Web-based Supplementary Materials for "Simulation-Based Hypothesis Testing of High Dimensional Means Under Covariance Heterogeneity", by Jinyuan Chang, Chao Zheng, Wen-Xin Zhou and Wen Zhou

This supplementary material includes technical proofs and additional numerical studies. Section A introduces notations and important auxiliary lemmas. In Section B, we report the derivations of theoretical justifications for the proposed tests for one-sample problems. The derivations of theoretical justification for the proposed two-sample testing procedures are given in Section C. The theoretical justification of the proposed two-step procedures, the proposed tests with screening, are displayed in Section D. In Section E, more simulations studies are reported. Section F presents more results for the analysis of the acute lymphoblastic leukemia data recorded in Chiaretti et al. (2004).

A Preliminaries

Throughout, we use C and c to denote positive constants that are independent of (n, m, p), which may take different values at each occurrence.

The proposed simulation-based procedures for testing the equality of means are founded on the idea of Gaussian approximation. In this section, we state some useful results on high dimensional Gaussian approximations recently established by Chernozhukov, Chetverikov and Kato (2013) with some discussions. In line with Section 2, let $\mathbf{X}, \mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^p$ be independent and identically distributed random vectors with mean $\boldsymbol{\mu}_1$ and covariance matrix $\boldsymbol{\Sigma}_1 = (\sigma_{1,k\ell})$. Write $\mathbf{X} = (X_1, \ldots, X_p)'$, $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})'$ and define

$$T_0 = \max_{1\le k\le p} \sqrt{n}\bar{X}_k \qquad \text{with} \qquad \bar{X}_k = n^{-1}\sum_{i=1}^n X_{ik}.$$

Moreover, let $\{\mathbf{G}_i = (G_{i1}, \ldots, G_{ip})'\}_{i=1}^n$ be a sequence of independent centred Gaussian random vectors in $\mathbb{R}^p$ with covariance matrix $\boldsymbol{\Sigma}_1$. The Gaussian analogue of T_0 can be defined as

$$Z_0 = \max_{1\le k\le p} n^{-1/2}\sum_{i=1}^n G_{ik}.$$

For $k = 1, \ldots, p$, let $\sigma_{1k}^2 = \mathrm{var}(X_k) = \sigma_{1,kk}$ and for $r \ge 1$, define the moments

$$\mu_r(\mathbf{X}) = \big(\mathbb{E}|\mathbf{U}|_\infty^r\big)^{1/r}, \quad \nu_r(\mathbf{X}) = \max_{1\le k\le p}\big(\mathbb{E}|U_k|^r\big)^{1/r}, \tag{A.1}$$

where $\mathbf{U} = \mathbf{U}(\mathbf{X}) = (U_1, \ldots, U_p)' = \mathbf{D}_1^{-1/2}\mathbf{X}$ and $\mathbf{D}_1 = \mathrm{diag}(\boldsymbol{\Sigma}_1)$. These quantities will play an important role in our theoretical analysis. For every $0 < t < 1$, set $u(t) = \max\{u_X(t), u_G(t)\}$, where $u_X(t)$ and $u_G(t)$ are the $(1-t)$-quantiles of $\max_{1\le i\le n} |\mathbf{U}_i|_\infty$ for $\mathbf{U}_i = \mathbf{D}_1^{-1/2}\mathbf{X}_i$ and $\max_{1\le i\le n} |\mathbf{D}_1^{-1/2}\mathbf{G}_i|_\infty$, respectively. Further, for $r \ge 4$ as in condition (M1), we write

$$\vartheta_{n,p} = \vartheta_{n,p}(r) = pn^{1-r/2}. \tag{A.2}$$

Lemmas 1 and 2 below are Theorem 2.2 and Lemma 3.1 in Chernozhukov, Chetverikov and Kato (2013).

Lemma 1. Assume that $\sigma_{1,kk}$ is bounded away from 0 and ∞. Then for any $0 < t < 1$,

$$\begin{aligned}\sup_{x\in\mathbb{R}} &|\mathbb{P}(T_0 \le x) - \mathbb{P}(Z_0 \le x)| \\ &\le C\Big[\big(\nu_3^{3/4} \vee \nu_4^{1/2}\big)\, n^{-1/8}\{\log(pn/t)\}^{7/8} + u(t)\, n^{-1/2}\{\log(pn/t)\}^{3/2} + t\Big],\end{aligned} \tag{A.3}$$

where $C > 0$ is a constant independent of n, p and t.

Lemma 2. Assume that there exist constants $C > c > 0$ such that for every $1 \le k \le p$, $c \le \sigma_{1,kk}, \sigma_{2,kk} \le C$. Then

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\max_{1\le k\le p} X_k \le x\right) - \mathbb{P}\left(\max_{1\le k\le p} Y_k \le x\right)\right| \le C'\Delta^{1/3}\{1 \vee \log(p/\Delta)\}^{2/3},$$

where $C' > 0$ is a constant depending only on c and C and $\Delta := \|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|_\infty$.

Remark: The upper bound given in (A.3) can be simplified under conditions (M1) and (M2), respectively. By Markov's inequality, we have for all $u > 0$,

$$\mathbb{P}\left(\max_{1\le i\le n} |\mathbf{D}^{-1/2}\mathbf{G}_i| > u\right) \le 2pn\{1-\Phi(u)\} \le u^{-1}\exp\{\log(pn) - u^2/2\}.$$

This implies that for every $0 < t < 1$, $u_G(t) \le \sqrt{2\log(pn/t)}$. Under condition (M1), once again using Markov's inequality yields, for every $u > 0$,

$$\mathbb{P}\left(\max_{1\le i\le n} |\mathbf{U}_i|_\infty > u\right) \le u^{-r}\,\mathbb{E}\left(\max_{1\le i\le n} |\mathbf{U}_i|_\infty^r\right) = u^{-r}\mu_{n,r}^r,$$

where $\mu_{n,r} := \{\mathbb{E}(\max_{1\le i\le n} |\mathbf{U}_i|_\infty^r)\}^{1/r}$. For $0 < t < 1$, it follows directly from the definition that $u_X(t) \le t^{-1/r}\mu_{n,r}$. In particular, taking

$$t = \min\left\{1,\ \left[\mu_{n,r} n^{-1/2}\{\log(pn)\}^{3/2}\right]^{r/(r+1)}\right\}$$

in (A.3) implies by the inequality $\mu_{n,r} = \{E(\max_{1\le i\le n} |\mathbf{U}_i|_\infty^r)\}^{1/r} \le (pn)^{1/r}\nu_r$ that the upper bound on the right side of (A.3) is of order

$$n^{-1/8}\{\log(pn)\}^{7/8} + \vartheta_{n,p}^{1/(r+1)}\{\log(pn)\}^{3/2}, \tag{A.4}$$

where $\vartheta_{n,p} = pn^{1-r/2}$ is as in (A.2).

On the other hand, if condition (M2) holds, then it is easy to see that $u_X(t)$ can be bounded by some multiple of $\{\log(pn/t)\}^{1/\gamma}$. Taking $t = n^{-1/2}$ gives the optimal rate of convergence in (A.3) which is of order

$$n^{-1/8}\{\log(pn)\}^{7/8} + n^{-1/2}\{\log(pn)\}^{3/2+1/\gamma}. \tag{A.5}$$

In the proofs below, (A.4) and (A.5) will be applied directly.

Recall that $\widehat{\boldsymbol{\Sigma}}_1 = (\hat\sigma_{1,k\ell})$ with $\hat\sigma_{1,k\ell} = n^{-1}\sum_{i=1}^n (X_{ik} - \bar X_k)(X_{i\ell} - \bar X_\ell)$ and $\hat\sigma_{1k}^2 = \hat\sigma_{1,kk}$. Furthermore, denote by $\mathbf{R}_1 = (r_{1,k\ell})$ the correlation matrix of $\mathbf{X}$ and its sample analogue is given by $\widehat{\mathbf{R}}_1 = (\hat r_{1,k\ell})$ with

$$\hat r_{1,k\ell} = \frac{\sum_{i=1}^n (X_{ik} - \bar X_k)(X_{i\ell} - \bar X_\ell)}{\sqrt{\sum_{i=1}^n (X_{ik} - \bar X_k)^2 \sum_{i=1}^n (X_{i\ell} - \bar X_\ell)^2}}.$$

The following lemma provides non-asymptotic bounds on the differences $\widehat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_1$ and $\widehat{\mathbf{R}}_1 - \mathbf{R}_1$ in the elementwise ℓ_∞-norm.

Lemma 3 (Probabilistic estimates)**.** Suppose that $n, p \ge 2$ and $\log(p) \le n$.

(i) Assume that condition (M1) holds. Then there exist constants $C_1, C_2 > 0$ independent of n and p such that, with probability at least $1 - C_1\{n^{-1} + \vartheta_{n,p}^{2/(r+2)}\}$,

$$\max\Big(\big|\mathbf{D}_1^{-1/2}\widehat{\boldsymbol{\Sigma}}_1\mathbf{D}_1^{-1/2} - \mathbf{R}_1\big|_\infty, \big|\widehat{\mathbf{R}}_1 - \mathbf{R}_1\big|_\infty\Big) \le C_2\Big[\nu_4^2\, n^{-1/2}\{\log(pn)\}^{1/2} + \nu_r^2\, \vartheta_{n,p}^{2/(r+2)} + \nu_r^2\, \vartheta_{n,p}^{2/r}\log(p)\Big]. \tag{A.6}$$

(ii) Assume that condition (M2) holds. Then there exist constants $C_3, C_4 > 0$ independent of n and p such that, with probability at least $1 - C_3 n^{-1}$,

$$\max\Big(\big|\mathbf{D}_1^{-1/2}\widehat{\boldsymbol{\Sigma}}_1\mathbf{D}_1^{-1/2} - \mathbf{R}_1\big|_\infty, \big|\widehat{\mathbf{R}}_1 - \mathbf{R}_1\big|_\infty\Big) \le C_4\Big[n^{-1/2}\{\log(pn)\}^{1/2} + n^{-1}\{\log(pn)\}^{2/\gamma}\Big]. \tag{A.7}$$

(iii) Assume that ν_4 in (A.1) is uniformly bounded. Then for $0 < t \le \sqrt{n}$,

$$\mathbb{P}\Big\{\sqrt{n}\big|\widehat{\mathbf{D}}_1^{-1/2}(\widehat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)\big|_\infty \ge t\Big\} \le Cp\exp(-ct^2) + n^{-1}, \tag{A.8}$$

where $\widehat{\mathbf{D}}_1 = \mathrm{diag}\,(\hat{\sigma}_{11}^2, \ldots, \hat{\sigma}_{1p}^2)$, $\hat{\boldsymbol{\mu}}_1 = (\bar{X}_1, \ldots, \bar{X}_p)'$ and $C, c > 0$ are constants independent of n and p.

Proof: Recall that for $i = 1, \ldots, n$, $\mathbf{U}_i = (U_{i1}, \ldots, U_{ip})'$ with $U_{ik} = (X_{ik} - \mu_{1k})/\sigma_{1k}$ and define $S_{kn} = \sum_{i=1}^n U_{ik}$ and $V_{kn} = (\sum_{i=1}^n U_{ik}^2)^{1/2}$. First, observe that

$$\frac{\hat{\sigma}_{1,k\ell}}{\sigma_{1k}\sigma_{1\ell}} = n^{-1}\sum_{i=1}^n (U_{ik} - n^{-1}S_{kn})(U_{i\ell} - n^{-1}S_{\ell n}) = n^{-1}\sum_{i=1}^n U_{ik}U_{i\ell} - n^{-2}S_{kn}S_{\ell n}. \tag{A.9}$$

Moreover, put $t_{kn} = n^{-1/2}S_{kn}/V_{kn}$ for $k = 1, \ldots, p$, such that the sample correlations can be expressed as

$$\hat{r}_{1,k\ell} = \frac{\sum_{i=1}^n U_{ik}U_{i\ell}}{V_{kn}V_{\ell n}}\big\{(1 - t_{kn}^2)(1 - t_{\ell n}^2)\big\}^{-1/2} - t_{kn}t_{\ell n}\big\{(1 - t_{kn}^2)(1 - t_{\ell n}^2)\big\}^{-1/2}. \tag{A.10}$$

By (A.9) and (A.10), lying in the heart of the proof is a careful analysis of the following quantity:

$$\Delta_1 = \max_{1\le k\le \ell\le p}\left| n^{-1}\sum_{i=1}^n U_{ik}U_{i\ell} - r_{1,k\ell}\right|. \tag{A.11}$$

Case 1. Assume that condition (M1) holds. We follow a standard procedure: first show Δ_1 is concentrated around its expectation $\mathbb{E}\Delta_1$, and then upper bound the expectation. Note that, $\mathbf{U}_1, \ldots, \mathbf{U}_n$ are independent and identical distributed random vectors in $\mathbb{R}^p$ with mean zero and covariance matrix $\mathbf{R}_1$. Applying Theorem 3.1 in Einmahl and Li (2008) with $s = \frac{1}{2}r$ and $\|\cdot\| = \|\cdot\|_\infty$ yields, for every $t > 0$,

$$\begin{aligned}\mathbb{P}\big(\Delta_1 \ge 2\mathbb{E}\Delta_1 + t\big) &\le \exp\{-(nt)^2/(3\nu_4^4 n)\} + C\,t^{-r/2}n^{1-r/2}\max_{1\le i\le n}\mathbb{E}\bigg(\max_{1\le k\le \ell\le p}|U_{ik}U_{i\ell}|^{r/2}\bigg)\\ &\le \exp\{-(nt)^2/(3\nu_4^4 n)\} + C\,t^{-r/2}n^{1-r/2}\max_{1\le i\le n}\mathbb{E}\bigg(\max_{1\le k\le p}|U_{ik}|^{r}\bigg)\\ &\le \exp\{-nt^2/(3\nu_4^4)\} + C\,t^{-r/2}\nu_r^r\,pn^{1-r/2},\end{aligned} \tag{A.12}$$

where by Lemma A.1 in Chernozhukov, Chetverikov and Kato (2013),

$$\begin{aligned}\mathbb{E}\Delta_1 &\le C\Bigg[n^{-1/2}\{\log(p)\}^{1/2}\Bigg\{\max_{1\le k\le \ell\le p}\sum_{i=1}^n \mathbb{E}(U_{ik}U_{i\ell})^2\Bigg\}^{1/2}\\ &\qquad + n^{-1}\log(p)\left\{E\left(\max_{1\le i\le n}\max_{1\le k\le \ell\le p}U_{ik}^2U_{i\ell}^2\right)\right\}^{1/2}\Bigg]\\ &\le C\left[\nu_4^2 n^{-1/2}\{\log(p)\}^{1/2} + \nu_r^2\,p^{2/r}n^{-1+2/r}\log(p)\right].\end{aligned} \tag{A.13}$$

