The Annals of Statistics
2017, Vol. 45, No. 3, 1185–1213
DOI: 10.1214/16-AOS1481

TESTS FOR COVARIANCE STRUCTURES WITH HIGH-DIMENSIONAL REPEATED MEASUREMENTS

BY PING-SHOU ZHONG*,[1] WEI LAN†,[2],
PETER X. K. SONG‡,[3] AND CHIH-LING TSAI§

Michigan State University, Southwestern University of Finance and Economics,†
University of Michigan‡ and University of California-Davis§*

In regression analysis with repeated measurements, such as longitudinal data and panel data, structured covariance matrices characterized by a small number of parameters have been widely used and play an important role in parameter estimation and statistical inference. To assess the adequacy of a specified covariance structure, one often adopts the classical likelihood-ratio test when the dimension of the repeated measurements (p) is smaller than the sample size (n). However, this assessment becomes quite challenging when p is bigger than n, since the classical likelihood-ratio test is no longer applicable. This paper proposes an adjusted goodness-of-fit test to examine a broad range of covariance structures under the scenario of "large p, small n." Analytical examples are presented to illustrate the effectiveness of the adjustment. In addition, large sample properties of the proposed test are established. Moreover, simulation studies and a real data example are provided to demonstrate the finite sample performance and the practical utility of the test.

1. Introduction. In the broad sense of repeated measures such as panel data and longitudinal data, the covariance matrix of repeated measurements plays an important role for statistical inference [see Davis (2002), Diggle et al. (2002) and Frees (2004)]. Since technological advances have led to increasingly high-dimensional data sets in various fields such as biological science, engineering, medicine and social science, among others, there is a practical need to model the dependence among repeated measurements to improve the estimation efficiency and prediction accuracy. In general, there are two types of covariance matrix to consider; one is structured covariance and the other is unstructured covariance (i.e., no assumption, except for symmetry, is made on the pattern of covariance). To estimate a high-dimensional unstructured covariance matrix, one can employ the shrinking, factoring, banding, tapering or thresholding approach to obtain a

Received May 2015; revised March 2016.
[1]Supported by NSF Grant DMS-13-09156.
[2]Supported in part by National Natural Science Foundation of China Grants 11401482 and 71532001.
[3]Supported in part by NIH Grant R01 ES024732 and NSF Grant DMS-15-13595.
MSC2010 subject classifications. Primary 62H15; secondary 62G20, 62G10.
Key words and phrases. Adjusted test, goodness-of-fit test, longitudinal data, panel data.

desirable sparse estimator. In contrast to the unstructured setting, researchers have considered several structured covariance matrices, such as autoregressive (AR), moving average (MA) and compound symmetry (CS) [e.g., see Zhao et al. (2007), Pourahmadi (2013), and Wiesel, Bibi and Globerson (2013)]. The number of unknown parameters in a structured covariance matrix can be significantly reduced so that computation is much easier than using regularization methods; this is particularly true for data with a small sample size and a large number of variables. However, an incorrectly specified covariance structure could lead to inaccurate predictions and misleading inferences. This motivates us to develop a test to assess the appropriateness of a structured covariance specification.

In this paper, our focus is on the development of a testing procedure for a variety of covariance structures in the context of repeated measurements with high-dimensional data. Let $(\mathbf{Y}_i, \mathbf{X}_i)$, $i = 1, \ldots, n$, be independent and identically distributed (IID) samples, where $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ is the response vector of the p-dimensional repeated measurements and $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{id})$ is the $p \times d$ matrix of predictors collected from the ith sample. We then consider the following repeated-measures model:

$$\mathbf{Y}_i = \boldsymbol{\mu} + \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

where $\boldsymbol{\mu}$ is a p-dimensional intercept, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top$ is a d-dimensional ($d < \infty$) vector of unknown regression coefficients and the errors $\boldsymbol{\varepsilon}_i$ are IID normally distributed random vectors with mean $E(\boldsymbol{\varepsilon}_i) = 0$ and covariance $\mathrm{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma} = \sigma^2 \mathbf{R}$, where $\mathbf{R}$ is the $p \times p$ correlation matrix and σ^2 is the scale parameter. Note that high-dimensionality in this paper refers to the dimension of repeated measurements and not to the dimension of $\boldsymbol{\beta}$. Furthermore, $\mathbf{X}_i$ and $\boldsymbol{\varepsilon}_i$ are assumed to be independent. To examine the covariance structure of $\boldsymbol{\Sigma}$, we test the following hypotheses:

$$H_0 : \boldsymbol{\Sigma} \in \mathscr{C} \quad \text{vs.} \quad H_a : \boldsymbol{\Sigma} \notin \mathscr{C}, \tag{1.2}$$

where $\mathscr{C} = \{\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2 \mathbf{R}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta, \sigma^2 > 0\}$ is a family of covariance matrices parameterized by the parameters σ^2 and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top \in \Theta \subset \mathbb{R}^q$ for $q < \infty$. When $q = 1$, we denote $\theta = \theta_1$. For example, the AR(1) structure, $\boldsymbol{\Sigma}(\theta) = \sigma^2 \mathbf{R}(\theta) = \sigma^2 (\rho^{|i-j|})_{i,j=1}^p$, is parameterized by σ^2 and the autocorrelation coefficient $\theta = \rho$. With a slight abuse of notation, the sphericity covariance structure is denoted as $\boldsymbol{\Sigma}(\theta) = \theta \mathbf{I}_p$ throughout the paper where $\mathbf{I}_p$ is the $p \times p$ identity matrix. In addition, the parametric structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is parameterized in a meaningful way so that $\boldsymbol{\Sigma}(\boldsymbol{\theta}_1) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_2)$ if and only if $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, that is, $\boldsymbol{\theta}$ is identifiable. The null hypothesis $\boldsymbol{\Sigma} \in \mathscr{C}$ represents that there exists some unique true $\boldsymbol{\theta}_0 \in \Theta$ for which $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$.

When the dimension p is less than the sample size n and the likelihood is available, the covariance structure in model (1.1) can be assessed via the likelihood ratio test by comparing the likelihood of the parameterized covariance with that of the unstructured covariance. Unfortunately, the likelihood ratio test is invalid when the dimension p is larger than the sample size n or $p/n \to 1$ [see, e.g., Cui, Zheng and Li (2013)]. To overcome this problem, several methods have been proposed in the nonregression setting, and they can mainly be classified into two types of methods. One is the so-called adjusted likelihood ratio test based on the large dimensional random matrix theory [see, e.g., Bai et al. (2009) and Li and Qin (2014)], while the other type is built upon a consistent estimator of the distance (in Frobenius or maximum norm) between the unstructured and parameterized covariances. Using the second type of method, Ledoit and Wolf (2002) modified two tests proposed by John (1971, 1972) to allow the data dimension p to increase in a polynomial order of the sample size n for normally distributed data. Later, Chen, Zhang and Zhong (2010) and Zou et al. (2014) proposed more robust methods that preclude the normality assumption and allow the data dimension p to be much larger than the sample size n. However, the extant methods mainly focused on testing relatively simple covariance structures, such as sphericity, under nonregression settings.

Despite this encouraging progress, it remains unclear how to test general covariance structures for high-dimensional data, since existing literature considers different tests for different covariance structures. After studying this issue thoroughly, we have found that the major challenge pertains to the effect induced by the estimation of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. It is worth noting that the estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ will usually not affect the asymptotic distribution of the covariance test in the fixed dimensional case. However, a great challenge arises in high-dimensional data due to the accumulation of errors from estimating parameters, which can impair the classical approach for obtaining the large sample properties of the covariance test. For example, Baltagi, Kao and Peng (2015) found that the estimation of fixed effect parameters results in bias for testing sphericity proposed by John (1971). Recently, Zou et al. (2014) showed that the estimation of the location parameter can affect the sign-based test for sphericity. Because they mainly focused on testing sphericity, the impact from estimating $\boldsymbol{\theta}$ on testing the general covariance structures has not been well studied yet. To address this major challenge, we develop a unified approach to analyze and accommodate this impact on test statistics. Accordingly, we find that, depending on the type of covariance structure, the estimator of the variance component $\boldsymbol{\theta}$ can result in a significant leading order effect on the asymptotic distribution of test statistics, whereas the estimation error of $\hat{\boldsymbol{\beta}}$ does not have such a leading order effect, under some mild assumptions.

The aim of this paper is to propose an adjusted goodness-of-fit test (namely the adjusted test hereafter) for assessing general covariance structures with high-dimensional repeated measurements. The proposed method analytically mitigates the detrimental influence of the plug-in estimator of the variance component $\boldsymbol{\theta}$ on the large sample properties of the test statistic. More importantly, the proposed method relaxes some restrictive assumptions on the underlying covariance structures that have previously limited the scope of application. For example, our method does not assume that $\mathrm{tr}(\mathbf{\Sigma}^4) = o\{\mathrm{tr}^2(\mathbf{\Sigma}^2)\}$ or $\mathbf{\Sigma}$ is sparse, which are extensively used in the existing literature [see, e.g., Chen, Zhang and Zhong (2010); Li and Chen (2012); Cai, Liu and Xia (2013)]. Although these assumptions appear reasonable in many applications such as the sphericity test, they may not be satisfied by some practically important covariance structures such as compound symmetry, that is, $\mathbf{\Sigma} = \mathbf{I}_p + \theta(\mathbf{1}\mathbf{1}^\top - \mathbf{I}_p)$ with $\mathbf{1} = (1, \dots, 1)^\top \in \mathbb{R}^p$ being a p-element vector of ones. Consequently, our method relaxes these assumptions and can be applied to a wide range of covariance structures, including sphericity, autoregression, moving average, compound symmetry and so on. Furthermore, the proposed method generalizes the high-dimensional covariance testing procedure to the regression model setting, which is of great importance in many practical applications.

The rest of the paper is organized as follows. Section 2 introduces two tests, the goodness-of-fit test $\hat{T}_n$ and an adjusted test, with two numerical examples. Then the main theoretical properties related to the two tests are presented in Section 3. Extensive simulation studies are reported in Section 4 and a real data example is illustrated in Section 5. The article concludes with a short discussion. All the technical details and additional simulation results are relegated to an associated supplemental material [Zhong et al. (2016)].

2. Test statistics. Let $\delta(\boldsymbol{\theta}) = \mathrm{tr}\{(\mathbf{\Sigma} - \mathbf{\Sigma}(\boldsymbol{\theta}))^2\}$ be the Frobenius distance between $\mathbf{\Sigma}$ and $\mathbf{\Sigma}(\boldsymbol{\theta})$ for some $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}_0$ be the minimizer of $\delta(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta$. In other words, $\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}\in\Theta} \delta(\boldsymbol{\theta})$ where Θ is the region of $\boldsymbol{\theta}$ such that $\mathbf{\Sigma}(\boldsymbol{\theta}) > 0$. Under the null hypothesis of (1.2), $\mathbf{\Sigma}$ belongs to the family of $\mathscr{C}$. Thus, $\boldsymbol{\theta}_0$ is the unique true value of $\boldsymbol{\theta}$ such that $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta}_0)$. In contrast, under the alternative hypothesis, $\mathbf{\Sigma}$ does not belong to the family of $\mathbf{\Sigma}(\boldsymbol{\theta})$. In this case, $\boldsymbol{\theta}_0$ is the value of $\boldsymbol{\theta} \in \Theta$ that minimizes the Frobenius distance between $\mathbf{\Sigma}$ and the parametric family $\mathscr{C}$. Accordingly, the hypotheses (1.2) are equivalent to

$$H_0 : \delta(\boldsymbol{\theta}_0) = 0 \quad \text{vs.} \quad H_a : \delta(\boldsymbol{\theta}_0) > 0. \tag{2.1}$$

For the ease of presentation, we assume $\boldsymbol{\mu} = 0$, $E(\mathbf{X}_i) = 0$ and $\mathrm{Var}(\mathbf{X}_{ij}) = \mathbf{\Sigma}_{X_j}$ for $j = 1, \dots, d$ in model (1.1). The general case with $\boldsymbol{\mu} \neq 0$ and $E(\mathbf{X}_i) \neq 0$ will be discussed at the end of the paper. When the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_0$ are known, we follow the spirit of Chen, Zhang and Zhong (2010) to obtain an unbiased estimator of $\delta(\boldsymbol{\theta}_0)$, which is

$$T_n(\boldsymbol{\theta}_0, \boldsymbol{\beta}) = \frac{1}{C_n^2} \sum_{i<j} \{\boldsymbol{\varepsilon}_i^\top(\boldsymbol{\beta}) \boldsymbol{\varepsilon}_j(\boldsymbol{\beta})\}^2 - \frac{2}{n} \sum_{i=1}^{n} \boldsymbol{\varepsilon}_i^\top(\boldsymbol{\beta}) \mathbf{\Sigma}(\boldsymbol{\theta}_0) \boldsymbol{\varepsilon}_i(\boldsymbol{\beta}) + \mathrm{tr}\{\mathbf{\Sigma}(\boldsymbol{\theta}_0)^2\},$$

where $\boldsymbol{\varepsilon}_i(\boldsymbol{\beta}) = \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}$ and $C_n^k = n!/\{(n-k)!k!\}$ $(k \leq n)$ is the binomial coefficient. We denote $T_n(\boldsymbol{\theta}_0, \boldsymbol{\beta})$ as T_n. In practice, however, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_0$ are often unknown. We replace them by their corresponding consistent estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}_0$ defined below, and this yields an estimator of $\delta(\boldsymbol{\theta}_0)$, which we name $\hat{T}_n$. For the sake of convenience, we denote $\boldsymbol{\varepsilon}_i(\boldsymbol{\beta})$ and $\boldsymbol{\varepsilon}_i(\hat{\boldsymbol{\beta}})$ as $\boldsymbol{\varepsilon}_i$ and $\hat{\boldsymbol{\varepsilon}}_i$, respectively, in the rest of paper. The test statistic T_n can also be used to test the correlation matrix $\mathbf{R}$ by replacing its $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ with $\sigma^{-1}\boldsymbol{\varepsilon}_i$ and $\mathbf{R}(\boldsymbol{\theta}_0)$, respectively. These replacements lead to a test statistic that is a scalar transformation of T_n (i.e., T_n/σ^4). It is worth noting that our proposed tests are based on the standardized test statistics. Accordingly, the standardized versions of T_n and T_n/σ^4 are exactly the same, and do not depend on σ^2. Consequently, T_n is applicable for testing correlation structures. Similar explanations are applicable to other test statistics presented later. Hereafter, without loss of generality, assume that σ^2 is known.

