Received: 18 October 2017 | Accepted: 14 September 2018

DOI: 10.1111/biom.12984

ORIGINAL ARTICLE

WILEY Biometrics

Diagonal likelihood ratio test for equality of mean vectors in high-dimensional data

Zongliang Hu[1] | **Tiejun Tong**[2] | **Marc G. Genton**[3]

[1]College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China

[2]Department of Mathematics, Hong Kong Baptist University, Hong Kong

[3]Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

Correspondence
Tiejun Tong, Department of Mathematics, Hong Kong Baptist University, Hong Kong
Email: tongt@hkbu.edu.hk

Abstract

We propose a likelihood ratio test framework for testing normal mean vectors in high-dimensional data under two common scenarios: the one-sample test and the two-sample test with equal covariance matrices. We derive the test statistics under the assumption that the covariance matrices follow a diagonal matrix structure. In comparison with the diagonal Hotelling's tests, our proposed test statistics display some interesting characteristics. In particular, they are a summation of the log-transformed squared t-statistics rather than a direct summation of those components. More importantly, to derive the asymptotic normality of our test statistics under the null and local alternative hypotheses, we do not need the requirement that the covariance matrices follow a diagonal matrix structure. As a consequence, our proposed test methods are very flexible and readily applicable in practice. Simulation studies and a real data analysis are also carried out to demonstrate the advantages of our likelihood ratio test methods.

KEYWORDS

high-dimensional data, Hotelling's test, likelihood ratio test, log-transformed squared t-statistic, statistical power, type I error

1 | INTRODUCTION

In high-dimensional data analysis, it is often necessary to test whether a mean vector is equal to another vector in the one-sample case, or to test whether two mean vectors are equal to each other in the two-sample case. One such example is to test whether two gene sets, or pathways, have equal expression levels under two different experimental conditions. Given the two normal random samples, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n_1} \in \mathbb{R}^p$ and $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n_2} \in \mathbb{R}^p$, one well-known method for testing whether their mean vectors are equal is Hotelling's T^2 test,

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}})^T S^{-1}(\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}}), \tag{1}$$

where $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{Y}}$ are the sample mean vectors and S is the pooled sample covariance matrix. Hotelling's T^2 test is well-behaved and has been extensively studied in the classical low-dimensional setting. However, this classic test may not perform well or may not even be applicable to high-dimensional data with a small sample size. Specifically, it suffers from the singularity problem because the sample covariance matrix S is singular when the dimension is larger than the sample size.

To overcome the singularity problem in Hotelling's T^2 test, Bai and Saranadasa (1996) replaced the sample covariance matrix in (1) with the identity matrix, so that their test statistic is essentially the same as $(\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}})^T(\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}})$. Following their method, Chen and Qin (2010) and Ahmad (2014) proposed some U-statistics for testing whether two mean vectors are equal. These test methods were referred to as the *unscaled Hotelling's tests* in Dong et al. (2016) As an alternative, Chen et al. (2011) and Li et al. (2016) proposed replacing the inverse sample covariance matrix S^{-1} with the regularized estimator $(S + \lambda I_p)^{-1}$ in Hotelling's test statistic, where I_p is the identity matrix and $\lambda > 0$ is a regularization

parameter. Lopes et al. (2011) proposed a random projection technique to estimate the sample covariance matrix. Specifically, they replaced S^{-1} in Hotelling's test statistic with $E_R^{-1}\{R(R^TSR)^{-1}R^T\}$, where R is a random matrix of size $p \times k$ and $E_R(\cdot)$ is the expectation operator over the distribution. The random projection technique was further explored by, for example, Thulin (2014), Srivastava (2016), and Wei et al. (2016). Dong et al. (2016) referred to the test methods in this category as the *regularized Hotelling's tests*.

In addition to the aforementioned methods, replacing the sample covariance matrix with a diagonal sample covariance matrix is another popular approach to improving Hotelling's T^2 test. In particular, Wu et al. (2006), Srivastava and Du (2008), and Srivastava (2009) considered the *diagonal Hotelling's test statistic*:

$$T^2_{\text{diag},2} = \frac{n_1 n_2}{n_1 + n_2}(\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}})^T\{\text{diag}(S)\}^{-1}(\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}}). \quad (2)$$

Recently, Srivastava et al. (2013), Feng et al. (2015), and Gregory et al. (2015) also considered the diagonal Hotelling's tests under the assumption of unequal covariance matrices. Their test statistics essentially follow the diagonal structure in (2), $(\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}})^T\{\text{diag}(S_1)/n_1 + \text{diag}(S_2)/n_2\}^{-1}(\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}})$, where S_1 and S_2 are two sample covariance matrices. Park and Park and Ayyala (2013) modified the diagonal Hotelling's test statistic (2) based on the idea of leave-out cross-validation. Dong et al. (2016) proposed a shrinkage-based Hotelling's test that replaced the diagonal elements of the sample covariance matrix in (2) with some improved variance estimates. To summarize, the diagonal Hotelling's tests are popular in practice for several reasons. First, since a diagonal matrix is always invertible for nonzero variance estimates, the singularity problem in the classic test is circumvented. Second, the diagonal Hotelling's tests are scale transformation invariant tests. As suggested by Park and Ayyala (2013), the scale transformation invariant tests usually provide a better performance than the orthogonal transformation invariant tests including the unscaled Hotelling's tests and the regularized Hotelling's tests, especially when the variances of the signal components are small and the variances of the noise components are large. Last but not least, a diagonal covariance matrix assumption is also popular in the high-dimensional literature, for example, in Dudoit et al. (2002), Bickel and Levina (2004), and Huang et al. (2010)

Note that Hotelling's test statistic originated from the likelihood ratio test in the classical setting when p is smaller than n. Recently, researchers have also applied the likelihood ratio test method to analyze high-dimensional data. For instance, Jiang and Yang (2013) and Jiang and Qi (2015) tested mean vectors and covariance matrices of normal distributions using the likelihood ratio test method, under the setting that p is smaller than n but in a way that allows $p/n \to 1$. Zhao and Xu (2016) proposed a generalized high-dimensional likelihood ratio test for the normal mean vector by a modified union-intersection method. Städler and Mukherjee (2017) provided a high-dimensional likelihood ratio test for the two-sample test based on sample splitting.

Following the diagonal matrix structure and the likelihood ratio test method, we propose a new test framework for high-dimensional data with a small sample size. Unlike the existing diagonal Hotelling's tests, in which the sample covariance matrix S was directly replaced with the diagonal matrix $\text{diag}(S)$, our likelihood ratio test statistics are a summation of the log-transformed squared t-statistics, rather than a direct summation of those components. When the sample size is small, the standard t tests may be unreliable due to the unstable variance estimates. As a remedy, our proposed tests use the log-transformed squared t-statistics and, consequently, provide more stable test statistics so that type I error rates are better controlled for small sample sizes. We demonstrate by simulation that our proposed tests are robust in terms of controlling the type I error rate at the nominal level in a wide range of settings.

The rest of the paper is organized as follows. In Section 2, we propose the diagonal likelihood ratio test method for the one-sample case. The asymptotic distributions of the test statistics are also derived as p tends to infinity under the null and local alternative hypotheses, respectively. In Section 3, we propose the diagonal likelihood ratio test method for the two-sample case and derive some asymptotic results, including the asymptotic null distribution and power. In Section 4, we conduct simulation studies to evaluate the proposed tests and to compare them with existing methods. We apply the proposed tests to a real data example in Section 5, and conclude the paper by providing a short summary and some future research directions in Section 6. The technical proofs are provided in the Web Appendix.

