Statistica Sinica **32** (2022), 1857-1879
doi:https://doi.org/10.5705/ss.202020.0362

HYPOTHESIS TESTING IN HIGH-DIMENSIONAL LINEAR REGRESSION: A NORMAL-REFERENCE SCALE-INVARIANT TEST

Tianming Zhu, Liang Zhang and Jin-Ting Zhang

National University of Singapore

Abstract: Recently, several non-scale-invariant and scale-invariant tests have been proposed for a general linear hypothesis testing problem for high-dimensional data, which include one-way and two-way MANOVA tests as special cases. Many of these tests impose strong assumptions on the underlying covariance matrix to ensure that their test statistics are asymptotically normally distributed. However, a simulation example and some theoretical justifications indicate that these assumptions are rarely satisfied in practice. As a result, these tests may not be able to maintain their nominal size well. To overcome this problem, we propose a normal-reference scale-invariant test. The test has good size control and power, without imposing strong assumptions on the underlying covariance or correlation matrix. A real-data example and several simulation studies demonstrate that the proposed test has much better size control and power than several non-scale-invariant and scale-invariant tests.

Key words and phrases: General linear hypothesis testing, high-dimensional linear regression, scale-invariant test.

1. Introduction

Modern data collecting and storing technologies mean that many variables are often observed on a few subjects in scientific fields such as biology, medicine, genetics, economics, finance, and so on, resulting in so-called high-dimensional data. Analyzing such data is challenging, because the dimension of the data may be much larger than the sample size. This study is motivated by a corneal surface data set described in Locantore et al. (1999). The data are from a consulting project on a keratoconus disease study with Ms. Nancy Tripoli and Dr. Kenneth L. Cohen of the Department of Ophthalmology, University of North Carolina at Chapel Hill. According to varying degrees of the keratoconus disease, when the corneas are misshaped, 150 corneal surfaces are classified into four groups, and each corneal surface has 6,912 measurements. Of interest is to

Corresponding author: Jin-Ting Zhang, Department of Statistics and Data Science, National University of Singapore, Singapore 117546. E-mail: stazjt2020@nus.edu.sg.

check whether the four corneal surface groups have the same mean corneal surface. Because the data dimension is much larger than the total sample size, this is a special case of a one-way multivariate analysis of variance (MANOVA) problem for high-dimensional data, which tests the equality of the mean vectors of several independent high-dimensional samples. The classical one-way MANOVA tests, such as Lawley–Hotelling's and Bartlett–Nanda–Pillai's trace tests, require that the data dimension be much smaller than the total sample size, and hence are not applicable.

It is well known that the one-way MANOVA problem is a special case of the general linear hypothesis testing (GLHT) problem in multivariate linear regression. We are interested in the GLHT problem in high-dimensional linear regression where both the response variable and the regression coefficient are high-dimensional. Mathematically, suppose $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)^\top$ is an $n \times p$ response matrix obtained by independently observing a p-dimensional response variable for n subjects, where n can be much smaller than p, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$ is a known $n \times k$ full-rank design matrix with $\text{rank}(\mathbf{X}) = k < n - 2$, $\mathbf{\Theta}$ is a $k \times p$ unknown parameter matrix, and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)^\top$ is an $n \times p$ error matrix, where $\boldsymbol{\epsilon}_i$, for $i = 1, \ldots, n$ are independent and identically distributed (i.i.d.) with mean vector $\mathrm{E}(\boldsymbol{\epsilon}_i) = \mathbf{0}$ and covariance matrix $\mathrm{Cov}(\boldsymbol{\epsilon}_i) = \mathbf{\Sigma}$. A high-dimensional linear regression model can then be expressed as

$$\mathbf{Y} = \mathbf{X}\mathbf{\Theta} + \boldsymbol{\epsilon}. \tag{1.1}$$

Of interest is to test the following GLHT problem:

$$\mathrm{H}_0 : \mathbf{C}\mathbf{\Theta} = \mathbf{0}, \quad \text{versus} \quad \mathrm{H}_1 : \mathbf{C}\mathbf{\Theta} \neq \mathbf{0}, \tag{1.2}$$

where $\mathbf{C}$ is a known matrix of size $q \times k$, with $\text{rank}(\mathbf{C}) = q < k$.

Note that the usual least squares estimator of $\mathbf{\Theta}$ is given by $\hat{\mathbf{\Theta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$. Then, the variation matrices due to the hypothesis and error, denoted as $\mathbf{S}_h$ and $\mathbf{S}_e$, respectively, can be expressed as

$$\begin{aligned} \mathbf{S}_h &= (\mathbf{C}\hat{\mathbf{\Theta}})^\top[\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top]^{-1}\mathbf{C}\hat{\mathbf{\Theta}} = \mathbf{Y}^\top\mathbf{H}\mathbf{Y}, \text{ and} \\ \mathbf{S}_e &= (\mathbf{Y} - \mathbf{X}\hat{\mathbf{\Theta}})^\top(\mathbf{Y} - \mathbf{X}\hat{\mathbf{\Theta}}) = \mathbf{Y}^\top(\mathbf{I}_n - \mathbf{P}_X)\mathbf{Y}, \end{aligned} \tag{1.3}$$

where $\mathbf{I}_n$ is the usual $n \times n$ identity matrix, and $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ and $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top[\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top]^{-1}\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ are two useful idempotent matrices of ranks k and q, respectively.

In high-dimensional linear regression settings, where the sample size n is much smaller than the data dimension p, the associated $\mathbf{S}_e : p \times p$ is, in general, degenerate, because its rank is not more than $n - k$, and hence is much smaller than p. Hence, the classical likelihood-ratio tests, such as Lawley–Hotelling's and Bartlett–Nanda–Pillai's trace tests, cannot be applied. Even when these tests can be well defined, they are no longer powerful (Bai and Saranadasa (1996)). To overcome this difficulty, several scale-invariant and non-scale-invariant tests have been proposed. Throughout this paper, a test is called scale-invariant if it is invariant under a scale transformation of the p-variables. Scale-invariant tests include those of Srivastava (2007), Yamada and Srivastava (2012), and Srivastava and Kubokawa (2013), among others, while non-scale-invariant tests include those of Takeda (1999), Fujikoshi, Himeno and Wakaki (2004), Srivastava and Fujikoshi (2006), Schott (2007), and Zhang, Guo and Zhou (2017), among others. Because it is quite common that the p-variables of high-dimensional data may have different scales, a scale-invariant test often has greater power than a non-scale-invariant test. Therefore, a scale-invariant test may be preferred to a non-scale-invariant test, although the former often requires some stronger assumptions. Except for Zhang, Guo and Zhou (2017), who proposed a simple and adaptive non-scale-invariant test for a GLHT problem in the one-way MANOVA setting without imposing strong assumptions on the underlying covariance matrix, most of the above tests impose strong assumptions on the underlying covariance or correlation matrix to ensure that their test statistics are asymptotically normally distributed. In practice, however, these assumptions are rarely satisfied or hardly checked. This means that, in practice, these tests may not maintain their size well, causing misleading conclusions.

To appreciate this, let us consider the scale-invariant one-way MANOVA test proposed by Yamada and Srivastava (2012) as an example. Let $\hat{\mathbf{\Sigma}} = (n-k)^{-1}\mathbf{S}_e$ be the usual unbiased estimator of $\mathbf{\Sigma}$, and let $\hat{\mathbf{D}} = \operatorname{diag}(\hat{\mathbf{\Sigma}})$ be a diagonal matrix formed by the diagonal entries of $\hat{\mathbf{\Sigma}}$. Let $\operatorname{tr}(\cdot)$ denote the usual trace operation. Then, the test statistic of Yamada and Srivastava (2012) can be expressed as

$$T_{\text{YS}} = \frac{\operatorname{tr}(\mathbf{S}_h\hat{\mathbf{D}}^{-1}) - (n-k)pq/(n-k-2)}{\{2q[\operatorname{tr}(\hat{\mathbf{R}}^2) - p^2/(n-k)]c_{n,p}\}^{1/2}}, \tag{1.4}$$

where

$$\hat{\mathbf{R}} = \hat{\mathbf{D}}^{-1/2}\hat{\mathbf{\Sigma}}\hat{\mathbf{D}}^{-1/2} \quad \text{and} \quad c_{n,p} = 1 + \frac{\operatorname{tr}(\hat{\mathbf{R}}^2)}{p^{3/2}} \tag{1.5}$$

are the sample correlation matrix and the so-called adjustment coefficient, respectively, used to improve the convergence of T_{YS} to the standard normal distribution. Yamada and Srivastava (2012) showed that under some regularity conditions, T_{YS} is asymptotically normally distributed. However, in a real data analysis, as indicated by the histograms of the simulated T_{YS} presented in Section S1 of the Supplementary Material, T_{YS} may not always be asymptotically normally distributed if it is blindly applied.

To overcome the above problem, we consider an extension of Zhang, Zhu and Zhang (2020, ZZZ) to the GLHT problem (1.2) in high-dimensional linear regression, resulting in a new scale-invariant test with the following test statistic:

$$T_{n,p} = \frac{n-k-2}{(n-k)pq}\mathrm{tr}(\mathbf{S}_h\hat{\mathbf{D}}^{-1}), \tag{1.6}$$

where $\mathbf{S}_h$ and $\hat{\mathbf{D}}$ are defined as in (1.3) and (1.4), respectively. It is seen from (1.4) and (1.6) that $T_{n,p}$ has a close relation with T_{YS}, and $T_{n,p}$ is always nonnegative, but T_{YS} takes both positive and negative values.

