Biometrics 000, 000–000 DOI: 000
000 0000

Web-based Supplementary Materials for "Diagonal Likelihood Ratio Test for Equality of Mean Vectors in High-Dimensional Data" by

Zongliang Hu[1], Tiejun Tong[2,*] and 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

*Email: tongt@hkbu.edu.hk

A. DLRT for One-Sample Test

A.1 *Derivation of Formula (4)*

Under the assumption $\Sigma = \text{diag}(\sigma_{11}^2, \sigma_{22}^2, \ldots, \sigma_{pp}^2)$, we can write the likelihood function as

$$L(\boldsymbol{\mu}, \Sigma) = \prod_{j=1}^{p} L_j(\mu_j, \sigma_{jj}^2),$$

where $L_j(\mu_j, \sigma_{jj}^2) = (2\pi)^{-\frac{n}{2}}(\sigma_{jj}^2)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(X_{ij}-\mu_j)^2/\sigma_{jj}^2}$ is the likelihood function for the jth component with data $X_{1j}, \ldots, X_{nj}$. Deriving the maximum likelihood estimator of $L(\boldsymbol{\mu}, \Sigma)$ is equivalent to finding the maximum likelihood estimators of $L_j(\mu_j, \sigma_{jj}^2)$ for $j = 1, 2, \ldots, p$, respectively.

It is known that the maximum of $L_j(\mu_j, \sigma_{jj}^2)$ is achieved when $\widehat{\mu}_j = \bar{X}_j = \sum_{i=1}^{n} X_{ij}/n$ and $s_j^2 = \sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2/(n-1)$. Hence, under the alternative hypothesis,

$$\max_{\boldsymbol{\mu},\Sigma} L(\boldsymbol{\mu}, \Sigma) = (2\pi)^{-\frac{np}{2}} e^{-\frac{np}{2}} \prod_{j=1}^{p} \Big\{\frac{1}{n}\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2\Big\}^{-\frac{n}{2}}.$$

Similarly, under the null hypothesis,

$$\max_{\Sigma} L(\boldsymbol{\mu}_0, \Sigma) = (2\pi)^{-\frac{np}{2}} e^{-\frac{np}{2}} \prod_{j=1}^{p} \Big\{\frac{1}{n}\sum_{i=1}^{n}(X_{ij} - \mu_{0j})^2\Big\}^{-\frac{n}{2}}.$$

From the above results, the likelihood ratio test statistic is given as

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

Further, we have

$$\Lambda_n^{2/n} = \prod_{j=1}^{p} \frac{\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2}{\sum_{i=1}^{n}(X_{ij} - \mu_{0j})^2} = \prod_{j=1}^{p} \frac{(n-1)s_j^2}{(n-1)s_j^2 + n(\bar{X}_j - \mu_{0j})^2}$$
$$= \prod_{j=1}^{p} \Big\{1 + \frac{n}{n-1}(\bar{X}_j - \mu_{0j})^2\Big\}^{-1}.$$

This leads to the test statistic

$$-2\log(\Lambda_n) = n\sum_{j=1}^{p} \log\Big\{1 + \frac{n}{n-1}(\bar{X}_j - \mu_{0j})^2/s_j^2\Big\} = n\sum_{j=1}^{p} \log\big(1 + t_{nj}^2/\nu_1\big),$$

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.

Lemma 1: *Let the gamma and digamma functions be the same as in Lemma 1. For any $\nu > 0$, we have the following integral equalities:*

$$\int_0^\infty (1+z^2)^{-(\nu+1)/2} dz = \frac{1}{2\sqrt{\nu}K(\nu)},$$
$$\int_0^\infty (1+z^2)^{-(\nu+1)/2} \log(1+z^2) dz = \frac{D(\nu)}{2\sqrt{\nu}K(\nu)},$$
$$\int_0^\infty (1+z^2)^{-(\nu+1)/2} \{\log(1+z^2)\}^2 dz = \frac{D^2(\nu) - 2D'(\nu)}{2\sqrt{\nu}K(\nu)},$$

where $K(\nu) = \Gamma\{(\nu+1)/2\}/\{\sqrt{\pi\nu}\Gamma(\nu/2)\}$.

We note that this lemma essentially follows the results of Lemmas 4 to 6 in Zhu and Galbraith (2010), and hence the proof is omitted.

A.2 *Proof of Lemma 1*

For simplicity, we omit the subscript n in the terms U_{nj} and t_{nj}. Noting that $\nu_1 = n-1$ and $U_j = n\log(1 + t_j^2/\nu_1)$, where $t_j = \sqrt{n}(\bar{X}_j - \mu_{0j})/s_j$, we have

$$E(U_j) = \frac{2n}{\sqrt{\nu_1 \pi}} \frac{\Gamma\{(\nu_1+1)/2\}}{\Gamma(\nu_1/2)} \int_0^\infty \log\left(1 + \frac{t^2}{\nu_1}\right)\left(1 + \frac{t^2}{\nu_1}\right)^{-\frac{\nu_1+1}{2}} dt,$$
$$E(U_j^2) = \frac{2n^2}{\sqrt{\nu_1 \pi}} \frac{\Gamma\{(\nu_1+1)/2\}}{\Gamma(\nu_1/2)} \int_0^\infty \left\{\log\left(1 + \frac{t^2}{\nu_1}\right)\right\}^2 \left(1 + \frac{t^2}{\nu_1}\right)^{-\frac{\nu_1+1}{2}} dt.$$

Letting $z = t/\sqrt{\nu_1}$ and by the method of substitution,

$$E(U_j) = 2n\sqrt{\nu_1}K(\nu_1) \int_0^\infty \log(1+z^2)(1+z^2)^{-\frac{\nu_1+1}{2}} dz,$$
$$E(U_j^2) = 2n^2\sqrt{\nu_1}K(\nu_1) \int_0^\infty \big\{\log(1+z^2)\big\}^2 (1+z^2)^{-\frac{\nu_1+1}{2}} dz.$$

By Lemma 1, we have $E(U_j) = nD(\nu_1)$ and $E(U_j^2) = n^2\big\{D^2(\nu_1) - 2D'(\nu_1)\big\}$. This shows that $E(U_j) = m_1$ and $E(U_j^2) = m_2$, consequently, $\mathrm{Var}(U_j) = m_2 - m_1^2$.