2 | ONE-SAMPLE TEST

2.1 | Diagonal LRT statistic

To illustrate the main idea of the diagonal likelihood ratio test method, we consider the one-sample test for a mean vector. Let $\boldsymbol{X}_i = (X_{i1}, X_{i2,}, \ldots, X_{ip})^T, i = 1, \ldots, n$, be independent and identically distributed (i.i.d.) random vectors from the multivariate normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu}$ is the population mean vector and Σ is the population covariance matrix. In the one-sample case, for a given vector $\boldsymbol{\mu}_0$, we test the hypothesis,

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \text{ versus } H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0. \quad (3)$$

Our new likelihood ratio test statistic is based on the assumption that the covariance matrix follows a diagonal

matrix structure, that is, $\Sigma = \text{diag}(\sigma_{11}^2, \ldots, \sigma_{pp}^2)$. In Web Appendix A.1, we show that the likelihood ratio test statistic for hypothesis (3) is

$$\Lambda_n = \frac{\max_{\Sigma} L(\boldsymbol{\mu}_0, \Sigma)}{\max_{\boldsymbol{\mu}, \Sigma} L(\boldsymbol{\mu}, \Sigma)} = \frac{\prod_{j=1}^{p} \left\{ \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)^2 \right\}^{n/2}}{\prod_{j=1}^{p} \left\{ \sum_{i=1}^{n} (X_{ij} - \mu_{0j})^2 \right\}^{n/2}},$$

where $\bar{X}_j = \sum_{i=1}^{n} X_{ij}/n$ are the sample means, and $s_j^2 = \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)^2/(n-1)$ are the sample variances. Taking the log transformation, we derive that

$$-2\log(\Lambda_n) = n \sum_{j=1}^{p} \log \left[1 + n \left(\bar{X}_j - \mu_{0j} \right)^2 / \left\{ (n-1) s_j^2 \right\} \right].$$

This suggests that the new test statistic is

$$T_1 = n \sum_{j=1}^{p} \log \left\{ 1 + \frac{n(\bar{X}_j - \mu_{0j})^2}{(n-1)s_j^2} \right\} = n \sum_{j=1}^{p} \log \left(1 + \frac{t_{nj}^2}{\nu_1} \right), \tag{4}$$

where $t_{nj} = \sqrt{n}(\bar{X}_j - \mu_{0j})/s_j$ are the standard t-statistics for the one-sample test with $\nu_1 = n - 1$ degrees of freedom. We refer to the diagonal likelihood ratio test statistic in (4) as the DLRT statistic.

Under the null hypothesis, it is easy to verify that $n \log\left(1 + t_{nj}^2/\nu_1\right) = t_{nj}^2 + O_p(1/n)$. Further, we have $T_1 = \sum_{j=1}^{p} t_{nj}^2 + O_p(p/n)$. So if p increases at such a rate that $p = o(n)$, then we have the following approximation:

$$T_1 \approx \sum_{j=1}^{p} t_{nj}^2 = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^T \left\{ \text{diag}(s_1^2, \ldots, s_p^2) \right\}^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0).$$

Thus, as a special case, our proposed DLRT statistic reduces to the diagonal Hotelling's test statistic in the one-sample case to which a direct summation of the squared t-statistics is applied.

2.2 | Null distribution

For ease of notation, let $U_{nj} = n \log(1 + t_{nj}^2/\nu_1)$ for $j = 1, \ldots, p$. In this section, we derive the asymptotic null distribution of the proposed DLRT statistic. To derive the limiting distribution, we first present a lemma; the proof is in Web Appendix A.2.

Lemma 1. *For the gamma function* $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, *let* $\Psi(x) = \Gamma'(x)/\Gamma(x)$ *be the digamma function. Also, let* $D(x) = \Psi\{(x+1)/2\} - \Psi(x/2)$, $m_1 = nD(\nu_1)$, *and* $m_2 = n^2\{D^2(\nu_1) - 2D'(\nu_1)\}$.

(*a*) *For any* $n \geq 2$, *we have* $E(U_{nj}) = m_1$ *and* $\text{Var}(U_{nj}) = m_2 - m_1^2$.

(*b*) *As* $n \to \infty$, *we have* $E(U_{nj}) \to 1$ *and* $\text{Var}(U_{nj}) \to 2$.

Despite T_1 has an additive form of log-transformed squared t-statistics, our derivation of its limiting distribution needs to account for the dependence among $\{U_{n1}, \ldots, U_{np}\}$. For example, the scaling parameter of T_1 may need to incorporate the information of $\text{Cov}(U_{n,j}, U_{n,j+k})$. We therefore need additional assumptions when establishing the asymptotic normality of the DLRT statistic. Let $\alpha(\mathcal{F}, \mathcal{G}) = \sup\{\|P(A \cap B) - P(A)P(B)\| : A \in \mathcal{F}, B \in \mathcal{G}\}$ be the strong mixing coefficient between two σ fields, $\mathcal{F}$ and $\mathcal{G}$, that measures the degree of dependence between the two σ fields. We also assume that the following two regularity conditions hold for the sequence $\{U_{nj}, j = 1, 2, \ldots\}$:

- **(C1)** Let $\alpha(r) = \sup\{\alpha(\mathcal{F}_1^k, \mathcal{F}_{k+r}^p) : 1 \leq k \leq p - r\}$, where $\mathcal{F}_a^b = \mathcal{F}_{a,n}^b = \sigma\{U_{nj} : a \leq j \leq b\}$. Assume that the stationary sequence $\{U_{nj}\}$ satisfies the strong mixing condition such that $\alpha(r) \downarrow 0$ as $r \to \infty$, where $\downarrow$ denotes the monotone decreasing convergence.
- **(C2)** Suppose that $\sum_{r=1}^{\infty} \alpha(r)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$, and for any $k \geq 0$, $\lim_{p\to\infty} \sum_{j=1}^{p-k} \text{Cov}(U_{nj}, U_{n,j+k})/(p-k) = \gamma(k)$ exists.

The following theorem establishes the asymptotic distribution of the DLRT statistic under the null hypothesis.

Theorem 1. *Let* $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ *be i.i.d. random vectors from* $N_p(\boldsymbol{\mu}, \Sigma)$. *If the sequence* $\{U_{nj}\}$ *is stationary and satisfies conditions (C1) and (C2), then under the null hypothesis, we have for any fixed* $n \geq 2$,

$$\frac{T_1 - pm_1}{\tau_1 \sqrt{p}} \xrightarrow{D} N(0, 1) \text{ as } p \to \infty$$

where $\xrightarrow{D}$ *denotes convergence in distribution, and* $\tau_1^2 = m_2 - m_1^2 + 2\sum_{k=1}^{\infty} \gamma(k)$.

The proof of Theorem 1 is given in Web Appendix A.3. In Theorem 1, we do not require Σ to follow a diagonal matrix structure. To derive the limiting distribution of the DLRT statistic under a general covariance matrix structure, we impose the mixing condition (C1) which implies a weak dependence structure in the data. Specifically, noting that $T_1 = \sum_{j=1}^{p} U_{nj}$, if the autocorrelation function of $\{U_{n1}, \ldots, U_{np}\}$ decays rapidly as the lag increases, T_1 will converge to the standard normal distribution under appropriate centering and scaling. Finally, we note that similar mixing conditions were also adopted in Gregory et al. (2015) and Zhao and Xu (2016) The asymptotic variance of $p^{1/2}T_1$, τ_1^2, depends on the autocovariance of the sequence $\{U_{n1}, U_{n2}, \ldots\}$ and is unknown. To establish the null distribution in practice, we need an estimate, $\widehat{\tau}_1^2$, to replace τ_1^2. In spectrum analysis, under the condition (C2), we note that $\sum_{k=-\infty}^{\infty} \gamma(k) = 2\pi f(0)$,

where $f(w)$ is a spectral density function defined as $f(w) = (2\pi)^{-1}\sum_{k=-\infty}^{\infty} e^{iwk}\gamma(k)$ for $w \in [-\pi, \pi]$. Therefore, we only need an estimate of $f(0)$.

The estimation of $f(w)$ has been extensively studied (eg, Bühlmann, 1996; Paparoditis and Politis, 2012). The traditional kernel estimator with a lag-window form is defined as

$$\widehat{f}(w) = (2\pi)^{-1}\sum_{k=-\infty}^{\infty} e^{iwk}\lambda(k/h)\widehat{\gamma}(k),$$

where $\widehat{\gamma}(k) = p^{-1}\sum_{j=1}^{p-k}(U_{nj} - \widetilde{T}_1)(U_{n(j+k)} - \widetilde{T}_1)$ is the sample autocovariance and $\widetilde{T}_1 = T_1/p$. We apply the Parzen window(Parzen, 1961) to determine the lag-window $\lambda(x)$ throughout the paper, where $\lambda(x) = 1 - 6x^2 - 6|x|^3$ if $|x| < 1/2$, and $\lambda(x) = 2(1 - |x|)^3$ if $1/2 \le |x| < 1$, and $\lambda(x) = 0$ if $|x| \ge 1$. Finally, we estimate τ_1^2 as

$$\widehat{\tau}_1^2 = 2\pi\widehat{f}(0) = 2\sum_{0<k\le h}\lambda(k/h)\widehat{\gamma}(k) + \gamma(0),$$

where h is the lag-window size, and $\gamma(0) = \mathrm{Var}(U_{nj}) = m_2 - m_1^2$.