To obtain an unbiased and consistent estimator of $\boldsymbol{\beta}$, we adopt the ordinary least squares method, which yields $\hat{\boldsymbol{\beta}} = \{\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{X}_i\}^{-1} \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{Y}_i$. For a given $\hat{\boldsymbol{\beta}}$, $\boldsymbol{\theta}_0$ is subsequently estimated by minimizing the following least squares distance between $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ [Browne (1974) and Bentler and Bonett (1980)]:

$$\hat{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_0(\hat{\boldsymbol{\beta}}) = \arg\min_{\boldsymbol{\theta}\in\Theta} \operatorname{tr}\{(\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))^2\}, \tag{2.2}$$

where $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} = n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^\top$. Then $\hat{\boldsymbol{\theta}}_0$ is a consistent estimator of $\boldsymbol{\theta}_0$, and a goodness-of-fit test for testing (1.2) takes the form $\hat{T}_n := T_n(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\beta}})$. However, as we will illustrate below and again in Section 3.1, the asymptotic distribution of $\hat{T}_n$ can be substantially affected by the underlying structure of $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, which causes a deficiency in this test.

We first define some notation before introducing the adjusted test statistic. Let θ_{0j} be the jth ($j = 1, \dots, q$) component of the q-dimensional vector $\boldsymbol{\theta}_0$. For $u = 0, 1, 2$, define $\mathbf{B}_{j,u} = (\partial\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)/\partial\theta_{0j})\boldsymbol{\Sigma}^u$ and let $\mathbf{V}_u = (v_{ij})$ be a $q \times q$ matrix whose (i, j)th component is $v_{ij} = \operatorname{tr}(\mathbf{B}_{i,u}\mathbf{B}_{j,u})$. In addition, let $\boldsymbol{\Omega}_0 = \mathbf{V}_0^{-1} = (w_{ij}(\boldsymbol{\theta}_0))$ be the inverse of the matrix $\mathbf{V}_0$, which can be estimated consistently. Note that q is the dimension of the parameters in $\boldsymbol{\theta}_0$, and hence $\mathbf{V}_0$ is a fixed dimensional matrix. Let $e_{i,k} = \boldsymbol{\varepsilon}_i^\top \mathbf{B}_{k,0}\boldsymbol{\varepsilon}_i$ and $\mathbf{e}_i = (e_{i,1}, \dots, e_{i,q})^\top$ be a q-dim vector. Note that $e_{i,k}$ is a random variable associated with the estimating equation for estimating $\boldsymbol{\theta}_0$ in (2.2). Moreover, define $Q_k(\boldsymbol{\varepsilon}_i) = e_{i,k} - \operatorname{tr}(\mathbf{B}_{k,0}\boldsymbol{\Sigma})$ as the centralized $e_{i,k}$ and $\mathbf{Q}(\boldsymbol{\varepsilon}_i) = \{Q_1(\boldsymbol{\varepsilon}_i), \dots, Q_q(\boldsymbol{\varepsilon}_i)\}^\top$. The corresponding estimated version will be denoted with hats. For example, $\hat{e}_{i,k} = \hat{\boldsymbol{\varepsilon}}_i^\top \hat{\mathbf{B}}_{k,0}\hat{\boldsymbol{\varepsilon}}_i$ is the estimated version of $e_{i,k}$, where $\hat{\mathbf{B}}_{k,0} = \partial\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_0)/\partial\theta_{0k}$, and $\hat{w}_{ij} = w_{ij}(\hat{\boldsymbol{\theta}}_0)$ is the estimate of $w_{ij} = w_{ij}(\boldsymbol{\theta}_0)$. We then propose the following adjusted test statistic:

$$\Lambda_n = \hat{T}_n - \hat{J}_{n3}, \tag{2.3}$$

where $\hat{J}_{n3} = n^{-1}\{(C_n^2)^{-1}\sum_{i<j}^n \hat{\mathbf{e}}_i^\top \hat{\boldsymbol{\Omega}}_0 \hat{\mathbf{e}}_j - n^{-1}\sum_{i=1}^n \hat{\mathbf{e}}_i^\top \hat{\boldsymbol{\Omega}}_0 \hat{\mathbf{e}}_i\}$. This adjustment term is associated with the bias due to the Taylor series expansion of $\hat{T}_n$ at $\boldsymbol{\theta}_0$, and more specific reasons for the adjustment are given in the rest of this section and Section 3.1.

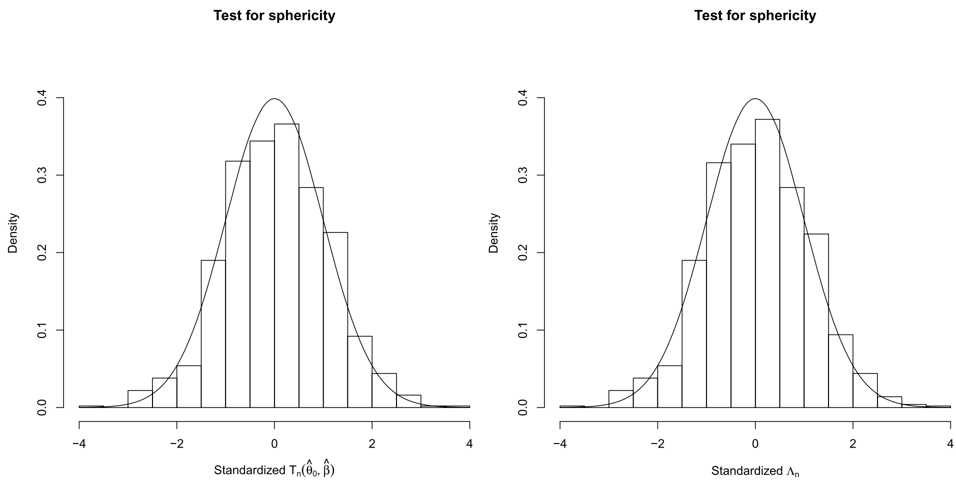

FIG. 1. *Histograms of the standardized* $\hat{T}_n$ (*left panel*) *and the standardized* Λ_n (*right panel*) *for testing the sphericity under* H_0. *The curves in both panels are the density of the standard normal distribution*.

Two examples are given below to illustrate the difference between the test statistic $\hat{T}_n$ and its adjusted version Λ_n. For the sake of simplicity, we sometimes use $\hat{T}_n$ and Λ_n to represent the standardized version of these tests, such as in numerical studies.

EXAMPLE 2.1 (Sphericity). Consider testing $H_0 : \mathbf{\Sigma} \in \mathscr{C}_{\mathrm{SP}}$ vs. $H_a : \mathbf{\Sigma} \notin \mathscr{C}_{\mathrm{SP}}$ where $\mathscr{C}_{\mathrm{SP}}$ is the sphericity family defined by $\mathscr{C}_{\mathrm{SP}} = \{\theta > 0 : \mathbf{\Sigma}(\theta) = \theta \mathbf{I}_p\}$. In this case, the adjustment term is given by

$$\hat{J}_{n3} = (np)^{-1}\left\{\frac{1}{C_n^2}\sum_{i<j}^{n} \hat{\boldsymbol{\varepsilon}}_i^\top \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_j^\top \hat{\boldsymbol{\varepsilon}}_j - \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{\varepsilon}}_i^\top \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i^\top \hat{\boldsymbol{\varepsilon}}_i\right\}.$$

Figure 1 displays the histograms of the standardized test statistics $\hat{T}_n$ and Λ_n in 1000 realizations under the null hypothesis H_0 by using the simulation setting in Example 4.1 of Section 4 with $p = 360$ and $n = 60$. It is worth noting that both tests are standardized by the mean and standard error of T_n under H_0. The result shows that $\hat{T}_n$ and Λ_n are very similar, and the adjustment term $\hat{J}_{n3}$ has little effect on $\hat{T}_n$. In addition, the standard normal curve (the solid curve) matches both histograms well.

EXAMPLE 2.2 (Compound symmetry). Consider testing $H_0 : \mathbf{\Sigma} \in \mathscr{C}_{\mathrm{CS}}$ versus $H_a : \mathbf{\Sigma} \notin \mathscr{C}_{\mathrm{CS}}$ where $\mathscr{C}_{\mathrm{CS}}$ is the collection of compound symmetry matrices defined by $\mathscr{C}_{\mathrm{CS}} = \{\mathbf{\Sigma}(\theta) : \mathbf{\Sigma}(\theta) = \mathbf{I}_p + \theta(\mathbf{1}\mathbf{1}^\top - \mathbf{I}_p)$ with $\theta \in (0, 1)\}$. After some algebraic

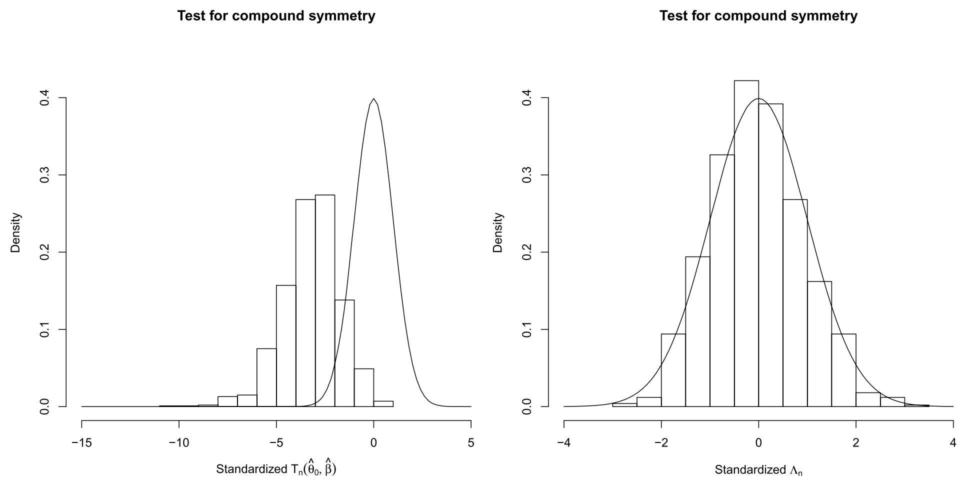

FIG. 2. *Histograms of the standardized* $\hat{T}_n$ (*left panel*) *and the standardized* Λ_n (*right panel*) *for testing the compound symmetry under* H_0. *The curves in both panels are the density of the standard normal distribution.*

simplification, we obtain that

$$\hat{J}_{n3} = \{np(p-1)\}^{-1}\left\{\frac{1}{C_n^2}\sum_{i<j}^{n}\hat{e}_{i,1}\hat{e}_{j,1} - \frac{1}{n}\sum_{i=1}^{n}\hat{e}_{i,1}^2\right\},$$

where $\hat{e}_{i,1} = \hat{\boldsymbol{\varepsilon}}_i^\top(\mathbf{1}\mathbf{1}^\top - \mathbf{I}_p)\hat{\boldsymbol{\varepsilon}}_i$. Figure 2 depicts the histograms of the standardized test statistics $\hat{T}_n$ and Λ_n in 1000 realizations under the null hypothesis H_0 using the simulation setting in Example 4.4 of Section 4 with $p = 360$ and $n = 60$. Clearly, their distributions look quite different, and from these figures we obtain three important observations. (i) The standardized $\hat{T}_n$ has a larger dispersion than the standard normal. This suggests that the variance of T_n underestimates the variance of $\hat{T}_n$. (ii) The center of the standardized $\hat{T}_n$ shifted from 0 to a negative number. This implies that the asymptotic mean of $\hat{T}_n$ is not 0 under the null hypothesis. (iii) The standard normal curve matches the histogram of Λ_n very well. These observations imply that the adjustment term $\hat{J}_{n3}$ plays a critical role in making $\hat{T}_n$ perform properly, which cannot be ignored. A detailed theoretical explanation for the use of this adjustment $\hat{J}_{n3}$ is given in Section 3.1.

3. Main results. In this section, we present theoretical properties of the test statistics, T_n, $\hat{T}_n$ and Λ_n, introduced in Section 2.

3.1. *Moments of* T_n *and* $\hat{T}_n$. Suppose that $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_0$ are known a priori. The following theorem presents the mean and variance of the test statistic T_n, which is the estimator of $\delta(\boldsymbol{\theta}_0)$.

THEOREM 1. *The mean and variance of T_n are, respectively,* $\mathrm{E}(T_n) = \delta(\boldsymbol{\theta}_0)$ *and*

$$\mathrm{Var}(T_n) = 4(n-2)(C_n^2)^{-1}\,\mathrm{tr}\big[\{\boldsymbol{\Sigma}(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))\}^2\big] + 2(C_n^2)^{-1}\big(\mathrm{tr}^2(\boldsymbol{\Sigma}^2) + \mathrm{tr}(\boldsymbol{\Sigma}^4) + 2\,\mathrm{tr}\big[\{\boldsymbol{\Sigma}(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))\}^2\big]\big).$$

In practice, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_0$ in T_n are often unknown. Hence, we need to employ the estimator $\hat{T}_n$ to assess the goodness-of-fit. To understand the asymptotic behavior of $\hat{T}_n$, we present the following two theorems in this subsection. Throughout the paper, we assume $p \to \infty$ as $n \to \infty$ for asymptotic analysis. We do not require explicit conditions on the data dimension p and sample size n. As long as the conditions and assumptions given in the paper are satisfied, the proposed method can be applied regardless of the relationship between p and n.

Let $r_n = \max_{1\le j\le d} \mathrm{tr}(\boldsymbol{\Sigma}_{X_j}^2)/\mathrm{tr}^2(\boldsymbol{\Sigma}_{X_j})$, $a_n = n^{-3/2}\,\mathrm{tr}(\boldsymbol{\Sigma}^2) r_n^{1/2}$ and $b_n = n^{-1/2} \times \mathrm{tr}^{1/2}[\{\boldsymbol{\Sigma}(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))\}^2]$. Then the following theorem provides an asymptotic connection between $\hat{T}_n = T_n(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\beta}})$ and $T_n = T_n(\boldsymbol{\theta}_0, \boldsymbol{\beta})$.

THEOREM 2. *Under conditions* (C1) *in Appendix*, *if $r_n \to 0$, then we have*

$$\hat{T}_n = \{T_n(\boldsymbol{\theta}_0, \boldsymbol{\beta}) - \mathbf{u}^\top(\boldsymbol{\theta}_0)\boldsymbol{\Omega}_0\mathbf{u}(\boldsymbol{\theta}_0)\}\{1 + o_p(1)\} + O_p(a_n) + o_p(b_n), \tag{3.1}$$

where $\mathbf{u}(\boldsymbol{\theta}_0) = (\mathrm{tr}[\mathbf{B}_{1,0}\{\mathbf{S}_n - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\}], \ldots, \mathrm{tr}[\mathbf{B}_{q,0}\{\mathbf{S}_n - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\}])^\top \in \mathbb{R}^q$ *and* $\mathbf{S}_n = n^{-1}\sum_{i=1}^n \boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top$.