We note that $n\log(1 + t^2/\nu_1) \leqslant nt^2/\nu_1 < 2t^2$, and that $(1 + t^2/\nu_1)^{-n/2}$ converges to $e^{-t^2/2}$

as $n \to \infty$. By the dominated convergence theorem,

$$\lim_{n\to\infty} \frac{2n}{\sqrt{\nu_1 \pi}} \frac{\Gamma\{(\nu_1+1)/2\}}{\Gamma(\nu_1/2)} \int_0^\infty \log\left(1+\frac{t^2}{\nu_1}\right)\left(1+\frac{t^2}{\nu_1}\right)^{-\frac{\nu_1+1}{2}} dt$$
$$= \lim_{n\to\infty} \frac{2}{\sqrt{\nu_1 \pi}} \frac{\Gamma\{(\nu_1+1)/2\}}{\Gamma(\nu_1/2)} \int_0^\infty t^2 e^{-\frac{t^2}{2}} dt$$
$$= \sqrt{\frac{2}{\nu_1}} \lim_{n\to\infty} \frac{\Gamma\{(\nu_1+1)/2\}}{\Gamma(\nu_1/2)} = 1,$$

where the last equation is obtained by Stirling's formula, $\Gamma(x) = \sqrt{2\pi}x^{x-1/2}e^{-x}\{1+O(1/x)\}$ as $x \to \infty$ (Spira, 1971). This shows that $m_1 \to 1$ as $n \to \infty$. Similarly, $m_2 \to 3$ as $n \to \infty$. Finally, $\mathrm{Var}(U_j) = m_2 - m_1^2 \to 2$ as $n \to \infty$.

A.3 *Proof of Theorem 1*

As the sequence $\{U_{nj}\}$ satisfies conditions (C1) and (C2), we only need to prove that $E|U_{nj} - E(U_{nj})|^{2+\delta} < \infty$ for any fixed n. We note that

$$E|U_{nj} - E(U_{nj})|^{2+\delta} = \frac{2n^{2+\delta}}{\sqrt{\nu_1 \pi}} \frac{\Gamma\{(\nu_1+1)/2\}}{\Gamma(\nu_1/2)} \int_0^\infty \left(1+\frac{t^2}{\nu_1}\right)^{-\frac{\nu_1+1}{2}} \left\{\log\left(1+\frac{t^2}{\nu_1}\right) - D(\nu_1)\right\}^{2+\delta} dt.$$

Then we only need to verify that, for any fixed $n \geqslant 2$,

$$\int_0^\infty (1+x^2)^{-\frac{\nu_1+1}{2}} \{\log(1+x^2) - D(\nu_1)\}^{2+\delta} dx < \infty.$$

The inequality clearly holds.

Finally, by the central limit theorem under the strong mixing condition (see Corollary 5.1 in Hall and Heyde (1980)), we have $(T_1 - pm_1)/(\tau_1\sqrt{p}) \xrightarrow{\mathcal{D}} N(0,1)$ as $p \to \infty$.

A.4 *Proof of Corollary 1*

(a) When the covariance matrix is a diagonal matrix, $\{t_{nj} = \sqrt{n}(\bar{X}_j - \mu_{0j})/s_j,\ j = 1, \dots, p\}$ are i.i.d. random variables. Consequently, $\{U_{nj},\ j = 1, \dots, p\}$ are also i.i.d. random variables with mean m_1 and variance $m_2 - m_1^2$. By the central limit theorem, we have
$(\widetilde{T}_1 - m_1)/\sqrt{p(m_2 - m_1^2)} \ \xrightarrow{\mathcal{D}}\ N(0,1)$ as $p \to \infty$.

(b) For simplicity, we omit the subscript n in the terms U_{nj} and t_{nj}. We note that $\log(1 + t_j^2/\nu_1)$ is sandwiched between $t_j^2/\nu_1 - (t_j^2)^2/2\nu_1^2 + \cdots + (-1)^{k+2}(t_j^2)^{k+1}/\{(k+1)\nu_1^{k+1}\}$ and

$t_j^2/\nu_1 - (t_j^2)^2/2\nu_1^2 + \cdots + (-1)^{k+3}(t_j^2)^{k+2}/\{(k+2)\nu_1^{k+2}\}$. For any $n > 2k+6$,

$$E(U_j) = m_1 = \frac{n}{\nu_1}E(t_j^2) - \frac{n}{2\nu_1^2}E\{(t_j^2)^2\} + \cdots + (-1)^{k+1}\frac{n}{k\nu_1^k}E\{(t_j^2)^k\} + O(1/n^k).$$

Noting that $E\{(t_j^2/\nu_1)^k\} = a_k$, and $m_2 - m_1 \to 2$ as $n \to \infty$, we have

$$\begin{aligned}\frac{T_1 - p\xi_k}{\sqrt{2p}} &= \frac{\sqrt{p}(\widetilde{T}_1 - m_1)}{\sqrt{2}} + \frac{\sqrt{p}(m_1 - \xi_k)}{\sqrt{2}} \\ &= \frac{\sqrt{p}(\widetilde{T}_1 - m_1)}{\sqrt{m_2 - m_1^2}}\frac{\sqrt{m_2 - m_1^2}}{\sqrt{2}} + \sqrt{p}O(1/n^k) \xrightarrow{\mathcal{D}} N(0,1).\end{aligned}$$

A.5 *Proof of Theorem 2*

First of all, we have

$$\frac{\widetilde{T}_1 - 1}{\sqrt{\hat{\tau}_1^2}} = \frac{\widetilde{T}_1 - \overline{m}_1}{\sqrt{\hat{\tau}_1^2}} + \frac{\sqrt{p}(\overline{m}_1 - 1)}{\sqrt{\hat{\tau}_1^2}}, \tag{1}$$

where $\overline{m}_1 = \sum_{j=1}^p \widetilde{m}_{1j}/p$ and $\widetilde{m}_{1j} = E(U_j|H_1) = E\big\{n\log\big(1 + t_j^2/\nu_1\big)|H_1\big\}$.

We note that $nt_j^2/\nu_1 - n(t_j^2)^2/(2\nu_1^2) \leqslant U_j \leqslant nt_j^2/\nu_1$. Then, for any $n > 6$, we have $nE(t_j^2|H_1)/\nu_1 - nE\big\{(t_j^2)^2|H_1\big\}/(2\nu_1^2) \leqslant E(U_j|H_1) \leqslant nE(t_j^2|H_1)/\nu_1$. Under the local alternative (5) and condition (6), we have $\widetilde{m}_{1j} = nE(t_j^2|H_1)/\nu_1 + O(1/n)$. We also note that t_j follows a noncentral t distribution with $\nu_1 = n-1$ degrees of freedom and a noncentrality parameter Δ_{1j}. Its second moment is $E(t_j^2|H_1) = \nu_1(1+\Delta_{1j}^2)/(\nu_1 - 2)$; hence, $\widetilde{m}_{1j} = 1 + \Delta_{1j}^2 + O(1/n)$. Consequently, $\overline{m}_1 = \sum_{j=1}^p \widetilde{m}_{1j}/p = 1 + \boldsymbol{\Delta}_1^T\boldsymbol{\Delta}_1/p + O(1/n)$.

