume the M-dependence. We introduce the M-dependence to relax the commonly assumed temporal independence in the literature. The assumption enables us to establish the asymptotic normality of the proposed test statistic given by (3) in Section 2.2 through the martingale central limit theorem. Moreover, the M-dependence combined with the stationarity in the spatial and temporal dependence, yields that the stopping time of the proposed stopping rule given by (7) in Section 2.3 converges to the Gumbel limiting distribution of a stationary Gaussian process under the null hypothesis. Under the alternative hypothesis, we apply the generalized Wald's lemma under M-dependence in Janson (1983) to study the expected detection delay. Despite technical challenges, it might be useful to consider dependence beyond the M-dependence. Similar to Li et al. (2019), an option is to require that the spatial and temporal dependence is weak and dominated by the dependence within M, when the timespan exceeds a critical value M. Another option is to apply the idea in Aue et al. (2009) to introduce a sequence of M-dependent random vectors $\{X_i^M, M \geq 1\}$ to approximate X_i so that $\sum_{M \geq 1} |X_i - X_i^M|_4 < \infty$, where

$|X|_4 = \{\mathrm{E}(X)^4\}^{1/4}$ for a random vector X. We leave for future research the validation of theoretical results, in particular Wald's lemma under such circumstances.

(C2). For any $h_1, h_2, h_3, h_4 \in \{0, \pm 1, \cdots, \pm M\}$, as $p \to \infty$,

$$\mathrm{tr}\{C(h_1)C(h_2)C(h_3)C(h_4)\} = o\Big[\mathrm{tr}\{C(h_1')C(h_2')\}\mathrm{tr}\{C(h_3')C(h_4')\}\Big],$$

where $\{h_1', h_2', h_3', h_4'\}$ is a permutation of $\{h_1, h_2, h_3, h_4\}$.

Without temporal dependence, (C2) becomes $\mathrm{tr}\{C^4(0)\} = o[\mathrm{tr}^2\{C^2(0)\}]$. It holds if all the eigenvalues of $C(0)$ are bounded, but may not hold under strong dependence such as the compound symmetry structure which means that all the variances are equal and all the covariances are equal. If the temporal dependence is present ($h \neq 0$), (C2) takes into account both spatial and temporal dependence. It can be shown that (C2) holds if the requirement of bounded eigenvalues is extended to the $np \times np$ covariance matrix of entire sequence $X = (X_1^T, X_2^T, \cdots, X_n^T)^T$. The condition cannot hold if the spatial and temporal dependence is too strong so that the covariance matrix of X has unbounded eigenvalues. The advantage of (C2) is that it does not impose any decay structures on $C(h)$ as long as the trace condition is satisfied. Moreover, it allows the dimension p to diverge without imposing its growth rate.

2.2 Test Statistic

Given the sequence of p-dimensional random vectors $\{X_i, 1 \leq i \leq n\}$, we need a test statistic, the expectation of which can measure the heterogeneity of covariance structure from the collected observations. Assuming for the moment that $\mu = 0$ in (2), we propose the following L_2-norm U-statistic

$$\hat{\mathcal{J}}_{n,M} \equiv \frac{1}{n^2}\sum_{i,j=1}^{n} W_M(i,j)(X_i^T X_j)^2, \tag{3}$$

where the weight function $W_M(i,j) \equiv \sum_{t=M+2}^{n-M-2} A_{t,M}(i,j)\mathrm{I}(|i-j| \geq M+1)$ and

$$\begin{aligned} A_{t,M}(i,j) &= \frac{n-t-M}{t-M-1}\mathrm{I}(i \leq t)\mathrm{I}(j \leq t) + \frac{t-M}{n-t-M-1}\mathrm{I}(t+1 \leq i)\mathrm{I}(t+1 \leq j) \\ &- \frac{(t-M)(n-t-M)}{t(n-t) - \frac{1}{2}M(M+1)}\{\mathrm{I}(i \leq t)\mathrm{I}(t+1 \leq j) + \mathrm{I}(t+1 \leq i)\mathrm{I}(j \leq t)\}. \end{aligned}$$

If $\mu \neq 0$, a centralized version of (3) is

$$\hat{\mathcal{J}}^*_{n,M} = \frac{1}{n^2}\sum_{i,j=1}^{n} W_M(i,j)\{(X_i - \hat{\mu})^T(X_j - \hat{\mu})\}^2, \tag{4}$$

where $\hat{\mu}$ is a consistent estimator of μ. As introduced in Section 2.3, the proposed stopping rule needs a training sample and $\hat{\mu}$ thus can be chosen as the sample mean of the training sample. While U-statistics have been widely implemented for hypothesis testing and offline change-point detection problems (Hoeffding, 1948; Chen and Qin, 2010; Li and Chen, 2012; Matteson and James, 2014), the extension to high-dimensional online change-point detection problem has not been explored yet.

Remark 1 We first assume a known M to present the main results of the proposed methods. We then provide a data-driven procedure for estimating M and establish the theoretical results based on the estimated M in Section 3.4.

Remark 2 Using the expression of $A_{t,M}(i,j)$, we write (3) into $\hat{\mathcal{J}}_{n,M} = n^{-2}\sum_{t=M+2}^{n-M-2}(t-M)(n-t-M)\hat{\mathcal{J}}_{n,M,t}$, where

$$\begin{aligned}\hat{\mathcal{J}}_{n,M,t} &= \sum_{i,j=1}^{t}\frac{\mathrm{I}(|i-j|\geq M+1)(X_i^TX_j)^2}{(t-M)(t-M-1)} + \sum_{i,j=t+1}^{n}\frac{\mathrm{I}(|i-j|\geq M+1)(X_i^TX_j)^2}{(n-t-M)(n-t-M-1)}\\ &- 2\sum_{i=1}^{t}\sum_{j=t+1}^{n}\frac{\mathrm{I}(|i-j|\geq M+1)(X_i^TX_j)^2}{t(n-t)-\frac{1}{2}M(M+1)}.\end{aligned}$$

The test statistic is constructed in several steps. At each t from $\{M+2,\cdots,n-M-2\}$, we partition the entire sequence $\{X_i, 1\leq i\leq n\}$ into $\{X_i, 1\leq i\leq t\}$ and $\{X_i, t+1\leq i\leq n\}$. Since $\mathrm{E}\{(X_i^TX_j)^2\mathrm{I}(|i-j|\geq M+1)\} = \mathrm{tr}(\Sigma_i\Sigma_j)$, the three terms in $\hat{\mathcal{J}}_{n,M,t}$ estimate the trace of covariance structure from the segments $\{1\leq i,j\leq t\}$ and $\{t+1\leq i,j\leq n\}$ and $\{1\leq i\leq t, t+1\leq j\leq n\}$, respectively. The normalization factors $(t-M)(t-M-1)$, $(n-t-M)(n-t-M-1)$ and $t(n-t)-1/2M(M+1)$ count the number of indices i and j within each segment, the distance of which is no less than $M+1$. Each normalization factor can be obtained by subtracting the diagonal and $2M$ off-diagonal elements from each of three blocks $\{1\leq i,j\leq t\}$ and $\{t+1\leq i,j\leq n\}$ and $\{1\leq i\leq t, t+1\leq j\leq n\}$ of the $n\times n$ matrix. For example, the factor $(t-M)(t-M-1)$ equals t^2 elements from the block $\{1\leq i,j\leq t\}$ minus $t+2\{(t-1)+\cdots+(t-M)\}$ elements from the diagonal and $2M$ off-diagonals. If t is chosen to be the change point τ, we can show that $\mathrm{E}(\hat{\mathcal{J}}_{n,M,\tau}) = \mathrm{tr}(\Sigma_\tau^2)+\mathrm{tr}(\Sigma_{\tau+1}^2)-2\mathrm{tr}(\Sigma_\tau\Sigma_{\tau+1}) = \mathrm{tr}\{(\Sigma_\tau-\Sigma_{\tau+1})^2\}$. However the change point is unknown in practice. We therefore accumulate $\hat{\mathcal{J}}_{n,M,t}$ from $t=M+2$ to $n-M-2$ to obtain the test statistic $\hat{\mathcal{J}}_{n,M}$, where each $\hat{\mathcal{J}}_{n,M,t}$ is multiplied by an additional weight factor $(t-M)(n-t-M)$ so that $\mathrm{E}(\hat{\mathcal{J}}_{n,M})$ is dominated by the term $(\tau-M)(n-\tau-M)\mathrm{E}(\hat{\mathcal{J}}_{n,M,\tau}) = (\tau-M)(n-\tau-M)\mathrm{tr}\{(\Sigma_\tau-\Sigma_{\tau+1})^2\}$ (Li et al., 2019).

The statistic $\hat{\mathcal{J}}_{n,M,t}$ can be related to the traditional cumulative sum (CUSUM) statistic subject to some bias correction terms. To appreciate this, we consider a special case of $M=0$ and obtain

$$\begin{aligned}\hat{\mathcal{J}}_{n,M,t} &= \frac{n^2}{t^2(n-t)^2}\mathrm{tr}\Big\{(\sum_{i=1}^{t}X_iX_i^T-\frac{t}{n}\sum_{i=1}^{n}X_iX_i^T)(\sum_{j=1}^{t}X_iX_i^T-\frac{t}{n}\sum_{j=1}^{n}X_iX_i^T)\Big\}\\ &+ \frac{1}{t^2(t-1)}\sum_{i,j=1}^{t}(X_i^TX_j)^2+\frac{1}{(n-t)^2(n-t-1)}\sum_{i,j=t+1}^{n}(X_i^TX_j)^2\\ &- \frac{1}{t(t-1)}\sum_{i=1}^{t}(X_i^TX_i)^2-\frac{1}{(n-t)(n-t-1)}\sum_{i=t+1}^{n}(X_i^TX_i)^2,\end{aligned}$$

where the first term on the right hand side contains the CUSUM statistic $\sum_{i=1}^{t} X_i X_i^T - t/n \sum_{i=1}^{n} X_i X_i^T$, and other terms are bias corrections. Especially when t is the change point τ, including the bias corrections leads to $\mathrm{E}(\hat{\mathcal{J}}_{n,M,t}) = \mathrm{tr}\{(\Sigma_\tau - \Sigma_{\tau+1})^2\}$.

Since the main task is to detect change in the covariance structure, we assume without further notice that $\mu = 0$ in (2), and focus on $\hat{\mathcal{J}}_{n,M}$ to facilitate theoretical investigation. All the established results can be readily extended to $\hat{\mathcal{J}}^*_{n,M}$ with $\mu \neq 0$.

Proposition 1 Assume (2) and (C1). Under the null hypothesis,

$$\mathrm{E}(\hat{\mathcal{J}}_{n,M}) = 0.$$

Under the alternative hypothesis,

$$\mu_{\hat{\mathcal{J}}_{n,M}} \equiv \mathrm{E}(\hat{\mathcal{J}}_{n,M}) = \frac{1}{n^2} \sum_{i,j=1}^{n} W_M(i,j) \mathrm{tr}(\Sigma_i \Sigma_j).$$

Since the expectation of $\hat{\mathcal{J}}_{n,M}$ under the alternative hypothesis differs from its expectation under the null hypothesis, it can be used to test heterogeneity of the covariance structure after we standardize it. This requires us to further derive the variance of the test statistic.

Proposition 2 Under (2) and (C1)–(C2), as $p \to \infty$, the variance of $\hat{\mathcal{J}}_{n,M}$ is

$$\sigma^2_{\hat{\mathcal{J}}_{n,M}} = \frac{4}{n^4} \sum_{i,j=1}^{n} \sum_{k,l=1}^{n} W_M(i,j) W_M(k,l) \mathrm{tr}^2\{C(i-k)C(l-j)\}\{1 + o(1)\}.$$

Under the null hypothesis, (C1) assumes that the spatial and temporal dependence is stationary. The leading order variance can be simplified as

$$\sigma^2_{\hat{\mathcal{J}}_{n,M},0} = \frac{4}{n^4} \sum_{i,j=1}^{n} \sum_{h_1,h_2} W_M(i,j) W_M(i-h_1, j+h_2) \mathrm{tr}^2\{C(h_1)C(h_2)\}, \quad (5)$$

where $h_1, h_2 \in \{0, \pm 1, \cdots, \pm M\}$.

2.3 Stopping Rule

We now move to the online framework. We intend to propose a stopping rule based on $\hat{\mathcal{J}}_{n,M}$ (3) for the hypotheses (1). However, there are two issues we need to address when using $\hat{\mathcal{J}}_{n,M}$. The first issue is nuisance parameters. There are two nuisance parameters related to $\hat{\mathcal{J}}_{n,M}$: the M for temporal dependence and the standard deviation of $\hat{\mathcal{J}}_{n,M}$ under the null hypothesis. While M is zero under temporal independence and $\sigma_{\hat{\mathcal{J}}_{n,M},0}$ in (5) becomes $2p/n^2\{\sum_{i,j} W_M^2(i,j)\}^{1/2}$ without temporal dependence and with the identity covariance matrix Σ, they are unknown in the presence of spatial and temporal dependence. Similar to Pollak and Siegmund (1991), we consider a training sample of size n_0 to provide estimation of both nuisance parameters. Estimating M based on the training sample is covered in Section 3.4. To estimate the standard deviation of $\hat{\mathcal{J}}_{n,M}$ under the null hypothesis, we need to estimate $\mathrm{tr}\{C(h_1)C(h_2)\}$ because it is the only unknown in (5). Based on a training sample, we estimate it by

$$\mathrm{tr}\{\widehat{C(h_1)C}(h_2)\} = \frac{1}{n^*}\sum_{s,t}^{*} X_{t+h_2}^T X_s X_{s+h_1}^T X_t, \tag{6}$$

where $\sum^*$ represents the sum of indices that are at least M apart in the training sample, and n^* be the corresponding number of indices. The consistency of the estimator is established in Theorem 3 of Section 3.3.

The second issue related to $\hat{\mathcal{J}}_{n,M}$ is the computational complexity. From (3), $\hat{\mathcal{J}}_{n,M}$ involves the weight function $W_M(i,j)$ which sums t from $M+2$ to $n-M-2$. As mentioned in Remark 2, at each t, the test statistic needs to compare the two covariance structures estimated separately from the two segments $\{X_i, 1 \le i \le t\}$ and $\{X_i, t+1 \le i \le n\}$. When n is large, it can be time consuming to compute $\hat{\mathcal{J}}_{n,M}$. To reduce the computational time, we consider a modified statistic

$$\hat{\mathcal{J}}_{n,M,H} = \frac{1}{H^2}\sum_{i,j=n-H+1}^{n} W_M(i,j)(X_i^T X_j)^2,$$

which, compared with the original $\hat{\mathcal{J}}_{n,M}$, is only based on the past H observations from the current time n and thus can be computationally more efficient. It is quite common to use a moving window H for the online change point detection; see, for example, Lai (1995) and Cao et al. (2019). More can be seen in Avanesov and Buzun (2018) and Avanesov (2019), where multiple window sizes are employed to attain a trade-off between the power of the change-point test and accuracy of the change-point estimation. While the motivation is to reduce the computational complexity, the impact of imposing H on the proposed method needs to be carefully addressed. We display its effect on our stopping rule explicitly through asymptotic results in Section 3 and simulation studies in Section 4. Some other guidelines in selecting the window size can be seen in Lai (1995).

We are now ready to propose the stopping rule

$$T_H(a,M) = \inf\left\{n - n_0 : \left|\frac{\hat{\mathcal{J}}_{n,M,H}}{\hat{\sigma}_{n_0,M,H}}\right| > a,\ n > n_0\right\}, \tag{7}$$

where n_0 is the size of a training sample, and $\hat{\sigma}^2_{n_0,M,H}$ is the estimator of the variance of $\hat{\mathcal{J}}_{n,M,H}$ under the null hypothesis. Using (6), we obtain

$$\hat{\sigma}^2_{n_0,M,H} = \frac{4}{H^4}\sum_{i,j=1}^{H}\sum_{h_1,h_2} W_M(i,j)W_M(i-h_1,j+h_2)\mathrm{tr}^2\{\widehat{C(h_1)C}(h_2)\}. \tag{8}$$

The proposed stopping rule terminates the detection procedure in a minimal number of new observations after the training sample n_0, when the absolute value of the standardized test statistic is above the threshold a. For online change point detection, a should be chosen to balance the tradeoff between false alarms and detection delay. If a is too small, the stopping rule may detect a change quickly but unavoidably generate a lot of false alarms if there is no change. On the other hand, if a is too large, the desire to avoid false alarms will lead to a significant delay between the change point and the termination time. A conventional method for choosing a is that the average run length can be controlled at any pre-specified value. In next section, we establish an explicit relationship between a and the average run length in Theorem 1, which allows us to quickly determine a with no need to run time-consuming Monte Carlo simulations. The proposed stopping rule also depends on the temporal dependence M, which is unknown in practice. In Section 3.4, we provide an algorithm to consistently estimate M. Our investigation shows that the stopping rule based on the estimated M performs as well as that based on the true M.

3. Asymptotic Results

In this section we examine the asymptotic performance of the proposed stopping rule $T_H(a, M)$ for large p and a.