Corollary 1. *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be i.i.d. random vectors from $N_p(\boldsymbol{\mu}, \Sigma)$ and assume that Σ is a diagonal matrix. Under the null hypothesis, we have the following asymptotic results:*

(*a*) *For any fixed $n \ge 2$, $(T_1 - pm_1)/\sqrt{p(m_2 - m_1^2)} \xrightarrow{D} N(0,1)$ as $p \to \infty$.*

(*b*) *If p increases at such a rate that $p = o(n^{2k})$, then for the given positive integer $k < \nu_1/2$,*

$$(T_1 - p\xi_k)/\sqrt{2p} \xrightarrow{D} N(0,1) \text{ as } (n,p) \to \infty,$$

where $\xi_k = n\{a_1 - a_2/2 + \cdots + (-1)^{k+1}a_k/k\}$ and $a_k = \prod_{i=1}^{k}\{(2i-1)/(\nu_1 - 2i)\}$.

The proof of Corollary 1 is given in Web Appendix A.4. This corollary defines asymptotic normality of the DLRT statistic for two scenarios under the diagonal covariance matrix assumption: the result from (a) establishes the asymptotic null distribution when n is fixed but p is large, and the result from (b) establishes the asymptotic null distribution when n and p are both large.

2.3 | Statistical power

To derive the asymptotic power of the proposed DLRT statistic for the one-sample test, we consider the local alternative

$$\boldsymbol{\mu} - \boldsymbol{\mu}_0 = \boldsymbol{\delta}_1/\sqrt{n}, \tag{5}$$

where $\boldsymbol{\delta}_1 = (\delta_{11}, \ldots, \delta_{1p})^T$. Assume that $\boldsymbol{\Delta}_1 = (\Delta_{11}, \ldots, \Delta_{1p})^T = (\delta_{11}/\sigma_{11}, \ldots, \delta_{1p}/\sigma_{pp})^T$, with all of the components uniformly bounded such that

$$\Delta_{1j}^2 \leqslant M_0, \text{ for } j = 1, \ldots, p, \tag{6}$$

where σ_{jj}^2 are the diagonal elements of Σ, and M_0 is a constant independent of n and p. Then we have the following theorem.

Theorem 2. *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be i.i.d. random vectors from $N_p(\boldsymbol{\mu}, \Sigma)$ and assume that p increases at such a rate that $p = o(n^2)$. Let z_α be the upper αth percentile such that $\Phi(z_\alpha) = 1 - \alpha$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. If the sequence $\{U_{nj}\}$ is stationary and satisfies conditions (C1) and (C2), then under the local alternative (5) and condition (6), the asymptotic power of the level α test is*

$$\beta(T_1) = 1 - \Phi\left(z_\alpha - \frac{\boldsymbol{\Delta}_1^T\boldsymbol{\Delta}_1/\sqrt{p}}{\sqrt{\tau_1^2}}\right) \text{ as } (n,p) \to \infty,$$

and hence $\beta(T_1) \to 1$ if $\sqrt{p} = o\big(\sum_{j=1}^{p}\delta_{1j}^2/\sigma_{jj}^2\big)$, and $\beta(T_1) \to \alpha$ if $\sum_{j=1}^{p}\delta_{1j}^2/\sigma_{jj}^2 = o(\sqrt{p})$.

The proof of Theorem 2 is given in Web Appendix A.5. If the true mean differences are dense but small such as the standardized signals $(\mu_{1j} - \mu_{0j})/\sigma_{jj} = \delta_0 p^{-1/2}$ with the constant $\delta_0 > 0$, then the asymptotic power will increase towards 1 as $(n,p) \to \infty$.

3 | TWO-SAMPLE TEST

In this section, we consider the two-sample test for mean vectors with equal covariance matrices. Let $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ip})^T$, $i = 1, \ldots, n_1$, be i.i.d. random vectors from $N_p(\boldsymbol{\mu}_1, \Sigma)$, and $\boldsymbol{Y}_k = (Y_{k1}, Y_{k2}, \ldots, Y_{kp})^T$, $k = 1, \ldots, n_2$, be i.i.d. random vectors from $N_p(\boldsymbol{\mu}_2, \Sigma)$, where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are two population mean vectors and Σ is the common covariance matrix. For ease of notation, let $N = n_1 + n_2$ and assume that $\lim_{N\to\infty} n_1/N = c \in (0,1)$. Let also $\bar{\boldsymbol{X}} = \sum_{i=1}^{n_1}\boldsymbol{X}_i/n_1$ and $\bar{\boldsymbol{Y}} = \sum_{k=1}^{n_2}\boldsymbol{Y}_k/n_2$ be two sample mean vectors, and

$$S = \frac{1}{N-2}\left\{\sum_{i=1}^{n_1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T + \sum_{k=1}^{n_2}(\boldsymbol{Y}_k - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_k - \bar{\boldsymbol{Y}})^T\right\}.$$

be the pooled sample covariance matrix.

In the two-sample case, we test the hypothesis

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ versus } H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2. \tag{7}$$

In Web Appendix B.1, we show that the DLRT statistic for hypothesis (7) is

$$T_2 = N\sum_{j=1}^{p}\log\left\{1+\frac{n_1n_2}{N(N-2)}\frac{(\bar{X}_j-\bar{Y}_j)^2}{s^2_{j,\text{pool}}}\right\} = N\sum_{j=1}^{p}\log\left(1+\frac{t^2_{Nj}}{\nu_2}\right), \quad (8)$$

where $t_{Nj}=\sqrt{n_1n_2/N}(\bar{X}_j-\bar{Y}_j)/s_{j,\text{pool}}$ are the standard t-statistics for the two-sample case with $\nu_2=N-2$ degrees of freedom, and $s^2_{j,\text{pool}}$ are the pooled sample variances, that is, the diagonal elements of S.

For ease of notation, let $V_{Nj}=N\log(1+t^2_{Nj}/\nu_2)$ for $j=1,\ldots,p$. The following theorem establishes the asymptotic null distribution of the DLRT statistic for the two-sample case under centering and scaling.

Theorem 3. *Let $\{\boldsymbol{X}_i\}_{i=1}^{n_1}$ and $\{\boldsymbol{Y}_k\}_{k=1}^{n_2}$ be i.i.d. random vectors from $N_p(\boldsymbol{\mu}_1,\Sigma)$ and $N_p(\boldsymbol{\mu}_2,\Sigma)$, respectively. If the sequence $\{V_{Nj}\}$ is stationary and satisfies conditions (C1) and (C2), then under the null hypothesis, we have for any fixed $N\geq 4$,*

$$\frac{T_2-pG_1}{\tau_2\sqrt{p}}\xrightarrow{\mathcal{D}}N(0,1)\ \text{ as }\ p\to\infty$$

where $\tau_2^2=G_2-G_1^2+2\sum_{k=1}^{\infty}\gamma(k)$, with $G_1=ND(\nu_2)$ and $G_2=N^2\{D^2(\nu_2)-2D'(\nu_2)\}$.

The proof of Theorem 3 is given in Web Appendix B.2. By imposing conditions (C1) and (C2) on the sequence $\{V_{Nj}\}$, Theorem 3 also does not require the assumption that each of the covariance matrices follows a diagonal matrix structure. Similar to the one-sample case, a consistent estimator for τ_2^2 is given as

$$\widehat{\tau}_2^2=2\sum_{0<k\leq h}\lambda(k/h)\widehat{\gamma}(k)+\gamma(0),$$

where $\lambda(x)$ is the Parzen window, h is the lag-window size, $\gamma(0)=\text{Var}(V_{Nj})=G_2-G_1^2$, and $\widehat{\gamma}(k)=p^{-1}\sum_{j=1}^{p-k}(V_{Nj}-\widetilde{T}_2)(V_{N(j+k)}-\widetilde{T}_2)$ is the sample autocovariance for $\{V_{Nj},j=1,2,\ldots,p\}$ and $\widetilde{T}_2=T_2/p$.