Together, (A.13) and (A.12) imply by taking $t = \max[\nu_4^2\, n^{-1/2}\{\log(n)\}^{1/2}, \nu_r^2\, \vartheta_{n,p}^{2/(r+2)}]$ that

$$\mathbb{P}\left[\Delta_1 > C\left\{\nu_4^2\, n^{-1/2}\{\log(pn)\}^{1/2} + \nu_r^2\, \vartheta_{n,p}^{2/(r+2)} + \nu_r^2\, \vartheta_{n,p}^{2/r}\log(p)\right\}\right] \le C\left\{n^{-1} + \vartheta_{n,p}^{2/(r+2)}\right\}. \tag{A.14}$$

Next we focus on the self-normalized sums t_{kn} in the second term on the right side of (A.10). Using Theorem 2.19 in de la Peña, Lai and Shao (2009) to the non-negative random variables U_{ik}^2 gives, for every $0 < \varepsilon < 1$,

$$\mathbb{P}\left\{V_{kn}^2 \le (1-\varepsilon)n\right\} \le \exp\left\{-n\varepsilon^2/(2\mathbb{E}U_{ik}^4)\right\} \le \exp\left\{-n\varepsilon^2/(2\nu_4^4)\right\}. \tag{A.15}$$

Moreover, it follows from Lemma 6.4 in Jing, Shao and Wang (2003) that for $t > 0$,

$$\mathbb{P}\left\{|S_{kn}| \ge t(4\sqrt{n} + V_{kn})\right\} \le 4\exp(-t^2/2). \tag{A.16}$$

Then it is concluded from (A.15) and (A.16) that, with probability at least $1 - 4\exp(-t^2/2) - \exp(-cn)$, $|t_{kn}| \le t(n^{-1/2} + 4V_{n,k}^{-1}) \le 7tn^{-1/2}$. Taking $t = \{2\log(pn)\}^{1/2}$ we obtain that

$$\mathbb{P}\left\{\max_{1\le k\le p} |t_{kn}| > Cn^{-1/2}\{\log(pn)\}^{1/2}\right\} \le Cn^{-1}.$$

The last display, joint with (A.10) and (A.14) proves the bound for $\|\widehat{\mathbf{R}}_1 - \mathbf{R}_1\|_\infty$ in (A.6).

Let $\Delta_2 = \max_{1\le k\le p}|n^{-1}S_{kn}|$. The arguments leading to (A.13) and (A.14) can be used to prove that, respectively, $\mathbb{E}\Delta_2 \le C[n^{-1/2}\{\log(p)\}^{1/2} + \nu_r\, p^{1/r} n^{-1+1/r}]$ and for $t > 0$,

$$\mathbb{P}\big(\Delta_2 \ge 2\mathbb{E}\Delta_2 + t\big) \le \exp(-nt^2/3) + C\, t^{-r}\nu_r^r\, pn^{1-r}.$$

This completes the proof of (A.6) in view of (A.9) and (A.14).

Case 2. Under condition (M2), it follows from Theorem 6 in Delaigle, Hall and Jin (2011) that for all $y > 0$ and $1 \le k \le \ell \le p$,

$$\mathbb{P}\left\{\left|\sum_{i=1}^n \big(U_{ik}U_{i\ell} - \mathbb{E}U_{ik}U_{i\ell}\big)\right| > y\right\} \le 2\exp\left\{-y^2/(4n)\right\} + C\exp(-cy^{\gamma/2}),$$

where $C, c > 0$ are constants depending only on K_1, K_2 and γ in (M2). This, combined with a union bound yields

$$\mathbb{P}\left(\Delta_1 > C\left[n^{-1/2}\{\log(pn)\}^{1/2} + n^{-1}\{\log(pn)\}^{2/\gamma}\right]\right) \le Cn^{-1},$$

where Δ_1 is as in (A.11). A completely analogous argument leads to

$$\mathbb{P}\left(\Delta_2 > C\left[n^{-1/2}\{\log(pn)\}^{1/2} + n^{-1}\{\log(pn)\}^{1/\gamma}\right]\right) \le Cn^{-1}.$$

Assembling the above calculations completes the proof of (A.7).

Finally we prove (A.8). It is well-known the t-statistic can be expressed as

$$\frac{\sqrt{n}(\bar{X}_k-\mu_{1k})}{\hat{\sigma}_{1k}}=\frac{S_{kn}/V_{kn}}{\{1-n^{-1}(S_{kn}/V_{kn})^2\}^{1/2}}. \tag{A.17}$$

Let $\varepsilon_n=\min[\frac{1}{2},\nu_4^2\,n^{-1/2}\{2\log(pn)\}^{1/2}]$. Inequalities (A.15) and (A.16) imply, for $0<t\le\sqrt{n}$,

$$\begin{aligned}
&\mathbb{P}\left\{\sqrt{n}\,|\widehat{\mathbf{D}}_1^{-1/2}(\widehat{\boldsymbol{\mu}}_1-\boldsymbol{\mu}_1)|_\infty\ge t\right\}\\
&=\mathbb{P}\left\{\max_{1\le k\le p}\frac{\sqrt{n}(\bar{X}_k-\mu_{1k})}{\hat{\sigma}_{1k}}\ge t\right\}\\
&\le p\max_{1\le k\le p}\mathbb{P}\left\{S_{kn}/V_{kn}\ge t(1+t^2n^{-1})^{-1/2}\right\}\\
&\le p\max_{1\le k\le p}\left[\mathbb{P}\left\{S_{kn}\ge\frac{t(4\sqrt{n}+V_{kn})}{(1+4\sqrt{2})(1+t^2n^{-1})^{1/2}}\right\}+\mathbb{P}\left\{V_{kn}^2\le(1-\varepsilon_n)n\right\}\right]\\
&\le 3p\exp(-ct^2)+n^{-1}.
\end{aligned}$$

A completely analogous argument will lead to the same bound for $\mathbb{P}\{\sqrt{n}\,|\widehat{\mathbf{D}}_1^{-1/2}(\widehat{\boldsymbol{\mu}}_1-\boldsymbol{\mu}_1)|_\infty\le -t\}$, and hence completes the proof of Lemma 3. □

B Proof of the asymptotic null properties: one-sample case

B.1 Proof of Proposition 1

We first introduce a proposition that gives non-asymptotic bounds on the distributional error between the one-sample test statistic $T_\nu^{(\mathrm{I})}$ and its Gaussian analogue under the null hypothesis for $\nu\in\{\mathrm{ns},\mathrm{s}\}$. We then turn to a proof of Theorem 1.

Proposition 1. *Let $\nu\in\{\mathrm{ns},s\}$, $\Theta_{\mathrm{ns}}^{(\mathrm{I})}=|\widetilde{\boldsymbol{\Sigma}}_1-\boldsymbol{\Sigma}_1|_\infty$ and $\Theta_{\mathrm{s}}^{(\mathrm{I})}=|\widetilde{\mathbf{R}}_1-\mathbf{R}_1|_\infty$. The estimator $\widetilde{\boldsymbol{\Sigma}}_1$ of $\boldsymbol{\Sigma}_1$ is such that diag$(\widetilde{\boldsymbol{\Sigma}}_1)=$ diag$(\widehat{\boldsymbol{\Sigma}}_1)$.*

(i) *Assume that* (M1) *holds. Then under $H_0^{(\mathrm{I})}$,*

$$\begin{aligned}
&\sup_{x\ge 0}\left|\mathbb{P}\{T_\nu^{(\mathrm{I})}>x\}-\mathbb{P}\{|\mathbf{W}_\nu^{(\mathrm{I})}|_\infty>x|\mathcal{X}_n\}\right|\\
&\le C_1\left[\{\Theta_\nu^{(\mathrm{I})}\}^{1/3}\{\log(pn)\}^{2/3}+n^{-1/8}\{\log(pn)\}^{7/8}+\vartheta_{n,p}^{1/(r+1)}\{\log(pn)\}^{3/2}\right],
\end{aligned}$$

where $\vartheta_{n,p}=pn^{1-r/2}$.

(ii) *Assume that* (M2) *holds. Then under* $H_0^{(\mathrm{I})}$,

$$\sup_{x\ge 0}\big|\mathbb{P}\{T_\nu^{(\mathrm{I})}>x\}-\mathbb{P}\{|\mathbf{W}_\nu^{(\mathrm{I})}|_\infty>x|\mathcal{X}_n\}\big| \le C_2\Big[\{\Theta_\nu^{(\mathrm{I})}\}^{1/3}\{\log(pn)\}^{2/3}+n^{-1/8}\{\log(pn)\}^{7/8}+n^{-1/2}\{\log(pn)\}^{3/2+1/\gamma}\Big].$$

The constants $C_1, C_2>0$ *are independent of* n,p,r *and* γ.

Proof. For the sake of clarity we give the proof only for the studentized statistic $T_{\mathrm{s}}:=T_{\mathrm{s}}^{(\mathrm{I})}$, as the non-studentized statistic $T_{\mathrm{ns}}^{(\mathrm{I})}$ can be dealt with in a similar way. To begin with, observe that for every $x\in\mathbb{R}$, $|x|=\max(x,-x)$. Therefore, for $t>0$,

$$\mathbb{P}\big(T_s>t\big)=\mathbb{P}\left\{\max_{1\le k\le p}\max\left(\frac{\sqrt{n}\bar{X}_k}{\hat{\sigma}_{1k}},-\frac{\sqrt{n}\bar{X}_k}{\hat{\sigma}_{1k}}\right)>t\right\}. \tag{B.18}$$

Define a new sequence of dilated random vectors $\mathbf{X}_1^{\mathrm{ext}},\ldots,\mathbf{X}_n^{\mathrm{ext}}$ taking values in $\mathbb{R}^{2p}$, given by $\mathbf{X}_i^{\mathrm{ext}}=(X_{i,1}^{\mathrm{ext}},\ldots,X_{i,2p}^{\mathrm{ext}})'=(\mathbf{X}_i',-\mathbf{X}_i')'$. In this notation, we have $\mathbb{P}(T_s>t)=\mathbb{P}(T_s^{\mathrm{ext}}>t)$, where

$$T_{\mathrm{s}}^{\mathrm{ext}}=\sqrt{n}\max_{1\le k\le 2p}\frac{\bar{X}_k^{\mathrm{ext}}}{\hat{\sigma}_k^{\mathrm{ext}}}\quad\text{with}\quad \bar{X}_k^{\mathrm{ext}}=n^{-1}\sum_{i=1}^n X_{i,k}^{\mathrm{ext}},\ \ (\hat{\sigma}_k^{\mathrm{ext}})^2=n^{-1}\sum_{i=1}^n(X_{i,k}^{\mathrm{ext}}-\bar{X}_k^{\mathrm{ext}})^2.$$

Hence, we only need to focus on $T_{\mathrm{s}}^+:=\sqrt{n}\max_{1\le k\le p}\bar{X}_k/\hat{\sigma}_{1k}$ without loss of generality.

Recall that, conditional on $\{\mathbf{X}_1,\ldots,\mathbf{X}_n\}$, $\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}=(W_{\mathrm{s},1}^{(\mathrm{I})},\ldots,W_{\mathrm{s},p}^{(\mathrm{I})})'$ is a centred Gaussian random vector with covariance matrix $\widetilde{\mathbf{R}}_1$. Put

$$Z_{\mathrm{s}}^+=\max_{1\le k\le p}W_{\mathrm{s},k}^{(\mathrm{I})}. \tag{B.19}$$

We shall show that, with high probability, the distribution of T_{s}^+ and the conditional distribution of Z_{s}^+ given $\mathcal{X}_n$ are close enough so that the conditional quantiles of Z_s^+ given $\mathcal{X}_n$ provide reasonable estimates of the quantiles of T_{s}^+. For this purpose, we introduce the following random variables that serve as intermediate approximations.

Let $\{\mathbf{G}_i=(G_{i1},\ldots,G_{ip})'\}_{i=1}^n$ be independent p-variate centered Gaussian random vectors with covariance matrix $\mathbf{R}_1$, and define

$$Z_{\mathrm{s}}^*=\max_{1\le k\le p}n^{-1/2}\sum_{i=1}^n G_{ik},\qquad T_{\mathrm{s}}^*=\max_{1\le k\le p}n^{-1/2}\sum_{i=1}^n U_{ik}, \tag{B.20}$$

where $U_{ik}=(X_{ik}-\mu_{1k})/\sigma_{1k}$. For the random variables Z_{s}^* and T_{s}^*, by Lemma 1 we have the following Berry-Esseen type bound:

$$d_n:=\sup_{x\in\mathbb{R}}\big|\mathbb{P}\big(T_{\mathrm{s}}^*\le x\big)-\mathbb{P}\big(Z_{\mathrm{s}}^*\le x\big)\big|, \tag{B.21}$$

where the order of d_n depends on the moment conditions imposed on $\mathbf{X}_i$ as described in (A.4) and (A.5).

Further, applying Theorem 3 in Chernozhukov, Chetverikov and Kato (2015) to $Z_{\rm s}^*$ gives, for any $\varepsilon > 0$,

$$\sup_{x\in\mathbb{R}} \mathbb{P}\big(|Z_{\rm s}^* - x| \le \varepsilon\big) \le 4\varepsilon \left[1 + \{2\log(p)\}^{1/2}\right]. \tag{B.22}$$

On the other hand, we see that under $H_0^{(\rm I)}$, $T_{\rm s}^* = \sqrt{n}\max_{1\le k\le p} \bar{X}_k/\sigma_{1k}$ and hence

$$\big|T_{\rm s}^+ - T_{\rm s}^*\big| \le \max_{1\le k\le p}\left|\frac{\hat\sigma_{1k}}{\sigma_{1k}} - 1\right| \cdot \max_{1\le k\le p}\frac{\sqrt{n}|\bar X_k|}{\hat\sigma_{1k}} \le \max_{1\le k\le p}\left|\frac{\hat\sigma_{1k}}{\sigma_{1k}} - 1\right| \cdot \sqrt{n}\big|\widehat{\mathbf{D}}_1^{-1/2}\widehat{\boldsymbol\mu}_1\big|_\infty. \tag{B.23}$$

For $t > 0$, define the event

$$\mathcal{E}_0(t) = \left\{\max_{1\le k\le p} |\hat\sigma_{1k}/\sigma_{1k} - 1| \le t\right\}. \tag{B.24}$$

On this event, we have $|T_{\rm s}^+ - T_{\rm s}^*| \le t\sqrt{n}|\widehat{\mathbf{D}}_1^{-1/2}\widehat{\boldsymbol\mu}_1|_\infty$. Together with (iii) in Lemma 3, this yields for any $\varepsilon > 0$ that

$$\mathbb{P}\big(|T_{\rm s}^+ - T_{\rm s}^*| > \varepsilon\big) \le Cp\exp\{-c(\varepsilon/t)^2\} + \mathbb{P}\{\mathcal{E}_0(t)^{\rm c}\}, \tag{B.25}$$

where $\mathcal{E}_0^{\rm c}$ denotes the complimentary set of $\mathcal{E}_0$.