REMARK 1. Theorem 2 indicates that the asymptotic behavior of $\hat{T}_n$ is not only determined by T_n but also affected by the other three terms on the right-hand side of (3.1). The second term is induced by the estimator $\hat{\boldsymbol{\theta}}_0$, while the third term $O_p(a_n)$ and the fourth term $o_p(b_n)$ are due to the estimation of $\boldsymbol{\beta}$. By Theorem 1, b_n^2 is no greater than $\mathrm{Var}(T_n)$. Hence, the order of the fourth term in (3.1) is smaller than that of T_n. In fact, $b_n \equiv 0$ under H_0. In addition, the third term in (3.1) can be ignored if the order of a_n is smaller than those of the first two terms in (3.1). This is a mild condition, which will be discussed before Theorem 5. Consequently, we ignore the last two terms in (3.1) in the evaluation of the asymptotic behavior of $\hat{T}_n$ and Λ_n.

REMARK 2. Note that the assumption $r_n \to 0$ in Theorem 2 does not hold when p is a fixed number. However, in the case of fixed p, a_n is of order $n^{-3/2}$ and b_n is of order $n^{-1/2}$ under the alternative H_1 with $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. Note that b_n is 0 under H_0. Therefore, the estimation error due to estimating $\boldsymbol{\beta}$ can be ignored when p is fixed.

To explore the asymptotic properties of $\hat{T}_n$, we define

$$J_n(\boldsymbol{\theta}_0, \boldsymbol{\beta}) := T_n(\boldsymbol{\theta}_0, \boldsymbol{\beta}) - \mathbf{u}^\top(\boldsymbol{\theta}_0)\boldsymbol{\Omega}_0\mathbf{u}(\boldsymbol{\theta}_0) - \delta(\boldsymbol{\theta}_0), \tag{3.2}$$

where $\delta(\boldsymbol{\theta}_0) = \mathrm{tr}\{(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))^2\}$, as previously defined in Section 2. Based on Remark 1, we have $J_n(\boldsymbol{\theta}_0, \boldsymbol{\beta}) = \hat{T}_n - \delta(\boldsymbol{\theta}_0)$ asymptotically. As a result, studying the asymptotic behavior of $\hat{T}_n$ is equivalent to studying that of $J_n(\boldsymbol{\theta}_0, \boldsymbol{\beta})$. For convenience, hereafter, we sometimes denote $J_n(\boldsymbol{\theta}_0, \boldsymbol{\beta})$ by J_n. After reformulation, J_n can be decomposed as a sum of three terms, namely, $J_n = J_{n1} + J_{n2} + J_{n3}$, where

$$J_{n1} = (C_n^2)^{-1} \sum_{i<j}^{n} \{(\boldsymbol{\varepsilon}_i^\top \boldsymbol{\varepsilon}_j)^2 - \boldsymbol{\varepsilon}_i^\top \boldsymbol{\Sigma} \boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_j^\top \boldsymbol{\Sigma} \boldsymbol{\varepsilon}_j + \mathrm{tr}(\boldsymbol{\Sigma}^2) - \mathbf{Q}^\top(\boldsymbol{\varepsilon}_i) \boldsymbol{\Omega}_0 \mathbf{Q}(\boldsymbol{\varepsilon}_j)\},$$

$$J_{n2} = \frac{2}{n} \sum_{i=1}^{n} (\boldsymbol{\varepsilon}_i^\top \{\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\} \boldsymbol{\varepsilon}_i - \mathrm{tr}[\boldsymbol{\Sigma}\{\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\}]) \quad \text{and}$$

$$J_{n3} = n^{-1} \left\{ (C_n^2)^{-1} \sum_{i<j}^{n} \mathbf{e}_i^\top \boldsymbol{\Omega}_0 \mathbf{e}_j - n^{-1} \sum_{i=1}^{n} \mathbf{e}_i^\top \boldsymbol{\Omega}_0 \mathbf{e}_i \right\}.$$

The term J_{n3} has a nonzero mean and is induced from the second term of (3.2). In addition, the terms J_{n1} and J_{n2} come from a recombination of the three terms of (3.2) after removing J_{n3}. It can be checked that J_{n1} is a degenerate U-statistic of order 2, J_{n2} is a U-statistic of order 1, and they are uncorrelated with means equal to zero. These facts allow us to effectively study the asymptotic behavior of the statistic $J_n - J_{n3} = J_{n1} + J_{n2}$, which is associated with the test statistic Λ_n.

To evaluate the mean and variance of J_n, we define the following notation:

$$\xi_1^2 = \mathrm{tr}^2(\boldsymbol{\Sigma}^2) + \mathrm{tr}(\boldsymbol{\Sigma}^4) + 2\,\mathrm{tr}\{(\boldsymbol{\Omega}_0 \mathbf{V}_1)^2\} - 4\,\mathrm{tr}(\boldsymbol{\Omega}_0 \mathbf{V}_2),$$

$$\xi_3^2 = \mathrm{tr}\{(\boldsymbol{\Omega}_0 \mathbf{V}_1)^2\} + 4\,\mathrm{tr}(\mathbf{W}_B^2) + 2 \sum_{k,l,s,t}^{q} w_{kl} w_{st} \mathrm{tr}(\mathbf{B}_{k,1} \mathbf{B}_{t,1} \mathbf{B}_{l,1} \mathbf{B}_{s,1}),$$

$$\xi_2^2 = \mathrm{tr}[\{\boldsymbol{\Sigma}(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))\}^2], \qquad \xi_{13} = \mathrm{tr}(\boldsymbol{\Omega}_0 \mathbf{V}_2) - \mathrm{tr}^2\{(\boldsymbol{\Omega}_0 \mathbf{V}_1)^2\},$$

$$\xi_{23} = \mathrm{tr}\{(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)) \boldsymbol{\Sigma} \mathbf{W}_B\}, \qquad \text{where } \mathbf{W}_B = \sum_{k,l=1}^{q} w_{kl} \mathbf{B}_{k,1} \mathbf{B}_{l,1}.$$

The following theorem presents the mean and variance of J_n.

THEOREM 3. *Under model* (1.1), $\mathrm{E}(J_n) = \mathrm{E}(J_{n3}) = -2\,\mathrm{tr}(\mathbf{W}_B)/n$ *and* $\mathrm{Var}(J_n) = V_{n1} + V_{n2} + V_{n3} + C_{n13} + C_{n23}$, *where* $V_{n1} := \mathrm{Var}(J_{n1}) = 2\xi_1^2/C_n^2$, $V_{n2} := \mathrm{Var}(J_{n2}) = 8\xi_2^2/n$, $V_{n3} := \mathrm{Var}(J_{n3}) = 8\xi_3^2/n^3$, $C_{n13} := 2\,\mathrm{Cov}(J_{n1}, J_{n3}) = 8\xi_{13}/\{nC_n^2\}$ *and* $C_{n23} := 2\,\mathrm{Cov}(J_{n2}, J_{n3}) = -32\xi_{23}/n^2$.

The above theorem shows that the mean of J_n is a negative number. Thus, the test statistic $\hat{T}_n$ has a smaller mean than the adjusted test Λ_n, which explains the location shift in Figure 2. To gain more insight about Theorem 3, we consider the following several analytical examples for five commonly used covariance structures $\mathscr{C}$, which include the sphericity family (SP), $\mathscr{C}_{\mathrm{SP}}$ defined in Example 2.1;

compound symmetry family (CS), $\mathscr{C}_{\mathrm{CS}}$ defined in Example 2.2; polynomial decay family (PN), $\mathscr{C}_{\mathrm{PN}}$, defined in Example 3.4; autoregressive with order 1 [AR(1)] defined by

$$\mathscr{C}_{\mathrm{AR}(1)} = \{\sigma^2\mathbf{R}(\theta) = \sigma^2(R_{ij}(\theta))_{p\times p} : R_{ij}(\theta) = \theta^{|i-j|} \text{ and } \theta \in (0,1)\};$$

and moving average model with order q [MA(q)] defined by

$$\mathscr{C}_{\mathrm{MA}(q)} = \left\{\sigma^2\mathbf{R}(\boldsymbol{\theta}) = \sigma^2(R_{ij}(\boldsymbol{\theta})) : R_{ij}(\boldsymbol{\theta}) = \frac{\gamma(h)}{\gamma(0)} \text{ if } |i-j| = h \le q \text{ and } R_{ij}(\boldsymbol{\theta}) = 0 \text{ if } |i-j| > q\right\},$$

where $\gamma(h) = \sum_{i=0}^{q-h}\theta_i\theta_{i+h}$ with $\theta_0 = 1$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^\top$. In particular, $\mathscr{C}_{\mathrm{MA}(1)}$ is

$$\mathscr{C}_{\mathrm{MA}(1)} = \left\{\sigma^2\mathbf{R}(\theta) = \sigma^2(R_{ij}(\theta)) : R_{ii}(\theta) = 1, R_{ij}(\theta) = \frac{\theta}{1+\theta^2} \text{ if } |i-j| = 1 \text{ and } R_{ij}(\theta) = 0 \text{ if } |i-j| > 1\right\}.$$

Note that, under H_0, $J_{n2} = 0$, and hence $V_{n2} = C_{n23} = 0$. Then by Theorem 3, we have $\mathrm{Var}(J_n) = V_{n1} + V_{n3} + C_{n13}$. Let $a \sim b$ denote that quantities a and b have the same order such that a/b is bounded below and above by some finite positive constants. We next present four examples to illustrate the asymptotic orders of J_n, J_{n1} and J_{n3}.

EXAMPLE 3.1 (Sphericity). Consider $\boldsymbol{\Sigma} \in \mathscr{C}_{\mathrm{SP}}$. Then $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\theta_0) = \theta_0\mathbf{I}_p$. It can be checked that $V_{n1} \sim p^2/n^2$, $V_{n3} \sim 1/n^3$ and $C_{n13} \sim 1/n^3$. Therefore, $V_{n1} \gg V_{n3}$, which leads to

$$\mathrm{Var}(J_n) = V_{n1}\{1 + o(1)\} = \frac{1}{C_n^2}\{2\,\mathrm{tr}^2(\boldsymbol{\Sigma}^2) + 2\,\mathrm{tr}(\boldsymbol{\Sigma}^4)\}\{1 + o(1)\}.$$

EXAMPLE 3.2 (Autoregressive and moving average). For $\boldsymbol{\Sigma} \in \mathscr{C}_{\mathrm{AR}(1)}$ or $\boldsymbol{\Sigma} \in \mathscr{C}_{\mathrm{MA}(q)}$, we can show that $V_{n1} \sim p^2/n^2$, $V_{n3} \sim 1/n^3$ and $C_{n13} \sim 1/n^3$. It follows that $V_{n1} \gg V_{n3}$. This yields

$$\mathrm{Var}(J_n) = V_{n1}\{1 + o(1)\} = \frac{1}{C_n^2}\{2\,\mathrm{tr}^2(\boldsymbol{\Sigma}^2) + 2\,\mathrm{tr}(\boldsymbol{\Sigma}^4)\}\{1 + o(1)\}.$$

EXAMPLE 3.3 (Compound symmetry). For $\boldsymbol{\Sigma} \in \mathscr{C}_{\mathrm{CS}}$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\theta_0) = \mathbf{I}_p + \theta_0(\mathbf{1}\mathbf{1}^\top - \mathbf{I}_p)$ with $\theta_0 \in (0,1)$. It can be shown that, up to a factor of $\{1 + o(1)\}$,

$$V_{n1} = \frac{1}{C_n^2}\{2\,\mathrm{tr}^2(\boldsymbol{\Sigma}^2) + 2\,\mathrm{tr}(\boldsymbol{\Sigma}^4) + 4p^{-4}(\mathbf{1}^\top\boldsymbol{\Sigma}\mathbf{1})^4 - 8p^{-2}(\mathbf{1}^\top\boldsymbol{\Sigma}^2\mathbf{1})^2\} \sim p^3/n^2$$

and $V_{n3} = 56(\mathbf{1}^\top \mathbf{\Sigma} \mathbf{1})^4/(p^4 n^3)\{1 + o(1)\} \sim p^4/n^3$. Consequently, if $p/n \to \infty$, then $V_{n3} \gg V_{n1}$ and $\mathrm{Var}(J_n) = V_{n3}\{1 + o(1)\}$; in contrast, if $p/n \to 0$, then $V_{n1} \gg V_{n3}$ and $\mathrm{Var}(J_n) = V_{n1}\{1 + o(1)\}$. In addition, under the null hypothesis H_0, the mean of J_n in Example 3.3 is $\mathrm{E}(J_n) = -2(\mathbf{1}^\top \mathbf{\Sigma} \mathbf{1})^2/(p^2 n) \sim p^2/n$, which is negative and cannot be ignored in comparison to the standard error of T_n with order p^2/n. These analytic findings are supported by the numerical evidence given in Example 2.2 of Section 2.

EXAMPLE 3.4 (Polynomial decay). Let $\mathscr{C}_{\mathrm{PN}} = \{\mathbf{\Sigma}(\theta) : \mathbf{\Sigma}(\theta) = (\sigma_{ij}(\theta))_{i,j=1}^p$: $\sigma_{ij}(\theta) = (1 + |i - j|)^{-\theta}\}$ for some $\theta \in (0, 0.5)$. It can be shown that, $V_{n1} \sim p^{(4-4\theta)}/n^2$ and $V_{n3} \sim p^{(4-4\theta)}/n^3$. Accordingly, $\mathrm{Var}(J_n) = V_{n1}\{1 + o(1)\} \sim p^{(4-4\theta)}/n^2$. It can also be verified that $\mathrm{E}(J_{n3}) \sim p^{(2-2\theta)}/n$, which has the same order as the standard deviation of J_n. Thus, the effect of J_{n3} cannot be ignored.

Examples 3.1–3.4 above reveal an important analytical insight about J_n. In Examples 3.1–3.2, J_{n1} is the dominant term, while in Examples 3.3 and 3.4, J_{n3} has a leading order impact if $p \gg n$. In other words, the estimation of $\boldsymbol{\theta}_0$ does not incur any leading order effects in the asymptotic variances and means in Examples 3.1–3.2, but such an effect becomes nonnegligible in Examples 3.3 and 3.4. We propose an adjusted test statistic Λ_n which was given in (2.3) that removes the effect of J_{n3} from the statistic $\hat{T}_n$ by subtracting out the estimator of J_{n3}. As a result, J_{n1} is guaranteed to be the leading order term in the proposed test Λ_n across all four Examples 3.1–3.4. In the next subsections, we present details regarding the asymptotic behavior of Λ_n.

