Biometrika (2019), **106**, 3, *pp.* 619–634
Printed in Great Britain
doi: 10.1093/biomet/asz011
Advance Access publication 24 May 2019

Homogeneity tests of covariance matrices with high-dimensional longitudinal data

BY PING-SHOU ZHONG

Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, Illinois 60607, U.S.A.
pszhong@uic.edu

RUNZE LI

Department of Statistics and The Methodology Center, Pennsylvania State University, University Park, Pennsylvania 16801, U.S.A.
rzli@psu.edu

AND SHAWN SANTO

Department of Statistics and Probability, Michigan State University, 619 Red Cedar Road, East Lansing, Michigan 48824, U.S.A.
santosha@stt.msu.edu

SUMMARY

This paper deals with the detection and identification of changepoints among covariances of high-dimensional longitudinal data, where the number of features is greater than both the sample size and the number of repeated measurements. The proposed methods are applicable under general temporal-spatial dependence. A new test statistic is introduced for changepoint detection, and its asymptotic distribution is established. If a changepoint is detected, an estimate of the location is provided. The rate of convergence of the estimator is shown to depend on the data dimension, sample size, and signal-to-noise ratio. Binary segmentation is used to estimate the locations of possibly multiple changepoints, and the corresponding estimator is shown to be consistent under mild conditions. Simulation studies provide the empirical size and power of the proposed test and the accuracy of the changepoint estimator. An application to a time-course microarray dataset identifies gene sets with significant gene interaction changes over time.

Some key words: High-dimensional data; Homogeneity test; Longitudinal data; Spatial and temporal dependence.

1. INTRODUCTION

In a typical time-course microarray dataset, thousands of gene expression values are measured repeatedly from the same subject at different stages in a developmental process (Tai & Speed, 2006). As a motivating example, Taylor et al. (2007) conducted a longitudinal study on 69 patients infected with the hepatitis C virus. The subjects' gene expression values were measured once before treatment and five times during the treatment regimen of pegylated alpha interferon and ribavirin. One purpose of the study was to identify which genes were regulated by treatment. The repeated measurements enable researchers to understand gene regulation over time. An important

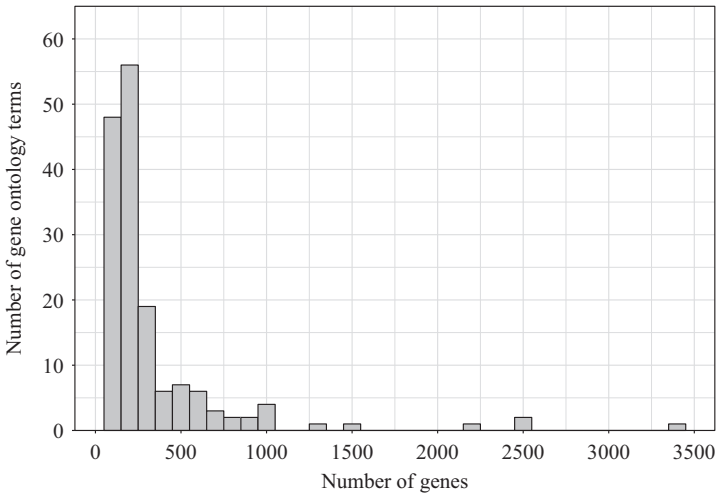

Fig. 1. Histogram of the number of genes among the 159 gene ontology terms analysed.

task in genomic studies is to identify gene sets with significant temporal changes (Storey et al., 2005). Much evidence has shown that gene interaction and coregulation play a critical role in the aetiology of various diseases (Shedden & Taylor, 2005). One application of our methods is to identify gene sets with significant changes in their covariance matrices, because the covariance matrix or its inverse can be used for quantifying interaction and coregulation among genes (Danaher et al., 2015).

Assume that $Y_{it} = (Y_{it1}, \ldots, Y_{itp})^{\mathrm{T}}$ is a p-dimensional random vector with mean μ_t and covariance Σ_t. In the aforementioned applications, the Y_{it} $(i = 1, \ldots, n;\ t = 1, \ldots, T)$ represent gene expressions for p genes in a gene set measured from the ith individual at the tth developmental stage, where n is the sample size and T is the total number of finite stages. The number of genes, p, in a given gene set ranges from a hundred to a few thousand, as illustrated by the histogram in Fig. 1; but n and T are small in the study, so p can be much larger than n and T. We focus on testing the homogeneity of covariance matrices:

$$H_0 : \Sigma_1 = \cdots = \Sigma_T \quad \text{versus} \quad H_1 : \Sigma_k \neq \Sigma_l \tag{1}$$

for some $1 \leqslant k \neq l \leqslant T$. The alternative in (1) can be written as a changepoint-type alternative:

$$H_1 : \Sigma_1 = \cdots = \Sigma_{k_1} \neq \Sigma_{k_1+1} = \cdots = \Sigma_{k_q} \neq \Sigma_{k_q+1} = \cdots = \Sigma_T, \tag{2}$$

where $k_1, \ldots, k_q$ with $1 \leqslant k_1 < \cdots < k_q < T$ are unknown locations of changepoints. This alternative is of interest in practice because it specifies the locations of changes. For example, researchers are often interested in understanding dynamic gene regulation. By identifying the changepoints, we can infer the change pattern of gene regulation, which is important for developing diagnostic and preventive tools for some diseases (Koh et al., 2014).

Testing the homogeneity of covariance matrices is a classical problem in multivariate analysis. Classical methods for testing (1) include the likelihood ratio test (Muirhead, 1982) and Box's M test (Box, 1949). Some resampling methods have been proposed by Zhang & Boos (1992) and Zhu et al. (2002). However, these methods are not valid for the aforementioned applications for the following reasons. First, they require n to be much larger than p, so they are not

applicable under the large-p, small-n paradigm. Second, these methods are only valid for independent samples without temporal dependence, but the independence assumption does not hold for high-dimensional longitudinal data because the repeated measurements obtained from the same individual are temporally dependent.

There is some existing research on testing (1) in the large-p, small-n scenario for independent samples. Li & Chen (2012) considered testing the equality of two covariance matrices for two independent samples. Schott (2007) and Srivastava & Yanagihara (2010) proposed test statistics for (1) based on estimators of the sum of the weighted pairwise Frobenius norm distances between any two covariance matrices. Zheng et al. (2015) and Yang & Pan (2017) applied random matrix theory to test the equality of two large-dimensional covariance matrices.

Some methods have also been presented in the neuroscience literature for the large-p and large-T setting with $T > p$, which is different from our large-p, small-n and small-T set-up. For example, Barnett & Onnela (2016) proposed a sieve bootstrap covariance changepoint detection method that requires removing both boundaries of a time series with length greater than p to avoid ill-conditioned covariance matrices. Laumann et al. (2017) discussed a method for detecting changes in covariances by assessing the stability of multivariate kurtosis via a simulation approach. Their method also requires $T > p$ to ensure the existence of the inverse of a sample covariance matrix. In addition to the aforementioned multivariate detection procedures, a marginal pairwise testing procedure was developed by Zalesky et al. (2014). Their approach relies on using a sliding window to detect changes in correlation coefficients between a pair of coordinates. The p-value for each pair is obtained by resampling residuals after fitting vector autoregressive models. It is then used to test the homogeneity of covariance matrices through multiple testing. Although no existing multivariate method can be applied directly to test (1) for temporally dependent data in the large-p, small-n and small-T setting, it would be an interesting future research topic to develop resampling techniques such as the phase-randomization method (Prichard & Theiler, 1994) for the high-dimensional set-up.

In this paper we propose a new method for testing the equality of covariance matrices for high-dimensional longitudinal data under the large-p, small-n and small-T scenario. The proposed method takes into account both spatial and temporal dependence. Spatial dependence refers to the dependence among different components of Y_{it}, and temporal dependence refers to the dependence between Y_{it} and Y_{is} for any two time-points $t \neq s$. The asymptotic distribution of the proposed test statistic is derived under mild conditions on dependence without any explicit requirement on the relationships between p, n and T.