For the two Gaussian maxima $Z_{\rm s}^+$ and $Z_{\rm s}^*$ given in (B.19) and (B.20), respectively, it follows from Lemma 2 that

$$\hat d_n := \sup_{x\in\mathbb{R}}\big|\mathbb{P}\big(Z_{\rm s}^* \le x\big) - \mathbb{P}\big(Z_{\rm s}^+ \le x|\mathcal{X}_n\big)\big| \le C\Theta_{\rm s}^{1/3}\{1 + \log(p/\Theta_s)\}^{2/3}, \tag{B.26}$$

where $\Theta_{\rm s} = \Theta_{\rm s}^{(\rm I)} = \|\widetilde{\mathbf{R}}_1 - \mathbf{R}_1\|_\infty$.

Consequently, combination of inequalities (B.21), (B.22), (B.25) and (B.26) gives, for every $x \in \mathbb{R}$ and $t, \varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\big(T_{\rm s}^+ > x\big)\\
&\le \mathbb{P}\big(T_{\rm s}^* > x - \varepsilon\big) + \mathbb{P}\big(|T_{\rm s}^+ - T_{\rm s}^*| > \varepsilon\big)\\
&\le \mathbb{P}\big(Z_{\rm s}^* > x - \varepsilon\big) + d_n + \mathbb{P}\big(|T_{\rm s}^+ - T_{\rm s}^*| > \varepsilon\big)\\
&\le \mathbb{P}\big(Z_{\rm s}^* > x\big) + C\varepsilon\{\log(p)\}^{1/2} + d_n + \mathbb{P}\big(|T_{\rm s}^+ - T_{\rm s}^*| > \varepsilon\big)\\
&\le \mathbb{P}\big(Z_{\rm s}^+ > x|\mathcal{X}_n\big) + C\varepsilon\{\log(p)\}^{1/2} + d_n + \hat d_n + Cp\exp\{-c(\varepsilon/t)^2\} + \mathbb{P}\{\mathcal{E}_0(t)^{\rm c}\}.
\end{aligned}$$

A similar argument leads to the reverse inequality, which together the previous display implies by taking $\varepsilon = Ct\{\log(pn)\}^{1/2}$ that

$$\sup_{x\in\mathbb{R}}\big|\mathbb{P}\big(T_{\rm s}^+ \le x\big) - \mathbb{P}\big(Z_{\rm s}^+ \le x|\mathcal{X}_n\big)\big| \le d_n + \hat d_n + C\left\{t\log(pn) + n^{-1}\right\} + \mathbb{P}\{\mathcal{E}_0(t)^{\rm c}\}, \tag{B.27}$$

where d_n and $\hat d_n$ are as in (B.21) and (B.26), respectively.

In the case where condition (M1) holds, by (i) in Lemma 3, taking $t \asymp \nu_4^2 n^{-1/2}\{\log(pn)\}^{1/2} + \nu_r^2 \vartheta_{n,p}^{2/(r+2)} + \nu_r^2 \vartheta_{n,p}^{2/r} \log(p)$ in (B.27) yields that $\mathbb{P}\{\mathcal{E}_0(t)^{\rm c}\} \le C\{n^{-1} + \vartheta_{n,p}^{2/(r+2)}\}$. Substituting this into (B.27) proves the first assertion of the proposition in view of (A.4).

Finally, when condition (M2) holds, (ii) in Lemma 3 implies by taking $t \asymp n^{-1/2}\{\log(pn)\}^{1/2} + n^{-1}\{\log(pn)\}^{2/\gamma}$ in (B.27) that $\mathbb{P}\{\mathcal{E}_0(t)^{\rm c}\} \le Cn^{-1}$. Together with (A.5) and (B.27), this completes the proof of the second assertion of the proposition.

□

B.2 Proof of Theorem 1

We now prove Theorem 1. As before, we only give the proof for the test that is based on the studentized statistic. For $\alpha \in (0,1)$ given, recall that $\mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})}$ is the conditional $(1-\alpha)$-quantile of $\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})} \sim N(0, \widehat{\mathbf{R}}_1)$ given $\mathcal{X}_n = \{\mathbf{X}_i\}_{i=1}^n$. Then it follows from (i), Proposition 1 and Lemma 3 that, under condition (M1),

$$\Big| P_{H_0^{(\mathrm{I})}}\{\Psi_{\mathrm{s},\alpha}^{(\mathrm{I})} = 1\} - P\big(|\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}|_\infty > \mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})} | \mathcal{X}_n\big)\Big| \xrightarrow{P} 0,$$

as $n \to \infty$. Applying Theorem 3 in Chernozhukov, Chetverikov and Kato (2015) and Comment 5 after it to $|\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}|_\infty$ implies by (**??**) that

$$\begin{aligned}
\alpha &\ge P\big(|\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}|_\infty > \mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})} | \mathcal{X}_n\big) \\
&\ge P\big(|\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}|_\infty > \mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})} - n^{-1} | \mathcal{X}_n\big) - P\big(\mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})} - n^{-1} \le |\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}|_\infty \le \mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})} | \mathcal{X}_n\big) \\
&\ge \alpha - Cn^{-1} E\big(|\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}|_\infty | \mathcal{X}_n\big) \ge \alpha - Cn^{-1}\{\log(p)\}^{1/2}.
\end{aligned}$$

The last two displays jointly complete the proof of the theorem. □

B.3 Proof of Theorem 2

As in the proof of Proposition 1, we start with $T_{\mathrm{s}} = T_{\mathrm{s}}^{(\mathrm{I})}$. A standard result on Gaussian maximum yields

$$\mathbb{E}\big(|\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}|_\infty | \mathcal{X}_n\big) \le \{2\log(p)\}^{1/2} + \{2\log(p)\}^{-1/2} \le [1 + \{2\log(p)\}^{-1}]\{2\log(p)\}^{1/2}, \tag{B.28}$$

where $\mathcal{X}_n = \{\mathbf{X}_i\}_{i=1}^n$. The following fundamental result (Borell, 1975) shows the concentration of $|\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}|_\infty$ around its mean $\mathbb{E}(|\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}|_\infty | \mathcal{X}_n)$ that for every $u > 0$,

$$\mathbb{P}\big\{|\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}|_\infty \ge \mathbb{E}\big(|\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}|_\infty | \mathcal{X}_n\big) + u \big| \mathcal{X}_n\big\} \le \exp(-u^2/2)$$

Together with (B.28), this implies

$$\mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})} \le [1+\{2\log(p)\}^{-1}]\sqrt{2\log(p)} + \sqrt{2\log(1/\alpha)}. \tag{B.29}$$

Next, recall that $T_{\mathrm{s}} = \max_{1\le k\le p}\max\big(\sqrt{n}\bar{X}_k/\hat{\sigma}_{1k}, -\sqrt{n}\bar{X}_k/\hat{\sigma}_{1k}\big)$, and let $\mathcal{E}_0(t)$ be as in (B.24) for some $0<t\le\frac{1}{2}$. Set $k_0 = \arg\max_{1\le k\le p}|\mu_{1k}|/\sigma_{1k}$, and assume without loss of generality that $\mu_{1k_0}>0$. Then, on the event $\mathcal{E}_0(t)$,

$$T_{\mathrm{s}} \ge \frac{\sqrt{n}\bar{X}_{k_0}}{\hat{\sigma}_{1k_0}} = \frac{\sqrt{n}(\bar{X}_{k_0}-\mu_{1k_0})}{\hat{\sigma}_{1k_0}} + \frac{\sqrt{n}\mu_{1k_0}}{\hat{\sigma}_{1k_0}} \ge \frac{\sqrt{n}(\bar{X}_{k_0}-\mu_{1k_0})}{\hat{\sigma}_{1k_0}} + (1+t)^{-1}\frac{\sqrt{n}\mu_{1k_0}}{\sigma_{1k_0}}.$$

By condition (M1) with $p = O(n^{r/2-1-\delta})$, we have $\vartheta_{n,p} = O(n^{-\delta})$. This, together wiht Lemma 3 implies by taking $t\asymp n^{-2\delta/(r+2)}$ that $\mathbb{P}\{\mathcal{E}_0(t)^{\mathrm{c}}\}\le C\{n^{-1}+n^{-2\delta/(r+2)}\}$. Further, choose $u=u_{n,p}$ in such a way that $(1+t)\big[1+\{\log(p)\}^{-1}+u\big] = 1+\varepsilon_n$, for $\varepsilon_n>0$ satisfying that $\varepsilon_n\to 0$ and $\varepsilon_n\sqrt{\log(p)}\to\infty$. Consequently,

$$\max_{1\le k\le p}\frac{\sqrt{n}|\mu_{1k}|}{\sigma_{1k}} \ge (1+t)\big[1+\{\log(p)\}^{-1}+u\big]\lambda(p,\alpha),$$

where $\lambda(p,\alpha)$ is as in Theorem 2. The last display together with Lemma 3 yields, as $n,p\to\infty$,

$$\begin{aligned}
\mathbb{P}\big(T_{\mathrm{s}} > \mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})}\big) &\ge \mathbb{P}\big(T_{\mathrm{s}}^{+} > [1+\{\log(p)\}^{-1}]\,\lambda(p,\alpha)\big)\\
&\ge \mathbb{P}\big(T_{\mathrm{s}}^{+} > [1+\{\log(p)\}^{-1}]\,\lambda(p,\alpha),\, \mathcal{E}_0(t)\,\big)\\
&\ge 1-\mathbb{P}\big\{\sqrt{n}(\bar{X}_{k_0}-\mu_{1k_0}) < -u\lambda(p,\alpha)\hat{\sigma}_{1k_0}\big\} - \mathbb{P}\big\{\mathcal{E}_0(t)^{\mathrm{c}}\big\}\\
&\ge 1-n^{-1}-C\exp\{-cu^2\lambda^2(p,\alpha)\} - \mathbb{P}\big\{\mathcal{E}_0(t)^{\mathrm{c}}\big\}\\
&\ge 1-n^{-1}-C\exp\{-cu^2\log(p)\} - \mathbb{P}\big\{\mathcal{E}_0(t)^{\mathrm{c}}\big\}\\
&\to 1.
\end{aligned} \tag{B.30}$$

Under condition (M2), taking $t\asymp n^{-1/2}\{\log(pn)\}^{1/2}+n^{-1}\{\log(pn)\}^{2/\gamma}$ instead gives $\mathbb{P}\{\mathcal{E}_0(t)^{\mathrm{c}}\}\le Cn^{-1}$. Then the conclusion follows directly from (B.29) and (B.30).

Next we consider the statistic $T_{\mathrm{ns}} = T_{\mathrm{ns}}^{(\mathrm{I})}$ without studentization. Analogously to (B.28) and the concentration inequality after it, now we have for $\mathbf{W}_{\mathrm{ns}}^{(\mathrm{I})}$,

$$\mathbb{E}\big(|\mathbf{W}_{\mathrm{ns}}^{(\mathrm{I})}|_\infty|\mathcal{X}_n\big) \le \max_{1\le k\le p}\hat{\sigma}_{1k}\cdot[1+\{2\log(p)\}^{-1}]\{2\log(p)\}^{1/2}$$

and for every $u>0$,

$$\mathbb{P}\big\{|\mathbf{W}_{\mathrm{ns}}^{(\mathrm{I})}|_\infty \ge \mathbb{E}\big(|\mathbf{W}_{\mathrm{ns}}^{(\mathrm{I})}|_\infty|\mathcal{X}_n\big)+u\,\big|\mathcal{X}_n\big\} \le \exp\bigg(-\frac{u^2}{2\max_{1\le k\le p}\hat{\sigma}_{1k}^2}\bigg),$$

The last two displays joint imply that on the event $\mathcal{E}_0(t)$,

$$\mathrm{cv}^{(\mathrm{I})}_{\mathrm{ns},\alpha} \le (1+t) \max_{1\le k\le p} \sigma_{1k} \left([1+\{2\log(p)\}^{-1}]\{2\log(p)\}^{1/2} + \{2\log(1/\alpha)\}^{1/2} \right). \tag{B.31}$$

Following the same argument that leads to (B.30), it suffices to estimate the probability that $\mathbb{P}\{\sqrt{n}(\bar{X}_k - \mu_{1k}) > y\}$ for $y>0$. Assuming the uniform boundedness of fourth moments, put $\widehat{X}_{ik} = (X_{ik}-\mu_{1k})I\{|X_{ik}-\mu_{1k}| \le \sigma_{1k}\sqrt{n}\}$, such that $|\mathbb{E}\widehat{X}_{ik} - \mu_{1k}| \le \sigma_{1k}\nu_4^4 n^{-3/2}$. Then it follows from Bernstein's inequality that for $y > 2\sigma_{1k}\nu_4^4 n^{-1}$,

$$\begin{aligned}
&\mathbb{P}\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_{ik} - \mu_{1k}) > y \right\} \\
&\le \mathbb{P}\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\widehat{X}_{ik} - \mathbb{E}\widehat{X}_{ik}\right) > \frac{y}{2} \right\} + n \max_{1\le i\le n} \mathbb{P}\left(|X_{ik}-\mu_{1k}| > \sigma_{1k}\sqrt{n}\right) \\
&\le \exp\left\{ -\frac{1}{C} \min\left(\frac{y^2}{\sigma_{1k}^2}, \frac{y}{\sigma_{1k}} \right) \right\} + \nu_4^4 n^{-1}.
\end{aligned}$$

The rest of the proof is similar to that for $T_{\mathrm{s}}^{(\mathrm{I})}$ and thus is omitted. The proof of Theorem 2 is then complete. □

C Proof of the asymptotic null properties: two-sample case

C.1 Proof of Theorem 3

Recall that $n = \min(n,m)$ and $\lambda = n/m$. The following result extends Proposition 1 and provides a non-asymptotic error bound between the distribution of the two-sample test statistic and that of its Gaussian analogue under the null hypothesis. Throughout the following, we write $\mathcal{X}_n = \{\mathbf{X}_i\}_{i=1}^n$ and $\mathcal{Y}_m = \{\mathbf{Y}_j\}_{j=1}^m$.

