Supplement to "Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing"

Xiufan Yu[1], Danning Li[2], Lingzhou Xue[3], and Runze Li[3]

[1]University of Notre Dame, [2]Northeast Normal University, and

[3]The Pennsylvania State University

In this supplement, we first present technical lemmas and complete proofs of the lemmas and theorems. We also present additional numerical results with regard to non-Gaussian data, various dependent structure, and comparisons with extreme-value-based tests, to serve as a complementary to the simulation studies in Section 5.

S.1 Lemmas

Lemma 1. Suppose $n_1/(n_1+n_2) \to \gamma$ for some constant $\gamma \in (0,1)$ as $\min\{n_1, n_2\} \to \infty$. Given Assumptions 1-2, under H_0, as $n_1, n_2, p \to \infty$,

$$\mathrm{Cov}\left(\frac{M_{n_1,n_2}}{\sigma_{01}}, \frac{\widetilde{T}_{n_1,n_2}}{\sigma_{02}}\right) = o(1), \tag{1}$$

where $\widetilde{T}_{n_1,n_2}$ are defined by (27).

Lemma 2. Suppose $n_1/(n_1+n_2) \to \gamma$ for some constant $\gamma \in (0,1)$ as $\min\{n_1, n_2\} \to \infty$. With Assumption 2, as $n_1, n_2 \to \infty$,

(i) if $\mu_{1i} = \mu_{2i}$,

$$\frac{\sqrt{2} M_i}{\nu_i^{1/2}} + 1 \xrightarrow{d} \chi_1^2, \tag{2}$$

where $\nu_i = \frac{2}{n_1(n_1-1)}\sigma_{1,ii}^2 + \frac{2}{n_2(n_2-1)}\sigma_{2,ii}^2 + \frac{4}{n_1 n_2}\sigma_{1,ii}\sigma_{2,ii}$.

(ii) if $\mu_{1i} \neq \mu_{2i}$,

$$\frac{M_i - (\mu_{1i} - \mu_{2i})^2}{\sqrt{4(\mu_{1i} - \mu_{2i})^2 \left(\frac{\sigma_{1,ii}}{n_1} + \frac{\sigma_{2,ii}}{n_2}\right)}} \xrightarrow{d} N(0,1). \tag{3}$$

Lemma 3. Suppose $n_1/(n_1+n_2) \to \gamma$ for some constant $\gamma \in (0,1)$ as $\min\{n_1, n_2\} \to \infty$. With Assumption 2, as $n_1, n_2 \to \infty$,

(i) if $\sigma_{1,ij} = \sigma_{2,ij}$,

$$\frac{\sqrt{2} T_{ij}}{\xi_{ij}^{1/2}} + 1 \xrightarrow{d} \chi_1^2, \tag{4}$$

where

$$\xi_{ij} = 2\Bigg(\frac{1}{n_1}\left(\sigma_{1,ij}^2 + \sigma_{1,ii}\sigma_{1,jj} + \Delta_1 \mathrm{tr}(\gamma_{1i}\gamma_{1j}^T \circ \gamma_{1i}\gamma_{1j}^T)\right) + \frac{1}{n_2}\left(\sigma_{2,ij}^2 + \sigma_{2,ii}\sigma_{2,jj} + \Delta_2 \mathrm{tr}(\gamma_{2i}\gamma_{2j}^T \circ \gamma_{2i}\gamma_{2j}^T)\right)\Bigg)^2 (1 + o(1)).$$

(ii) if $\sigma_{1,ij} \neq \sigma_{2,ij}$,

$$\frac{T_{ij} - (\sigma_{1,ij} - \sigma_{2,ij})^2}{\psi_{ij}^{1/2}} \xrightarrow{d} N(0,1), \tag{5}$$

where

$$\psi_{ij} = 4(\sigma_{1,ij} - \sigma_{2,ij})^2 \Bigg(\frac{1}{n_1}\left(\sigma_{1,ij}^2 + \sigma_{1,ii}\sigma_{1,jj} + \Delta_1 \mathrm{tr}(\gamma_{1i}\gamma_{1j}^T \circ \gamma_{1i}\gamma_{1j}^T)\right) + \frac{1}{n_2}\left(\sigma_{2,ij}^2 + \sigma_{2,ii}\sigma_{2,jj} + \Delta_2 \mathrm{tr}(\gamma_{2i}\gamma_{2j}^T \circ \gamma_{2i}\gamma_{2j}^T)\right)\Bigg).$$

S.2 Proofs

S.2.1 Proof of Lemma 1

Proof. Without loss of generality, we assume the common mean vector under H_0 is $\boldsymbol{\mu} = \mathbf{0}$. We rewrite the statistics M_{n_1,n_2} and $\widetilde{T}_{n_1,n_2}$ into the form of two-sample U-statistics as follows.

$$M_{n_1,n_2} = \frac{4}{n_1(n_1-1)n_2(n_2-1)} \sum_{u<v}^{n_1} \sum_{k<l}^{n_2} H_1(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l), \tag{6}$$

and

$$\widetilde{T}_{n_1,n_2} = \frac{4}{n_1(n_1-1)n_2(n_2-1)} \sum_{u<v}^{n_1} \sum_{k<l}^{n_2} H_2(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l). \tag{7}$$

where

$$H_1(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l) = (\mathbf{X}_u'\mathbf{X}_v) + (\mathbf{Y}_k'\mathbf{Y}_l) - \frac{(\mathbf{X}_u'\mathbf{Y}_k) + (\mathbf{X}_v'\mathbf{Y}_l) + (\mathbf{X}_u'\mathbf{Y}_l) + (\mathbf{X}_v'\mathbf{Y}_k)}{2},$$

and

$$H_2(\mathbf{X}_u, \mathbf{X}_v, \mathbf{Y}_k, \mathbf{Y}_l) = (\mathbf{X}_u'\mathbf{X}_v)^2 + (\mathbf{Y}_k'\mathbf{Y}_l)^2 - \frac{(\mathbf{X}_u'\mathbf{Y}_k)^2 + (\mathbf{X}_v'\mathbf{Y}_l)^2 + (\mathbf{X}_u'\mathbf{Y}_l)^2 + (\mathbf{X}_v'\mathbf{Y}_k)^2}{2}.$$

To simplify notation, we define $\mathbf{A} = \boldsymbol{\Gamma}_1^T\boldsymbol{\Gamma}_1 = (a_{ij})_{1\le i,j\le m_1}$, $\mathbf{B} = \boldsymbol{\Gamma}_2^T\boldsymbol{\Gamma}_2 = (b_{ij})_{1\le i,j\le m_2}$, and $\mathbf{D} = \boldsymbol{\Gamma}_1^T\boldsymbol{\Gamma}_2 = (d_{ij})_{1\le i\le m_1, 1\le j\le m_2}$. Recall that under the null hypothesis, we have $\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1^T = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}$, and $\boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_2^T = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$. The covariance of the two U-statistics $M_{n_1,n_2}/\sigma_{01}$ and $\widetilde{T}_{n_1,n_2}/\sigma_{02}$ can be obtained by

$$\mathrm{Cov}\left(\frac{M_{n_1,n_2}}{\sigma_{01}}, \frac{\widetilde{T}_{n_1,n_2}}{\sigma_{02}}\right) = \binom{n_1}{2}^{-1}\binom{n_2}{2}^{-1} \sum_{0\le i,j\le 2} \binom{2}{i}\binom{n_1-2}{2-i}\binom{2}{j}\binom{n_2-2}{2-j} \mathrm{cov}_{(i,j)}, \tag{8}$$

where $\mathrm{cov}_{(i,j)} = \sigma_{01}^{-1}\sigma_{02}^{-1}\,\mathrm{Cov}\left(H_1(\mathbf{X}_{u_1}, \mathbf{X}_{v_1}, \mathbf{Y}_{k_1}, \mathbf{Y}_{l_1}), H_2(\mathbf{X}_{u_2}, \mathbf{X}_{v_2}, \mathbf{Y}_{k_2}, \mathbf{Y}_{l_2})\right)$ with i being the number of integers common to (u_1, v_1) and (u_2, v_2), and j being the number of integers common to (k_1, l_1) and (k_2, l_2). After simple calculation, we have

$$
\begin{aligned}
\mathrm{cov}_{(0,0)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_3, \mathbf{X}_4, \mathbf{Y}_3, \mathbf{Y}_4)\right) = 0, \\
\mathrm{cov}_{(1,0)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_1, \mathbf{X}_3, \mathbf{Y}_3, \mathbf{Y}_4)\right) = 0, \\
\mathrm{cov}_{(0,1)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_3, \mathbf{X}_4, \mathbf{Y}_1, \mathbf{Y}_3)\right) = 0, \\
\mathrm{cov}_{(1,1)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_1, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_3)\right) \\
&= \frac{\sigma_{01}^{-1}\sigma_{02}^{-1}}{4} \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} d_{ij}^3 E z_{11i}^3 E z_{21j}^3, \\
\mathrm{cov}_{(2,0)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_3, \mathbf{Y}_4)\right) \\
&= \sigma_{01}^{-1}\sigma_{02}^{-1} \sum_{i,j=1}^{m_1} a_{ij}^3 E z_{11i}^3 E z_{12j}^3, \\
\mathrm{cov}_{(0,2)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_3, \mathbf{X}_4, \mathbf{Y}_1, \mathbf{Y}_2)\right) \\
&= \sigma_{01}^{-1}\sigma_{02}^{-1} \sum_{i,j=1}^{m_2} b_{ij}^3 E z_{21i}^3 E z_{22j}^3, \\
\mathrm{cov}_{(2,1)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_3)\right) \\
&= \sigma_{01}^{-1}\sigma_{02}^{-1} \left(\sum_{i,j=1}^{m_1} a_{ij}^3 E z_{11i}^3 E z_{12j}^3 + \frac{1}{2} \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} d_{ij}^3 E z_{11i}^3 E z_{21j}^3 \right), \\
\mathrm{cov}_{(1,2)} &= \sigma_{01}^{-1}\sigma_{02}^{-1}\ \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_1, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_2)\right) \\
&= \sigma_{01}^{-1}\sigma_{02}^{-1} \left(\sum_{i,j=1}^{m_2} b_{ij}^3 E z_{21i}^3 E z_{22j}^3 + \frac{1}{2} \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} d_{ij}^3 E z_{11i}^3 E z_{21j}^3 \right), \\
\mathrm{cov}_{(2,2)} &= \sigma_{01}^{-1}\sigma_{02}^{-1} \mathrm{Cov}\left(H_1(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2), H_2(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2)\right) \\
&= \sigma_{01}^{-1}\sigma_{02}^{-1} \left(\sum_{i,j=1}^{m_1} a_{ij}^3 E z_{11i}^3 E z_{12j}^3 + \sum_{i,j=1}^{m_2} b_{ij}^3 E z_{21i}^3 E z_{22j}^3 + \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} d_{ij}^3 E z_{11i}^3 E z_{21j}^3 \right).
\end{aligned}
$$

Therefore, (8) becomes

$$
\left| \mathrm{Cov}\left(\frac{M_{n_1,n_2}}{\sigma_{01}}, \frac{\widetilde{T}_{n_1,n_2}}{\sigma_{02}} \right) \right| \leq \frac{C}{n_1 n_2 \sigma_{01}\sigma_{02}} \left| \sum_{i,j=1}^{m_1} a_{ij}^3 + \sum_{i,j=1}^{m_2} b_{ij}^3 + \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} d_{ij}^3 \right|, \tag{9}
$$

where C is a sufficiently large positive constant. What's more,

$$\left|\sum_{i,j=1}^{m_1} a_{ij}^3\right| \le \max_{1\le i,j\le m_1} |a_{ij}| \cdot \sum_{i,j=1}^{m} a_{ij}^2 \le \lambda_{\max}^{\frac{1}{2}}(\mathbf{A}^T\mathbf{A}) \cdot \mathrm{tr}(\mathbf{A}^2) = \lambda_{\max}(\mathbf{\Sigma}) \cdot \mathrm{tr}(\mathbf{\Sigma}^2).$$

Similarly,

$$\left|\sum_{i,j=1}^{m_2} b_{ij}^3\right| \le \lambda_{\max}(\mathbf{\Sigma}) \cdot \mathrm{tr}(\mathbf{\Sigma}^2), \quad \left|\sum_{i=1}^{m_1}\sum_{j=1}^{m_2} d_{ij}^3\right| \le \lambda_{\max}(\mathbf{\Sigma}) \cdot \mathrm{tr}(\mathbf{\Sigma}^2).$$

Further by Assumption 1, we have $\lambda_{\max}(\mathbf{\Sigma}) = o(\mathrm{tr}^{\frac{1}{2}}(\mathbf{\Sigma}^2))$, the proof of Lemma 1 is complete.