We also propose a method for estimating the location of changepoints $k_1, \ldots, k_q$ among covariance matrices. There exists some work on identifying changepoints in high-dimensional means, but the literature for high-dimensional covariances is very small. Aue et al. (2009) laid groundwork by considering a p-dimensional multivariate, possibly high-dimensional, time series set-up where T diverges, $n = 1$ and $p < T$. Their test statistic involves the inverse of a $p \times p$ sample covariance matrix, which is singular if $p > T$. Thus, their method is not applicable to high-dimensional longitudinal data. In the case of finite p and n but diverging T, one major concern is that the changepoint estimator is not consistent (Hinkley, 1970) and only the ratios k_i/T $(i = 1, \ldots, q)$ are consistent. When p is finite but $n \to \infty$, it has been shown that changepoints can be estimated consistently. However, it is not clear how the data dimension affects the rate of convergence. We study the rate of convergence of our proposed changepoint estimator and find that it depends on the data dimension, sample size, noise level and signal strength. Consistency of the changepoint estimator is possible even in the high-dimensional case. Furthermore, we develop a binary segmentation procedure for identifying the locations of multiple changepoints, whose consistency is also established.

Our work is related to, but different from, that of Li & Chen (2012), who considered a test for the equality of two covariance matrices with two independent samples. First, we consider a general homogeneity test of covariance matrices with more than two populations, while Li and Chen dealt with only a two-sample case. Second, Li and Chen's test was for two independent samples, but our proposed method can accommodate both temporal and spatial dependence. Moreover, our method is designed to test for the existence of changepoints among high-dimensional covariance matrices for longitudinal data. Therefore, the test procedure considered in this paper differs from that in Li & Chen (2012).

This paper makes the following contributions. From a methodological perspective, the proposed test procedure provides a novel solution to changepoint detection problems in the large-p, small-n and small-T scenario. The test statistic combines the strengths of the maximum and Frobenius norms, and is powerful against the alternative. Second, we propose a method for estimating locations of changepoints among high-dimensional covariance matrices. The proposed changepoint detection and identification procedures are widely applicable without any sparsity assumption. We establish the asymptotic distribution of a test statistic for data with general temporal and spatial dependence. The identification procedure for multiple changepoints is shown to be consistent. Our results reveal the effects of data dimension, sample size, and signal-to-noise ratio on the rate of convergence of the changepoint estimator. The proposed methods formally address two challenges that are unsolved in the existing covariance changepoint literature: the large-p, small-n and small-T issue, and spatial and temporal dependence.

2. BASIC SETTING

Let $Y_{it} = (Y_{it1}, \ldots, Y_{itp})^{\mathrm{T}}$ be the observed p-dimensional random vector for the ith individual at time-point $t = 1, \ldots, T$, where $T \geqslant 2$ and $i = 1, \ldots, n$. Assume that Y_{it} follows the model

$$Y_{it} = \mu_t + \varepsilon_{it}, \tag{3}$$

where μ_t is a p-dimensional unknown mean vector and $\varepsilon_{it} = (\varepsilon_{it1}, \ldots, \varepsilon_{itp})^{\mathrm{T}}$ is a multivariate normally distributed random error vector with mean zero and covariance $\mathrm{var}(\varepsilon_{it}) = \Sigma_t$. A generalization to the non-Gaussian set-up is given in the Supplementary Material. In addition, it is assumed that $\varepsilon_{it} = \Gamma_t Z_i$, where Γ_t is a $p \times m$ matrix with $m \geqslant pT$ and Z_i is an m-dimensional standard multivariate normally distributed random vector, so that $\mathrm{cov}(\varepsilon_{is}, \varepsilon_{jt}) = \Gamma_s \Gamma_t^{\mathrm{T}} = C_{st}$ if $i = j \in \{1, \ldots, n\}$ and $\mathrm{cov}(\varepsilon_{is}, \varepsilon_{jt}) = 0$ if $i \neq j$. The random errors $\{\varepsilon_{it}\}_{i=1}^{n}$ are independent, but $\{\varepsilon_{it}\}_{t=1}^{T}$ depend on each other. Of interest is to test whether any changepoints among covariances occur at some time-points $t \in \{1, \ldots, T-1\}$. We test the hypothesis H_0 versus H_1 specified in (1) and (2). If H_0 is rejected, we further estimate the locations of changepoints.

3. HOMOGENEITY TESTS OF COVARIANCE MATRICES

At each $t \in \{1, \ldots, T-1\}$, we define a measure $D_t = w^{-1}(t) \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} \mathrm{tr}\{(\Sigma_{s_1} - \Sigma_{s_2})^2\}$, where $w(t) = t(T-t)$. The measure D_t characterizes the differences between the covariances before t and after t. Clearly, $D_t = 0$ for all $t \in \{1, \ldots, T-1\}$ under H_0, and $D_t > 0$ for any t under H_1. Therefore, $\max_{1 \leqslant t \leqslant T-1} D_t = 0$ under H_0 and $\max_{1 \leqslant t \leqslant T-1} D_t > 0$ under H_1. Thus, D_t is useful for distinguishing the null and alternative hypotheses.

Measure D_t is different from the measure $S_{1,T} = \sum_{s_1=1}^{T-1} \sum_{s_2=s_1+1}^{T} \mathrm{tr}\{(\Sigma_{s_1} - \Sigma_{s_2})^2\}$ in Schott (2007), who used $S_{1,T}$ in constructing a homogeneity test as specified in (1) for independent

samples. In fact, for any $t \in \{1, \ldots, T-1\}$, $D_t = S_{1,T} - (S_{1,t} + S_{t+1,T})$, where $S_{1,t}$ and $S_{t+1,T}$ quantify the differences between covariances only before time t and only after time t, respectively. These are not useful for measuring the differences between covariances before and after time t. Measure D_t removes both $S_{1,t}$ and $S_{t+1,T}$ from $S_{1,T}$.

To construct an unbiased estimator of D_t, we need an unbiased estimator of $\mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})$. We make use of U-statistic-type estimators because they avoid bias that is not ignorable in a high-dimensional setting (Bai & Saranadasa, 1996; Chen & Qin, 2010). Otherwise, bias correction could be a challenge and require conditions on the data dimension and sample size that limit the scope of applications. Let $\widetilde{\sum}$ denote summation over mutually different indices of sample subjects. For example, $\widetilde{\sum}_{i,j,k}$ means summation over $\{(i,j,k) \in \{1,\ldots,n\} : i \neq j, j \neq k, k \neq i\}$. For any $s_1, s_2 \in \{1,\ldots,T\}$, define $U_{s_1s_2,0} = (1/P_n^2)\sum_{i\neq j}^{n}(Y_{is_1}^{\mathrm{T}}Y_{js_2})^2$ as an unbiased estimator of $\mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2}) + \mu_{s_1}^{\mathrm{T}}\Sigma_{s_2}\mu_{s_1} + \mu_{s_2}^{\mathrm{T}}\Sigma_{s_1}\mu_{s_2} + (\mu_{s_1}^{\mathrm{T}}\mu_{s_2})^2$, where $P_n^k = n!/(n-k)!$. To remove the nuisance terms $\mu_{s_1}^{\mathrm{T}}\Sigma_{s_2}\mu_{s_1}$ and $(\mu_{s_1}^{\mathrm{T}}\mu_{s_2})^2$, we define $U_{s_1s_2,1} = (1/P_n^3)\widetilde{\sum}_{i,j,k} Y_{is_1}^{\mathrm{T}}Y_{js_2}Y_{js_2}^{\mathrm{T}}Y_{ks_1}$ as an unbiased estimator of $\mu_{s_1}^{\mathrm{T}}\Sigma_{s_2}\mu_{s_1} + (\mu_{s_1}^{\mathrm{T}}\mu_{s_2})^2$; similarly, $U_{s_2s_1,1}$ is an unbiased estimator of $\mu_{s_2}^{\mathrm{T}}\Sigma_{s_1}\mu_{s_2} + (\mu_{s_1}^{\mathrm{T}}\mu_{s_2})^2$. To remove the nuisance term $(\mu_{s_1}^{\mathrm{T}}\mu_{s_2})^2$, we define $U_{s_1s_2,2} = (1/P_n^4)\widetilde{\sum}_{i,j,k,l} Y_{is_1}^{\mathrm{T}}Y_{js_2}Y_{ks_1}^{\mathrm{T}}Y_{ls_2}$ as an unbiased estimator of $(\mu_{s_1}^{\mathrm{T}}\mu_{s_2})^2$. A computationally efficient formulation of $U_{s_1s_2,1}$ and $U_{s_1s_2,2}$ is given in the Appendix. Finally, we define an unbiased estimator for $\mathrm{tr}(\Sigma_{s_1}\Sigma_{s_2})$ as

$$U_{s_1s_2} = U_{s_1s_2,0} - U_{s_1s_2,1} - U_{s_2s_1,1} + U_{s_1s_2,2}. \tag{4}$$