The main contributions of this work, compared with ZZZ, are as follows. First, we consider the GLHT problem in high-dimensional linear regression, which not only includes the two-sample problem of ZZZ as a special case, but also includes one-way and two-way MANOVA tests and their post hoc and contrast tests as special cases. Therefore, the application range of this work is much larger than that of ZZZ. Second, the test statistic (1.6) presented here is more complicated than that in ZZZ, and includes a constant factor $(n-k-2)/(n-k)$ that does not appear in the test statistic of ZZZ. This constant factor is used to take the sample size n and the number of the predictors k into account. It has a strong impact on the performance of the proposed test. It allows the proposed test to work well, even when the sample size is not that large compared with $\log(p)$, as demonstrated in a simulation study presented in the Supplementary Material. Third, following ZZZ, we assume Condition (2.3), which guarantees that the test statistic $T_{n,p}$ can be well approximated by its simplified version $T^*_{n,p}$, obtained by replacing the diagonal matrix $\hat{\mathbf{D}}$ with the underlying diagonal matrix $\mathbf{D}$. This condition requires that the sample size n should not be too small compared with $\log(p)$. However, in constrast to ZZZ, we establish (2.1) and (2.2) under our own Conditions C1–C4, with their nontrivial proofs presented in Section S2 of the Supplementary Material. Fourth, following ZZZ, we show that the test statistic $T_{n,p}$ and a χ^2-type mixture have the same normal and nonnormal limiting distributions. However, the χ^2-type mixture (2.8) obtained here is much more complicated than that of ZZZ, as is the resulting nonnormal limiting distribution. Furthermore, the proof of Theorem 1 is much more involved than that of Theorem 1 of ZZZ. Fifth, intensive simulation studies presented here and in the Supplementary Material demonstrate that, in general, in terms of size control and power, the proposed test outperforms the tests proposed by Fujikoshi, Himeno and Wakaki (2004), Srivastava and Fujikoshi (2006), Yamada and Srivastava (2012), and Li, Aue and Paul (2020).

The rest of the paper is organized as follows. The main results are presented in Section 2. As two special cases of the GLHT problem, one-way and two-way MANOVA tests are briefly discussed in Section 3. Simulation studies and an application to the corneal surface data set are given in Sections 4 and 5, respectively. We conclude the paper in Section 6. Additional discussions, additional simulation studies, and technical proofs of the main results are relegated to the Supplementary Material.

2. Main Results

2.1. Asymptotic null distribution

To test the GLHT problem (1.2), it is necessary to derive the null distribution of $T_{n,p}$. To this end, we first consider $T^*_{n,p} = (n-k-2)/((n-k)pq)\mathrm{tr}(\mathbf{S}_h\mathbf{D}^{-1})$, which is obtained by replacing $\hat{\mathbf{D}}$ in the expression (1.6) with $\mathbf{D} = \mathrm{diag}(\mathbf{\Sigma})$, the population version of $\hat{\mathbf{D}}$. Let $\hat{\sigma}_{rr}$, for $r = 1, \ldots, p$, and σ_{rr}, for $r = 1, \ldots, p$, be the diagonal entries of $\hat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$, respectively. Then, $\hat{\mathbf{D}} = \mathrm{diag}(\hat{\sigma}_{11}, \ldots, \hat{\sigma}_{pp})$ and $\mathbf{D} = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{pp})$. Under Conditions C1–C4, it is shown in Section S6.1 of the Supplementary Material that as $n, p \to \infty$, we have

$$\max_{1\le r\le p}\left|\frac{\sigma_{rr}}{\hat{\sigma}_{rr}} - 1\right| = O_p\left[n^{-1}\log(p)\right], \tag{2.1}$$

and

$$T_{n,p} = T^*_{n,p}\left\{1 + O_p\left[n^{-1}\log(p)\right]\right\}, \tag{2.2}$$

where $O_p(\cdot)$ denotes the "bounded in probability" operation. That is, $T_{n,p}$ and $T^*_{n,p}$ have the same distribution asymptotically, provided that

$$\log(p) = o(n). \tag{2.3}$$

Remark 1. Condition (2.3) requires that the sample size n should not be too small compared with $\log(p)$. This condition is actually weaker than the condition

$$n = O(p^\delta), \quad \delta > \frac{1}{2}, \quad \text{as } p \to \infty, \tag{2.4}$$

imposed as Assumption A0 in Yamada and Srivastava (2012), because in this case, we have $\log(p)/n = [\log(p)/p^\delta][p^\delta/n] \to 0$ as $p \to \infty$. It is also weaker than the condition as $n, p \to \infty, p/n \to c,$ where c is some constant, as imposed in Bai and Saranadasa (1996, Theorem 2.1) and Li, Aue and Paul (2020, Condition C2) for high-dimensional hypothesis testing, because in this case, as $n, p \to \infty$, we have $\log(p)/n = [\log(p)/p](p/n) \to 0$. Condition (2.3) is crucial to the proposed test. When it is not satisfied, we cannot estimate the component variances σ_{rr} accurately, as indicated by (2.1), and the study of the distribution of $T_{n,p}$ cannot be reduced to that of $T^*_{n,p}$, as indicated by (2.2). That is, the resulting test will be less accurate in terms of size control. This is partially verified by a small simulation study presented in Section S2 of the Supplementary Material.

For further study, we can write

$$T^*_{n,p} = \frac{n-k-2}{n-k}\left[T^*_{n,p,0} + 2S_{n,p} + (pq)^{-1}\mathrm{tr}(\ \boldsymbol{\Omega}\mathbf{D}^{-1})\right], \tag{2.5}$$

where

$$\begin{aligned} T^*_{n,p,0} &= (pq)^{-1}\mathrm{tr}(\boldsymbol{\epsilon}^\top\mathbf{H}\boldsymbol{\epsilon}\mathbf{D}^{-1}), \quad S_{n,p} = (pq)^{-1}\mathrm{tr}[(\mathbf{X}\boldsymbol{\Theta})^\top\mathbf{H}\boldsymbol{\epsilon}\mathbf{D}^{-1}], \\ \boldsymbol{\Omega} &= (\mathbf{C}\boldsymbol{\Theta})^\top[\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top]^{-1}(\mathbf{C}\boldsymbol{\Theta}). \end{aligned} \tag{2.6}$$

It is clear that under the null hypothesis, as $n \to \infty$, we have

$$T^*_{n,p} \stackrel{d}{=} \frac{n-k-2}{n-k}T^*_{n,p,0} = T^*_{n,p,0}[1+o(1)], \tag{2.7}$$

where $\stackrel{d}{=}$ denotes equality in distribution. Therefore, studying the asymptotic distribution of $T^*_{n,p}$ under the null hypothesis is equivalent to studying the asymptotic distribution of $T^*_{n,p,0}$. Let $\mathbf{R} = \mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2}$ be the population correlation matrix associated with $\boldsymbol{\Sigma}$. Let $\lambda_{p,1}, \ldots, \lambda_{p,p}$ be the eigenvalues of $\mathbf{R}$ in descending order. When the rows of the measurement error matrix $\boldsymbol{\epsilon}$ are normally distributed, the distribution of $T^*_{n,p,0}$ is the same as that of the following chi-square-type mixture,

$$T^*_{p,0} \stackrel{d}{=} (pq)^{-1}\sum_{r=1}^{p}\lambda_{p,r}A_r, \quad A_r \stackrel{i.i.d.}{\sim} \chi^2_q, \tag{2.8}$$

where χ^2_v denotes a chi-square distribution with v degrees of freedom. Note that the distribution of $T^*_{p,0}$ is not connected to the sample size n. For further study, the first three cumulants of $T^*_{p,0}$ are given by $\mathrm{E}(T^*_{p,0}) = 1$,

$$\mathrm{Var}(T^*_{p,0}) = \frac{2\mathrm{tr}(\mathbf{R}^2)}{p^2q}, \quad \text{and} \quad \mathrm{E}[T^*_{p,0} - \mathrm{E}(T^*_{p,0})]^3 = \frac{8\mathrm{tr}(\mathbf{R}^3)}{p^3q^2}. \tag{2.9}$$

In addition, the skewness of $T^*_{p,0}$ is given by

$$\frac{\mathrm{E}[T^*_{p,0} - \mathrm{E}(T^*_{p,0})]^3}{\mathrm{Var}^{3/2}(T^*_{p,0})} = \sqrt{\frac{8}{d^*}}, \quad \text{where } \ d^* = q\frac{\mathrm{tr}^3(\mathbf{R}^2)}{\mathrm{tr}^2(\mathbf{R}^3)}. \tag{2.10}$$

To derive the limiting distributions of $T^*_{n,p,0}$ and $T^*_{p,0}$ when both n and p tend to infinity, we set $\rho_{p,r} = \lambda_{p,r}/[\mathrm{tr}(\mathbf{R}^2)]^{1/2}$, for $r = 1, \ldots, p$, which are the eigenvalues of $\mathbf{R}/[\mathrm{tr}(\mathbf{R}^2)]^{1/2}$ in descending order, and impose the following conditions:

C1. In the linear regression model (1.1), we can write $\boldsymbol{\epsilon} = \mathbf{V}\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is a $p \times p$ matrix such that $\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma} = \boldsymbol{\Sigma}$, and $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)^\top$, with $\mathbf{v}_1, \ldots, \mathbf{v}_n$ being i.i.d. with $\mathrm{E}(\mathbf{v}_1) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{v}_1) = \mathbf{I}_p$.

C2. Assume $\mathrm{E}(v_{1r}^4) = 3 + \Delta < \infty$ where v_{1r} is the rth component of $\mathbf{v}_1$, Δ is some constant, and $\mathrm{E}(v_{11}^{\ell_1} \cdots v_{1p}^{\ell_p}) = 0$ (or 1) when there is one ℓ_r equal to one or there are two v_k equal to two whenever $\ell_1 + \cdots + \ell_p = 4$, where $\ell_1, \ldots, \ell_p$ are nonnegative integers.

C3. For $\mathbf{H} = (h_{ij})$, $h_{ij} = O(n^{-1})$, for $i, j = 1, \ldots, n$.

C4. There exist two constants c_1 and c_2 such that $0 < c_1 \leq \min_{1\leq r\leq p} \sigma_{rr} \leq \max_{1\leq r\leq p} \sigma_{rr} \leq c_2 < \infty$, for all p.

C5. There exist real numbers $\rho_r, r = 1, 2, \ldots$, such that $\lim_{p\to\infty} \rho_{p,r} = \rho_r, r = 1, 2, \ldots$, uniformly and $\lim_{p\to\infty} \sum_{r=1}^p \rho_{p,r} = \sum_{r=1}^\infty \rho_r < \infty$.