Under conditions (C1) and (C2), for any consistent estimator $\hat{\tau}_1^2$ for τ_1^2,

$$\begin{aligned}\frac{T_1 - p}{\sqrt{p\hat{\tau}_1^2}} &= \frac{\sqrt{p}(\widetilde{T}_1 - \overline{m}_1)}{\sqrt{\hat{\tau}_1^2}} + \frac{\boldsymbol{\Delta}_1^T\boldsymbol{\Delta}_1/\sqrt{p}}{\sqrt{\hat{\tau}_1^2}} + \frac{\sqrt{p}O(1/n)}{\sqrt{\hat{\tau}_1^2}} \\ &\xrightarrow{\mathcal{D}} N(0,1) + \frac{\boldsymbol{\Delta}_1^T\boldsymbol{\Delta}_1/\sqrt{p}}{\sqrt{\tau_1^2}}.\end{aligned}$$

We note that $\boldsymbol{\Delta}_1^T\boldsymbol{\Delta}_1/\sqrt{p} = \sum_{j=1}^p(\delta_{1j}/\sigma_{jj})^2/\sqrt{p}$. Thus, if $\sqrt{p} = o(\sum_{j=1}^p \delta_{1j}^2/\sigma_{jj}^2)$, then the power of the one-sample test will increase towards 1 as $(n,p) \to \infty$. On the other hand, if $\sum_{j=1}^p \delta_{1j}^2/\sigma_{jj}^2 = o(\sqrt{p})$, then the test will have little power as $(n,p) \to \infty$.

B. DLRT for Two-Sample Test

B.1 *Derivation of Formula (8)*

For the two-sample test, we derive the DLRT statistic based on the assumption that the two covariance matrices are equal and they follow a diagonal matrix structure, i.e., $\Sigma_1 = \Sigma_2 = \Sigma = \mathrm{diag}(\sigma_{11}^2, \sigma_{22}^2, \ldots, \sigma_{pp}^2)$.

Under the alternative hypothesis, the maximum of the likelihood function is

$$\max_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\Sigma} L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma | H_1) = (2\pi e)^{-\frac{p}{2}(n_1+n_2)} \prod_{j=1}^{p} \Big[\frac{1}{n_1+n_2}\Big\{\sum_{i=1}^{n_1}(X_{ij}-\bar{X}_j)^2 + \sum_{k=1}^{n_2}(Y_{kj}-\bar{Y}_j)^2\Big\}\Big]^{-\frac{(n_1+n_2)}{2}}.$$

Similarly, under the null hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$,

$$\max_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\Sigma} L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma | H_0) = (2\pi e)^{-\frac{p}{2}(n_1+n_2)} \prod_{j=1}^{p} \Big[\frac{1}{n_1+n_2}\Big\{\sum_{i=1}^{n_1}(X_{ij}-\widehat{\mu}_{0j})^2 + \sum_{k=1}^{n_2}(Y_{kj}-\widehat{\mu}_{0j})^2\Big\}\Big]^{-\frac{(n_1+n_2)}{2}},$$

where $\widehat{\mu}_{0j} = (n_1\bar{X}_j + n_2\bar{Y}_j)/(n_1+n_2)$, $\bar{X}_j = \sum_{i=1}^{n_1} X_{ij}/n_1$, and $\bar{Y}_j = \sum_{k=1}^{n_2} Y_{kj}/n_2$.

We also note that

$$\sum_{i=1}^{n_1}(X_{ij}-\widehat{\mu}_{0j})^2 + \sum_{k=1}^{n_2}(Y_{kj}-\widehat{\mu}_{0j})^2 = \sum_{i=1}^{n_1}(X_{ij}-\bar{X}_j)^2 + \sum_{k=1}^{n_2}(Y_{kj}-\bar{Y}_j)^2 + \frac{n_1 n_2^2}{(n_1+n_2)^2}(\bar{X}_j-\bar{Y}_j)^2 + \frac{n_2 n_1^2}{(n_1+n_2)^2}(\bar{Y}_j-\bar{X}_j)^2.$$

We have

$$\Lambda_N = \frac{\max_{\boldsymbol{\mu_1},\boldsymbol{\mu_2},\Sigma} L(\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \Sigma | H_0)}{\max_{\boldsymbol{\mu_1},\boldsymbol{\mu_2},\Sigma} L(\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \Sigma | H_1)} = \prod_{j=1}^{p} \frac{\big\{\frac{n_1+n_2-2}{n_1+n_2}s_{j,\mathrm{pool}}^2 + \frac{n_1 n_2}{(n_1+n_2)^2}(\bar{X}_j-\bar{Y}_j)^2\big\}^{-\frac{(n_1+n_2)}{2}}}{\Big[\frac{1}{n_1+n_2}\big\{\sum_{i=1}^{n_1}(X_{ij}-\bar{X}_j)^2 + \sum_{k=1}^{n_2}(Y_{kj}-\bar{Y}_j)^2\big\}\Big]^{-\frac{(n_1+n_2)}{2}}},$$

where $s_{j,\mathrm{pool}}^2 = \{(n_1-1)s_{1j}^2 + (n_2-1)s_{2j}^2\}/(n_1+n_2-2)$ are the pooled sample variances with $s_{1j}^2 = \sum_{i=1}^{n_1}(X_{ij}-\bar{X}_j)^2/(n_1-1)$ and $s_{2j}^2 = \sum_{k=1}^{n_2}(Y_{kj}-\bar{Y}_j)^2/(n_2-1)$. This leads to the test statistic

$$-2\log(\Lambda_N) = (n_1+n_2)\sum_{j=1}^{p}\log\Big\{1 + \frac{\frac{n_1 n_2}{n_1+n_2}(\bar{X}_j-\bar{Y}_j)^2}{s_{j,\mathrm{pool}}^2}\frac{1}{n_1+n_2-2}\Big\} = N\sum_{j=1}^{p}\log\Big(1+\frac{t_{Nj}^2}{\nu_2}\Big),$$

where $t_{Nj} = \sqrt{n_1 n_2/(n_1+n_2)}(\bar{X}_j - \bar{Y}_j)/s_{j,\text{pool}}$ are the standard t-statistics for the two-sample test with $\nu_2 = N-2$ degrees of freedom.

B.2 *Proof of Theorem 3*

We first show that $E(V_{Nj}) = G_1 \to 1$ and $\text{Var}(V_{Nj}) = G_2 - G_1^2 \to 2$ as $N \to \infty$. For simplicity, we omit the subscript N in the terms V_{Nj} and t_{Nj}. As $V_j = N\log\big(1+t_j^2/\nu_2\big)$, we have

$$E(V_j) = \frac{2N}{\sqrt{\nu_2\pi}}\frac{\Gamma\{(\nu_2+1)/2\}}{\Gamma(\nu_2/2)}\int_0^\infty \log\left(1+\frac{t^2}{\nu_2}\right)\left(1+\frac{t^2}{\nu_2}\right)^{-\frac{\nu_2+1}{2}}dt,$$
$$E(V_j^2) = \frac{2N^2}{\sqrt{\nu_2\pi}}\frac{\Gamma\{(\nu_2+1)/2\}}{\Gamma(\nu_2/2)}\int_0^\infty \left\{\log\left(1+\frac{t^2}{\nu_2}\right)\right\}^2\left(1+\frac{t^2}{\nu_2}\right)^{-\frac{\nu_2+1}{2}}dt.$$

Let $z = t/\sqrt{\nu_2}$, we have

$$E(V_j) = 2N\sqrt{\nu_2}K(\nu_2)\int_0^\infty \log(1+z^2)(1+z^2)^{-\frac{\nu_2+1}{2}}dz,$$
$$E(V_j^2) = 2N^2\sqrt{\nu_2}K(\nu_2)\int_0^\infty \big\{\log(1+z^2)\big\}^2(1+z^2)^{-\frac{\nu_2+1}{2}}dz.$$

By Lemma 1, we have $E(V_j) = G_1 = ND(\nu_2)$ and $E(V_j^2) = G_2 = N^2\{D^2(\nu_2) - 2D'(\nu_2)\}$.