Corollary 2. *Let $\{\boldsymbol{X}_i\}_{i=1}^{n_1}$ and $\{\boldsymbol{Y}_k\}_{k=1}^{n_2}$ be i.i.d. random vectors from $N_p(\boldsymbol{\mu}_1,\Sigma)$ and $N_p(\boldsymbol{\mu}_2,\Sigma)$, respectively, and assume that Σ is a diagonal matrix. Under the null hypothesis, we have the following asymptotic results:*

(*a*) *For any fixed $N\geq 4$, $(T_1-pG_1)/\sqrt{p(G_2-G_1^2)}\xrightarrow{\mathcal{D}}N(0,1)$ as $p\to\infty$.*

(*b*) *If p increases at such a rate that $p=o(N^{2k})$, then for the given positive integer $k<\nu_2/2$,*

$$(T_2-p\eta_k)/\sqrt{2p}\xrightarrow{\mathcal{D}}N(0,1)\ \text{ as }\ (N,p)\to\infty,$$

where $\eta_k=N\{b_1-b_2/2+\cdots+(-1)^{k+1}b_k/k\}$ and $b_k=\prod_{i=1}^{k}\{(2i-1)/(\nu_2-2i)\}$.

The proof of Corollary 2 is given in Web Appendix B.3. This corollary defines asymptotic normality of the DLRT statistic for two scenarios under the diagonal covariance matrix assumption: the result from (a) establishes the asymptotic null distribution when N is fixed but p is large, and the result from (b) establishes the asymptotic null distribution when N and p are both large.

When $\boldsymbol{\mu}_1\neq\boldsymbol{\mu}_2$, we consider the local alternative

$$\boldsymbol{\mu}_1-\boldsymbol{\mu}_2=\sqrt{\frac{N}{n_1n_2}}\boldsymbol{\delta}_2, \quad (9)$$

where $\boldsymbol{\delta}_2=(\delta_{21},\ldots,\delta_{2p})^T$. We assume that $\boldsymbol{\Delta}_2=(\Delta_{21},\ldots,\Delta_{2p})^T=(\delta_{21}/\sigma_{11},\ldots,\delta_{2p}/\sigma_{pp})^T$, with all of the components uniformly bounded such that

$$\Delta_{2j}^2\leqslant M_1,\ \text{for } j=1,\ldots,p, \quad (10)$$

where σ_{jj}^2 are the diagonal elements of Σ, and M_1 is a constant independent of N and p. The following theorem establishes the asymptotic power of our proposed DLRT statistic for the two-sample test.

Theorem 4. *Let $\{\boldsymbol{X}_i\}_{i=1}^{n_1}$ and $\{\boldsymbol{Y}_k\}_{k=1}^{n_2}$ be i.i.d. random vectors from $N_p(\boldsymbol{\mu}_1,\Sigma)$ and $N_p(\boldsymbol{\mu}_2,\Sigma)$, respectively. Assume that p increases at such a rate that $p=o(N^2)$. If the sequence $\{V_{Nj},j=1,2,\ldots\}$ is stationary and satisfies conditions (C1) and (C2), then under the local alternative (9) and condition (10), the asymptotic power of the level α test is*

$$\beta(T_2)=1-\Phi\left(z_\alpha-\frac{\boldsymbol{\Delta}_2^T\boldsymbol{\Delta}_2/\sqrt{p}}{\sqrt{\tau_2^2}}\right)\ \text{ as }\ (N,p)\to\infty,$$

and hence, $\beta(T_2)\to 1$ if $\sqrt{p}=o\big(\sum_{j=1}^{p}\delta_{2j}^2/\sigma_{jj}^2\big)$, and $\beta(T_2)\to\alpha$ if $\sum_{j=1}^{p}\delta_{2j}^2/\sigma_{jj}^2=o(\sqrt{p})$.

4 | MONTE CARLO SIMULATION STUDIES

In this section, we carry out simulations to evaluate the performance of our DLRT method. For ease of presentation, we consider the proposed DLRT test for the two-sample case only. We compare DLRT with five existing tests from the

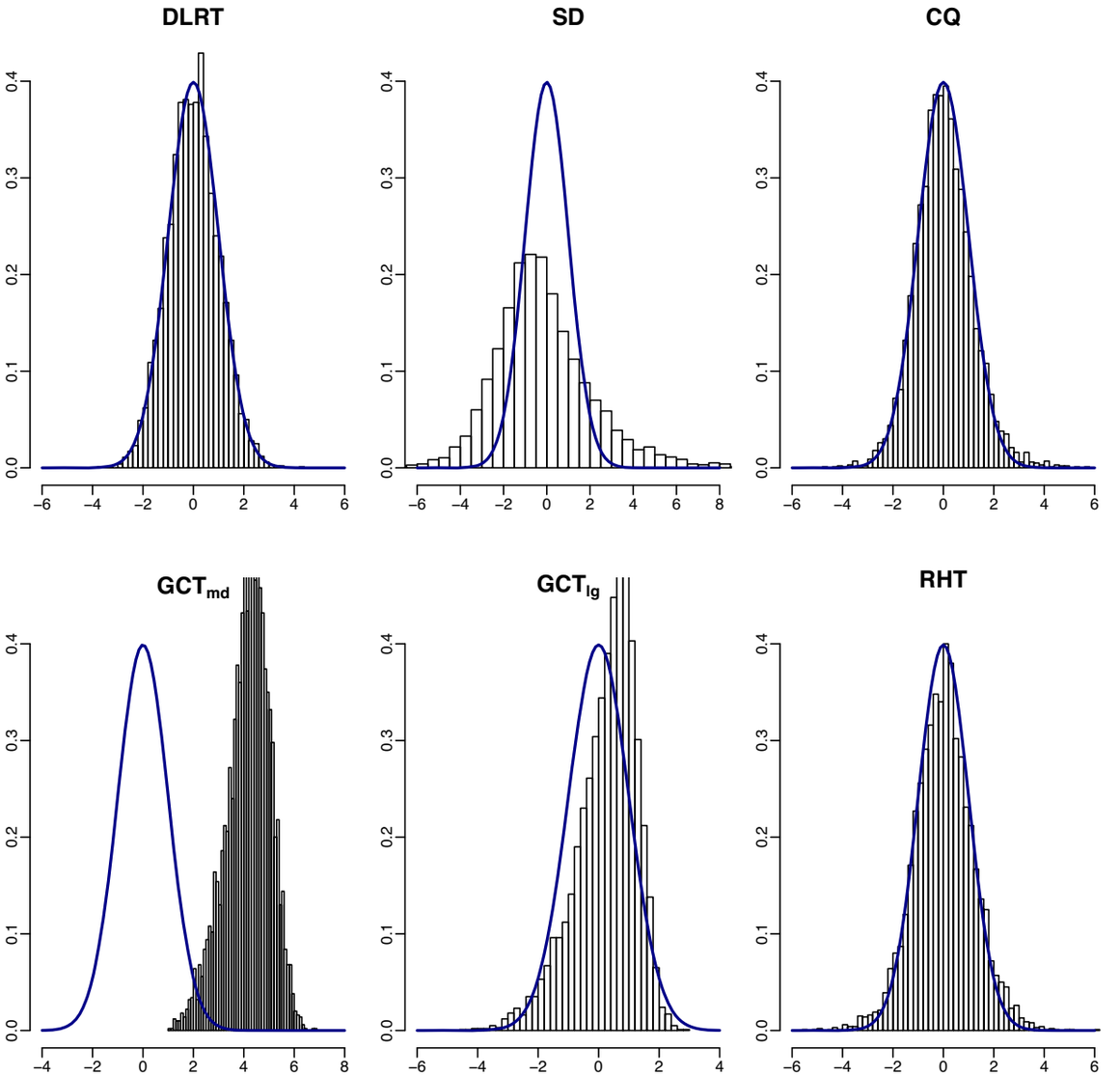


FIGURE 1 Comparison with the standard normal distribution under the null hypothesis for the DLRT, SD, CQ, GCT_{md}, GCT_{lg}, and RHT tests with $n_1 = n_2 = 3$ and $p = 500$. The histograms are based on 5000 simulations. This figure appears in color in the electronic version of this article.

aforementioned three categories: one unscaled Hotelling's test including the CQ test from Chen and Qin (2010), one regularized Hotelling's test including the RHT test from Chen et al. (2011), and two diagonal Hotelling's tests including the SD test from Srivastava and Du (2008), and the GCT test from Gregory et al. (2015). Gregory et al. (2015) considered two different versions of the GCT test with centering corrections that allowed the dimension to grow at either a moderate or large order of the sample size, which are denoted as GCT_{md} and GCT_{lg}, respectively. The lag-window size throughout the simulations is $h = 5$.