Proposition 2. *Let $\nu \in \{\mathrm{ns}, \mathrm{s}\}$, $\Theta^{(\mathrm{II})}_{\mathrm{ns}} = \|\widetilde{\mathbf{\Sigma}}_{1,2} - \mathbf{\Sigma}_{1,2}\|_\infty$ and $\Theta^{(\mathrm{II})}_{\mathrm{s}} = \|\widetilde{\mathbf{R}}_{1,2} - \mathbf{R}_{1,2}\|_\infty$. The estimators $\widetilde{\mathbf{\Sigma}}_1$ and $\widetilde{\mathbf{\Sigma}}_2$ of $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, respectively, are such that diag$(\widetilde{\mathbf{\Sigma}}_1) =$ diag$(\widehat{\mathbf{\Sigma}}_1)$ and diag$(\widetilde{\mathbf{\Sigma}}_2) =$ diag$(\widehat{\mathbf{\Sigma}}_2)$.*

(i) *Assume that* (M1) *holds. Then under $H_0^{(\mathrm{II})}$,*

$$\begin{aligned}
&\sup_{x\ge 0} \left| \mathbb{P}\{T^{(\mathrm{II})}_\nu > x\} - \mathbb{P}\{|\mathbf{W}^{(\mathrm{II})}_\nu|_\infty > x | \mathcal{X}_n, \mathcal{Y}_m\} \right| \\
&\qquad \le C_1 \left[(\Theta^{(\mathrm{II})}_\nu)^{1/3} \{\log(pn)\}^{2/3} + n^{-1/8}\{\log(pn)\}^{7/8} + \vartheta_{n,p}^{1/(r+1)} \{\log(pn)\}^{3/2} \right]
\end{aligned}$$

for $\vartheta_{n,p}$ as in (A.2).

(ii) *Assume that* (M2) *holds. Then under* $H_0^{(\mathrm{II})}$,

$$\sup_{x\geq 0}\left|\mathbb{P}\{T_\nu^{(\mathrm{II})}>x\}-\mathbb{P}\{|\mathbf{W}_\nu^{(\mathrm{II})}|_\infty>x|\mathcal{X}_n,\mathcal{Y}_m\}\right|$$
$$\leq C_2\Big[(\Theta_\nu^{(\mathrm{II})})^{1/3}\{\log(pn)\}^{2/3}+n^{-1/8}\{\log(pn)\}^{7/8}+n^{-1/2}\{\log(pn)\}^{3/2+1/\gamma}\Big].$$

The constants $C_1, C_2 > 0$ *and independent of* n, m, p, r *and* γ.

As a direct consequence of Proposition 2, the proof for the validity of the two-sample tests based on the statistics $T_{\mathrm{ns}}^{(\mathrm{II})}$ and $T_{\mathrm{s}}^{(\mathrm{II})}$ is almost identical to that of Theorem 1, and hence is omitted.

Proof of Proposition 2. The basic idea is essentially in line with that for proving Proposition 1, and we shall only focus on the studentized test statistic $T_{\mathrm{s}} = T_{\mathrm{s}}^{(\mathrm{II})}$ with slight abuse of notation. First, define a pooled sequence of random vectors $\{\boldsymbol{\xi}_i = (\xi_{i1},\ldots,\xi_{ip})'\}_{i=1}^N$ as follows:

$$\xi_{ik}=\begin{cases} X_{ik}-\mu_{1k}, & 1\leq i\leq n,\\ -\lambda(Y_{i-n,k}-\mu_{2k}), & n+1\leq i\leq N,\end{cases}$$

for $N = n + m$ and $\lambda = n/m$ as specified in Section 2. It is easy to see that $\mathbb{E}\boldsymbol{\xi}_i = \mathbf{0}$ and

$$\mathbb{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i')=\begin{cases} \boldsymbol{\Sigma}_1, & 1\leq i\leq n,\\ \lambda^2\boldsymbol{\Sigma}_2, & n+1\leq i\leq N.\end{cases}$$

Moreover, define weighted versions of $\boldsymbol{\xi}_i$:

$$\tilde{\boldsymbol{\xi}}_i=(\tilde{\xi}_{i1},\ldots,\tilde{\xi}_{ip})'=\mathbf{D}^{-1/2}\boldsymbol{\xi}_i \qquad \text{and} \qquad \hat{\boldsymbol{\xi}}_i=(\hat{\xi}_{i1},\ldots,\hat{\xi}_{ip})'=\widehat{\mathbf{D}}^{-1/2}\boldsymbol{\xi}_i,$$

where $\mathbf{D} = \mathrm{diag}(s_1^2,\ldots,s_p^2)$ and $\widehat{\mathbf{D}} = \mathrm{diag}(\hat{s}_1^2,\ldots,\hat{s}_p^2)$ with

$$s_k^2=\frac{n}{N}\left(\sigma_{1k}^2+\lambda\sigma_{2k}^2\right) \quad \text{and} \quad \hat{s}_k^2=\frac{n}{N}\left(\hat{\sigma}_{1k}^2+\lambda\hat{\sigma}_{2k}^2\right). \tag{C.32}$$

In the above notation, we have under $H_0^{(\mathrm{II})}$, $T_{\mathrm{s}} = T_{\mathrm{s}}^{(\mathrm{II})} = N^{-1/2}\max_{1\leq k\leq p}|\sum_{i=1}^N \hat{\xi}_{ik}|$. According to the discussions below (B.18), it is sufficient to focus on on the following statistic:

$$T_{\mathrm{s}}^{+}=\max_{1\leq k\leq p} N^{-1/2}\sum_{i=1}^N \hat{\xi}_{ik}.$$

In the definition of T_{s}^{+}, replacing the variance estimates $\hat{s}_k^2$ with their population analogues leads to $T_{\mathrm{s}}^{*} = N^{-1/2}\max_{1\leq k\leq p}\sum_{i=1}^N \tilde{\xi}_{ik}$, satisfying

$$\left|T_{\mathrm{s}}^{+}-T_{\mathrm{s}}^{*}\right|\leq\max_{1\leq k\leq p}\left|\frac{\hat{s}_k}{s_k}-1\right|\sqrt{n}\left|\widehat{\mathbf{D}}_{1,2}^{-1/2}(\widehat{\boldsymbol{\mu}}_1-\widehat{\boldsymbol{\mu}}_2)\right|_\infty,$$

where $\widehat{\mathbf{D}}_{1,2} = \mathrm{diag}\,(\widehat{\mathbf{\Sigma}}_1 + \lambda\widehat{\mathbf{\Sigma}}_2)$. For $0 < t \le \frac{1}{2}$, define the subset

$$\mathcal{E}_{n,m}(t) = \left\{ \max_{1\le k\le p} \left| \frac{\hat{\sigma}^2_{\nu k}}{\sigma^2_{\nu k}} - 1 \right| \le t, \quad \nu = 1,2 \right\}. \tag{C.33}$$

On this event, we have $1 - t \le (\hat{s}_k/s_k)^2 \le 1 + t$ for all $k = 1, \ldots, p$. By (C.32), we apply the inequality $1 + \frac{1}{2}u - \frac{1}{2}u^2 \le (1+u)^{1/2} \le 1 + \frac{1}{2}u$ that holds for all $u \ge -1$ to obtain

$$\mathcal{E}_{n,m}(t) \subseteq \left\{ \max_{1\le k\le p} \left| \frac{\hat{s}_k}{s_k} - 1 \right| \le \frac{t(1+t)}{2} \right\}.$$

Next, let $\{(G_{i1}, \ldots, G_{ip})'\}_{i=1}^N$ be independent centred Gaussian random vectors with the same covariance matrix as $\tilde{\boldsymbol{\xi}}_i$, and define

$$Z^*_{\mathrm{s}} = \max_{1\le k\le p} N^{-1/2} \sum_{i=1}^N G_{ik}. \tag{C.34}$$

Observe that for each k, $N^{-1}\sum_{i=1}^N \mathbb{E}\tilde{\xi}^2_{ik} = 1$. Hence we use Lemma 1 to bound the Kolmogorov distance between T^*_{s} and Z^*_{s}, i.e.

$$d_{n,m} = \sup_{x\in\mathbb{R}} \left| \mathbb{P}\big(T^*_{\mathrm{s}} \le x\big) - \mathbb{P}\big(Z^*_{\mathrm{s}} \le x\big) \right|,$$

and the anti-concentration inequality (B.22) holds for Z^*_{s} given in (C.34).

Conditional on $\mathcal{X}_n$ and $\mathcal{Y}_m$, the p-vector $\mathbf{W}^{(\mathrm{II})}_{\mathrm{s}} = (W^{(\mathrm{II})}_{\mathrm{s},1}, \ldots, W^{(\mathrm{II})}_{\mathrm{s},p})'$ is centred Gaussian with covariance matrix $\widetilde{\mathbf{R}}_{1,2} = \widetilde{\mathbf{D}}^{-1/2}_{1,2}(\widetilde{\mathbf{\Sigma}}_1 + \lambda\widetilde{\mathbf{\Sigma}}_2)\widetilde{\mathbf{D}}^{-1/2}_{1,2}$, where $\widetilde{\mathbf{D}}_{1,2} = \mathrm{diag}\,(\widetilde{\mathbf{\Sigma}}_1 + \lambda\widetilde{\mathbf{\Sigma}}_2) = \mathrm{diag}\,(\widehat{\mathbf{\Sigma}}_1 + \lambda\widehat{\mathbf{\Sigma}}_2) = \widehat{\mathbf{D}}_{1,2}$. Note that Z^*_s given in (C.34) is the maximum of the p coordinates of a centred Gaussian random vector with covariance matrix $\mathbf{R}_{1,2}$. Once again, using Lemma 2 gives

$$\hat{d}_{n,m} = \sup_{x\in\mathbb{R}} \left| \mathbb{P}\big(Z^*_{\mathrm{s}} \le x\big) - \mathbb{P}\big(Z^+_{\mathrm{s}} \le x | \mathcal{X}_n, \mathcal{Y}_m\big) \right| \le C\Theta^{1/3}_{\mathrm{s}}\{1 + \log(p/\Theta_{\mathrm{s}})\}^{2/3},$$

where $Z^+_{\mathrm{s}} = \max_{1\le k\le p} W^{(\mathrm{II})}_{\mathrm{s},k}$ and $\Theta_{\mathrm{s}} = \Theta^{(\mathrm{II})}_{\mathrm{s}} = |\widetilde{\mathbf{R}}_{1,2} - \mathbf{R}_{1,2}|_\infty$.

The rest of the proof is similar to the one-sample case and thus is omitted. □

C.2 Proof of Theorem 4

To begin with, we note that inequality (B.29) holds for $\mathrm{cv}^{(\mathrm{II})}_{\mathrm{s},\alpha}$, the conditional $(1-\alpha)$-quantile of $\mathbf{W}^{(\mathrm{II})}_{\mathrm{s}}$. Define $k_0 = \arg\max_{1\le k\le p} \sigma^{-1}_k |\mu_{1k} - \mu_{2k}|$ with $\sigma^2_k = \sigma^2_{1k}/n + \sigma^2_{2k}/m$. Put $\hat{\sigma}^2_k = \hat{\sigma}^2_{1k}/n + \hat{\sigma}^2_{2k}/m$, and assume without loss of generality that $\mu_{1k_0} > \mu_{2k_0}$. Then on the event $\mathcal{E}_{n,m}(t)$ given in (C.33) for $0 < t \le \frac{1}{2}$,

$$T_{\mathrm{s}}^{(\mathrm{II})} \ge \frac{(\bar{X}_{k_0} - \mu_{1k_0}) - (\bar{Y}_{k_0} - \mu_{2k_0})}{\hat{\sigma}_{k_0}} + \frac{\mu_{1k_0} - \mu_{2k_0}}{\hat{\sigma}_{k_0}}$$
$$\ge \frac{(\bar{X}_{k_0} - \mu_{1k_0}) - (\bar{Y}_{k_0} - \mu_{2k_0})}{\hat{\sigma}_{k_0}} + (1 + t/2)^{-1} \frac{\mu_{1k_0} - \mu_{2k_0}}{\sigma_{k_0}}.$$

A chain of inequalities that are similar to those in (B.30) hold for $\mathbb{P}\{T_{\mathrm{s}}^{(\mathrm{II})} > \mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{II})}\}$, and thus the conclusion in (ii) follows.

For the non-studentized statistic $T_{\mathrm{ns}}^{(\mathrm{II})}$, an argument similar to that leading to (B.31) implies $\mathrm{cv}_{\mathrm{ns},\alpha}^{(\mathrm{II})} \le [1 + \{\log(p)\}^{-1}]\sqrt{n}\,\lambda(p,\alpha) \cdot \max_{1 \le k \le p} \hat{\sigma}_k$. The rest of the proof is almost immediate. □

D Proof of Theorem 5

Let $\varrho = \sqrt{2} + \frac{\sqrt{2}}{2\log p} + \sqrt{\frac{2\log(1/\alpha)}{\log p}}$. First we prove the limiting null property for $\Psi_{\mathrm{s},\alpha}^{f,(\mathrm{I})}$. To begin with, define the event $\mathcal{E}_{1n} = \{\widehat{\mathcal{S}}_1 = \{1,\ldots,p\}\}$ and observe that under $H_0^{(\mathrm{I})}$, $\mathbb{P}_{H_0^{(\mathrm{I})}}(\mathcal{E}_{1n}^{\mathrm{c}}) \le \mathbb{P}_{H_0^{(\mathrm{I})}}\{T_{\mathrm{s}}^{(\mathrm{I})} > \varrho\sqrt{\log(p)}\}$. This, together with (B.29) and Theorem 1, implies $\limsup_{n\to\infty} \mathbb{P}_{H_0^{(\mathrm{I})}}(\mathcal{E}_{1n}^{\mathrm{c}}) \le \alpha$. On the other hand, it is easy to see that $\Psi_{\nu,\alpha}^{f,(\mathrm{I})} = 0$ on the event $\mathcal{E}_{1n}$ for $\nu \in \{\mathrm{s},\mathrm{ns}\}$. Together with the inequality $\mathbb{P}_{H_0^{(\mathrm{I})}}\{\Psi_{\nu,\alpha}^{f,(\mathrm{I})} = 1\} \le \mathbb{P}_{H_0^{(\mathrm{I})}}\{\Psi_{\nu,\alpha}^{f,(\mathrm{I})} = 1, \mathcal{E}_{1n}\} + \mathbb{P}_{H_0^{(\mathrm{I})}}(\mathcal{E}_{1n}^{\mathrm{c}})$, this completes the proof of (i).