The estimator $U_{s_1s_2}$ is a generalization of the estimator for the trace of the covariance given by Chen et al. (2010) and Li & Chen (2012). For $t = 1,\ldots,T-1$, an unbiased estimator of D_t is

$$\hat{D}_{nt} = \frac{1}{w(t)}\sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T}(U_{s_1s_1} + U_{s_2s_2} - U_{s_1s_2} - U_{s_2s_1}).$$

To study the asymptotic variance of $\hat{D}_{nt}$ for $t = 1,\ldots,T-1$, define

$$V_{0t} = \sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*}\sum_{\substack{u,v,\\ k,l\in\{1,2\}}}(-1)^{|u-v|+|k-l|}\,\mathrm{tr}^2(C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}})$$

and

$$V_{1t} = \sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*}\sum_{u,k\in\{1,2\}}(-1)^{|u-k|}\,\mathrm{tr}\big\{(\Sigma_{s_1} - \Sigma_{s_2})C_{s_uh_k}(\Sigma_{h_1} - \Sigma_{h_2})C_{s_uh_k}^{\mathrm{T}}\big\},$$

where $\sum_{\substack{s_1,s_2,\\ h_1,h_2}}^{*} = \sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T}\sum_{h_1=1}^{t}\sum_{h_2=t+1}^{T}$. If no temporal dependence exists, then $C_{s_uh_k} = 0$ for any $s_u \neq h_k$, and $V_{0t} = \sum_{s_1,s_2}^{*}\sum_{u,v\in\{1,2\}}\mathrm{tr}^2(\Sigma_{s_u}\Sigma_{s_v})$ where $\sum_{s_1,s_2}^{*} = \sum_{s_1=1}^{t}\sum_{s_2=t+1}^{T}$. Up to a scale factor, this V_{0t} is the portion of the variance of $\hat{D}_{nt}$ for the case with independent samples under H_0.

The asymptotic setting considered in this paper is $p(n) \to \infty$ as $n \to \infty$, where p is considered to be a function of n. We do not require a specific relationship between p and n. Instead, for any

$t \in \{1, \ldots, T-1\}$ we have two regularity conditions. Writing $A^{\otimes 2} = AA^{\mathrm{T}}$ for any matrix A, the conditions are as follows:

Condition 1. $\mathrm{tr}\{(\Gamma_{s_2}^{\mathrm{T}} C_{s_1 h_1} \Gamma_{h_2})^{\otimes 2}\} = o(V_{0t})$ for any $s_1, s_2, h_1, h_2 \in \{1, \ldots, T\}$;

Condition 2. $\mathrm{tr}[\{(\Gamma_{s_1} + \Gamma_{s_2})^{\mathrm{T}}(\Sigma_{s_1} - \Sigma_{s_2})(\Gamma_{s_1} - \Gamma_{s_2})\}^{\otimes 2}] = o(nV_{1t})$ for $s_1 \in \{1, \ldots, t\}$ and $s_2 \in \{t+1, \ldots, T\}$.

Condition 1 generalizes Condition 2, which was imposed by Li & Chen (2012), to a T-sample test with temporal dependence. If there is no temporal dependence, Condition 1 can be simplified to $\mathrm{tr}(\Sigma_{s_2}\Sigma_{s_1}\Sigma_{h_2}\Sigma_{s_1}) = o(V_{0t})$. In general, the left-hand side of the equality in Condition 1 is bounded by $\{\mathrm{tr}(\Sigma_{h_2}\Sigma_{h_1}\Sigma_{h_2}\Sigma_{h_1})\mathrm{tr}(\Sigma_{s_2}\Sigma_{s_1}\Sigma_{s_2}\Sigma_{s_1})\}^{1/2}$, which is of order $O(p)$ if all the eigenvalues of Σ_t are bounded. If the temporal dependence is not overwhelming so that $V_{0t} \asymp p^{\delta}$ for any $\delta > 1$, then Condition 1 holds. To appreciate this point, consider a null hypothesis case with $C_{st} = (1 - r_{st,n})\Sigma$ for $s, t \in \{1, \ldots, T\}$. Here $1 - r_{st,n}$ measures the temporal correlation. If $r_{st,n}$ is small for all s and t, then the temporal dependence among $\{Y_{it}\}_{t=1}^{T}$ is strong. Let $r_n = \sum_{s_1,s_2,h_1,h_2}^{*} \sum_{u,v,k,l \in \{1,2\}} (-1)^{|u-v|+|k-l|} r_{s_u h_k, n} r_{s_v h_l, n}$. If $r_{st,n} \to 0$ for all s and t, then $V_{0t} \asymp r_n \mathrm{tr}^2(\Sigma^2) \asymp r_n p^2$ provided all the eigenvalues of Σ are bounded. If the temporal dependence is not too strong so that $1/p = o(r_n)$, then Condition 1 holds as $p \to \infty$. Intuitively, Condition 1 implies that spatial and temporal dependence cannot be too strong.

Condition 2 is automatically true under H_0 because its left-hand side equals zero; hence it is not needed under H_0. If there is no temporal dependence, it can be shown that the left-hand side of Condition 2 is $\mathrm{tr}\{(\Sigma_{s_1}^2 - \Sigma_{s_2}^2)^2\}$, whose order is no greater than V_{1t}. Therefore, Condition 2 is not needed for data without temporal dependence. This condition implies that the alternatives should not be too far away from the null hypothesis; otherwise, the alternatives would be easy to detect because the test statistics would diverge to infinity.

Theorem 1 states the mean and variance of $\hat{D}_{nt}$. The proof is given in the Supplementary Material.

THEOREM 1. *The expectation of $\hat{D}_{nt}$ is $E(\hat{D}_{nt}) = D_t$. Under Condition 1, the leading-order variance of $\hat{D}_{nt}$ is $\sigma_{nt}^2 = w^{-2}(t)(4V_{0t}/n^2 + 8V_{1t}/n)$.*

Based on Theorem 1, we observe that $E(\hat{D}_{nt}) = D_t = 0$ under H_0. Under alternative H_1 in (2), it is clear that $E(\hat{D}_{nt}) > 0$ for all t under H_1. Therefore, $\hat{D}_{nt}$ is able to distinguish the null and alternative hypotheses in (1) and (2).

If $T = 2$ and no temporal dependence exists, V_{0t} and V_{1t} are, respectively, simplified to $V_{01} = \mathrm{tr}^2(\Sigma_1^2) + 2\,\mathrm{tr}^2(\Sigma_1\Sigma_2) + \mathrm{tr}^2(\Sigma_2^2)$ and $V_{11} = \sum_{s_1,s_2}^{*} \sum_{u=1}^{2} \mathrm{tr}[\{\Sigma_{s_u}(\Sigma_{s_1} - \Sigma_{s_2})\}^2]$, the same as the expressions obtained by Li & Chen (2012). For a general case with temporal dependence, $V_{01} = \mathrm{tr}^2(\Sigma_1^2) + 2\,\mathrm{tr}^2(\Sigma_1\Sigma_2) + \mathrm{tr}^2(\Sigma_2^2) - 4\{\mathrm{tr}^2(\Sigma_1 C_{21}) + \mathrm{tr}^2(\Sigma_2 C_{12})\} + 2\{\mathrm{tr}^2(C_{12}C_{12}^{\mathrm{T}}) + \mathrm{tr}^2(C_{12}C_{12})\}$. The last four terms in V_{01}, because of the temporal dependence, are not included in Li and Chen's test. However, in general these four terms are not ignorable. Therefore, Li and Chen's procedure is not suitable for temporally dependent data even in the two-sample case.

We now study the asymptotic distribution of $\hat{D}_{nt}$. The following theorem establishes the asymptotic normality of $\hat{D}_{nt}$. The proof is given in the Supplementary Material.

THEOREM 2. *Under Conditions 1 and 2, $\sigma_{nt}^{-1}(\hat{D}_{nt} - D_t) \to N(0,1)$ in distribution as $n \to \infty$, where σ_{nt}^2 is defined in Theorem 1.*

We do not require explicit conditions on p and n in Theorem 2. The asymptotic normality holds provided Conditions 1 and 2 hold. In particular, we only need Condition 1 under the null

hypothesis. Thus, our test is valid under Condition 1 without Condition 2, which is needed only for studying the power of the test. The normality assumption in model (3) is not essential and can be relaxed to a multivariate model as considered in Chen et al. (2010) and Li & Chen (2012). See the Supplementary Material for the generalization to the non-Gaussian case.