4.1 | Normal data

In the first simulation, we generate $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n_1}$ from $N_p(\boldsymbol{\mu}_1, \Sigma)$, and $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n_2}$ from $N_p(\boldsymbol{\mu}_2, \Sigma)$. For simplicity, let $\boldsymbol{\mu}_1 = \boldsymbol{0}$. Under the alternative hypothesis, we assume that the first p_0 elements in $\boldsymbol{\mu}_2$ are nonzero, where $p_0 = \beta p$ with $\beta \in [0, 1]$ being the tuning parameter that controls the signal sparsity. When $\beta = 0$, the null hypothesis holds. The common covariance matrix is $\Sigma = D^T R D$, where R is the correlation matrix and D is a diagonal matrix such that $D = \text{diag}(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{pp})$. To account for the heterogeneity of variances, $\sigma_{11}^2, \ldots, \sigma_{pp}^2$ are randomly sampled from the scaled chi-square distribution $\chi_5^2/5$. For the dependence structure in the matrix R, we consider the following three scenarios:

(a) Independent (IND) structure: R is the $p \times p$ identity matrix.
(b) Short range dependence (SRD) structure: $R = (\rho^{|i-j|})_{p \times p}$ follows the first-order autoregressive structure, in which

TABLE 1 Type I error rates over 2000 simulations for the DLRT, GCT_{md}, GCT_{lg}, SD, CQ, RHT tests under three dependence structures. The significance level is $\alpha = 0.05$. Two different correlation, $\rho = 0.3$ or 0.6, are considered for the SRD structure

		$p = 100$			$p = 500$		
Cov	**Method\(n_1, n_2)**	**(3, 3)**	**(5, 5)**	**(15, 15)**	**(3, 3)**	**(5, 5)**	**(15, 15)**
IND	DLRT	0.060	0.056	0.058	0.055	0.043	0.048
	SD	0.403	0.111	0.044	0.342	0.038	0.024
	GCT_{md}	0.608	0.233	0.063	0.983	0.922	0.151
	GCT_{lg}	0.101	0.124	0.092	0.047	0.110	0.060
	CQ	0.092	0.066	0.056	0.084	0.058	0.051
	RHT	0.097	0.068	0.058	0.105	0.070	0.053
SRD ($\rho = 0.3$)	DLRT	0.067	0.058	0.054	0.054	0.053	0.061
	SD	0.387	0.105	0.038	0.309	0.046	0.023
	GCT_{md}	0.548	0.203	0.071	0.986	0.893	0.143
	GCT_{lg}	0.116	0.119	0.094	0.05	0.106	0.070
	CQ	0.076	0.054	0.052	0.078	0.059	0.051
	RHT	0.099	0.067	0.058	0.098	0.060	0.060
SRD ($\rho = 0.6$)	DLRT	0.072	0.076	0.078	0.080	0.072	0.078
	SD	0.351	0.082	0.031	0.28	0.041	0.023
	GCT_{md}	0.421	0.152	0.091	0.980	0.749	0.116
	GCT_{lg}	0.130	0.164	0.129	0.064	0.122	0.095
	CQ	0.079	0.050	0.056	0.080	0.051	0.053
	RHT	0.110	0.067	0.060	0.099	0.063	0.048
LRD	DLRT	0.061	0.065	0.054	0.052	0.071	0.056
	SD	0.383	0.112	0.034	0.332	0.054	0.025
	GCT_{md}	0.594	0.222	0.067	0.983	0.913	0.143
	GCT_{lg}	0.104	0.152	0.095	0.047	0.109	0.077
	CQ	0.092	0.058	0.052	0.082	0.062	0.057
	RHT	0.108	0.058	0.060	0.102	0.063	0.052

the correlation among the observations decays exponentially with distance. We consider $\rho = 0.3$ or 0.6 to represent two different levels of correlation.

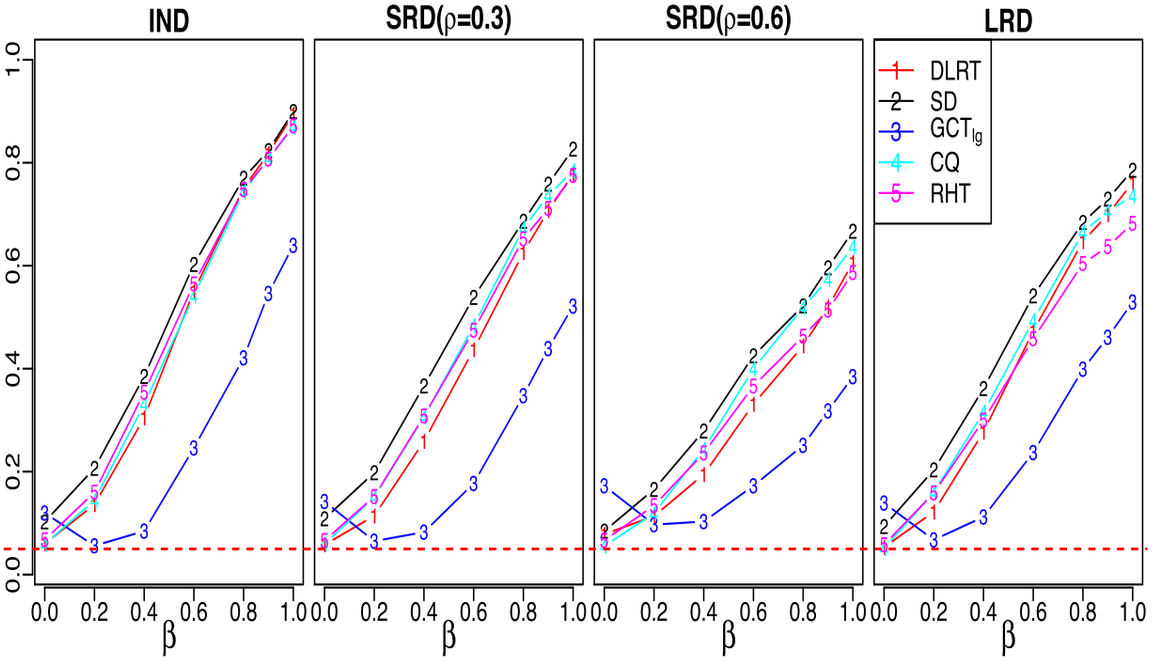
(c) Long range dependence (LRD) structure: we follow the same setting as in Gregory et al. (2015) Specifically, we consider the (i, j)th element of R as $r_{ij} = [(k+1)^{2H} + (k-1)^{2H} - 2k^{2H}]/2$ with $k = |j - i|$, and the self-similarity parameter as $H = 0.625$.

For the power comparison, we set the jth nonzero component in μ_2 as $\mu_{2j} = \theta\sigma_{jj}, j = 1, \ldots, p_0$, where θ is the effect size of the corresponding component. The other parameters are set as $(n_1, n_2, \theta) \times p = \{(3, 3, 0.5)$ or $(5, 5, 0.5)$ or $(15, 15, 0.25)\} \times \{100$ or $500\}$, respectively.