□

S.2.2 Proof of Lemma 2

Proof. By definition, it is straightforward to show that

$$\begin{aligned}\mathrm{var}(M_i) =& \frac{2}{n_1(n_1-1)}\sigma_{1,ii}^2 + \frac{2}{n_2(n_2-1)}\sigma_{2,ii}^2 + \frac{4}{n_1 n_2}\sigma_{1,ii}\sigma_{2,ii} \\ &+ \frac{4}{n_1}(\mu_{1i}-\mu_{2i})^2\sigma_{1,ii} + \frac{4}{n_2}(\mu_{1i}-\mu_{2i})^2\sigma_{2,ii}.\end{aligned}$$

Note that M_i can be re-written as follows.

$$\begin{aligned}M_i =& \frac{1}{n_1(n_1-1)}\sum_{u\neq v}^{n_1} X_{ui}X_{vi} + \frac{1}{n_2(n_2-1)}\sum_{u\neq v}^{n_2} Y_{ui}Y_{vi} - \frac{2}{n_1 n_2}\sum_{u}^{n_1}\sum_{v}^{n_2} X_{ui}Y_{vi} \\ =& \frac{1}{n_1(n_1-1)}\sum_{u\neq v}^{n_1}(X_{ui}-\mu_{1i})(X_{vi}-\mu_{1i}) + \frac{1}{n_2(n_2-1)}\sum_{u\neq v}^{n_2}(Y_{ui}-\mu_{2i})(Y_{vi}-\mu_{2i}) \\ &- \frac{2}{n_1 n_2}\sum_{u}^{n_1}\sum_{v}^{n_2}(X_{ui}-\mu_{1i})(Y_{vi}-\mu_{2i}) + \frac{2}{n_1}\sum_{u}^{n_1}(\mu_{1i}-\mu_{2i})(X_{ui}-\mu_{1i}) \\ &+ \frac{2}{n_2}\sum_{v}^{n_2}(\mu_{2i}-\mu_{1i})(Y_{vi}-\mu_{2i}) + (\mu_{1i}-\mu_{2i})^2 \\ :=& M_{i,0} + \frac{2}{n_1}\sum_{u}^{n_1}(\mu_{1i}-\mu_{2i})(X_{ui}-\mu_{1i}) + \frac{2}{n_2}\sum_{v}^{n_2}(\mu_{2i}-\mu_{1i})(Y_{vi}-\mu_{2i}) + (\mu_{1i}-\mu_{2i})^2,\end{aligned} \tag{10}$$

where

$$M_{i,0} := \frac{1}{n_1(n_1-1)}\sum_{u\neq v}^{n_1}(X_{ui}-\mu_{1i})(X_{vi}-\mu_{1i}) + \frac{1}{n_2(n_2-1)}\sum_{u\neq v}^{n_2}(Y_{ui}-\mu_{2i})(Y_{vi}-\mu_{2i})$$
$$-\frac{2}{n_1 n_2}\sum_{u}^{n_1}\sum_{v}^{n_2}(X_{ui}-\mu_{1i})(Y_{vi}-\mu_{2i}). \tag{11}$$

We first show that $\sqrt{2}M_{i,0}\nu_i^{-1/2}+1 \xrightarrow{d} \chi_1^2$ as $n_1, n_2 \to \infty$. Without loss of generality, we consider $\mu_{1i}=\mu_{2i}=0$ in $M_{i,0}$. Therefore, (11) becomes

$$\begin{aligned}
&M_{i,0}\\
&= \frac{1}{n_1(n_1-1)}\sum_{u\neq v}^{n_1}X_{ui}X_{vi} + \frac{1}{n_2(n_2-1)}\sum_{u\neq v}^{n_2}Y_{ui}Y_{vi} - \frac{2}{n_1n_2}\sum_{u}^{n_1}\sum_{v}^{n_2}X_{ui}Y_{vi}\\
&= \frac{\sigma_{1,ii}}{n_1-1}\left[\left(\frac{1}{\sqrt{n_1}}\sum_{u}^{n_1}\frac{X_{ui}}{\sigma_{1,ii}^{1/2}}\right)^2 - \frac{1}{n_1}\sum_{u}^{n_1}\frac{X_{ui}^2}{\sigma_{1,ii}}\right] + \frac{\sigma_{2,ii}}{n_2-1}\left[\left(\frac{1}{\sqrt{n_2}}\sum_{u}^{n_2}\frac{Y_{ui}}{\sigma_{2,ii}^{1/2}}\right)^2 - \frac{1}{n_2}\sum_{u}^{n_2}\frac{Y_{ui}^2}{\sigma_{2,ii}}\right]\\
&\quad - \frac{2\sigma_{1,ii}^{1/2}\sigma_{2,ii}^{1/2}}{\sqrt{n_1n_2}}\cdot\frac{1}{\sqrt{n_1}}\sum_{u}^{n_1}\frac{X_{ui}}{\sigma_{1,ii}^{1/2}}\cdot\frac{1}{\sqrt{n_2}}\sum_{v}^{n_2}\frac{Y_{vi}}{\sigma_{2,ii}^{1/2}}\\
&= \left(\frac{\sigma_{1,ii}^{1/2}}{\sqrt{n_1}}\cdot\frac{1}{\sqrt{n_1}}\sum_{u}^{n_1}\frac{X_{ui}}{\sigma_{1,ii}^{1/2}} - \frac{\sigma_{2,ii}^{1/2}}{\sqrt{n_2}}\cdot\frac{1}{\sqrt{n_2}}\sum_{u}^{n_2}\frac{Y_{ui}}{\sigma_{2,ii}^{1/2}}\right)^2 - \frac{\sigma_{1,ii}}{n_1-1}\cdot\frac{1}{n_1}\sum_{u}^{n_1}\frac{X_{ui}^2}{\sigma_{1,ii}}\\
&\quad - \frac{\sigma_{2,ii}}{n_2-1}\cdot\frac{1}{n_2}\sum_{u}^{n_2}\frac{Y_{ui}^2}{\sigma_{2,ii}} + \frac{\sigma_{1,ii}}{n_1(n_1-1)}\left(\frac{1}{\sqrt{n_1}}\sum_{u}^{n_1}\frac{X_{ui}}{\sigma_{1,ii}^{1/2}}\right)^2 + \frac{\sigma_{2,ii}}{n_2(n_2-1)}\left(\frac{1}{\sqrt{n_2}}\sum_{u}^{n_2}\frac{Y_{ui}}{\sigma_{2,ii}^{1/2}}\right)^2.
\end{aligned} \tag{12}$$

By the central limit theorem and the law of large numbers, as $n_1, n_2 \to \infty$,

$$\frac{1}{\sqrt{n_1}}\sum_{u}^{n_1}\frac{X_{ui}}{\sigma_{1,ii}^{1/2}} \xrightarrow{d} N(0,1), \quad \frac{1}{\sqrt{n_2}}\sum_{u}^{n_2}\frac{Y_{ui}}{\sigma_{2,ii}^{1/2}} \xrightarrow{d} N(0,1), \quad \frac{1}{n_1}\sum_{u}^{n_1}\frac{X_{ui}^2}{\sigma_{1,ii}} \xrightarrow{p} 1, \quad \frac{1}{n_2}\sum_{u}^{n_2}\frac{Y_{ui}^2}{\sigma_{2,ii}} \xrightarrow{p} 1.$$

What's more,

$$\frac{\sqrt{2}\sigma_{1,ii}}{n_1\nu_i^{1/2}} = \frac{\sqrt{2}\sigma_{1,ii}}{n_1\cdot\sqrt{2}\left(\frac{\sigma_{1,ii}}{n_1}+\frac{\sigma_{2,ii}}{n_2}\right)(1+o(1))} \to \frac{\sigma_{1,ii}}{\sigma_{1,ii}+\frac{1}{1-\gamma}\sigma_{2,ii}},$$

$$\frac{\sqrt{2}\sigma_{2,ii}}{n_2\nu_i^{1/2}} = \frac{\sqrt{2}\sigma_{2,ii}}{n_2 \cdot \sqrt{2}\left(\frac{\sigma_{1,ii}}{n_1} + \frac{\sigma_{2,ii}}{n_2}\right)(1+o(1))} \to \frac{\sigma_{2,ii}}{(1-\gamma)\sigma_{1,ii} + \sigma_{2,ii}}.$$

Also because $\{X_{ui}\}_{u=1}^{n_1}$ and $\{Y_{vi}\}_{v=1}^{n_2}$ are independent, and note that

$$\frac{\sigma_{1,ii}}{\sigma_{1,ii} + \frac{1}{1-\gamma}\sigma_{2,ii}} + \frac{\sigma_{2,ii}}{(1-\gamma)\sigma_{1,ii} + \sigma_{2,ii}} = 1, \tag{13}$$

we have $\sqrt{2}M_{i,0}\nu_i^{-1/2} + 1 \xrightarrow{d} \chi_1^2$ as $n_1, n_2 \to \infty$. According to (10), $M_i = M_{i,0}$ if $\mu_{1i} = \mu_{2i}$. We complete the proof of (2).

If $\mu_{1i} \neq \mu_{2i}$, based on (10),

$$M_i - (\mu_{1i} - \mu_{2i})^2 = M_{i,0} + \frac{2}{n_1}(\mu_{1i} - \mu_{2i})\sum_u^{n_1}(X_{ui} - \mu_{1i}) + \frac{2}{n_2}(\mu_{2i} - \mu_{1i})\sum_v^{n_2}(Y_{vi} - \mu_{2i})$$

$$= M_{i,0} + \frac{2\sigma_{1,ii}^{1/2}(\mu_{1i} - \mu_{2i})}{\sqrt{n_1}} \cdot \frac{1}{\sqrt{n_1}}\sum_u^{n_1}\frac{X_{ui} - \mu_{1i}}{\sigma_{1,ii}^{1/2}} + \frac{2\sigma_{2,ii}^{1/2}(\mu_{2i} - \mu_{1i})}{\sqrt{n_2}} \cdot \frac{1}{\sqrt{n_2}}\sum_v^{n_2}\frac{Y_{vi} - \mu_{2i}}{\sigma_{2,ii}^{1/2}}$$

Since

$$\frac{M_{i,0}}{\sqrt{4(\mu_{1i} - \mu_{2i})^2\left(\frac{\sigma_{1,ii}}{n_1} + \frac{\sigma_{2,ii}}{n_2}\right)}} = \frac{M_{i,0}}{\nu_i^{1/2}} \cdot \left(\frac{2\left(\frac{\sigma_{1,ii}}{n_1} + \frac{\sigma_{2,ii}}{n_2}\right)^2(1+o(1))}{4(\mu_{1i} - \mu_{2i})^2\left(\frac{\sigma_{1,ii}}{n_1} + \frac{\sigma_{2,ii}}{n_2}\right)}\right)^{1/2} \to 0,$$

$$\frac{\left|2\sigma_{1,ii}^{1/2}(\mu_{1i} - \mu_{2i})/\sqrt{n_1}\right|}{\sqrt{4(\mu_{1i} - \mu_{2i})^2\left(\frac{\sigma_{1,ii}}{n_1} + \frac{\sigma_{2,ii}}{n_2}\right)}} = \left(\frac{\sigma_{1,ii}}{n_1\left(\frac{\sigma_{1,ii}}{n_1} + \frac{\sigma_{2,ii}}{n_2}\right)}\right)^{1/2} \to \left(\frac{\sigma_{1,ii}}{\sigma_{1,ii} + \frac{1}{1-\gamma}\sigma_{2,ii}}\right)^{1/2},$$

$$\frac{\left|2\sigma_{2,ii}^{1/2}(\mu_{2i} - \mu_{1i})/\sqrt{n_2}\right|}{\sqrt{4(\mu_{1i} - \mu_{2i})^2\left(\frac{\sigma_{1,ii}}{n_1} + \frac{\sigma_{2,ii}}{n_2}\right)}} = \left(\frac{\sigma_{2,ii}}{n_2\left(\frac{\sigma_{1,ii}}{n_1} + \frac{\sigma_{2,ii}}{n_2}\right)}\right)^{1/2} \to \left(\frac{\sigma_{2,ii}}{(1-\gamma)\sigma_{1,ii} + \sigma_{2,ii}}\right)^{1/2},$$

and together with the central limit theorem,

$$\frac{M_i - (\mu_{1i} - \mu_{2i})^2}{\sqrt{4(\mu_{1i} - \mu_{2i})^2\left(\frac{\sigma_{1,ii}}{n_1} + \frac{\sigma_{2,ii}}{n_2}\right)}} \xrightarrow{d} N(0,1) \quad \text{as } n_1, n_2 \to \infty.$$

This finishes the proof of Lemma 2. $\square$

S.2.3 Proof of Lemma 3

Proof. As discussed by Li and Chen (2012), the third and fourth-moment summation terms in A_{ij}, B_{ij} and C_{ij} are of the smaller order than the leading second-moment terms. After centering each variable, removing those terms from T_{ij} would not affect its asymptotic behavior. As a result, without loss of generality, we assume $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = 0$ and study the simplified statistic

$$\widetilde{T}_{ij} = \frac{1}{n_1(n_1-1)}\sum_{u\neq v}^{n_1} X_{ui}X_{uj}X_{vi}X_{vj} + \frac{1}{n_2(n_2-1)}\sum_{u\neq v}^{n_2} Y_{ui}Y_{uj}Y_{vi}Y_{vj} - \frac{2}{n_1n_2}\sum_{u}^{n_1}\sum_{v}^{n_2} X_{ui}X_{uj}Y_{vi}Y_{vj}. \tag{14}$$

$\widetilde{T}_{ij}$ and T_{ij} share the same asymptotic properties.