Under H_0, $D_t = 0$ for all $t \in \{1, \ldots, T-1\}$. Theorem 2 indicates that $\sigma_{nt,0}^{-1}\hat{D}_{nt}$ converges to $N(0,1)$ in distribution where $\sigma_{nt,0}^2 = 4V_{0t}/\{nw(t)\}^2$ is the variance of $\hat{D}_{nt}$ under H_0. An asymptotic α-level rejection region is $R_t = \{\sigma_{nt,0}^{-1}\hat{D}_{nt} > z_\alpha\}$, where z_α is the upper α-quantile of the standard normal distribution. For each $t \in \{1, \ldots, T-1\}$, one can use R_t to test the hypothesis in (1). Provided that one test based on $\hat{D}_{nt}$ rejects the null hypothesis, one might suspect that changepoints exist among the covariance matrices. Accordingly, t in $\hat{D}_{nt}$ could be considered as a tuning parameter, and it is hard to decide which t should be used for testing in practice. To make the proposed method free of any tuning parameter and adaptive to unknown changepoints, we propose the following statistic for testing the hypothesis in (1):

$$\mathcal{M}_n = \max_{1 \leqslant t \leqslant T-1} \hat{\sigma}_{nt,0}^{-1}\hat{D}_{nt}, \tag{5}$$

where $\hat{\sigma}_{nt,0}^2 = 4\hat{V}_{0t}/\{nw(t)\}^2$. The estimator $\hat{V}_{0t}$ can be constructed by replacing $\mathrm{tr}(C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}})$ in V_{0t} with $U_{s_us_v,h_kh_l}$, an unbiased estimator of $\mathrm{tr}(C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}})$. Define $U_{s_us_v,h_kh_l} = U_{s_us_v,h_kh_l,0} - U_{s_us_v,h_kh_l,1} - U_{s_vs_u,h_lh_k,1} + U_{s_us_v,h_kh_l,2}$, where $U_{s_us_v,h_kh_l,0} = (1/P_n^2)\sum_{i\neq j=1}^{n} Y_{is_u}^{\mathrm{T}}Y_{js_v}Y_{ih_k}^{\mathrm{T}}Y_{jh_l}$ is an unbiased estimator of $\mathrm{tr}(C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}}) + \mu_{s_v}^{\mathrm{T}}C_{s_uh_k}\mu_{h_l} + \mu_{s_u}^{\mathrm{T}}C_{s_vh_l}\mu_{h_k} + \mu_{s_u}^{\mathrm{T}}\mu_{s_v}\mu_{h_k}^{\mathrm{T}}\mu_{h_l}$, $U_{s_us_v,h_kh_l,1} = (1/P_n^3)\widetilde{\sum}_{i,j,g} Y_{is_u}^{\mathrm{T}}Y_{js_v}Y_{ih_k}^{\mathrm{T}}Y_{gh_l}$ is an unbiased estimator of $\mu_{s_v}^{\mathrm{T}}C_{s_uh_k}\mu_{h_l} + \mu_{s_u}^{\mathrm{T}}\mu_{s_v}\mu_{h_k}^{\mathrm{T}}\mu_{h_l}$, and $U_{s_us_v,h_kh_l,2} = (1/P_n^4)\times \widetilde{\sum}_{i,j,g,f} Y_{is_u}^{\mathrm{T}}Y_{js_v}Y_{gh_k}^{\mathrm{T}}Y_{fh_l}$ is an unbiased estimator of $\mu_{s_u}^{\mathrm{T}}\mu_{s_v}\mu_{h_k}^{\mathrm{T}}\mu_{h_l}$. A computationally efficient formulation of the estimators $U_{s_us_v,h_kh_l,q}$ $(q = 1, 2)$ is similar to that for $U_{s_1,s_2,q}$ defined in (4).

Under H_0 and Condition 1, following a derivation similar to that in Lemma S4 in the Supplementary Material, the leading order of the $\mathrm{cov}(\hat{D}_{nt}, \hat{D}_{nq})$ can be shown to be $Q_{n,tq}$, where $Q_{n,tq} = \sum_{s_1=1}^{t}\sum_{h_1=1}^{q}\sum_{s_2=t+1}^{T}\sum_{h_2=q+1}^{T} V_{n0}(s_1,s_2,h_1,h_2)/\{w(t)w(q)\}$ and $V_{n0}(s_1,s_2,h_1,h_2) = (4/n^2)\sum_{u,v,k,l\in\{1,2\}}(-1)^{|u-v|+|k-l|}\,\mathrm{tr}^2(C_{s_uh_k}C_{s_vh_l}^{\mathrm{T}})$. Then the covariance between $\sigma_{nt,0}^{-1}\hat{D}_{nt}$ and $\sigma_{nq,0}^{-1}\hat{D}_{nq}$ is $Q_{n,tq}/\sqrt{(Q_{n,tt}Q_{n,qq})}$, which is the correlation between $\hat{D}_{nt}$ and $\hat{D}_{nq}$.

Let V_{nD} be a correlation matrix whose (t,s) component is $Q_{n,ts}/\sqrt{(Q_{n,tt}Q_{n,ss})}$ for $t,s \in \{1, \ldots, T-1\}$. Assume that V_{nD} converges to V_D as $n \to \infty$. The following theorem provides the asymptotic distribution of $\mathcal{M}_n$.

THEOREM 3. *Under Condition 1, we have that under H_0, $\mathcal{M}_n \to W$ in distribution as $n \to \infty$, where $W = \max_{1\leqslant t\leqslant T-1} Z_t$ and $Z = (Z_1, \ldots, Z_{T-1})^{\mathrm{T}}$ is a multivariate normally distributed random vector with mean zero and covariance V_D.*

According to Theorem 3, an α-level test for (1) rejects the null hypothesis if $\mathcal{M}_n > W_\alpha$, where W_α is the α-quantile of W such that $\mathrm{pr}(W > W_\alpha) = \alpha$. Let Z_n be a $N(0, \hat{V}_{nD})$-distributed random vector with the (t,s) component of $\hat{V}_{nD}$ estimated by $\hat{Q}_{n,ts}/\sqrt{(\hat{Q}_{n,tt}\hat{Q}_{n,ss})}$, where

$$\hat{Q}_{n,ts} = \frac{4}{n^2w(t)w(s)}\sum_{s_1=1}^{t}\sum_{h_1=1}^{s}\sum_{s_2=t+1}^{T}\sum_{h_2=s+1}^{T}\sum_{u,v,k,l\in\{1,2\}}(-1)^{|u-v|+|k-l|}\,U_{s_us_v,h_kh_l}^2$$

with $U_{s_u s_v, h_k h_l}$ defined just below (5). Simulations suggest that the plug-in estimates of the correlation matrix $\hat{V}_{nD}$ are reliable when the sample size is approximately 40 or above. See the Supplementary Material for a detailed comparison of $\hat{V}_{nD}$ and V_{nD}. The quantile W_α can be approximated by $W_{n,\alpha}$ obtained from the multivariate normal distribution by finding the quantile $w_{n,\alpha} = (W_{n,\alpha}, \ldots, W_{n,\alpha})^{\mathrm{T}}$ satisfying $\mathrm{pr}(Z_n < w_{n,\alpha}) = 1 - \alpha$. The quantile $w_{n,\alpha}$ can be computed using the `R` (R Development Core Team, 2019) package `mvtnorm` (Genz et al., 2018), and no simulation is needed to find $W_{n,\alpha}$.

The lower bound for the power based on $\mathcal{M}_n$ is

$$\mathrm{pr}(\mathcal{M}_n > W_\alpha) \geqslant \max_{1 \leqslant t \leqslant T-1} \mathrm{pr}(\hat{\sigma}_{nt,0}^{-1} \hat{D}_{nt} > W_\alpha) = \max_{1 \leqslant t \leqslant T-1} \Phi\left(-\frac{\sigma_{nt,0}}{\sigma_{nt}} W_\alpha + \frac{D_t}{\sigma_{nt}} \right), \tag{6}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. If D_t/σ_{nt} dominates W_α, the right-hand side of (6) is the maximum power of the test using R_t constructed on a single $\hat{D}_{nt}$, so the test based on $\mathcal{M}_n$ is more powerful than any test based on a single $\hat{D}_{nt}$.