3.2. *Asymptotic distribution of* Λ_n. In this subsection, we study the asymptotic normality of Λ_n under general conditions on $\mathbf{\Sigma}$ and $\mathbf{\Sigma}(\boldsymbol{\theta})$. Assume that $(\sqrt{n}\alpha_p)^{-1}$ is the convergence rate of $\hat{\boldsymbol{\theta}}_0$ to $\boldsymbol{\theta}_0$. It can be verified that $\alpha_p = \sqrt{p}$ for the sphericity, AR(1), and moving average structures and $\alpha_p = 1$ for the compound symmetry structure. In addition, $\alpha_p = \log(p)$ for the polynomial decay structure given in Example 3.4. We next postulate the following two assumptions:

(A1) $\mathrm{tr}\{(\Delta_{\mathbf{\Sigma}} \mathbf{\Sigma})^2\} = o(\xi_1^2)$, where $\Delta_{\mathbf{\Sigma}} = \mathbf{\Sigma} - \sum_{k,l}^q w_{kl}\, \mathrm{tr}(\mathbf{B}_{k,1})\{\partial \mathbf{\Sigma}(\boldsymbol{\theta}_0)/\partial \theta_{0l}\}$.

(A2) (i) $\max_{k,l,m}\{\partial w_{kl}(\boldsymbol{\theta}_0)/\partial \theta_{0m}\}\{\mathrm{tr}(\mathbf{B}_{k,1}^2)\,\mathrm{tr}(\mathbf{B}_{l,1}^2)\}^{1/2} = O(\alpha_p \xi_1)$ and
(ii) $\max_{k,l,m} w_{kl}[\mathrm{tr}\{[\{\partial^2 \mathbf{\Sigma}(\boldsymbol{\theta}_0)/\partial \theta_{0l}\partial \theta_{0m}\}\mathbf{\Sigma}]^2\}\,\mathrm{tr}(\mathbf{B}_{k,1}^2)]^{1/2} = O(\alpha_p \xi_1)$.

Assumption (A1) is satisfied by most of the commonly used covariance structures. This assumption is similar to the condition $\mathrm{tr}(\mathbf{\Sigma}^4) = o\{\mathrm{tr}^2(\mathbf{\Sigma}^2)\}$ given in Chen, Zhang and Zhong (2010), and it is mainly used for showing the asymptotic normality of Λ_n. However, assumption (A1) is weaker than that of Chen, Zhang and Zhong (2010), and it can be satisfied by more general covariance structures such as compound symmetry and polynomial decay. Assumption (A2) ensures that the adjustment of $\hat{T}_n$ satisfies $\hat{J}_{n3} - J_{n3} = o_p(J_{n1})$. Both assumptions are satisfied for the five types of covariance structures discussed in Examples 3.5–3.8 given below.

EXAMPLE 3.5 (Sphericity). Under H_0, let $\mathbf{\Sigma} = \mathbf{\Sigma}(\theta_0) = \theta_0 \mathbf{I}_p$. Then following from Example 3.1, $\xi_1^2 = 2\theta_0^4 p^2\{1 + o(1)\}$ and $\Delta_{\mathbf{\Sigma}} = \theta_0 \mathbf{I}_p - p^{-1}(\theta_0 p)\mathbf{I}_p = 0$. Thus, $\mathrm{tr}\{(\Delta_{\mathbf{\Sigma}} \mathbf{\Sigma})^2\} = 0$. Hence, assumption (A1) is satisfied. Recall that $\alpha_p = \sqrt{p}$. In addition, since $\partial^2 \mathbf{\Sigma}(\theta_0)/\partial\theta_0^2 = 0$, assumption (A2)(ii) holds. One can further verify that $\partial w_{11}(\theta_0)/\partial\theta_0 = 0$, thus, assumption (A2)(i) holds.

EXAMPLE 3.6 (Autoregressive and moving average). We first consider the AR(1) covariance matrix under H_0, that is, $\mathbf{\Sigma} = \mathbf{\Sigma}(\theta_0) = (\theta_0^{|i-j|})_{p\times p}$ with $\theta_0 \in (0, 1)$. It follows from Example 3.2 that $\xi_1^2 \sim p^2$ and $\mathrm{tr}\{(\Delta_{\mathbf{\Sigma}} \mathbf{\Sigma})^2\} = O(p)$. Accordingly, (A1) is satisfied. Recall that $\alpha_p = \sqrt{p}$. Also, note that $\mathrm{tr}[\{(\partial^2 \mathbf{\Sigma}(\theta_0)/\partial\theta_0^2)\mathbf{\Sigma}\}^2] = O(p)$, $\mathrm{tr}(\mathbf{B}_{1,1}^2) = O(p)$, $\partial w_{11}(\theta_0)/\partial\theta_0 = O(1/p)$ and $w_{11} = O(1/p)$. Thus, assumption (A2) holds. By a similar argument, one can verify that assumptions (A1) and (A2) hold for the MA(q) covariance structure.

EXAMPLE 3.7 (Compound symmetry). Under H_0, $\mathbf{\Sigma} = \mathbf{\Sigma}(\theta_0) = \mathbf{I}_p + \theta_0(\mathbf{1}\mathbf{1}^\top - \mathbf{I}_p)$, where $\theta_0 \in (0, 1)$. It follows from Example 3.3 that $\xi_1^2 \sim p^3$ and $\Delta_{\mathbf{\Sigma}} = \mathbf{I}_p$. Then

$$\mathrm{tr}\{(\Delta_{\mathbf{\Sigma}} \mathbf{\Sigma})^2\} = \mathrm{tr}(\mathbf{\Sigma}^2) = \theta_0^2(p^2 - p) + p.$$

We then have $\mathrm{tr}\{(\Delta_{\mathbf{\Sigma}} \mathbf{\Sigma})^2\} \sim p^2$, which is a small order of ξ_1^2. Hence, assumption (A1) is satisfied. Recall that $\alpha_p = 1$. In addition, we know that $\partial w_{11}(\theta_0)/\partial\theta_0 = 0$ and $\partial^2 \mathbf{\Sigma}(\theta_0)/\partial\theta_0^2 = 0$, which together imply that assumption (A2) holds.

EXAMPLE 3.8 (Polynomial decay). Under H_0, $\mathbf{\Sigma} \in \mathscr{C}_{\mathrm{PN}}$ for $\theta \in (0, 0.5)$, where $\mathscr{C}_{\mathrm{PN}}$ has been defined in Example 3.4. It follows from Example 3.4 that $\xi_1^2 \sim p^{4-4\theta}$. After algebraic simplification, we obtain the leading order terms of w_{11} and $\mathrm{tr}(\mathbf{B}_{1,1})$ as $w_{11} = (1 - 2\theta)(2 - 2\theta)p^{2\theta-2}/\log^2(p)$ and $\mathrm{tr}(\mathbf{B}_{1,1}) = \{(1 - 2\theta)(2 - 2\theta)\}^{-1} p^{2-2\theta} \log(p)$. Then $w_{11} \,\mathrm{tr}(\mathbf{B}_{1,1}) \sim 1/\log(p)$ and $\mathrm{tr}\{(\Delta_{\mathbf{\Sigma}} \mathbf{\Sigma})^2\} \le \mathrm{tr}(\mathbf{\Sigma}^2 \Delta_{\mathbf{\Sigma}}^2) \le \mathrm{tr}(\mathbf{\Sigma}^2)\,\mathrm{tr}(\Delta_{\mathbf{\Sigma}}^2) \sim p^{2-2\theta}\,\mathrm{tr}(\Delta_{\mathbf{\Sigma}}^2)$. In addition, it can be shown that $\mathrm{tr}(\Delta_{\mathbf{\Sigma}}^2)$ is of smaller order than $p^{2-2\theta}$. In sum, $\mathrm{tr}\{(\Delta_{\mathbf{\Sigma}} \mathbf{\Sigma})^2\}$ is of smaller order than ξ_1^2. Thus, assumption (A1) is satisfied. As a result, it can be demonstrated that $\partial w_{11}(\theta)/\partial\theta \sim p^{2\theta-2}/\log^2(p)$, $\mathrm{tr}(\mathbf{B}_{1,1}^2) \sim \log^2(p) p^{4-4\theta}$ and $\mathrm{tr}[[\{\partial^2 \mathbf{\Sigma}(\theta)/\partial\theta^2\}\mathbf{\Sigma}]^2\} \sim \log^4(p) p^{4-4\theta}$. Furthermore, we can show that the convergence rate of $\hat{\theta}_0$ is $1/\{\sqrt{n}\log(p)\}$. Accordingly, $\alpha_p = \log(p)$. Combining the above results together, assumption (A2) holds.

We next establish the asymptotic normality of Λ_n under assumptions (A1) and (A2).

THEOREM 4. *Assume that assumptions* (A1)–(A2) *and condition* (C1) *in Appendix hold. If* $na_n = o(\xi_1)$ *and* $r_n \to 0$, *then we have*

$$\sigma_{\Lambda_n,1}^{-1}\{\Lambda_n - \delta(\boldsymbol{\theta}_0)\} \xrightarrow{d} N(0,1), \tag{3.3}$$

where $\delta(\boldsymbol{\theta}_0) = \operatorname{tr}\{(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))^2\}$, a_n *and* r_n *are defined above Theorem* 2 *and* $\sigma_{\Lambda_n,1}^2 = V_{n1} + V_{n2} = 2\xi_1^2/C_n^2 + 8\xi_2^2/n$ *with* ξ_1^2 *and* ξ_2^2 *given in Theorem* 3.

In order to understand the connection between Λ_n and $\hat{T}_n$, we further consider the asymptotic behavior of Λ_n under more restrictive assumptions for $\boldsymbol{\Sigma}$ given below:

(A3) $\operatorname{tr}(\boldsymbol{\Sigma}^4) = o\{\operatorname{tr}^2(\boldsymbol{\Sigma}^2)\}$.

(A4) Assume that $w_{kl} \sim L_p$ such that $L_p \operatorname{tr}(\mathbf{B}_{k,u}^2) = o\{\operatorname{tr}(\boldsymbol{\Sigma}^2)\}$ for $k, l = 1, \ldots, q$ and $u = 1, 2$.

Assumption (A3) is the same as that considered in Chen and Qin (2010) and Chen, Zhang and Zhong (2010). Assumption (A4) is a mild assumption such that the estimator of $\boldsymbol{\theta}_0$ does not yield the leading order effect on the asymptotic variance of T_n. It is easy to show that sphericity, autoregressive and moving average covariance matrices satisfy (A3) and (A4), but compound symmetry and polynomial decay structure given in Example 3.4 do not. It is also worth noting that (A3) and (A4) guarantee that the condition $na_n = o(\xi_1)$, used in Theorem 4, holds. For example, in Examples 3.1–3.2, we have $\xi_1 \sim \operatorname{tr}(\boldsymbol{\Sigma}^2)$, and this together with the fact of $r_n < 1$ leads to $na_n = n^{-1/2} r_n^{1/2} \operatorname{tr}(\boldsymbol{\Sigma}^2) = o(\xi_1)$. Moreover, under (A3) and (A4), the following theorem demonstrates that the asymptotic distribution of Λ_n is the same as that of $\hat{T}_n$.

THEOREM 5. *Under assumptions* (A3)–(A4) *and condition* (C1) *in the Appendix, we have that* $\sigma_{\Lambda_n,1}^{-1} \hat{J}_{n3} = \sigma_{\Lambda_n,1}^{-1}(\Lambda_n - \hat{T}_n) = o_p(1)$, *where* $\sigma_{\Lambda_n,1}^2 = 2\operatorname{tr}^2(\boldsymbol{\Sigma}^2)/C_n^2 + 8\xi_2^2/n$.

Theorem 5 provides the theoretical basis for the phenomenon observed in the histograms of $\hat{T}_n$ and Λ_n in Figure 1; namely that they are asymptotically equivalent under the sphericity covariance structure. In contrast, Figure 2, under the compound symmetry covariance structure, exhibits a considerable discrepancy between the histograms of $\hat{T}_n$ and Λ_n, since (A3) and (A4) are not satisfied.

REMARK 3. To simplify the expression of $\operatorname{Var}(\Lambda_n)$, we imposed the normality assumption on the error terms $\boldsymbol{\varepsilon}_i$ in model (1.1). This allows us to avoid the evaluation of the remainder terms of $\operatorname{Var}(\Lambda_n)$ due to nonnormality. However, this normality condition can be relaxed for Theorem 5, since (A3) and (A4) ensure that the remainder terms of $\operatorname{Var}(\Lambda_n)$ associated with nonnormality are of smaller orders than $\mathrm{tr}^2(\boldsymbol{\Sigma}^2)/n^2 + \xi_2^2/n$. Based on these findings, one can show that the asymptotic normality of Λ_n in Theorem 4 holds for nonnormal errors $\boldsymbol{\varepsilon}_i$ under assumptions (A3)–(A4) and condition (C1). A detailed discussion of nonnormality is given in Section 3 of the supplemental material.

REMARK 4. In the fixed p case, under H_0, $\hat{T}_n - E(J_n)$ and Λ_n are degenerate U-statistics that follow a weighted chi-square distribution rather than a normal distribution [see Serfling (1980), page 194]. Specifically, under H_0, both $n\{\hat{T}_n - E(J_n)\}$ and $n\Lambda_n$ converge to a weighted chi-square distribution $\sum_{k=1}^{\infty} \lambda_k(\chi_{1,k}^2 - 1)$, where the λ_j's are eigenvalues of the kernel $h(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j) = (\boldsymbol{\varepsilon}_i^\top \boldsymbol{\varepsilon}_j)^2 - \boldsymbol{\varepsilon}_i^\top \boldsymbol{\Sigma}\boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_j^\top \boldsymbol{\Sigma}\boldsymbol{\varepsilon}_j + \mathrm{tr}(\boldsymbol{\Sigma}^2) - \mathbf{Q}^\top(\boldsymbol{\varepsilon}_i)\boldsymbol{\Omega}_0\mathbf{Q}(\boldsymbol{\varepsilon}_j)$ and the $\chi_{1,k}^2$s are independent chi-square random variables with one degree of freedom. However, the weighted chi-square distribution is not easy to use in practice. Hence, we apply the Satterthwaite's method to approximate the weighted chi-square distribution by $g_1\chi_{g_2}^2 - g_3$, where $g_1 = \sum_k \lambda_k^2 / \sum_k \lambda_k$, $g_2 = (\sum_k \lambda_k)^2 / \sum_k \lambda_k^2$ and $g_3 = \sum_k \lambda_k$. It can be shown that $E\{h(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_1)\} = \sum_k \lambda_k = \mathrm{tr}^2(\boldsymbol{\Sigma}) + \mathrm{tr}(\boldsymbol{\Sigma}^2) - 2\,\mathrm{tr}(\boldsymbol{\Omega}_0\mathbf{V}_1)$ and $E\{h^2(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2)\} = \sum_k \lambda_k^2 = 2\xi_1^2$, where ξ_1^2 was defined directly above Theorem 3. Plugging in the estimators of $\boldsymbol{\Sigma}$, $\boldsymbol{\Omega}_0$, $\mathbf{V}_1$ and $\mathbf{V}_2$ in $E\{h(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_1)\}$ and $E\{h^2(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2)\}$, we then obtain the estimators for g_1, g_2 and g_3. For more details, see Section 4.2 of the supplemental material.