C6. $\lim\limits_{p\to\infty} \mathrm{tr}(\mathbf{R}^2)/p = a \in (0, \infty)$ and $\lim\limits_{p\to\infty} \mathrm{tr}(\mathbf{R}^4)/p^2 = 0$.

Condition C1 specifies a factor model for the measurement error matrix $\boldsymbol{\epsilon}$. Condition C2 assumes that the components of $\mathbf{v}_1$ have finite fourth moments and are nearly independent. For normally distributed $\mathbf{v}_i$, for $i = 1, \ldots, n$, with zero mean vector and identity covariance matrix, Condition C2 is satisfied with $\Delta = 0$. Condition C3 is satisfied as long as $n^{-1}\mathbf{X}^\top\mathbf{X}$ tends to a positive-definite matrix. Condition C4 is regular for scale-invariant tests. Condition C5 ensures the existence of the limits of the eigenvalues of $\mathbf{R}/[\mathrm{tr}(\mathbf{R}^2)]^{1/2}$ as $n, p \to \infty$, and that the limit and summation operations in $\lim_{n,p\to\infty} \sum_{r=1}^p \rho_{p,r}$ are exchangeable. It guarantees that the distributions of the normalized $T^*_{n,p,0}$ and $T^*_{p,0}$, namely,

$$\tilde{T}^*_{n,p,0} = \frac{T^*_{n,p,0} - 1}{\sqrt{2p^{-2}q^{-1}\mathrm{tr}(\mathbf{R}^2)}}, \quad \text{and } \ \tilde{T}^*_{p,0} = \frac{T^*_{p,0} - 1}{\sqrt{2p^{-2}q^{-1}\mathrm{tr}(\mathbf{R}^2)}},$$

are not normally distributed. Condition C6 is imposed by Srivastava and Kubokawa (2013), and is a key condition for the asymptotic normality of $\tilde{T}^*_{n,p,0}$ and $\tilde{T}^*_{p,0}$. In practice, Condition C6 may not be satisfied or is rarely be checked. Fortunately, Condition C6 is not a necessary condition for us to conduct the proposed test with the W–S χ^2-approximation. Let $\xrightarrow{L}$ denote convergence in distribution.

Theorem 1.

(a) *Under Conditions* C1–C5*, as* $n, p \to \infty$*, we have*

$$\tilde{T}^*_{n,p,0} \xrightarrow{L} \zeta, \quad \textit{and} \quad \tilde{T}^*_{p,0} \xrightarrow{L} \zeta, \tag{2.11}$$

where $\zeta \stackrel{d}{=} \sum_{r=1}^{\infty} \rho_r (A_r - q)/\sqrt{2q},\ A_r \stackrel{i.i.d.}{\sim} \chi^2_q$.

(b) *Under Conditions* C1–C4 *and* C6*, as* $n, p \to \infty$*, we have*

$$\tilde{T}^*_{n,p,0} \xrightarrow{L} N(0,1), \quad \textit{and} \quad \tilde{T}^*_{p,0} \xrightarrow{L} N(0,1). \tag{2.12}$$

Therefore, under the conditions of either (a) or (b), as $n, p \to \infty$*, we have*

$$\sup_x |\Pr(T^*_{n,p,0} \le x) - \Pr(T^*_{p,0} \le x)| \to 0. \tag{2.13}$$

Theorem 1 provides a theoretical justification for us to use the distribution of $T^*_{p,0}$ to approximate the distribution of $T^*_{n,p,0}$. Note that $T^*_{p,0}$ is obtained when the rows of the measurement error matrix $\boldsymbol{\epsilon}$ are i.i.d. normally distributed. Thus, we call the distribution of $T^*_{p,0}$ the normal-reference distribution of $T^*_{n,p,0}$, and the resulting test the normal-reference scale-invariant test.

2.2. Implementation

To implement the proposed normal-reference scale-invariant test, under Conditions C1–C4, (2.3), and the null hypothesis, by (2.2), we may approximate the distribution of $T_{n,p}$ with that of $T^*_{n,p}$, and by (2.7), we may approximate the distribution of $T^*_{n,p}$ with that of $T^*_{n,p,0}$. Theorem 1 shows that under Conditions C1–C4 and (2.3), if Condition C5 or C6 is satisfied, we may approximate the distribution of $T^*_{n,p,0}$ with that of $T^*_{p,0}$; see the numerical comparison of the distributions of $T_{n,p}, T^*_{n,p}, T^*_{n,p,0}$, and $T^*_{p,0}$ in Section S3 of the Supplementary Material. Note that $T^*_{p,0}$ is a χ^2-type mixture with unknown coefficients $\lambda_{p,r}$, for $r = 1, \ldots, p$, and it is always nonnegative and usually skewed. This indicates that the null distribution of $T_{n,p}$ is generally skewed, although it may be asymptotically normal under some regularity conditions. Therefore, the normal approximation to the null distribution of $T_{n,p}$ as used in Fujikoshi, Himeno and Wakaki (2004), Srivastava and Fujikoshi (2006), Yamada and Srivastava (2012), and Srivastava and Kubokawa

(2013) is not always appropriate because a normal distribution is always symmetric and bell-shaped. To overcome this difficulty, we suggest approximating the null distribution of $T_{n,p}$ by the well-known W–S χ^2-approximation. Under Conditions C1–C4, (2.3) and the null hypothesis, by (2.2), as $n, p \to \infty$, we have that $\mathrm{E}(T_{n,p})$ tends to $\mathrm{E}(T^*_{n,p}) = 1$. Hence, by the W–S χ^2-approximation, the distribution of $T_{n,p}$ can be approximated by that of the following random variable:

$$G \stackrel{d}{=} \frac{\chi^2_d}{d}, \tag{2.14}$$

the expectation of which is also one. The approximation parameter d is usually called the approximate degrees of freedom of $T_{n,p}$. It can be determined by matching the variances of $T_{n,p}$ and G under the null hypothesis. By (2.14), it is easy to find that $\mathrm{Var}(G) = 2/d$, and if the conditions of Theorem 1(a) or (b) are satisfied, under Condition (2.3) and the null hypothesis, using the information from (2.9), for large samples, we have $\mathrm{Var}(T_{n,p}) = 2(n-k-2)^2/[(n-k)^2p^2q]\mathrm{tr}(\mathbf{R}^2)$ approximately. Thus, equating the variances of $T_{n,p}$ and G under the null hypothesis approximately leads to

$$d = \frac{(n-k)^2p^2q}{(n-k-2)^2\mathrm{tr}(\mathbf{R}^2)}. \tag{2.15}$$

For any nominal significance level $\alpha > 0$, let $\chi^2_v(\alpha)$ denote the upper 100α percentile of the χ^2_v distribution. Let $\hat{d}$ be the ratio-consistent estimator of d in (2.15). Then, the proposed test can be conducted using the approximate critical value $\chi^2_{\hat{d}}(\alpha)/\hat{d}$ or the approximate p-value $\Pr(\chi^2_{\hat{d}}/\hat{d} \geq T_{n,p})$.

By Theorem 5 of Zhang et al. (2020), we have $q \leq d^* \leq p^2q/\mathrm{tr}(\mathbf{R}^2) \leq pq$ where d^* is given in (2.10). This indicates that when the dimension p is finite, both d^* and d are finite. In addition, by Theorem 4 of Zhang et al. (2020), $T^*_{p,0}$ is asymptotically normal if and only if $d^* \to \infty$. Therefore, when $T^*_{p,0}$ is asymptotically normal, the associated $d \to \infty$ as well, so that G is also asymptotically normal. When d is finite, on the other hand, we have that d^* is also finite, in which case, neither G nor $T^*_{p,0}$ is asymptotically normal. Thus, the proposed test with the W–S χ^2-approximation is adaptive to the shape of its null distribution. This advantage is not shared by the tests studied by Fujikoshi, Himeno and Wakaki (2004), Srivastava and Fujikoshi (2006), Yamada and Srivastava (2012), and Srivastava and Kubokawa (2013), where only the asymptotic normal distributions of their test statistics are considered.

We now briefly describe how to implement the proposed normal-reference scale-invariant test in practice. To this end, all we need is to find a ratio-consistent estimator of d or $\mathrm{tr}(\mathbf{R}^2)$, as in the following theorem.

Theorem 2. *Under Conditions* C1–C4 *and* (2.3)*, as $n, p \to \infty$, we have $\widehat{tr(\mathbf{R}^2)}/tr(\mathbf{R}^2) \to 1$ and $\hat{d}/d \to 1$ in probability, where*

$$\widehat{tr(\mathbf{R}^2)} = \frac{(n-k)^2}{(n-k-1)(n-k+2)}\left[tr(\hat{\mathbf{R}}^2) - \frac{p^2}{n-k}\right], \tag{2.16}$$

where $\hat{\mathbf{R}}$ *is given in* (1.5) *and*

$$\hat{d} = \frac{(n-k)^2 p^2 q}{(n-k-2)^2 \widehat{tr(\mathbf{R}^2)}}. \tag{2.17}$$

The proof of Theorem 2 is given in Section S6.3 of the Supplementary Material. When $n \to \infty$, the constant factor $(n-k)^2(n-k-1)^{-1}(n-k+2)^{-1}$ in (2.16) tends to one, and hence can be replaced with one, as in Yamada and Srivastava (2012) and Srivastava and Kubokawa (2013), among others.

2.3. Asymptotic power

In this section, we investigate the asymptotic power of $T_{n,p}$ based on (2.2), (2.5), and (2.6), where $T^*_{n,p,0}, S_{n,p}$, and $\mathbf{\Omega}$ are defined. Following ZZZ, the asymptotic power of $T_{n,p}$ is studied under the following local alternative:

$$\text{as } n, p \to \infty, \quad \mathrm{Var}(S_{n,p}) = o[\mathrm{Var}(T^*_{n,p,0})], \tag{2.18}$$

where $\mathrm{Var}(S_{n,p}) = (pq)^{-2}\mathrm{tr}(\mathbf{R}\mathbf{D}^{-1/2}\ \mathbf{\Omega}\mathbf{D}^{-1/2})$. This is the case when the information in the local alternative is weak compared with the variance of $T^*_{n,p,0}$, so that the variance of $T_{n,p}$ is about the same as that of $T^*_{n,p,0}$.

Theorem 3.