The test statistic $\mathcal{M}_n$ can be improved under sparse alternatives where the differences among the Σ_t reside in only a few components. Inspired by Fan et al. (2015), we propose a power-enhanced test statistic, $\mathcal{M}_n^*$, to improve the power under sparse alternatives. Let $\bar{Y}_{s_1 v} = \sum_{i=1}^{n} Y_{i s_1 v}/n$ be the sample mean and $\hat{\sigma}_{s_1,uv} = \sum_{i=1}^{n-1} (Y_{i s_1 u} - \bar{Y}_{s_1 u})(Y_{i s_1 v} - \bar{Y}_{s_1 v})/(n-1)$ the sample covariance between components $u, v \in \{1, \ldots, p\}$. Define $\hat{D}_{nt,uv}$ as a plug-in estimator of $D_{nt,uv} = \sum_{s_1=1}^{t} \sum_{s_2=t+1}^{T} (\sigma_{s_1,uv} - \sigma_{s_2,uv})^2$, where $\sigma_{s_1,uv}$ is the (u,v) component of Σ_{s_1}. Let $C_{s_k h_t}^{(uv)}$ be the (u,v) component of $C_{s_k h_t}$, and define $G_{s_k s_l h_t h_s}^{(uv)} = \{C_{s_k h_s}^{(vu)} C_{s_k h_s}^{(uv)} + C_{s_k h_s}^{(vv)} C_{s_k h_s}^{(uu)}\}\{C_{s_l h_t}^{(vu)} C_{s_l h_t}^{(uv)} + C_{s_l h_t}^{(vv)} C_{s_l h_t}^{(uu)}\} + \{C_{s_k h_t}^{(vu)} C_{s_k h_t}^{(uv)} + C_{s_k h_t}^{(vv)} C_{s_k h_t}^{(uu)}\}\{C_{s_l h_s}^{(vu)} C_{s_l h_s}^{(uv)} + C_{s_l h_s}^{(vv)} C_{s_l h_s}^{(uu)}\}$. The variance of $\hat{D}_{nt,uv}$ under H_0 is $\sigma_{nt,uv0}^2 = \sum_{h_1, h_2}^{*} \sum_{s,t \in \{1,2\}}^{k,l} (-1)^{|k-l|+|s-t|} G_{s_k s_l h_t h_s}^{(uv)}/n^2$. The power-enhanced test statistic is

$$\mathcal{M}_n^* = \max_{1 \leqslant t \leqslant T-1} \left\{ \hat{\sigma}_{nt,0}^{-1} \hat{D}_{nt} + \lambda_n \sum_{u \leqslant v} I(\hat{D}_{nt,uv} > \delta_{n,p} \hat{\sigma}_{nt,uv0}) \right\},$$

where $\hat{\sigma}_{nt,uv0}^2$ is a plug-in estimator of $\sigma_{nt,uv0}^2$ and $I(\cdot)$ denotes the indicator function. The choices of the tuning parameters $\delta_{n,p}$ and λ_n are discussed in the Supplementary Material. A numerical simulation in the Supplementary Material illustrates the performance of $\mathcal{M}_n^*$ under sparse alternatives.

4. CHANGEPOINT IDENTIFICATION

If H_0 is rejected, then there exist changepoints among the covariances Σ_t. We first consider an alternative with one changepoint:

$$H_1^* : \Sigma_1 = \cdots = \Sigma_{k_1} \neq \Sigma_{k_1+1} = \cdots = \Sigma_T, \tag{7}$$

where k_1 is the true changepoint, whose location is estimated by

$$\hat{k}_1 = \underset{1 \leqslant t \leqslant T-1}{\arg\max} \ \hat{D}_{nt}. \tag{8}$$

Define the weight function

$$r(t;k) = \begin{cases} (T-k)/(T-t), & 1 \leqslant t \leqslant k, \\ k/t, & k+1 \leqslant t \leqslant T-1. \end{cases}$$

For any fixed value $k \in \{1, \ldots, T-1\}$, the function $r(t;k)$ achieves its maximum value at $t = k$. Let $\beta_n = \max_{1 \leqslant t \leqslant T-1} \max\{\sqrt{V_{0t}}, \sqrt{(nV_{1t})}\}$ and $\Delta_n = \mathrm{tr}\{(\Sigma_1 - \Sigma_T)^2\}$. The next theorem establishes the rate of convergence of the changepoint estimator $\hat{k}_1$ obtained by (8) under the alternative H_1^*.

THEOREM 4. *Under the alternative H_1^* in (7), $E(\hat{D}_{nt}) = D_t = r(t;k_1)\Delta_n$ and D_t attains its maximum at $t = k_1$. Moreover, $\hat{k}_1 - k_1 = O_{\mathrm{p}}\{\beta_n/(n\Delta_n)\}$.*

Since $r(t;k_1)$ achieves its maximum at $t = k_1$, the first part of Theorem 4 indicates that $t = k_1$ maximizes $E(\hat{D}_{nt})$ as a function of t. This is the rationale for estimating k_1 through (8). When the data dimension is fixed, $\hat{k}_1 - k_1 = O_{\mathrm{p}}(1/\sqrt{n})$. The effect of the data dimension is reflected in both β_n and Δ_n. Here β_n can be considered as noise and Δ_n can be viewed as the signal. If the signal level is larger than the noise level, the rate of convergence of $\hat{k}_1$ is faster than $O_{\mathrm{p}}(1/\sqrt{n})$. On the other hand, if β_n is not smaller than $n\Delta_n$, $\hat{k}_1$ is not consistent.

Next, we consider the alternative H_1 with multiple changepoints $k_1 < \cdots < k_q$, as specified in (2). Under H_1^*, we have shown in Theorem 4 that the maximum of D_t is attained at changepoint k_1.

THEOREM 5. *Under H_1 in (2), the maximum value of D_t is attained at one of the changepoints among $k_1 < \cdots < k_q$.*

If we estimate the multiple changepoints by repeatedly applying the estimation method in (8) for the population version D_t to all subsequences with nonzero D_t, Theorem 5 ensures that we find all the true changepoints. This property is important for applying the binary segmentation method to identify multiple changepoints, as demonstrated by E. Venkatraman in a 1992 unpublished technical report from the Department of Statistics at Stanford University.

To describe the proposed binary segmentation method, we first define some notation. Let $[I_t]$ represent the quantities computed based on the data within the time interval I_t, a subset of $[1, T]$. For example, $\mathcal{M}_n[t_1, t_2]$ is the test statistic defined in (5) calculated based on $Y[t_1, t_2]$, the data collected between times $t = t_1$ and $t = t_2$ for $t_1 < t_2$. Specifically, $\mathcal{M}_n[t_1, t_2] = \max_{t_1 \leqslant t < t_2} \hat{\sigma}_{nt,0}^{-1}[t_1, t_2]\hat{D}_{nt}[t_1, t_2]$.

The binary segmentation method can be summarized as follows. Let α_n be a number specified in Theorem 6. In the first step, compute $\mathcal{M}_n[1, T]$. If $\mathcal{M}_n[1, T] < W_{\alpha_n}[1, T]$, where $W_{\alpha_n}[1, T]$ is the cut-off quantile estimated based on $Y[1, T]$, we accept the null hypothesis and stop. Otherwise, we identify the changepoint, say $\hat{k}_1$, using (8). Next, we compute $\mathcal{M}_n$ for both subsequences $Y[1, \hat{k}_1]$ and $Y[\hat{k}_1 + 1, T]$. For each subsequence, we repeat the first step until no changepoints can be identified or the number of repeated measurements in the subsequence is less than two.

Let I_t be any interval of the form $[k_f + 1, k_g]$ (where $f \in \{0, \ldots, q-1\}$ and $g \in \{2, \ldots, q+1\}$ with $f + 1 < g$, $k_0 = 0$ and $k_{q+1} = T$) that contains at least one changepoint k_j for $j \in \{1, \ldots, q\}$. Define $\mathrm{mSNR} = \min_{I_t} \max_{k_s \in I_t} \sigma_{nk_s,0}^{-1}[I_t]D_{k_s}[I_t]$, the smallest maximum signal-to-noise ratio among all segmentations I_t.