3.1 Average Run Length

Let E_∞ and P_∞ denote the expectation and probability, respectively, under the null hypothesis. Let

$$g(t/H, a) = 2\log(t/H) + 1/2\log\log(t/H) + \log(4/\sqrt{\pi}) - a\sqrt{2\log(t/H)}.$$

The average run length is defined to be the expected value of the stopping time under the null hypothesis. The following theorem establishes the average run length or $\mathrm{E}_\infty\{T_H(a, M)\}$ for the proposed stopping rule (7).

Theorem 1 Assume (2) and (C1)–(C2). As $n_0 \to \infty$, $p \to \infty$, and both H and $a \to \infty$ satisfying $H = o\{\exp(a^2/2)\}$,

$$\mathrm{E}_\infty\{T_H(a, M)\} = \left(H + \int_H^\infty \exp\left[-2\exp\left\{g(t/H, a)\right\}\right] dt\right)\{1 + o(1)\}.$$

As shown in the proof of Theorem 1, the average run length is readily obtained by establishing the cumulative distribution function of $T_H(a, M)$ as $a \to \infty$. Since the randomness of $T_H(a, M)$ is determined by $\hat{\mathcal{J}}_{n,M,H}/\hat{\sigma}_{n_0,M,H}$, the cumulative distribution of the former can be derived by establishing the asymptotic distribution of the latter when p and $H \to \infty$. Here the condition $H = o\{\exp(a^2/2)\}$ specifies the growth rate of H with respect to a. It is imposed to ensure that the probability the procedure stops within the window H goes to zero exponentially fast.

Theorem 1 states that the average run length depends on the threshold a and the window size H. In particular, it increases as a increases. This can also be seen from the proposed stopping rule (7), where raising a makes the standardized test statistic less likely to go beyond the a when there is no change point. The practical usefulness of Theorem 1 is that with any pre-specified average run length and H, we can quickly determine the value of a by solving the nonlinear equation using a software such as the uniroot function in R rather than running time-consuming Monte Carlo simulations.

3.2 Expected Detection Delay

When there is a change point τ, the proposed stopping rule is conventionally examined by the expected detection delay, $\mathrm{E}_\tau\{T_H(a,M)-(\tau-n_0)|T_H(a,M)>\tau-n_0\}$ with $\tau\ge n_0$. In the literature, it is customary to consider the expected detection delay for the so-called immediate change point; see, for example, Siegmund and Venkatraman (1995) and Xie and Siegmund (2013). In terms of our setup, it refers to the change that occurs immediately after the training sample n_0 and the corresponding expected detection delay is $\mathrm{E}_0\{T_H(a,M)\}$. The main reason to consider the expected detection delay of the immediate change point is that for many stopping rules, the supremum of all the expected detection delays attains at the immediate change point. It is therefore important to see if such property attains is still held by our proposed stopping rule. We establish the following theorem which confirms this conclusion. More importantly, the theorem provides an upper bound for the expected detection delays.

Theorem 2 Consider $\tau\ge n_0$ and assume the same conditions in Theorem 1,

$$\sup_{n_0\le\tau<\infty}\mathrm{E}_\tau\{T_H(a,M)-(\tau-n_0)|T_H(a,M)>\tau-n_0\}=\mathrm{E}_0\{T_H(a,M)\},\quad\text{and}$$

$$\mathrm{E}_0\{T_H(a,M)\}\le(M+2)+\left(\frac{a\cdot H\cdot\sigma_{H,M,0}}{\log H\cdot||\Sigma_{\tau+1}-\Sigma_\tau||_F^2}+\frac{M(M+1)||\Sigma_\tau||_F}{||\Sigma_{\tau+1}-\Sigma_\tau||_F}\right)^{1/2}\{1+o(1)\},$$

where $\sigma_{H,M,0}$ is obtained by replacing n with H in (5), and $||\cdot||_F$ represents the matrix Frobenius Norm.

Theorem 2 demonstrates the impact of some key factors on the expected detection delay. First, a larger M could lead to a greater expected detection delay, showing the adverse effect of the dependence on change-point detection. Second, the impact of the threshold a on the expected detection delay essentially depends on the choice of the average run length, because a is obtained by solving the equation in Theorem 1 in which the window size H and the average run length are pre-specified by the user. Generally speaking, a larger user-chosen average run length leads to a higher value of a and thus a greater expected detection delay. Finally, by applying the result $\sigma_{H,M,0}=O(||\Sigma_\tau||_F^2)$ from the proof of Theorem 1, the impact of $\sigma_{H,M,0}\,||\Sigma_{\tau+1}-\Sigma_\tau||_F^{-2}$ can be demonstrated to be

$$\frac{\sigma_{H,M,0}}{||\Sigma_{\tau+1}-\Sigma_\tau||_F^2}=O\left(\frac{||\Sigma_\tau||_F^2}{||\Sigma_{\tau+1}-\Sigma_\tau||_F^2}\right).$$

The result shows that the expected detection delay can be significantly reduced by increasing the ratio of the change in covariance structure to the original covariance.

Remark 3 It requires a minimum change in the covariance structure, for the proposed stopping rule to detect the change point. To understand this, we consider the configuration with the immediate change after the training sample. As the window continuously moves to the right, the number of observations with Σ_τ decreases but the number of observations with $\Sigma_{\tau+1}$ increases. If the detection procedure has not yet stopped when the last observation with Σ_τ begins to leave the window, it probably won't be able to stop because the process ends up with all the H observations having the same $\Sigma_{\tau+1}$. Theorem 2 actually provides a minimum change the proposed stopping rule requires. By noticing that the right hand side of the inequality in Theorem 2 must be no more than H, the change of covariance structure

$$||\Sigma_{\tau+1} - \Sigma_\tau||_F \geq \frac{M(M+1)||\Sigma_\tau||_F}{2(H-M-2)^2} + \sqrt{\frac{M^2(M+1)^2||\Sigma_\tau||_F^2}{4(H-M-2)^4} + \frac{a \cdot H \cdot \sigma_{H,M,0}}{(H-M-2)^2 \log H}},$$

where the quantity on the right hand side is therefore the minimum change in the covariance structure the proposed stopping rule is able to detect. To provide an insight of the result, we consider $\Sigma_\tau = I_p$ where $p = 1000$, $\Sigma_{\tau+1} = (\rho^{|i-j|})$ where $0 < \rho < 1$ and $M = 0$. Further, we choose $H = 100$ and obtain $a = 3.58$ by solving the equation in Theorem 1 so that the the average run length is controlled around 5000. We can obtain the minimum ρ for the stopping rule to detect the change is 0.065.

Remark 4 In sequential change-point analysis, an optimal procedure detects the change as soon as possible while maintaining the false alarm at a pre-specified level. There are Bayesian and minimax formulations for investigating the optimality of a detection procedure. Under the two formulations, the optimality of CUSUM and Shiryayev-Roberts procedures has been well studied in univariate and multivariate settings (Shiryayev, 1961, 1963; Lorden, 1971; Pollak, 1985; Moustakides, 1986; Tartakovsky and Veeravalli, 2008; Pollak and Tartakovsky, 2009; Polunchenko and Tartakovsky, 2010). For the high-dimensional online change-point detection problem (1), we employ a formulation proposed by Pollak (1985) which is asymptotically equivalent to the minimax formulation proposed by Lorden (1971). More specifically, an optimal detection procedure T should minimize the supremum average detection delay

$$\sup_{n_0 \leq \tau < \infty} \mathrm{E}_\tau\{T - (\tau - n_0) | T > \tau - n_0\},$$

subject to the constraint that the average run length $\mathrm{E}_\infty(T) \geq \gamma$ where γ is a pre-specified lower bound. Similar to univariate and multivariate settings, a lower bound of the above supremum average detection delay is needed in order to study the optimality of detection procedures. Note that an upper bound of the proposed stopping rule (7) has been specified in Theorem 2. Its performance in detecting covariance changes can therefore be evaluated by comparing the upper bound with the lower bound. Especially, if the upper bound matches the lower bound, the proposed detection procedure would be optimal. However, to the best of our knowledge, such a lower bound has not been established in literature. Nevertheless, there exists a closely related work in Chan (2017) where the author studies a slightly different problem of detecting multivariate normal mean shifts, and shows that the lower bound of the supremum average detection delay varies with the sparsity of data streams undergoing mean shifts as the data dimensionality diverges to infinity. By analogy with Chan (2017), when $\Sigma_{\tau+1}$ differs from Σ_τ in a large number of components, the lower bound of the supremum average detection delay would be trivially given by 1 and the proposed stopping rule may reach the lower bound as it accumulates all the differences through $||\Sigma_{\tau+1} - \Sigma_\tau||_F$. On the other hand, when $\Sigma_{\tau+1}$ differs from Σ_τ only in a sparse number of components, the proposed stopping rule cannot be optimal because the components without the change do not contribute to $||\Sigma_{\tau+1} - \Sigma_\tau||_F$ but to $||\Sigma_\tau||_F$ which leads to a large $||\Sigma_\tau||_F / ||\Sigma_{\tau+1} - \Sigma_\tau||_F$ and thus a long detection delay.

3.3 Training Sample

A training sample primarily provides estimation of unknown nuisance parameters for the proposed stopping rule. Because of its importance, it is worth discussing the availability of the training sample in practice. In many biological studies, prior regulatory networks or pathway information for different biological processes are available through massive data sets (Li and Li, 2008). Such data sets can be used as a training sample if the contained prior knowledge or information matches the initial covariance structure of the considered online detection process. Under other circumstances, a training sample can be historical observations from previous experimental runs subject to similar experimental conditions, after their stationarity of the covariance structure has been confirmed. Suppose $\{X_i, 1 \le i \le n_0\}$ are such historical observations. To check their stationarity in the covariance structure, it is equivalent to considering the hypotheses

$$\begin{aligned} H_0^* &: \Sigma_1 = \cdots = \Sigma_{n_0}, \qquad \text{against} \\ H_1^* &: \Sigma_1 = \cdots = \Sigma_{\tau_1} \neq \Sigma_{\tau_1+1} = \cdots = \Sigma_{\tau_q} \neq \Sigma_{\tau_q+1} = \cdots = \Sigma_{n_0}, \end{aligned} \tag{9}$$

where $1 \le \tau_1 < \cdots < \tau_q < n_0$ are unknown change points. To test the null hypothesis, we consider the test statistic $\hat{\mathcal{J}}_{n_0,M}$ which is obtained by replacing n with n_0 in (3). The rationale of using $\hat{\mathcal{J}}_{n_0,M}$ is that its expectation can distinguish the alternative from the null hypothesis. The following theorem establishes the asymptotic normality of $\hat{\mathcal{J}}_{n_0,M}$.

Theorem 3 Assume (2) and (C1)–(C2). As $n_0 \to \infty$, $(\hat{\mathcal{J}}_{n_0,M} - \mu_{\hat{\mathcal{J}}_{n_0,M}})/\sigma_{\hat{\mathcal{J}}_{n_0,M}}$ converges in distribution to the standard normal $N(0,1)$, where $\mu_{\hat{\mathcal{J}}_{n_0,M}}$ and $\sigma_{\hat{\mathcal{J}}_{n_0,M}}$ are given by Propositions 1 and 2, respectively, with n replaced by n_0. In particular, under H_0^* of (9), $\hat{\mathcal{J}}_{n_0,M}/\hat{\sigma}_{\hat{\mathcal{J}}_{n_0,M},0}$ converges in distribution to the standard normal $N(0,1)$, where $\hat{\sigma}_{\hat{\mathcal{J}}_{n_0,M},0}$ is defined in (8) with H replaced by n_0.

From Theorem 3, we reject H_0^* of (9) with a significance level α if $\hat{\mathcal{J}}_{n_0,M}/\hat{\sigma}_{\hat{\mathcal{J}}_{n_0,M},0} > z_\alpha$, where z_α is the upper α-quantile of the standard normal. Otherwise, we fail to reject H_0^* and hereby obtain a training sample for the proposed stopping rule.

Asymptotic results in Theorem 3 and in Theorem 4 of Section 3.4 require a diverging training sample size n_0 to provide consistent estimation for $\mathrm{tr}\{C(h_1)C(h_2)\}$ in (5) and M in the stopping rule (7). In practice, one can consider a relatively large n_0 for satisfactory performance of the proposed stopping rule. For example, we choose $n_0 = 200$ in the simulation studies of Section 4.

3.4 Stopping Rule with Estimated M

The unknown M in the stopping rule (7) can be estimated through the training sample $X_1, \cdots, X_{n_0}$. From (C1), we know that $\mathrm{cov}(X_i, X_j) = C(i-j)$ is zero if and only if $|i-j| > M$, or equivalently, $\mathrm{tr}\{C(h)C^T(h)\}$ is zero if and only if $|h| > M$. We thus estimate M through the following steps.

- Using (6), we compute $\mathrm{tr}\{\widehat{C(h)C^T}(h)\}/\mathrm{tr}\{\widehat{C(0)C}(0)\}$ with h starting from 0.

- We terminate the process when the first non-negative integer h^* satisfies

$$\frac{\text{tr}\{\widehat{C(h^*)C^T(h^*)}\}}{\text{tr}\{\widehat{C(0)C(0)}\}} \leq \epsilon,$$

where ϵ is a small constant and can be chosen to be 0.05 in practice.

- We then estimate M by $\hat{M} = h^* - 1$.

Let $T_H(a, \hat{M})$ be the stopping rule obtained by replacing M with $\hat{M}$ in (7). The following theorem shows that $T_H(a, \hat{M})$ performs asymptotically as well as $T_H(a, M)$ under both null and alternative hypotheses.

Theorem 4 Assume the same conditions in Theorems 1 and 2. As the training sample size $n_0 \to \infty$,

$$\text{E}_\infty\{T_H(a, \hat{M})\} - \text{E}_\infty\{T_H(a, M)\} \to 0, \quad \text{E}_0\{T_H(a, \hat{M})\} - \text{E}_0\{T_H(a, M)\} \to 0.$$

4. Simulation Studies

In this section we present simulation results to examine the empirical performance of the proposed stopping rule.

4.1 Accuracy of the Theoretical Average Run Length

We first evaluate the performance of the stopping rule under the null hypothesis. The random vectors X_i for $i = 1, 2, \cdots$ are generated from

$$X_i = \sum_{l=0}^{M} \Gamma_l \, \epsilon_{i-l}, \tag{10}$$

where the $p \times p$ matrix $\Gamma_l = \{0.6^{|i-j|}(M - l + 1)^{-1}\}$ for $i, j = 1, \cdots, p$, and $l = 0, \cdots, M$. Each ϵ_i is a p-variate random vector with mean 0 and identity covariance I_p, and all ϵ_is are mutually independent. If $M = 0$, all X_is are mutually independent from (10) and each individual X_i has the covariance matrix $\Gamma_0\Gamma_0^T$. If $M \neq 0$, $\text{cov}(X_i, X_j) = \sum_{l=0}^{M-(i-j)} \Gamma_{i-j+l}\Gamma_l^T$ for $i - j = 0, \cdots, M$. Here we consider the normally distributed ϵ_i. Non-Gaussian ϵ_i can be also considered and the obtained results are similar to those of Gaussian ϵ_i. We choose the dimension $p = 200$, 400 and 1000, the size of historical data $n_0 = 200$, the window size $H = 100$ and 150, and dependence $M = 0, 1, 2$, respectively.

To examine the accuracy of the theoretical average run length, we first specify its value and obtain the corresponding a by solving the equation in Theorem 1. Based on the a, we obtain the Monte Carlo average run length by taking the average of the stopping times from 1000 simulations. Table 1 compares the theoretical average run lengths with the corresponding Monte Carlo average run lengths under different combinations of H, p and M. All the Monte Carlo average run lengths are reasonably close to the theoretical average run lengths, subject to some random variations from simulations under different M and p.

	$H = 100$								
Theoretical	$p = 200$			$p = 400$			$p = 1000$		
(a, ARL)	$M = 0$	1	2	$M = 0$	1	2	$M = 0$	1	2
$(3.04, 1002)$	1178	1151	1194	1245	1284	1317	1302	1295	1335
$(3.42, 3008)$	3067	3148	2986	3690	3614	3529	3850	3954	3617
$(3.58, 5038)$	5118	4527	4253	5799	5923	5212	6570	6102	5878
	$H = 150$								
	$p = 200$			$p = 400$			$p = 1000$		
	$M = 0$	1	2	$M = 0$	1	2	$M = 0$	1	2
$(2.88, 1005)$	1044	1127	1149	1069	1173	1308	1145	1198	1270
$(3.29, 3033)$	3240	3156	3202	3505	3795	3628	3652	3759	3931
$(3.46, 5118)$	5120	5097	5280	6083	5820	6156	6162	6586	6794

Table 1: The comparison between theoretical average run lengths and Monte Carlo average run lengths based on 1000 simulations. For each average run length and window size H, the threshold a is obtained by solving the equation in Theorem 1.

4.2 Accuracy of the Upper Bound for Expected Detection Delay

We next evaluate the performance of the stopping rule under the alternative hypothesis. In particular, we examine the accuracy of the upper bound for the expected detection delay in Theorem 2. In the simulation studies, we consider an immediate change, namely the change occurring immediately after the historical data of size $n_0 = 200$. Before the change point τ, the observations X_i for $i = 1, \cdots, 200$ are generated from (10) where $\Gamma_l = I(M - l + 1)^{-1}$ with I being the identity matrix. After the change, $\Gamma_l = Q(M - l + 1)^{-1}$ in (10) where the $p \times p$ matrix Q is modeled by one of the following patterns.