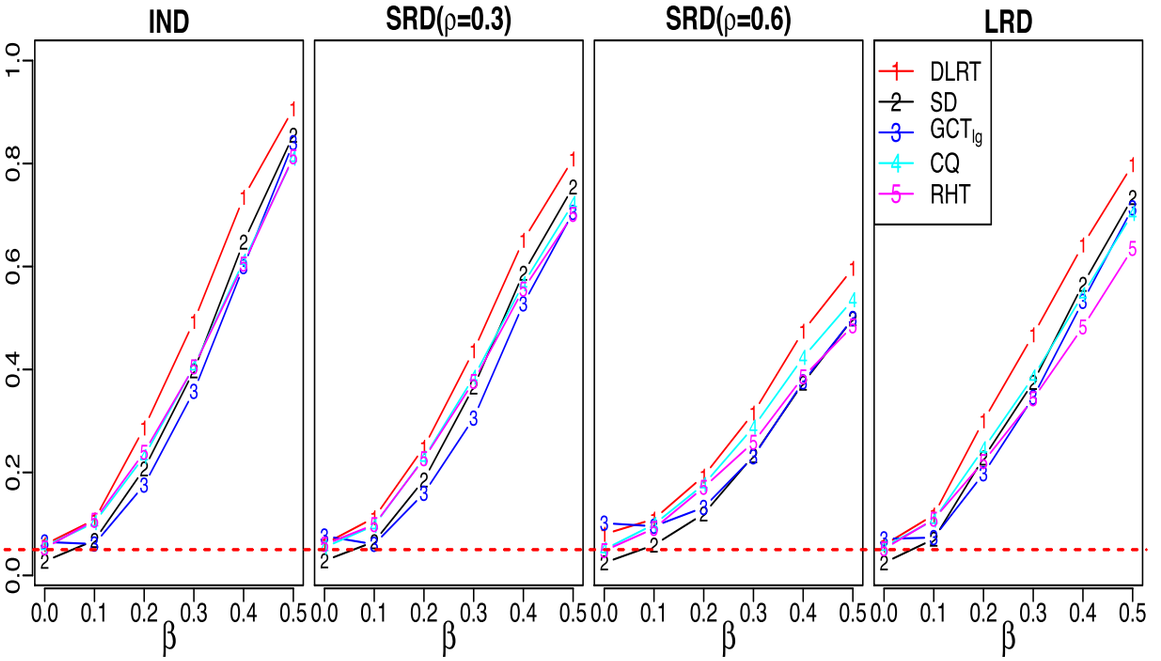
Figure 1 shows the simulated null distributions of the DLRT, SD, GCT_{md}, GCT_{lg}, CQ, and RHT tests under the independent structure, when the sample size is small (eg, $n_1 = n_2 = 3$) and the dimension is large. The histograms are based on 5000 simulations. For comparison, their limiting distributions are also plotted. However, the null distributions of the other three tests, and especially the GCT_{md} test, are either skewed or shifted away from the standard normal distribution.

We summarize the type I error rates from the simulations for each of the six tests, with different sample sizes and dependence structures, in Table 1. When the variables are uncorrelated or weakly correlated with each other, the type I error rates of DLRT are closer to the nominal level ($\alpha = 0.05$) than the other five tests under most settings. In addition, DLRT provides a more stable test statistic and better control over the type I error rate when the sample size is not large; the SD, GCT_{lg}, CQ, and RHT tests have inflated type I error rates when the sample size is relatively small (eg, $n_1 = n_2 = 3$). The GCT_{md} test in particular fails to keep the type I error rate within the nominal level under each setting, and performs more poorly when the sample size is small and the dimension is large. Therefore, we exclude the GCT_{md} test from the following power comparison.

Figure 2 presents the simulated power of the DLRT, SD, GCT_{lg}, CQ, and RHT tests at the significance level $\alpha = 0.05$. When the dimension is low (eg, $p = 100$), the DLRT, CQ, and RHT tests are able to control the type I error rates well, whereas the SD and GCT_{lg} tests suffer from inflated type I error rates. In particular for the GCT_{lg} test, it exhibits a relatively low power when the sample size is small. This coincides with the findings in Figure 1. As the dimension is large and the sample size is not small, the DLRT, SD, CQ, and RHT tests control the type I error rate close to the nominal level, whereas the GCT_{lg} test still fails. DLRT also provides a higher power in most settings. To conclude, DLRT performs comparably to the existing tests for normal data.