THEOREM 6. *Assume that $\alpha_n \to 0$ and* mSNR *diverges so that $W_{\alpha_n} = o(\mathrm{mSNR})$. For all I_t, if $\beta_n[I_t] = o(nD_{k_s}[I_t])$ for some changepoints $k_s \in I_t$, then $\lim_{n\to\infty} \mathrm{pr}(\hat{q} = q; \hat{k}_j = k_j, j = 1, \ldots, q) = 1$.*

Table 1. *Empirical size and power of the proposed test: percentages of simulation replications that reject the null hypothesis under settings* (I) *and* (II)

			$T = 5$			$T = 8$		
			p			p		
Setting	δ	n	500	750	1000	500	750	1000
(I)		40	4.6	4.8	6.4	4.8	4.8	4.4
	0 (size)	50	4.6	5.2	5.4	4.4	5.8	4.6
		60	6.0	4.4	4.2	5.4	4.2	3.6
		40	21.4	27.6	24.8	35.6	34.6	34.2
	0.05	50	37.0	36.0	36.0	49.8	48.8	52.0
		60	45.6	49.2	46.2	59.6	65.6	65.0
		40	99.6	100	99.8	100	100	100
	0.10	50	100	100	100	100	100	100
		60	100	100	100	100	100	100
(II)		40	4.4	5.4	5.0	4.4	4.0	4.8
	0 (size)	50	5.6	4.6	4.8	6.0	5.2	5.6
		60	4.8	4.6	4.2	3.6	5.6	5.0
		40	33.4	35.8	38.2	50.2	52.0	51.6
	0.10	50	44.2	48.6	47.0	68.4	70.6	74.0
		60	65.4	63.6	60.4	87.0	89.6	88.0
		40	99.8	99.8	99.6	100	100	100
	0.20	50	99.8	100	100	100	100	100
		60	100	100	100	100	100	100

The first assumption, $W_{\alpha_n} = o(\text{mSNR})$, is a very mild condition, which ensures the consistency of the proposed test at each step of the binary segmentation. The second assumption, $\beta_n[I_t] = o(nD_{k_s}[I_t])$, is needed to ensure the consistency of the changepoint estimators. Theorem 6 implies that the proposed binary segmentation procedure consistently estimates the number and locations of changepoints.

5. SIMULATION STUDIES

In this section, we present multiple simulation studies to demonstrate the finite-sample performance of the proposed method. The data were generated from the model

$$Y_{it} = \mu_t + \sum_{h=0}^{L} A_{t,h}\eta_{i(t-h)} \quad (i = 1, \ldots, n;\ t = 1, \ldots, T),$$

where $A_{t,h}$ is a $p \times p$ matrix, $\mu_t = 0$, and η_{it} are p-dimensional multivariate normally distributed random vectors with mean 0 and covariance I_p. Let $t \geqslant s$. This implies that $\text{cov}(Y_{it}, Y_{is}) = \sum_{h=t-s}^{L} A_{t,h}A_{s,h-(t-s)}^{\mathrm{T}}$ if $t - s \leqslant L$ and $\text{cov}(Y_{it}, Y_{is}) = 0$ if $t - s > L$, and allows dependence among components within the vector Y_{it} and dependence among $\{Y_{it}\}_{t=1}^{T}$ at different time-points. In the simulation studies, we set $n = 40, 50$ and 60, $p = 500, 750$ and 1000, $T = 5$ and 8, and $L = 3$. The simulation results reported in Tables 1 and 2 were based on 500 replications. The results in Table 3 were based on 100 simulation replications. More simulation results for non-Gaussian random vectors and sparse alternatives, as well as a numerical comparison with a pairwise-based method, are presented in the Supplementary Material.

Table 2. *Percentages of correct changepoint identification among all rejected hypotheses under settings* (I) *and* (II)

			$T=5$			$T=8$		
			p			p		
Setting	δ	n	500	750	1000	500	750	1000
(I)		40	41.12	37.96	40.65	30.18	29.88	37.58
	0.05	50	51.35	45.81	43.33	39.52	39.34	41.54
		60	52.63	53.28	52.17	49.33	49.70	55.08
		40	93.17	96.60	95.19	93.79	93.80	96.40
	0.10	50	98.00	98.60	98.20	98.40	97.20	99.00
		60	99.20	99.80	99.40	99.80	98.60	99.00
(II)		40	49.10	45.51	55.50	43.12	47.15	47.10
	0.10	50	65.00	61.51	55.98	53.80	61.19	58.65
		60	72.78	69.72	64.57	67.59	72.99	75.23
		40	90.58	90.98	89.16	95.80	96.00	95.60
	0.20	50	93.37	92.60	93.20	98.60	98.20	99.40
		60	97.00	96.20	96.40	99.80	99.80	99.80

Let $k_1 = [T/2]$ be the largest integer no greater than $T/2$. For $t \in \{1, \ldots, k_1\}$, we set $A_{t,h} = A^{(1)}$ for $h \in \{0, \ldots, L\}$. For $t \in \{k_1+1, \ldots, T\}$ and $h \in \{0, \ldots, L\}$, $A_{t,h} = A^{(2)}$. Two simulation settings were used for the generation of the A matrices. In setting (I), we set $A^{(1)} = \{0.6^{|i-j|} I(|i-j| < p/5)\}$ and $A^{(2)} = \{(0.6+\delta)^{|i-j|} I(|i-j| < p/5)\}$. If $\delta = 0$, $A^{(1)}$ and $A^{(2)}$ are the same and the covariances of Y_{it} are the same for all t. Hence the null hypothesis, H_0, is true. If $\delta \neq 0$, the null hypothesis is false and k_1 is the true changepoint. In setting (II), we set $A^{(1)} = \{(|i-j|+1)^{-2} I(|i-j| < p/5)\}$ and $A^{(2)} = \{(|i-j|+\delta+1)^{-2} I(|i-j| < p/5)\}$. As in setting (I), a value of $\delta = 0$ corresponds to the null hypothesis being true. If $\delta \neq 0$, k_1 is the underlying true changepoint for the covariance matrices.

Table 1 reports the empirical size and power of the proposed test for the homogeneity of covariance matrices under setting (I) at nominal level 0.05. We observe that the size of the proposed test is reasonably close to the nominal level. The power increases as n increases, as δ increases, and as T increases. Table 1 also provides the empirical size and power of the proposed test under simulation setting (II). The results in setting (II) are very similar to those in setting (I).

When the null hypothesis is false under settings (I) and (II), the percentages of correct identification are summarized in Table 2. The percentage of correct identification is the percentage of simulation replications that estimate the location of the changepoint correctly among all those that reject the null hypothesis. When $T = 5$ the true changepoint is $k_1 = 2$, and when $T = 8$ the true changepoint is $k_1 = 4$. In both settings, for almost all the cases, the percentages increase as n and δ increase.

To assess the performance of the proposed binary segmentation procedure for identifying multiple changepoints, we generated data using simulation set-up (II) with two changepoints, k_1 and k_2. When $T = 5$, $k_1 = 2$ and $k_2 = 4$. When $T = 8$, $k_1 = 4$ and $k_2 = 6$. For $t \in \{k_{j-1} + 1, \ldots, k_j\}$, we set $A_{t,h} = A^{(j)}$ for $h \in \{0, \ldots, L\}$ and $j = 1, 2, 3$ with $k_0 = 0$ and $k_3 = T$. Here, $A^{(1)}$ and $A^{(2)}$ were taken to be the same as in setting (II), and we set $A^{(3)} = A^{(1)}$. The values of δ were chosen to be 0.15 and 0.25. The average true positives and average true negatives are summarized in Table 3. The true positives are the correctly identified changepoints, and the true negatives are the correctly identified time-points at which no covariance change exists. For $T = 5$, the maximum number of true positives and true negatives for each is 2. For $T = 8$, the maximum numbers of true positives and true negatives are 2 and 5, respectively.