We note that $N\log(1+t^2/\nu_2) \leqslant Nt^2/\nu_2 \leqslant 2t^2$, and that $(1+t^2/\nu_2)^{-N/2}$ converges to $e^{-t^2/2}$ as $N \to \infty$. By the dominated convergence theorem,

$$\lim_{N\to\infty}\frac{2N}{\sqrt{\nu_2\pi}}\frac{\Gamma\{(\nu_2+1)/2\}}{\Gamma(\nu_2/2)}\int_0^\infty \log\left(1+\frac{t^2}{\nu_2}\right)\left(1+\frac{t^2}{\nu_2}\right)^{-\frac{\nu_1+1}{2}}dt$$
$$= \lim_{N\to\infty}\frac{2}{\sqrt{\nu_2\pi}}\frac{\Gamma\{(\nu_2+1)/2\}}{\Gamma(\nu_2/2)}\int_0^\infty t^2e^{-\frac{t^2}{2}}dt$$
$$= \sqrt{\frac{2}{\nu_2}}\lim_{N\to\infty}\frac{\Gamma\{(\nu_2+1)/2\}}{\Gamma(\nu_2/2)} = 1.$$

This shows that $E(V_j) = G_1 \to 1$ as $N \to \infty$. Similarly, $E(V_j^2) = G_2 \to 3$ as $N \to \infty$. Finally, $\text{Var}(V_j) = G_2 - G_1^2 \to 2$ as $N \to \infty$.

Second, we prove that $(T_2 - pG_1)/(\tau_2\sqrt{p}) \xrightarrow{\mathcal{D}} N(0,1)$ as $p \to \infty$. As the sequence $\{V_j\}$ satisfies conditions (C1) and (C2), we only need to prove that, for any fixed $N \geqslant 4$, $E|V_j -$

$E(V_j)|^{2+\delta} < \infty$. We note that

$$E|V_j - E(V_j)|^{2+\delta} = \frac{2N^{2+\delta}}{\sqrt{\nu_2 \pi}} \frac{\Gamma\{(\nu_2+1)/2\}}{\Gamma(\nu_2/2)} \int_0^\infty \Big(1+\frac{t^2}{\nu_2}\Big)^{-\frac{\nu_2+1}{2}} \Big\{ \log\Big(1+\frac{t^2}{\nu_2}\Big) - D(\nu_2)\Big\}^{2+\delta} dt.$$

Then, by the similar arguments to those in the proof of Theorem 1, we have $E|V_j - E(V_j)|^{2+\delta} < \infty$.

B.3 *Proof of Corollary 2*

(a) When the covariance matrix is a diagonal matrix, $\{t_{Nj} = \sqrt{n_1 n_2/(n_1+n_2)}(\bar{X}_j - \bar{Y}_j)/s_{j,\text{pool}}, j = 1,\ldots,p\}$ are i.i.d. random variables. Consequently, $\{V_{Nj}, j=1,\ldots,p\}$ are also i.i.d. random variables with mean G_1 and variance $G_2 - G_1^2$. By the central limit theorem, we have $(T_2 - pG_1)/\sqrt{p(G_2-G_1^2)} \xrightarrow{\mathcal{D}} N(0,1)$ as $p \to \infty$.

(b) For simplicity, we omit the subscript N in the terms t_{Nj} and V_{Nj}. We note that $\log(1 + t_j^2/\nu_2)$ is sandwiched between $t_j^2/\nu_2 - (t_j^2)^2/2\nu_2^2 + \cdots + (-1)^{k+2}(t_j^2)^{k+1}/\{(k+1)\nu_2^{k+1}\}$ and $t_j^2/\nu_2 - (t_j^2)^2/2\nu_2^2 + \cdots + (-1)^{k+3}(t_j^2)^{k+2}/\{(k+2)\nu_2^{k+2}\}$. Then, for any $N > 2k+6$,

$$E(V_j) = G_1 = \frac{N}{\nu_2}E(t_j^2) - \frac{N}{2\nu_2^2}E\{(t_j^2)^2\} - \cdots - \frac{N}{k\nu_2^k}E\{(t_j^2)^k\} + O(1/N^k).$$

Following the proof of Theorem 3, we have $G_2 - G_1^2 \to 2$ as $N \to \infty$. Further,

$$\begin{aligned}\frac{T_2 - p\xi_k}{\sqrt{2p}} &= \frac{\sqrt{p}(T_2/p - G_1)}{\sqrt{2}} + \frac{\sqrt{p}(G_1 - \eta_k)}{\sqrt{2}} \\ &= \frac{\sqrt{p}(T_2/p - G_1)}{\sqrt{G_2 - G_1^2}}\frac{\sqrt{G_2 - G_1^2}}{\sqrt{2}} + \sqrt{p}O(1/N^k) \xrightarrow{\mathcal{D}} N(0,1).\end{aligned}$$

B.4 *Proof of Theorem 4*

First of all, we have

$$\frac{\widetilde{T}_2 - 1}{\sqrt{\hat{\tau}_2^2}} = \frac{\widetilde{T}_2 - \overline{G}_1}{\sqrt{\hat{\tau}_2^2}} + \frac{\overline{G}_1 - 1}{\sqrt{\hat{\tau}_2^2}}, \tag{2}$$

where $\overline{G}_1 = \sum_{j=1}^p \widetilde{G}_{1j}/p$ and $\widetilde{G}_{1j} = E(V_j|H_1) = E\{N\log\big(1+t_j^2/\nu_2\big)|H_1\}$.

We note that $N\{t_j^2/\nu_2 - (t_j^2)^2/(2\nu_2^2)\} \leqslant V_j \leqslant Nt_j^2/\nu_2$. Then, for any $N > 6$, we have $N[E(t_j^2|H_1)/\nu_2 - E\{(t_j^2)^2|H_1\}/(2\nu_2^2)] \leqslant E(V_j|H_1) \leqslant NE(t_j^2|H_1)/\nu_2$. Under the local alternative (9) and condition (10), we have $\widetilde{G}_{1j} = NE(t_j^2|H_1)/\nu_2 + O(1/N)$. We also note that t_j follows a noncentral t distribution with $\nu_2 = N - 2$ degrees of freedom and a noncentrality parameter Δ_{2j}. Its second moment is $E(t_j^2|H_1) = \nu_2(1 + \Delta_{2j}^2)/(\nu_2 - 2)$; hence, $\widetilde{G}_{1j} = 1 + \Delta_{2j}^2 + O(1/N)$. Consequently, $\overline{G}_1 = 1 + \boldsymbol{\Delta}_2^T \boldsymbol{\Delta}_2/p + O(1/N)$.