(a). Q satisfies $QQ^T = \Sigma$, where $\Sigma_{ij} = \rho^{|i-j|}$ for $1 \le i, j \le p$.

(b). Each row of Q has only three non-zero elements that are randomly chosen from $\{1, \cdots, p\}$ with magnitude ρ multiplied by a random sign.

(c). Q satisfies $QQ^T = \Sigma$, where $\Sigma_{ii} = 1$ for $i = 1, \cdots, p$, and $\Sigma_{ij} = \rho$ for $i \ne j$.

Models (a)–(c) specify the bandable, sparse and strong covariance matrices, respectively. We choose $\rho = 0.6, 0.7, 0.8$ to obtain different magnitudes in the covariance change, and choose the dimension $p = 1000$, the window size $H = 100$ and 150, and dependence $M = 0, 1, 2$, respectively. Moreover, the threshold $a = 3.58$ when $H = 100$ and $a = 3.46$ when $H = 150$ so that the theoretical average run length is controlled around 5000. Table 2 compares the theoretical bound for the expected detection delay in Theorem 2 with the corresponding Monte Carlo expected detection delay based on 1000 simulations. As we can see, each Monte Carlo expected detection delay is no more than its theoretical upper bound. Furthermore, both Monte Carlo expected detection delays and theoretical bounds decrease as ρ increases with the same M and H, but increase as M increases with the same ρ and H. The simulation results are consistent with the theoretical findings in Theorem 2.

		$\rho = 0.6$			0.7			0.8		
		$M = 0$	1	2	$M = 0$	1	2	$M = 0$	1	2
		Model (a)								
$H = 100$	Monte Carlo	16.18	20.14	24.04	11.31	14.34	16.98	8.11	10.31	12.44
	Theoretical	20.59	23.63	25.99	16.23	18.79	20.83	12.46	14.61	16.38
$H = 150$	Monte Carlo	17.49	21.62	25.45	12.34	15.38	18.56	8.90	11.32	13.37
	Theoretical	24.36	28.10	31.04	19.11	22.21	24.70	14.59	17.13	19.22
		Model (b)								
$H = 100$	Monte Carlo	4.36	5.84	7.16	3.58	4.71	5.87	3.10	4.13	5.06
	Theoretical	7.42	9.09	10.58	6.07	7.40	8.76	5.11	6.42	7.60
$H = 150$	Monte Carlo	4.79	6.38	7.68	3.85	5.13	6.58	3.27	4.50	5.32
	Theoretical	8.53	10.38	11.92	6.80	8.45	9.88	5.74	7.19	8.45
		Model (c)								
$H = 100$	Monte Carlo	2.84	3.90	4.94	2.68	3.68	4.78	2.63	3.69	4.72
	Theoretical	3.04	4.15	6.23	2.89	3.99	5.05	2.78	3.87	4.92
$H = 150$	Monte Carlo	2.96	3.94	5.09	2.89	3.93	4.91	2.72	3.76	4.84
	Theoretical	3.25	4.40	5.51	3.07	4.20	5.30	2.94	4.05	5.13

Table 2: The comparison between theoretical upper bounds for expected detection delays and Monte Carlo expected detection delays based on 1000 simulations with the average run length controlled around 5000.

4.3 Accuracy of the Data-Driven Procedure for M

We also examine the data-driven procedure proposed in Section 3.4 for estimating M. For each simulation, a training sample of 200 observations is generated from (10) with $p = 1000$, where the 1000×1000 matrix $\Gamma_l = \{0.6^{|i-j|}(M - l + 1)^{-1}\}$ for $i, j = 1, \cdots, 1000$, and $l = 0, \cdots, M$. We terminate the estimating procedure when $\mathrm{tr}\{\widehat{C(h)C^T}(h)\}/\mathrm{tr}\{\widehat{C(0)C}(0)\}$ is less than or equal to a small constant $\epsilon = 0.02$. From (10), $\mathrm{tr}\{\widehat{C(h)C^T}(h)\}$ estimates $\mathrm{tr}\{C(h)C^T(h)\} = \sum_{l,q=0}^{M-h} \mathrm{tr}(\Gamma_{q+h}^T\Gamma_{l+h}\Gamma_l^T\Gamma_q)$ and $\mathrm{tr}\{\widehat{C(0)C}(0)\}$ estimates $\mathrm{tr}\{C(0)C(0)\} = \sum_{l,q=0}^{M} \mathrm{tr}(\Gamma_q^T\Gamma_l\Gamma_l^T\Gamma_q)$. Figure 1 illustrates the histograms of selected M based on 100 simulations when the actual $M = 0, 1, 2$ and 3. The proposed data-driven procedure demonstrates its satisfactory performance for estimating the M, while the probability of identifying the correct M becomes smaller as M gets greater. This is because $\mathrm{tr}\{C(0)C(0)\} = \sum_{l,q=0}^{M} \mathrm{tr}(\Gamma_q^T\Gamma_l\Gamma_l^T\Gamma_q)$ can be much greater than $\mathrm{tr}\{C(M)C^T(M)\} = \mathrm{tr}(\Gamma_M^T\Gamma_M\Gamma_l^T\Gamma_q)$ when M is larger. For example, if $M = 3$, $\mathrm{tr}\{C(M)C^T(M)\}/\mathrm{tr}\{C(0)C(0)\} \approx 1/16$, which is slightly greater than $\epsilon = 0.02$. As a result, the data-driven procedure in Section 3.4 could end up with $h^* = M$ or estimates M by $h^* - 1 = 2$. As demonstrated in the last panel of Figure 1, 70 out of 100 simulations correctly identify $M = 3$, but 23 out of 100 simulations identity 2.

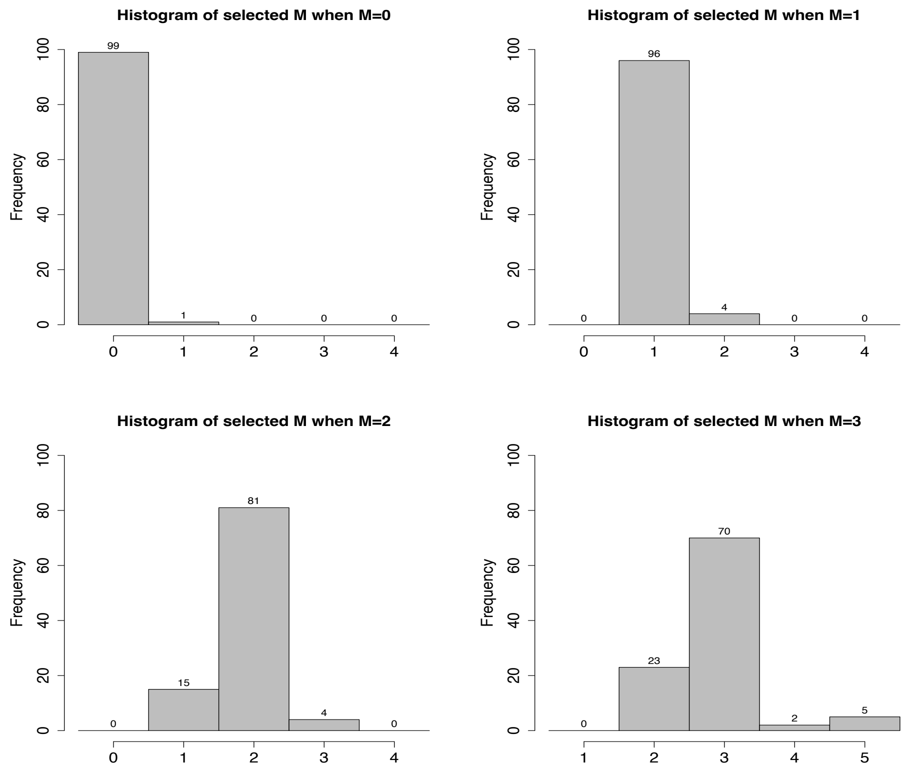

Figure 1: Histograms of selected M by the proposed data-driven procedure when the actual $M = 0, 1, 2$ and 3. The results are based on 100 simulations.

4.4 Comparison with Other Methods

In the last part of simulation studies, we compare our proposed online change-point detection method with the one in Avanesov (2019). Note that Avanesov (2019) assumes temporal independence of data. We consider the following simulation setup similar to Avanesov (2019). Without any change point, there are 520 vectors which are independently generated from the multivariate normal $\mathrm{N}(0, I_p)$. With a change point, there are 400 vectors before the change point which are independently generated from the multivariate normal $\mathrm{N}(0, I_p)$. There are 120 vectors after the change point which are independently generated from $\mathrm{N}(0, \Sigma)$, where Σ is the inverse of the tridiagonal matrix with diagonal elements equal to 1 and off-diagonal elements equal to 0.4. The dimensionality of data is $p = 50$.

To detect the change point, our method uses an additional historical data of size $n_0 = 200$ and a fixed window size $H = 100$. The method in Avanesov (2019) uses multiple window sizes $(15, 30, 60)$. Another difference is that the method in Avanesov (2019) controls the probability $\alpha = 0.05$ to raise a false alarm among the simulated 520 observations, but our method controls the average run length. To compare the two methods based on the same criterion, we obtain the threshold $a = 3.95$ in the stopping rule (7) through 1000 Monte

Carlo simulations, such that the probability of the false alarm is $\alpha = 0.05$ when there is no change point in the covariance structure among the 520 observations. The threshold in Avanesov (2019) for $\alpha = 0.05$ is obtained through a bootstrap calibration scheme. Based on the 1000 simulations, the power of our method is 1 and the Monte Carlo expected detection delay is 12.4. The power is greater than Avanesov (2019) which is 0.82 and the Monte Carlo expected detection delay is smaller than Avanesov (2019) which is 41.1.

5. Case Study

Resting-state fMRI is a method to explore brain's internal dynamic networks. We apply the proposed method to a resting-state fMRI data set obtained from the 2017 Human Connectome Project (HCP) data release. The data consist of 300 independent component analysis (ICA) component nodes ($p = 300$) repeatedly observed over 1200 time points, collected for each of 1003 subjects. The publicly accessible data set together with details about data acquisition and preprocessing procedures can be found in HCP website (http://www.humanconnectome.org).

We detect the change in a real-time manner, in the sense that we pretend the observations in the data set continually arrive in time. At each time, we determine whether the process should be terminated through the proposed stopping rule. Note that the proposed stopping is designed only for detecting the covariance change. When a detection process involves a change in the mean, it cannot be detected by the proposed stopping rule. Despite such a limitation, we still apply the stopping rule to the data set for the covariance change as the main interest of using the resting-state fMRI is to study the dynamic nature of brain connectivity (Cribben et al., 2013; Jeong et al., 2016).

While there are 1003 subjects in the data set, we randomly choose two subjects 103010 and 130417 to demonstrate the practical usefulness of the proposed method. The proposed stopping rule needs a training sample. We pretend that the first 200 observations of each time series are historical, and further justify their stationarity in the covariance structure through the testing procedure in Section 3.3. Here we use relatively large training sample size 200 to attain precise estimation of nuisance parameters. Based on the training sample, we estimate M by 5 for the subject 103010 and 6 for the subject 130417 using the method in Section 3.4 and obtain the sample mean $\hat{\mu}$ and the sample standard deviation of the test statistic using (8) in Section 2.3. Choosing the threshold $a = 3.58$ so that the average run length is controlled around 5000, we apply the proposed stopping rule with the window size $H = 100$ and terminate the process at the time 287 for the subject 103010 and the time 245 for the subject 130417.

With each of the stopping times 287 and 245, we pull out the observations from time 1 to the stopping time and conduct some post analyses. The first analysis is change-point estimation. Similar to Bai (2010), the change point is estimated by

$$\hat{\tau} = \arg \max_{1<t<T_H(a,\hat{M})} \hat{\mathcal{J}}_{t,\hat{M},H},$$

where $\hat{\mathcal{J}}_{t,\hat{M},H}$ is obtained by replacing $W_M(i,j)$ in $\hat{\mathcal{J}}_{n,M,H}$ with $A_{t,\hat{M}}(i,j)$ defined in (3). The rationale of using the above estimator is that the expectation of $\hat{J}_{t,\hat{M},H}$ always attains its maximum at the true change point, as mentioned in Remark 2 of Section 2.2. The

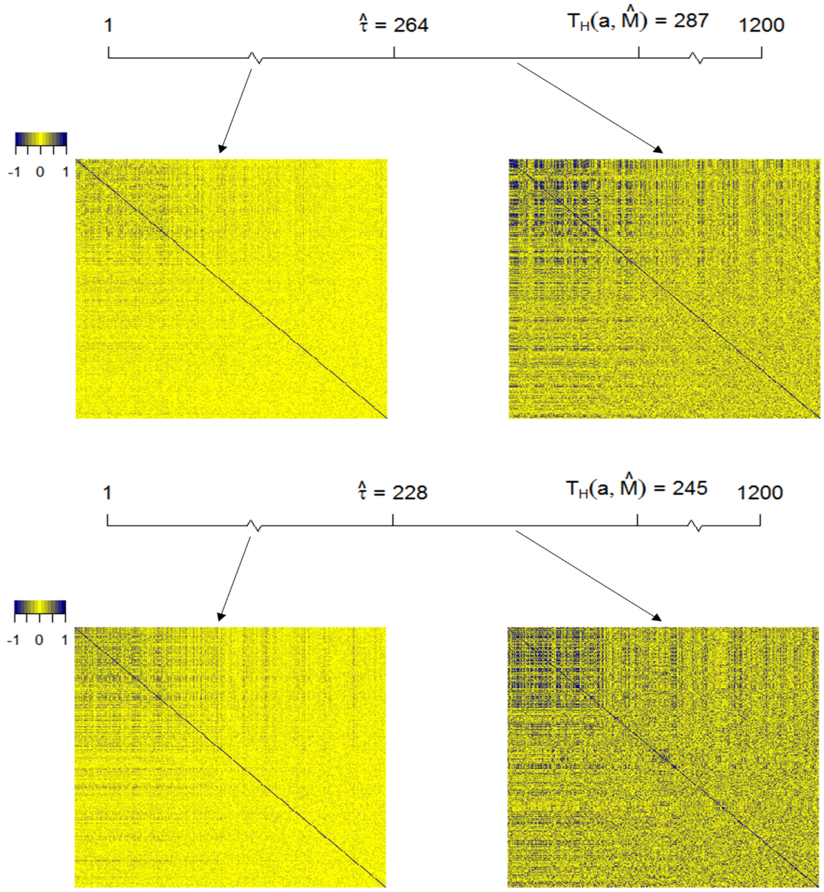

Figure 2: Online change-point detection in the covariance structure of subject 103010 (upper panel) and subject 130417 (lower panel). Each panel illustrates the estimated correlation matrices before and after the estimated change point.

estimated change points are 264 for the subject 103010 and 228 for the subject 130417. With the two stopping times 287 and 245, the corresponding detection delays are 23 for the subject 103010 and 17 for the subject 130417, showing that the proposed stopping rule can quickly terminate the process when the brain's network change occurs.

The second analysis is illustrating the actual change in the brain's network. For each subject, we estimate the correlation matrices before and after the estimated change point using the *glasso*. The obtained results for the two subjects are summarized in Figure 2, which clearly illustrates the brain's internal networks become stronger after the estimated change points. The results are consistent with recent studies that during the resting state, brain's networks activate when a subject focuses on internal tasks, and exhibit dynamic changes within time scales of seconds to minutes (Allen et al., 2014; Calhoun et al., 2014;

Chang and Glover, 2010; Cribben et al., 2012; Handwerker et al., 2012; Hutchison et al., 2013; Jeong et al., 2016; Monti et al., 2014).

6. Discussion

We propose a new procedure to detect the anomaly in the covariance structure of high-dimensional online data. The procedure is implementable when data are non-Gaussian, and involve both spatial and temporal dependence. We investigate its theoretical properties by deriving an explicit expression for the average run length and an upper bound for the expected detection delay. The established average run length can be employed to obtain the level of the threshold in the stopping rule without running time-consuming Monte Carlo simulations. The derived upper bound demonstrates the impact of data dependence and magnitude of change in the covariance structure on the expected detection delay. The theoretical properties are examined and justified by the empirical studies through both simulation and a real application.