By definition, it is straightforward to show that $\mathrm{var}(\widetilde{T}_{ij}) = \psi_{ij} + \xi_{ij}$. Following similar arguments as in the proof of Lemma 2, we first rewrite $\widetilde{T}_{ij}$ in the form of

$$\widetilde{T}_{ij} = T_{ij}^{(1)} + T_{ij}^{(2)}, \tag{15}$$

where

$$\begin{aligned} T_{ij}^{(1)} := & \frac{1}{n_1(n_1-1)}\sum_{u\neq v}^{n_1}(X_{ui}X_{uj}-\sigma_{1,ij})(X_{vi}X_{vj}-\sigma_{1,ij}) \\ & + \frac{1}{n_2(n_2-1)}\sum_{u\neq v}^{n_1}(Y_{ui}Y_{uj}-\sigma_{2,ij})(Y_{vi}Y_{vj}-\sigma_{2,ij}) \\ & - \frac{2}{n_1n_2}\sum_{u}^{n_1}\sum_{v}^{n_2}(X_{ui}X_{uj}-\sigma_{1,ij})(Y_{vi}Y_{vj}-\sigma_{2,ij}), \end{aligned} \tag{16}$$

and

$$T_{ij}^{(2)} := (\sigma_{1,ij}-\sigma_{2,ij})^2 + \frac{2}{n_1}(\sigma_{1,ij}-\sigma_{2,ij})\sum_{u=1}^{n_1}(X_{ui}X_{uj}-\sigma_{1,ij})$$
$$+ \ \frac{2}{n_2}(\sigma_{2,ij}-\sigma_{1,ij})\sum_{v=1}^{n_2}(Y_{vi}Y_{vj}-\sigma_{2,ij}). \tag{17}$$

Following the same procedure as in the proof of Lemma 2, we are able to prove

$$\sqrt{2}T_{ij}^{(1)}\xi_{ij}^{-1/2}+1 \xrightarrow{d} \chi_1^2$$

as $n_1, n_2 \to \infty$. When $\sigma_{1,ij} = \sigma_{2,ij}$, $\widetilde{T}_{ij} = T_{ij}^{(1)}$, which proves (4). If $\sigma_{1,ij} \neq \sigma_{2,ij}$, we have $\psi_{ij}^{-1/2}(T_{ij}^{(2)} - (\sigma_{1,ij}-\sigma_{2,ij})^2) \xrightarrow{d} N(0,1)$ and $\psi_{ij}^{-1/2}T_{ij}^{(1)} = (\xi_{ij}/\psi_{ij})^{1/2}\cdot T_{ij}^{(1)}\xi_{ij}^{-1/2} \xrightarrow{p} 0$ as $n_1, n_2 \to \infty$. Together with (15), we complete the proof of Lemma 3. □

S.2.4 Proof of Theorem 1

Proof. By Slutsky's theorem and Lemma 2, under the null hypothesis $H_{0m} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, we have $\sqrt{2}M_i\widehat{\nu}_i^{-1/2}+1 \xrightarrow{d} \chi_1^2$ as $n_1, n_2 \to \infty$. What's more, for each $i = 1, \cdots, p$, by Lemma 1 of Chen, Li and Zhong (2019) and Theorem 3 of Petrov (1954), we have

$$P\left(\sqrt{2}M_i\widehat{\nu}_i^{-1/2}+1 > \delta_{n,p}|H_{0m}\right) = 2(1-\Phi(\delta_p^{1/2}))(1+O(n^{-1/6})+O(\delta_p^{3/2}n^{-1/2})).$$

Therefore, with $\delta_p = 2\log p$ and assuming $\log p = o(n^{-1/3})$,

$$P\left(\sqrt{2}M_i\widehat{\nu}_i^{-1/2}+1 > \delta_{n,p}|H_{0m}\right) \leq \delta_p^{-1/2}\exp(-\log p)(1+o(1)) = o(p^{-1}).$$

As a result, as $n_1, n_2, p \to \infty$,

$$P\left(J_m = 0|H_{0m}\right) = P\left(\max_{1\leq i\leq p}\sqrt{2}M_i\widehat{\nu}_i^{-1/2}+1 \leq \delta_{n,p}|H_{0m}\right)$$
$$=1 - P\left(\max_{1\leq i\leq p}\sqrt{2}M_i\widehat{\nu}_i^{-1/2}+1 > \delta_{n,p}|H_{0m}\right)$$

$$\geq 1 - p \cdot \max_{1\leq i\leq p} P\left(\sqrt{2}M_i\widehat{\nu}_i^{-1/2} + 1 > \delta_{n,p}|H_{0m}\right) \to 1.$$

Chen and Qin (2010) proved that under the null hypothesis H_{0m}, $\widehat{\sigma}_{01}^{-1}\sum_{i=1}^{p} M_i \xrightarrow{d} N(0,1)$ as $n_1, n_2, p \to \infty$. As a result, under H_{0m},

$$M_{PE} = \frac{1}{\widehat{\sigma}_{01}}\sum_{i=1}^{p} M_i + J_m \xrightarrow{d} N(0,1) \quad \text{as } n_1, n_2, p \to \infty.$$

The proof of Theorem 1 is complete.

□

S.2.5 Proof of Theorem 2

Proof. By Slutsky's theorem and Lemma 3, under the null hypothesis $H_{0c} : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, we have $\sqrt{2}T_{ij}\widehat{\xi}_{ij}^{-1/2} + 1 \xrightarrow{d} \chi_1^2$ as $n_1, n_2 \to \infty$. Lemma 2 of Chen, Guo and Qiu (2019) proves that $P\left(\max_{1\leq i,j\leq p}\sqrt{2}T_{ij}\widehat{\xi}_{ij}^{-1/2} + 1 > 4\log p|H_{0c}\right) \to 0$ as $n_1, n_2, p \to \infty$. Therefore, with $\eta_p = 4\log p$,

$$P\left(J_c = 0|H_{0c}\right) = P\left(\max_{1\leq i,j\leq p}\sqrt{2}T_{ij}\widehat{\xi}_{ij}^{-1/2} + 1 \leq \eta_{n,p}|H_{0c}\right) \to 1 \quad \text{as } n_1, n_2, p \to \infty.$$

Li and Chen (2012) proved that under the null hypothesis H_{0c} , $\widehat{\sigma}_{02}^{-1}\sum_{i=1}^{p}\sum_{j=1}^{p} T_{ij} \xrightarrow{d} N(0,1)$ as $n_1, n_2, p \to \infty$. As a result, under H_{0c},

$$T_{PE} = \frac{1}{\widehat{\sigma}_{02}}\sum_{i=1}^{p}\sum_{j=1}^{p} T_{ij} + J_c \xrightarrow{d} N(0,1) \quad \text{as } n_1, n_2, p \to \infty.$$

The proof of Theorem 2 is complete.

□

S.2.6 Proof of Theorem 3

Proof. For the power enhancement properties of M_{PE},

$$\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d\cup\mathcal{G}_m^s} P\left(M_{PE}\geq z_\alpha\right)\geq \min\left\{\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d} P\left(M_{PE}\geq z_\alpha\right), \inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(M_{PE}\geq z_\alpha\right)\right\}.$$

By definition, $J_m=\sqrt{p}\sum_{i=1}^p M_i\widehat{\nu}_i^{-1/2}\mathcal{I}\{\sqrt{2}M_i\widehat{\nu}_i^{-1/2}+1>\delta_p\}\geq 0$ as long as n and p are not too small such that $\delta_p>1$. Hence, as $n_1,n_2,p\to\infty$,

$$\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d} P\left(M_{PE}\geq z_\alpha\right)\geq \inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d} P\left(\frac{1}{\widehat{\sigma}_{01}}\sum_{i=1}^p M_i\geq z_\alpha\right) \tag{18}$$

where the right-hand-side of (18) approaches 1 as shown in Chen and Qin (2010). It suffices to show that

$$\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(M_{PE}\geq z_\alpha\right)\to 1. \tag{19}$$

Let

$$S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)=\left\{1\leq i\leq p:\frac{(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}\geq 16\delta_p\right\}$$

and

$$\widehat{S}_m=\left\{1\leq i\leq p:\frac{\sqrt{2}M_i}{\widehat{\nu}_i^{1/2}}+1>\delta_p\right\},$$

then $J_m=\sqrt{p}\sum_{i=1}^p M_i\widehat{\nu}_i^{-1/2}\mathcal{I}\{\sqrt{2}M_i\widehat{\nu}_i^{-1/2}+1>\delta_p\}=\sqrt{p}\sum_{i\in\widehat{S}_m}M_i\widehat{\nu}_i^{-1/2}$. By definition, $\{J_m\leq\sqrt{p}(\delta_p-1)/\sqrt{2}\}=\{\widehat{S}_m=\varnothing\}$ and $\mathcal{G}_m^s=\{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2):S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\neq\varnothing\}$.

By the large deviation and $\log p=o(n^{1/5})$, for any $\epsilon>0$,

$$\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s}\max_{1\leq i\leq p} P\left(\left|\frac{\widehat{\nu}_i}{\nu_i}-1\right|\geq\epsilon\right)=o(p^{-1}).$$

Note that,

$$\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\subset\widehat{S}_m\right)=\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(\bigcap_{i\in S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)}\left\{\frac{\sqrt{2}M_i}{\widehat{\nu}_i^{1/2}}+1>\delta_p\right\}\right)$$

$$\geq \inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(\bigcap_{i\in S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)}\left\{\left\{\frac{M_i-(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}+\frac{(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}>\frac{\delta_p\sqrt{1+\epsilon}}{\sqrt{2}}\right\}\bigcap\left\{\left|\frac{\widehat{\nu}_i}{\nu_i}-1\right|\leq\epsilon\right\}\right\}\right)$$

$$\geq \inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s}\left\{P\left(\bigcap_{i\in S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)}\left\{\frac{M_i-(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}+\frac{(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}>\frac{\delta_p\sqrt{1+\epsilon}}{\sqrt{2}}\right\}\right)\right.$$
$$\left.-P\left(\bigcup_{i\in S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)}\left\{\left|\frac{\widehat{\nu}_i}{\nu_i}-1\right|\geq\epsilon\right\}\right)\right\}$$

$$\geq \inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(\bigcap_{i\in S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)}\left\{\frac{M_i-(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}>-\frac{(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}+\frac{\delta_p\sqrt{1+\epsilon}}{\sqrt{2}}\right\}\right)$$
$$-p\max_{1\leq i\leq p}P\left(\left|\frac{\widehat{\nu}_i}{\nu_i}-1\right|\geq\epsilon\right)$$

$$\geq 1-\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(\bigcup_{i\in S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)}\left\{\left|\frac{M_i-(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}\right|>\frac{(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}-\delta_p\frac{\sqrt{1+\epsilon}}{\sqrt{2}}\right\}\right)$$
$$-p\max_{1\leq i\leq p}P\left(\left|\frac{\widehat{\nu}_i}{\nu_i}-1\right|\geq\epsilon\right)$$

$$\geq 1-\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} p\cdot\max_{1\leq i\leq p}P\left(\left|\frac{M_i-(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}\right|>\frac{(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}-\delta_p\frac{\sqrt{1+\epsilon}}{\sqrt{2}}\right)$$
$$-p\max_{1\leq i\leq p}P\left(\left|\frac{\widehat{\nu}_i}{\nu_i}-1\right|\geq\epsilon\right).$$