Under conditions (C1) and (C2), for any consistent estimator $\hat{\tau}_2^2$ for τ_2^2,

$$\begin{aligned}\frac{T_2 - p}{\sqrt{p\hat{\tau}_2^2}} &= \frac{\sqrt{p}(\widetilde{T}_2 - \overline{G}_1)}{\sqrt{\hat{\tau}_2^2}} + \frac{\boldsymbol{\Delta}_2^T \boldsymbol{\Delta}_2/\sqrt{p}}{\sqrt{\hat{\tau}_2^2}} + \frac{\sqrt{p}O(1/N)}{\sqrt{\hat{\tau}_2^2}} \\ &\xrightarrow{\mathcal{D}} N(0,1) + \frac{\boldsymbol{\Delta}_2^T \boldsymbol{\Delta}_2/\sqrt{p}}{\sqrt{\tau_2^2}}.\end{aligned}$$

We note that $\boldsymbol{\Delta}_2^T \boldsymbol{\Delta}_2/\sqrt{p} = p^{-1/2}\sum_{j=1}^p(\delta_{2j}^2/\sigma_{jj}^2)$. Thus, if $\sqrt{p} = o(\sum_{j=1}^p \delta_{2j}^2/\sigma_{jj}^2)$, then the power of the two-sample test will increase towards 1 as $(N,p) \to \infty$. On the other hand, if $\sum_{j=1}^p \delta_{2j}^2/\sigma_{jj}^2 = o(\sqrt{p})$, then the test will have little power as $(N,p) \to \infty$.

C. Additional Simulations

C.1 *Simulations for the comparison between the DLRT and PA tests*

As mentioned by the reviewer, Park and Ayyala (2013) also proposed a scale invariant test, referred to as the PA test, which uses the idea of leave-out cross validation. To evaluate the performance of the DLRT and PA tests, we conduct simulations based on the same settings as Section 4.1 in the main paper.

Figure S1 presents the power of the DLRT and PA tests with the nominal level of $\alpha = 0.05$. When the sample size is small (e.g., $n = 5$), the PA test has some inflated type I error rate, and also suffers from a low power of detection. When the sample size is not small (e.g., $n = 15$), the DLRT and PA tests are able to simultaneously control the type I error rate, and at the same time keep a high power of detection. To conclude, when the sample size is not large, DLRT performs better than or at least as well as the PA test for normal distributed data.

[Figure 1 about here.]

In addition, the PA test is not a shift invariant test, which indicates that if the mean vectors under the null hypothesis are not located at the origin, the PA test may not keep the type I error rate at the nominal level. To further demonstrate this point, let $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = (100, \cdots, 100)^T \in R^p$. We consider $p = 100$, 300 or 500 and $n_1 = n_2 = 5$ or 15, respectively. The other settings are the same as Section 4.1 in the main paper. The simulation results are reported in Table S1. From Table S1, we can see that the PA test suffers from significant inflated type I error rate, whereas, the DLRT test still has a well control type I error rate.

[Table 1 about here.]

C.2 *Additional simulations with highly heterogeneous variances*

In Section 4.1 in the main paper, the variances, $\sigma_{11}^2, \ldots, \sigma_{pp}^2$, are randomly sampled from the scaled chi-square distribution $\chi_5^2/5$. To account for highly heterogeneous variances, we let $\sigma_{11}^2 < \sigma_{11}^2 < \ldots < \sigma_{pp}^2$ be equally spaced on the interval $[0.01, 150]$. The other settings are all the same as Section 4.1. Accordingly, the variances of the signal components are small, whereas the variances of the noise components are large.

Figure S2 shows the simulation results. When the sample size is small, e.g., $n = 5$, the SD, PA, $\mathrm{GCT}_{\mathrm{lg}}$ tests suffer from inflated type I error rates, whereas, the DLRT is able to control the type I error rate and exhibits high power than the CQ and RHT tests. As the dimension is large and the sample size is not small, the DLRT and PA tests are both able to control the type I error rates and at the same time exhibit high power than the other tests such as CQ and RHT. In addition, from Figure S2, we can also see that the diagonal Hotelling's tests like our proposed DLRT perform better than the unscaled Hotelling's tests like CQ, and the regularized Hotelling's tests like RHT. As mentioned in the Introduction of the main paper, when the variances of the signal components are small and the variances of the noise components are large, the scale transformation invariant tests such as the diagonal

Hotelling's tests usually provide a better performance than the orthogonal transformation invariant tests such as the unscaled Hotelling's tests and the regularized Hotelling's tests.

[Figure 2 about here.]

C.3 *Additional simulations with general multivariate data*

The proposed new test assumes a natural ordering of the components in the p-dimensional random vector, e.g., 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.

We have also conducted simulations to evaluate the performance of DLRT after reordering the components by the best permutation algorithm. Specifically, we first 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 $N_p(\boldsymbol{\mu}_2, \Sigma)$, where the common covariance matrix, Σ, follows the same setting as Section 4.1 of our main paper. 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. We consider two dependence structures including the independent (IND) structure and the short range dependence (SRD) structure. To represent different levels of correlation in the SRD structure, we consider two levels of correlation $\rho = 0.3$ and 0.6, respectively. For the power comparison, we set the jth nonzero component in $\boldsymbol{\mu}_2$ 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 $(n_1, n_2, \theta) \times p = \{(30, 30, 0.25)\} \times \{500 \text{ or } 800\}$, respectively. In addition, we let $\widetilde{X}_i = P\boldsymbol{X}_i$

and $\widetilde{Y}_k = P\boldsymbol{Y}_k$, where $P \in R^{p\times p}$ is a permutation matrix. Consequently, the components of $\widetilde{X}_i$ and $\widetilde{Y}_k$ do not follow a natural ordering.

Next, we apply the BP-DLRT, SD, PA, BP-GCT$_{\rm lg}$, CQ and RHT tests to the data $\{\widetilde{X}_i\}_{i=1}^{n_1}$ and $\{\widetilde{Y}_k\}_{k=1}^{n_2}$, where BP-DLRT and BP-GCT$_{\rm lg}$ are formed by first reordering $\widetilde{X}_i$ and $\widetilde{Y}_k$ with the best permutation method, and then applying the DLRT and GCT$_{\rm lg}$ to the reordered variables (note that GCT$_{\rm lg}$ also assumes a natural ordering).

The simulation results are shown in Table S2 with nominal level of 5%. When the correlation is weak, the type I error rates of BP-DLRT are still more close to the nominal level than the BP-GCT$_{\rm lg}$ test. The power of the BP-DLRT test is usually larger than the SD, PA, BP-GCT$_{\rm lg}$, CQ and RHT tests. When the correlation becomes stronger, we note that BP-GCT$_{\rm lg}$ and BP-GCT$_{\rm lg}$ both have inflated type I error rates.