The proposed test statistic (3) takes the Frobenius norm to measure the magnitude of change in the covariance matrix. Since the Frobenius norm sums the squared elements of a matrix, the proposed stopping rule would be advantageous for detecting a change point if changes of the covariance matrix happen in a large number of elements. On the other hand, choosing the Frobenius norm would not be ideal if changes of the covariance matrix happen in a sparse number of elements. This can be seen from Theorem 2 demonstrating that the elements without undergoing changes do not contribute to $||\Sigma_{\tau+1} - \Sigma_\tau||_F$ but to $||\Sigma_\tau||_F$ which leads to a large $||\Sigma_\tau||_F / ||\Sigma_{\tau+1} - \Sigma_\tau||_F$ and thus a long detection delay. One way to reduce the detection delay is to replace the test statistic by

$$\hat{\mathcal{J}}_{n,M,H} = \frac{2}{H^2} \sum_{1 \le k \le l \le p} \sum_{i,j=n-H+1}^{n} W_M(i,j) Y_{i,kl} Y_{j,kl},$$

where $Y_{i,kl} = X_{i,k} X_{i,l}$ and $X_{i,k}$ is the kth component of the p-dimensional random vector X_i, and remove the elements $Y_{i,kl} Y_{j,kl}$ with no change. Another way is to choose different matrix norms. For instance, the statistics in Avanesov and Buzun (2018) and Avanesov (2019) estimate the precision matrix and covariance matrix before and after a time point and then take the sup-norm to measure the difference between the two matrices. Based on a similar idea, one can choose the operator norm and define a statistic

$$L_{n,H}(t) = ||\hat{\Sigma}^l_{n,H}(t) - \hat{\Sigma}^r_{n,H}(t)||_2$$

where $||\cdot||_2$ is the matrix operator norm, and $\hat{\Sigma}^l_{n,H}(t)$ and $\hat{\Sigma}^r_{n,H}(t)$ are estimated covariance matrices before and after a chosen time t from the observations $\{X_i, n-H+1 \le i \le n\}$ within a window of size H. Since the sup-norm is the largest element of a matrix in absolute value and the operator norm is the largest eigenvalue in absolute value when a matrix is symmetric, stopping rules based on the sup-norm or operator norm would be advantageous if changes of the covariance matrix happen in a sparse number of elements.

In univariate settings, the two classical methods for sequential change-point detection are the CUSUM (Page, 1954; Lorden, 1971) and Shiryayev-Roberts procedures (Shiryayev, 1963; Roberts, 1966), where the former is based on the max-type likelihood ratio statistic and the latter is based on the sum-type likelihood ratio statistic. There is a preference for the Shiryayev-Roberts procedure when the change in magnitude is weak and the change occurs at large values, and conversely the CUSUM procedure when the change in magnitude is strong and the change occurs immediately after the training sample. In high-dimensional settings, in addition to the stopping rule (7) based on the sum-type U-statistic (3), we may also consider a stopping rule based on a max-type U-statistic

$$T_H^*(b, M) = \inf\left\{n - n_0 : \max_{n-M-H-2\le t\le n-M-2} \left|\frac{\hat{\mathcal{J}}_{t,M,H}}{\hat{\sigma}_{\hat{\mathcal{J}}_{t,M,H},0}}\right| > b,\ n > n_0\right\}, \tag{11}$$

where the statistic $\hat{\mathcal{J}}_{t,M,H}$ can be obtained by replacing $W_M(i,j)$ in $\hat{\mathcal{J}}_{n,M,H}$ with $A_{t,M}(i,j)$ defined in (3). Since the two stopping rules (7) and (11) are analogous to the Shiryayev-Roberts and CUSUM procedures respectively, our future research is to extend the superior and inferior aspects of the sum-type and max-type stopping rules from univariate settings to high-dimensional settings.

Except for covariance change, high-dimensional online data may also encounter mean change. Another interesting problem is to consider the hypotheses

$$\begin{aligned} H_0: &\quad \mu_1 = \mu_2 = \cdots \quad \text{and} \quad \Sigma_1 = \Sigma_2 = \cdots \quad \text{against} \\ H_1: &\quad \mu_1 = \cdots = \mu_{\tau_1} \neq \mu_{\tau_1+1} = \cdots \quad \text{or} \quad \Sigma_1 = \cdots = \Sigma_{\tau_2} \neq \mu_{\tau_2+1} = \cdots\ , \end{aligned}$$

where τ_1 and τ_2 are unknown change points for the mean and covariance, respectively. To test the above hypotheses, we may consider a test statistic

$$\frac{\hat{\mathcal{J}}_{n,M,H}}{\hat{\sigma}_{\mathcal{J},n_0,M,H}} + \frac{\hat{\mathcal{L}}_{n,M,H}}{\hat{\sigma}_{\mathcal{L},n_0,M,H}},$$

where the first term is the proposed test statistic in (7) for the covariance change. The second term is proposed for the mean change with

$$\hat{\mathcal{L}}_{n,M,H} = \frac{1}{H}\sum_{i,j=n-H+1}^{n} W_M(i,j) X_i^T X_j,$$

where the weight function $W_M(i,j)$ is defined in (3). Similar to Proposition 1, we can show that for a given n and H, $\mathrm{E}(\hat{\mathcal{L}}_{n,M,H}) = H^{-1}\sum_{i,j=n-H+1}^{n} W_M(i,j)\mu_i^T\mu_j$ which is zero when there is no change for the mean but positive otherwise. Moreover, by analogy with (8), the estimated variance of $\hat{\mathcal{L}}_{n,M,H}$ is

$$\hat{\sigma}^2_{\mathcal{L},n_0,M,H} = \frac{2}{H^2}\sum_{i,j=1}^{H}\sum_{h_1,h_2} W_M(i,j)W_M(i-h_1, j+h_2)\mathrm{tr}\{\widehat{C(h_1)C}(h_2)\}.$$

The rational of considering the sum of $\hat{\mathcal{J}}_{n,M,H}/\hat{\sigma}_{\mathcal{J},n_0,M,H}$ and $\hat{\mathcal{L}}_{n,M,H}/\hat{\sigma}_{\mathcal{L},n_0,M,H}$ is to detect the change of mean and covariance matrix simultaneously. The corresponding stopping rule analogous to (7) can be proposed as

$$T_H^*(a, M) = \inf\left\{n - n_0 : \left|\frac{\hat{\mathcal{J}}_{n,M,H}}{\hat{\sigma}_{\mathcal{J},n_0,M,H}} + \frac{\hat{\mathcal{L}}_{n,M,H}}{\hat{\sigma}_{\mathcal{L},n_0,M,H}}\right| > a^*,\ n > n_0\right\},$$

where the value of threshold a^* can be determined by establishing its explicit relationship with the average run length. Due to a similar structure of $T_H^*(a, M)$ to (7), we expect that its average run length and detection delay would be similar to the results in Theorems 1 and 2.

Acknowledgements

The authors thank the Editor Prof. Harchaoui and anonymous referees for their constructive comments which have significantly improved the quality of the paper. The authors are partially supported by NSF grant DMS-1916239, and NIH grant R15GM139115.

Appendix A. Proofs

In this appendix we prove the propositions and theorems of the paper.

A. 1. Proof of Proposition 1

From (3), X_i and X_j in $\hat{\mathcal{J}}_{n,M}$ are M apart because of the indicator function in $W_M(i,j)$. From (C1), X_i and X_j are independent. Using the model (2), we obtain $\mathrm{E}(\hat{\mathcal{J}}_{n,M}) = n^{-2}\sum_{i,j} W_M(i,j)\mathrm{tr}(\Sigma_i\Sigma_j)$, which is the expectation under the alternative hypothesis. Under the null hypothesis, $\mathrm{E}(\hat{\mathcal{J}}_{n,M}) = \mathrm{tr}(\Sigma^2)n^{-2}\sum_{i,j} W_M(i,j) = 0$ because $\sum_{i,j} W_M(i,j) = 0$. ■

A. 2. Proof of Proposition 2

Note that $\mathrm{var}(\hat{\mathcal{J}}_{n,M}) = \mathrm{E}(\hat{\mathcal{J}}_{n,M}^2) - \mathrm{E}^2(\hat{\mathcal{J}}_{n,M})$, where $\mathrm{E}^2(\hat{\mathcal{J}}_{n,M})$ can be obtained from Proposition 1. We thus only need to derive $\mathrm{E}(\hat{\mathcal{J}}_{n,M}^2)$, which, from (2), is

$$\begin{aligned}
\mathrm{E}(\hat{\mathcal{J}}_{n,M}^2) &= \frac{1}{n^4}\sum_{i,j}\sum_{k,l} W_M(i,j)W_M(k,l)\mathrm{tr}(\Sigma_i\Sigma_j)\mathrm{tr}(\Sigma_k\Sigma_l) + \frac{4}{n^4}\sum_{i,j}\sum_{k,l} W_M(i,j)W_M(k,l) \\
&\times \mathrm{tr}^2\{C(i-k)C(l-j)\} + \frac{1}{n^4}\sum_{i,j}\sum_{k,l} W_M(i,j)W_M(k,l) \\
&\times \Big[16\mathrm{tr}\{\Sigma_l C(i-k)\Sigma_j C(k-i)\} + 4\mathrm{tr}\{C(k-i)C(j-l)C(k-i)C(j-l)\} \\
&+ 8\beta\mathrm{tr}(\Gamma_i^T\Gamma_j\Gamma_j^T\Gamma_i \circ \Gamma_k^T\Gamma_l\Gamma_l^T\Gamma_k) + 8\beta\mathrm{tr}(\Gamma_i^T\Gamma_j\Gamma_k^T\Gamma_l \circ \Gamma_i^T\Gamma_j\Gamma_k^T\Gamma_l) \\
&+ \sum_m 2\beta^2\mathrm{tr}(\Gamma_j^T\Gamma_i e_m e_m^T\Gamma_k^T\Gamma_l \circ \Gamma_j^T\Gamma_i e_m e_m^T\Gamma_k^T\Gamma_l)\Big],
\end{aligned}$$

where for any square matrices A and B, the symbol $A \circ B = (a_{ij}b_{ij})$, and e_m is the unit vector with the only non-zero element at the mth component. Applying (C2) and subtracting $\mathrm{E}^2(\hat{\mathcal{J}}_{n,M})$ in Proposition 1, we have

$$\mathrm{E}(\hat{\mathcal{J}}_{n,M}^2) = \frac{4}{n^4}\sum_{i,j}\sum_{k,l} W_M(i,j)W_M(k,l)\mathrm{tr}^2\{C(i-k)C(l-j)\}\{1+o(1)\}.$$

■

A. 3. Proof of Theorem 1

We need to derive the cumulative distribution function of $T_H(a,M)$. From (7),

$$\mathrm{P}_\infty\{T_H(a,M) \le t\} = \mathrm{P}_\infty\left(\max_{0\le i\le t}\left|\frac{\hat{\mathcal{J}}_{n_0+i,M,H}}{\hat{\sigma}_{n_0,M,H}}\right| > a\right).$$

The cumulative distribution function of $T_H(a,M)$ thus depends on that of $\hat{\mathcal{J}}_{n_0+i,M,H}/\hat{\sigma}_{n_0,M,H}$, which will be shown to converge to a stationary Gaussian process.

To simplify notation, let $\hat{\mathcal{J}}_{n_0+i,M,H} \equiv \hat{\mathcal{J}}_{i,M}$, and $\hat{\sigma}_{n_0,M,H} \equiv \hat{\sigma}_0$. The Gaussian process is established by showing (i) the joint asymptotic normality of $(\hat{\sigma}_0^{-1}\hat{\mathcal{J}}_{i_1,M},\dots,\hat{\sigma}_0^{-1}\hat{\mathcal{J}}_{i_d,M})'$ for any $i_1 < i_2 < \cdots < i_d$. (ii) the tightness of $\hat{\sigma}_0^{-1}\hat{\mathcal{J}}_{i,M}$. To prove (i), we apply the Cramér-Wold device to show that for any non-zero $a_1,\cdots,a_d$, $\sum_{l=1}^{d}\hat{\sigma}_0^{-1}a_l\hat{\mathcal{J}}_{i_l,M}$ is asymptotic normal. Since the proof is similar to that of Theorem 3, we omit it. We thus only need to prove (ii).

Toward this end, we first obtain the leading order of $\mathrm{var}(\hat{\mathcal{J}}_{i,M})$. The following lemma is proved in Section A. 4.

Lemma 1 *As $H \to \infty$,*

$$\begin{aligned} var(\hat{\mathcal{J}}_{i,M}) &= \frac{4}{H^4}\sum_{i,j=1}^{H}\sum_{h_1,h_2} W_M(i,j)W_M(i-h_1,j+h_2)tr^2\{C(h_1)C(h_2)\} \\ &= \frac{4}{H^4}\sum_{h_1,h_2} tr^2\{C(h_1)C(h_2)\}\left(\frac{6\pi^2-51}{18}H^4\right)\{1+o(1)\}. \end{aligned}$$

Let $i_1, i_2 \in \{1,\dots,t\}$ and $i_d \equiv i_2 - i_1$. We next consider $\mathrm{cov}(\hat{\mathcal{J}}_{i_1,M},\hat{\mathcal{J}}_{i_2,M})$, which equals $\mathrm{E}(\hat{\mathcal{J}}_{i_1,M}\hat{\mathcal{J}}_{i_2,M})$ when there is no any change point. For $i_d \in \{1,\dots,H-1\}$ and $H - i_d = O(H)$, the leading order of $\mathrm{cov}(\hat{\mathcal{J}}_{i_1,M},\hat{\mathcal{J}}_{i_2,M})$ depends on i_d. Following similar derivations of $\mathrm{var}(\hat{\mathcal{J}}_{i,M})$, we can obtain that under Under (C1) and as $H \to \infty$,

$$\mathrm{cov}(\hat{\mathcal{J}}_{i_1,M},\hat{\mathcal{J}}_{i_2,M}) = \frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\left\{\frac{6\pi^2-51}{18}(H-i_d)^4\right\}\{1+o(1)\}.$$

For $i_d \in \{1,\dots,H-1\}$ and $H - i_d = o(H)$, or for $i_d \geq H$, $\mathrm{cov}(\hat{\mathcal{J}}_{i_1,M},\hat{\mathcal{J}}_{i_2,M})$ can be shown is the smaller order of $\mathrm{var}(\hat{\mathcal{J}}_{i,M})$, i.e.

$$\mathrm{cov}(\hat{\mathcal{J}}_{i_1,M},\hat{\mathcal{J}}_{i_2,M}) = o\left[\frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\left(\frac{6\pi^2-51}{18}H^4\right)\right].$$

We need to show the tightness of $\sigma_0^{-1}\hat{\mathcal{J}}_{i,M}$. Then the tightness of $\hat{\sigma}_0^{-1}\hat{\mathcal{J}}_{i,M}$ can be established by the Slutsky theorem because $\hat{\sigma}_0$ is ratio-consistent to σ_0 according to Theorem 3. Consider $i = q^*\cdot t$, for $q^* = i/t \in (0,1)$, with $i = 1,\dots,t$. It is equivalent to show the tightness of $G(i/t)$, where $G(i/t) = G(q^*) = \sigma_0^{-1}\hat{\mathcal{J}}_{i,M}$. For $0 < q^* < r^* < 1$,

$$\begin{aligned} \mathrm{E}|G(r^*) - G(q^*)|^2 &= \sigma_0^{-1}\mathrm{E}|\hat{\mathcal{J}}_{i_1,M} - \hat{\mathcal{J}}_{i_1,M}|^2 \\ &= \sigma_0^{-1}\{\mathrm{E}(\hat{\mathcal{J}}^2_{i_1,M}) + \mathrm{E}(\hat{\mathcal{J}}^2_{i_2,M}) - 2\mathrm{E}(\hat{\mathcal{J}}_{i_1,M}\hat{\mathcal{J}}_{i_2,M})\}. \end{aligned}$$

When there is no any change point,

$$\begin{aligned} &\mathrm{E}(\hat{\mathcal{J}}^2_{i_1,M}) = \mathrm{E}(\hat{\mathcal{J}}^2_{i_2,M}) = \mathrm{var}(\hat{\mathcal{J}}_{i,M}) \\ = \;& \frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\left(\frac{6\pi^2-51}{18}H^4\right)\{1+o(1)\}. \end{aligned}$$

For any $i_1, i_2 \in \{1, \ldots, t\}$, and $i_2 - i_1 = i_d \in \{1, \ldots, H-1\}$, as $H \to \infty$,

$$\begin{aligned} \mathrm{E}|G(r^*) - G(q^*)|^2 &\leq C \frac{(4/H^4)\sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} 2\{H^4 - (H - i_d)^4\}}{(4/H^4)\sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} H^4} \\ &\leq C\left(\frac{i_d}{H}\right). \end{aligned}$$

For $i_d \geq H$,

$$\mathrm{E}|G(r^*) - G(q^*)|^2 \leq \frac{C(4/H^4)\sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} 2\{H^4 + o(H^4)\}}{(4/H^4)\sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} H^4} \leq C.$$

Therefore, by Chebyshev's inequality, if $1 \leq i_d \leq H-1$,

$$\mathrm{P}(|G(r^*) - G(q^*)| \geq \lambda) \leq \frac{\mathrm{E}|G(r^*) - G(q^*)|^2}{\lambda^2} \leq (C/\lambda^2)(i_d/H).$$

Let $H/t = d$, then

$$i_d/H = (i_2 - i_1)/H = (r^* - q^*)t/H = (r^* - q^*)/d,$$

and $\{(r^* - q^*)/d\} \in (0,1)$. If $i_d \geq H$, or equivalently $\{(r^* - q^*)/d\} \geq 1$,

$$\mathrm{P}(|G(r^*) - G(q^*)| \geq \lambda) \leq \frac{\mathrm{E}|G(r^*) - G(q^*)|^2}{\lambda^2} \leq C/\lambda^2.$$

Let

$$f_d\{(q^*, r^*]\} = \begin{cases} (r^* - q^*)/d, & if \quad r^* - q^* < d \\ 1, \quad if \quad r^* - q^* \geq d, \end{cases}$$

then

$$\mathrm{P}(|G(r^*) - G(q^*)| \geq \lambda) \leq (C/\lambda^2) f_d\{(q^*, r^*]\}.$$