Further that let $\epsilon<1$

$$\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s}\max_{1\leq i\leq p}P\left(\left|\frac{M_i-(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}\right|>\frac{(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}-\delta_p\frac{\sqrt{1+\epsilon}}{\sqrt{2}}\right)$$
$$\leq\max_{1\leq i\leq p}P\left(\left|\frac{M_{i,0}}{\nu_i^{1/2}}\right|>\delta_p\right)$$
$$+\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s}\max_{1\leq i\leq p}P\left(\left|\frac{2\sigma_{1,ii}^{1/2}(\mu_{1i}-\mu_{2i})}{\sqrt{n_1}\nu_i^{1/2}}\cdot\frac{1}{\sqrt{n_1}}\sum_u^{n_1}\frac{X_{ui}-\mu_{1i}}{\sigma_{1,ii}^{1/2}}\right|>\frac{1}{2}\left(\frac{(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}-\delta_p(1+\frac{\sqrt{1+\epsilon}}{\sqrt{2}})\right)\right)$$
$$+\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s}\max_{1\leq i\leq p}P\left(\left|\frac{2\sigma_{2,ii}^{1/2}(\mu_{2i}-\mu_{1i})}{\sqrt{n_2}\nu_i^{1/2}}\cdot\frac{1}{\sqrt{n_2}}\sum_v^{n_2}\frac{Y_{vi}-\mu_{2i}}{\sigma_{2,ii}^{1/2}}\right|>\frac{1}{2}\left(\frac{(\mu_{1i}-\mu_{2i})^2}{\nu_i^{1/2}}-\delta_p(1+\frac{\sqrt{1+\epsilon}}{\sqrt{2}})\right)\right)$$
$$\leq 2\exp(-\frac{\delta_p}{2})\frac{1}{\sqrt{\delta_p}}$$
$$+\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s}\max_{1\leq i\leq p}P\left(\left|\frac{1}{\sqrt{n_1}}\sum_u^{n_1}\frac{X_{ui}-\mu_{1i}}{\sigma_{1,ii}^{1/2}}\right|>\frac{\sqrt{n_1}v_i^{1/4}}{4\sigma_{1,ii}^{1/2}}\left(\frac{|\mu_{1i}-\mu_{2i}|}{\nu_i^{1/4}}-\frac{\delta_p^{1/2}}{4}(1+\frac{\sqrt{1+\epsilon}}{\sqrt{2}})\right)\right)$$

$$+\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s}\max_{1\le i\le p} P\left(\left|\frac{1}{\sqrt{n_2}}\sum_v^{n_2}\frac{Y_{vi}-\mu_{2i}}{\sigma_{2,ii}^{1/2}}\right| > \frac{\sqrt{n_2}v_i^{1/4}}{4\sigma_{2,ii}^{1/2}}\left(\frac{|\mu_{1i}-\mu_{2i}|}{\nu_i^{1/4}}-\frac{\delta_p^{1/2}}{4}(1+\frac{\sqrt{1+\epsilon}}{\sqrt{2}})\right)\right)$$

$$\le 2\exp(-\frac{\delta_p}{2})\frac{1}{\sqrt{\delta_p}}+\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s}\max_{1\le i\le p} P\left(\left|\frac{1}{\sqrt{n_1}}\sum_u^{n_1}\frac{X_{ui}-\mu_{1i}}{\sigma_{1,ii}^{1/2}}\right| > \frac{2^{1/4}}{4}\left(\delta_p^{1/2}(4-\frac{1}{2})\right)\right)$$

$$+\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s}\max_{1\le i\le p} P\left(\left|\frac{1}{\sqrt{n_2}}\sum_v^{n_2}\frac{Y_{vi}-\mu_{2i}}{\sigma_{2,ii}^{1/2}}\right| > \frac{2^{1/4}}{4}\left(\delta_p^{1/2}(4-\frac{1}{2})\right)\right) = o(p^{-1}).$$

It implies, $\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\subset\widehat{S}_m\right)\to 1$. Furthermore,

$$\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(J_m\le\sqrt{p}\cdot\frac{\delta_p-1}{\sqrt{2}}\right)=\sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(\widehat{S}_m=\varnothing\right)$$
$$=\sup_{\{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2):S_m(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\neq\varnothing\}} P\left(\widehat{S}_m=\varnothing\right)\to 0,$$

i.e., $\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(J_m>\sqrt{p}(\delta_p-1)/\sqrt{2}\right)\to 1$. Therefore, as $n_1,n_2,p\to\infty$,

$$\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(M_{PE}\ge z_\alpha\right)\ge\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^s} P\left(\sqrt{p}\cdot\frac{\delta_p-1}{\sqrt{2}}+\frac{1}{\widehat{\sigma}_{01}}\sum_{i=1}^p M_i\ge z_\alpha\right)\to 1.$$

(19) is proved. In terms of power enhancement for T_{PE},

$$\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^d\cup\mathcal{G}_c^s} P\left(T_{PE}\ge z_\alpha\right)\ge\min\left\{\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^d} P\left(T_{PE}\ge z_\alpha\right),\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(T_{PE}\ge z_\alpha\right)\right\}.$$

Let

$$S_c(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)=\left\{(i,j):\frac{(\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}}\ge 16\eta_p, 1\le i,j\le p\right\},$$

and

$$\widehat{S}_c=\left\{(i,j):\frac{\sqrt{2}T_{ij}}{\widehat{\xi}_{ij}^{1/2}}+1>\eta_p, 1\le i,j\le p\right\},$$

then $J_c=\sqrt{p}\sum_{i=1}^p\sum_{j=1}^p T_{ij}\widehat{\xi}_{ij}^{-1/2}\mathcal{I}\{\sqrt{2}T_{ij}\widehat{\xi}_{ij}^{-1/2}+1>\eta_p\}=\sqrt{p}\sum_{(i,j)\in\widehat{S}_c}T_{ij}\widehat{\xi}_{ij}^{-1/2}$. Note that

$J_c > 0$ so long as n and p are not too small such that $\eta_{n,p} > 1$. As $n_1, n_2, p \to \infty$,

$$\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^d} P\left(T_{PE} \ge z_\alpha\right) \ge \inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^d} P\left(\frac{1}{\widehat{\sigma}_{02}} \sum_{i=1}^{p}\sum_{j=1}^{p} T_{ij} \ge z_\alpha\right) \tag{20}$$

where the right-hand-side of (20) approaches 1 as shown in Li and Chen (2012). It remains to prove that

$$\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^s} P\left(T_{PE} \ge z_\alpha\right) \to 1 \quad \text{as } n_1, n_2, p \to \infty. \tag{21}$$

Similarly,

$$\begin{aligned}
&\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^s} P\left(S_c(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2) \subset \widehat{S}_c\right) = \inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^s} P\left(\bigcap_{(i,j)\in S_c(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)} \left\{\frac{\sqrt{2}T_{ij}}{\widehat{\xi}_{ij}^{1/2}} + 1 > \eta_p\right\}\right) \\
&\ge \inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^s} P\left(\left\{\bigcap_{(i,j)\in S_c(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)} \left\{\frac{T_{ij} - (\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}} + \frac{(\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}} > \frac{\eta_p\sqrt{1+\epsilon}}{\sqrt{2}}\right\} \bigcap \left\{\left|\frac{\widehat{\xi}_{ij}}{\xi_{ij}} - 1\right| \le \epsilon\right\}\right\}\right) \\
&\ge 1 - P\left(\max_{1\le i,j\le p}\left|\frac{T_{ij}^{(1)}}{\xi_{ij}^{1/2}}\right| > \eta_p\right) - p^2 \max_{1\le i,j\le p} P\left(\left|\frac{\widehat{\xi}_{ij}}{\xi_{ij}} - 1\right| \ge \epsilon\right) \\
&- \sup_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^s} \max_{1\le i,j\le p} P\left(\left|\frac{2(\sigma_{1,ij}-\sigma_{2,ij})}{n_1\xi_{ij}^{1/2}} \sum_{u=1}^{n_1}(X_{ui}X_{uj} - \sigma_{1,ij})\right| > \frac{1}{2}\left(\frac{(\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}} - \eta_p^{1/2}(1+\frac{\sqrt{1+\epsilon}}{\sqrt{2}})\right)\right) \\
&- \sup_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^s} \max_{1\le i,j\le p} P\left(\left|\frac{2(\sigma_{2,ij}-\sigma_{1,ij})}{n_2\xi_{ij}^{1/2}} \sum_{u=1}^{n_1}(Y_{ui}Y_{uj} - \sigma_{2,ij})\right| > \frac{1}{2}\left(\frac{(\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}} - \eta_p^{1/2}(1+\frac{\sqrt{1+\epsilon}}{\sqrt{2}})\right)\right).
\end{aligned}$$

Let $\epsilon < 1$. Together with $\left(\max_{1\le i,j\le p}\left|\frac{T_{ij}^{(1)}}{\xi_{ij}^{1/2}}\right| > \eta_p\right) \to 0$,

$$\begin{aligned}
&\sup_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^s} \max_{1\le i,j\le p} P\left(\left|\frac{2(\sigma_{1,ij}-\sigma_{2,ij})}{n_1\xi_{ij}^{1/2}} \sum_{u=1}^{n_1}(X_{ui}X_{uj} - \sigma_{1,ij})\right| > \frac{1}{2}\left(\frac{(\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}} - 2\eta_p^{1/2}\right)\right) \\
&\qquad = o(p^{-2}) \\
&\sup_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^s} \max_{1\le i,j\le p} P\left(\left|\frac{2(\sigma_{2,ij}-\sigma_{1,ij})}{n_2\xi_{ij}^{1/2}} \sum_{u=1}^{n_1}(Y_{ui}Y_{uj} - \sigma_{2,ij})\right| > \frac{1}{2}\left(\frac{(\sigma_{1,ij}-\sigma_{2,ij})^2}{\xi_{ij}^{1/2}} - 2\eta_p^{1/2}\right)\right) \\
&\qquad = o(p^{-2})
\end{aligned}$$

and by the similar proof of Lemma 3 from Cai et al. (2013), we have

$$\sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s}\max_{1\leq i,j\leq p} P\left(\left|\frac{\widehat{\xi}_{ij}}{\xi_{ij}}-1\right|\geq\epsilon\right)=o(p^{-2})\quad \text{for } \epsilon>0,$$

we have, $\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(S_c(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\subset\widehat{S}_c\right)\to 1$. Furthermore,

$$\begin{aligned}
&\sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(J_c\leq\sqrt{p}\cdot\frac{\eta_p-1}{\sqrt{2}}\right)=\sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(\widehat{S}_c=\varnothing\right)\\
=&\sup_{\{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2):S_c(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\neq\varnothing\}} P\left(\widehat{S}_c=\varnothing\right)\to 0,
\end{aligned}$$

i.e., $\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(J_c>\sqrt{p}(\eta_p-1)/\sqrt{2}\right)\to 1$. Therefore, as $n_1,n_2,p\to\infty$,

$$\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(T_{PE}\geq z_\alpha\right)\geq\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathcal{G}_c^s} P\left(\sqrt{p}\cdot\frac{\eta_p-1}{\sqrt{2}}+\frac{1}{\widehat{\sigma}_{02}}\sum_{i=1}^p\sum_{j=1}^p T_{ij}\geq z_\alpha\right)\to 1.$$

(21) is proved. The proof of Theorem 3 is complete. □

S.2.7 Proof of Theorem 4

Proof. From Theorems 1 and 2, under the null hypothesis H_0,

$$P\left(J_m=0\right)\to 1,\quad P\left(J_c=0\right)\to 1\quad \text{as } n_1,n_2,p\to\infty.$$

Therefore, for any $x_1,x_2\in\mathbb{R}$,

$$\begin{aligned}
&\left|P\left(M_{PE}\leq x_1,T_{PE}\leq x_2\right)-P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}}\leq x_1,\frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}}\leq x_2\right)\right|\\
\leq&\left|P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}}+J_m\leq x_1,\frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}}+J_c\leq x_2\right)-P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}}+J_m\leq x_1,\frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}}\leq x_2\right)\right|\\
&+\left|P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}}+J_m\leq x_1,\frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}}\leq x_2\right)-P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}}\leq x_1,\frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}}\leq x_2\right)\right|\\
\leq&\left|P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}}+J_m\leq x_1,\frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}}+J_c\leq x_2\middle|J_c=0\right)\right.\\
&\left.-P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}}+J_m\leq x_1,\frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}}\leq x_2\middle|J_c=0\right)\right|\times P\left(J_c=0\right)+P\left(J_c\neq 0\right)
\end{aligned}$$

$$
\begin{aligned}
&+\left|P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}}+J_m\le x_1,\frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}}\le x_2\middle|J_m=0\right)\right.\\
&\left.-P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}}\le x_1,\frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}}\le x_2\middle|J_m=0\right)\right|\times P\left(J_m=0\right)+P\left(J_m\ne 0\right)\\
&\longrightarrow 0\quad\text{under } H_0\quad\text{as } n_1,n_2,p\to\infty. \qquad (22)
\end{aligned}
$$

It remains to prove that under H_0,

$$
P\left(\frac{M_{n_1,n_2}}{\widehat{\sigma}_{01}}\le x_1,\frac{T_{n_1,n_2}}{\widehat{\sigma}_{02}}\le x_2\right)\to\Phi(x_1)\Phi(x_2)\quad\text{as } n_1,n_2,p\to\infty. \qquad (23)
$$

Together with Lemma 1, it suffices to show that for any $a,b\in\mathbb{R}$, $aM_{n_1,n_2}/\widehat{\sigma}_{01}+bT_{n_1,n_2}/\widehat{\sigma}_{02}$ converge to a normal distribution. From the discussions of Chen and Qin (2010) and Li and Chen (2012), we know that $\widehat{\sigma}_{02}^{-1}\left(T_{n_1,n_2}-\widetilde{T}_{n_1,n_2}\right)\xrightarrow{p}0$ and $\widehat{\sigma}_{0i}/\sigma_{0i}\xrightarrow{p}1$ for $i=1,2$. As a result, we only need to show that under H_0, $aM_{n_1,n_2}/\sigma_{01}+b\widetilde{T}_{n_1,n_2}/\sigma_{02}$ is asymptotically normally distributed. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the common mean vector and covariance matrix under the null hypothesis. Without loss of generality, we assume $\boldsymbol{\mu}=\mathbf{0}$.