In conclusion, when the natural ordering among the components is unknown, the BP-DLRT test can be recommended for the un-ordered variables. In particular, when the correlation among the components are not strong, the BP-DLRT test is able to provide a satisfactory power performance.

[Table 2 about here.]

D. Properties of the DLRT, SD, PA, GCT, CQ and RHT tests

D.1 *A summary of asymptotic power*

In this paper we focus on the testing problem with dense but weak signals, the asymptotic power of the considered tests depend heavily on the total amount of the signals. To specify the necessary order of the respective tests, we have reported their asymptotic power in Table S3. In the third column of Table S3, we present the necessary order of the signal for each test.

Specifically, we assume $\kappa_0 = n_1n_2/(n_1+n_2)$, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are two population mean vectors,

Σ and R be the common covariance matrix and correlation matrix, respectively. Let $D_\sigma = \text{diag}(\sigma_{11}^2, \cdots, \sigma_{pp}^2)$ be the diagonal matrix of Σ, $\tau_D^2 = 2\pi f_d(0)$ and $\tau_G^2 = 2\pi f_g(0)$, where $f_d(\cdot)$ and $f_g(\cdot)$ are the spectrum density functions of $\{V_{Nj}\}_{j=1}^p$ and $\{t_{Nj}\}_{j=1}^p$, respectively. Let

$$H_\lambda = \frac{1}{1+\gamma\Theta_1(\lambda,\gamma)}\Sigma + \lambda I_p,$$

where

$$\Theta_1(\lambda,\gamma) = \frac{1-\lambda m_F(-\lambda)}{1-\gamma\{1-\lambda m_F(-\lambda)\}},$$

with $\gamma = p/n$, $\lambda > 0$, and $m_F(z)$ is the solution of Marchenko–Pastur equation (Chen et al., 2011). In addition, let

$$\Theta_2(\lambda,\gamma) = \frac{1-\lambda m_F(-\lambda)}{[1-\gamma\{1-\lambda m_F(-\lambda)\}]^3} - \lambda\frac{m_F(-\lambda)-\lambda m_F'(-\lambda)}{[1-\gamma\{1-\lambda m_F(-\lambda)\}]^4},$$

where $m_F'(z)$ is the derivative of $m_F(z)$.

The asymptotic power and the necessary order of the signal for each test are reported in Table S3.

[Table 3 about here.]

D.2 *A summary of the transformation invariance properties*

The transformation invariant is a desirable property when constructing the test statistics. To cater for demand, let $\boldsymbol{X} \in R^p$ is a random vector, we consider the following three types of transformation:

(1) the orthogonal transformation: $\boldsymbol{X} \to \Gamma\boldsymbol{X}$, where $\Gamma^T\Gamma = I$ is an identity matrix;
(2) the scale transformation: $\boldsymbol{X} \to D\boldsymbol{X}$, where $D = \text{diag}(d_1, \ldots, d_p)$ without any zero entries;
(3) the shift transformation: $\boldsymbol{X} \to \boldsymbol{d}_0 + \boldsymbol{X}$, where $\boldsymbol{d}_0 \in R^p$ is a constant vector.

We summarize the transformation invariance properties of the DLRT, SD, PA, GCT, CQ and RHT tests into Table S4. Interestingly, none of the six tests possesses all the three invariance properties. Future research may be warranted to investigate whether there exists a new test that possesses all the three invariance properties.

[Table 4 about here.]

References

Bickel, P. J. and Levina, E. (2008a). Regularized estimation of large covariance matrices. *The Annals of Statistics* **36,** 199–227.

Bickel, P. J. and Levina, E. (2008b). Covariance regularization by thresholding. *The Annals of Statistics* **36,** 2577–2604.

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. *Journal of the American Statistical Association* **106,** 1345–1360.

Gilbert, J. R., Moler, C., and Schreiber, R. (1992). Sparse matrices in MATLAB: design and implementation. *SIAM Journal on Matrix Analysis and Applications* **13,** 333–356.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

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

Rajaratnam, B., Massam, H., and Carvalho, C. M. (2008). Flexible covariance estimation in graphical Gaussian models. *The Annals of Statistics* **36,** 2818–2849.

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

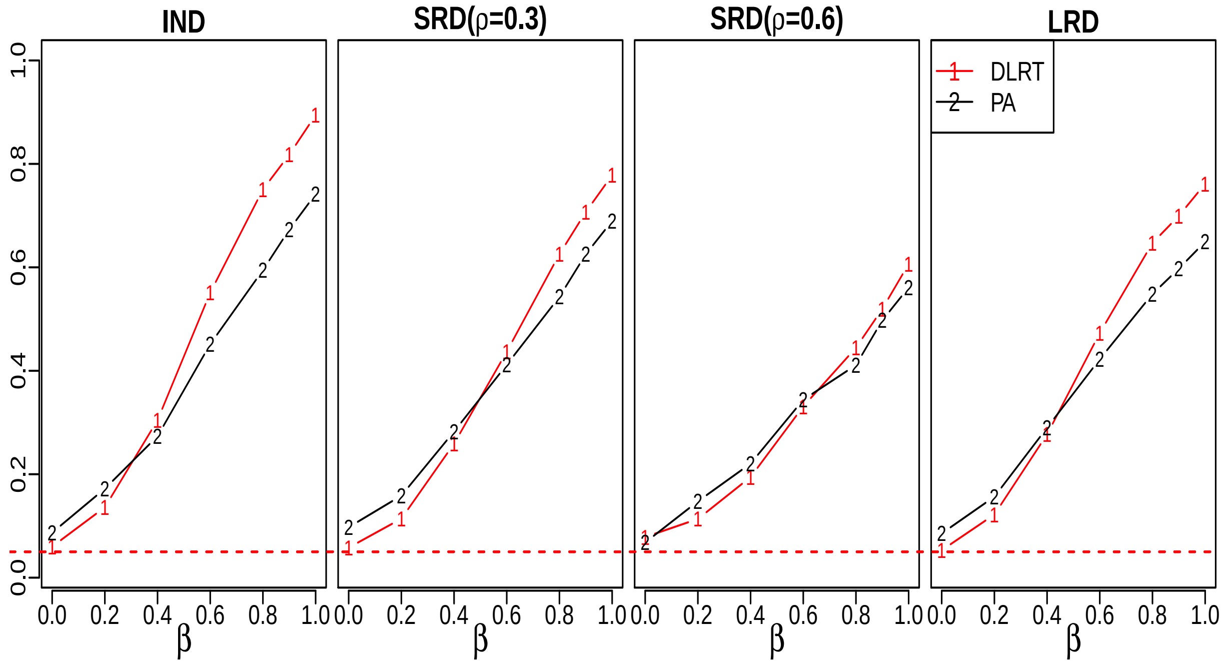
Rothman, A. J., Levina, E., and Zhu, J. (2010). A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika* **97,** 539–550.