Let $\xi_i = G(i/t) - G\{(i-1)/m\}$, for $i = 1, \ldots, t$. Then $S_i = \xi_1 + \cdots + \xi_i = G(i/t)$ with $S_0 = 0$. Therefore,

$$\mathrm{P}(|S_{i_2} - S_{i_1}| \geq \lambda) \leq (C/\lambda^2) f_d\{(q^*, r^*]\}.$$

For any $0 < p^* < q^* < r^* < 1$, $G(p^*) = S_{i_0}$, $G(q^*) = S_{i_1}$ and $G(r^*) = S_{i_2}$, respectively. Let $m^* = |G(q^*) - G(p^*)| \wedge |G(r^*) - G(q^*)|$. Then

$$\begin{aligned} \mathrm{P}(m^* \geq \lambda) &= \mathrm{P}\big[\{|G(q^*) - G(p^*)| \geq \lambda\} \cap \{|G(r^*) - G(q^*)| \geq \lambda\}\big] \\ &\leq \mathrm{P}^{1/2}(|S_{i_1} - S_{i_0}| \geq \lambda) \cdot \mathrm{P}^{1/2}(|S_{i_2} - S_{i_1}| \geq \lambda) \\ &\leq (C/\lambda) f_d^{1/2}\{(p^*, q^*]\}(C/\lambda) f_d^{1/2}\{(q^*, r^*]\} \\ &\leq (C/\lambda^2)[f_d\{(p^*, q^*]\} + f_d\{(q^*, r^*]\}]. \end{aligned}$$

If $q^* - p^* < d$ and $r^* - q^* < d$, or equivalently $r^* - p^* < 2d$,

$$\mathrm{P}(m^* \geq \lambda) \leq (C/\lambda^2)\left\{\frac{q^* - p^*}{d} + \frac{r^* - q^*}{d}\right\} \leq (C/\lambda^2)\left(\frac{r^* - p^*}{d}\right).$$

If $q^* - p^* < d$ and $r^* - q^* \geq d$, or $q^* - p^* \geq d$ and $r^* - q^* < d$, but $r^* - p^* < 2d$,

$$\begin{aligned} \mathrm{P}(m^* \geq \lambda) &\leq (C/\lambda^2)\left\{\frac{q^* - p^*}{d} + 1\right\} \\ &\leq (C/\lambda^2)\left(\frac{q^* - p^*}{d} + \frac{r^* - q^*}{d}\right) \leq (C/\lambda^2)\left(\frac{r^* - p^*}{d}\right). \end{aligned}$$

If $q^* - p^* < d$ and $r^* - q^* \geq d$, or $q^* - p^* \geq d$ and $r^* - q^* < d$, but $r^* - p^* \geq 2d$,

$$\mathrm{P}(m^* \geq \lambda) \leq (C/\lambda^2)\left\{\frac{q^* - p^*}{d} + 1\right\} \leq 2C/\lambda^2.$$

If $q^* - p^* \geq d$ and $r^* - q^* \geq d$, and $r^* - p^* \geq 2d$,

$$\mathrm{P}(m^* \geq \lambda) \leq 2C/\lambda^2.$$

Let

$$\mu_{\alpha,d}\{(p^*, r^*]\} = \begin{cases} (\frac{r^*-p^*}{d})^{\frac{1}{2\alpha}}, & if \quad r^* - p^* < 2d \\ 2^{\frac{1}{2\alpha}}, & if \quad r^* - p^* \geq 2d, \end{cases}$$

where $\alpha > \frac{1}{2}$. Then $\mu_{\alpha,d}\{(p^*, r^*]\}$ is a finite measure on $T = (0, 1]$. For any $\epsilon > 0$ and $p^*, q^*, r^* \in T = (0, 1]$,

$$\mathrm{P}(m^* \geq \lambda) \leq (C/\lambda^2)\mu_{\alpha,d}^{2\alpha}\{(p^*, r^*]\}.$$

Let

$$L(G) = \sup_{p^* \leq q^* \leq r^*} m^* = \max_{i_0 \leq i_1 \leq i_2} |S_{i_1} - S_{i_0}| \wedge |S_{i_2} - S_{i_1}|.$$

Using Theorem 10.3 in Bilingsley (1999), we conclude

$$P\{L(G) \geq \lambda\} \leq \frac{KC}{\lambda^2}\mu_{\alpha,d}^{2\alpha}\{(0, 1]\},$$

where K is a constant. As $t \gg H$, $d = H/t$ is close to zero, and $2d < (1 - 0)$. Hence, $\mu_{\alpha,d}^{2\alpha}\{(0, 1]\} = 2$, and

$$P\{L(G) \geq \lambda\} \leq \frac{2KC}{\lambda^2}.$$

From (10.4) in Bilingsley (1999), we obtain

$$\max_{1 \leq i \leq t} |S_i| \leq L(G) + |S_t|.$$

Since $\mathrm{E}|S_t|^2 = \sigma_0^{-2}\mathrm{E}(\hat{\mathcal{J}}_{t,M}^2) = 1$, we have

$$\begin{aligned} \mathrm{P}(\max_{1 \leq i \leq t} |S_i| \geq \lambda) &\leq \mathrm{P}\left\{L(G) \geq \frac{1}{2}\lambda\right\} + P\left(|S_t| \geq \frac{1}{2}\lambda\right) \\ &\leq \frac{2KC}{(\frac{1}{2}\lambda)^2} + \frac{\mathrm{E}|S_t|^2}{(\frac{1}{2}\lambda)^2} \leq \frac{KC}{\lambda^2}. \end{aligned}$$

If λ goes to infinity, the above probability converges to zero. Therefore, S_i is tight or equivalently $\hat{\mathcal{J}}_{i,M}/\sigma_0$ is tight.

Let $q = i/H$ and let $Y(q) = Y(i/H) \equiv \hat{\mathcal{J}}_{i,M}/\sigma_0$. For $0 \leq p \leq q$, consider $|p-q| \to 0$, then we have, as $H \to \infty$,

$$|p-q| \to 0 \Rightarrow |i_1 - i_2|/H \to 0 \Rightarrow i_d/H \to 0 \Rightarrow i_d = o(H).$$

If $i_d = o(H)$,

$$\begin{aligned}
\operatorname{cov}\{Y(p), Y(q)\} &= \sigma_0^{-2}\mathrm{E}(\hat{\mathcal{J}}_{i_1,M}\hat{J}_{i_2,M}) \\
&= \frac{(4/H^4)\sum_{h_1,h_2}\operatorname{tr}^2\{C(h_1)C(h_2)\}\{(H-i_d)^4\}}{(4/H^4)\sum_{h_1,h_2}\operatorname{tr}^2\{C(h_1)C(h_2)\}H^4}\{1+o(1)\} \\
&= \{(H-i_d)^4/H^4\}\{1+o(1)\} = 1 - 4(i_d/H) + o\{(i_d/H)\} \\
&= 1 - 4|p-q| + o\{|p-q|\}.
\end{aligned}$$

On the other hand, if $|p-q| \to \infty$ or $i_d/H \to \infty$, $\operatorname{cov}\{Y(p), Y(q)\} = 0$.

As a result, $\{Y(q), q \geq 0\}$ converges to $\{Z(q), q \geq 0\}$, which is a stationary Gaussian process with zero mean, unit variance and covariance function of the form

$$r(|p-q|) = \operatorname{cov}\{Z(p), Z(q)\} = 1 - 4|p-q| + o(|p-q|),$$

as $|p-q| \to 0$. On the other hand, as $|p-q| \to \infty$, $r(|p-q|)\log(|p-q|) \to 0$.

Let $Q = t/H$. From Finch (2003), as $Q \to \infty$, $\max_{0\leq q\leq Q}|Z(q)|$ has the Gumbel distribution so that

$$\mathrm{P}_\infty\left\{\max_{0\leq q\leq Q}|Z(q)| \leq a\right\} = \exp\left[-2\exp\left\{g(t/H,a)\right\}\right],$$

where

$$g(t/H,a) = 2\log(t/H) + 1/2\log\log(t/H) + \log(4/\sqrt{\pi}) - a\sqrt{2\log(t/H)}.$$

As a result, when $t > H$,

$$\mathrm{P}_\infty\{T_H(a,M) \leq t\} = 1 - \exp\left[-2\exp\left\{g(t/H,a)\right\}\right].$$

When $t = H$ and as $H \to \infty$,

$$\mathrm{P}_\infty\left\{\max_{0\leq q\leq 1}|Z(q)| \leq a\right\} = \exp\left\{-(2\sqrt{\pi})^{-1}H\exp(-a^2/2)\right\},$$

which has the order of $1 - 1/(2\sqrt{\pi})H\exp(-a^2/2)$ because $H = o\{\exp(a^2/2)\}$. As a result,

$$\mathrm{P}_\infty\{T_H(a,M) \leq H\} = 1/(2\sqrt{\pi})H\exp(-a^2/2),$$

which decays to zero as $H = o\{\exp(a^2/2)\}$.

We next derive the expectation of $T_H(a, M)$. Since the support of $T_H(a, M)$ is non-negative, we have

$$\mathrm{E}_\infty\{T_H(a, M)\} = \int_0^\infty \{1 - F_{T_H(a,M)}(t)\}dt,$$

where $F_{T_H(a,M)}(t)$ is the cumulative distribution function of $T_H(a, M)$ evaluated at t. From the above results, we have

$$\begin{aligned}
\mathrm{E}_\infty\{T_H(a, M)\} &= \int_0^H \{1 - F_{T_H(a,M)}(t)\}dt + \int_H^\infty \{1 - F_{T_H(a,M)}(t)\}dt \\
&= \left(H + \int_H^\infty \exp\left[-2\exp\left\{g(t/H, a)\right\}\right]dt\right)\{1 + o(1)\}.
\end{aligned}$$

■

A. 4. Proof of Lemma 1

From the expression of $W_M(i, j)$, we write

$$\begin{aligned}
W_M(i,j) &= \sum_{t=M+2}^{H-M-2} \frac{H-t-M}{t-M-1} I(i \le t)I(j \le t)I(|i-j| \ge M+1) \\
&+ \frac{t-M}{H-t-M-1} I(i \ge t+1)I(j \ge t+1)I(|i-j| \ge M+1) \\
&- 2\frac{(t-M)(H-t-M)}{t(H-t) - \frac{1}{2}M(M+1)} I(i \le t)I(j \ge t+1)I(|i-j| \ge M+1) \\
&= W_{M,1}(i,j) + W_{M,2}(i,j) + W_{M,3}(i,j)
\end{aligned}$$

Using (5), we obtain

$$\begin{aligned}
\mathrm{var}(\hat{\mathcal{J}}_{i,M}) &= \frac{4}{H^4} \sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} \sum_{i,j=1}^{H} \{W_{M,1}(i,j)W_{M,1}(i-h_1, j+h_2) \\
&+ W_{M,2}(i,j)W_{M,2}(i-h_1, j+h_2) + W_{M,3}(i,j)W_{M,3}(i-h_1, j+h_2) \\
&+ R_{W,1}(i,j) + R_{W,2}(i,j)\},
\end{aligned} \tag{12}$$

where

$$R_{W,1}(i,j) = \{W_{M,1}(i,j)W_{M,2}(i-h_1, j+h_2) + W_{M,2}(i,j)W_{M,1}(i-h_1, j+h_2)\},$$

and

$$\begin{aligned}
R_{W,2}(i,j) &= \{W_{M,1}(i,j)W_{M,3}(i-h_1, j+h_2) + W_{M,3}(i,j)W_{M,1}(i-h_1, j+h_2)\} \\
&+ W_{M,2}(i,j)W_{M,3}(i-h_1, j+h_2) + W_{M,3}(i,j)W_{M,2}(i-h_1, j+h_2)\}.
\end{aligned}$$

Since $h_1, h_2 \in \{0, \pm 1, \ldots, \pm M\}$, they are finite constant. As $H \to \infty$, we obtain the leading order of the first term on the right hand side of (12) as

$$
\begin{aligned}
& \frac{4}{H^4} \sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} \sum_{i,j=1}^{H} \{W_{M,1}(i,j) W_{M,1}(i-h_1, j+h_2)\} \\
= & \frac{4}{H^4} \sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} \sum_{i,j=1}^{H} \{W_{M,1}(i,j) W_{M,1}(i,j)\} \\
= & \frac{4}{H^4} \sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} \sum_{t=M+2}^{H-M-2} (H-t-M)^2 \\
+ & \frac{8}{H^4} \sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} \sum_{t_1=M+2}^{H-M-3} (H-t_1-M)(t_1-M) \sum_{t_2=t_1+1}^{H-M-2} \frac{H-t_2-M}{t_2-M-1} \\
= & I_{(1)} + I_{(2)},
\end{aligned}
$$

where

$$
\begin{aligned}
I_{(1)} = & \frac{4}{H^4} \sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} \Bigg\{ \sum_{t=1}^{H-2M-3} t^2 - 2(H-2M-1) \sum_{t=1}^{H-2M-3} t \\
+ & (H-2M-3)(H-2M-1)^2 \Bigg\} \\
= & \frac{4}{H^4} \sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} \left(\frac{1}{3} H^3\right) \{1+o(1)\},
\end{aligned}
$$

and by Euler-Mascheroni constant,

$$
\begin{aligned}
I_{(2)} = & \frac{8}{H^4} \sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} \sum_{t_1=M+2}^{H-M-3} (H-t_1-M)(t_1-M) \Bigg\{ -(H-t_1-M-2) \\
+ & (H-2M-1) \sum_{t_2=t_1+1}^{H-M-2} \frac{1}{t_2-M-1} \Bigg\} \\
= & -\frac{8}{H^4} \sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} \sum_{t_1=M+2}^{H-M-3} (H-t_1-M)(H-t_1-M-2)(t_1-M) \\
+ & \frac{8}{H^4} \sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} \sum_{t_1=M+2}^{H-M-3} (H-2M-1)(H-t_1-M)(t_1-M) \\
\times & \log\left(\frac{H-2M-3}{t_1-M-1}\right) \\
= & I_{(21)} + I_{(22)}.
\end{aligned}
$$

We see that,

$$
\begin{aligned}
I_{(21)} &= -\frac{8}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\sum_{t_1=1}^{H-2M-4}(H-t_1-2M-3)(H-t_1-2M-1)(t_1+1)\\
&= -\frac{8}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\Bigg[\sum_{t_1=1}^{H-2M-4}t_1^3-(2H-4M-5)\sum_{t_1=1}^{H-2M-4}t_1^2\\
&+ \{(H-2M-2)(H-2M-3)-(H-2M-1)\}\sum_{t_1=1}^{H-2M-4}t_1\\
&+ (H-2M-1)(H-2M-3)(H-2M-4)\Bigg]\\
&= -\frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\left(\frac{1}{6}H^4\right)\{1+o(1)\},
\end{aligned}
$$

and

$$
\begin{aligned}
I_{(22)} &= \frac{8}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\Bigg\{(H-2M-1)\int_1^{H-2M-4}t^2\log\left(\frac{t}{H-2M-3}\right)dt\\
&- (H-2M-1)(H-2M-2)\int_1^{H-2M-4}t\log\left(\frac{t}{H-2M-3}\right)dt\\
&- (H-2M-1)^2\int_1^{H-2M-4}\log\left(\frac{t}{H-2M-3}\right)dt\Bigg\}\{1+o(1)\}\\
&= \frac{8}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\Bigg\{-\frac{1}{9}(H-2M-1)(H-2M-4)^3\\
&+ \frac{1}{4}(H-2M-1)(H-2M-2)(H-2M-4)^2\Bigg\}\{1+o(1)\}\\
&= \frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\left(\frac{5}{18}H^4\right)\{1+o(1)\}.
\end{aligned}
$$

By combining all the above results, the first term on the right hand side of (12) is

$$
\begin{aligned}
&\frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\sum_{i,j=1}^{H}\{W_{M,1}(i,j)W_{M,1}(i-h_1,j+h_2)\}\\
=&\ \frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\left(\frac{1}{9}H^4\right)\{1+o(1)\}.
\end{aligned}
$$

By the same idea, the second term on the right hand side of (12) is

$$
\begin{aligned}
&\frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\sum_{i,j=1}^{H}W_{M,2}(i,j)W_{M,2}(i-h_1,j+h_2)\}\\
=&\ \frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\left(\frac{1}{9}H^4\right)\{1+o(1)\}.
\end{aligned}
$$

For the third term on the right hand side of (12), we have

$$\begin{aligned}
& \frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\sum_{i,j=1}^{H} W_{M,3}(i,j)W_{M,3}(i-h_1,j+h_2)\} \\
= & \frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\sum_{i,j=1}^{H} W_{M,3}(i,j)W_{M,3}(i,j)\}\{1+o(1)\} \\
= & \frac{16}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\sum_{t=M+2}^{H-M-2} t(H-t)\{1+o(1)\} \\
+ & \frac{32}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\sum_{t_1=M+2}^{H-M-3}\sum_{t_2=t_1+1}^{H-M-2} t_1(H-t_2)\{1+o(1)\} = II_{(1)}+II_{(2)},
\end{aligned}$$

where

$$\begin{aligned}
II_{(1)} &= \frac{16}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\{-\sum_{t=1}^{H-2M-3} t^2+(H-2M-2)\sum_{t=1}^{H-2M-3} t \\
&+ (H-2M-3)(M+1)(H-M-1)\} \\
&= \frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\left(\frac{2}{3}H^3\right)\{1+o(1)\},
\end{aligned}$$

and

$$\begin{aligned}
II_{(2)} &= \frac{32}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}H\sum_{t=1}^{H-2M-4}(H-t-2M-3)(t+M+1) \\
&- \frac{16}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\sum_{t=1}^{H-2M-4}(H+t)(H-t-2M-3)(t+M+1) \\
&= II_{(21)}+II_{(22)}.
\end{aligned}$$