For ease of notation, let $\mathbf{W}_i$ be a new random variable taking values of

$$
\mathbf{W}_i=\mathbf{X}_i,\ i=1,\cdots,n_1;\quad \mathbf{W}_{n_1+i}=\mathbf{Y}_i,\ i=1,\cdots,n_2.
$$

Let $\mathcal{F}_0=\{\varnothing,\Omega\}$, $\mathcal{F}_m=\sigma\{Y_1,\ldots,Y_m\}$ for $m=1,2,\ldots,n_1+n_2$, and let $E_m(\cdot)$ denote the conditional expectation given $\mathcal{F}_m$., i.e., $E_m(\cdot)=E(\,\cdot\,|\mathcal{F}_m)$. Consider

$$
D_{n,m}=(E_m-E_{m-1})M_{n_1,n_2},\quad G_{n,m}=(E_m-E_{m-1})\widetilde{T}_{n_1,n_2}.
$$

To be more specific,

$$
\begin{aligned}
D_{n,m}=&\frac{2}{n_1(n_1-1)}\sum_{i=1}^{m-1}\mathbf{W}_i'\mathbf{W}_m\mathcal{I}\{m\le n_1\}+\frac{2}{n_2(n_2-1)}\sum_{i=n_1+1}^{m-1}\mathbf{W}_i'\mathbf{W}_m\mathcal{I}\{m>n_1\}\\
&-\frac{2}{n_1n_2}\sum_{i=1}^{n_1}\mathbf{W}_i'\mathbf{W}_m\mathcal{I}\{m>n_1\}
\end{aligned}
$$

$$
\begin{aligned}
G_{n,m} =& \frac{2}{n_1(n_1-1)}\left[\mathbf{W}_m' Q_{1,m-1}\mathbf{W}_m - \mathrm{tr}(Q_{1,m-1}\boldsymbol{\Sigma})\right]\mathcal{I}\{m \le n_1\} \\
&+ \frac{2}{n_2(n_2-1)}\left[\mathbf{W}_m' Q_{2,m-1}\mathbf{W}_m - \mathrm{tr}(Q_{2,m-1}\boldsymbol{\Sigma})\right]\mathcal{I}\{m > n_1\} \\
&- \frac{2}{n_1 n_2}\left[\mathbf{W}_m' Q_{1,n_1}\mathbf{W}_m - \mathrm{tr}(Q_{1,n_1}\boldsymbol{\Sigma})\right]\mathcal{I}\{m > n_1\},
\end{aligned}
$$

where $Q_{1,m-1} = \sum_{i=1}^{m-1}(\mathbf{W}_i\mathbf{W}_i' - \boldsymbol{\Sigma})$ and $Q_{2,n_1+l-1} = \sum_{i=1}^{l-1}(\mathbf{W}_{n_1+i}\mathbf{W}_{n_1+i}' - \boldsymbol{\Sigma})$.

It's easy to verify that

$$
aM_{n_1,n_2}/\sigma_{01} + b\widetilde{T}_{n_1,n_2}/\sigma_{02} = \sum_{i=1}^{n_1+n_2}\left(aD_{n,m}/\sigma_{01} + bG_{n,m}/\sigma_{02}\right),
$$

and for any n, $\{aD_{n,m}/\sigma_{01} + bG_{n,m}/\sigma_{02}, 1 \le m \le n\}$ is a martingale difference sequence with respect to the σ-fields $\{\mathcal{F}_m, 1 \le m \le n\}$. By Martingale central limit theorem, we only need to show that

$$
\frac{\sum_{m=1}^{n_1+n_2} E_{m-1}(aD_{n,m}/\sigma_{01} + bG_{n,m}/\sigma_{02})^2}{s_n^2} \xrightarrow{p} 1, \tag{24}
$$

and

$$
\frac{\sum_{m=1}^{n_1+n_2} E(aD_{n,m}/\sigma_{01} + bG_{n,m}/\sigma_{02})^4}{s_n^4} \to 0, \tag{25}
$$

where $s_n^2 = \mathrm{var}(aM_{n_1,n_2}/\sigma_{01} + b\widetilde{T}_{n_1,n_2}/\sigma_{02}) = a^2 + b^2 + o(1)$.

Chen and Qin (2010) and Li and Chen (2012) proved that

$$
\frac{1}{\sigma_{01}^2}\sum_{m=1}^{n_1+n_2} E_{m-1}(D_{n,m}^2) \xrightarrow{p} 1, \quad \frac{1}{\sigma_{02}^2}\sum_{m=1}^{n_1+n_2} E_{m-1}(G_{n,m}^2) \xrightarrow{p} 1.
$$

and

$$
\frac{1}{\sigma_{01}^4}\sum_{m=1}^{n_1+n_2} E(D_{n,m}^4) \to 0, \quad \frac{1}{\sigma_{02}^4}\sum_{m=1}^{n_1+n_2} E(G_{n,m}^4) \to 0.
$$

Therefore,

$$
\frac{1}{s_n^4}\sum_{m=1}^{n_1+n_2} E\left(\frac{aD_{n,m}}{\sigma_{01}} + \frac{bG_{n,m}}{\sigma_{02}}\right)^4 \le \frac{8}{s_n^4}\left(\frac{a^4}{\sigma_{01}^4}\sum_{m=1}^{n_1+n_2} E(D_{n,m}^4) + \frac{b^4}{\sigma_{02}^4}\sum_{m=1}^{n_1+n_2} E(G_{n,m}^4)\right) \to 0,
$$

(25) is proved. In order to prove (24), we only need to show

$$\frac{1}{\sigma_{01}\sigma_{02}} \sum_{m=1}^{n_1+n_2} E_{m-1}(D_{n,m}G_{n,m}) \xrightarrow{p} 0. \tag{26}$$

Rigorous calculation suggests that for a sufficiently large positive constant C, we have

$$\begin{aligned} \left|E\left(\sum_{m=1}^{n_1+n_2} E_{m-1}(D_{n,m}G_{n,m})\right)\right| &= \left|\sum_{m=1}^{n_1+n_2} E\left(E_{m-1}(D_{n,m}G_{n,m})\right)\right| \\ &\leq C\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^2 \lambda_{\max}(\mathbf{\Sigma})\mathrm{tr}(\mathbf{\Sigma}^2), \end{aligned}$$

and

$$\begin{aligned} E\left(\sum_{m=1}^{n_1+n_2} E_{m-1}(D_{n,m}G_{n,m})\right)^2 &\leq (n_1+n_2)\sum_{m=1}^{n_1+n_2} E\left(E_{m-1}^2(D_{n,m}G_{n,m})\right) \\ &\leq \frac{C\lambda_{\max}^2(\mathbf{\Sigma})\mathrm{tr}^2(\mathbf{\Sigma}^2)}{(n_1+n_2)^4}. \end{aligned}$$

Therefore,

$$E\left(\sum_{m=1}^{n_1+n_2} E_{m-1}(D_{n,m}G_{n,m})\right) = o\left(\sigma_{01}\sigma_{02}\right),$$

and

$$\mathrm{var}\left(\sum_{m=1}^{n_1+n_2} E_{m-1}(D_{n,m}G_{n,m})\right) = o\left(\sigma_{01}^2\sigma_{02}^2\right).$$

Hence (26) holds, which gives (24). Further by the martingale central limit theorem (Hall and Heyde, 2014), we have (23). Together with (22), we finish the proof of Theorem 4. □

S.2.8 Proof of Theorem 5

Proof. (i) (Asymptotically accurate size) Without loss of generality, we assume the common mean vector under the null hypothesis H_0 equals to zero. As discussed by Li and Chen (2012), the third and fourth-moment summation terms in A_{n_1}, B_{n_2} and C_{n_1,n_2} are all of small order than the leading second-moment terms. As a result, after centering each datum, removing those terms from T_{n_1,n_2} would not affect its asymptotic behaviors. It brings us a lot of convenience for theoretical analysis and greatly alleviates the computational burden at the same time. We consider the simplified T_{n_1,n_2} statistic.

$$\widetilde{T}_{n_1,n_2} = \frac{1}{n_1(n_1-1)} \sum_{u \neq v}^{n_1} (\mathbf{X}_u' \mathbf{X}_v)^2 + \frac{1}{n_2(n_2-1)} \sum_{u \neq v}^{n_2} (\mathbf{Y}_u' \mathbf{Y}_v)^2 - \frac{2}{n_1 n_2} \sum_{u}^{n_1} \sum_{v}^{n_2} (\mathbf{X}_u' \mathbf{Y}_v)^2 . \quad (27)$$

Lemma 1 presents a crucial intermediate result that the statistics M_{n_1,n_2} and $\widetilde{T}_{n_1,n_2}$ are asymptotically uncorrelated. As an implication of Lemma 1 and further by the martingale central limit theorem (Hall and Heyde, 2014), we are able to obtain the asymptotic joint null distribution of M_{n_1,n_2} and T_{n_1,n_2}. In combination with Theorems 1 and 2, the joint limiting null distribution of M_{PE} and T_{PE} is obtained and summarized in Theorem 4. Then the asymptotically accurate size of the simultaneous test J_{n_1,n_2} directly follows from the asymptotic independence of Theorem 4.

(ii) (Asymptotically consistent power) Note that

$$\begin{aligned} &\inf_{\{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^d\cup\mathcal{G}_c^s\}\cup\{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d\cup\mathcal{G}_m^s\}} P\left(J_{n_1,n_2} \geq q_\alpha\right) \\ &\geq \min\left\{ \inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^d\cup\mathcal{G}_c^s} P\left(J_{n_1,n_2} \geq q_\alpha\right), \inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d\cup\mathcal{G}_m^s} P\left(J_{n_1,n_2} \geq q_\alpha\right) \right\}. \end{aligned}$$

It suffices to show

$$\inf_{\mathcal{G}_k^d\cup\mathcal{G}_k^s} P\left(J_{n_1,n_2} \geq q_\alpha\right) \to 1, \quad (28)$$

for $k = c, m$. Since both $-2\log\left(1-\Phi(T_{PE})\right)$ and $-2\log\left(1-\Phi(M_{PE})\right)$ are always non-negative, we have

$$\inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^d\cup\mathcal{G}_c^s} P\left(J_{n_1,n_2} \geq q_\alpha\right) \geq \inf_{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in\mathcal{G}_c^d\cup\mathcal{G}_c^s} P\left(T_{PE} \geq \Phi^{-1}\left(1-\exp(-q_\alpha/2)\right)\right),$$

and

$$\inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d\cup\mathcal{G}_m^s} P\left(J_{n_1,n_2} \geq q_\alpha\right) \geq \inf_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)\in\mathcal{G}_m^d\cup\mathcal{G}_m^s} P\left(M_{PE} \geq \Phi^{-1}\left(1-\exp(-q_\alpha/2)\right)\right).$$

Together with the power analysis in Theorem 3, (28) is proved. □

S.3 Additional Numerical Results

S.3.1 Non-Gaussian Data

This subsection includes additional numerical results with the intention to exam the test performance in regard to non-Gaussian data. Tables S.1 presents the empirical power for testing H_m and H_c. Tables S.2 and S.3 include power analysis for testing H_b. These results show a similar pattern to the Gaussian scenarios as in Tables 2-4, implying that the proposed approaches are insensitive to non-Gaussianity.