(a) Under Conditions C1–C5, (2.3)*, and the local alternative* (2.18)*, as $n, p \to \infty$, we have*

$$\Pr\left[T_{n,p} \geq \chi^2_{\hat{d}}(\alpha)/\hat{d}\right] = \Pr\left[\zeta \geq \frac{\chi^2_d(\alpha) - d}{\sqrt{2d}} - \frac{tr(\ \mathbf{\Omega}\mathbf{D}^{-1})}{\sqrt{2q\,tr(\mathbf{R}^2)}}\right][1 + o(1)],$$

where ζ is defined in Theorem 1 *and d is given in* (2.15)*.*

(b) Under Conditions C1–C4, C6, (2.3)*, and the local alternative* (2.18)*, as $n, p \to \infty$, we have*

$$\Pr\left[T_{n,p} \geq \frac{\chi^2_{\hat{d}}(\alpha)}{\hat{d}}\right] = \Phi\left[-z_\alpha + \frac{tr(\,\mathbf{\Omega}\mathbf{D}^{-1})}{\sqrt{2q\,tr(\mathbf{R}^2)}}\right][1+o(1)], \tag{2.19}$$

where z_α *denotes the upper* 100α*-percentile of* $N(0,1)$.

Remark 2. Under the conditions of Theorem 3(b), the asymptotic power (2.19) of $T_{n,p}$ is identical to that of T_{YS} (Yamada and Srivastava (2012, Theorem 3.1)). Yamada and Srivastava (2012, Proposition 4.1) showed that in this case, when the p-variables of the high-dimensional data are independent and have different variances, the asymptotic power of T_{YS} is greater than that of the non-scale-invariant test T_{SF} proposed by Srivastava and Fujikoshi (2006). It follows that the asymptotic power of $T_{n,p}$ is also greater than that of T_{SF}. However, under the conditions of Theorem 3(a), such a theoretical power comparison is not immediately available, because the associated asymptotic power of T_{SF} is not available. Nevertheless, the simulation results presented in Section 4 and in Section S4 of the Supplementary Material indicate that when the variances of the p-variables of the high-dimensional data are different and these p-variables are not necessarily independent, the empirical power of $T_{n,p}$ is, in general, larger than that of T_{SF} under various simulation settings.

3. Two Special Cases

As mentioned in Section 1, the GLHT problem (1.2) under the high-dimensional linear regression model (1.1) includes the one-way and two-way MANOVA problems as special cases. In this section, we briefly describe the two problems, and show how to re-write them in the form of the GLHT problem (1.2) under the high-dimensional linear regression model (1.1) so that the associated normal-reference scale-invariant tests can be obtained easily.

3.1. One-way MANOVA

The one-way MANOVA problem for high-dimensional data aims to check whether g high-dimensional samples have the same mean vector. It can be briefly described as follows. Suppose for each $i = 1,\ldots,g$, we have the following p-dimensional sample:

$$\mathbf{y}_{i1},\ldots,\mathbf{y}_{in_i} \text{ are i.i.d. with } \mathrm{E}(\mathbf{y}_{i1}) = \boldsymbol{\mu}_i \text{ and } \mathrm{Cov}(\mathbf{y}_{i1}) = \mathbf{\Sigma}, \tag{3.1}$$

and the above g samples are independent of each other. Of interest is to test the following hypotheses:

$$\mathrm{H}_0: \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_g, \quad \text{versus} \quad \mathrm{H}_1: \mathrm{H}_0 \text{ is false.} \tag{3.2}$$

The above one-way MANOVA problem can be re-written into the form of the GLHT problem (1.2) under the high-dimensional linear regression model (1.1). To re-write (3.2) into the form of the GLHT problem (1.2), we need only set $\boldsymbol{\Theta} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_g)^\top$ and $\mathbf{C} = (\mathbf{I}_q, -\mathbf{1}_q)$, with $q = g - 1$. Accordingly, the associated high-dimensional linear regression model (1.1) can be defined as follows. The sample size is $n = \sum_{i=1}^g n_i$, the response matrix $\mathbf{Y}$ and the design matrix $\mathbf{X}$ are given by $\mathbf{Y} = (\mathbf{y}_{11}, \ldots, \mathbf{y}_{1n_1}, \ldots, \mathbf{y}_{g1}, \ldots, \mathbf{y}_{gn_g})^\top$ and $\mathbf{X} = \mathrm{diag}(\mathbf{1}_{n_1}, \mathbf{1}_{n_2}, \ldots, \mathbf{1}_{n_g})$, a block diagonal matrix of size $n \times g$, respectively, while the measurement error matrix is defined as $\boldsymbol{\epsilon} = (\mathbf{y}_{11} - \boldsymbol{\mu}_1, \ldots, \mathbf{y}_{1n_1} - \boldsymbol{\mu}_1, \ldots, \mathbf{y}_{g1} - \boldsymbol{\mu}_g, \ldots, \mathbf{y}_{gn_g} - \boldsymbol{\mu}_g)^\top$.

With $\mathbf{Y}$, $\mathbf{X}$, and $\mathbf{C}$ as defined above, the associated variation matrices due to the hypothesis and error in (1.3) are then given by $\mathbf{S}_h = \sum_{i=1}^g n_i(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})^\top$ and $\mathbf{S}_e = \sum_{i=1}^g \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)^\top$, where $\bar{\mathbf{y}}_i = n_i^{-1} \sum_{j=1}^{n_i} \mathbf{y}_{ij}$, for $i = 1, \ldots, g$, and $\bar{\mathbf{y}} = n^{-1} \sum_{i=1}^g n_i \bar{\mathbf{y}}_i$. Thus,

$$\hat{\mathbf{R}} = \hat{\mathbf{D}}^{-1/2} \hat{\boldsymbol{\Sigma}} \hat{\mathbf{D}}^{-1/2}, \tag{3.3}$$

with $\hat{\boldsymbol{\Sigma}} = (n-g)^{-1}\mathbf{S}_e = (n-g)^{-1} \sum_{i=1}^g \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)^\top$ and $\hat{\mathbf{D}} = \mathrm{diag}(\hat{\boldsymbol{\Sigma}})$. It follows that the test statistic $T_{n,p}$ (1.6) can be simplified as $T_{n,p} = (n-g-2) \sum_{i=1}^g n_i (\bar{\mathbf{y}}_i - \bar{\mathbf{y}})^\top \hat{\mathbf{D}}^{-1} (\bar{\mathbf{y}}_i - \bar{\mathbf{y}}) / ((n-g)p(g-1))$. Note that a ratio-consistent estimator of $\widehat{\mathrm{tr}(\mathbf{R}^2)}$ is still given by (2.16), but with $\hat{\mathbf{R}}$ defined in (3.3). Accordingly, with the current $\widehat{\mathrm{tr}(\mathbf{R}^2)}$, a ratio-consistent estimator $\hat{d}$ of d is still given by (2.17), with $k = g$ and $q = g - 1$. The proposed normal-reference scale-invariant test with the W–S χ^2-approximation for the one-way MANOVA problem (3.2) can then be conducted accordingly, as described in Section 2.2.

3.2. Two-way MANOVA

The two-way MANOVA problem aims to test whether one of the main effects or the interaction effects of two factors are the same. It can be briefly defined as follows. Consider an experiment with two factors A and B, each having a and b levels, with a total of ab factorial combinations or cells. Suppose at the (i, j)th cell, we have the following p-dimensional sample:

$$\mathbf{y}_{ij1}, \ldots, \mathbf{y}_{ijn_{ij}} \text{ are i.i.d. with } \mathrm{E}(\mathbf{y}_{ij1}) = \boldsymbol{\mu}_{ij} \text{ and } \mathrm{Cov}(\mathbf{y}_{ij1}) = \boldsymbol{\Sigma}, \tag{3.4}$$

for $i = 1, \ldots, a; j = 1, \ldots, b$. All ab samples are independent of each other. In the two-way MANOVA problem, the cell mean vectors $\boldsymbol{\mu}_{ij}$ are usually decomposed into the form

$$\boldsymbol{\mu}_{ij} = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij}, \quad i = 1, \ldots, a; j = 1, \ldots, b, \tag{3.5}$$

where $\boldsymbol{\mu}_0$ is the grand mean vector, $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$ are the ith and the jth main effects of the factors A and B, respectively, and $\boldsymbol{\gamma}_{ij}$ is the (i,j)th interaction effect between factors A and B. Consider the following three null hypotheses:

$$\begin{aligned} \mathrm{H}_{0A} &: \boldsymbol{\alpha}_1 = \cdots = \boldsymbol{\alpha}_a, \\ \mathrm{H}_{0B} &: \boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_b, \\ \mathrm{H}_{0AB} &: \boldsymbol{\gamma}_{11} = \cdots = \boldsymbol{\gamma}_{1b} = \cdots = \boldsymbol{\gamma}_{a1} = \cdots = \boldsymbol{\gamma}_{ab}, \end{aligned} \tag{3.6}$$

against their usual alternative hypotheses that invalidate their associated null hypotheses. The first two hypotheses aim to test whether the main effects of the two factors are statistically significant, while the last one tests whether the interaction effect between the two factors is statistically significant.