We see that

$$\begin{aligned}
II_{(21)} &= \frac{32}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}H\Bigg\{-\sum_{t=1}^{H-2M-4} t^2+(H-3M-4)\sum_{t=1}^{H-2M-4} t \\
&+ (H-2M-4)(M+1)(H-2M-3)\Bigg\} \\
&= \frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\left(\frac{4}{3}H^4\right)\{1+o(1)\},
\end{aligned}$$

and

$$
\begin{aligned}
II_{(22)} &= \frac{16}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\Bigg[\sum_{t=1}^{H-2M-4}t^3+(3M+4)\sum_{t=1}^{H-2M-4}t^2\\
&- \{H(H-2M-3)-H(M+1)+(M+1)(H-2M-3)\}\sum_{t=1}^{H-2M-4}t\\
&- H(H-2M-3)(M+1)\Bigg]\\
&= -\frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}H^4\{1+o(1)\}.
\end{aligned}
$$

By combining the above results, the third term on the right hand side of (12) is

$$
\begin{aligned}
&\frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\}\sum_{i,j=1}^{H}W_{M,3}(i,j)W_{M,3}(i-h_1,j+h_2)\}\\
=\;&\frac{4}{H^4}\sum_{h_1,h_2}\mathrm{tr}^2\{C(h_1)C(h_2)\left(\frac{1}{3}H^4\right)\{1+o(1)\}.
\end{aligned}
$$

We now evaluate the terms involving $R_{W,1}(i,j)$ and $R_{W,2}(i,j)$ in (12), where

$$
\begin{aligned}
&\sum_{i,j=1}^{H}R_{W,1}(i,j)\\
=\;&2\sum_{i,j=1}^{H}W_{M,1}(i,j)W_{M,2}(i,j)\{1+o(1)\}\\
=\;&2\sum_{i,j=1}^{H}\sum_{t_1>t_2=M+2}^{H-M-2}\frac{(H-t_1-M)(t_2-M)}{(t_1-M-1)(H-t_2-M-1)}\\
\times\;&I(t_2+1\le i\le t_1)I(t_2+1\le j\le t_1)I(|i-j|\ge M+1)\{1+o(1)\}\\
=\;&\int_{2M+4}^{H-M-2}\int_{M+2}^{t_1-M-2}\frac{2(H-t_1-M)(t_2-M)(t_1-t_2-M)(t_1-t_2-M-1)}{(t_1-M-1)(H-t_2-M-1)}dt_2dt_1\\
=\;&\left(\frac{\pi^2}{3}-\frac{59}{18}\right)H^4\{1+o(1)\}.
\end{aligned}
$$

Next, we consider $R_{W,2}(i,j)$. Note that

$$
R_{W,2}(i,j) = 2W_{M,1}(i,j)W_{M,3}(i,j)\{1+o(1)\}+2W_{M,2}(i,j)W_{M,3}(i,j)\{1+o(1)\},
$$

where

$$
\begin{aligned}
& \sum_{i,j}^{H} W_{M,1}(i,j) W_{M,3}(i,j) \\
= \quad & -2 \sum_{i,j}^{H} \sum_{t_1 > t_2 = M+2}^{H-M-2} \frac{H - t_1 - M}{t_1 - M - 1} \frac{(t_2 - M)(H - t_2 - M)}{t_2(H - t_2) - \frac{1}{2}M(M+1)} \\
\times \quad & I(i \le t_2) I(t_2 + 1 \le j \le t_1) I(|i - j| \ge M + 1) \\
= \quad & -2 \sum_{t_2 = M+2}^{H-M-3} \sum_{t_1 = t_2+1}^{H-M-2} \frac{H - t_1 - M}{t_1 - M - 1} \cdot \frac{(t_2 - M)(H - t_2 - M)}{t_2(H - t_2) - \frac{1}{2}M(M+1)} \left\{ t_2(t_1 - t_2) - \frac{1}{2} M(M+1) \right\} \\
= \quad & -2 \int_{M+2}^{H-2M-3} \int_{t_2+M+1}^{H-M-2} \frac{H - t_1 - M}{t_1 - M - 1} \left\{ t_2(t_1 - t_2) - \frac{1}{2} M(M+1) \right\} dt_1 dt_2 \{1 + o(1)\} \\
= \quad & -\frac{1}{36} H^4 \{1 + o(1)\}.
\end{aligned}
$$

By the same idea, we have $\sum_{i,j}^{H} W_{M,2}(i,j) W_{M,3}(i,j) = -1/36 H^4\{1 + o(1)\}$. Therefore, $\sum_{i,j}^{H} R_{W,2}(i,j) = -1/9 H^4\{1 + o(1)\}$.

As a result, the leading order of the variance in Lemma 1 is

$$
\begin{aligned}
\mathrm{var}(\hat{\mathcal{J}}_{i,M}) \quad &= \quad \frac{4}{H^4} \sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} \left(\frac{2}{9} H^4 + \frac{1}{3} H^4 + \frac{6\pi^2 - 59}{18} H^4 - \frac{1}{9} H^4 \right) \{1 + o(1)\} \\
&= \quad \frac{4}{H^4} \sum_{h_1,h_2} \mathrm{tr}^2\{C(h_1)C(h_2)\} \left(\frac{6\pi^2 - 51}{18} H^4 \right) \{1 + o(1)\}.
\end{aligned}
$$

■

A. 5. Proof of Theorem 2

We first prove that the supremum of the expected detection delays attains at the immediate change point $\tau = n_0$. Equivalently, we need to show that for any $\tau > n_0$,

$$\mathrm{E}_\tau\{T_H(a,M) - (\tau - n_0) | T_H(a,M) > \tau - n_0\} \le \mathrm{E}_0\{T_H(a,M)\}.$$

To simplify notation, we let $\tau^* = \tau - n_0$ and $T^* = T_H(a,M) - \tau^*$. Then to show the above inequality, we only need to show that

$$\mathrm{E}_\tau\{T^* | T^* > 0\} \le \mathrm{E}_0\{T^*\}.$$

Since

$$\mathrm{E}_\tau\{T^* | T^* > 0\} = \int_0^\infty \{1 - \mathrm{P}_\tau(T^* < t | T^* > 0)\} dt, \quad \text{and}$$

$$\mathrm{E}_0\{T^*\} = \int_0^\infty \{1 - \mathrm{P}_0(T^* < t)\} dt,$$

we only need to show that

$$\mathrm{P}_\tau(T^* < t | T^* > 0) \ge \mathrm{P}_0(T^* < t), \tag{13}$$

First, the probability on the left hand side of (13) is

$$\mathrm{P}_\tau(T^* < t | T^* > 0) = \frac{\mathrm{P}_\tau(T^* < t) - \mathrm{P}_\tau(T^* < 0)}{1 - \mathrm{P}_\tau(T^* < 0)}, \tag{14}$$

where

$$\mathrm{P}_\tau\{T^* < t\} = \mathrm{P}_\tau\left(\max_{0 \le i \le t+\tau^*} \left| \frac{\hat{\mathcal{J}}_{n_0+i,M,H}}{\hat{\sigma}_{n_0,M,H}} \right| > a \right), \quad \text{and}$$

$$\mathrm{P}_\tau\{T^* < 0\} = \mathrm{P}_\tau\left(\max_{0 \le i \le \tau^*} \left| \frac{\hat{\mathcal{J}}_{n_0+i,M,H}}{\hat{\sigma}_{n_0,M,H}} \right| > a \right).$$

From the above two probabilities, we can define two events

$$A = \{ \max_{0 \le i \le t+\tau^*} \left| \frac{\hat{\mathcal{J}}_{n_0+i,M,H}}{\hat{\sigma}_{n_0,M,H}} \right| > a \}, \quad \text{and} \quad B = \{ \max_{0 \le i \le \tau^*} \left| \frac{\hat{\mathcal{J}}_{n_0+i,M,H}}{\hat{\sigma}_{n_0,M,H}} \right| > a \}.$$

Second, the probability on the right hand side of (13) is

$$\begin{aligned} \mathrm{P}_0\{T^* < t\} &= \mathrm{P}_0\left(\max_{0 \le i \le t} \left| \frac{\hat{\mathcal{J}}_{n_0+i,M,H}}{\hat{\sigma}_{n_0,M,H}} \right| > a \right) \\ &= \mathrm{P}_\tau\left(\max_{\tau^* \le i \le t+\tau^*} \left| \frac{\hat{\mathcal{J}}_{n_0+i,M,H}}{\hat{\sigma}_{n_0,M,H}} \right| > a \right). \end{aligned} \tag{15}$$

The last equation holds because both probabilities are based on the observations after the change points 0 and $\tau - n_0$, respectively, and the observations have the same distribution. From (15), we define the event

$$C = \{ \max_{\tau^* \le i \le t+\tau^*} \left| \frac{\hat{\mathcal{J}}_{n_0+i,M,H}}{\hat{\sigma}_{n_0,M,H}} \right| > a \}.$$

From the above defined events A, B and C, we see that $A = B \cup C$. Therefore, $\mathrm{P}(A) = \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(B \cap C)$. From the definition of the events B and C, we see that if B occurs, then $T^* < 0$ or the stopping time $T_H(a, M) < \tau^*$. Then C cannot occur. Therefore $\mathrm{P}(A) = \mathrm{P}(B) + \mathrm{P}(C)$. Moreover, from the definitions of A, B, C, (14) becomes

$$\mathrm{P}_\tau(T^* < t | T^* > 0) = \frac{\mathrm{P}_\tau(A) - \mathrm{P}_\tau(B)}{1 - \mathrm{P}_\tau(B)} = \frac{\mathrm{P}_\tau(C)}{1 - \mathrm{P}_\tau(B)} = \frac{\mathrm{P}_0\{T^* < t\}}{1 - \mathrm{P}_\tau(B)},$$

where the last equation holds by using (15). Then (13) can be proved accordingly. This completes the proof that the supremum of the expected detection delays attains at the immediate change point $\tau = n_0$.

We next establish the upper bound for the expected detection delays. To simplify notation, we let $\hat{\mathcal{J}}_T \equiv \hat{\mathcal{J}}_{n,M,H}$, which is the test statistic evaluated at the stopping time $T_H(a, M)$. Using (3), we see that

$$\mathrm{E}(\hat{\mathcal{J}}_T) = \mathrm{E}\left\{ \frac{1}{H^2} \sum_{i,j=1}^{H} W_M(i,j) (X_{n_0-H+T_H+i}^T X_{n_0-H+T_H+j})^2 \right\}.$$

The following lemma is proved in Section A. 6.

Lemma 2 *Under the same conditions in Theorem 2,*

$$\begin{aligned} E(\hat{\mathcal{J}}_T) &= \frac{\log H}{H}\Big[E\big\{(T_H - M)(T_H - M - 1)\big\} tr\{(\Sigma_{\tau+1} - \Sigma_\tau)^2\} \\ &+ M(M+1) tr\{\Sigma_\tau(\Sigma_{\tau+1} - \Sigma_\tau)\}\Big]\{1 + o(1)\}. \end{aligned}$$

From Lemma 2, we obtain

$$\begin{aligned} &\frac{\log H}{H}\mathrm{E}\big\{(T_H - M)(T_H - M - 1)\big\}\mathrm{tr}\{(\Sigma_{\tau+1} - \Sigma_\tau)^2\}\{1 + o(1)\} - (|\mathrm{E}(\hat{\mathcal{J}}_T)| - a \cdot \sigma_{H,M,0}) \\ \leq\ & a \cdot \sigma_{H,M,0} + \frac{\log H}{H} M(M+1)|\mathrm{tr}\{\Sigma_\tau(\Sigma_{\tau+1} - \Sigma_\tau)\}|\{1 + o(1)\}. \end{aligned} \tag{16}$$

Let $\hat{\mathcal{J}}_{T-1}$ denote the test statistic evaluated at $T_H - 1$. From the stopping rule (7), we have

$$\mathrm{E}|\hat{\mathcal{J}}_{T-1}| \leq a \cdot \sigma_{H,M,0}.$$

By Jensen's inequality and triangle inequality, we also have

$$\mathrm{E}|\hat{\mathcal{J}}_{T-1}| \geq |\mathrm{E}(\hat{\mathcal{J}}_T)| - |\mathrm{E}(\hat{\mathcal{J}}_T - \hat{\mathcal{J}}_{T-1})|.$$

Combining the above two inequality, we obtain

$$|\mathrm{E}(\hat{\mathcal{J}}_T)| - a \cdot \sigma_{H,M,0} \leq |\mathrm{E}(\hat{\mathcal{J}}_T - \hat{\mathcal{J}}_{T-1})|. \tag{17}$$

Based on similar derivations,

$$|\mathrm{E}(\hat{\mathcal{J}}_T - \hat{\mathcal{J}}_{T-1})| = \frac{2\log H}{H}\mathrm{E}(T_H - M - 1)\mathrm{tr}\{(\Sigma_{\tau+1} - \Sigma_\tau)^2\}\{1 + o(1)\}. \tag{18}$$

Combining (16), (17) and (18), we obtain

$$\begin{aligned} &\frac{\log H}{H}\mathrm{E}\{(T_H - M - 2)^2\}\mathrm{tr}\{(\Sigma_{\tau+1} - \Sigma_\tau)^2\}\{1 + o(1)\} \\ \leq\ & a \cdot \sigma_{H,M,0} + \frac{\log H}{H} M(M+1)|\mathrm{tr}\{\Sigma_\tau(\Sigma_{\tau+1} - \Sigma_\tau)\}|\{1 + o(1)\}. \end{aligned}$$

Using the Jensen's inequality and Cauchy-Schwarz inequality, we have

$$\begin{aligned} \mathrm{E}(T_H - M - 2) &\leq \sqrt{\mathrm{E}\{(T_H - M - 2)^2\}} \\ &\leq \left[\frac{a \cdot \sigma_{H,M,0} \cdot H}{\log H \cdot \mathrm{tr}\{(\Sigma_{\tau+1} - \Sigma_\tau)^2\}} + \frac{M(M+1)\sqrt{\mathrm{tr}(\Sigma_\tau^2)}}{\sqrt{\mathrm{tr}\{(\Sigma_{\tau+1} - \Sigma_\tau)^2\}}}\right]^{\frac{1}{2}}\{1 + o(1)\}. \end{aligned}$$

This completes the proof of Theorem 2. ■

A. 6. Proof of Lemma 2

To simplify notation, we let $Y_i = X_{n_0-H+T_H+i}X_{n_0-H+T_H+i}^T$. Since $\mathrm{E}(Y_i) = \Sigma_\tau$ if $n_0 - H + T_H + i \le n_0$ and $\mathrm{E}(Y_i) = \Sigma_{\tau+1}$ if $n_0 - H + T_H + i > n_0$, we write $\hat{\mathcal{J}}_T$ as

$$\begin{aligned}
\hat{\mathcal{J}}_T &= \frac{1}{H^2}\sum_{i,j=1}^{H-T_H} W_M(i,j)\mathrm{tr}(Y_iY_j) + \frac{1}{H^2}\sum_{i,j=H-T_H+1}^{H} W_M(i,j)\mathrm{tr}(Y_iY_j) \\
&+ \frac{2}{H^2}\sum_{i=1}^{H-T_H}\sum_{j=H-T_H+1}^{H} W_M(i,j)\mathrm{tr}(Y_iY_j) \\
&= \hat{\mathcal{J}}_{T,1} + \hat{\mathcal{J}}_{T,2} + \hat{\mathcal{J}}_{T,3}, \qquad (19)
\end{aligned}$$

where

$$\begin{aligned}
W_M(i,j) &= \sum_{t=M+2}^{H-M-2}\Big\{\frac{H-t-M}{t-M-1} I(i\le t)I(j\le t)I(|i-j|\ge M+1) \\
&+ \frac{t-M}{H-t-M-1} I(i\ge t+1)I(j\ge t+1)I(|i-j|\ge M+1) \\
&- \frac{2(t-M)(H-t-M)}{t(H-t)-\frac{1}{2}M(M+1)} I(i\le t)I(j\ge t+1)I(|i-j|\ge M+1)\Big\}.
\end{aligned}$$

Since $|i-j| \ge M+1$, it requires $T_H \ge M+2$ so that $\sum_{i,j=H-T_H+1}^{H} W_M(i,j)\mathrm{tr}(Y_iY_j) \ne 0$. We first evaluate $\mathrm{E}(\hat{\mathcal{J}}_{T,1})$ in (19), where $\mathrm{tr}(Y_iY_j)$ is only based on the training sample $\{X_i, 1 \le i \le n_0\}$. Since the stopping time T_H is relative to $\mathcal{F}_i = \sigma\{X_{n_0+1}, \dots, X_{n_0+i}\}$ and the training sample $\{X_i, 1 \le i \le n_0\}$ is independent of $\{X_{n_0+1}, \cdots\}$, we obtain

$$\mathrm{E}(\hat{\mathcal{J}}_{T,1}) = \frac{1}{H^2}\mathrm{E}\Big\{\sum_{i,j=1}^{H-T_H} W_M(i,j)\Big\}\mathrm{E}\{\mathrm{tr}(Y_iY_j)\} = \frac{1}{H^2}\mathrm{E}\Big\{\sum_{i,j=1}^{H-T_H} W_M(i,j)\mathrm{tr}(\Sigma_\tau^2)\Big\}.$$

There are three summations in $\mathrm{E}(\hat{\mathcal{J}}_{T,1})$: one with respect to t within $W_M(i,j)$, and the other two with respect to the indices i and j. We first evaluate the summations with respect to i and j and then the summation with respect to the index t. To do so, we let C_1 and C_2 be some constants satisfying $C_1 \le t \le C_2$ and apply the following results:

$$\begin{aligned}
\sum_{i,j=C_1}^{C_2} I(i\le t)I(j\le t)I(|i-j|\ge M+1) &= (t-C_1-M+1)(t-C_1-M), \\
\sum_{i,j=C_1}^{C_2} I(i\ge t+1)I(j\ge t+1)I(|i-j|\ge M+1) &= (C_2-t-M)(C_2-t-M-1), \\
\sum_{i,j=C_1}^{C_2} I(i\le t)I(j\ge t+1)I(|i-j|\ge M+1) &= (t-C_1+1)(C_2-t)-M(M+1)/2.
\end{aligned}$$

Note that the diagonal elements and the elements from the first to the Mth off-diagonal of an $H\times H$ matrix represent the elements with $|i-j| \le M$. The three equations can be established by subtracting those elements from each block matrix.