3.3. *Asymptotic null distributions.* In this subsection, we consider the asymptotic null distributions of Λ_n for testing five commonly used covariance structures; namely, sphericity (SP) $\mathscr{C}_{\mathrm{SP}}$, autoregressive (AR) $\mathscr{C}_{\mathrm{AR}(1)}$, moving average (MA) $\mathscr{C}_{\mathrm{MA}(q)}$, compound symmetry (CS) $\mathscr{C}_{\mathrm{CS}}$ and polynomial decay $\mathscr{C}_{\mathrm{PN}}$. Corollary 1 establishes the asymptotic normality of Λ_n under H_0 for testing the above five covariance matrices.

COROLLARY 1. *Let $\mathscr{C}$ be one of the covariance structures among $\mathscr{C}_{\mathrm{SP}}$, $\mathscr{C}_{\mathrm{AR}(1)}$, $\mathscr{C}_{\mathrm{MA}(q)}$, $\mathscr{C}_{\mathrm{CS}}$ and $\mathscr{C}_{\mathrm{PN}}$. Assume that $r_n \to 0$. Under the null hypothesis that $H_0 : \boldsymbol{\Sigma} \in \mathscr{C}$, we have*

$$\sigma_{\Lambda_n,0}^{-1}\Lambda_n \xrightarrow{d} N(0,1), \tag{3.4}$$

where $\sigma_{\Lambda_n,0}^2 = 2\xi_1^2/C_n^2$.

Note that assumptions (A1)–(A2) for testing the five covariance structures, under the null hypothesis, are satisfied. In addition, it can be verified that condition (C1) holds under the null hypothesis. Thus, we do not need include these restrictions in Corollary 1.

To conduct the test based on Λ_n, one needs to estimate $\sigma_{\Lambda_n,0}^2$ in Corollary 1. A natural approach is to replace $\boldsymbol{\Sigma}$ by $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}_0} := \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_0)$ in $\sigma_{\Lambda_n,0}^2$, which leads to an estimator $\hat{\sigma}^2_{\Lambda_n,0} = 2\hat{\xi}_1^2/C_n^2$ with $\hat{\xi}_1^2 = \mathrm{tr}^2\{\mathbf{\Sigma}^2_{\hat{\boldsymbol{\theta}}_0}\} + \mathrm{tr}\{\mathbf{\Sigma}^4_{\hat{\boldsymbol{\theta}}_0}\} + 2\,\mathrm{tr}\{(\hat{\mathbf{\Omega}}_0\hat{\mathbf{V}}_1)^2\} - 4\mathrm{tr}(\hat{\mathbf{\Omega}}_0\hat{\mathbf{V}}_2)$. Here, $\hat{\mathbf{\Omega}}_0$, $\hat{\mathbf{V}}_1$ and $\hat{\mathbf{V}}_1$ are the corresponding estimates of $\mathbf{\Omega}_0$, $\mathbf{V}_1$ and $\mathbf{V}_2$. In summary, we reject the null hypothesis if $\hat{\sigma}^{-1}_{\Lambda_n,0}\Lambda_n \geq z_\alpha$, where z_α stands for the upper αth quantile of a standard normal distribution. Based on Theorem 4 and Corollary 1, the type I error of the proposed test Λ_n is $P(\sigma^{-1}_{\Lambda_n,0}\Lambda_n \geq z_\alpha | \delta(\boldsymbol{\theta}_0) = 0)$. Accordingly, we only need to evaluate the size at only one point $\delta(\boldsymbol{\theta}_0) = 0$. Note that the asymptotic theories developed are point-wise and dependent on the value of $\boldsymbol{\theta}_0$. Consequently, the accuracy of the asymptotic approximation of the proposed test may vary with the value of $\boldsymbol{\theta}_0$.

If assumptions (A3)–(A4) hold, under the null hypothesis H_0, Theorem 5 indicates that the resulting variance of Λ_n is $\sigma^2_{\Lambda_n,0} = 2\,\mathrm{tr}^2(\mathbf{\Sigma}^2)/C_n^2$. Hence, one can reject the null hypothesis if $\hat{\sigma}^{-1}_{\Lambda_n,0}\Lambda_n \geq z_\alpha$, where $\hat{\sigma}_{\Lambda_n,0} = 2\,\mathrm{tr}(\mathbf{\Sigma}^2_{\hat{\boldsymbol{\theta}}_0})/n$.

Corollary 1 shows that the proposed test statistic Λ_n is applicable to test for SP, AR(1), MA(q), CS or polynomial decay covariance structures. However, the applicability of Λ_n is not limited to the above five covariance structures. It could be generalized to many other families if conditions (A1)–(A2) or (A3)–(A4) are satisfied.

3.4. *Asymptotic power.* In this subsection, we study the asymptotic power of the adjusted test Λ_n. Let $\Phi(\cdot)$ be the cumulative distribution function of the standard normal distribution and let $\gamma_n\{\delta(\boldsymbol{\theta}_0)\} = \delta(\boldsymbol{\theta}_0)/\max(\sigma_{\Lambda_n,0}, \sigma_{\Lambda_n,1})$, where $\sigma_{\Lambda_n,0}$ and $\sigma_{\Lambda_n,1}$ are defined, respectively, in Corollary 1 and Theorem 4.

THEOREM 6. *Suppose that assumptions* (A1)–(A2) *and condition* (C1) *in Appendix hold. Then the power function* $\mathscr{B}\{\delta(\boldsymbol{\theta}_0)\} = P_{H_a}(\sigma^{-1}_{\Lambda_n,0}\Lambda_n > z_\alpha)$ *of the test* Λ_n *satisfies*

$$\lim_{n\to\infty}\left|\mathscr{B}\{\delta(\boldsymbol{\theta}_0)\} - \Phi\left(-\frac{\sigma_{\Lambda_n,0}}{\sigma_{\Lambda_n,1}}z_\alpha + \frac{\delta(\boldsymbol{\theta}_0)}{\sigma_{\Lambda_n,1}}\right)\right| = 0.$$

Furthermore, the adjusted test Λ_n *is consistent if* $\gamma_n\{\delta(\boldsymbol{\theta}_0)\} \to \infty$.

To illustrate the properties of $\gamma_n\{\delta(\boldsymbol{\theta}_0)\}$ and the consistency of the proposed test, we present the following two examples.

EXAMPLE 3.9. Consider the test under the scenario that the eigenvalues of $\mathbf{\Sigma}$ and $\mathbf{\Sigma}(\theta_0)$ are bounded away from both 0 and infinity. In this case, assumptions (A3)–(A4) hold. Then, applying Theorem 5, we have $\sigma^2_{\Lambda_n,1} = 2(C_n^2)^{-1}\,\mathrm{tr}^2(\mathbf{\Sigma}^2) + 8n^{-1}\mathrm{tr}[\{\mathbf{\Sigma}(\mathbf{\Sigma} - \mathbf{\Sigma}(\theta_0))\}^2]$ and $\sigma^2_{\Lambda_n,0} = 2\,\mathrm{tr}^2(\mathbf{\Sigma}^2)/C_n^2$. Hence,

$$\max(\sigma^2_{\Lambda_n,0}, \sigma^2_{\Lambda_n,1}) = O\{n^{-2}\,\mathrm{tr}^2(\mathbf{\Sigma}^2)\} + O\{n^{-1}\delta(\theta_0)\}.$$

Consequently, $\gamma_n\{\delta(\theta_0)\} \to \infty$ as long as $n\delta(\theta_0)/\mathrm{tr}(\mathbf{\Sigma}^2) \to \infty$. For example, the sphericity covariance structure satisfies assumptions (A3) and (A4), and in order to test $H_0 : \mathbf{\Sigma}(\theta) = \theta \mathbf{I}_p$, we note that $\theta_0 = \mathrm{tr}(\mathbf{\Sigma})/p$ minimizes $\delta(\theta)$. According to the above discussion, the test Λ_n is consistent as long as

$$\frac{n\delta(\theta_0)}{\mathrm{tr}(\mathbf{\Sigma}^2)} = \frac{n\,\mathrm{tr}\{(\mathbf{\Sigma} - p^{-1}\,\mathrm{tr}(\mathbf{\Sigma})\mathbf{I}_p)^2\}}{\mathrm{tr}(\mathbf{\Sigma}^2)} = n\left(1 - \frac{\mathrm{tr}^2(\mathbf{\Sigma})}{p\,\mathrm{tr}(\mathbf{\Sigma}^2)}\right) \to \infty.$$

This holds as $1 - \mathrm{tr}^2(\mathbf{\Sigma})/\{p\,\mathrm{tr}(\mathbf{\Sigma}^2)\} > 0$, which measures the departure from the null sphericity hypothesis. The finding is consistent with the results of Theorem 4(i) given by Chen, Zhang and Zhong (2010).

EXAMPLE 3.10. Consider testing $H_0 : \mathbf{\Sigma} \in \mathscr{C}_{\mathrm{CS}}$ against a sequence of alternatives $\mathbf{\Sigma}$ satisfying $\mathrm{tr}(\mathbf{\Sigma}^2) \sim p^2$. Following Example 3.3, we have $\sigma^2_{\Lambda_n,0} \le n^{-2}\,\mathrm{tr}^2(\mathbf{\Sigma}^2)$. In addition, $\mathrm{tr}[\{\mathbf{\Sigma}(\mathbf{\Sigma} - \mathbf{\Sigma}(\theta_0))\}^2] \le \mathrm{tr}(\mathbf{\Sigma}^2)\delta(\theta_0) = O\{p^2\delta(\theta_0)\}$, which implies that $\sigma^2_{\Lambda_n,1} \le 2\,\mathrm{tr}^2(\mathbf{\Sigma}^2)/C_n^2 + O\{n^{-1}p^2\delta(\theta_0)\} = O\{n^{-2}\,\mathrm{tr}^2(\mathbf{\Sigma}^2)\} + O\{n^{-1}p^2\delta(\theta_0)\}$. Hence,

$$\max(\sigma_{\Lambda_n,0}, \sigma_{\Lambda_n,1}) = O\{n^{-1}\mathrm{tr}(\mathbf{\Sigma}^2)\} + O\{n^{-1/2}p\delta^{1/2}(\theta_0)\}.$$

Note that $\mathrm{tr}(\mathbf{\Sigma}^2) \sim p^2$. Accordingly, $\gamma_n\{\delta(\theta_0)\} \to \infty$ if $n\delta(\theta_0)/\mathrm{tr}(\mathbf{\Sigma}^2) \to \infty$. Using the fact that $\theta_0 = \{\mathbf{1}^\top \mathbf{\Sigma}\mathbf{1} - \mathrm{tr}(\mathbf{\Sigma})\}/\{p(p-1)\}$ is the minimizer of $\delta(\theta)$, the test Λ_n is consistent if

$$\frac{n\delta(\theta_0)}{\mathrm{tr}(\mathbf{\Sigma}^2)} = n\{1 - (\mathbf{1}^\top \mathbf{\Sigma}\mathbf{1})^2/\{p^2\,\mathrm{tr}(\mathbf{\Sigma}^2)\}\}\{1 + o(1)\} \to \infty.$$

The above equation holds when $1 - (\mathbf{1}^\top \mathbf{\Sigma}\mathbf{1})^2/\{p^2\,\mathrm{tr}(\mathbf{\Sigma}^2)\} > 0$.

To further study the asymptotic power of Λ_n for testing sphericity, we next compare Λ_n with the test proposed by Ledoit and Wolf (2002), the test proposed by Chen, Zhang and Zhong (2010) and the Sign test introduced by Zou et al. (2014). For simplicity, we name the test proposed by Ledoit and Wolf (2002) as the LW test in the rest of the paper. According to Corollary 2 of Zou et al. (2014), their Sign test is asymptotically as powerful as that of Chen, Zhang and Zhong (2010). Thus, we only compare the power of our proposed test Λ_n with that of Sign test and the LW test.

Following the setting of Zou et al. (2014), we test for sphericity against a sequence of local alternatives

$$H_{n,a} : \mathbf{\Sigma} = \theta_0(\mathbf{I}_p + \mathbf{D}_{n,p}), \tag{3.5}$$

where $\mathbf{D}_{n,p}$ has zero diagonal elements such that $n\,\mathrm{tr}(\mathbf{D}^2_{n,p})/p$ is a positive constant and $\theta_0 > 0$. Under the above local alternatives, we have $\delta(\theta_0) = \mathrm{tr}\{(\mathbf{\Sigma} - \theta_0\mathbf{I}_p)^2\} = \theta_0^2\,\mathrm{tr}(\mathbf{D}^2_{n,p}) \sim p/n$, and $\mathrm{tr}[\{\mathbf{\Sigma}(\mathbf{\Sigma} - \mathbf{\Sigma}(\theta_0))\}^2] \le 2\theta_0^4\{\mathrm{tr}(\mathbf{D}^2_{n,p}) + \mathrm{tr}(\mathbf{D}^4_{n,p})\} \le 2\theta_0^4\{\mathrm{tr}(\mathbf{D}^2_{n,p}) + \mathrm{tr}^2(\mathbf{D}^2_{n,p})\} \sim p/n + (p/n)^2$. As a result, ξ_2^2/n is of order $O(p/n^2 + p^2/n^3)$, which is a small order of $\sigma^2_{\Lambda_n,0} \sim p^2/n^2$. Thus, we obtain

$\sigma^2_{\Lambda_n,1} = \sigma^2_{\Lambda_n,0}\{1+o(1)\} = (2p\theta_0^2/n)^2\{1+o(1)\}$. Therefore, the asymptotic power function of Λ_n given in Theorem 6 for testing sphericity can be simplified to

$$\mathscr{B}\{\delta(\theta_0)\} = \Phi\biggl(-z_\alpha + \frac{n\,\mathrm{tr}(\mathbf{D}^2_{n,p})}{2p}\biggr).$$

On the other hand, using Theorem 2 of Zou et al. (2014), we obtain the asymptotic power function of the Sign test as

$$\mathscr{B}^{\mathrm{Sign}}\{\delta(\theta_0)\} = \Phi\biggl(-z_\alpha + \frac{n\,\mathrm{tr}(\mathbf{D}^2_{n,p})}{2p}\biggr).$$

Applying Proposition 2 in Ledoit and Wolf (2002), the asymptotic power function of the LW test is

$$\mathscr{B}^{\mathrm{LW}}\{\delta(\theta_0)\} = \Phi\biggl(-z_\alpha + \frac{n\delta(\theta_0)}{2\,\mathrm{tr}^2(\mathbf{\Sigma})/p}\biggr) = \Phi\biggl(-z_\alpha + \frac{n\,\mathrm{tr}(\mathbf{D}^2_{n,p})}{2\,\mathrm{tr}^2(\mathbf{\Sigma})/(p\theta_0^2)}\biggr).$$

Note that $\mathrm{tr}(\mathbf{\Sigma}) = p\theta_0$ under the above local alternatives. By comparing the above power functions $\mathscr{B}\{\delta(\theta_0)\}$, $\mathscr{B}^{\mathrm{Sign}}\{\delta(\theta_0)\}$ and $\mathscr{B}^{\mathrm{LW}}\{\delta(\theta_0)\}$, we have the following theorem.