Table 3. *Average true positives and average true negatives for identifying multiple changepoints using the proposed binary segmentation method; standard errors are included after each number. For $T = 5$, the maximum number of true positives and true negatives for each is 2; for $T = 8$, the maximum numbers of true positives and true negatives are 2 and 5, respectively*

			$\delta = 0.15$				$\delta = 0.25$			
T	p	n	ATP	SE	ATN	SE	ATP	SE	ATN	SE
		40	1.10	0.36	1.90	0.30	1.81	0.39	1.92	0.27
	500	50	1.36	0.48	1.87	0.37	1.94	0.24	1.98	0.14
		60	1.57	0.50	1.92	0.27	2.00	0.00	1.92	0.28
		40	1.11	0.37	1.82	0.41	1.76	0.43	1.94	0.24
5	750	50	1.38	0.49	1.92	0.27	2.00	0.00	1.96	0.24
		60	1.47	0.50	1.90	0.30	2.00	0.00	1.98	0.14
		40	1.15	0.36	1.90	0.30	1.87	0.34	1.95	0.22
	1000	50	1.22	0.42	1.96	0.20	1.96	0.20	1.96	0.20
		60	1.54	0.50	1.96	0.20	2.00	0.00	1.98	0.14
		40	1.40	0.49	4.84	0.40	1.91	0.29	4.90	0.30
	500	50	1.62	0.49	4.85	0.36	1.97	0.17	4.92	0.27
		60	1.78	0.42	4.89	0.32	2.00	0.00	4.95	0.22
		40	1.52	0.50	4.82	0.39	1.90	0.30	4.85	0.36
8	750	50	1.67	0.47	4.83	0.38	1.97	0.17	4.94	0.24
		60	1.81	0.40	4.90	0.34	2.00	0.00	4.90	0.30
		40	1.43	0.50	4.82	0.44	1.88	0.33	4.92	0.27
	1000	50	1.68	0.47	4.80	0.40	1.99	0.10	4.96	0.25
		60	1.84	0.37	4.92	0.27	2.00	0.00	4.94	0.24

ATP, average true positives; ATN, average true negatives; SE, standard error.

The results in Table 3 show that the proposed binary segmentation procedure performs well as the sample size, n, increases and as the signal, δ, increases.

6. AN EMPIRICAL STUDY

In this section, we apply our proposed method to a time-course gene expression dataset collected by Taylor et al. (2007). The goal was to identify gene sets with significant changes in covariances over time and estimate their respective changepoints, should any exist. The data come from a study where peripheral blood mononuclear cells were collected from 69 patients with the hepatitis C virus. The cells were collected once before treatment, day 0, and five times during treatment, on days 1, 2, 7, 14 and 28. The treatment consisted of pegylated alpha interferon and ribavirin. More information about the experiment can be found in Taylor et al. (2007).

Prior to the application of our method, the data were pre-processed. The gene expressions with low quality measurements were removed if the corresponding Microarray Suite 5.0 signal transcript was classified as absent. We kept only individuals with gene expression arrays at all six time-points. After pre-processing, our dataset consisted of 46 individuals with gene expression arrays at days 0, 1, 2, 7, 14 and 28. The original dataset is available at `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7123`.

The genes were grouped into gene sets that were defined by gene ontology, which classifies genes according to their attributes in three biological domains: molecular function, biological process, and cellular component (Ashburner et al., 2000). For instance, the gene ontology term labelled 0006468 is related to introducing a phosphate group onto a protein; hence, this gene

Table 4. *Significant gene ontology terms, test statistic values, number of genes in each gene ontology term, identified changepoints, and estimated local false discovery rates*

GO	Number of genes	Test statistic value	Changepoints	Local FDR
0006511	132	11.10	4, 5	0.012
0030054	136	9.92	1, 4, 5	0.044
0042493	128	9.54	5	0.064
0008219	122	9.34	4, 5	0.076
0006357	167	9.13	1, 4	0.090
0005765	116	8.93	4	0.103
0019904	117	8.87	4, 5	0.106
0008285	148	8.75	1, 2, 5	0.115
0048471	263	8.04	1, 4, 5	0.168
0005739	661	8.04	4, 5	0.168

GO, gene ontology; Local FDR, local false discovery rate.

ontology term would consist of all the genes that have a role in that biological process. A given gene can be a member of multiple gene ontologies. For example, in our processed dataset, gene ontology 0006468 consists of 221 genes and gene ontology 0007155 consists of 134 genes, with 64 genes in common. After filtering the dataset according to the procedure above, 159 gene ontology terms were analysed. We applied our method to gene ontology terms with a minimum of 100 genes. Figure 1 displays the number of genes in the 159 gene ontology terms. Each gene set analysed had a gene count much larger than the sample size of 46 patients.

Let $Y_{it}^{(g)}$ $(i = 1, \ldots, 46, t = 1, \ldots, 6)$ be the gene expression data for the gth gene ontology term of the ith individual at time t, where $t = 1$ represents day 0, before treatment, and $t = 2, 3, 4, 5, 6$ represent the times during the treatment of hepatitis C virus with pegylated alpha interferon and ribavirin. Assume model (3) for each gene ontology term, $Y_{it}^{(g)} = \mu_t^{(g)} + \varepsilon_{it}^{(g)}$ for $g = 1, \ldots, 159$, where $\mu_t^{(g)}$ is an unknown mean vector and $\mathrm{var}(\varepsilon_{it}^{(g)}) = \Sigma_t^{(g)}$. The assumptions on $\varepsilon_{it}^{(g)}$ in model (3) incorporate temporal dependence so that $\{\varepsilon_{it}^{(g)}\}_{t=1}^{T}$ are dependent over time. For each gene ontology term, we tested whether the covariance matrices $\Sigma_t^{(g)}$ are the same across all t. In addition, the changepoints were identified for those gene ontology terms found to be significant.

For the gth gene ontology term, we computed $\hat{D}_{nt}^{(g)}/\hat{\sigma}_{(g),nt,0}$ for $t = 1, \ldots, 5$ and the covariance matrix estimator $\hat{V}_{n,D}^{(g)}$. Let $\tilde{\mathcal{M}}_n^{(g)}$ be the maximum of the standardized test statistics $\{\hat{V}_{n,D}^{(g)}\}^{-1/2}\{\hat{\sigma}_{(g),n1,0}^{-1}\hat{D}_{n1}^{(g)}, \ldots, \hat{\sigma}_{(g),n5,0}^{-1}\hat{D}_{n5}^{(g)}\}^{\mathrm{T}}$. For each gene ontology term, the local false discovery rate was estimated using $\{\tilde{\mathcal{M}}_n^{(g)}\}_{g=1}^{159}$ based on the method proposed by Efron (2007). As suggested in Efron (2007), a cut-off value of 0.20 was used for the local false discovery rate procedure. There were 10 gene ontology terms that had a local false discovery rate less than or equal to 0.20. These 10 significant gene ontology terms and their corresponding number of genes, test statistic value, estimated changepoints, and local false discovery rate are given in Table 4. Among those gene ontology terms listed in Table 4, term 0008285 is associated with the reduction or stoppage of cell proliferation. This is of interest, as Kannan et al. (2011) had noted that the hepatitis C virus reduces cell proliferation. Thus, the results here suggest that treatment using pegylated alpha interferon and ribavirin has some effect on the covariances of those genes that play a role in cellular proliferation.

After identifying ten significant gene ontology terms, we applied binary segmentation to identify all changepoints. We discovered that eight terms have a changepoint at $t = 5$, day 14, eight have a changepoint at $t = 4$, day 7, and four terms have a changepoint at $t = 1$, day 0. Recall that

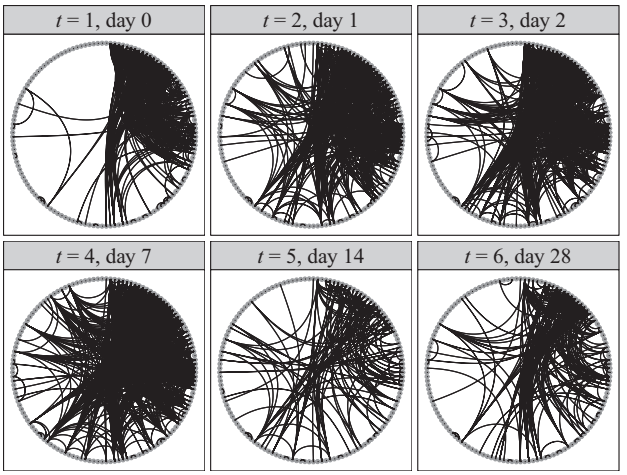

Fig. 2. Correlation network map for gene ontology term 0030054. Each dot represents a gene within the gene ontology. A link between dots indicates an interaction between genes.

a changepoint at time $t = 5$ implies that the covariance matrix at time $t = 5$ is not equal to that at time $t = 6$. Hence, most of the identified changes in the covariance matrices occurred by the initial day of treatment or later in the treatment cycle. These findings complement those of Taylor et al. (2007), who observed that for the majority of the genes that were altered in expression, the changes occurred in the early days of treatment and again, marginally, between treatment days 7 and 28. To illustrate the changes in the covariance matrices, Fig. 2 displays the correlation networks of gene ontology term 0030054 at the six time-points. We see that the correlation networks change at time-points 1, 4 and 5, which is consistent with the identified changepoints reported in Table 4.