We are ready to evaluate $\mathrm{E}(\hat{\mathcal{J}}_{T,1})$. Note that $T_H \geq M+2$. We apply the above three equations to obtain

$$\begin{aligned}
\frac{1}{H^2}\sum_{i,j=1}^{H-T_H} W_t(i,j) &= \frac{1}{H^2}\sum_{t=M+2}^{H-T_H}\Bigg[\frac{H-t-M}{t-M-1}(t-M)(t-M-1) \\
&+ \frac{t-M}{H-t-M-1}(H-T_H-t-M)(H-T_H-t-M-1) \\
&+ 2\Big\{-1+\frac{M(H-\frac{3}{2}M-\frac{1}{2})}{t(H-t)-\frac{1}{2}M(M+1)}\Big\}\Big\{t(H-T_H-t)-\frac{1}{2}M(M+1)\Big\}\Bigg] \\
&+ \frac{1}{H^2}\sum_{t=H-T_H+1}^{H-M-2}\frac{H-t-M}{t-M-1}(H-T_H-M)(H-T_H-M-1),
\end{aligned}$$

where, as $H\to\infty$, we approximate the sum by integral to obtain

$$\begin{aligned}
&\frac{1}{H^2}\sum_{t=M+2}^{H-T_H}\Bigg[(H-t-M)(t-M)+\frac{(t-M)(H-T_H-t-M)(H-T_H-t-M-1)}{H-t-M-1} \\
+&\ 2\Big\{-1+\frac{M(H-\frac{3}{2}M-\frac{1}{2})}{t(H-t)-\frac{1}{2}M(M+1)}\Big\}\Big\{t(H-T_H-t)-\frac{1}{2}M(M+1)\Big\}\Bigg] \\
=&\ T_H(T_H-2M-1)\frac{\log H}{H}\{1+o(1)\}.
\end{aligned}$$

And

$$\begin{aligned}
&\frac{1}{H^2}\sum_{t=H-T_H+1}^{H-M-2}\frac{(H-t-M)(H-T_H-M)(H-T_H-M-1)}{t-M-1} \\
=&\ \frac{(-2M-1)(T_H-M-2)}{H}\{1+o(1)\}.
\end{aligned}$$

As a result, we obtain $H^{-2}\sum_{i,j=1}^{H-T_H} W_t(i,j) = T_H(T_H-2M-1)H^{-1}\log H\{1+o(1)\}$. This shows that as $H\to\infty$,

$$\mathrm{E}(\hat{\mathcal{J}}_{T,1}) = \frac{\log H}{H}\mathrm{E}\big\{T_H(T_H-2M-1)\big\}\mathrm{tr}(\Sigma_\tau^2)\{1+o(1)\}. \tag{20}$$

We next consider $\hat{\mathcal{J}}_{T,2}$ in (19). To this end, we first write

$$\begin{aligned}
\hat{\mathcal{J}}_{T,2} &= \frac{2}{H^2}\sum_{i<j=H-T_H+1}^{H} I(j-i\geq M+1)\Big\{\sum_{t=j}^{H-M-2}\frac{H-t-M}{t-M-1}+\sum_{t=M+2}^{i-1}\frac{t-M}{H-t-M-1} \\
&- \sum_{t=i}^{j-1}\frac{(t-M)(H-t-M)}{t(H-t)-\frac{1}{2}M(M+1)}\Big\}\mathrm{tr}(Y_iY_j).
\end{aligned}$$

As $H \to \infty$, the leading order can be derived to be

$$\begin{aligned}
\mathrm{E}(\hat{\mathcal{J}}_{T,2}) &= \frac{2}{H^2}\mathrm{E}\Big[\sum_{i<j=H-T_H+1}^{H} I(j-i \geq M+1)\Big\{(3+2M-H)+(H-2M-1) \\
&\times \sum_{t=j}^{H-M-2}\frac{1}{t-M-1}+(H-2M-1)\sum_{t=M+2}^{i-1}\frac{1}{H-t-M-1}\Big\}\mathrm{tr}(Y_iY_j)\Big]\{1+o(1)\} \\
&= \frac{2}{H^2}\mathrm{E}\Big[\sum_{i<j=H-T_H+1}^{H} I(j-i \geq M+1)\Big\{(3+2M-H) \\
&+ (H-2M-1)\sum_{t=j-M-1}^{H-2M-3}\frac{1}{t}+(H-2M-1)\sum_{t=H-M-i}^{H-2M-3}\frac{1}{t}\Big\}\mathrm{tr}(Y_iY_j)\Big]\{1+o(1)\} \\
&= \frac{2}{H^2}\mathrm{E}\Big\{\Big(\sum_{i<j=H-T_H+1}^{H} I(j-i \geq M+1)H\sum_{t=H-M-i}^{H-2M-3}\frac{1}{t}\Big)\mathrm{tr}(Y_iY_j)\Big\}\{1+o(1)\}.
\end{aligned}$$

Note that $1 \leq H-M-i \leq T-M$. As $H \to \infty$, we approximate $\sum_{t=H-M-i}^{H-2M-3} 1/t$ by integral and obtain the leading order

$$\begin{aligned}
\mathrm{E}(\hat{\mathcal{J}}_{T,2}) &= \frac{2\log H}{H}\mathrm{E}\Big\{\sum_{i<j=H-T_H+1}^{H} I(j-i \geq M+1)\mathrm{tr}(Y_iY_j)\Big\}\{1+o(1)\} \\
&= \frac{2\log H}{H}\mathrm{E}\Big\{\sum_{i<j=1}^{T_H} I(j-i \geq M+1)\mathrm{tr}(Y_{n_0+i}Y_{n_0+j})\Big\}\{1+o(1)\} \\
&= \frac{2\log H}{H}\mathrm{E}\Big\{\sum_{i<j=1}^{T_H} I(j-i \geq M+1)\mathrm{tr}(\Sigma_{\tau+1}^2)\Big\}\{1+o(1)\} \\
&+ \frac{2\log H}{H}\mathrm{E}\Big[\sum_{i<j=1}^{T_H} I(j-i \geq M+1)\mathrm{tr}\{(Y_{n_0+i}-\Sigma_{\tau+1})(Y_{n_0+j}-\Sigma_{\tau+1})\}\Big]\{1+o(1)\} \\
&+ \frac{2\log H}{H}\mathrm{E}\Big[\sum_{i,j=1}^{T_H} I(|j-i| \geq M+1)\mathrm{tr}\{(\Sigma_{\tau+1})(Y_{n_0+j}-\Sigma_{\tau+1})\}\Big]\{1+o(1)\}, \quad (21)
\end{aligned}$$

where from first line to second line, we have changed the index by letting $i' = i-H+T_H$ and $j' = j-H+T_H$. To simplify notation, we redefine i' as i and j' as j.

We next evaluate each of the three terms (the last three lines) in (21). For the first term, the only random variable is T_H. After summing over i and j, we obtain

$$\frac{2\log H}{H}\mathrm{E}\Big\{\sum_{i<j=1}^{T_H} I(j-i \geq M+1)\mathrm{tr}(\Sigma_{\tau+1}^2)\Big\} = \frac{\log H}{H}\mathrm{E}\Big\{(T_H-M)(T_H-M-1)\Big\}\mathrm{tr}(\Sigma_{\tau+1}^2). \quad (22)$$

We write the second term in (21) as

$$\begin{aligned}
& \frac{2\log H}{H}\mathrm{E}\Big[\sum_{i<j=1}^{T_H} I(j-i\geq M+1)\mathrm{tr}\{(Y_{n_0+i}-\Sigma_{\tau+1})(Y_{n_0+j}-\Sigma_{\tau+1})\}\Big] \\
= & \frac{\log H}{H}\sum_{r_1,r_2=1}^{p}\mathrm{E}\Big\{(\sum_{i=1}^{T_H} Y^*_{n_0+i,r_1r_2})^2\Big\} - \frac{\log H}{H}\sum_{r_1,r_2=1}^{p}\mathrm{E}\Big(\sum_{i=1}^{T_H} Y^{*2}_{n_0+i,r_1r_2}\Big) \\
- & \frac{2\log H}{H}\sum_{r_1,r_2=1}^{p}\mathrm{E}\Big\{\Big(\sum_{i=1}^{T_H-M}\sum_{q=1}^{M} + \sum_{i=T_H-M+1}^{T_H-1}\sum_{q=1}^{T_H-i}\Big)Y^*_{n_0+i,r_1r_2}Y^*_{n_0+i+q,r_1r_2})\Big\}, \quad (23)
\end{aligned}$$

where $Y^*_{n_0+i} = Y_{n_0+i} - \Sigma_{\tau+1}$ and $Y^*_{n_0+i,r_1r_2}$ is the element on the r_1th row and r_2th column of $Y^*_{n_0+i}$. The last term on the right hand side exists if $M \geq 1$. Using (vii) of Corollary 1.1 in Janson (1983), we obtain

$$\begin{aligned}
\sum_{r_1,r_2=1}^{p}\mathrm{E}\Big\{(\sum_{i=1}^{T_H} Y^*_{n_0+i,r_1r_2})^2\Big\} &= \mathrm{E}(T_H)\mathrm{tr}^2(\Sigma_{\tau+1}) + 2\mathrm{E}\Big(\sum_{i=1}^{T_H-M}\sum_{q=1}^{M} + \sum_{i=T_H-M+1}^{T_H-1}\sum_{q=1}^{T_H-i}\Big) \\
&\times [\mathrm{tr}^2\{C_{\tau+1}(q)\} + \mathrm{tr}\{C_{\tau+1}(q)C^T_{\tau+1}(q)\}] + o\{\mathrm{E}(T_H)\mathrm{tr}(\Sigma^2_{\tau+1})\}.
\end{aligned}$$

Moreover, using (v) of Corollary 1.1 in Janson (1983), we obtain

$$\sum_{r_1,r_2=1}^{p}\mathrm{E}\Big(\sum_{i=1}^{T_H} Y^{*2}_{n_0+i,r_1r_2}\Big) = \mathrm{E}(T_H)\mathrm{tr}^2(\Sigma_{\tau+1}) + o\{\mathrm{E}(T_H)\mathrm{tr}(\Sigma^2_{\tau+1})\}.$$

To evaluate the last term on the right hand side of (23), we let $\mathcal{F}_i = \sigma\{X_{n_0+1},\ldots,X_{n_0+i}\}$ if $i \geq 1$ and $\mathcal{F}_i = \{\emptyset,\Omega\}$ if $i = 0$. Then $\{\mathcal{F}_i\}_0^\infty$ is an increasing sequence of σ-fields on a probability space $(\Omega,\mathcal{F}_\infty,P)$, and $\{X_{n_0+i}\}_{i=1}^\infty$ is a stationary and M-dependent sequence of random vectors adapted to $\{\mathcal{F}_i\}_{i=1}^\infty$, and $\{X_{n_0+i+j}\}_{j=M+1}^\infty$ is independent of $\{\mathcal{F}_i\}$ for every i. Moreover, $\{Y_{n_0+i,r_1r_2}\}_{i=1}^\infty$ is a sequence of stationary and M-dependent random variables adapted to $\{\mathcal{F}_i\}_{i=1}^\infty$, and $\{Y_{n_0+i+j,r_1r_2}\}_{j=M+1}^\infty$ is independent of $\{\mathcal{F}_i\}$ for every i. As a result, $\{\sum_{q=1}^M Y^*_{n_0+i,r_1r_2}Y^*_{n_0+i+q,r_1r_2}\}_{i=1}^\infty$ is a sequence of stationary and $2M$-dependent random variables adapted to $\{\mathcal{F}_{i+M}\}_{i=1}^\infty$, and $\mathrm{E}\{\sum_{q=1}^M (Y_{n_0+i,r_1r_2}-\sigma_{\tau+1,r_1r_2})(Y_{n_0+i+q,r_1r_2}-\sigma_{\tau+1,r_1r_2})\} = \sum_{q=1}^M \mathrm{Cov}(Y_{n_0+i},Y_{n_0+i+q})$. Again, using (v) of Corollary 1.1 in Janson (1983), we obtain

$$\begin{aligned}
& \sum_{r_1,r_2=1}^{p}\mathrm{E}\Big\{\Big(\sum_{i=1}^{T_H-M}\sum_{q=1}^{M} + \sum_{i=T_H-M+1}^{T_H-1}\sum_{q=1}^{T_H-i}\Big)Y^*_{n_0+i,r_1r_2}Y^*_{n_0+i+q,r_1r_2})\Big\} \\
= & \mathrm{E}\Big(\sum_{i=1}^{T_H-M}\sum_{q=1}^{M} + \sum_{i=T_H-M+1}^{T_H-1}\sum_{q=1}^{T_H-i}\Big)[\mathrm{tr}^2\{C_{\tau+1}(q)\} + \mathrm{tr}\{C_{\tau+1}(q)C^T_{\tau+1}(q)\}] \\
+ & o\{\mathrm{E}(T_H)\mathrm{tr}(\Sigma^2_{\tau+1})\}.
\end{aligned}$$

Combining the above results for (23), we obtain

$$\begin{aligned}&\frac{2\log H}{H}\mathrm{E}\Big[\sum_{i<j=1}^{T_H} I(j-i\geq M+1)\mathrm{tr}\{(Y_{n_0+i}-\Sigma_{\tau+1})(Y_{n_0+j}-\Sigma_{\tau+1})\}\Big]\\ =\ & o\Big\{H^{-1}\mathrm{E}(T_H)\log(H)\mathrm{tr}(\Sigma_{\tau+1}^2)\Big\}. \qquad (24)\end{aligned}$$

For the last term in (21), we apply (v) of Corollary 1.1 in Janson (1983) to obtain

$$\begin{aligned}&\frac{2\log H}{H}\mathrm{E}\Big[\sum_{i,j=1}^{T_H} I(|j-i|\geq M+1)\mathrm{tr}\{(\Sigma_{\tau+1})(Y_{n_0+j}-\Sigma_{\tau+1})\}\Big]\\ =\ & o\Big\{H^{-1}\mathrm{E}(T_H)\log(H)\mathrm{tr}(\Sigma_{\tau+1}^2)\Big\}. \qquad (25)\end{aligned}$$

By combining (22), (24) and (25), $\mathrm{E}(\hat{\mathcal{J}}_{T,2})$ in (21) becomes

$$\mathrm{E}(\hat{\mathcal{J}}_{T,2})=\frac{\log H}{H}\mathrm{E}\Big\{(T_H-M)(T_H-M-1)\Big\}\mathrm{tr}(\Sigma_{\tau+1}^2)\Big\{1+o(1)\Big\}. \qquad (26)$$

At last, we consider $\mathrm{E}(\hat{\mathcal{J}}_{T,3})$ in (19), where Y_i is the observation in the training sample and Y_j is observation after the training sample. Since the stopping time T_H is relative to $\mathcal{F}_i=\sigma\{X_{n_0+1},\ldots,X_{n_0+i}\}$ and Y_i and Y_j are independent, we obtain

$$\mathrm{E}(\hat{\mathcal{J}}_{T,3}) \;=\; \frac{2}{H^2}\sum_{j=H-T_H+1}^{H}\sum_{t=M+2}^{H-M-2}\sum_{i=1}^{H-T_H} W_M(i,j)\mathrm{tr}(\Sigma_\tau Y_j).$$

Note that $T_H\geq M+2$. We write

$$\begin{aligned}&\frac{2}{H^2}\sum_{j=H-T_H+1}^{H}\sum_{t=M+2}^{H-M-2}\sum_{i=1}^{H-T_H} W_M(i,j)\mathrm{tr}(\Sigma_\tau Y_j)\\ =\ & \frac{2}{H^2}\sum_{j=H-T_H+1}^{H}\sum_{t=M+2}^{H-T_H}\Big\{\frac{(t-M)(H-T_H-t)}{H-t-M-1}-\frac{t(t-M)(H-t-M)}{t(H-t)-\frac{1}{2}M(M+1)}\Big\}\mathrm{tr}(\Sigma_\tau Y_j)\\ +\ & \frac{2}{H^2}\sum_{j=H-T_H+1}^{H}\sum_{t=H-T_H+1}^{H-M-2}\Big\{\frac{(H-t-M)(H-T_H)}{t-M-1}I(t\geq j)\\ -\ & \frac{(t-M)(H-t-M)(H-T_H)}{t(H-t)-\frac{1}{2}M(M+1)}I(t\leq j-1)\Big\}\mathrm{tr}(\Sigma_\tau Y_j).\end{aligned}$$

The leading order of $\mathrm{E}(\hat{\mathcal{J}}_{T,3})$ is contributed by

$$\begin{aligned}&\frac{2}{H^2}\mathrm{E}\Big[\sum_{j=H-T_H+1}^{H}\sum_{t=M+2}^{H-T_H}\Big\{\frac{(t-M)(H-T_H-t)}{H-t-M-1}-\frac{t(t-M)(H-t-M)}{t(H-t)-\frac{1}{2}M(M+1)}\Big\}\mathrm{tr}(\Sigma_\tau Y_j)\Big]\\ =\ & -\frac{2\log H}{H}\mathrm{E}\Big\{T_H(T_H-2M-1)+\frac{M(M+1)}{2}\Big\}\mathrm{tr}(\Sigma_\tau\Sigma_{\tau+1})\{1+o(1)\},\end{aligned}$$

where to obtain the right hand side, we have approximated $\sum_{t=M+2}^{H-T_H}$ by the integral as $H \to \infty$, and similar to (25), we have applied (v) of Corollary 1.1 in Janson (1983).