THEOREM 7. *For testing sphericity under a sequence of alternatives $H_{n,a}$ defined in* (3.5), *the proposed test Λ_n, the test of Chen, Zhang and Zhong* (2010), *the Sign test of Zou et al.* (2014) *and the LW test of Ledoit and Wolf* (2002) *have the same asymptotic power.*

For testing sphericity, Theorem 7 shows the asymptotic equivalence of the four tests in power performance. However, the proposed test Λ_n is able to test a broad range of covariance structures (e.g., see Examples 3.2–3.4). It is also worth noting that all four test statistics considered in Theorem 7 are designed for general alternative hypotheses. If we are interested in specific alternatives such as the spiked covariance structures defined in Onatski, Moreira and Hallin (2013, 2014), then likelihood-ratio type statistics may be more powerful than the tests considered in Theorem 7 [see Onatski, Moreira and Hallin (2013, 2014)].

4. Simulation results. We consider five simulation examples to evaluate the finite sample performance of the goodness-of-fit test $\hat{T}_n$ and the adjusted test Λ_n. In all simulation experiments, the following model is considered:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \qquad \text{for } i = 1, \dots, n,$$

where $\boldsymbol{\beta} = (0.5, 1)^\top$ and $\mathbf{X}_i = (\mathbf{X}_{i1}, \mathbf{X}_{i2})$ is a $p \times 2$ matrix, and $\mathbf{X}_{i1}$ and $\mathbf{X}_{i2}$ are independently generated from the p-dimensional multivariate normal distribution with mean $\mathbf{0}$ and covariance $(0.6^{|j-k|})_{p\times p}$ $(j, k = 1, \dots, p)$. The dimensions p of the response vector are chosen to be 270, 385, 500, 625, 670, 785, 900 and 1025.

The sample sizes n are chosen as $n = 60, 80, 100$ and 120. Based on 1000 realizations, five covariance structures of $\boldsymbol{\varepsilon}_i$ (i.e., sphericity, autoregressive, moving average, compound symmetry and polynomial decay) are examined by the tests $\hat{T}_n$ and Λ_n. Empirical sizes and powers are reported with the nominal level $\alpha = 0.05$. In testing sphericity, we also compare $\hat{T}_n$ and Λ_n with the Sign test introduced by Zou et al. (2014) and the LW test proposed by Ledoit and Wolf (2002).

EXAMPLE 4.1 (Sphericity). Consider the following null and alternative hypotheses:

$$H_0 : \boldsymbol{\Sigma} \in \mathscr{C}_{\mathrm{SP}} \quad \text{versus} \quad H_a : \boldsymbol{\Sigma} \notin \mathscr{C}_{\mathrm{SP}}.$$

To evaluate the performance of the tests, the jth component of the random error $\boldsymbol{\varepsilon}_i$ is generated independently according to $\varepsilon_{ij} = u_i + z_{ij}$ for $i = 1, \dots, n$ and $j = 1, \dots, p$, where $u_i \sim N(0, \nu^2)$, $z_{ij} \sim N(0, 1)$, u_i and z_{ij} are independent, ν is a nonnegative constant and $u_i = 0$ as $\nu = 0$. To study the size of the test, the data set is simulated by setting $\nu = 0$. In contrast, the data sets are generated by setting $\nu = 0.08$ and 0.10, respectively, for assessing the power of the test. Tables 1 and 2, respectively, report empirical sizes and powers of Λ_n, $\hat{T}_n$, the Sign test and the LW test.

TABLE 1

Empirical sizes of Λ_n, $\hat{T}_n$, the Sign test introduced by Zou et al. (2014) and the LW test proposed by Ledoit and Wolf (2002) for testing sphericity in Example 4.1

			p							
ν	**n**		**270**	**385**	**500**	**625**	**670**	**785**	**900**	**1025**
0 (size)	60	Λ_n	0.048	0.047	0.040	0.037	0.045	0.046	0.050	0.044
		$\hat{T}_n$	0.047	0.047	0.040	0.037	0.045	0.046	0.050	0.044
		Sign	0.061	0.055	0.049	0.046	0.056	0.048	0.064	0.051
		LW	0.060	0.047	0.053	0.047	0.058	0.050	0.065	0.058
	80	Λ_n	0.044	0.059	0.044	0.040	0.052	0.044	0.054	0.040
		$\hat{T}_n$	0.044	0.059	0.044	0.040	0.052	0.044	0.054	0.040
		Sign	0.046	0.068	0.053	0.056	0.056	0.041	0.063	0.045
		LW	0.051	0.060	0.054	0.055	0.060	0.042	0.059	0.049
	100	Λ_n	0.045	0.049	0.048	0.042	0.049	0.043	0.033	0.048
		$\hat{T}_n$	0.045	0.049	0.048	0.042	0.049	0.043	0.032	0.048
		Sign	0.045	0.053	0.051	0.050	0.057	0.050	0.043	0.057
		LW	0.047	0.053	0.056	0.047	0.053	0.050	0.044	0.052
	120	Λ_n	0.049	0.050	0.051	0.042	0.056	0.055	0.040	0.042
		$\hat{T}_n$	0.049	0.050	0.051	0.042	0.056	0.055	0.040	0.042
		Sign	0.045	0.057	0.057	0.046	0.062	0.054	0.042	0.038
		LW	0.048	0.052	0.059	0.047	0.060	0.055	0.042	0.042

TABLE 2
Empirical powers of Λ_n, $\hat{T}_n$, the Sign test introduced by Zou et al. (2014) and the LW test proposed by Ledoit and Wolf (2002) for testing sphericity in Example 4.1

			p							
ν	***n***		**270**	**385**	**500**	**625**	**670**	**785**	**900**	**1025**
0.08	60	Λ_n	0.077	0.125	0.142	0.194	0.213	0.237	0.281	0.318
		$\hat{T}_n$	0.076	0.125	0.142	0.194	0.213	0.236	0.281	0.316
		Sign	0.086	0.137	0.151	0.210	0.222	0.250	0.300	0.323
		LW	0.083	0.135	0.160	0.210	0.226	0.265	0.314	0.346
	80	Λ_n	0.099	0.130	0.219	0.261	0.281	0.346	0.417	0.457
		$\hat{T}_n$	0.099	0.129	0.219	0.260	0.281	0.346	0.417	0.457
		Sign	0.110	0.142	0.239	0.273	0.291	0.354	0.419	0.468
		LW	0.116	0.143	0.241	0.297	0.293	0.373	0.440	0.476
	100	Λ_n	0.139	0.201	0.253	0.368	0.375	0.423	0.514	0.598
		$\hat{T}_n$	0.139	0.199	0.253	0.367	0.375	0.422	0.514	0.597
		Sign	0.136	0.207	0.273	0.361	0.382	0.413	0.513	0.585
		LW	0.142	0.217	0.269	0.375	0.381	0.449	0.539	0.610
	120	Λ_n	0.170	0.255	0.334	0.407	0.463	0.568	0.641	0.710
		$\hat{T}_n$	0.170	0.253	0.333	0.407	0.463	0.567	0.641	0.710
		Sign	0.172	0.255	0.335	0.411	0.458	0.555	0.624	0.713
		LW	0.181	0.262	0.342	0.412	0.473	0.578	0.656	0.728
0.10	60	Λ_n	0.190	0.307	0.408	0.518	0.527	0.645	0.686	0.773
		$\hat{T}_n$	0.188	0.306	0.406	0.517	0.525	0.645	0.686	0.773
		Sign	0.198	0.317	0.416	0.521	0.523	0.631	0.674	0.769
		LW	0.221	0.326	0.429	0.557	0.548	0.656	0.703	0.793
	80	Λ_n	0.293	0.420	0.555	0.703	0.735	0.814	0.877	0.892
		$\hat{T}_n$	0.292	0.419	0.555	0.703	0.734	0.814	0.877	0.892
		Sign	0.297	0.412	0.546	0.701	0.723	0.795	0.866	0.891
		LW	0.302	0.442	0.571	0.719	0.743	0.821	0.882	0.905
	100	Λ_n	0.377	0.542	0.704	0.842	0.856	0.910	0.951	0.981
		$\hat{T}_n$	0.375	0.541	0.704	0.841	0.856	0.910	0.951	0.981
		Sign	0.371	0.526	0.687	0.831	0.858	0.902	0.953	0.976
		LW	0.375	0.558	0.708	0.845	0.865	0.917	0.959	0.975
	120	Λ_n	0.431	0.645	0.823	0.903	0.944	0.971	0.980	0.994
		$\hat{T}_n$	0.430	0.644	0.822	0.903	0.943	0.971	0.980	0.994
		Sign	0.413	0.636	0.807	0.890	0.939	0.968	0.979	0.996
		LW	0.447	0.660	0.830	0.904	0.946	0.976	0.985	0.997

EXAMPLE 4.2 (Autoregressive). Consider the hypotheses

$$H_0 : \mathbf{\Sigma} \in \mathscr{C}_{\mathrm{AR}(1)} \quad \text{vs.} \quad H_a : \mathbf{\Sigma} \notin \mathscr{C}_{\mathrm{AR}(1)}.$$

TABLE 3
Empirical size and power of Λ_n for testing autoregressive covariance structure in Example 4.2

ν	n	p = 270	385	500	625	670	785	900	1025
0 (size)	60	0.045	0.043	0.036	0.050	0.027	0.048	0.033	0.050
	80	0.033	0.059	0.047	0.044	0.045	0.033	0.057	0.050
	100	0.042	0.046	0.056	0.048	0.044	0.050	0.041	0.049
	120	0.049	0.049	0.047	0.055	0.046	0.056	0.056	0.033
0.10	60	0.124	0.171	0.211	0.287	0.315	0.368	0.441	0.474
	80	0.154	0.221	0.304	0.414	0.453	0.536	0.624	0.678
	100	0.207	0.304	0.396	0.515	0.553	0.635	0.721	0.795
	120	0.257	0.390	0.500	0.632	0.677	0.747	0.840	0.882
0.15	60	0.593	0.782	0.892	0.942	0.953	0.979	0.988	0.994
	80	0.759	0.919	0.963	0.989	0.989	0.999	1.000	0.999
	100	0.874	0.970	0.999	0.999	0.999	1.000	1.000	1.000
	120	0.929	0.989	0.998	1.000	1.000	1.000	1.000	1.000
0.20	60	0.978	0.998	1.000	1.000	1.000	1.000	1.000	1.000
	80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	120	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

To study the performance of the adjusted test Λ_n, each jth component of the random error $\boldsymbol{\varepsilon}_i$ is generated from the following model:

$$\varepsilon_{ij} = 0.5\varepsilon_{i(j-1)} + u_{ij} + \eta_i \qquad \text{for } j = 1, \dots, p \text{ and } i = 1, \dots, n,$$

where $\varepsilon_{i0} \sim N(0, 1 - \nu^2)$, $u_{ij} \sim N(0, 0.75(1 - \nu^2))$, $\eta_i \sim N(0, \nu^2)$, $1 - \nu^2 > 0$, ν is a nonnegative number and obviously $\eta_i = 0$ when $\nu = 0$. In addition, u_{ij}, η_i and ε_{i0} are independent. To examine the size of the test, the data set is simulated by setting $\nu = 0$. In contrast, the data sets are generated by setting $\nu = 0.1, 0.15$ and 0.2, respectively, to evaluate the power of the test. Simulation results of the test Λ_n are reported in Table 3.

EXAMPLE 4.3 (Moving average). Consider the hypotheses

$$H_0 : \boldsymbol{\Sigma} \in \mathscr{C}_{\mathrm{MA}(1)} \quad \text{versus} \quad H_a : \boldsymbol{\Sigma} \notin \mathscr{C}_{\mathrm{MA}(1)},$$

where $\mathscr{C}_{\mathrm{MA}(1)}$ is defined in Section 3.1. To investigate the performance of test Λ_n, the random errors $\boldsymbol{\varepsilon}_i$ are generated from the following model:

$$\varepsilon_{ij} = u_{ij} + \rho u_{i(j-1)} + \eta_i \qquad \text{for } j = 1, \dots, p \text{ and } i = 1, \dots, n,$$

where $u_{ij} \sim N\{0, (1 - \nu^2)/(1 + \rho^2)\}$ for $j = 0, \dots, p$, $\eta_i \sim N(0, \nu^2)$, $\rho \in (0, 1)$, $1 - \nu^2 > 0$, ν is a nonnegative constant, and clearly $\eta_i = 0$ when $\nu = 0$. Furthermore, u_{ij}, u_{i0} and η_i are mutually independent. To study the size of the test, the

TABLE 4
Empirical size and power of Λ_n for testing moving average covariance structure in Example 4.3

		p							
ν	***n***	**270**	**385**	**500**	**625**	**670**	**785**	**900**	**1025**
0 (size)	60	0.057	0.060	0.052	0.059	0.038	0.050	0.037	0.053
	80	0.046	0.041	0.052	0.051	0.050	0.050	0.049	0.047
	100	0.056	0.044	0.043	0.049	0.039	0.058	0.048	0.049
	120	0.049	0.045	0.047	0.057	0.054	0.040	0.044	0.050
0.10	60	0.173	0.225	0.342	0.427	0.430	0.537	0.625	0.756
	80	0.230	0.327	0.473	0.569	0.637	0.736	0.851	0.910
	100	0.273	0.437	0.584	0.748	0.777	0.894	0.939	0.967
	120	0.375	0.578	0.722	0.827	0.885	0.967	0.984	0.996
0.15	60	0.880	0.975	0.999	1.000	1.000	1.000	1.000	1.000
	80	0.965	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	100	0.992	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	120	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.20	60	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	120	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

data set is simulated by setting $\rho = 0.5$ and $\nu = 0$. In contrast, the data sets are generated by setting $\rho = 0.5$ and $\nu = 0.1, 0.15$ and 0.2, respectively, to evaluate the power of the test. Simulation results of the test Λ_n are presented in Table 4.