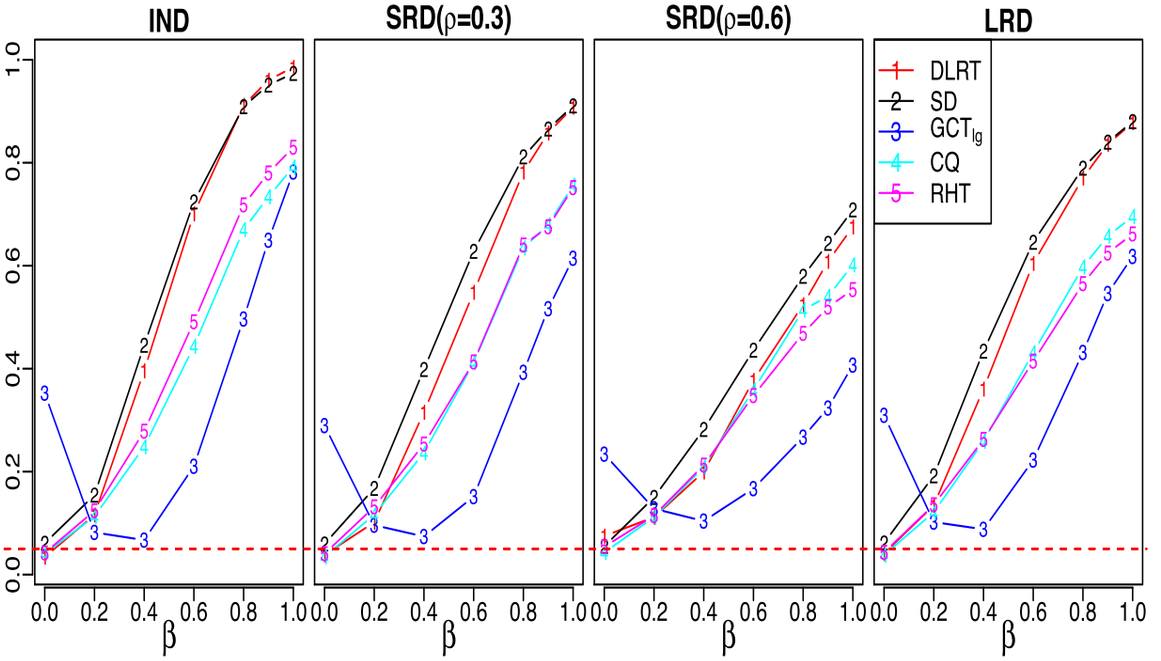


(a) $n_1 = n_2 = 5, p = 100$

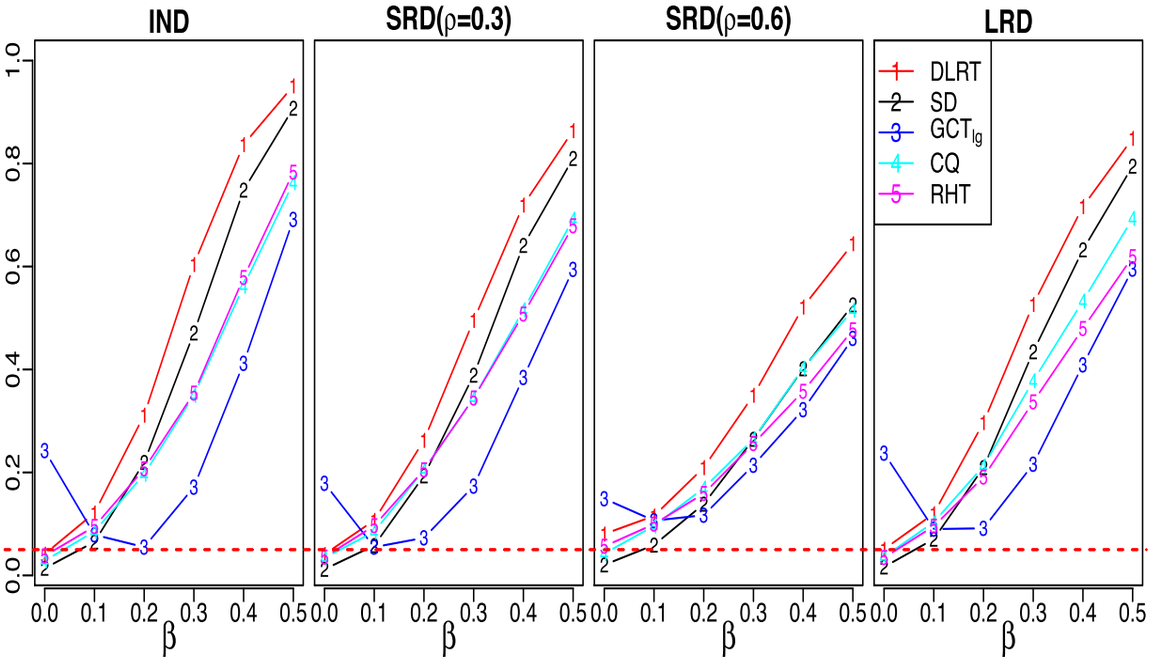


(b) $n_1 = n_2 = 15, p = 500$

FIGURE 2 Power comparisons among the DLRT, SD, GCT_{lg}, CQ, and RHT tests with ($n_1 = n_2 = 5, p = 100$) or ($n_1 = n_2 = 15, p = 500$), respectively. The horizontal dashed red lines represent the significance level of $\alpha = 0.05$. The results are based on 2000 simulations with data from the normal distribution. This figure appears in color in the electronic version of this article, and color refers to that version.



(a) $n_1 = n_2 = 5, p = 100$



(b) $n_1 = n_2 = 15, p = 500$

FIGURE 3 Power comparisons among the DLRT, SD, GCT_{lg}, CQ, and RHT tests with ($n_1 = n_2 = 5$, $p = 100$) or ($n_1 = n_2 = 15$, $p = 500$), respectively. The horizontal dashed red lines represent the significance level of $\alpha = 0.05$. The results are based on 2000 simulations with data from a heavy-tailed distribution. This figure appears in color in the electronic version of this article, and color refers to that version.

4.2 | Heavy-tailed data

To evaluate the robustness of DLRT, we also conduct simulations with heavy-tailed data. Following Gregory et al. (2015), the data are generated based on a "double" Pareto distribution with parameters a and b. The algorithm is as follows:

(i) Generate two independent random variables U and V, where U is from the Pareto distribution with the cumulative distribution function $F(x) = 1 - (1 + x/b)^{-a}$ for $x \geq 0$, and V is a binary random variable with $P(V = 1) = P(V = -1) = 0.5$. Then $Z = UV$ follows the double Pareto distribution with parameters a and b.

(ii) Generate random vectors $\{\boldsymbol{X}_i^{(0)} = (x_{i1}, \ldots, x_{ip})^T\}_{i=1}^{n_1}$, and random vectors $\{\boldsymbol{Y}_k^{(0)} = (y_{k1}, \ldots, y_{kp})^T\}_{k=1}^{n_2}$, where all the components of $\boldsymbol{X}_i^{(0)}$ and $\boldsymbol{Y}_k^{(0)}$ are sampled independently from the double Pareto distribution with parameters $a = 16.5$ and $b = 8$.

(iii) Let $\boldsymbol{X}_i = \boldsymbol{\mu}_1 + \Sigma^{1/2}\boldsymbol{X}_i^{(0)}/c_0$ and $\boldsymbol{Y}_k = \boldsymbol{\mu}_2 + \Sigma^{1/2}\boldsymbol{Y}_k^{(0)}/c_0$, where $c_0^2 = 512/899$ is the variance of the double Pareto distribution with $a = 16.5$ and $b = 8$, and $\Sigma = D^T R D$ with $D = \text{diag}(\sigma_{11}, \ldots, \sigma_{pp})$. Consequently, $\boldsymbol{X}_i$ and $\boldsymbol{Y}_k$ have a common correlation matrix $\boldsymbol{R}$.

For the matrix $\boldsymbol{R}$, we also consider three scenarios: (a) the IND structure, (b) the SRD structure, and (c) the LRD structure. In each scenario, the generating algorithms for $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and Σ follow the simulation procedure described in Section 4.1. The parameters used in the algorithms are $(n_1, n_2, \theta) \times p = \{(5, 5, 0.5) \text{ or } (15, 15, 0.25)\} \times \{100 \text{ or } 500\}$, respectively.

Figure 3 presents the simulation results for the five tests with heavy-tailed data at the significance level $\alpha = 0.05$. When the dimension is large and the sample size is small, the DLRT, SD, and RHT tests control the type I error rate well, whereas the GCT_{lg} test exhibits a substantially inflated type I error rate and a low power for detection. One possible explanation is that the GCT_{lg} statistic involves the estimation of high order moments which leads to instability when the sample size is small. DLRT is again more powerful than the CQ and RHT tests in most settings. In summary, it is evident that the DLRT test provides a more robust performance with heavy-tailed data than the existing five tests, especially when the dimension is large.

5 | BRAIN CANCER DATA ANALYSIS

In this section, we apply DLRT to a data set from The Cancer Genome Atlas (TCGA). This data set contains the copy number measurements from genomic locations of the probes on chromosomes in 92 long-term survivors and 138 short-term survivors with a brain cancer called glioblastoma multiforme. The long-term brain cancer survivors lived for more than two years after their first diagnosis, and the short-term survivors lived for less than two years after their first diagnosis. According to Olshen et al. (2004) and Baladandayuthapani et al. (2010), the copy number variations between the patient groups will occur across multiple probes rather than at a single probe. That is, the signal structure is dense-but-small rather than sparse-but-strong. To identify the particular regions in the genome where the genes were differentially expressed, we apply the following tests: the DLRT, SD, GCT_{lg}, CQ, and RHT tests. Gregory et al. (2015) separated the whole chromosome into 26 segments of varying lengths. We focus our analysis on one segment of the q arm of chromosome 1, which contains measurements of probes at 400 locations. The copy number data at 400 locations are summarized in "chr1qseg.rda" which is available from the R package "highD2pop."

TABLE 2 Empirical power for testing the equality of gene expressions in the TCGA data, when $p = 100, 200$, or 400. The nominal level is α and the sample sizes of the two classes are $n_1 = n_2 = 8$

		DLRT	SD	GCT_{lg}	CQ	RHT
$\alpha = 0.05$	$p = 100$	0.165	0.121	0.082	0.143	0.189
	$p = 200$	0.158	0.096	0.094	0.116	0.194
	$p = 400$	0.144	0.081	0.119	0.106	0.136
$\alpha = 0.10$	$p = 100$	0.274	0.201	0.165	0.233	0.345
	$p = 200$	0.261	0.157	0.190	0.192	0.354
	$p = 400$	0.258	0.154	0.212	0.181	0.295

To compare the performance of the tests, we first perform the two-sample t-tests to screen top p significant genes, and then calculate the empirical power with $p = 100$, 200, or 400, respectively. To determine the empirical critical values corresponding to a given nominal level α, we bootstrap two distinct classes from the short-term survival group to compute the test statistics. Since both classes are partitioned from the short-term survival group, the null hypothesis can be regarded as the truth. Therefore, we repeat the procedure 10 000 times for each test method, and select the $(10\,000\alpha)$th largest value of the test statistics as the empirical critical values. To determine the empirical power, we bootstrap one class from the short-term survival group and another class from the long-term survival group. For both classes, we consider $n_1 = n_2 = 8$ for computing the empirical critical values and power.

Table 2 shows the empirical power of the DLRT, SD, GCT_{lg}, CQ, and RHT tests. We note that the DLRT test performs nearly as well as the RHT test, and it has a higher empirical power than the other three tests under all the settings.

6 | CONCLUSION

In the classical low-dimensional setting, Hotelling's T^2 test is an important and useful tool for testing the equality of one or two mean vectors from multivariate normal distributions. However, this classic test may not be applicable when

the dimension is larger than the sample size, as the sample covariance matrix is no longer invertible. This motivates the development of new methods to address the testing problems for high-dimensional data with a small sample size. According to how the covariance matrices are estimated, most available methods can be classified into three categories: the unscaled Hotelling's tests, the regularized Hotelling's tests, and the diagonal Hotelling's tests.