Table S.1: Empirical Power (%) against H_m and H_c with Gamma Distributed $\{Z_{k,i}\}$ in the Data Generating Process

H	Method	$N = 100$					$N = 200$				
		$p = 100$	200	500	800	1000	100	200	500	800	1000
H_m^d	M_{n_1,n_2}	46.88	46.22	47.72	46.80	46.88	86.68	88.84	90.38	91.02	91.20
	M_{PE}	48.46	46.98	48.22	47.06	47.34	87.06	88.96	90.44	91.08	91.20
	T_{n_1,n_2}	5.50	5.02	5.12	4.90	5.04	5.34	5.38	4.94	5.24	4.96
	T_{PE}	5.52	5.02	5.12	4.90	5.04	5.34	5.38	4.94	5.24	4.96
	S_{n_1,n_2}	33.60	30.44	30.64	29.80	29.40	75.70	76.64	78.30	79.58	79.22
	C_{n_1,n_2}	33.98	30.44	30.38	29.44	29.34	76.06	77.04	79.22	80.26	80.20
	J_{n_1,n_2}	40.22	37.10	38.12	37.38	37.66	80.24	81.24	83.26	84.50	83.98
H_m^s	M_{n_1,n_2}	42.84	32.98	23.46	19.52	18.70	82.98	73.38	52.78	45.22	39.70
	M_{PE}	78.34	79.32	79.12	80.50	79.96	99.50	99.64	99.68	99.82	99.80
	T_{n_1,n_2}	5.18	5.10	5.18	5.22	4.98	5.22	5.30	5.22	5.24	5.14
	T_{PE}	5.22	5.10	5.18	5.22	4.98	5.22	5.30	5.22	5.24	5.14
	S_{n_1,n_2}	76.52	77.50	77.72	79.30	78.96	99.38	99.58	99.64	99.80	99.76
	C_{n_1,n_2}	76.48	77.52	77.84	79.08	79.10	99.38	99.56	99.62	99.80	99.76
	J_{n_1,n_2}	77.62	78.42	78.66	79.74	79.68	99.40	99.58	99.62	99.80	99.76
H_c^d	M_{n_1,n_2}	5.00	4.90	5.02	4.98	5.08	5.28	5.10	5.12	5.20	5.36
	M_{PE}	5.58	5.36	5.32	5.22	5.26	5.62	5.32	5.22	5.28	5.42
	T_{n_1,n_2}	58.08	58.72	59.36	58.28	59.90	97.12	97.94	98.54	98.32	98.62
	T_{PE}	58.10	58.74	59.36	58.28	59.90	97.12	97.94	98.54	98.32	98.62
	S_{n_1,n_2}	36.16	37.14	37.28	36.48	38.54	91.02	92.44	92.90	93.40	93.84
	C_{n_1,n_2}	37.56	37.92	37.14	36.46	39.10	91.56	92.92	93.44	94.24	94.86
	J_{n_1,n_2}	45.54	46.50	45.62	44.68	48.40	93.26	94.92	95.36	95.70	95.94
H_c^s	M_{n_1,n_2}	5.26	5.42	5.40	5.08	5.50	5.16	5.50	5.20	5.50	5.32
	M_{PE}	5.60	5.52	5.46	5.10	5.52	5.36	5.56	5.22	5.54	5.36
	T_{n_1,n_2}	36.26	23.22	11.70	8.88	8.56	67.72	47.68	19.18	12.88	11.32
	T_{PE}	44.98	30.20	14.74	10.92	10.28	86.90	75.10	45.64	33.48	27.58
	S_{n_1,n_2}	34.64	22.08	10.34	8.60	8.56	81.02	67.20	40.14	29.66	24.32
	C_{n_1,n_2}	35.68	22.06	10.42	8.94	9.16	81.62	67.32	39.80	29.58	24.46
	J_{n_1,n_2}	39.86	26.86	13.68	10.56	10.02	83.52	70.46	43.26	32.72	26.32

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%. (2) H_m^d and H_m^s stands for the type of alternative hypotheses (dense/sparse) in regards of the mean differences of the two populations. (3) H_c^d and H_c^s stands for the type of alternative hypotheses (dense/sparse) in regards of the covariance differences of the two populations.

Table S.2: Empirical Power (%) against H_b with Gamma Distributed $\{Z_{k,i}\}$ in the Data Generating Process

H_b	N	Method	$p = 100$	200	500	800	1000
$H_m^d \cap H_c^d$	100	M_{n_1,n_2}	45.42	43.76	44.96	45.08	46.04
		M_{PE}	46.72	44.48	45.30	45.48	46.30
		T_{n_1,n_2}	57.56	59.30	60.08	60.70	60.20
		T_{PE}	57.58	59.32	60.08	60.70	60.20
		S_{n_1,n_2}	57.62	57.46	57.24	58.92	57.90
		C_{n_1,n_2}	55.94	55.42	55.32	56.60	55.94
		J_{n_1,n_2}	72.64	73.44	74.46	76.24	75.94
	200	M_{n_1,n_2}	84.34	85.66	88.48	89.48	89.38
		M_{PE}	84.64	85.86	88.56	89.52	89.38
		T_{n_1,n_2}	97.52	98.28	98.26	98.56	98.34
		T_{PE}	97.52	98.28	98.26	98.56	98.34
		S_{n_1,n_2}	98.12	98.94	99.08	99.28	99.14
		C_{n_1,n_2}	97.74	98.60	98.88	99.06	99.02
		J_{n_1,n_2}	99.46	99.74	99.76	99.88	99.72
$H_m^d \cap H_c^s$	100	M_{n_1,n_2}	38.44	40.82	41.32	39.57	38.92
		M_{PE}	40.26	41.70	41.58	39.79	39.02
		T_{n_1,n_2}	36.34	22.80	11.84	8.71	8.36
		T_{PE}	44.40	29.18	14.46	10.29	9.46
		S_{n_1,n_2}	48.44	39.58	28.40	27.93	24.86
		C_{n_1,n_2}	47.86	39.04	28.52	27.64	24.54
		J_{n_1,n_2}	58.30	50.18	40.00	36.93	34.56
	200	M_{n_1,n_2}	77.00	83.46	84.40	83.39	82.84
		M_{PE}	77.90	83.72	84.54	83.39	82.90
		T_{n_1,n_2}	67.78	45.94	19.46	13.44	10.98
		T_{PE}	85.94	72.74	43.38	30.28	26.40
		S_{n_1,n_2}	91.26	87.96	78.74	74.72	71.94
		C_{n_1,n_2}	90.62	87.10	79.38	75.28	72.80
		J_{n_1,n_2}	94.34	92.36	85.74	81.28	79.42

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%.
(2) H_b stands for the type of alternative hypotheses (dense/sparse) in regards of the mean and covariance differences of the two populations.

Table S.3: Empirical Power (%) against H_b with Gamma Distributed $\{Z_{k,i}\}$ in the Data Generating Process

H_b	N	Method	$p = 100$	200	500	800	1000
$H_m^s \cap H_c^d$	100	M_{n_1,n_2}	41.38	32.02	22.14	19.02	16.66
		M_{PE}	76.86	77.74	78.46	78.32	79.40
		T_{n_1,n_2}	58.16	60.80	57.92	58.98	59.72
		T_{PE}	58.18	60.82	57.94	58.98	59.72
		S_{n_1,n_2}	83.56	84.80	85.26	84.86	86.18
		C_{n_1,n_2}	83.42	84.96	85.58	84.82	86.36
		J_{n_1,n_2}	87.96	88.68	88.98	88.38	89.44
	200	M_{n_1,n_2}	82.22	71.14	51.92	43.64	37.84
		M_{PE}	99.04	99.40	99.62	99.74	99.84
		T_{n_1,n_2}	97.18	97.64	98.42	98.60	98.22
		T_{PE}	97.20	97.64	98.42	98.60	98.22
		S_{n_1,n_2}	99.90	99.96	99.98	99.96	99.98
		C_{n_1,n_2}	99.94	99.98	99.98	99.98	99.98
		J_{n_1,n_2}	99.94	99.98	100.00	99.98	99.98
$H_m^s \cap H_c^s$	100	M_{n_1,n_2}	33.30	29.82	19.76	16.76	14.60
		M_{PE}	68.78	72.24	70.54	68.42	68.48
		T_{n_1,n_2}	36.72	22.80	11.56	8.52	8.02
		T_{PE}	45.42	30.00	14.88	10.38	9.24
		S_{n_1,n_2}	76.90	75.44	70.92	67.68	67.72
		C_{n_1,n_2}	77.02	75.18	70.62	67.68	67.54
		J_{n_1,n_2}	79.56	78.04	72.96	69.54	69.44
	200	M_{n_1,n_2}	73.42	64.90	46.66	36.84	33.44
		M_{PE}	96.56	97.46	98.08	98.16	98.18
		T_{n_1,n_2}	67.58	46.46	19.00	12.94	10.56
		T_{PE}	86.06	73.48	47.22	32.64	27.40
		S_{n_1,n_2}	98.84	98.72	98.52	98.58	98.40
		C_{n_1,n_2}	98.84	98.72	98.62	98.54	98.38
		J_{n_1,n_2}	99.06	99.02	98.74	98.74	98.46

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%.
(2) H_b stands for the type of alternative hypotheses (dense/sparse) in regards of the mean and covariance differences of the two populations.

S.3.2 Various Dependent Structure

In this subsection, we conduct additional experiments with various covariance matrices to examine the test performance in the presence of different dependent structure. Under H_0, we generate samples $\{\mathbf{X}_u\}_{u=1}^{n_1}$ and $\{\mathbf{Y}_v\}_{v=1}^{n_2}$ from $N_p(\mathbf{0}, \boldsymbol{\Sigma}^{*(i)})$. We consider the following two covariance matrices:

(i) $\boldsymbol{\Sigma}^{*(1)}$ is a block diagonal matrix with each block being $0.5\mathbf{I}_5 + 0.5\mathbb{1}_5\mathbb{1}_5'$.

(ii) $\boldsymbol{\Sigma}^{*(2)} = (\widetilde{\boldsymbol{\Sigma}} + \delta\mathbf{I}_p)/(1+\delta)$, where $\widetilde{\boldsymbol{\Sigma}} = (\widetilde{\sigma}_{ij})_{p\times p}$ with $\widetilde{\sigma}_{ii} = 1$, $\widetilde{\sigma}_{ij} = 0.5 * \text{Bernoulli}(1, 0.05)$ for $i < j$, $\widetilde{\sigma}_{ij} = \widetilde{\sigma}_{ji}$, and $\delta = |\lambda_{\min}(\widetilde{\boldsymbol{\Sigma}})| + 0.05$.

$\boldsymbol{\Sigma}^{*(1)}$ has a block structure with moderate within-block correlations. $\boldsymbol{\Sigma}^{*(2)}$ is a randomly sparse matrix. Table S.4 reports the empirical size of simulation with the two covariance structure. Generally speaking, the size is well controlled at the nominal level, except for a slightly higher result when both N and p are small.

Table S.4: Empirical Size (%) with Various Covariance Structure

		$\boldsymbol{\Sigma}^{*(1)}$					$\boldsymbol{\Sigma}^{*(2)}$				
N	Method	$p = 100$	200	500	800	1000	100	200	500	800	1000
100	M_{n_1,n_2}	5.56	5.36	5.56	5.62	5.16	5.70	5.20	5.32	4.86	5.30
	M_{PE}	6.10	5.74	6.02	5.84	5.36	6.28	5.84	5.66	5.14	5.76
	T_{n_1,n_2}	5.64	4.66	4.82	5.12	5.52	5.02	5.28	5.06	4.98	4.88
	T_{PE}	5.64	4.66	4.82	5.12	5.52	5.02	5.28	5.06	4.98	4.88
	S_{n_1,n_2}	5.88	5.48	5.32	5.34	5.48	5.76	6.04	5.58	5.84	4.96
	C_{n_1,n_2}	6.78	6.08	5.74	5.84	6.00	6.50	6.12	5.84	5.52	5.18
	J_{n_1,n_2}	6.56	5.74	5.74	5.00	5.84	6.24	5.82	5.70	5.72	5.44
200	M_{n_1,n_2}	5.38	5.34	5.60	5.12	5.56	5.80	5.58	5.70	5.78	5.24
	M_{PE}	5.54	5.64	5.66	5.14	5.60	6.22	5.84	5.84	5.84	5.30
	T_{n_1,n_2}	4.94	5.58	5.60	5.24	5.50	5.08	5.02	4.86	5.32	5.22
	T_{PE}	4.94	5.58	5.60	5.24	5.50	5.08	5.02	4.86	5.32	5.22
	S_{n_1,n_2}	5.22	5.60	5.34	5.12	5.42	5.78	5.22	5.58	4.68	5.50
	C_{n_1,n_2}	6.28	6.28	6.02	5.76	6.02	6.48	6.12	5.94	5.76	5.34
	J_{n_1,n_2}	6.02	5.68	5.64	5.22	5.52	6.18	5.84	5.98	5.56	5.30

Note: This table reports the frequencies of rejection by each method under the null hypothesis H_0 based on 5000 independent replications conducted at the significance level 5%.