Similarly, the above three two-way MANOVA problems can be re-written in the form of the GLHT problem (1.2) under the high-dimensional linear regression model (1.1). First, we express the three hypotheses (3.6) in the form of the GLHT problem (1.2). To this end, we set $\boldsymbol{\Theta} = (\boldsymbol{\mu}_{11}, \boldsymbol{\mu}_{12}, \ldots, \boldsymbol{\mu}_{ab})^\top$ and, based on Section 2 of Zhang (2011), the coefficient matrix $\mathbf{C}$ for the three hypotheses (3.6) can be set as $\mathbf{C}_a = \mathbf{H}_a\mathbf{A}_a$, $\mathbf{C}_b = \mathbf{H}_b\mathbf{A}_b$, and $\mathbf{C}_{ab} = \mathbf{H}_{ab}\mathbf{A}_{ab}$, with $\mathbf{H}_a = (\mathbf{I}_{a-1}, -\mathbf{1}_{a-1})$, $\mathbf{H}_b = (\mathbf{I}_{b-1}, -\mathbf{1}_{b-1})$, $\mathbf{H}_{ab} = \mathbf{H}_a \otimes \mathbf{H}_b$, and $\mathbf{A}_a = (\mathbf{I}_a - \mathbf{1}_a\mathbf{u}^\top) \otimes \mathbf{v}^\top$, $\mathbf{A}_b = \mathbf{u}^\top \otimes (\mathbf{I}_b - \mathbf{1}_b\mathbf{v}^\top)$, $\mathbf{A}_{ab} = (\mathbf{I}_a - \mathbf{1}_a\mathbf{u}^\top) \otimes (\mathbf{I}_b - \mathbf{1}_b\mathbf{v}^\top)$, where $\otimes$ is the Kronecker product operation, and $\mathbf{u} = (u_1, \ldots, u_a)^\top$ and $\mathbf{v} = (v_1, \ldots, v_b)^\top$ are two vectors used to impose constraints on the parameters $\boldsymbol{\mu}_0, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j$, and $\boldsymbol{\gamma}_{ij}$ in (3.5), because otherwise they are not uniquely defined (Zhang (2011)). There are two methods for specifying the weight vectors $\mathbf{u}$ and $\mathbf{v}$: the equal-weight method, which specifies $\mathbf{u}$ and $\mathbf{v}$ as $u_i = 1/a, v_j = 1/b$, for $i = 1, \ldots, a$ and $j = 1, \ldots, b$, and the size-adapted-weight method, which specifies $\mathbf{u}$ and $\mathbf{v}$ as $u_i = \sum_{j=1}^b n_{ij}/n$, for $i = 1, \ldots, a$, and $v_j = \sum_{i=1}^a n_{ij}/n$, for $j = 1, \ldots, b$. When the two-way MANOVA design is balanced, that is, when all the cell sizes n_{ij}, for $i = 1, \ldots, a$ and $j = 1, \ldots, b$, are the same, the size-adapted-weight method reduces to the equal-weight method. The matrices $\mathbf{C}_a, \mathbf{C}_b$, and $\mathbf{C}_{ab}$ are full rank, having ranks $a-1, b-1$, and $(a-1)(b-1)$ respectively. The two-way MANOVA hypotheses (3.6) can then be written in the form of the GLHT problem (1.2) using $\mathbf{C}_a$, $\mathbf{C}_b$, and $\mathbf{C}_{ab}$. Second, the associated high-dimensional linear regression model (1.1) can be defined as follows. The sample size is $n = \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}$. The observation matrix $\mathbf{Y}$ and the design matrix $\mathbf{X}$ are defined as $\mathbf{Y} = (\mathbf{y}_{111}, \ldots, \mathbf{y}_{11n_{11}}, \mathbf{y}_{121}, \ldots, \mathbf{y}_{12n_{12}}, \ldots, \mathbf{y}_{ab1}, \ldots, \mathbf{y}_{abn_{ab}})^\top$ and $\mathbf{X} = \mathrm{diag}(\mathbf{1}_{n_{11}}, \mathbf{1}_{n_{12}}, \ldots, \mathbf{1}_{n_{ab}})$, a block diagonal matrix of size $n \times (ab)$, respectively, while the measurement error matrix is $\boldsymbol{\epsilon} = (\mathbf{y}_{111} - \boldsymbol{\mu}_{11}, \ldots, \mathbf{y}_{11n_{11}} - \boldsymbol{\mu}_{11}, \mathbf{y}_{121} - \boldsymbol{\mu}_{12}, \ldots, \mathbf{y}_{12n_{12}} - \boldsymbol{\mu}_{12}, \ldots, \mathbf{y}_{ab1} - \boldsymbol{\mu}_{ab}, \ldots, \mathbf{y}_{abn_{ab}} - \boldsymbol{\mu}_{ab})^\top$.

With $\mathbf{Y}, \mathbf{X}$, and $\mathbf{C}$ (which can be $\mathbf{C}_a, \mathbf{C}_b$, or $\mathbf{C}_{ab}$) defined above, the variation matrices due to the hypothesis and error, as defined in (1.3), can be simplified as $\mathbf{S}_h = (\mathbf{C}\hat{\boldsymbol{\Theta}})^\top(\mathbf{C}\mathbf{W}_n\mathbf{C}^\top)^{-1}\mathbf{C}\hat{\boldsymbol{\Theta}}$ and $\mathbf{S}_e = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(\mathbf{y}_{ijk} - \hat{\boldsymbol{\mu}}_{ij})(\mathbf{y}_{ijk} - \hat{\boldsymbol{\mu}}_{ij})^\top$, where $\hat{\boldsymbol{\Theta}} = (\hat{\boldsymbol{\mu}}_{11}, \hat{\boldsymbol{\mu}}_{12}, \ldots, \hat{\boldsymbol{\mu}}_{ab})^\top$ with $\hat{\boldsymbol{\mu}}_{ij} = n_{ij}^{-1}\sum_{k=1}^{n_{ij}} \mathbf{y}_{ijk}$, for $i = 1, \ldots, a$ and $j = 1, \ldots, b$, and $\mathbf{W}_n = (\mathbf{X}^\top\mathbf{X})^{-1} = \mathrm{diag}(n_{11}^{-1}, \ldots, n_{ab}^{-1})$. Then, we have

$$\hat{\mathbf{R}} = \hat{\mathbf{D}}^{-1/2}\hat{\boldsymbol{\Sigma}}\hat{\mathbf{D}}^{-1/2}, \tag{3.7}$$

with $\hat{\boldsymbol{\Sigma}} = (n - ab)^{-1}\mathbf{S}_e$ and $\hat{\mathbf{D}} = \mathrm{diag}(\hat{\boldsymbol{\Sigma}})$. With the above, the test statistic $T_{n,p}$ can be simplified as $T_{n,p} = (n-k-2)/((n-k)pq)\mathrm{tr}[(\mathbf{C}\hat{\boldsymbol{\Theta}})^\top(\mathbf{C}\mathbf{W}_n\mathbf{C}^\top)^{-1}\mathbf{C}\hat{\boldsymbol{\Theta}}\hat{\mathbf{D}}^{-1}]$, where $k = ab$ and $q = \mathrm{rank}(\mathbf{C})$. Then, a ratio-consistent estimator of $\widehat{\mathrm{tr}(\mathbf{R}^2)}$ is still given by (2.16), but with $\hat{\mathbf{R}}$ defined in (3.7) and with $k = ab$. Accordingly, with the current $\widehat{\mathrm{tr}(\mathbf{R}^2)}$, a ratio-consistent estimator $\hat{d}$ of d is still given by (2.17) with the current q and $k = ab$. The proposed normal-reference scale-invariant test with the W–S χ^2-approximation for any of the two-way MANOVA problems (3.6) can then be conducted accordingly.

4. Simulation Studies

In this section, we conduct two simulation studies to evaluate the performance of the proposed test $T_{n,p}$ in terms of size control and power against several existing competitors for the GLHT problem (1.2).

Our simulation data are generated from the data structure specified in Condition C1. Given $\mathbf{X}$ and $\boldsymbol{\Theta}$, we generate $\mathbf{Y}$ by $\mathbf{Y} = \mathbf{X}\boldsymbol{\Theta} + (\mathbf{v}_1, \ldots, \mathbf{v}_n)^\top\boldsymbol{\Sigma}^{1/2}$, where $\mathbf{v}_i = (v_{i1}, \ldots, v_{ip})^\top$, for $i = 1, \ldots, n$, are i.i.d. random variables, with entries generated from the following three models:

Model 1: $v_{ir}, r = 1, \ldots, p \overset{i.i.d.}{\sim} \mathrm{N}(0, 1)$.

Model 2: $v_{ir} = z_{ir}/\sqrt{2}$, with $z_{ir}, r = 1, \ldots, p \overset{i.i.d.}{\sim} t_4$.

Model 3: $v_{ir} = (z_{ir} - 2)/2$, with $z_{ir}, r = 1, \ldots, p \overset{i.i.d.}{\sim} \chi_2^2$.

It is clear that the above three models generate the variables v_{ir} with three distributions: normal; nonnormal, but symmetric; and nonnormal and skewed. To specify the covariance matrix $\boldsymbol{\Sigma}$, we set $\boldsymbol{\Sigma} = \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2}$, where $\mathbf{D} = \mathrm{diag}(d_1^2, \ldots,$

d_p^2) with $d_i = (p - i + 1)/p$, for $i = 1, \ldots, p$, and $\mathbf{R} = (r_{ij}) : p \times p$ with $r_{ij} = (-1)^{i+j}\rho^{0.2|i-j|}$. Note that the value of the tuning parameter $\rho \in (0, 1)$ controls the correlation between the p-variables of the simulated data: the larger the value of ρ, the larger is the correlation between the p-variables. Note too that for $i, j \in \{1, \ldots, p\}$, r_{ij} decreases with increasing $|i-j|$. The other tuning parameters are specified as follows. We set $p \in \{200, 500, 1000\}$ and $\rho \in \{0.01, 0.55, 0.95\}$ so that the p-variables of the simulated data are nearly uncorrelated, moderately correlated, and highly correlated. To measure the overall performance of the tests in terms of size control, we adopt the average relative error (Zhang (2011)): $\mathrm{ARE} = 100M^{-1}\sum_{j=1}^{M} |\hat{\alpha}_j - \alpha|/\alpha$, where $\hat{\alpha}_j$, for $j = 1, \ldots, M$, are the empirical sizes under consideration. Throughout, the nominal size is $\alpha = 5\%$ and the number of replications is 10,000.