Next we study the power of $\Psi_{\mathrm{s},\alpha}^{f,(\mathrm{I})}$ under $H_1^{(\mathrm{I})}$. Let $\widehat{T}_{\mathrm{s}}^{(\mathrm{I})} = \max_{k \notin \widehat{\mathcal{S}}_1} \sqrt{n}|\bar{X}_k|/\hat{\sigma}_{1k}$ and define the event $\mathcal{E}_{2n} = \{T_{\mathrm{s}}^{(\mathrm{I})} = \widehat{T}_{\mathrm{s}}^{(\mathrm{I})}\}$. By (B.29) and (B.30), we have $\lim_{n\to\infty} \mathbb{P}_{H_1^{(\mathrm{I})}}\{T_{\mathrm{s}}^{(\mathrm{I})} > \varrho\sqrt{\log(p)}\} \to 1$, which further implies

$$\lim_{n\to\infty} \mathbb{P}_{H_1^{(\mathrm{I})}}(\mathcal{E}_{2n}^{\mathrm{c}}) = 0. \tag{D.35}$$

Further, since $\widehat{\mathcal{S}}_1 \subseteq \{1,\ldots,p\}$, $\mathbb{P}\{\max_{k\notin\widehat{\mathcal{S}}_1} |W_{\mathrm{s},k}^{(\mathrm{I})}| > \mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})}\} \le \mathbb{P}\{|\mathbf{W}_{\mathrm{s}}^{(\mathrm{I})}|_\infty > \mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})}\} \le \alpha$, and hence $\mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})}(\widehat{\mathcal{S}}_1) \le \mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})}$. Together with (D.35), this leads to

$$\begin{aligned}
\mathbb{P}_{H_1^{(\mathrm{I})}}\{\Psi_{\mathrm{s},\alpha}^{f,(\mathrm{I})} = 1\} &\ge \mathbb{P}_{H_1^{(\mathrm{I})}}\{\Psi_{\mathrm{s},\alpha}^{f,(\mathrm{I})} = 1,\, \mathcal{E}_{2n}\} \\
&\ge \mathbb{P}_{H_1^{(\mathrm{I})}}\{T_{\mathrm{s}}^{(\mathrm{I})} > \mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})},\, \mathcal{E}_{2n}\} \\
&\ge \mathbb{P}_{H_1^{(\mathrm{I})}}\{T_{\mathrm{s}}^{(\mathrm{I})} > \mathrm{cv}_{\mathrm{s},\alpha}^{(\mathrm{I})}\} - \mathbb{P}_{H_1^{(\mathrm{I})}}(\mathcal{E}_{2n}^{\mathrm{c}}) \\
&\to 1
\end{aligned}$$

as $n \to \infty$. Similarly, we can construct the result for its non-studentized analogue $\Psi_{\mathrm{ns},\alpha}^{f,(\mathrm{I})}$. Hence, we completes the proof of Theorem 5. □

E Proof of Theorem 6

The proof of Theorem 6 is almost identical to that of Theorem 5, and therefore is omitted. □

F More simulation results

F.1 Additional Models

In the section, we introduce two addtional models for both one-sample and two-sample cases, which will be considered in simulations in the following sections.

One-Sample models:

(a) Model $4^{(\mathrm{I})}$ (Block diagonal $\mathbf{\Sigma}_1$): $\sigma_{1,kk}$ are independent and identically drawn from $\mathrm{Unif}(2,3)$, $\sigma_{1,k\ell} = 0.7$ for $10(t-1)+1 \le k \ne \ell \le 10t$, where $t = 1, \ldots, \lfloor p/10 \rfloor$, and $\sigma_{1,k\ell} = 0$ otherwise.

(b) Model $5^{(\mathrm{I})}$ (Moving average process with Beta distributed innovations): For $i = 1, \ldots, n$ and $k = 1, \ldots, p$, we considered $X_{ik} = \rho_1 Z_{i,k} + \rho_2 Z_{i,k+1} + \cdots + \rho_p Z_{i,k+p-1} + \mu_k$ where ρ_ℓ are independent and identically drawn from $0.6\mathrm{Unif}(-1,1) + 0.4\delta_0$ for $\ell = 1, \ldots, p$, where δ_0 is the point mass at 0 and $\{Z_{i,k}\}$ are independent random variables with a common centered $\mathrm{Beta}(2,1)$ distribution.

Two-samples models:

(a) Model $4^{(\mathrm{II})}$ (Long range dependence): Let $\theta_{11}, \ldots, \theta_{1p}, \theta_{21}, \ldots, \theta_{2p}$ be independent and identically drawn from $\mathrm{Unif}(1,2)$; for $q = 1, 2$, we took $\sigma_{q,kk} = \theta_{qk}$ and $\sigma_{q,k\ell} = \rho_\alpha(|k-\ell|)$ for $k \ne \ell$, where $\rho_\alpha(e) = \frac{1}{2}\{(e+1)^{2H} + (e-1)^{2H} - 2e^{2H}\}$ with $H = 0.9$.

(b) Model $5^{(\mathrm{II})}$ (Moving average process with Gamma distributed innovations): Generate $X_{ik} = \rho_{1,1} Z_{i,k} + \rho_{1,2} Z_{i,k+1} + \cdots + \rho_{1,p} Z_{i,k+p-1} + \mu_{1k}$ and $Y_{jk} = \rho_{2,1} \widetilde{Z}_{j,k} + \rho_{2,2} \widetilde{Z}_{j,k+1} + \cdots + \rho_{2,p} \widetilde{Z}_{j,k+p-1} + \mu_{2k}$, where $\rho_{1,\ell}$ are i.i.d. drawn from $0.6\mathrm{Unif}(-1,1) + 0.4\delta_0$ and $\rho_{2,\ell}$ are i.i.d. drawn from $0.8\mathrm{Unif}(-1,1) + 0.2\delta_0$ for $\ell = 1, \ldots, p$, where δ_0 is the point mass at 0, $\{Z_{i,k}\}$ and $\{\widetilde{Z}_{j,k}\}$ are independent random variables with a common centered $\mathrm{Gamma}(1,4)$ and $\mathrm{Gamma}(4,1)$ distributions, respectively.

F.2 More results on empirical size

In this subsection, we report the empirical size (Table S1 to S3) of the proposed tests at the 0.05 nominal level the for $p = 120, 360$ and 1080, along with those of the competing tests. Table S1 and S2 presents the empirical size under the additional models in Section F.1 and Table S3 presents the empirical size with larger sample size ($n = 200$).

	Model $4^{(\mathrm{I})}$			Model $5^{(\mathrm{I})}$		
tests / p	120	360	1080	120	360	1080
			$n=40$			
$\Psi_{\mathrm{ns},\alpha}$	0.035	0.033	0.020	0.044	0.023	0.023
$\Psi_{\mathrm{s},\alpha}$	0.094	0.114	0.216	0.094	0.190	0.252
$\Psi^{f}_{\mathrm{ns},\alpha}$	0.041	0.045	0.030	0.055	0.030	0.032
$\Psi^{f}_{\mathrm{s},\alpha}$	0.093	0.152	0.226	0.095	0.205	0.305
ZCX	1	1	1	0.389	1	1
HC	0.128	0.240	0.315	0.142	0.232	0.329
			$n=80$			
$\Psi_{\mathrm{ns},\alpha}$	0.043	0.032	0.025	0.034	0.039	0.036
$\Psi_{\mathrm{s},\alpha}$	0.073	0.092	0.091	0.083	0.090	0.125
$\Psi^{f}_{\mathrm{ns},\alpha}$	0.053	0.043	0.034	0.046	0.054	0.045
$\Psi^{f}_{\mathrm{s},\alpha}$	0.088	0.091	0.106	0.091	0.098	0.157
ZCX	1	1	1	0.593	1	1
HC	0.082	0.127	0.128	0.072	0.119	0.139

Table S1: Empirical sizes of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^{f}_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^{f}_{\mathrm{s},\alpha}$) for the one-sample problem (1.1), along with those of the tests by Zhong, Chen and Xu (2013) (ZCX), and Donoho and Jin (2004) (HC) at 5% nominal significance. Models with Gaussian data and block diagonal, and the moving average model with Beta distributed innovations are considered when $n = 40, 80$ and $p = 120, 360, 1080$.

F.3 Empirical power of one-sample case

In this subsection, we report the empirical powers (Figures S1 to S5) of the proposed tests against different alternatives in Model $1^{(\mathrm{I})}$ to Model $5^{(\mathrm{I})}$ for $p = 120, 360$, along with those of the test by Zhong, Chen and Xu (2013) (ZCX). We also show the power performance (Figure S6) for $p = 1080$ with the additional one-sample models in Section F.1 .

F.4 Empirical power of two-sample case

In this subsection, we report the empirical powers of the proposed tests (Figures S7 to S11) against different alternatives in Model $1^{(\mathrm{II})}$ to Model $5^{(\mathrm{II})}$ for $p = 120, 360$, along with those of the tests by Chen and Qin (2010) (CQ), Cai, Liu and Xia (2014) (CLX) at 5% nominal significance. We also show the power performance (Figure S12) for p = 1080 with the additional two-sample models in Section F.1.

	Model $4^{(\mathrm{II})}$			Model $5^{(\mathrm{II})}$		
tests / p	120	360	1080	120	360	1080
			$(n,m)=(40,40)$			
$\Psi_{\mathrm{ns},\alpha}$ 0.051	0.036	0.030	0.034	0.030	0.032	
$\Psi_{\mathrm{s},\alpha}$	0.096	0.095	0.130	0.094	0.141	0.158
$\Psi^f_{\mathrm{ns},\alpha}$	0.063	0.045	0.044	0.044	0.038	0.045
$\Psi^f_{\mathrm{s},\alpha}$	0.092	0.093	0.160	0.092	0.174	0.180
HC	0.086	0.157	0.191	0.108	0.191	0.219
CQ	0.046	0.059	0.051	0.046	0.047	0.051
CLX	0.085	0.093	0.116	0.081	0.127	0.157
			$(n,m)=(80,80)$			
$\Psi_{\mathrm{ns},\alpha}$	0.042	0.038	0.049	0.041	0.036	0.035
$\Psi_{\mathrm{s},\alpha}$	0.064	0.071	0.090	0.070	0.090	0.097
$\Psi^f_{\mathrm{ns},\alpha}$	0.052	0.044	0.063	0.046	0.056	0.045
$\Psi^f_{\mathrm{s},\alpha}$	0.073	0.084	0.093	0.086	0.092	0.099
HC	0.066	0.083	0.115	0.065	0.100	0.109
CQ	0.038	0.044	0.049	0.051	0.053	0.053
CLX	0.058	0.060	0.083	0.059	0.087	0.098

Table S2: Empirical sizes of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^f_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^f_{\mathrm{s},\alpha}$) for the two-sample problem (1.2), along with those of the tests by Donoho and Jin (2004) (HC), Chen and Qin (2010) (CQ), and Cai, Liu and Xia (2014) (CLX) at 5% nominal significance. Models with Gaussian data and long range dependence covariance matrices, and the moving average processes model with Gamma distributed innovations are considered when $n = m = 40, 80$ and $p = 120, 360, 1080$.

F.5 Numerical experiments on perfect correlated variables

We conducted extra numerical experiments to demonstrate that the proposed tests can be applied to perfectly correlated variables. We have the results summarized in Tables S4 and S5. The model under considerations are as following, where the settings are similar to Section 4.

(i) (One-sample problem): We considered independent random variables $\mathbf{X}_i = \mathbf{\Gamma}\mathbf{Z}_i$ with $\mathbf{\Gamma}\mathbf{\Gamma}' = \mathbf{\Sigma}_1$ where $\mathbf{Z}_i$ consists of p independent centred Gamma$(4,1)$ random variables. For the covariance matrix $\mathbf{\Sigma}_1 = (\sigma_{1,k\ell})_{1\le k,\ell\le p}$, let $\sigma_{1,kk} = 1$; randomly drew an index set $\mathcal{J} = \{s_1, \cdots, s_{\lfloor p/20\rfloor}\} \subset \{1,\ldots,\lfloor p/10\rfloor\}$, let $\sigma_{1,k\ell} = 1$ for $10(t-1)+1 \le k \ne \ell \le 10t$ given $t \in \mathcal{J}$ and let $\sigma_{1,k\ell} = 0.7$ for $10(t-1)+1 \le k \ne \ell \le 10t$ when $t \in \{1,\ldots,\lfloor p/10\rfloor\} \setminus \mathcal{J}$; and $\sigma_{1,k\ell} = 0$ otherwise.

	Model $1^{(I)}$/ $1^{(II)}$			Model $4^{(I)}$/ $2^{(II)}$		
tests / p	120	360	1080	120	360	1080
One Sample Tests						
$\Psi_{\text{ns},\alpha}$	0.051	0.045	0.046	0.062	0.057	0.051
$\Psi_{\text{s},\alpha}$	0.071	0.077	0.082	0.079	0.085	0.084
$\Psi^{f}_{\text{ns},\alpha}$	0.038	0.031	0.039	0.051	0.045	0.043
$\Psi^{f}_{\text{s},\alpha}$	0.057	0.060	0.067	0.066	0.073	0.063
ZCX	0.077	0.068	0.084	1	1	1
[0.5ex] HC	0.045	0.063	0.075	0.056	0.080	0.071
Two Sample Tests						
$\Psi_{\text{ns},\alpha}$	0.061	0.067	0.066	0.053	0.072	0.056
$\Psi_{\text{s},\alpha}$	0.076	0.079	0.71	0.075	0.073	0.077
$\Psi^{f}_{\text{ns},\alpha}$	0.051	0.049	0.049	0.042	0.052	0.045
$\Psi^{f}_{\text{s},\alpha}$	0.063	0.067	0.058	0.063	0.063	0.065
HC	0.054	0.073	0.061	0.047	0.065	0.062
CQ	0.056	0.043	0.047	0.055	0.050	0.062
CLX	0.078	0.103	0.117	0.055	0.067	0.064

Table S3: Empirical sizes of the proposed tests (non-studentized without screening $\Psi_{\text{ns},\alpha}$, studentized without screening $\Psi_{\text{s},\alpha}$, non-studentized with screening $\Psi^{f}_{\text{ns},\alpha}$, and studenzied with screening $\Psi^{f}_{\text{s},\alpha}$) for the one-sample and two-sample problems, along with those of the tests by Zhong, Chen and Xu (2013) (ZCX), Donoho and Jin (2004) (HC), Chen and Qin (2010) (CQ), and Cai, Liu and Xia (2014) (CLX) at 5% nominal significance. Models with block diagonal and non-sparse covariance matrices are considered, where $n = 200$ or $(n, m) = (200, 200)$ and $p = 120, 360, 1080$.

(ii) (Two-sample problem): We considered independent and identically distributed p-variate random vectors $\mathbf{X}_i \sim t_{\omega_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \mathbf{Y}_j \sim t_{\omega_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ for $i = 1, \cdots, n$ and $j = 1, \cdots, m$, where $t_{\omega_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $t_{\omega_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ are the non-central multivariate t-distributions with non-central parameters $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$, degree freedoms $\omega_1 = 5, \omega_2 = 7$. For the covariance matrices $\boldsymbol{\Sigma}_1 = (\sigma_{1,k\ell})_{1\le k,\ell\le p}$ and $\boldsymbol{\Sigma}_2 = (\sigma_{2,k\ell})_{1\le k,\ell\le p}$, we randomly drew an index set $\mathcal{J} = \{s_1, \cdots, s_{\lfloor p/20\rfloor}\} \subset \{1, \ldots, \lfloor p/10\rfloor\}$, and let $\sigma_{1,kk} = \sigma_{2,kk} = 1$, let $\sigma_{1,k\ell} = \sigma_{2,k\ell} = 1$ for $10(t-1)+1 \le k \ne \ell \le 10t$ given $t \in \mathcal{J}$ and let $\sigma_{1,k\ell} = 0.7$ and $\sigma_{2,k\ell} = 0.85$ for $10(t-1)+1 \le k \ne \ell \le 10t$ when $t \in \{1, \ldots, \lfloor p/10\rfloor\} \setminus \mathcal{J}$, and $\sigma_{1,k\ell} = \sigma_{2,k\ell} = 0$ otherwise.