In terms of the power, we generate samples from $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$. Following Section 5, we investigate the test performance with respect to three types of alternatives: H_m, H_c and H_b. To examine H_m and H_c we consider sparsely and densely differed means (H_m^s/H_m^d) and covariances (H_c^s/H_c^d), and to examine H_b we consider four scenarios that are combinations of H_m^s/H_m^d and H_c^s/H_c^d. For H_m^s and H_m^d, we set $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}^{*(k)}$, $k = 1, 2$, and set $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ in the same way as in Section 5. For H_c^s and H_c^d, we set both mean vectors are zero vectors and consider different sparsity pattern in $\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1$ to depict the sparse and dense alternatives. Specifically, to examine the testing power against H_c^s, we use covariance pairs $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}^{*(k)} + \varepsilon \mathbf{I}_p$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}^{*(k)} + \varepsilon \mathbf{I}_p + \mathbf{U}$, where $\mathbf{U}$ is a symmetric sparse matrix with 8 random nonzero entries, and $\varepsilon = |\min\{\lambda_{\min}(\boldsymbol{\Sigma}^{*(k)} + \mathbf{U}), \lambda_{\min}(\boldsymbol{\Sigma}^{*(k)})\}| + 0.05$ is to guarantee the positive definiteness of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. The generalization of $\mathbf{U}$ is introduced in Section 5. To examine the testing power against H_c^d, we let $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}^{*(k)}$ and $\boldsymbol{\Sigma}_2 = (1 - \tau)\boldsymbol{\Sigma}^{*(k)} + \tau \mathbf{I}_p$ with $\tau = 0.5$. For each setup, we report the empirical power based on 5,000 replications at significance level 5%. The results are reported in Tables S.5 - S.8. The outcomes resound with the empirical findings in Section 5, implying that our proposed tests yield satisfactory performance under different covariance structure provided that the covariance matrices satisfy the regularity conditions, see Equation (2.5).

Table S.5: Empirical Power (%) against H_m with Various Covariance Structure

H_m	N	Method	$\mathbf{\Sigma}^{*(1)}$					$\mathbf{\Sigma}^{*(2)}$				
			$p=100$	200	500	800	1000	100	200	500	800	1000
H_m^d	100	M_{n_1,n_2}	31.26	32.32	30.50	31.00	31.72	40.58	40.98	40.56	41.22	41.06
		M_{PE}	32.52	33.12	30.94	31.26	32.12	42.32	42.16	41.14	41.42	41.32
		T_{n_1,n_2}	5.52	5.16	5.04	4.88	4.64	5.56	5.06	5.38	4.90	4.66
		T_{PE}	5.52	5.16	5.04	4.88	4.64	5.56	5.06	5.38	4.90	4.66
		S_{n_1,n_2}	23.10	21.88	19.20	18.86	18.14	28.74	26.90	25.56	25.36	24.46
		C_{n_1,n_2}	28.10	27.54	24.86	24.72	24.50	35.34	34.66	33.66	32.60	32.18
		J_{n_1,n_2}	27.78	27.54	24.34	24.84	24.14	35.16	33.82	32.74	32.48	30.94
	200	M_{n_1,n_2}	63.26	64.86	67.20	68.06	69.26	81.26	81.74	84.92	83.80	84.30
		M_{PE}	64.00	65.20	67.28	68.08	69.32	81.88	81.94	85.02	83.84	84.34
		T_{n_1,n_2}	5.76	5.00	4.76	4.66	5.02	4.94	4.76	4.88	4.48	4.78
		T_{PE}	5.76	5.00	4.76	4.66	5.02	4.94	4.76	4.88	4.48	4.78
		S_{n_1,n_2}	49.68	48.86	50.52	50.10	50.36	68.16	66.82	70.08	68.06	68.62
		C_{n_1,n_2}	57.04	57.40	58.96	59.46	60.10	76.28	74.82	78.58	76.28	76.90
		J_{n_1,n_2}	55.48	55.72	57.20	57.68	58.82	74.38	73.12	75.92	73.66	74.88
H_m^s	100	M_{n_1,n_2}	26.02	21.20	34.56	27.94	12.82	36.36	29.30	45.92	36.14	16.56
		M_{PE}	77.68	77.50	91.50	90.12	78.72	77.56	78.40	95.98	95.50	79.72
		T_{n_1,n_2}	6.04	5.04	5.10	4.60	5.78	5.36	4.58	4.88	5.02	5.00
		T_{PE}	6.04	5.04	5.10	4.60	5.78	5.36	4.58	4.88	5.02	5.00
		S_{n_1,n_2}	76.34	76.12	90.82	89.58	78.16	75.64	76.84	95.64	95.00	78.76
		C_{n_1,n_2}	77.36	77.02	91.22	89.82	78.28	76.96	77.60	95.84	95.28	79.38
		J_{n_1,n_2}	77.22	76.94	91.06	89.74	78.48	77.14	77.60	95.82	95.18	79.56
	200	M_{n_1,n_2}	61.46	47.74	76.44	66.32	25.40	76.54	64.98	90.18	81.22	33.98
		M_{PE}	99.42	99.74	99.98	99.98	99.76	99.26	99.48	100.00	100.00	99.70
		T_{n_1,n_2}	6.10	5.12	5.10	4.76	4.96	5.18	5.32	4.92	5.32	5.52
		T_{PE}	6.10	5.12	5.10	4.76	4.96	5.18	5.32	4.92	5.32	5.52
		S_{n_1,n_2}	99.40	99.68	100.00	99.98	99.76	99.16	99.42	100.00	100.00	99.72
		C_{n_1,n_2}	99.40	99.70	100.00	99.98	99.76	99.22	99.50	100.00	100.00	99.72
		J_{n_1,n_2}	99.42	99.70	100.00	99.98	99.76	99.22	99.44	100.00	100.00	99.74

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%.
(2) H_m stands for the type of alternative hypotheses (dense/sparse) in regards of the mean differences of the two populations.

Table S.6: Empirical Power (%) against H_c with Various Covariance Structure

H_c	N	Method	$\boldsymbol{\Sigma}^{*(1)}$					$\boldsymbol{\Sigma}^{*(2)}$				
			$p=100$	200	500	800	1000	100	200	500	800	1000
H_c^d	100	M_{n_1,n_2}	5.70	5.84	5.42	5.00	5.32	5.38	5.90	5.52	5.12	4.94
		M_{PE}	6.46	6.36	5.78	5.26	5.54	6.24	6.64	5.92	5.32	5.26
		T_{n_1,n_2}	97.06	97.74	98.60	98.20	98.88	35.72	38.28	38.82	39.14	38.74
		T_{PE}	97.06	97.74	98.60	98.20	98.88	35.72	38.28	38.82	39.14	38.74
		S_{n_1,n_2}	90.86	93.24	94.12	93.48	94.64	19.44	21.18	21.58	20.86	21.30
		C_{n_1,n_2}	94.90	95.96	97.24	96.46	97.42	27.86	30.22	30.70	29.60	30.04
		J_{n_1,n_2}	93.32	95.06	96.24	95.52	96.44	27.22	29.32	29.82	28.96	29.18
	200	M_{n_1,n_2}	6.30	5.62	5.46	5.60	5.46	5.50	5.82	5.18	4.80	5.62
		M_{PE}	6.64	5.84	5.58	5.70	5.62	5.80	6.12	5.32	4.88	5.78
		T_{n_1,n_2}	100.00	100.00	100.00	100.00	100.00	80.42	83.54	85.76	87.54	87.84
		T_{PE}	100.00	100.00	100.00	100.00	100.00	80.42	83.54	85.76	87.54	87.84
		S_{n_1,n_2}	100.00	100.00	100.00	100.00	100.00	61.60	64.58	67.88	70.40	71.32
		C_{n_1,n_2}	100.00	100.00	100.00	100.00	100.00	72.02	75.04	78.12	80.12	80.48
		J_{n_1,n_2}	100.00	100.00	100.00	100.00	100.00	69.58	72.62	75.58	77.28	77.86
H_c^s	100	M_{n_1,n_2}	5.70	5.60	5.10	5.46	5.10	6.28	5.46	5.54	5.78	5.24
		M_{PE}	6.44	6.18	5.42	5.66	5.28	7.16	6.06	5.98	6.20	5.48
		T_{n_1,n_2}	33.34	16.40	8.66	7.56	6.60	38.26	18.64	10.28	7.50	7.14
		T_{PE}	73.18	55.90	26.58	15.70	11.38	76.88	63.98	38.00	22.94	17.56
		S_{n_1,n_2}	69.92	53.28	24.28	14.22	10.64	73.26	60.70	35.44	21.80	16.48
		C_{n_1,n_2}	72.10	54.78	26.08	14.98	11.30	75.46	62.60	36.74	22.80	17.28
		J_{n_1,n_2}	72.10	55.00	25.80	15.30	11.18	75.14	62.40	36.94	23.08	17.58
	200	M_{n_1,n_2}	5.02	5.62	5.38	5.62	5.36	5.96	5.48	5.64	5.30	5.06
		M_{PE}	5.38	6.02	5.44	5.70	5.46	6.32	5.72	5.78	5.40	5.14
		T_{n_1,n_2}	76.00	39.28	14.32	10.22	9.26	80.60	46.68	16.86	11.02	9.96
		T_{PE}	99.60	98.94	98.46	98.38	98.58	99.66	99.42	98.88	98.68	98.62
		S_{n_1,n_2}	99.48	98.86	98.44	98.38	98.50	99.66	99.38	98.80	98.66	98.54
		C_{n_1,n_2}	99.52	98.92	98.46	98.40	98.60	99.66	99.40	98.86	98.66	98.60
		J_{n_1,n_2}	99.52	98.94	98.46	98.40	98.54	99.62	99.38	98.86	98.66	98.58

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%.
(2) H_c stands for the type of alternative hypotheses (dense/sparse) in regards of the covariance differences of the two populations.