EXAMPLE 4.4 (Compound symmetry). Consider the following hypotheses:

$$H_0 : \mathbf{\Sigma} \in \mathscr{C}_{\mathrm{CS}} \quad \text{vs.} \quad H_a : \mathbf{\Sigma} \notin \mathscr{C}_{\mathrm{CS}}.$$

To evaluate the performance of both tests Λ_n and $\hat{T}_n$, the random errors $\boldsymbol{\varepsilon}_i$ are generated from the following model:

$$\varepsilon_{ij} = \nu \varepsilon_{i(j-1)} + \eta_i + u_{ij} \qquad \text{for } j = 1, \dots, p \text{ and } i = 1, \dots, n,$$

where $\varepsilon_{i0} \sim N(0, 1 - \rho^2)$, $u_{ij} \sim N(0, (1 - \nu^2)(1 - \rho^2))$ and $\eta_i \sim N(0, \rho^2)$ with $\rho^2 < 1$ and $\nu^2 < 1$. In addition, u_{ij}, η_i and ε_{i0} are mutually independent. To check the size of the test, the data set is simulated by setting $\rho = 0.5$ and $\nu = 0$. In contrast, the data sets are generated by setting $\rho = 0.5$ and $\nu = 0.5, 0.7$ and 0.8, respectively, for examining the power of the test. Simulation results of the tests Λ_n and $\hat{T}_n$ are reported in Table 5.

EXAMPLE 4.5 (Polynomial decay). Consider the hypotheses

$$H_0 : \mathbf{\Sigma} \in \mathscr{C}_{\mathrm{PN}} \quad \text{versus} \quad H_a : \mathbf{\Sigma} \notin \mathscr{C}_{\mathrm{PN}}.$$

TABLE 5

Empirical size and power of Λ_n and empirical size of $\hat{T}_n$ for testing compound symmetry covariance structure in Example 4.4

			p							
ν	n		270	385	500	625	670	785	900	1025
0 (size)	60	Λ_n	0.051	0.053	0.043	0.035	0.053	0.058	0.049	0.055
		$\hat{T}_n$	0.001	0.001	0.000	0.000	0.001	0.000	0.001	0.000
	80	Λ_n	0.044	0.040	0.049	0.050	0.052	0.047	0.041	0.058
		$\hat{T}_n$	0.000	0.001	0.000	0.000	0.001	0.001	0.000	0.000
	100	Λ_n	0.050	0.052	0.038	0.041	0.051	0.043	0.052	0.046
		$\hat{T}_n$	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	120	Λ_n	0.061	0.049	0.049	0.037	0.050	0.041	0.051	0.042
		$\hat{T}_n$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.5	60	Λ_n	0.492	0.391	0.284	0.231	0.210	0.181	0.176	0.153
	80	Λ_n	0.785	0.639	0.483	0.412	0.406	0.343	0.308	0.270
	100	Λ_n	0.928	0.818	0.738	0.622	0.593	0.523	0.498	0.424
	120	Λ_n	0.990	0.941	0.873	0.791	0.763	0.679	0.649	0.564
0.7	60	Λ_n	0.989	0.959	0.912	0.875	0.869	0.787	0.753	0.721
	80	Λ_n	1.000	0.998	0.991	0.976	0.968	0.953	0.918	0.900
	100	Λ_n	1.000	1.000	1.000	0.999	0.996	0.995	0.990	0.978
	120	Λ_n	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.994
0.8	60	Λ_n	1.000	0.999	0.996	0.982	0.993	0.982	0.965	0.965
	80	Λ_n	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.997
	100	Λ_n	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	120	Λ_n	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

To evaluate the performance of both tests Λ_n and $\hat{T}_n$, the random errors $\boldsymbol{\varepsilon}_i$ are generated from a multivariate normal distribution with mean zero and covariance $\boldsymbol{\Sigma} = (\sigma_{jl})_{j,l=1}^{p}$, where

$$\sigma_{jl} = (1-\nu)\big(1+|j-l|\big)^{-\theta} + \nu 0.5^{|j-l|},$$

$0 \le \nu < 1$ and $\theta = 0.2$. We set $\nu = 0$ to examine the size of the test, and we set $\nu = 0.4, 0.5$ and 0.6 to evaluate the power of the test. Simulation results of the tests Λ_n and $\hat{T}_n$ are presented in Table 6.

In Tables 1 and 2, the results for testing sphericity indicate that all four tests control the size well, and their powers increase when either n or p gets larger. The powers of Λ_n and $\hat{T}_n$ are comparable to those of the Sign test but slightly less than those of the LW test. However, the LW test does not perform well under nonnormality (see Table 1 in the supplemental material). It is worth noting that the proposed tests are not restricted to testing sphericity, and they are applicable for testing other covariance structures given in Examples 4.2–4.5. Furthermore,

TABLE 6
Empirical size and power of Λ_n and empirical size of $\hat{T}_n$ for testing polynomial decay covariance structure in Example 4.5

			p							
ν	***n***		**270**	**385**	**500**	**625**	**670**	**785**	**900**	**1025**
0 (size)	60	Λ_n	0.045	0.047	0.051	0.035	0.044	0.042	0.049	0.045
		$\hat{T}_n$	0.009	0.010	0.006	0.013	0.012	0.006	0.014	0.003
	80	Λ_n	0.052	0.052	0.037	0.047	0.034	0.033	0.044	0.046
		$\hat{T}_n$	0.007	0.005	0.010	0.008	0.010	0.009	0.007	0.008
	100	Λ_n	0.044	0.035	0.034	0.047	0.050	0.044	0.046	0.042
		$\hat{T}_n$	0.007	0.005	0.004	0.005	0.002	0.007	0.008	0.008
	120	Λ_n	0.052	0.040	0.048	0.037	0.044	0.038	0.047	0.057
		$\hat{T}_n$	0.000	0.004	0.004	0.004	0.004	0.003	0.004	0.001
0.4	60	Λ_n	0.099	0.095	0.096	0.091	0.064	0.094	0.083	0.072
	80	Λ_n	0.170	0.188	0.174	0.175	0.185	0.157	0.158	0.197
	100	Λ_n	0.306	0.319	0.302	0.319	0.306	0.307	0.304	0.289
	120	Λ_n	0.424	0.472	0.480	0.466	0.452	0.484	0.475	0.451
0.5	60	Λ_n	0.251	0.263	0.268	0.274	0.285	0.267	0.262	0.261
	80	Λ_n	0.472	0.497	0.553	0.551	0.525	0.572	0.568	0.559
	100	Λ_n	0.703	0.778	0.800	0.788	0.817	0.815	0.819	0.817
	120	Λ_n	0.868	0.908	0.927	0.933	0.945	0.948	0.952	0.945
0.6	60	Λ_n	0.536	0.583	0.657	0.692	0.698	0.698	0.733	0.723
	80	Λ_n	0.827	0.911	0.924	0.953	0.947	0.959	0.958	0.964
	100	Λ_n	0.962	0.989	0.990	0.999	0.997	0.998	0.997	0.997
	120	Λ_n	0.998	0.999	1.000	1.000	1.000	0.999	1.000	1.000

following an anonymous referee's suggestion, we have compared Λ_n with the corrected likelihood ratio (CLR) test proposed by Cui, Zheng and Li (2013) for testing sphericity and compound symmetry when $p < n$; see Tables 5 and 6 in the supplemental material. The results show that the proposed tests are either slightly better than or comparable to the CLR test. Per an anonymous referee's suggestion, we have also compared the four tests when $\mathbf{\Sigma} - \mathbf{\Sigma}(\theta_0)$ is sparse. The simulation results indicate that these tests are comparable, and they are not presented here due to space limitations.

From the results of Tables 1–6, we have the following findings for Λ_n. (i) Under the null hypothesis, the test Λ_n controls the size well at the nominal level 0.05, across five commonly used covariance structures. (ii) Under the alternative hypothesis, the power rises as the sample size increases. (iii) The power also rises toward 1 when ν increases, which implies that the test Λ_n is consistent. The above findings corroborate theoretical results in Theorems 4–6.

Tables 1 and 2 present the results for testing the sphericity structure in Example 4.1 via the adjusted test Λ_n and the test $\hat{T}_n$, respectively. We find that the discrepancy between these two tests is very small. In addition, we compare Λ_n and $\hat{T}_n$ for testing the covariance structures in Examples 4.2 and 4.3. The results show that these two tests perform very similarly. Therefore, we do not present the results of $\hat{T}_n$. The above findings are supported by the theoretical result in Theorem 5, since the covariance structures in Examples 4.1 to 4.3 satisfy assumptions (A3) and (A4). However, the covariance structure with the compound symmetry in Example 4.4 and the polynomial decay structure in Example 4.5 only satisfy assumptions (A1) and (A2), but not (A3) and (A4). Hence, Tables 5 and 6 show the different performances for Λ_n and $\hat{T}_n$ in accordance with Theorems 4 and 5. Specifically, Λ_n controls the size well, while $\hat{T}_n$ has an extremely small size.

To examine the robustness of the adjusted test Λ_n against the normality assumption of random errors, we have conducted Monte Carlo studies via various nonnormally distributed errors. The results show that Λ_n maintains the size reasonably well and is a consistent test (see Section 3 in the supplemental material). Based on our simulation studies, we conclude that Λ_n is a reliable and powerful test and it can be used for testing a broad range of covariance structures in repeated-measures models.

5. Real data analysis. In this section, we apply the adjusted test to analyze Canadian weather data, collected from 35 weather stations [see Ramsay and Silverman (2005)]. The data set contains daily temperature and precipitation averaged across the period from 1960 to 1994 for each of 35 weather stations. In addition, these 35 weather stations are located in 4 different climate zones: Atlantic, Pacific, Continental and Arctic.

Let $Y_i(t_k)$ and $V_i(t_k)$ be, respectively, precipitation (in log 10 scale) and temperature on day t_k ($k = 1, \dots, 365$) at the ith ($i = 1, \dots, 35$) station, where t_1 was set to be day 100 in the original data set. Furthermore, let $u(i)$ be the climate zone of the ith station with values $1, \dots, 4$ corresponding to Atlantic, Pacific, Continental and Arctic, respectively. To establish the relationship between precipitation and temperature, we consider the following functional linear model [see Ramsay and Silverman (2005)]:

$$Y_i(t_k) = \eta_{u(i)}(t_k) + Z_i(t_k)\beta(t_k) + \varepsilon_i(t_k), \qquad i = 1, \dots, 35, k = 1, \dots, 365, \tag{5.1}$$

where $\eta_u(t_k)$ ($u = 1, 2, 3, 4$) represent the climate zone effects, $Z_i(t_k)$ are zone-adjusted temperatures, $\beta(t_k)$ are unknown coefficients and $\varepsilon_i(t_k)$ are random errors with the homogeneous variance across t_k. It is worth noting that both temperature and precipitation are affected by climate zones. Hence, we employ the following method to remove the climate zone effects from the temperature and obtain $Z_i(t_k)$. Specifically, consider $V_i(t_k) = \alpha(t_k) + \theta_{u(i)}(t_k) + Z_i(t_k)$, where $\theta_u(t)$ represents the climate zone effect satisfying $\sum_{i=1}^4 \theta_u(t) = 0$. Then apply Fourier series expansions and obtain $\alpha(t) = \sum_{m=1}^5 c_{\alpha,m}\phi_l(t)$ and $\theta_u(t) = \sum_{m=1}^5 c_{\theta,um}\phi_m(t)$, where

$\phi_m(t)$ are Fourier basis functions for $m = 1, \dots, 5$. The resulting least squares estimates $\hat{c}_{\alpha,um}$ and $\hat{c}_{\theta,um}$ of the coefficients $c_{\alpha,m}$ and $c_{\theta,um}$ yield the zone-adjusted temperature $Z_i(t_k) = V_i(t_k) - \sum_{m=1}^{M} \hat{c}_{\alpha,m}\phi_m(t_k) - \sum_{m=1}^{M} \hat{c}_{\theta,u(i)m}\phi_m(t_k)$.

Using the Fourier series approximation, we can represent model (5.1) as an approximate repeated-measures model. Let $\mathbf{B}(t) = (B_1(t), \dots, B_5(t))^T$ be a set of Fourier basis functions. Then $\eta_u(t_k) = \sum_{s=1}^{5} b_{0,us}B_s(t_k)$ and $\beta(t_k) = \sum_{s=1}^{5} b_{1,s}B_s(t_k)$. In addition, let $\mathbf{D}_u$ be a 4-dimensional row vector, where the uth component is 1 and the rest of the components are zero. For example, if $u = 1$, then $\mathbf{D}_1 = (1, 0, 0, 0)$. Moreover, let $\mathbf{B}$ be a $p \times 5$ matrix such that $\mathbf{B} = (\mathbf{B}(t_1), \dots, \mathbf{B}(t_p))^\top$, where $p = 365$. It follows that model (5.1) can be approximated as

$$\mathbf{Y}_i = (\mathbf{D}_{u(i)} \otimes \mathbf{B})\boldsymbol{\beta}_0 + \mathbf{W}_i\boldsymbol{\beta}_1 + \tilde{\boldsymbol{\varepsilon}}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \tag{5.2}$$

where $\mathbf{Y}_i = (Y_i(t_1), \dots, Y_i(t_p))^\top$ is a $p \times 1$ vector, $\mathbf{X}_i = (\mathbf{D}_{u(i)} \otimes \mathbf{B}, \mathbf{W}_i)$, $\otimes$ represents the Kronecker product, $\mathbf{D}_{u(i)} \otimes \mathbf{B}$ is a $p \times 20$ matrix, $\mathbf{W}_i = (Z_i(t_1)\mathbf{B}(t_1), \dots, Z_i(t_p)\mathbf{B}(t_p))^\top$ is a $p \times 5$ matrix, $\boldsymbol{\beta} = (\boldsymbol{\beta}_0^\top, \boldsymbol{\beta}_1^\top)^\top$, $\boldsymbol{\beta}_0 = (b_{0,11}, \dots, b_{0,15}, \dots, b_{0,41}, \dots, b_{0,45})^\top$ is a 20×1 vector and $\boldsymbol{\beta}_1 = (b_{1,1}, \dots, b_{1,5})^\top$ is a 5×1 vector and $\boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_i(t_1), \dots, \boldsymbol{\varepsilon}_i(t_p))^\top$ is a $p \times 1$ vector. Furthermore, assume that $\boldsymbol{\varepsilon}_i$ are IID random vectors with mean 0 and covariance $\boldsymbol{\Sigma}$.