4.1. Simulation 1

In this simulation, we aim to evaluate the performance of $T_{n,p}$ against the non-scale-invariant tests developed by Fujikoshi, Himeno and Wakaki (2004) and Srivastava and Fujikoshi (2006), and against one scale-invariant test studied by Yamada and Srivastava (2012) and Srivastava and Kubokawa (2013), denoted as T_{FHW}, T_{SF}, and T_{YS}, respectively, for the one-way MANOVA problem described in (3.1) and (3.2) in Section 3.1. We also include an alternative version of T_{YS}, denoted as T_{YS}^*, obtained by removing the adjustment coefficient $c_{n,p}$ from the definition (1.4) of T_{YS}, that is,

$$T_{\mathrm{YS}} = \frac{T_{\mathrm{YS}}^*}{\sqrt{c_{n,p}}}, \tag{4.1}$$

which assesses the impact of $c_{n,p}$ on the performance of T_{YS}. For simplicity, consider $g = 3$ and set $n_1 = 0.8n_0, n_2 = n_0$, and $n_3 = 1.2n_0$, with $n_0 \in \{80, 100, 120\}$. Furthermore, set $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \mathbf{0}$ for the null hypothesis, and $\boldsymbol{\mu}_1 = \mathbf{0}$, $\boldsymbol{\mu}_2 = \delta\mathbf{h}$, and $\boldsymbol{\mu}_3 = 1.5\delta\mathbf{h}$ for the alternative hypothesis, where the tuning parameter δ controls the difference between $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and $\boldsymbol{\mu}_3$, while $\mathbf{h} = \mathbf{h}_0/\sqrt{\mathbf{h}_0^\top \mathbf{h}_0}$, with $\mathbf{h}_0 = (1, \ldots, p)^\top$. The values of δ are chosen as $0.02, 0.04$, and 0.06, respectively for $\rho = 0.01, 0.55$, and 0.95.

Table 1 displays the empirical sizes of T_{FHW}, T_{SF}, T_{YS}, T_{YS}^*, and $T_{n,p}$ with the associated ARE values listed in the last row. In terms of size control, $T_{n,p}$ performs well in all the configurations, and outperforms the other four tests, especially when $\rho = 0.55$ and 0.95. When $\rho = 0.01$, all five tests perform reasonably well. However, when $\rho = 0.55$ and 0.95, in terms of size control, T_{FHW}, T_{SF}, and

Table 1. Empirical sizes (in %, Simulation 1).

			$\rho = 0.01$					$\rho = 0.55$					$\rho = 0.95$				
Model	p	n_0	T_{FHW}	T_{SF}	T_{YS}	T^*_{YS}	$T_{n,p}$	T_{FHW}	T_{SF}	T_{YS}	T^*_{YS}	$T_{n,p}$	T_{FHW}	T_{SF}	T_{YS}	T^*_{YS}	$T_{n,p}$
1	200	80	5.69	5.72	5.10	6.24	5.75	6.38	6.39	3.17	6.48	5.48	7.57	7.58	0.74	7.66	5.75
		100	5.78	5.82	4.93	6.04	5.41	6.69	6.70	3.45	6.71	5.64	7.04	7.05	0.56	6.77	5.20
		120	5.68	5.71	4.58	5.73	5.15	6.42	6.42	3.16	6.17	5.32	7.74	7.74	0.67	7.25	5.58
	500	80	5.35	5.40	4.34	5.59	5.19	6.10	6.16	3.18	5.84	5.22	6.93	6.96	0.57	6.97	5.44
		100	5.18	5.21	4.15	5.17	4.85	6.06	6.06	3.40	6.19	5.43	6.73	6.77	0.47	6.67	5.14
		120	5.65	5.71	4.62	5.61	5.18	6.60	6.62	3.62	6.31	5.62	7.15	7.21	0.70	7.23	5.94
	1,000	80	5.48	5.52	4.40	5.57	5.45	6.24	6.28	3.73	6.20	5.69	6.92	6.95	0.59	7.14	5.76
		100	5.70	5.71	4.65	5.81	5.62	5.92	5.97	3.51	5.96	5.48	6.91	6.93	0.57	7.00	5.68
		120	5.31	5.34	4.21	4.98	4.81	6.02	6.03	3.80	6.14	5.68	6.98	6.99	0.69	7.21	5.86
2	200	80	5.46	5.50	4.89	5.97	5.52	6.61	6.62	3.23	6.52	5.49	7.30	7.32	0.63	7.63	5.85
		100	5.58	5.59	4.75	5.88	5.27	6.68	6.71	3.34	6.84	5.76	7.37	7.38	0.64	7.26	5.35
		120	5.54	5.55	4.44	5.56	5.00	6.81	6.84	3.15	6.31	5.43	7.45	7.45	0.53	7.30	5.58
	500	80	4.84	4.87	4.38	5.55	5.25	6.43	6.47	3.24	5.93	5.26	7.05	7.06	0.70	6.87	5.44
		100	5.43	5.46	4.77	5.76	5.46	5.78	5.78	3.48	5.93	5.26	6.99	7.01	0.64	7.07	5.79
		120	5.09	5.11	4.43	5.37	4.94	6.59	6.62	3.57	5.97	5.43	7.11	7.15	0.49	7.00	5.33
	1,000	80	5.00	5.07	3.92	5.39	5.13	5.95	6.00	3.44	5.77	5.38	6.76	6.78	0.63	6.87	5.66
		100	5.11	5.12	4.79	5.98	5.67	5.90	5.94	4.00	6.52	5.95	7.04	7.08	0.64	7.01	5.55
		120	5.23	5.25	4.09	5.10	4.82	5.65	5.69	3.24	5.24	4.86	7.30	7.31	0.70	6.74	5.45
3	200	80	5.33	5.36	4.56	5.66	5.20	6.88	6.90	3.31	6.80	5.74	6.59	6.61	0.56	6.60	4.99
		100	5.59	5.60	5.00	6.17	5.41	6.63	6.67	3.63	7.01	5.85	7.23	7.25	0.61	7.19	5.32
		120	5.51	5.52	4.51	5.49	4.93	6.89	6.91	3.54	6.73	5.75	7.13	7.17	0.49	7.43	5.56
	500	80	5.75	5.79	4.44	5.77	5.51	6.17	6.20	3.53	6.05	5.39	7.51	7.54	0.58	7.01	5.59
		100	5.10	5.12	4.38	5.35	5.01	6.19	6.24	3.55	6.30	5.50	6.95	6.97	0.64	7.00	5.69
		120	5.58	5.59	4.69	5.85	5.42	6.59	6.63	3.57	6.32	5.66	6.89	6.90	0.70	6.83	5.42
	1,000	80	5.37	5.44	4.38	5.79	5.60	5.63	5.65	3.35	5.62	5.12	7.16	7.19	0.86	7.08	5.94
		100	5.29	5.33	4.48	5.42	5.20	5.69	5.74	3.26	5.62	5.19	6.78	6.84	0.76	6.54	5.31
		120	5.51	5.54	4.55	5.67	5.46	6.00	6.02	3.81	6.20	5.71	6.72	6.76	0.54	7.15	5.67
ARE			**8.48**	**9.04**	**9.46**	**12.97**	**6.30**	**25.56**	**26.12**	**30.92**	**24.21**	**10.05**	**41.70**	**42.19**	**87.48**	**41.10**	**11.01**

T^*_{YS} are all rather liberal, while T_{YS} is very conservative. In particular, when $\rho = 0.95$, the empirical sizes of T_{YS} are less than 1%, which are unacceptable. Therefore, $T_{\text{FHW}}, T_{\text{SF}}, T_{\text{YS}}$, and T^*_{YS} are less or no longer applicable when the data are moderately or highly correlated, but that is not the case for $T_{n,p}$.

To explore, in terms of size control, why T_{FHW}, T_{SF}, T_{YS}, and T^*_{YS} perform well when $\rho = 0.01$, but perform much worse when $\rho = 0.55$ and 0.95, we list the values of $\hat{d}$ and $c_{n,p}$ under Model 1 in Table 2. Because the values of $\hat{d}$ and $c_{n,p}$ do not depend on the model, we do not present their values under Models 2 and 3 for in order to conserve space. When $\rho = 0.01$, the values of $\hat{d}$ are large ($296 \sim 1{,}494$) and the values of $c_{n,p}$ are small ($1.13 \sim 1.18$), showing that the underlying distributions of the five tests are nearly normal. In this case, the normal approximations to the associated underlying null distributions are adequate, so that all five tests perform well in terms of size control. However,

Table 2. Values of $\hat{d}$ and $c_{n,p}$ under Model 1 (Simulation 1)

		$\rho = 0.01$		$\rho = 0.55$		$\rho = 0.95$	
p	n_0	$\hat{d}$	$c_{n,p}$	$\hat{d}$	$c_{n,p}$	$\hat{d}$	$c_{n,p}$
	80	299.10	1.16	50.11	1.64	5.58	6.25
200	100	297.44	1.14	49.79	1.63	5.54	6.25
	120	296.37	1.14	49.58	1.62	5.51	6.26
	80	747.16	1.16	123.62	1.46	11.74	4.99
500	100	742.89	1.14	122.87	1.45	11.66	4.99
	120	740.17	1.12	122.38	1.43	11.61	4.97
	80	1493.57	1.18	246.17	1.40	22.23	4.04
1,000	100	1485.20	1.15	244.73	1.37	22.10	4.02
	120	1479.88	1.13	243.76	1.35	22.01	4.01

when $\rho = 0.55$ and 0.95, the values of $\hat{d}$ are small or moderate ($5 \sim 247$), but the values of $c_{n,p}$ are large ($1.35 \sim 6.26$), showing that the underlying distributions of the five tests are less or not normal. In these cases, the normal approximations to the associated underlying null distributions are less or not adequate. This explains why $T_{\text{FHW}}, T_{\text{SF}}$, T_{YS}, and T^*_{YS} do not perform well. In particular, the poor performance of T_{YS} when $\rho = 0.55$ and 0.95 is also due to the blind application of the adjustment coefficient $c_{n,p}$. This can be seen clearly by comparing the empirical sizes of T_{YS} and T^*_{YS} under various settings.