Spira, R. (1971). Calculation of the gamma function by Stirling's formula. *Mathematics of Computation* **25,** 317–322.

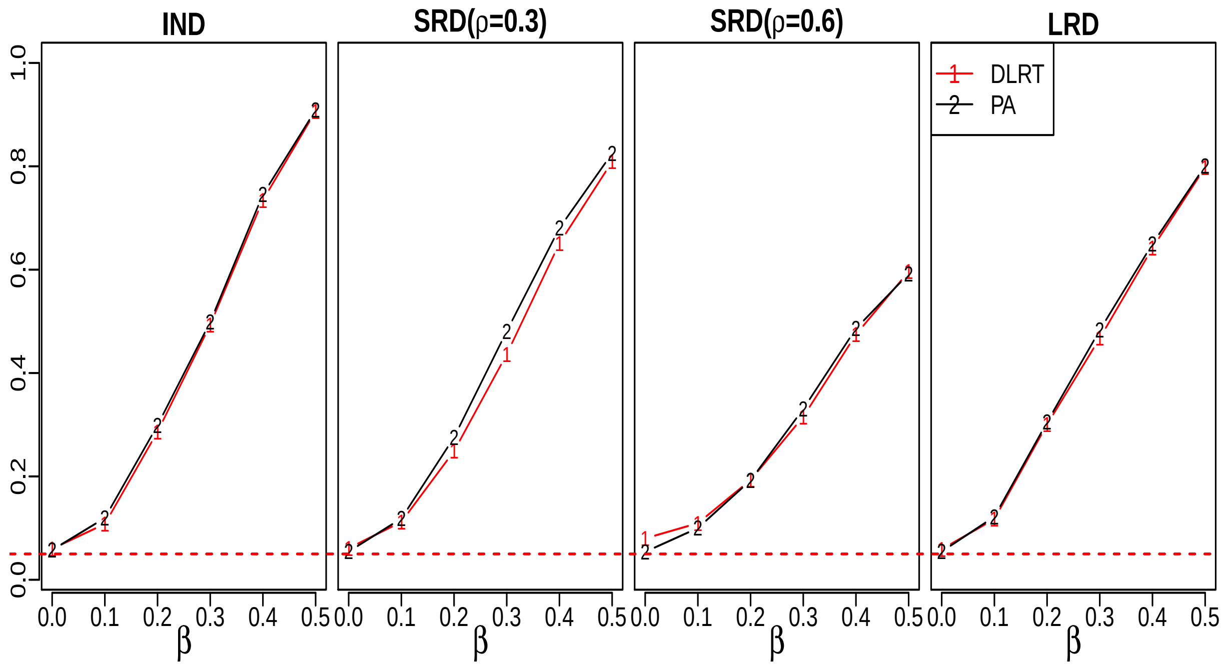
Wagaman, A. and Levina, E. (2009). Discovering sparse covariance structures with the

Isomap. *Journal of Computational and Graphical Statistics* **18,** 551–572.

Zhu, D. and Galbraith, J. W. (2010). A generalized asymmetric Student-t distribution with application to financial econometrics. *Journal of Econometrics* **157,** 297–305.