The alternatives for examining the empirical powers were similar to those in Sections 4.1 and 4.2. That is, under the alternative, there were $\lfloor kp^r \rfloor$ coordinates presenting signals whose strengths are $\{2\beta\sigma_{1,\ell\ell}\log(p)/n\}^{1/2}$ for the one-sample problem and $\{2\beta\sigma_{\ell\ell}\log(p)(1/n+1/m)\}^{1/2}$ with $\sigma_{\ell\ell}$ the ℓth diagonal entry of the pooled covariance $\boldsymbol{\Sigma}_{1,2}$. Particularly, we considered four alternatives: $r = 0$ (with $k = 8$) with $\beta = 0.2, 0.6$ and $\beta = 0.01$ with $r = 0.5, 0.85$. From these simulation studies, we observe that the proposed tests are applicable to variables with perfect correlations and the empirical performance are consistent with results reported in Section 4.

		p	$\Psi_{\text{ns},\alpha}$	$\Psi_{\text{s},\alpha}$	$\Psi^{f}_{\text{ns},\alpha}$	$\Psi^{f}_{\text{s},\alpha}$
Size	$\beta = 0$	120	0.048	0.054	0.058	0.100
		360	0.038	0.044	0.086	0.134
		1080	0.036	0.048	0.098	0.152
Power	$\beta = 0.01, r = 0.85$	120	0.056	0.102	0.152	0.214
		360	0.038	0.088	0.116	0.248
		1080	0.060	0.150	0.146	0.320
	$\beta = 0.01, r = 0.5$	120	0.044	0.088	0.148	0.241
		360	0.050	0.108	0.136	0.234
		1080	0.024	0.148	0.122	0.334
	$\beta = 0.2, r = 0\ (k = 8)$	120	0.090	0.118	0.214	0.272
		360	0.084	0.136	0.206	0.288
		1080	0.046	0.164	0.178	0.400
	$\beta = 0.6, r = 0\ (k = 8)$	120	0.292	0.330	0.506	0.530
		360	0.242	0.308	0.470	0.530
		1080	0.186	0.278	0.416	0.518

Table S4: Empirical performance of the proposed tests (non-studentized without screening $\Psi_{\text{ns},\alpha}$, studentized without screening $\Psi_{\text{s},\alpha}$, non-studentized with screening $\Psi^{f}_{\text{ns},\alpha}$, and studenzied with screening $\Psi^{f}_{\text{s},\alpha}$) for perfectly correlated variables in one-sample problems $H_0^{(\mathrm{I})} : \boldsymbol{\mu}_1 = \mathbf{0}$ *versus* $H_1^{(\mathrm{I})} : \boldsymbol{\mu}_1 \neq \mathbf{0}$. The empirical powers are against the alternatives with different levels of the signal strength (β) and sparsity ($1 - r$) at 5% nominal significance. In the simulation, $n = 80$.

F.6 Numerical experiments on utilizing regularized covariance estimations

Though the proposed tests can directly employ the sample covariance matrices without any structural assumptions, they may still benefit from using regularized covariance estimations when the covariance matrices do have special structures such as banding or sparsity. We illustrate this by conducting two small simulations, which are displayed in Tables S6 and S7, where we considered the Model$^{(\mathrm{I})}$ with banding structures. It can be conclude from the results that using the banding estimator for the covariance matrices in the proposed tests will increase the performace of empirical sizes.

		p	$\Psi_{\mathrm{ns},\alpha}$	$\Psi_{\mathrm{s},\alpha}$	$\Psi^f_{\mathrm{ns},\alpha}$	$\Psi^f_{\mathrm{s},\alpha}$
Size	$\beta=0$	120	0.040	0.070	0.052	0.078
		360	0.040	0.050	0.048	0.086
		1080	0.042	0.090	0.060	0.094
Power	$\beta=0.01, r=0.85$	120	0.056	0.094	0.074	0.088
		360	0.042	0.084	0.062	0.070
	$\beta=0.01, r=0.5$	120	0.060	0.082	0.052	0.060
		360	0.052	0.070	0.046	0.060
	$\beta=0.2, r=0$ $(k=8)$	120	0.092	0.128	0.140	0.198
		360	0.084	0.114	0.132	0.182
	$\beta=0.6, r=0$ $(k=8)$	120	0.472	0.564	0.536	0.588
		360	0.416	0.520	0.514	0.548

Table S5: Empirical performance of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^f_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^f_{\mathrm{s},\alpha}$) for perfectly correlated variables in two-sample problems $H_0^{(\mathrm{II})}:\boldsymbol{\mu}_1=\boldsymbol{\mu}_2$ *versus* $H_1^{(\mathrm{II})}:\boldsymbol{\mu}_1\neq\boldsymbol{\mu}_2$. The empirical powers are against the alternatives with different levels of the signal strength (β) and sparsity $(1-r)$ at 5% nominal significance. In the simulation, $n=m=80$.

	Non-Banding		Banding	
tests / p	400	800	400	800
		$n=40$		
$\Psi_{\mathrm{ns},\alpha}$	0.022	0.034	0.024	0.034
$\Psi_{\mathrm{s},\alpha}$	0.218	0.300	0.168	0.186
$\Psi^f_{\mathrm{ns},\alpha}$	0.010	0.020	0.020	0.018
$\Psi^f_{\mathrm{s},\alpha}$	0.178	0.264	0.146	0.156
		$n=80$		
$\Psi_{\mathrm{ns},\alpha}$	0.054	0.038	0.046	0.042
$\Psi_{\mathrm{s},\alpha}$	0.086	0.130	0.074	0.080
$\Psi^f_{\mathrm{ns},\alpha}$	0.042	0.028	0.044	0.036
$\Psi^f_{\mathrm{s},\alpha}$	0.072	0.108	0.064	0.056

Table S6: Empirical sizes of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^f_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^f_{\mathrm{s},\alpha}$) for the one-sample problems under the Model $1^{(\mathrm{I})}$ (sparse covariance) with sample covariance matrices and its banding estimates, where $n=40,80$ and $p=400,800$.

	Non-Banding			Banding		
tests / p	120	360	1080	120	360	1080
$\Psi_{\mathrm{ns},\alpha}$	0.051	0.045	0.046	0.048	0.052	0.0480
$\Psi_{\mathrm{s},\alpha}$	0.071	0.077	0.082	0.066	0.074	0.0680
$\Psi^f_{\mathrm{ns},\alpha}$	0.038	0.031	0.039	0.040	0.040	0.044
$\Psi^f_{\mathrm{s},\alpha}$	0.057	0.060	0.067	0.052	0.054	0.058

Table S7: Empirical sizes of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^f_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^f_{\mathrm{s},\alpha}$) for the one-sample problems under the Model $1^{(\mathrm{I})}$ (sparse covariance) with sample covariance matrices and its banding estimates, where $n = 200$ and $p = 120, 360, 1080$. at 5% nominal significance.

G More results on real data analysis

In the following tables, for the ALL data, we display the top 15 gene-sets in the CC and MF categories that were identified to te the diseases associated by our propose two-step test without studentization, $\Psi^f_{\mathrm{ns},\alpha}$, but not by the test in Chen and Qin (2010).

GO ID	GO term description
GO:0000323	lytic vacuole
GO:0005758	mitochondrial intermembrane space
GO:0009295	nucleoid
GO:0030863	cortical cytoskeleton
GO:0032991	macromolecular complex
GO:1990234	transferase complex
GO:0005765	lysosomal membrane
GO:0009898	cytoplasmic side of plasma membrane
GO:0016529	sarcoplasmic reticulum
GO:0044439	peroxisomal part
GO:0048770	pigment granule
GO:0045178	basal part of cell
GO:0005768	endosome
GO:0009986	cell surface
GO:0031907	microbody lumen

Table S8: Top 15 gene-sets in the CC category that were identified to be BCR/ABL associated by the proposed two-step test, $\Psi^f_{\mathrm{ns},\alpha}$, but not by the test by Chen and Qin (2010). The significance level was at 0.05 and the FDR was controlled at 0.015.

References

Borell, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Inventiones mathematicae*, **30**, 207–216.

GO ID	GO term description
GO:0004252	serine-type endopeptidase activity
GO:0015485	cholesterol binding
GO:0042625	ATPase activity, coupled to transmembrane movement of ions
GO:0042626	ATPase activity, coupled to transmembrane movement of substances
GO:0061135	endopeptidase regulator activity
GO:1901681	sulfur compound binding
GO:0004860	protein kinase inhibitor activity
GO:0008013	beta-catenin binding
GO:0008022	protein C-terminus binding
GO:0008528	G-protein coupled peptide receptor activity
GO:0016684	oxidoreductase activity, acting on peroxide as acceptor
GO:0001882	nucleoside binding
GO:0016788	hydrolase activity, acting on ester bonds
GO:0030374	ligand-dependent nuclear receptor transcription coactivator activity
GO:0032182	ubiquitin-like protein binding

Table S9: Top 15 gene-sets in the MF category that were identified to be BCR/ABL associated by the proposed two-step test, $\Psi^{f}_{\mathrm{ns},\alpha}$, but not by the test by Chen and Qin (2010). The significance level was at 0.05 and the FDR was controlled at 0.015.

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

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*, **76**, 349–372.

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

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, **41**, 2786–2819.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields*, **162**, 47–70.

Chiaretti, S., Li, X., Gentleman, R., Vitale, A., Vignetti, M., Mandelli, F., Ritz, J. and Foa, R. (2004). Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. *Blood*, **103**, 2771–2778.

de la Peña, V. H., Lai, T. L. and Shao, Q.-M. (2009). *Self-Normalized Processes: Limit Theory and Statistical Applications.* Springer, Berlin.

Delaigle, A., Hall, P. and Jin, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student's t-statistic. *Journal of the Royal Statistical Society, Series B*, **73**, 283–301.

Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, **32**, 962–994.

Einmahl, U. and Li, D. (2008). Characterization of LIL behavior in Banach space. *Transactions of the American Mathematical Society*, **360**, 6677–6693.

Li, J. and Siegumnd, D. (2015). Higher criticism: p-values and criticism. *The Annals of Statistics*, **43**, 1323–1350.

Jing, B.-Y., Shao, Q.-M. and Wang, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *The Annals of Probability*, **31**, 2167–2215.

Zhong, P.-S., Chen, S. X. and Xu, M. (2013). Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. *The Annals of Statistics*, **41**, 2820–2851.

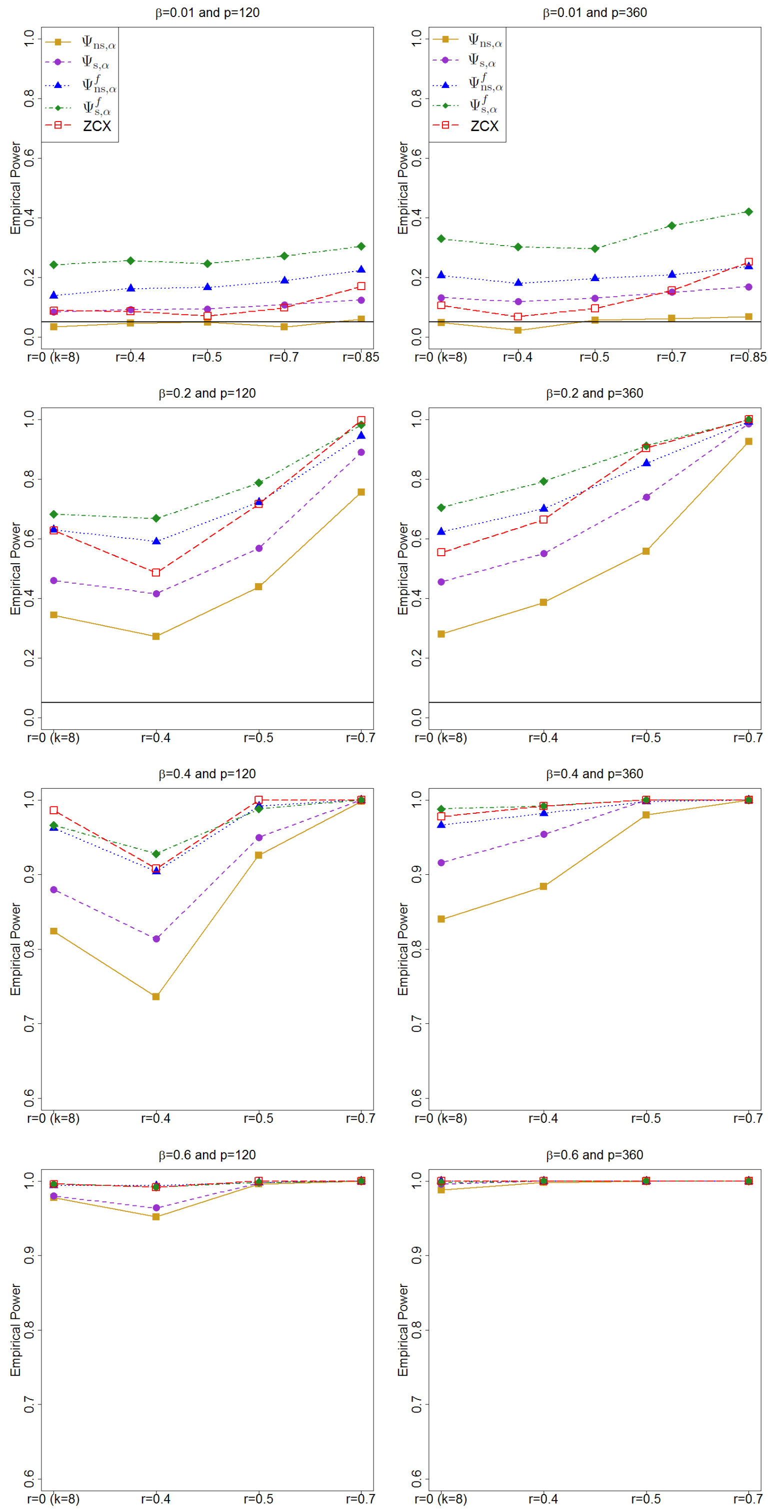


Figure S1: Empirical powers of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized with screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^{f}_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^{f}_{\mathrm{s},\alpha}$) against alternatives with different levels of signal strength (β) and sparsity ($1-r$) for the one-sample problem (1.1), along with the power of the test by Zhong, Chen and Xu (2013) (ZCX) at 5% nominal significance for the Gaussian data and bandable covariance matrices in Model $1^{(\mathrm{I})}$ when $n = 80$ and $p = 120, 360$.