The goal of this study is to predict precipitation via model (5.2). Specifically, we predict $\mathbf{Y}_i^{(2)} = (Y_i(t_{336}), \dots, Y_i(t_{365}))^\top$ by using $\mathbf{Y}_i^{(1)} = (Y_i(t_1), \dots, Y_i(t_{335}))^\top$. To this end, we correspondingly partition $\mathbf{X}_i$, $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\Sigma}$ as follows:

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{X}_i^{(1)} \\ \mathbf{X}_i^{(2)} \end{pmatrix}, \qquad \boldsymbol{\varepsilon}_i = \begin{pmatrix} \boldsymbol{\varepsilon}_i^{(1)} \\ \boldsymbol{\varepsilon}_i^{(2)} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Accordingly, the best linear unbiased predictor (BLUP) of $\mathbf{Y}_i^{(2)}$ given $\mathbf{Y}_i^{(1)}$ is

$$\hat{\mathbf{Y}}_i^{(2)} = \mathbf{X}_i^{(2)}\hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{21}(\hat{\boldsymbol{\theta}}_0)\boldsymbol{\Sigma}_{11}^{-1}(\hat{\boldsymbol{\theta}}_0)\big(\mathbf{Y}_i^{(1)} - \mathbf{X}_i^{(1)}\hat{\boldsymbol{\beta}}\big), \tag{5.3}$$

where $\hat{\boldsymbol{\beta}}$ is the least squares estimate of $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\theta}}_0$ is the least squares estimate of $\boldsymbol{\theta}_0$, and both of them are calculated via the first 335 observations.

Since the accuracy of prediction relies on the covariance structure, we next employ the adjusted test Λ_n to identify the best $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ from the four covariance structures: sphericity, MA(1), AR(1) and compound symmetry presented in Examples 3.1 to 3.3. The resulting test statistics and their associated p-values in the parentheses are 17.09 (<0.0001), 6.85 (<0.0001), 4.43 (<0.0001) and 0.29 (0.38), respectively. In sum, the compound symmetry is likely to be the most appropriate covariance for predictions.

To assess the efficacy of the covariance structure determined by Λ_n, we investigate the performance of predictions obtained from the four covariance structures. We follow the approach of out-of-sample forecasting by randomly partitioning the data into a training set of 30 stations and a validation (or test) set of 5 stations. For

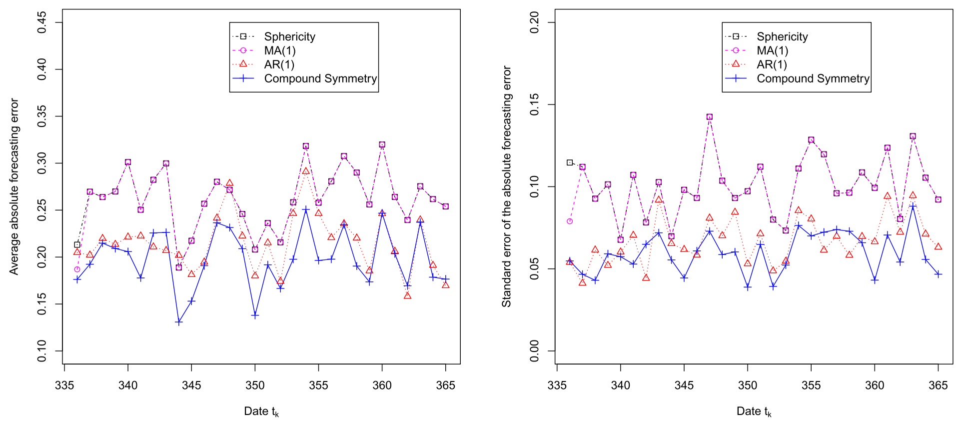

FIG. 3. *Average* (*left panel*) *and standard errors* (*right panel*) *of absolute prediction errors obtained from four types of covariance structures*: *sphericity*, MA(1), AR(1) *and compound symmetry.*

each training data set, we use the first 335 repeated measurements to estimate the unknown parameters. For each test data set $\mathscr{T}_S^{(j)}$ $(j = 1, \dots, 30)$, the performance of the prediction at date t_k is measured by the following absolute prediction error:

$$\mathrm{AE}_k^{(j)} = \frac{1}{5} \sum_{i \in \mathscr{T}_S^{(j)}} \left| \hat{Y}_i(t_k) - Y_i(t_k) \right| \qquad \text{for } k = 336, \dots, 365.$$

Subsequently, based on the four types of covariance structures, we employ the BLUP (5.3) to make predictions and then calculate their absolute prediction errors. The averages and standard errors of $\mathrm{AE}_k^{(j)}$ from 30 replicates at each date t_k are plotted in Figure 3. They clearly show that the predictions using the compound symmetry covariance exhibit the smallest average prediction errors and almost the smallest standard errors across all 30 validation days, followed by AR(1), MA(1) and Sphericity. This finding indicates that Λ_n is a reliable test for determining an adequate covariance and for making forecasts.

6. Discussion. In regression analysis, assessing the appropriateness of the covariance model assumption plays an important role in making valid inferences and desirable predictions. This paper considers the regression model with high dimensional repeated measurements and proposes an adjusted test that enables us to examine a wide range of covariance structures. The proposed test can be applied or extended to test whether the covariance can be approximated by a linear combination of several types of covariance structures [see, e.g., Anderson (1973)]. This warrants further investigation.

For the ease of presentation, we have assumed $\boldsymbol{\mu} = 0$ and $E(\mathbf{X}_i) = 0$ in this paper. In general, if we relax these assumptions, we could construct a test statistic as follows. Let $\tilde{\mathbf{X}}_i = \mathbf{X}_i - \bar{\mathbf{X}}$ and $\tilde{\mathbf{Y}}_i = \mathbf{Y}_i - \bar{\mathbf{Y}}$, where $\bar{\mathbf{Y}}$ and $\bar{\mathbf{X}}$ are sample means. Let $\tilde{\boldsymbol{\beta}} = (\sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{X}}_i)^{-1} \sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{Y}}_i$ and $\tilde{\boldsymbol{\varepsilon}}_i = \tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i \tilde{\boldsymbol{\beta}}$. Define

$\tilde{\boldsymbol{\theta}}_0 = \operatorname{argmin}_\theta \operatorname{tr}\{(\tilde{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\beta}}} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))^2\}$, where $\tilde{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\beta}}} = n^{-1}\sum_{i=1}^n \tilde{\boldsymbol{\varepsilon}}_i\tilde{\boldsymbol{\varepsilon}}_i^\top$. In addition, define $\tilde{T}_{n1}$ and $\tilde{J}_{n3}$ as the statistics $\hat{T}_{n1}$ and $\hat{J}_{n3}$, respectively, except replacing $\hat{\boldsymbol{\varepsilon}}_i$ with $\tilde{\boldsymbol{\varepsilon}}_i$ and $\hat{\boldsymbol{\theta}}_0$ with $\tilde{\boldsymbol{\theta}}_0$. Moreover, define

$$\tilde{T}_{n2} = \frac{1}{3!C_n^3}\sum_{i\neq j\neq k}\tilde{\boldsymbol{\varepsilon}}_i^T\{\tilde{\boldsymbol{\varepsilon}}_j\tilde{\boldsymbol{\varepsilon}}_j^T - \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}}_0)\}\tilde{\boldsymbol{\varepsilon}}_k \quad \text{and} \quad \tilde{T}_{n3} = \frac{1}{4!C_n^4}\sum_{i\neq j\neq k\neq l}\tilde{\boldsymbol{\varepsilon}}_i^T\tilde{\boldsymbol{\varepsilon}}_j\tilde{\boldsymbol{\varepsilon}}_k^T\tilde{\boldsymbol{\varepsilon}}_l.$$

Then the test statistic $\tilde{\Lambda}_n = \tilde{T}_{n1} - 2\tilde{T}_{n2} + \tilde{T}_{n3} - \tilde{J}_{n3}$ can be used to test the null hypothesis of H_0 in the paper. The asymptotic distribution of $\tilde{\Lambda}_n$ is the same as that of Λ_n, since $\tilde{T}_{n2}$ and $\tilde{T}_{n3}$ have no leading order effects. A detailed proof can be found in Section 5 of the supplemental material.

APPENDIX

Technical conditions. To facilitate the theoretical proof, the following technical conditions are considered:

(C1) Assume that $\tilde{\boldsymbol{\theta}}_0$ is in a small neighborhood of $\boldsymbol{\theta}_0$. (i) For any $i, j \in \{1, \dots, q\}$,

$$\operatorname{tr}\left\{\frac{\partial^2\boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}}_0)}{\partial\theta_{0i}\partial\theta_{0j}}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) - \boldsymbol{\Sigma})\right\} = o\left\{\operatorname{tr}\left(\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial\theta_{0i}}\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial\theta_{0j}}\right)\right\};$$

$$\operatorname{tr}\left\{\left(\frac{\partial^2\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial\theta_{0i}\partial\theta_{0j}}(\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) - \boldsymbol{\Sigma})\right)^2\right\} = O\left\{\operatorname{tr}^2\left(\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial\theta_{0i}}\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial\theta_{0j}}\right)\right\}.$$

(ii) For any $i, j, k \in \{1, \dots, q\}$,

$$\operatorname{tr}\left\{\left(\frac{\partial^3\boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}}_0)}{\partial\theta_{0i}\partial\theta_{0j}\partial\theta_{0k}}\boldsymbol{\Sigma}\right)^u\right\} = O\left\{\operatorname{tr}\left\{\left(\frac{\partial^2\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial\theta_{0i}\partial\theta_{0j}}\boldsymbol{\Sigma}\right)^u\right\}\right\} \qquad \text{for } u = 1, 2;$$

$$\operatorname{tr}\left\{\left(\frac{\partial^2\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial\theta_{0i}\partial\theta_{0j}}\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\right)^2\right\} = O\left\{\operatorname{tr}^2\left(\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial\theta_{0i}}\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)}{\partial\theta_{0j}}\right)\right\}.$$

Condition (C1)(i) is not needed under H_0: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. This condition specifies that $\boldsymbol{\Sigma}$ is not too far away from $\mathscr{C}$. Otherwise, distinguishing $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ becomes trivial since the distance between $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ is so large. (C1)(ii) extends condition (a) in Browne (1974) to accommodate the high-dimensional setup, which is used to obtain the asymptotic expression of $\hat{\boldsymbol{\theta}}_0$.

Acknowledgments. We are grateful to the Editors, Associate Editor and three referees for their constructive and insightful comments that helped improve the quality of the paper.

SUPPLEMENTARY MATERIAL

Supplement to "Tests for covariance structures with high-dimensional repeated measurements" (DOI: 10.1214/16-AOS1481SUPP; .pdf). This supplemental material provides technical proofs of the main results in Section 3, some asymptotic results on the proposed test statistics for nonnormally distributed random vectors, and the proof of the asymptotic normality of the test statistic given in Section 6. We also provide the proof of Remark 4 and present additional numerical simulation experiments to compare our proposed test statistics with some existing methods.

REFERENCES

ANDERSON, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141. MR0331612

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444

BALTAGI, B. H., KAO, C. and PENG, B. (2015). On testing for sphericity with non-normality in a fixed effects panel data model. *Statist. Probab. Lett.* **98** 123–130. MR3306453

BENTLER, P. M. and BONETT, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* **88** 588–606.

BROWNE, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statist. J.* **8** 1–24. MR0370949

CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697

CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863

CUI, H., ZHENG, S. and LI, R. (2013). Test of significance on high dimensional covariance matrix structures. Technical report.

DAVIS, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. Springer, New York. MR1883764

DIGGLE, P. J., HEAGERTY, P., LIANG, K.-Y. and ZEGER, S. L. (2002). *Analysis of Longitudinal Data*, 2nd ed. Oxford Univ. Press, New York. MR2049007

FREES, E. W. (2004). *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. Cambridge Univ. Press, Cambridge. MR2087947

JOHN, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127. MR0275568

JOHN, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59** 169–173. MR0312619

LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169

LI, J. and CHEN, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40** 908–940. MR2985938

LI, W. and QIN, Y. (2014). Hypothesis testing for high-dimensional covariance matrices. *J. Multivariate Anal.* **128** 108–119. MR3199831

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2014). Signal detection in high dimension: The multispiked case. *Ann. Statist.* **42** 225–254. MR3189485

POURAHMADI, M. (2013). *High-Dimensional Covariance Estimation*. Wiley, Hoboken, NJ. MR3235948

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

WIESEL, A., BIBI, O. and GLOBERSON, A. (2013). Time varying autoregressive moving average models for covariance estimation. *IEEE Trans. Signal Process.* **61** 2791–2801. MR3064093

ZHAO, W., LI, H., HOU, W. and WU, R. (2007). Wavelet-based parametric functional mapping of developmental trajectories with high-dimensional data. *Genetics* **176** 1879–1892.

ZHONG, P.-S., LAN, W., SONG, P. X. K. and TSAI, C.-L. (2016). Supplement to "Tests for covariance structures with high-dimensional repeated measurements." DOI:10.1214/16-AOS1481SUPP.

ZOU, C., PENG, L., FENG, L. and WANG, Z. (2014). Multivariate sign-based high-dimensional tests for sphericity. *Biometrika* **101** 229–236. MR3180668

P.-S. ZHONG
DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824
USA
E-MAIL: pszhong@stt.msu.edu

W. LAN
CENTER OF STATISTICAL RESEARCH
SOUTHWESTERN UNIVERSITY OF FINANCE AND ECONOMICS
CHENGDU, SICHUAN 611130
CHINA
E-MAIL: facelw@gmail.com

P. X. K. SONG
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109
USA
E-MAIL: pxsong@umich.edu

C.-L. TSAI
GRADUATE SCHOOL OF MANAGEMENT
UNIVERSITY OF CALIFORNIA, DAVIS
DAVIS, CALIFORNIA 95616
USA
E-MAIL: cltsai@ucdavis.edu