As a result,

$$\mathrm{E}(\hat{\mathcal{J}}_{T,3}) = -\frac{2\log H}{H}\mathrm{E}\Big\{T_H(T_H - 2M - 1) + \frac{M(M+1)}{2}\Big\}\mathrm{tr}(\Sigma_\tau \Sigma_{\tau+1})\{1 + o(1)\}. \quad (27)$$

From (19), $\mathrm{E}(\hat{\mathcal{J}}_T)$ can be derived by adding (20), (26) and (27) together. This completes the proof of Lemma 2.

■

A. 7. Proof of Theorem 3

The asymptotic normality of $\hat{\mathcal{J}}_{n_0,M}$ can be established by the martingale central limit theorem. Toward this end, we let $\mathscr{F}_0 = \{\emptyset, \Omega\}$, $\mathscr{F}_k = \sigma\{X_1, ..., X_k\}$ with $k = 1, 2, ..., n_0$, and $\mathrm{E}_k(\cdot)$ denote the conditional expectation given $\mathscr{F}_k$. Define $D_{n_0,k} = (\mathrm{E}_k - \mathrm{E}_{k-1})\hat{\mathcal{J}}_{n_0,M}$ and it is easy to see that $\hat{\mathcal{J}}_{n_0,M} - \mu_{\hat{\mathcal{J}}_{n_0,M}} = \sum_{k=1}^{n_0} D_{n_0,k}$.

We further define $S_{n_0,m} = \sum_{k=1}^{m} D_{n_0,k} = \mathrm{E}_m\hat{\mathcal{J}}_{n_0,M} - \mu_{\hat{\mathcal{J}}_{n_0,M}}$. We can show that for $q \geq m$, $\mathrm{E}(S_{n_0,q}|\mathscr{F}_m) = S_{n_0,m}$. To this end, we note that $S_{n_0,q} = \mathrm{E}_q\hat{\mathcal{J}}_{n_0,M} - \mu_{\hat{\mathcal{J}}_{n_0,M}} = \mathrm{E}_m\hat{\mathcal{J}}_{n_0,M} - \mu_{\hat{\mathcal{J}}_{n_0,M}} + \mathrm{E}_q\hat{\mathcal{J}}_{n_0,M} - \mathrm{E}_m\hat{\mathcal{J}}_{n_0,M} = S_{n_0,m} + (\mathrm{E}_q\hat{\mathcal{J}}_{n_0,M} - \mathrm{E}_m\hat{\mathcal{J}}_{n_0,M})$. Then

$$\begin{aligned}
\mathrm{E}(S_{n_0,q}|\mathscr{F}_m) &= S_{n_0,m} + \mathrm{E}\{\mathrm{E}_q(\hat{\mathcal{J}}_{n_0,M})|\mathscr{F}_m\} - \mathrm{E}\{\mathrm{E}_m(\hat{\mathcal{J}}_{n_0,M})|\mathscr{F}_m\} \\
&= S_{n_0,m} + \mathrm{E}\{\mathrm{E}_m(\hat{\mathcal{J}}_{n_0,M})\} - \mathrm{E}\{\mathrm{E}_m(\hat{\mathcal{J}}_{n_0,M})\} \\
&= S_{n_0,m}.
\end{aligned}$$

As a result, we see that $\{S_{n_0,k}, \mathscr{F}_k\}$ is a martingale and accordingly, $\{D_{n_0,k}, 1 \leq k \leq n_0\}$ is a martingale difference sequence with respect to the σ-fields $\{\mathscr{F}_k, 1 \leq k \leq n_0\}$

Based on similar derivations for Lemmas 2 and 3 in Li and Chen (2012), we can show that under (2.2) in the main paper and Conditions 1–2, as $n_0 \to \infty$,

$$\frac{\sum_{k=1}^{n_0}\mathrm{E}_{k-1}(D_{n_0,k}^2)}{\sigma^2_{\hat{\mathcal{J}}_{n_0,M}}} \xrightarrow{p} 1. \qquad \text{And}$$

$$\frac{\sum_{k=1}^{n_0}\mathrm{E}(D_{n_0,k}^4)}{\sigma^4_{\hat{\mathcal{J}}_{n_0,M}}} \to 0.$$

The above two results are sufficient conditions for the martingale central limit theorem. This thus completes the first part of Theorem 3.

To show the second part of Theorem 3, we only need to show the ratio consistency of $\hat{\sigma}_{\hat{\mathcal{J}}_{n_0,M},0}$ defined in (8) to $\sigma_{\hat{\mathcal{J}}_{n_0,M},0}$ under the null hypothesis. From the expression (6), we apply (2) such that under the null hypothesis,

$$\begin{aligned}
\mathrm{E}\bigg(\frac{1}{n^*}\sum_{s,t}^{*} X_{t+h_2}^T X_s X_{s+h_1}^T X_t\bigg) &= \frac{1}{n^*}\sum_{s,t}^{*}\mathrm{E}(Z^T\Gamma_{t+h_2}^T\Gamma_s Z Z^T \Gamma_{s+h_1}^T\Gamma_t Z) \\
&= \mathrm{tr}\{C(h_1)C(h_2)\}.
\end{aligned}$$

This shows that $\mathrm{E}[\mathrm{tr}\{\widehat{C(h_1)C}(h_2)\}] = \mathrm{tr}\{C(h_1)C(h_2)\}$. Similarly, under (C1–C2), we have $\mathrm{var}[\mathrm{tr}\{\widehat{C(h_1)C}(h_2)\}] = o[\mathrm{tr}^2\{C(h_1)C(h_2)\}]$. This implies that under the null hypothesis,

$$\mathrm{tr}\{\widehat{C(h_1)C}(h_2)\}/\mathrm{tr}\{C(h_1)C(h_2)\} \xrightarrow{p} 1.$$

The second part of Theorem 3 is then proved by applying the continuous mapping theorem. ■

A. 8. Proof of Theorem 4

We first show that $\mathrm{P}(\hat{M} = M) \to 1$ as $n_0 \to \infty$. Note that the event that $\hat{M} > M$ is equivalent to the event that

$$\mathrm{tr}\{\widehat{C(M+1)C}(-M-1)\}/\mathrm{tr}\{\widehat{C(0)C}(0)\} > \epsilon.$$

Therefore, $\mathrm{P}(\hat{M} > M)$ is equivalent to

$$\mathrm{P}\left[\mathrm{tr}\{\widehat{C(M+1)C}(-M-1)\}/\mathrm{tr}\{\widehat{C(0)C}(0)\} > \epsilon\right].$$

It is also equivalent to $\mathrm{P}\left[\mathrm{tr}\{\widehat{C(M+1)C}(-M-1)\}/\mathrm{tr}\{C(0)C(0)\} > \epsilon\right]$ as

$$\mathrm{tr}\{\widehat{C(0)C}(0)\}/\mathrm{tr}\{C(0)C(0)\} \xrightarrow{p} 1$$

from the proof of Theorem 3.

From (6), we can show that $\mathrm{E}[\mathrm{tr}\{\widehat{C(M+1)C}(-M-1)\}] = 0$ and

$$\mathrm{var}\left[\mathrm{tr}\{\widehat{C(M+1)C}(-M-1)\}/\mathrm{tr}\{C(0)C(0)\}\right] = O(n^{-2}).$$

Using Chebyshev's inequality, we can show that as $n_0 \to \infty$,

$$\mathrm{P}\left[\mathrm{tr}\{\widehat{C(M+1)C}(-M-1)\}/\mathrm{tr}\{C(0)C(0)\} > \epsilon\right] \to 0,$$

or equivalently, $\mathrm{P}(\hat{M} > M) \to 0$. Similarly, we can show that $\mathrm{P}(\hat{M} < M) \to 0$. We then establish $\mathrm{P}(\hat{M} = M) \to 1$ as $n_0 \to \infty$.

To prove $\mathrm{E}_\infty\{T_H(a, \hat{M})\} - \mathrm{E}_\infty\{T_H(a, M)\} \to 0$, we only need to show that for every t, as $n_0 \to \infty$,

$$\mathrm{P}_\infty\{T_H(a, \hat{M}) \le t\} - \mathrm{P}_\infty\{T_H(a, M) \le t\} \to 0.$$

Toward this end, we notice that

$$\mathrm{P}_\infty\{T_H(a, \hat{M}) \le t\} = \mathrm{P}_\infty\{T_H(a, \hat{M}) \le t, \hat{M} = M\} + \mathrm{P}_\infty\{T_H(a, \hat{M}) \le t, \hat{M} \ne M\},$$

where the second term converges to zero because $\mathrm{P}(\hat{M} = M) \to 1$ as $n_0 \to \infty$.

Similarly, we can show $\mathrm{E}_0\{T_H(a, \hat{M})\} - \mathrm{E}_0\{T_H(a, M)\} \to 0$. ■

References

Elena Allen, Eswar Damaraju, Sergey Plis, Erik Erhardt, Tom Eichele and Vince Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, 24(3): 663-676, 2014.

Alexander Aue, Siegfried Hörmann, Lajos Horváth and Matthew Reimherr. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B): 4046-4087, 2009.

Valeriy Avanesov and Nazar Buzun. Change-point detection in high-dimensional covariance structure. *Electronic Journal of Statistics*, 12: 3254-3294, 2018.

Valeriy Avanesov. Structural break analysis in high-dimensional covariance structure. Technical Report, arXiv:1803.00508, 2019.

Jushan Bai. Common breaks in means and variances for panel data. *Journal of Econometrics*, 157: 78-92, 2010.

Zhidong Bai and Hewa Saranadasa. Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6: 311–329, 1996.

Ioana-Alexandra Bara, Carol Fung and Thang Dinh. Enhancing Twitter spam accounts discovery using cross-account pattern mining. *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 491-496, IEEE, 2015.

Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 1999.

Vince D Calhoun, Robyn Miller, Godfrey Pearlson and Tulay Adalı. The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron*, 84(2): 262-274, 2014.

Yang Cao, Andrew Thompson, Meng Wang and Yao Xie. Sketching for sequential change-point detection. *EURASIP Journal on Advances in Signal Processing*, 42, 2019.

Hock Peng Chan. Optimal sequential detection in multi-stream data. *The Annals of Statistics*, 45(6): 2736-2763, 2017.

Catie Chang and Gary H Glover. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage*, 50(1): 81-98, 2010.

Song Xi Chen and Ying-Li Qin. A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2): 808-835, 2010.

Ivor Cribben, Ragnheidur Haraldsdottir, Lauren Y Atlas, Tor D Wager and Martin A Lindquist. Dynamic connectivity regression: determining state-related changes in brain connectivity. *Neuroimage*, 61(4): 907-920, 2012.

Ivor Cribben, Tor D Wager and Martin A Lindquist. Detecting functional connectivity change points for single-subject fMRI data. *Frontiers in Computational Neuroscience*, 7: 143, 2013.

Steven Finch. *Mathematical Constants*. Cambridge University Press, 2003.

Daniel A Handwerker, Vinai Roopchansingh, Javier Gonzalez-Castillo and Peter A Bandettini. Periodic changes in fMRI connectivity. *Neuroimage*, 63(3): 1712-1719, 2012.

Wassily Hoeffding. A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics*, 19: 293-325, 1948.

R Matthew Hutchison, Thilo Womelsdorf, Elena A Allen, Peter A Bandettini, Vince D Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H Duyn, Gary H Glover, Javier Gonzalez-Castillo, Daniel A Handwerker, Shella Keilholz, Vesa Kiviniemi, David A Leopold, Francesco de Pasquale, Olaf Sporns, Martin Walter, Catie Chang. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80: 360-378, 2013.

Svante Janson. Renewal theory for M-dependent variables. *The Annals of Probability*, 11(3): 558-568, 1983.

Seok-OhJeong, Chongwon Pae and Hae-Jeong Park. Connectivity-based change point detection for large-size functional networks. *NeuroImage*, 143: 353-363, 2016.

Moritz Jirak. Uniform change point tests in high dimension. *The Annals of Statistics*, 43: 2451-2483, 2015.

Tze Leung Lai. Sequential changepoint detection in quality control and dynamical systems. *Journal of Royal Statistical Society, Series B*, 57: 613-658, 1995.

Jun Li and Song Xi Chen. Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics*, 40(2): 908-940, 2012.

Caiyan Li and Hongzhe Li. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, 24: 1175-1182, 2008.

Jun Li and Lingjun Li. onlineCOV: online change point detection in high-dimensional covariance structure. R package Version 0.1.3.

Jun Li, Minya Xu, Ping-Shou Zhong and Lingjun Li. Change point detection in the mean of high-dimensional time series data under dependence. Technical Report, arXiv:1903.07006, 2019.

Gary Lorden. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 42(6): 1897-1908, 1971.

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

Florence Merlevède, Magda Peligrad, and Sergey Utev. Recent advances in invariance principles for stationary sequences. *Probability surveys*, 3: 1-36, 2006.

Ricardo Pio Monti, Peter Hellyer, David Sharp, Robert Leech, Christoforos Anagnostopoulos and Giovanni Montana. Estimating time-varying brain connectivity networks from functional MRI time series. *NeuroImage*, 103: 427-443, 2014.

George V. Moustakides. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, 14(4): 1379-13877, 1986.

Ewan Stafford Page. Continuous Inspection Schemes. *Biometrika*, 41(1/2): 100-115, 1954.

Moshe Pollak. Optimal detection of a change in distribution. *The Annals of Statistics*, 13(1): 206-227, 1985.

Moshe Pollak and David Siegmund. Sequential detection of a change in a normal mean when the initial value is unknown. *The Annals of Statistics*, 19(1): 394-416, 1991.

Moshe Pollak and Alexander G. Tartakovsky. Optimality properties of the Shiryaev-Roberts procedure. *Statistica Sinica*, 19: 1729-1739, 2009.

Aleksey S Polunchenko and Alexander G. Tartakovsky. On optimality of the Shiryaev-Roberts procedure for detecting a change in distribution. *The Annals of Statistics*, 38(6): 3445-3457, 2010.

SW Roberts. A comparison of some control chart procedures. *Technometrics*, 8(3): 411-430, 1966.

Albert Nikolayevich Shiryaev. The problem of the most rapid detection of a disturbance in a stationary process. *Soviet Math. Dokl.*, 2: 795-799, 1961.

Albert Nikolayevich Shiryaev. On optimal methods in earliest detection problems. *Theory of Probability and its Applications*, 8: 26-51, 1963.

David Siegmund. *Sequential Analysis: Tests and Confidence Intervals*. Springer Science and Business Media, 1985.

David Siegmund and E. S. Venkatraman. Using the generalized likelihood ratio statistic for sequential detection of a change-point. *The Annals of Statistics*, 23(1): 255-271, 1995.

Alexander G. Tartakovsky and Venugopal V. Veeravalli. Asymptotically optimal quickest change detection in distributed sensor systems. *Sequential Analysis*, 27(4): 441-475, 2008.

Gaël Varoquaux, Alexandre Gramfort, Jean Baptiste Poline and Bertrand Thirion. Brain covariance selection: better individual functional connectivity models using population prior. *Advances in Neural Information Processing Systems*, 2334–2342, 2010.

Abraham Wald. *Sequential Analysis*. Courier Corporation, 1973.

Yao Xie and David Siegmund. Sequential multi-sensor change-point detection. *Annals of Statistics*, 41: 670-692, 2013.