Table S.7: Empirical Power (%) against H_b with Various Covariance Structure

H_b	N	Method	$\mathbf{\Sigma}^{*(1)}$					$\mathbf{\Sigma}^{*(2)}$				
			$p=100$	200	500	800	1000	100	200	500	800	1000
$H_m^d \cap H_c^d$	100	M_{n_1,n_2}	36.06	36.48	36.88	35.84	36.94	42.96	41.80	44.42	42.38	42.20
		M_{PE}	37.68	37.24	37.26	36.06	37.24	44.52	42.38	44.92	42.78	42.62
		T_{n_1,n_2}	97.08	97.94	98.80	98.54	98.62	36.70	39.40	38.86	39.68	39.36
		T_{PE}	97.08	97.94	98.80	98.54	98.62	36.70	39.40	38.86	39.68	39.36
		S_{n_1,n_2}	94.66	95.42	96.50	96.12	96.44	44.04	43.86	43.36	42.66	43.80
		C_{n_1,n_2}	97.22	97.70	98.40	98.18	98.42	55.48	55.20	56.38	54.90	56.24
		J_{n_1,n_2}	97.20	97.82	98.66	98.40	98.56	60.40	61.30	61.66	61.90	62.20
	200	M_{n_1,n_2}	69.52	73.08	74.64	75.82	77.12	83.66	85.30	86.44	87.38	86.90
		M_{PE}	70.16	73.26	74.70	75.90	77.14	84.14	85.42	86.50	87.42	86.98
		T_{n_1,n_2}	100.00	100.00	100.00	100.00	100.00	83.34	86.54	86.82	87.42	86.58
		T_{PE}	100.00	100.00	100.00	100.00	100.00	83.34	86.54	86.82	87.42	86.58
		S_{n_1,n_2}	100.00	100.00	100.00	100.00	100.00	92.94	94.44	95.06	95.48	95.26
		C_{n_1,n_2}	100.00	100.00	100.00	100.00	100.00	96.06	96.82	97.06	97.54	97.28
		J_{n_1,n_2}	100.00	100.00	100.00	100.00	100.00	97.56	98.20	98.30	98.60	98.36
$H_m^d \cap H_c^s$	100	M_{n_1,n_2}	13.46	13.48	12.96	11.38	11.08	16.70	14.08	11.30	9.68	9.88
		M_{PE}	14.70	14.12	13.36	11.66	11.32	18.00	14.58	11.78	9.98	10.22
		T_{n_1,n_2}	35.12	17.74	8.64	7.90	6.46	38.42	19.26	9.68	8.18	6.80
		T_{PE}	74.00	58.34	28.14	16.44	11.84	76.80	63.86	37.68	23.42	17.62
		S_{n_1,n_2}	71.46	57.24	27.86	16.08	12.28	75.32	61.94	36.84	22.72	17.48
		C_{n_1,n_2}	74.22	60.36	31.38	19.42	15.28	77.72	64.60	39.72	24.96	20.02
		J_{n_1,n_2}	74.50	60.76	32.06	20.04	15.84	78.46	65.66	40.32	26.02	20.38
	200	M_{n_1,n_2}	26.04	24.38	23.24	23.26	22.36	34.10	28.98	20.40	16.04	15.64
		M_{PE}	26.68	24.80	23.44	23.36	22.46	35.00	29.34	20.54	16.14	15.82
		T_{n_1,n_2}	76.02	39.28	15.34	9.86	9.22	80.96	46.22	17.36	11.54	10.76
		T_{PE}	99.34	98.72	98.28	98.68	98.86	99.54	99.26	98.86	98.84	98.86
		S_{n_1,n_2}	99.32	98.78	98.38	98.78	98.88	99.56	99.30	98.94	98.86	98.92
		C_{n_1,n_2}	99.36	98.86	98.40	98.80	99.02	99.62	99.42	98.98	98.94	98.94
		J_{n_1,n_2}	99.40	98.86	98.44	98.86	99.00	99.60	99.44	99.00	98.94	99.00

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%.
(2) H_b stands for the type of alternative hypotheses (dense/sparse) in regards of the mean and covariance differences of the two populations.

Table S.8: Empirical Power (%) against H_b with Various Covariance Structure

H_b	N	Method	$\mathbf{\Sigma}^{*(1)}$					$\mathbf{\Sigma}^{*(2)}$				
			$p=100$	200	500	800	1000	100	200	500	800	1000
$H_m^s \cap H_c^d$	100	M_{n_1,n_2}	31.14	25.60	38.98	32.32	15.02	37.96	29.90	48.14	40.34	17.04
		M_{PE}	77.00	78.76	92.04	92.58	79.78	77.78	78.38	95.58	96.12	79.08
		T_{n_1,n_2}	97.12	97.68	98.56	98.60	98.64	36.88	37.44	39.56	38.52	40.30
		T_{PE}	97.12	97.68	98.56	98.60	98.64	36.88	37.44	39.56	38.52	40.30
		S_{n_1,n_2}	97.66	98.52	99.50	99.66	98.72	80.06	80.26	95.80	96.66	81.98
		C_{n_1,n_2}	98.72	99.20	99.80	99.90	99.28	82.86	83.36	96.52	97.52	84.80
		J_{n_1,n_2}	98.72	99.16	99.78	99.90	99.30	83.80	84.02	96.86	97.54	85.32
	200	M_{n_1,n_2}	70.34	56.58	84.24	74.90	29.50	79.44	68.30	92.04	85.36	36.90
		M_{PE}	99.20	99.58	99.96	100.00	99.84	99.10	99.52	100.00	100.00	99.78
		T_{n_1,n_2}	100.00	100.00	100.00	100.00	100.00	82.02	80.46	87.32	86.20	87.54
		T_{PE}	100.00	100.00	100.00	100.00	100.00	82.02	80.46	87.32	86.20	87.54
		S_{n_1,n_2}	100.00	100.00	100.00	100.00	100.00	99.72	99.82	100.00	100.00	99.92
		C_{n_1,n_2}	100.00	100.00	100.00	100.00	100.00	99.82	99.82	100.00	100.00	99.96
		J_{n_1,n_2}	100.00	100.00	100.00	100.00	100.00	99.86	99.84	100.00	100.00	99.92
$H_m^s \cap H_c^s$	100	M_{n_1,n_2}	12.72	10.80	13.22	11.48	7.60	15.52	11.34	12.30	9.84	6.28
		M_{PE}	19.56	16.06	19.96	18.10	10.74	26.16	18.54	17.26	12.68	7.50
		T_{n_1,n_2}	33.90	17.44	9.14	7.48	7.10	37.82	19.40	10.46	8.02	7.46
		T_{PE}	72.82	57.84	27.18	16.02	12.16	76.16	64.66	36.86	23.52	18.32
		S_{n_1,n_2}	71.42	57.62	33.50	22.30	13.88	76.18	65.32	39.06	25.00	18.12
		C_{n_1,n_2}	74.42	60.64	36.54	25.44	15.78	79.16	67.50	42.32	27.60	19.60
		J_{n_1,n_2}	74.68	61.50	37.16	25.96	16.06	79.26	68.10	42.74	28.30	19.70
	200	M_{n_1,n_2}	23.94	18.34	25.98	20.88	10.62	30.90	20.58	21.64	15.92	8.12
		M_{PE}	43.42	36.92	52.24	47.68	28.56	58.56	44.62	42.44	28.92	13.12
		T_{n_1,n_2}	77.12	39.68	15.48	10.30	9.14	80.68	46.20	17.82	13.36	10.36
		T_{PE}	99.30	98.86	98.46	98.74	98.36	99.68	99.24	98.84	98.92	98.98
		S_{n_1,n_2}	99.36	99.02	98.76	99.08	98.54	99.72	99.28	99.18	98.94	98.96
		C_{n_1,n_2}	99.40	99.04	98.82	99.10	98.56	99.76	99.38	99.20	99.00	99.00
		J_{n_1,n_2}	99.48	99.10	98.86	99.12	98.64	99.78	99.36	99.20	99.04	99.00

Note: (1) This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%.
(2) H_b stands for the type of alternative hypotheses (dense/sparse) in regards of the mean and covariance differences of the two populations.

S.3.3 Comparisons between the Power-Enhanced Tests and Other Tests Designed for Sparse Alternatives

The numerical results in Section 5 have showed that the mean test M_{n_1,n_2} and the covariance test T_{n_1,n_2} are powerful under their respective dense alternatives H_m^d and H_c^d, but may lack power under the sparse alternatives H_m^s and H_c^s. By adding the PE components, the proposed power-enhanced tests M_{PE} and T_{PE} exhibit substantially enhanced power under H_m^s and H_c^s while retain high power under H_m^d and H_c^d, respectively. In this subsection, we compare the power-enhanced tests with other tests that are designed for sparse alternatives, in specific, the extreme-value-based mean test M_{CLX} (Cai et al., 2013) and the extreme-value-based covariance test T_{CLX} (Cai et al., 2014).

We adopt the same simulation settings as those in Section 5. The empirical size and power are reported in Table S.9. From the table we can see that under the sparse alternatives H_m^s and H_c^s, M_{PE} and T_{PE} seem to be less powerful than M_{CLX} and T_{CLX}. Under the dense alternatives H_m^d and H_c^d, M_{PE} and T_{PE} outperform M_{CLX} and T_{CLX} by a large margin. M_{CLX} is powerful in detecting differences in the mean vectors but not covariance matrices, whereas T_{CLX} performs well in testing the covariance matrices but cannot distinguish two different mean vectors. In contrast, J_{n_1,n_2} remains high power under various alternatives.

Table S.9: Comparisons between Power-Enhanced Tests and Extreme-Value-Based Tests

H	Method	$p = 100$	200	500	800	1000
H_0	M_{PE}	5.68	5.56	5.62	5.18	5.30
	M_{CLX}	3.90	5.06	5.04	5.06	5.76
	T_{PE}	4.78	4.72	5.20	4.98	4.98
	T_{CLX}	4.70	4.38	4.34	4.34	4.52
	J_{n_1,n_2}	5.50	5.12	5.24	5.30	5.08
H_m^d	M_{PE}	87.34	89.10	90.86	91.58	91.34
	M_{CLX}	46.64	36.36	24.92	21.36	20.40
	T_{PE}	5.36	4.80	5.62	4.64	4.96
	T_{CLX}	4.64	4.46	4.64	4.68	4.68
	J_{n_1,n_2}	80.32	82.20	83.32	84.46	84.64
H_m^s	M_{PE}	99.22	99.68	100.00	99.98	99.74
	M_{CLX}	99.22	99.68	100.00	99.98	99.74
	T_{n_1,n_2}	5.10	4.64	4.76	4.66	5.04
	T_{CLX}	4.80	4.36	4.02	4.56	4.54
	J_{n_1,n_2}	99.18	99.64	100.00	99.98	99.74
H_c^d	M_{PE}	5.48	5.26	5.36	5.32	5.40
	M_{CLX}	4.36	4.86	5.78	5.18	4.92
	T_{PE}	97.46	98.24	98.52	98.22	98.46
	T_{CLX}	42.86	37.28	30.58	25.74	23.82
	J_{n_1,n_2}	94.34	94.68	95.62	95.50	95.64
H_c^s	M_{PE}	5.56	5.60	5.46	5.02	5.18
	M_{CLX}	4.58	4.80	5.04	5.46	5.02
	T_{PE}	99.68	99.48	98.70	99.04	99.16
	T_{CLX}	100.00	100.00	100.00	100.00	99.98
	J_{n_1,n_2}	99.62	99.44	98.74	98.98	99.18
$H_m^d \cap H_c^d$	M_{PE}	84.86	85.74	88.02	89.26	89.22
	M_{CLX}	43.46	34.22	24.08	20.12	18.92
	T_{PE}	98.16	98.26	98.26	98.38	98.48
	T_{CLX}	45.70	38.70	30.56	25.22	25.28
	J_{n_1,n_2}	99.64	99.88	99.88	99.88	99.90
$H_m^d \cap H_c^s$	M_{PE}	77.54	83.14	85.22	83.40	82.48
	M_{CLX}	37.20	31.26	20.22	17.50	15.60
	T_{PE}	99.82	99.46	98.84	98.94	99.00
	T_{CLX}	99.98	100.00	100.00	100.00	100.00
	J_{n_1,n_2}	99.92	99.82	99.52	99.42	99.42
$H_m^s \cap H_c^d$	M_{PE}	99.04	99.40	99.62	99.74	99.84
	M_{CLX}	99.66	99.90	100.00	100.00	100.00
	T_{PE}	97.20	97.64	98.42	98.60	98.22
	T_{CLX}	39.30	33.68	24.04	19.32	18.66
	J_{n_1,n_2}	99.94	99.98	100.00	99.98	99.98
$H_m^s \cap H_c^s$	M_{PE}	96.56	97.46	98.08	98.16	98.18
	M_{CLX}	98.54	99.48	99.56	99.62	99.72
	T_{PE}	86.06	73.48	47.22	32.64	27.40
	T_{CLX}	98.36	97.72	95.60	92.08	91.34
	J_{n_1,n_2}	99.06	99.02	98.74	98.74	98.46

Note: This table reports the frequencies of rejection by each method under the alternative hypothesis based on 5000 independent replications conducted at the significance level 5%.

References

Cai, T., Liu, W. and Xia, Y. (2013), 'Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings', *Journal of the American Statistical Association* **108**(501), 265–277.

Cai, T. T., Liu, W. and Xia, Y. (2014), 'Two-sample test of high dimensional means under dependence', *Journal of the Royal Statistical Society: Series B: Statistical Methodology* **76**(2), 349–372.

Chen, S. X., Guo, B. and Qiu, Y. (2019), 'Multi-level thresholding test for high dimensional covariance matrices', *arXiv preprint arXiv:1910.13074* .

Chen, S. X., Li, J. and Zhong, P.-S. (2019), 'Two-sample and anova tests for high dimensional means', *The Annals of Statistics* **47**(3), 1443–1474.

Chen, S. X. and Qin, Y.-L. (2010), 'A two-sample test for high-dimensional data with applications to gene-set testing', *The Annals of Statistics* **38**(2), 808–835.

Hall, P. and Heyde, C. C. (2014), *Martingale Limit Theory and its Application*, Academic Press.

Li, J. and Chen, S. X. (2012), 'Two sample tests for high-dimensional covariance matrices', *The Annals of Statistics* **40**(2), 908–940.

Petrov, V. V. (1954), 'Generalization of cramér's limit theorem', *Uspekhi Matematicheskikh Nauk* **9**(4), 195–202.