Now, we compare the empirical powers of the five tests. Figure 1 displays the empirical powers of the five tests under consideration (T_{FHW}: black solid curves with triangles, T_{SF}: red dashed curves with diamonds, T_{YS}: green dotted curves with squares, T^*_{YS}: orange long-dashed curves with crosses, and $T_{n,p}$: blue dot-dashed curves with circles). It is seen that the two non-scale-invariant tests T_{FHW} and T_{SF} have almost no power, while the three scale-invariant tests $T_{\text{YS}}, T^*_{\text{YS}}$, and $T_{n,p}$ have nontrivial power under various configurations. This is because the diagonal entries of $\mathbf{\Sigma}$ are not equal, and the two scale-invariant tests do not take this information into account, while the three scale-invariant tests do. In addition, when $\rho = 0.01$, the empirical powers of $T_{\text{YS}}, T^*_{\text{YS}}$, and $T_{n,p}$ are, in general, comparable, because their empirical sizes are also comparable, in gerenal. Furthermore, when $\rho = 0.55$ and 0.95, the empirical powers of T^*_{YS} are slightly (res., much) larger than those of $T_{n,p}$ (res., T_{YS}) because the empirical sizes of $T^*_{n,p}$ are also slightly (res., much) larger than those of $T_{n,p}$ (res., T_{YS}). Thus, when the underlying null distributions of T^*_{YS} and T_{YS} are actually not normal, the empirical size and power of T^*_{YS} may be artificially enlarged by the blind application of the normal approximation, while that of T_{YS} may be substantially reduced by the

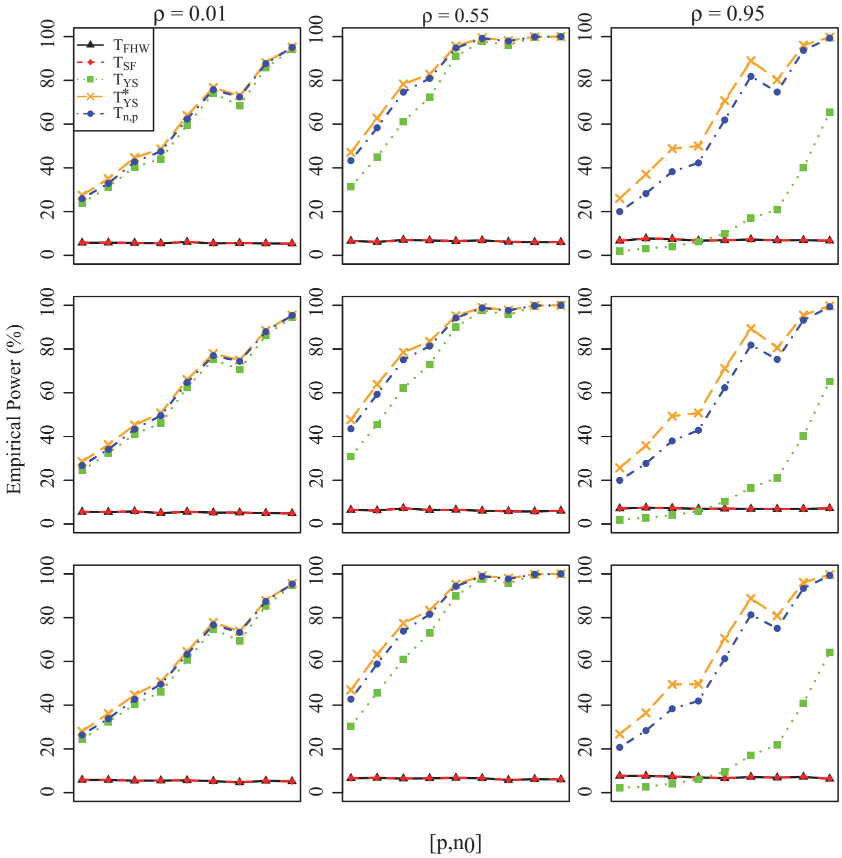

Figure 1. Empirical power (in %, Simulation 1) of the five tests (T_{FHW}: solid curves with triangles, T_{SF}: dashed curves with diamonds, T_{YS}: dotted curves with squares, T^*_{YS}: long-dashed curves with crosses, and $T_{n,p}$: dot-dashed curves with circles) associated with parameters $[p, n_0]$ from the settings under Model 1 (first row), Model 2 (second row), and Model 3 (third row).

blind use of the adjustment coefficient $c_{n,p}$. A theoretical explanation of the effect of $c_{n,p}$ on T_{YS} can be found in Section S5 of the Supplementary Material.

4.2. Simulation 2

In this simulation study, we continue to compare the performance of $T_{n,p}$ against $T_{\text{FHW}}, T_{\text{SF}}$, T_{YS}, and T^*_{YS} for the two-way MANOVA problem, as described in (3.4) and (3.6) in Section 3.2. To this end, we set $a = 2, b = 3$, and $\mathbf{n} = (n_{11}, n_{12}, \ldots, n_{ab}) \in \{\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3\}$, with $\mathbf{n}_1 = (30, 45)_3$, $\mathbf{n}_2 = (40, 60)_3$, and $\mathbf{n}_3 = (50, 75)_3$, where $\mathbf{v}_r$ denotes the vector obtained by repeating the vector $\mathbf{v}$ r times. For example, $(30, 45)_3 = (30, 45, 30, 45, 30, 45)$. For the null hypothesis,

Table 3. Empirical sizes (in %, Simulation 2) for testing the interaction effect.

			$\rho = 0.01$					$\rho = 0.55$					$\rho = 0.95$				
Model	p	**n**	T_{FHW}	T_{SF}	T_{YS}	T^*_{YS}	$T_{n,p}$	T_{FHW}	T_{SF}	T_{YS}	T^*_{YS}	$T_{n,p}$	T_{FHW}	T_{SF}	T_{YS}	T^*_{YS}	$T_{n,p}$
1	200	80	5.87	5.88	4.66	6.09	5.29	6.69	6.71	3.30	6.46	5.56	7.03	7.06	0.62	6.83	5.05
		100	5.93	5.96	4.65	5.75	5.20	6.55	6.58	3.60	7.07	6.02	6.88	6.89	0.61	7.01	5.46
		120	5.61	5.64	4.84	5.77	5.35	6.94	6.96	3.30	6.39	5.54	7.32	7.36	0.35	7.13	5.27
	500	80	5.66	5.73	4.25	5.46	5.13	6.49	6.53	3.28	6.31	5.62	6.84	6.87	0.53	7.01	5.48
		100	5.96	5.98	4.84	5.88	5.47	6.23	6.28	3.69	6.33	5.72	7.09	7.09	0.59	6.91	5.32
		120	5.66	5.70	4.77	5.66	5.36	6.16	6.18	3.64	6.49	5.80	7.07	7.11	0.56	7.06	5.61
	1,000	80	5.08	5.11	4.19	5.76	5.52	5.88	5.92	3.40	5.74	5.31	7.24	7.28	0.81	7.17	5.98
		100	4.93	4.98	3.95	5.15	4.73	5.66	5.67	3.73	6.19	5.63	6.63	6.67	0.71	6.75	5.54
		120	5.28	5.30	4.61	5.52	5.27	5.98	6.02	3.55	5.65	5.17	6.88	6.91	0.57	6.67	5.51
2	200	80	5.19	5.20	4.66	5.83	5.23	6.75	6.79	3.63	7.19	6.24	6.94	6.97	0.40	6.94	5.20
		100	5.55	5.56	4.72	5.67	5.11	6.59	6.60	3.24	6.51	5.48	7.04	7.06	0.63	7.19	5.42
		120	5.67	5.68	4.41	5.37	4.99	6.54	6.56	3.44	6.59	5.80	7.21	7.24	0.60	6.86	5.32
	500	80	5.41	5.47	4.45	5.84	5.42	6.51	6.52	3.45	6.11	5.57	7.36	7.39	0.68	7.51	5.98
		100	5.38	5.39	4.46	5.46	5.05	6.11	6.11	3.42	6.33	5.55	6.60	6.62	0.69	6.80	5.22
		120	5.23	5.24	4.80	5.97	5.49	5.90	5.94	3.24	5.71	5.12	7.21	7.22	0.60	7.21	5.64
	1,000	80	5.14	5.20	4.05	5.37	5.18	6.23	6.26	3.47	5.87	5.35	7.21	7.22	0.67	7.33	6.03
		100	5.24	5.27	4.52	5.73	5.42	5.90	5.94	3.61	5.95	5.48	6.55	6.57	0.54	6.53	5.36
		120	5.51	5.56	4.67	5.68	5.42	6.20	6.22	3.89	6.22	5.72	7.13	7.15	0.43	6.92	5.66
3	200	80	5.84	5.86	4.51	5.55	4.99	6.73	6.75	3.54	7.23	6.23	7.22	7.24	0.55	7.32	5.48
		100	5.74	5.75	5.07	6.25	5.61	6.61	6.67	3.45	6.65	5.79	7.46	7.46	0.72	7.05	5.57
		120	5.33	5.34	4.40	5.41	4.98	6.43	6.45	3.35	6.41	5.63	7.61	7.62	0.66	7.28	5.49
	500	80	5.11	5.16	4.14	5.33	5.08	5.98	6.02	3.28	6.26	5.69	7.08	7.11	0.70	7.42	5.62
		100	5.37	5.40	4.54	5.68	5.19	6.28	6.33	3.43	6.01	5.28	7.00	7.02	0.65	7.16	5.82
		120	5.56	5.59	4.90	6.02	5.60	6.31	6.33	3.42	6.16	5.45	6.82	6.83	0.60	7.14	5.79
	1,000	80	5.23	5.28	4.12	5.49	5.23	6.12	6.18	3.70	6.00	5.55	7.60	7.63	0.65	6.75	5.57
		100	5.43	5.47	4.21	5.19	4.99	6.15	6.17	3.46	5.72	5.30	6.88	6.90	0.80	6.60	5.58
		120	5.45	5.47	4.55	5.63	5.30	6.26	6.28	4.17	6.20	5.69	7.11	7.12	0.89	6.84	5.70
ARE			**9.26**	**9.79**	**9.78**	**12.97**	**5.36**	**26.06**	**26.64**	**29.87**	**25.74**	**12.07**	**41.49**	**41.93**	**87.55**	**40.29**	**10.87**

$\boldsymbol{\mu}_{ij} = \mathbf{0}$, for $i = 1, \ldots, a$ and $j = 1, \ldots, b$, while for the alternative hypothesis, $\boldsymbol{\mu}_{ij} = ij\delta\mathbf{h}/(ab)$, for $i = 1, \ldots, a$ and $j = 1, \ldots, b$, where $\mathbf{h}$ is defined as in the previous simulations. The values of δ are set as follows. To test H_{0A} in (3.6), $\delta = 0.23, 0.35, 0.65$ respectively for $\rho = 0.01, 0.55, 0.95$; to test H_{0B} in (3.6), $\delta = 0.04, 0.06, 0.11$ respectively for $\rho = 0.01, 0.55, 0.95$; and to test H_{0AB} in (3.6), $\delta = 0.035, 0.055, 0.10$ respectively for $\rho = 0.01, 0.55, 0.95$. Finally, to specify the C-matrices for the three hypotheses (3.6), the size-adapted-weight method described in Section 3.2 is employed.