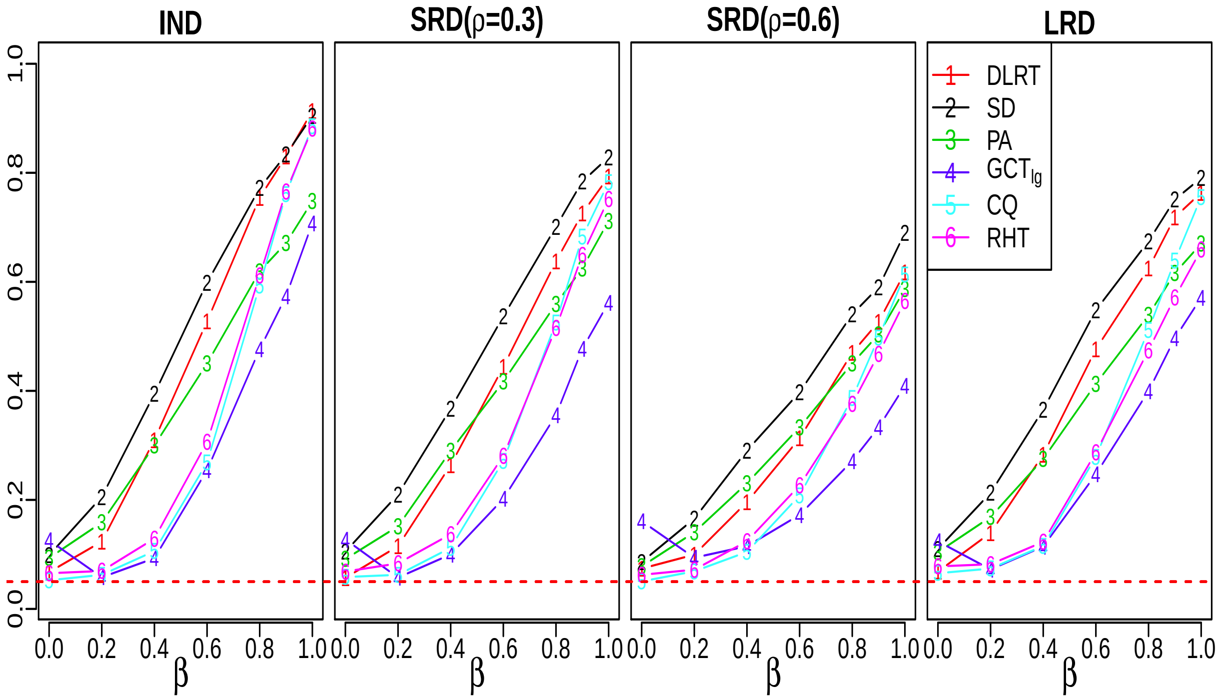


(a) $n_1 = n_2 = 5, p = 100$

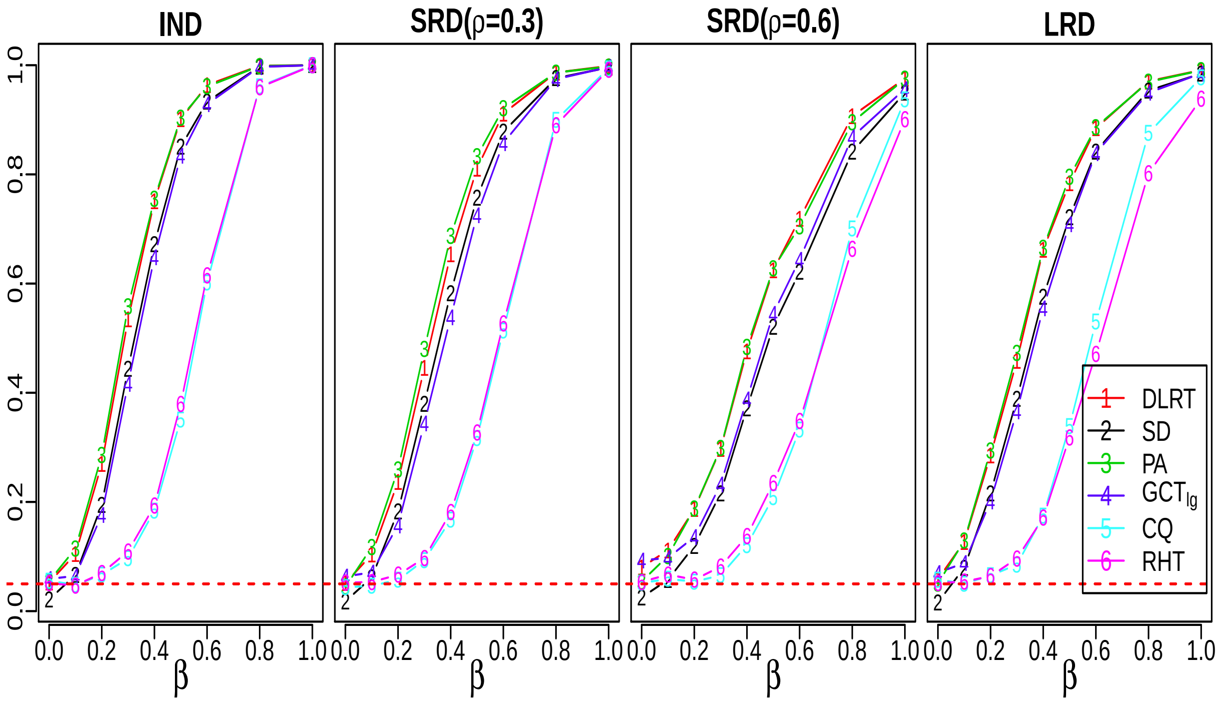


(b) $n_1 = n_2 = 15, p = 500$

Figure S1: Power comparisons between the DLRT and PA 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 nominal level of $\alpha = 0.05$. The results are based on 2000 simulations with data from the normal distribution.



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



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

Figure S2: Power comparisons among the DLRT, SD, PA, $\mathrm{GCT_{lg}}$, 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 nominal level of $\alpha = 0.05$. The results are based on 2000 simulations with data from the normal distribution.

Table S1: Type I error rates over 2000 simulations for the DLRT and PA tests under the IND or SRD structure. The nominal level is $\alpha = 0.05$. In the SRD structure, we consider the correlation, $\rho = 0.6$.

		$n_1 = n_2 = 5$			$n_1 = n_2 = 15$		
Cov	Method	$p = 100$	$p = 300$	$p = 500$	$p = 100$	$p = 300$	$p = 500$
IND	DLRT	0.056	0.052	0.043	0.058	0.049	0.048
	PA	0.981	0.985	0.987	0.763	0.795	0.779
SRD $\rho = 0.6$	DLRT	0.076	0.080	0.072	0.078	0.081	0.078
	PA	0.969	0.980	0.967	0.601	0.649	0.631

Table S2: Power comparisons over 2000 simulations for the BP-DLRT, BP-GCT$_{\text{lg}}$, PA, SD, CQ and RHT tests under the IND or SRD structure, respectively. Two different correlation, $\rho = 0.3$ or 0.6, are considered for the SRD structure. The nominal level is 5%.

		$p = 500$			$p = 800$		
Cov	Method	$\beta = 0$	$\beta = 0.15$	$\beta = 0.30$	$\beta = 0$	$\beta = 0.15$	$\beta = 0.30$
IND	BP-DLRT	0.051	0.509	0.967	0.044	0.695	0.998
	SD	0.032	0.478	0.966	0.030	0.654	0.997
	PA	0.050	0.544	0.974	0.045	0.745	0.999
	BP-GCT$_{\text{lg}}$	0.063	0.393	0.947	0.045	0.612	0.995
	CQ	0.049	0.466	0.920	0.051	0.568	0.979
	RHT	0.050	0.434	0.905	0.054	0.566	0.979
SRD ($\rho = 0.3$)	BP-DLRT	0.064	0.536	0.937	0.071	0.717	0.993
	SD	0.039	0.430	0.896	0.032	0.585	0.986
	PA	0.048	0.499	0.924	0.061	0.678	0.993
	BP-GCT$_{\text{lg}}$	0.081	0.430	0.908	0.077	0.637	0.991
	CQ	0.049	0.426	0.844	0.056	0.544	0.966
	RHT	0.064	0.381	0.791	0.058	0.502	0.958
SRD ($\rho = 0.6$)	BP-DLRT	0.089	0.311	0.722	0.082	0.486	0.913
	SD	0.042	0.251	0.690	0.039	0.035	0.837
	PA	0.055	0.343	0.748	0.053	0.446	0.901
	BP-GCT$_{\text{lg}}$	0.112	0.265	0.678	0.093	0.419	0.889
	CQ	0.054	0.301	0.690	0.056	0.359	0.836
	RHT	0.056	0.235	0.546	0.045	0.327	0.765

Table S3: Asymptotic power and the necessary order of signal for the DLRT, SD, PA, GCT, CQ and RHT tests.

test method	asymptotic power	the necessary order of signal
DLRT	$\Phi\Big(-z_\alpha+\kappa_0\frac{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T D_\sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)}{\sqrt{p\tau_D^2}}\Big)$	$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T D_\sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)=O(\sqrt{p\tau_D^2})$
SD	$\Phi\Big(-z_\alpha+\kappa_0\frac{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T D_\sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)}{\sqrt{2\mathrm{tr}(R^2)}}\Big)$	$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T D_\sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)=O(\sqrt{2\mathrm{tr}(R^2)})$
PA	$\Phi\Big(-z_\alpha+\kappa_0\frac{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T D_\sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)}{\sqrt{2\mathrm{tr}(R^2)}}\Big)$	$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T D_\sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)=O(\sqrt{2\mathrm{tr}(R^2)})$
GCT	$\Phi\Big(-z_\alpha+\kappa_0\frac{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T D_\sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)}{\sqrt{p\tau_G^2}}\Big)$	$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T D_\sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)=O(\sqrt{p\tau_G^2})$
CQ	$\Phi\Big(-z_\alpha+\kappa_0\frac{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T D_\sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)}{\sqrt{2\mathrm{tr}(\Sigma^2)}}\Big)$	$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T D_\sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)=O(\sqrt{2\mathrm{tr}(\Sigma^2)})$
RHT	$\Phi\Big(-z_\alpha+\kappa_0\frac{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T H_\lambda^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)}{\sqrt{2p\Theta_2(\lambda,\gamma)}}\Big)$	$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T H_\lambda^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)=O(\sqrt{2p\Theta_2(\lambda,\gamma)})$

Table S4: The transformation invariance properties of the DLRT, SD, PA, GCT, CQ and RHT tests

	DLRT	SD	PA	GCT	CQ	RHT
orthogonal invariance	No	No	No	No	Yes	Yes
scale invariance	Yes	Yes	Yes	Yes	No	No
shift invariance	Yes	Yes	No	Yes	No	Yes