In this paper, we proposed a new test framework based on the likelihood ratio test for both one- and two-sample cases. The proposed test statistics are derived under the assumption that the covariance matrices follow a diagonal matrix structure. Our tests use the log-transformed squared t-statistics and provide more stable test statistics than the standard t-statistics when the sample size is small. Through simulation studies, we showed that DLRT is also more robust than the existing test methods when the data are heavy-tailed or weakly correlated. In other words, when the dimension is large and the sample size is small, DLRT is able to keep the type I error rate within the nominal level and, at the same time, maintains a high power for detection.

The proposed new test assumes a natural ordering of the components in the p-dimensional random vector, for example, the correlation among the components are related to their positions, and hence we can take into account the additional structure information to avoid an estimation of the full covariance matrix. When the ordering of the components is not available, we propose to reorder the components from the sample data using some well known ordering methods before applying our proposed tests. For instance, with the best permutation algorithm in Rajaratnam and Salzman (2013), the strongly correlated elements can be reordered close to each other. For other ordering methods of random variables, one may refer to, for example, Gilbert et al. (1992), Wagaman and Levina (2009), and the references therein.

When the sample size is relatively small and the correlation is very high, our proposed tests will have slightly inflated type I error rates, especially when the dimension is also large. This is mainly because the test statistics are derived under the assumption that the covariance matrices follow a diagonal matrix structure. When the diagonal matrix assumption is violated, the asymptotic null distributions may not follow the standard normal distribution, or the asymptotic properties may require more restrictive assumptions including a larger sample size. To overcome these limitations, future research is warranted to improve our current version of DLRT or to derive more accurate asymptotic distributions when the underlying assumptions are violated.

We also note that our current paper has focused on testing high-dimensional mean vectors under the parametric setting. More recently, some nonparametric tests have also been developed in the literature for the same testing problems; see, for example, Wang et al. (2015), Ghosh and Biswas (2016), and Chakraborty and Chaudhuri (2017).

ACKNOWLEDGMENTS

The authors thank the editor, the associate editor, and three reviewers for their constructive comments that have led to a substantial improvement of the paper. Tiejun Tong's research was supported by the Hong Kong RGC Grant (No. HKBU12303918), the National Natural Science Foundation of China (No. 11671338), the Health and Medical Research Fund (No. 04150476), and three Hong Kong Baptist University Grants (RC-IG-FNRA/17-18/13, FRG1/17-18/045, and FRG2/17-18/020). Marc G. Genton's research was supported by the King Abdullah University of Science and Technology (KAUST).

ORCID

Tiejun Tong http://orcid.org/0000-0003-0947-3990
Marc G. Genton http://orcid.org/0000-0001-6467-2998

REFERENCES

Ahmad, M. R. (2014). A U-statistic approach for a high-dimensional two-sample mean testing problem under non-normality and Behrens-Fisher setting. *Ann Inst Stat Math* 66, 33–61.

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

Baladandayuthapani, V., Ji, Y., Talluri, R., Nieto-Barajas, L. E., and Morris, J. S. (2010). Bayesian random segmentation models to identify shared copy number aberrations for array CGH data. *J Am Stat Assoc* 105, 1358–1375.

Bickel, P. J. and Levina, E. (2004). Some theory for Fisher's linear discriminant function, 'naive Bayes', and some alternatives when there are many more variables than observations. *Bernoulli* 10, 989–1010.

Bühlmann P. (1996). Locally adaptive lag–window spectral estimation. *J Time Ser Anal* 17, 247–270.

Chakraborty, A. and Chaudhuri, P. (2017). Tests for high-dimensional data based on means, spatial signs and spatial ranks. *Ann Stat* 45, 771–799.

Chen, S. and Qin, Y. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann Stat* 38, 808–835.

Chen, L. S., Paul, D., Prentice, R. L., and Wang, P. (2011). A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. *J Am Stat Assoc.* 106, 1345–1360.

Dong, K., Pang, H., Tong, T., and Genton, M. G. (2016). Shrinkage-based diagonal Hotelling's tests for high-dimensional small sample size data. *J Multivariate Anal* 143, 127–142.

Dudoit, S., Fridlyand, J., and Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *J Am Stat Assoc* 97, 77–87.

Feng, L., Zou, C., Wang, Z., and Zhu, L. (2015). Two-sample Behrens-Fisher problem for high-dimensional data. *Stat Sin* 25, 1297–1312.

Gilbert, J. R., Moler, C., and Schreiber, R. (1992). Sparse matrices in MATLAB: Design and implementation. *SIAM J Matrix Anal Appl* 13, 333–356.

Ghosh, A. K. and Biswas, M. (2016). Distribution-free high-dimensional two-sample tests based on discriminating hyperplanes. *TEST* 25, 525–547.

Gregory, K. B., Carroll, R. J., Baladandayuthapani, V., and Lahiri, S. N. (2015). A two-sample test for equality of means in high dimension. *J Am Stat Assoc* 110, 837–849.

Huang, S., Tong, T., and Zhao, H. (2010). Bias-corrected diagonal discriminant rules for high-dimensional classification. *Biometrics* 66, 1096–1106.

Jiang, T. and Qi, Y. (2015). Likelihood ratio tests for high-himensional normal distributions. *Scand J Stat* 42, 988–1009.

Jiang, T. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Annal Stat* 41, 2029–2074.

Li, H., Aue, A., Paul, D., Peng, J., and Wang, P. (2016). An adaptable generalization of Hotelling's T^2 test in high dimension. *arXiv preprint arXiv:1609.08725*.

Lopes, M. E., Jacob, L., and Wainwright, M. J. (2011). A more powerful two-sample test in high dimensions using random projection. *Adv Neural Inf Process Syst* 24, 1206–1214.

Olshen, A. B., Venkatraman, E. S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5, 557–572.

Paparoditis, E. and Politis, D. N. (2012). Nonlinear spectral density estimation: Thresholding the correlogram. *J Time Ser Anal* 33, 386–397.

Park, J. and Ayyala, D. N. (2013). A test for the mean vector in large dimension and small samples. *J Stat Plan Inference* 143, 929–943.

Parzen, E. (1961). Mathematical considerations in the estimation of spectra. *Technometrics* 3, 167–190.

Rajaratnam, B. and Salzman, J. (2013). Best permutation analysis. *J Multivariate Anal* 121, 193–223.

Srivastava, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *J Multivariate Anal* 100, 518–532.

Srivastava, R., Li, P., and Ruppert, D. (2016). RAPTT: An exact two-sample test in high dimensions using random projections. *J Comput Graph Stat* 23, 954–970.

Srivastava, M. S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *J Multivariate Anal* 99, 386–402.

Srivastava, M. S., Katayama, S., and Kano, Y. (2013). A two sample test in high dimensional data. *J Multivariate Anal* 114, 349–358.

Städler, N. and Mukherjee, S. (2017). Two-sample testing in high dimensions. *J R Stat Soc Ser B* 79, 225–246.

Thulin, M. (2014). A high-dimensional two-sample test for the mean using random subspaces. *Comput Stat Data Anal* 74, 26–38.

Wagaman, A. S. and Levina, E. (2009). Discovering sparse covariance structures with the Isomap. *J Comput Graph Stat* 18, 551–572.

Wang, L., Peng, B., and Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J Am Stat Assoc* 110, 1658–1669.

Wei, S., Lee, C., Wichers, L., and Marron, J. S. (2016). Direction-projection-permutation for high-dimensional hypothesis tests. *J Comput Graph Stat* 25, 549–569.

Wu, Y., Genton, M. G., and Stefanski, L. A. (2006). A multivariate two-sample mean test for small sample size and missing data. *Biometrics* 62, 877–885.

Zhao, J. and Xu, X. (2016). A generalized likelihood ratio test for normal mean when p is greater than n. *Comput Stat Data Anal* 99, 91–104.

SUPPORTING INFORMATION

Web Appendix referenced in Sections 2 and 3 is available with this paper at the Biometrics website on Wiley Online Library. The R code for implementing our new DLRT method is also available at the Biometrics website on Wiley Online Library.

How to cite this article: Hu Z, Tong T, Genton MG. Diagonal likelihood ratio test for equality of mean vectors in high-dimensional data. *Biometrics.* 2019;75:256–267. https://doi.org/10.1111/biom.12970