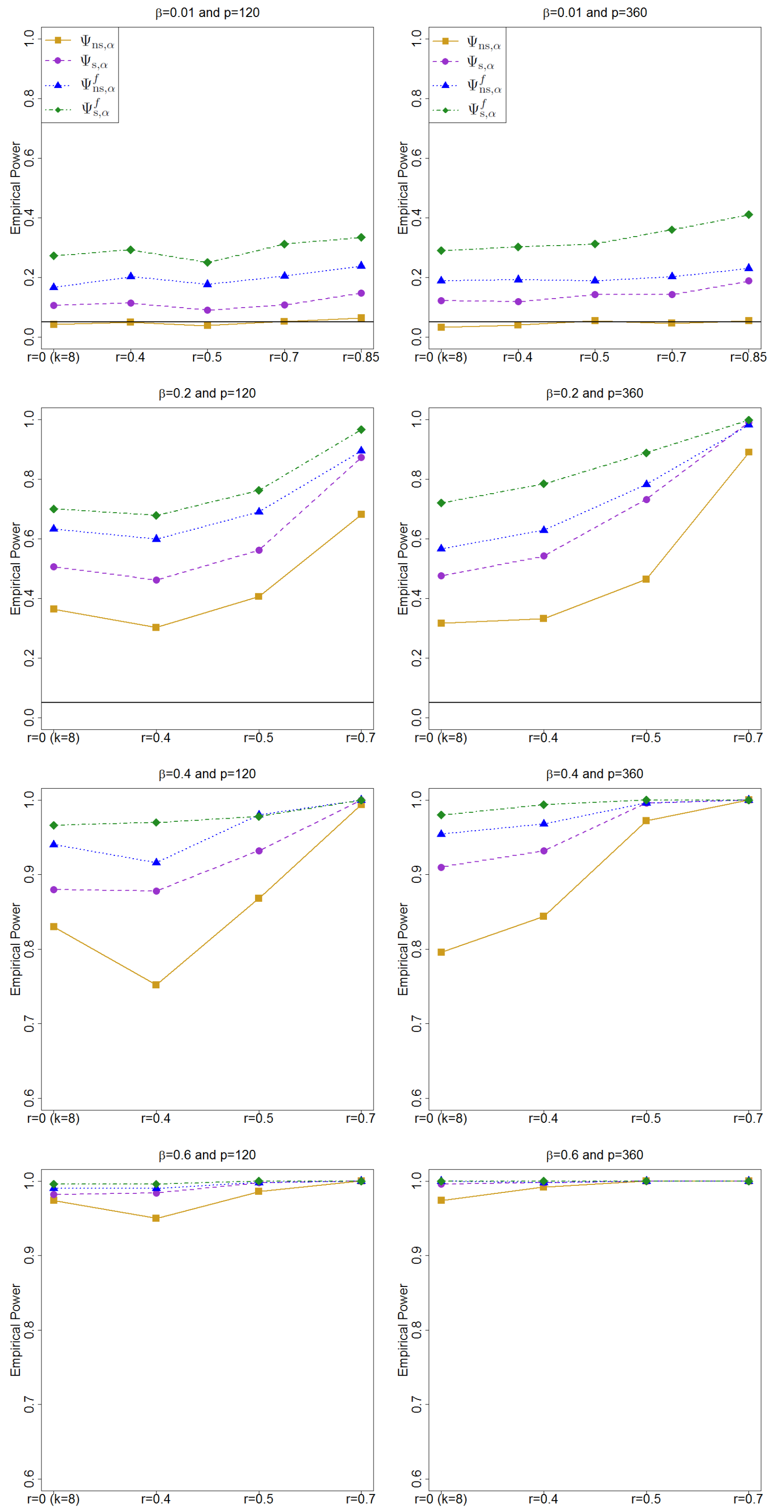


Figure S2: Empirical powers of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^{f}_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^{f}_{\mathrm{s},\alpha}$) against alternatives with different levels of the signal strength (β) and sparsity $(1-r)$ for the one-sample problem (1.1) at 5% nominal significance for the Gaussian data and long range dependence covariance matrices in Model $2^{(\mathrm{I})}$. For simulations when $n = 80$ and $p = 120, 360$.

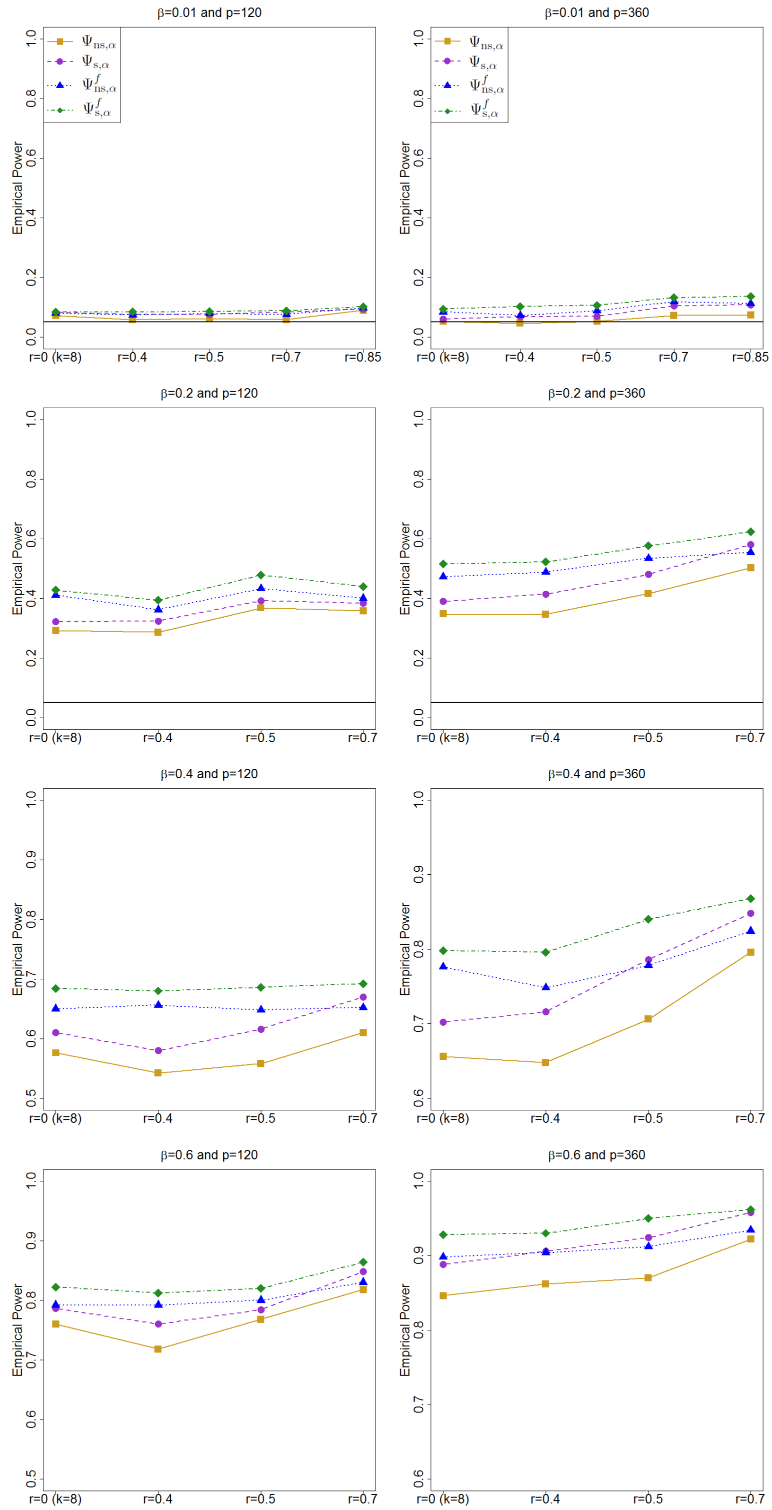


Figure S3: Empirical powers of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^f_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^f_{\mathrm{s},\alpha}$) against alternatives with different levels of the signal strength (β) and sparsity $(1-r)$ for the one-sample problem (1.1) at 5% nominal significance for the autoregressive process model, Model $3^{(\mathrm{I})}$, with t-distributed innovations when $n = 80$ and $p = 120, 360$.

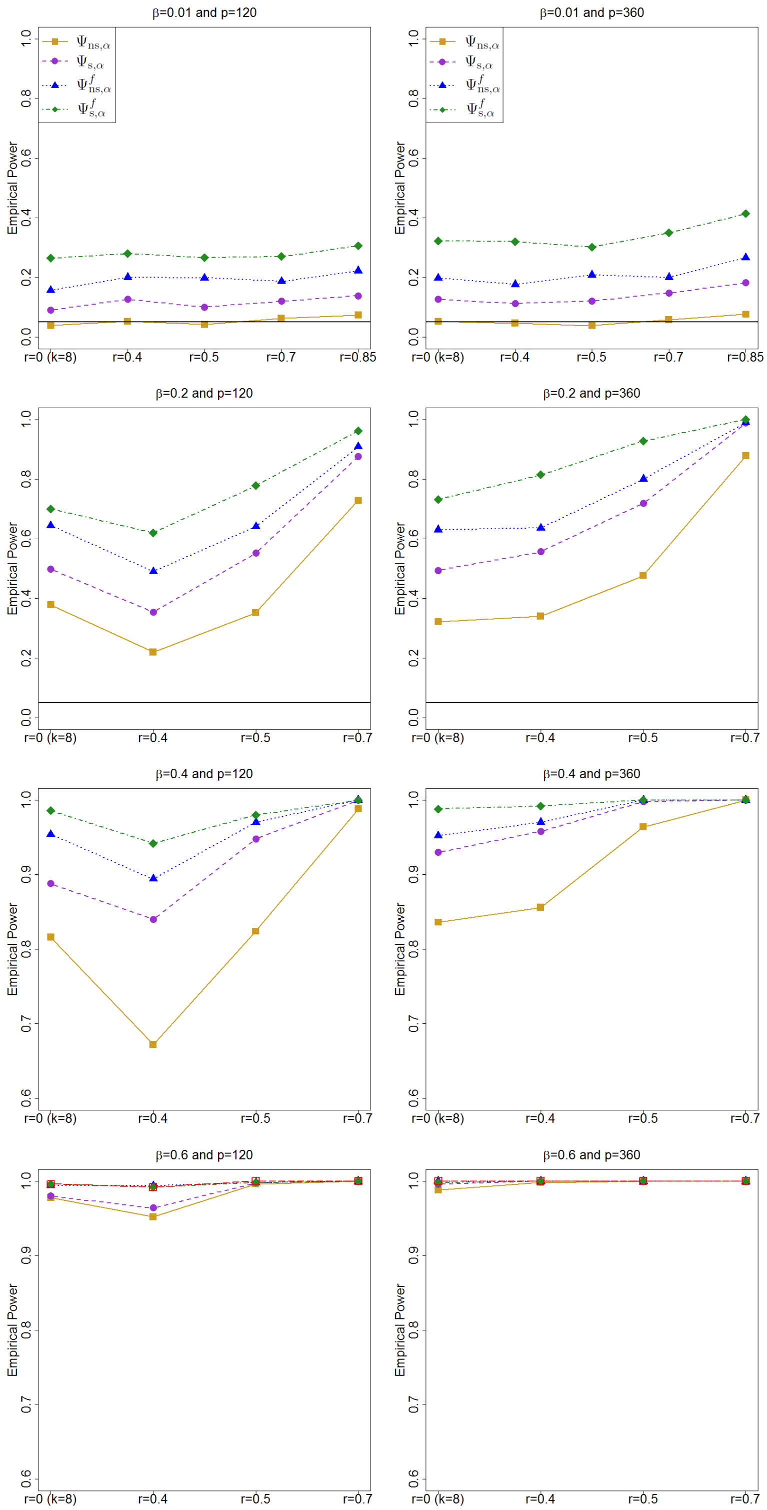


Figure S4: Empirical powers of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized with screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^{f}_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^{f}_{\mathrm{s},\alpha}$) against alternatives with different levels of the signal strength (β) and sparsity $(1-r)$ for the one-sample problem (1.1), along with the power of the test by Zhong, Chen and Xu (2013) (ZCX) at 5% nominal significance for the Gaussian data and bandable covariance matrices in Model $4^{(\mathrm{I})}$ when $n = 80$ and $p = 120, 360$.