ACKNOWLEDGEMENT

We are grateful to the editor, associate editor and two referees for their insightful, constructive and careful comments that have significantly improved the paper. We thank Ms Amanda Applegate for her help with the English language. This research was supported by the U.S. National Science Foundation, National Institute of Drug Abuse and National Institutes of Health, and was partially supported by the National Natural Science Foundation of China.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes proofs of the main results, extensions and simulation results and an R package and its installation instructions.

APPENDIX

Computation of the proposed statistics

To save on computational costs, we can rewrite $U_{s_1s_2,1}$ and $U_{s_1s_2,2}$ defined in (4) in a computationally efficient form as follows. First, we consider $U_{s_1s_2,1}$, which can be rewritten as

$$P_n^3 U_{s_1s_2,1} = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} Y_{is_1}^{\mathrm{T}} Y_{js_2} \right)^2 - \sum_{i,j=1}^{n} (Y_{is_1}^{\mathrm{T}} Y_{js_2})^2 - 2 \sum_{k \neq j=1}^{n} Y_{js_1}^{\mathrm{T}} Y_{js_2} Y_{js_2}^{\mathrm{T}} Y_{ks_1}.$$

Therefore, the computational complexity of $U_{s_1s_2,1}$ with respect to the sample subjects is of the order of n^2, not n^3. To write $U_{s_1s_2,2}$ in a computationally efficient form, we first define

$$V_{s_1s_2,1} = (1/P_n^3)\widetilde{\sum}_{i,j,k} Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{js_1}^{\mathrm{T}} Y_{ks_2}.$$

Similar to $U_{s_1s_2,1}$, we can write $V_{s_1s_2,1}$ as

$$P_n^3 V_{s_1s_2,1} = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} Y_{is_1}^{\mathrm{T}} Y_{js_2} \right) \left(\sum_{i=1}^{n} Y_{is_2}^{\mathrm{T}} Y_{js_1} \right) - \sum_{i,j=1}^{n} Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{js_1}^{\mathrm{T}} Y_{is_2}$$
$$- \sum_{k \neq j=1}^{n} Y_{js_1}^{\mathrm{T}} Y_{js_2} Y_{js_1}^{\mathrm{T}} Y_{ks_2} - \sum_{i \neq j=1}^{n} Y_{is_1}^{\mathrm{T}} Y_{js_2} Y_{js_1}^{\mathrm{T}} Y_{js_2}.$$

The computational complexity of $V_{s_1s_2,1}$ with respect to the sample subjects is also of the order of n^2. Finally, we can write $U_{s_1s_2,2}$ as

$$P_n^4 U_{s_1s_2,2} = \left(\sum_{i \neq j=1}^{n} Y_{is_1}^{\mathrm{T}} Y_{js_2} \right)^2 - P_n^3 (U_{s_1s_2,1} + U_{s_2s_1,1} + 2V_{s_1s_2,1}) - P_n^2 U_{s_1s_2,0}$$
$$- \sum_{i \neq j=1}^{n} (Y_{is_1}^{\mathrm{T}} Y_{js_2})(Y_{is_2}^{\mathrm{T}} Y_{js_1}).$$

Based on the above expression for $P_n^4 U_{s_1s_2,2}$, we can see that the computational complexity of $U_{s_1s_2,2}$ with respect to the sample subjects is of the order of n^2. In summary, the computational cost of the proposed statistic $U_{s_1s_2}$ with respect to the sample subjects is of the order of n^2.

References

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. et al. (2000). Gene ontology: Tool for the unification of biology. *Nature Genet.* **25**, 25–9.

Aue, A., Hormann, S., Horvath, L. & Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *Ann. Statist.* **37**, 4046–87.

Bai, Z & Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6**, 311–29.

Barnett, I & Onnela, J.-P. (2016). Change point detection in correlation networks. *Sci. Rep.* **6**, 18893.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* **36**, 317–46.

Chen, S. X. & Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38**, 808–35.

Chen, S. X., Zhang, L. & Zhong, P.-S. (2010). Testing high dimensional covariance matrices. *J. Am. Statist. Assoc.* **105**, 810–9.

Danaher, P., Paul, D. & Wang, P. (2015). Covariance-based analyses of biological pathways. *Biometrika* **102**, 533–44.

Efron, B. (2007). Size, power and false discovery rates. *Ann. Statist.* **35**, 1351–77.

Fan, J., Liao, Y. & Yao, J. (2015). Power enhancement in high dimensional cross-sectional tests. *Econometrica* **83**, 1497–541.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F. & Hothorn, T. (2018). *mvtnorm: Multivariate Normal and t Distributions.* R package version 1.0-8.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57**, 1–17.

Kannan, R. P., Hensley, L. L., Evers, L. E., Lemon, S. M. & McGivern, D. R. (2011). Hepatitis C virus infection causes cell cycle arrest at the level of initiation of mitosis. *J. Virol.* **85**, 7989–8001.

Koh, W., Pan, W., Gawad, C., Fan, H. C., Kerchner, G. A., Wyss-Coray, T., Blumenfeld, Y. J., El-Sayed, Y. Y. & Quake, S. R. (2014). Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proc. Nat. Acad. Sci.* **111**, 7361–6.

LAUMANN, T. O., SNYDER, A. Z., MITRA, A., GORDON, E. M., GRATTON, C., ADEYEMO, B., GILMORE, A. W., NELSON, S. M., BERG, J. J., GREENE, D. J. et al. (2017). On the stability of BOLD fMRI correlations. *Cereb. Cortex* **27**, 4719–32.

LI, J. & CHEN, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40**, 908–40.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. New York: John Wiley.

PRICHARD, D. & THEILER, J. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Phys. Rev. Lett.* **73**, 951–4.

R DEVELOPMENT CORE TEAM (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.

SCHOTT, J. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample size. *Comp. Statist. Data Anal.* **51**, 6535–42.

SHEDDEN, K. & TAYLOR, J. (2005). Differential correlation detects complex associations between gene expression and clinical outcomes in lung adenocarcinomas. In *Methods of Microarray Data Analysis*, J. S. Shoemaker & S. M. Lin, eds. Boston: Springer, pp. 121–31.

SRIVASTAVA, M. & YANAGIHARA, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *J. Mult. Anal.* **101**, 1319–29.

STOREY, J., XIAO, W., LEEK, J., TOMPKINS, R, & DAVIS, R. (2005). Significance analysis of time course microarray experiments. *Proc. Nat. Acad. Sci.* **102**, 12837–42.

TAI, Y. & SPEED, T. P. (2006). A multivariate empirical Bayes statistic for replicated microarray time course data. *Ann. Statist.* **34**, 2387–412.

TAYLOR, M. W., TSUKAHARA, T., BRODSKY, L., SCHALEY, J., SANDA, C., STEPHENS. M. J., MCCLINTICK, J. N., EDENBERG, H. J., LI, L., TAVIS, J. E. et al. (2007). Changes in gene expression during pegylated interferon and ribavirin therapy of chronic hepatitis C virus distinguish responders from nonresponders to antiviral therapy. *J. Virol.* **81**, 3391–401.

YANG, Q. & PAN, G. (2017). Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices. *J. Am. Statist. Assoc.* **112**, 188–200.

ZALESKY, A., FORNITO, A., COCCHI, L., GOLLO, L. L. & BREAKSPEAR, M. (2014). Time-resolved resting-state brain networks. *Proc. Nat. Acad. Sci.* **111**, 10341–6.

ZHANG, J. & BOOS, D. D. (1992). Bootstrap critical values for testing homogeneity of covariance matrices. *J. Am. Statist. Assoc.* **87**, 425–9.

ZHENG, S., BAI, Z. &YAO, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43**, 546–91.

ZHU, L.-X., NG, K. & JING, P. (2002). Resampling methods for homogeneity tests of covariance matrices. *Statist. Sinica* **12**, 769–83.

[*Received on* 19 *August* 2016. *Editorial decision on* 10 *November* 2018]