To save space, we report only the empirical sizes of $T_{\text{FHW}}, T_{\text{SF}}$, T_{YS}, T^*_{YS}, and $T_{n,p}$ when testing the interaction effects between the two factors in Table 3. In terms of size control, $T_{n,p}$ again performs well and, in general, outperforms the other four tests, especially when $\rho = 0.55$ and 0.95. The other four tests perform well when $\rho = 0.01$, but not when $\rho = 0.55$ and 0.95. In particular, when $\rho = 0.95$,

Table 4. One-way MANOVA for the corneal surface data.

method	statistic	p-value	$\hat{d}$	$c_{n,p}$
T_{FHW}	6.40	7.69×10^{-11}	-	-
T_{SF}	6.42	6.68×10^{-11}	-	-
T_{YS}	2.35	0.01	-	16.78
T^*_{YS}	9.64	2.82×10^{-22}	-	-
$T_{n,p}$	5.57	1.08×10^{-7}	8.97	-

T_{YS} performs poorly. These conclusions are similar to those drawn from Table 1.

Based on the above simulation studies, in terms of size control and power, our test $T_{n,p}$ outperforms its competitors, in general, and is recommended, regardless of whether the data are nearly uncorrelated, moderately correlated, or highly correlated.

5. Application to the corneal surface data set

We now apply $T_{n,p}$, $T_{\text{FHW}}, T_{\text{SF}}$, T_{YS}, and T^*_{YS}, to the corneal surface data set introduced in Section 1, which contains four cornea groups. The first is a normal cornea group with 43 healthy corneas. The other three nonnormal cornea groups are unilateral suspect, suspect map, and clinical keratoconus groups, consisting of $14, 21$, and 72 corneas, respectively, with varying degrees of keratoconus, a disease that misshapes the cornea.

In the keratoconus study that yielded the above data set, each corneal surface has 6,912 measurements taken at a polar grid of 256×27 points, with 256 radial directions and 27 locations at each radial. Because not all 6,912 measurements for each corneal surface were obtained, the measurements were often subject to measurement errors, and the polar grids may be different for different corneas. Therefore, before further analysis, the corneal surfaces were reconstructed using a Legendre–Fourier basis system (Locantore et al. (1999)), and were evaluated at a given common polar grid of 100×20 points, resulting in 150 vectors of length 2,000, which represent well the corneal surface data set.

To apply $T_{n,p}$, $T_{\text{FHW}}, T_{\text{SF}}$, T_{YS}, and T^*_{YS}, to check whether the four corneal surface groups have the same mean corneal surface, we first check the equality of the covariance matrices of the four groups using the test proposed by Srivastava and Yanagihara (2010). The resulting p-value is nearly one, showing that it may be reasonable to assume that the four groups of the corneal surface data set have the same covariance matrix.

Table 4 displays the test results of the one-way MANOVA for the corneal

Table 5. P-values of some contrast tests for the corneal surface data.

hypothesis	T_{FHW}	T_{SF}	T_{YS}	T^*_{YS}	$T_{n,p}$
Nor vs Uni	0.75	0.75	0.54	0.67	0.59
Nor vs Sus	0.27	0.27	0.51	0.52	0.41
Uni vs Sus	0.73	0.73	0.55	0.69	0.62
Uni vs Cli	0.04	0.04	0.26	3.5×10^{-3}	0.02
Sus vs Cli	1.82×10^{-14}	1.50×10^{-14}	0.01	1.47×10^{-21}	8.49×10^{-6}
Nor vs Cli	3.73×10^{-13}	3.12×10^{-13}	1.14×10^{-3}	3.51×10^{-36}	2.29×10^{-7}

surface data using T_{FHW}, T_{SF}, T_{YS}, T^*_{YS}, and $T_{n,p}$. Although all five tests strongly reject the null hypothesis, their p-values are very different: the p-value of T^*_{YS} is the smallest, followed by those of T_{FHW}, T_{SF}, and $T_{n,p}$; the p-value of T_{YS} is the largest. Because $\hat{d} = 8.97$ is very small and $c_{n,p} = 16.78$ is very large, the normal approximation to the underlying null distributions of these test statistics is unlikely to be adequate. Thus, only the p-value of $T_{n,p}$ is reliable, as suggested by the simulation results presented in Section 4.

Because the above one-way MANOVA test is highly significant, it is now of interest to consider some post hoc or contrast tests for the corneal surface data. Here, several contrast tests are considered to check whether any two cornea groups have different mean corneal surfaces. The test results of these contrast tests using T_{FHW}, T_{SF}, T_{YS}, T^*_{YS}, and $T_{n,p}$ are displayed in Table 5, where the normal, unilateral suspect, suspect map, and clinical keratoconus cornea groups are labeled as "Nor," "Uni," "Sus," and "Cli," respectively. All five tests suggest that the contrast tests "Nor vs. Uni," "Nor vs. Sus," and "Uni vs. Sus" are not significant, while the contrast tests "Sus vs. Cli" and "Nor vs. Cli" are highly significant. However, for the contrast test "Uni vs. Cli," T_{YS} fails to reject the null hypothesis at the 5% significance level, whereas the other four tests do reject the hypothesis. Thus, all five tests, except T_{YS}, give largely consistent conclusions about these contrast tests. Nevertheless, their p-values for these contrast tests can be substantially different. For example, for the contrast tests "Sus vs. Cli" and "Nor vs. Cli," the p-values of $T_{\text{FHW}}, T_{\text{SF}}$, and T^*_{YS} are 1,000,000 times smaller than those of $T_{n,p}$. Fortunately, it is again very easy to conclude that the p-values of $T_{n,p}$ for these contrast tests are more trustworthy than those of the other four tests. This is because $\hat{d} = 2.99$ is very small and $c_{n,p} = 16.78$ is much larger than one, showing that the normal approximation to the underlying null distributions of the other four tests are unlikely to be adequate.

6. Conclusion

In this paper, we have proposed and studied a normal-reference scale-invariant test for the GLHT problem in high-dimensional linear regression, where the dimension of the data can be much larger than the total sample size. Simulation studies and a real-data example demonstrate that under some mild conditions, the proposed test performs well, regardless of whether the data are nearly uncorrelated, moderately correlated, or highly correlated, and it has much better size control and power than several existing competitors. Therefore, it is recommended for real-data applications. Admittedly, the good performance of the proposed test requires the condition "$\log(p) = o(n)$" be satisfied, while for a non-scale-invariant test, this condition is not required. This implies that the proposed test requires a relatively larger sample size to work well than a non-scale-invariant test does, as demonstrated by the simulation results presented in Section S2 of the Supplementary Material. Thus, when the sample size n is too small compared with $\log(p)$, a non-scale-invariant test can outperform the proposed test in terms of size control.

Supplementary Material

The online Supplementary Material provides histograms of simulated T_{YS}, several additional simulation studies, a discussion about some asymptotic properties of the adjustment coefficient $c_{n,p}$, and technical proofs of the main results.

Acknowledgments

The work was supported by the National University of Singapore academic research grant R-155-000-187-114. The authors thank the co-editor, associate editor and two anonymous reviewers for their constructive comments and helpful suggestions.

References

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

Fujikoshi, Y., Himeno, T. and Wakaki, H. (2004). Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large compared to the sample size. *Journal of the Japan Statistical Society* **34**, 19–26.

Li, H., Aue, A. and Paul, D. (2020). High-dimensional general linear hypothesis tests via non-linear spectral shrinkage. *Bernoulli* **26**, 2541–2571.

Locantore, N., Marron, J., Simpson, D., Tripoli, N., Zhang, J., Cohen, K. et al. (1999). Robust principal component analysis for functional data. *Test* **8**, 1–73.

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

Srivastava, M. S. (2007). Multivariate theory for analyzing high dimensional data. *Journal of the Japan Statistical Society* **37**, 53–86.

Srivastava, M. S. and Fujikoshi, Y. (2006). Multivariate analysis of variance with fewer observations than the dimension. *Journal of Multivariate Analysis* **97**, 1927–1940.

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

Srivastava, M. S. and Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis* **101**, 1319–1329.

Takeda, J. (1999). Comparison of highly multivariate small samples–estimation of effective dimensionality in the Dempster's approximation test. *Japanese Journal of Applied Statistics* **28**, 163–177.

Yamada, T. and Srivastava, M. S. (2012). A test for multivariate analysis of variance in high dimension. *Communications in Statistics-Theory and Methods* **41**, 2602–2615.

Zhang, J.-T. (2011). Two-way MANOVA with unequal cell sizes and unequal cell covariance matrices. *Technometrics* **53**, 426–439.

Zhang, J.-T., Guo, J. and Zhou, B. (2017). Linear hypothesis testing in high-dimensional one-way MANOVA. *Journal of Mulitvariate Analysis* **155**, 200–216.

Zhang, J.-T., Guo, J., Zhou, B. and Cheng, M.-Y. (2020). A simple two-sample test in high dimensions based on L^2-norm. *Journal of American Statistical Association* **115**, 1011–1027.

Zhang, L., Zhu, T. and Zhang, J.-T. (2020). A simple scale-invariant two-sample test for high-dimensional data. *Econometrics and Statistics* **14**, 131–144.

Tianming Zhu

Department of Statistics and Data Science, National University of Singapore, Singapore 117546.

E-mail: tianming_zhu@nus.edu.sg

Liang Zhang

Department of Statistics and Data Science, National University of Singapore, Singapore 117546.

E-mail: liangz@u.nus.edu

Jin-Ting Zhang

Department of Statistics and Data Science, National University of Singapore, Singapore 117546.

E-mail: stazjt2020@nus.edu.sg

(Received April 2020; accepted February 2021)