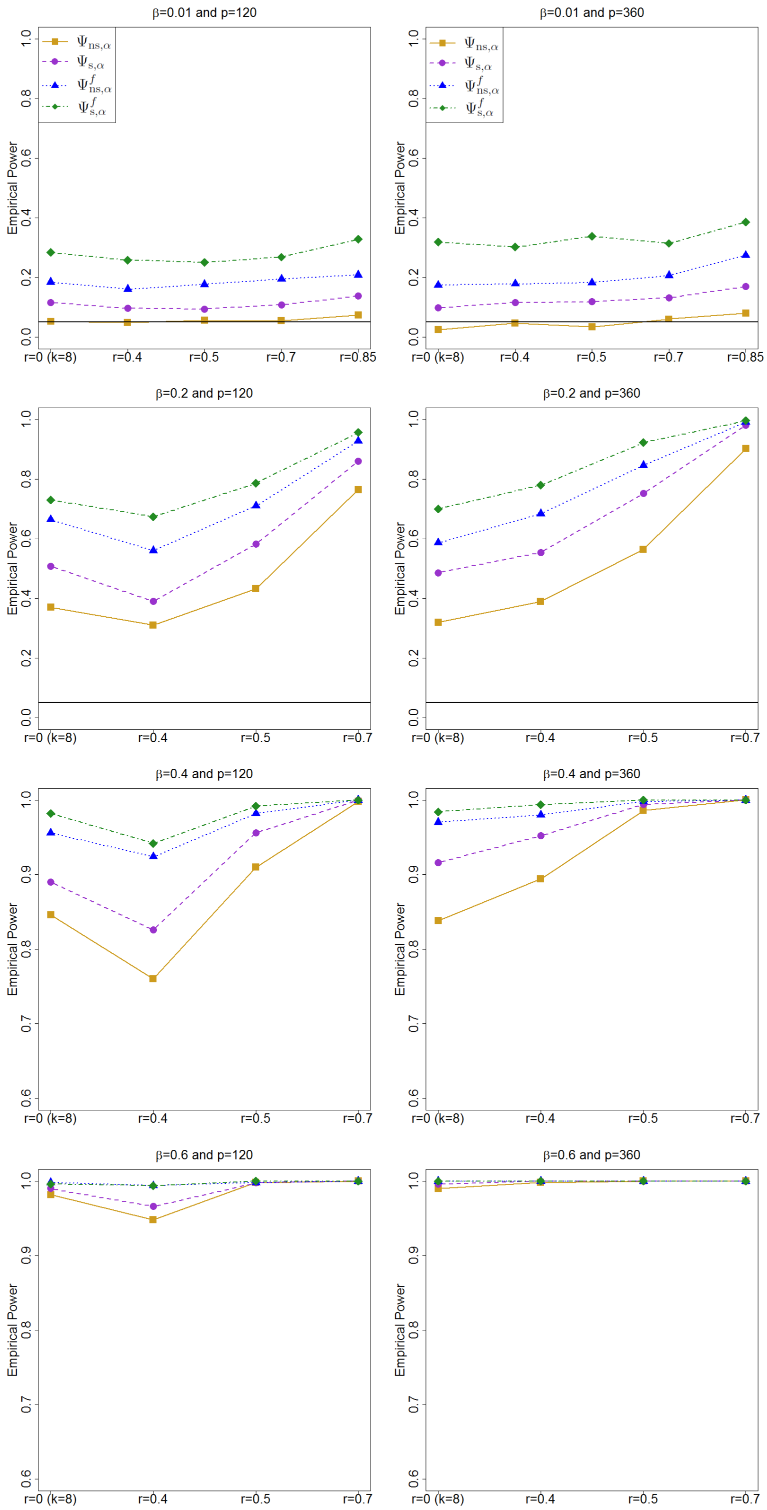
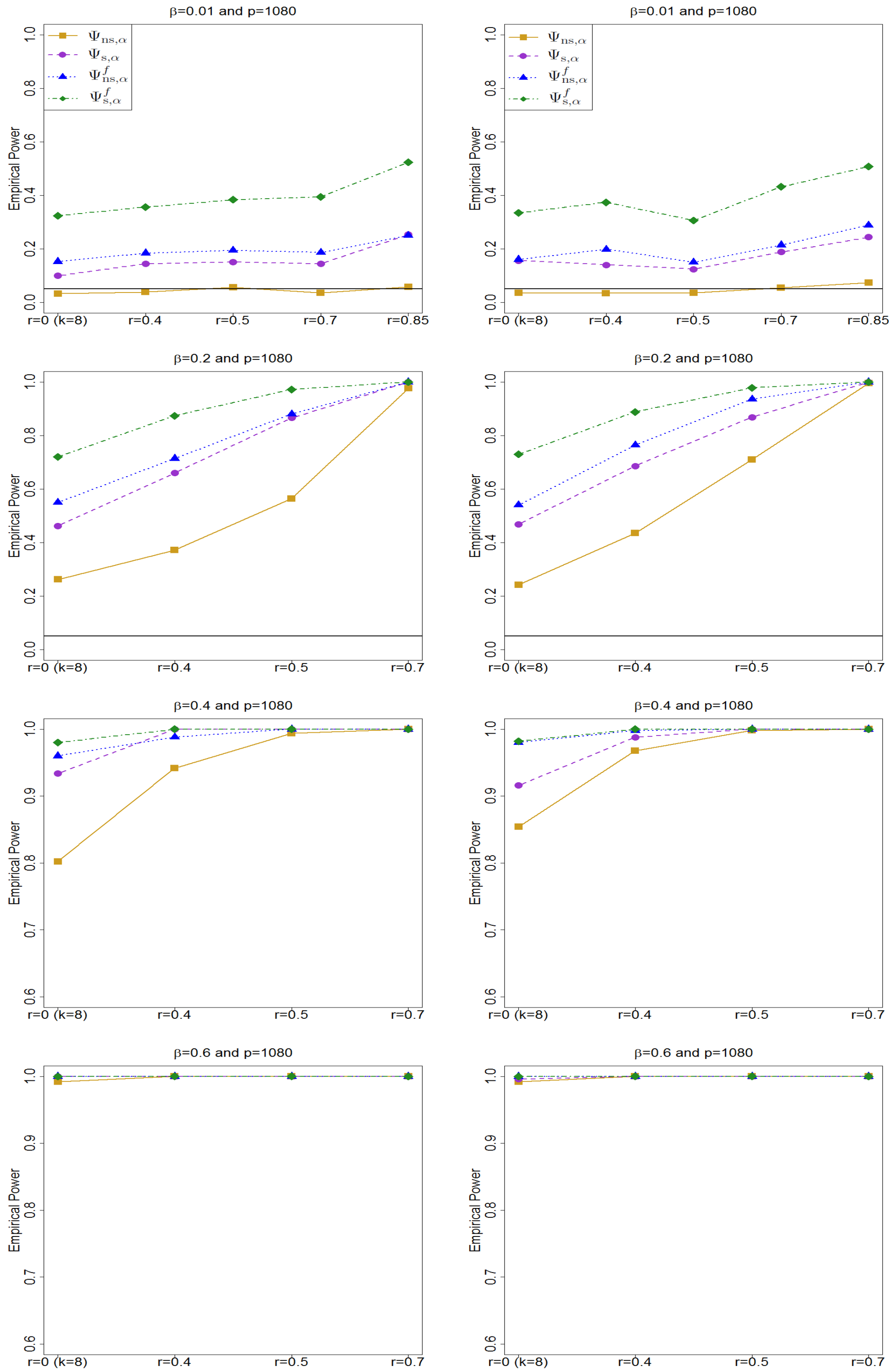


Figure S5: Empirical powers of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^f_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^f_{\mathrm{s},\alpha}$) against alternatives with different levels of the signal strength (β) and sparsity $(1-r)$ for the one-sample problem (1.1) at 5% nominal significance for the moving average process with Beta distributed innovations in Model $5^{(\mathrm{I})}$ when $n = 80$ and $p = 120, 360$.

f02



(a) Model $2^{(\mathrm{I})}$ (b) Model$5^{(\mathrm{I})}$

Figure S6: Empirical powers of the proposed tests (non-studentized without screening $\Psi_{\text{ns},\alpha}$, studentized without screening $\Psi_{\text{s},\alpha}$, non-studentized with screening $\Psi^{f}_{\text{ns},\alpha}$, and studenzied with screening $\Psi^{f}_{\text{s},\alpha}$) against alternatives with different levels of the signal strength (β) and sparsity $(1-r)$ for the one-sample problem (1.1) when $n = 80$ and $p = 1080$ at 5% nominal significance for the Gaussian data and block diagonal covariance in Model $4^{(\mathrm{I})}$(column (a)), and the moving average model with Beta distributed innovations in Model $5^{(\mathrm{I})}$ (column (b)).

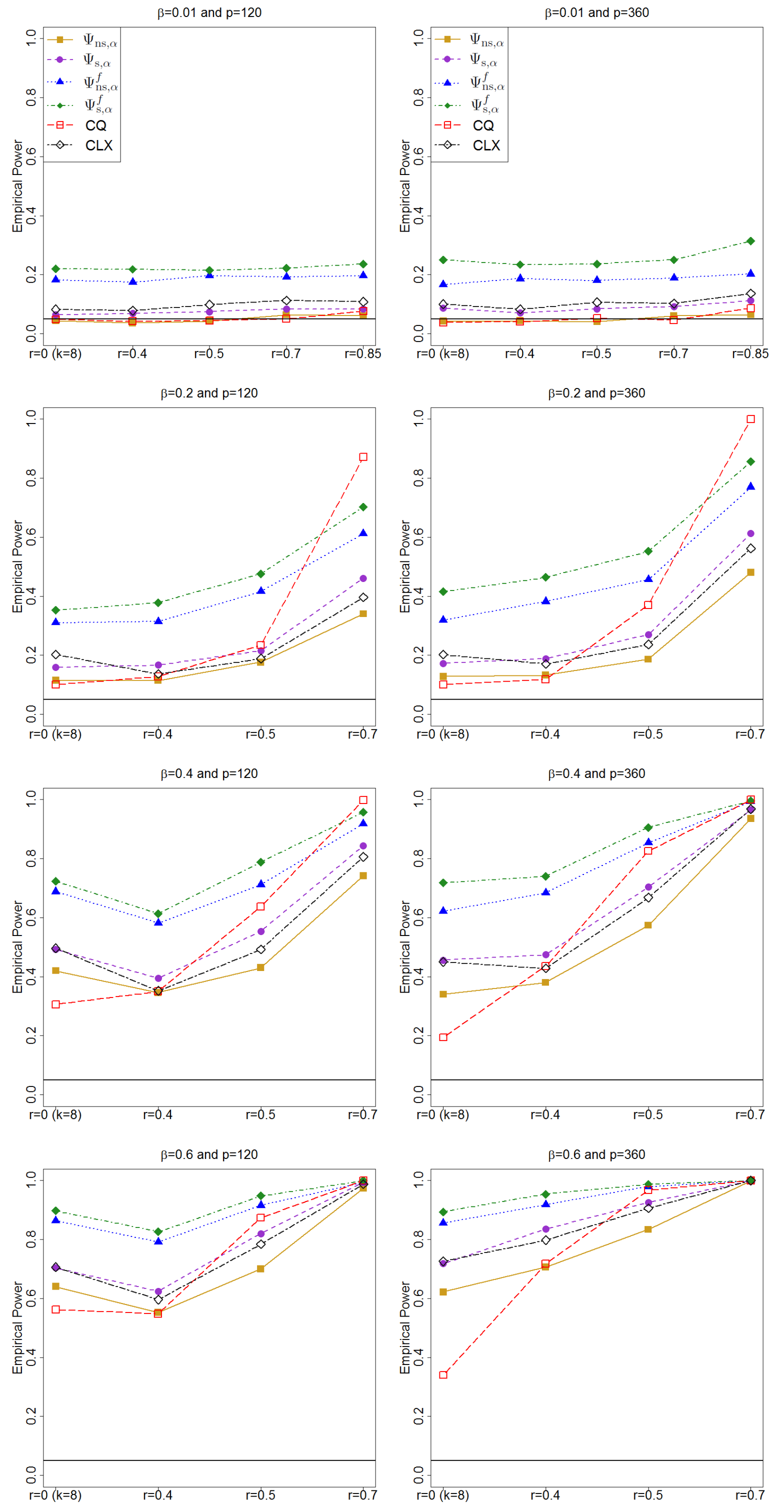


Figure S7: Empirical powers of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^{f}_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^{f}_{\mathrm{s},\alpha}$) against the alternatives with different levels of the signal strength (β) and sparsity $(1-r)$ for the two-sample problem (1.2), along of those of the tests by Chen and Qin (2010) (CQ) and Cai, Liu and Xia (2014) (CLX) at 5% nominal significance for the Gaussian data and block diagonal covariances in Model $1^{(\mathrm{II})}$ when $n=m=80$ and $p=120,360$.

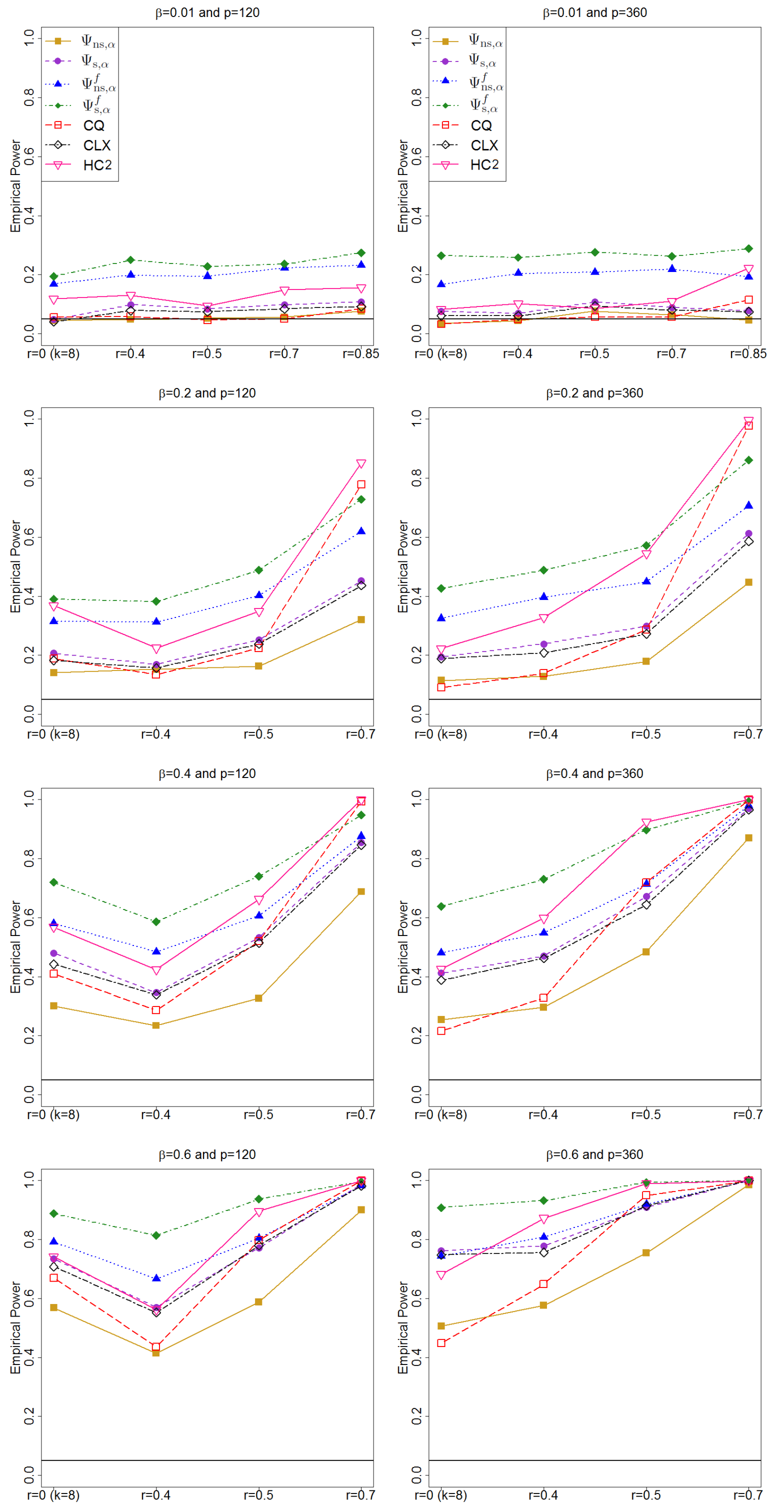


Figure S8: Empirical powers of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized with screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^f_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^f_{\mathrm{s},\alpha}$) against alternatives with different levels of the signal strength (β) and sparsity ($1-r$) for the two-sample problem (1.2), along of those of the tests by Chen and Qin (2010) (CQ), Cai, Liu and Xia (2014) (CLX) and Delaigle, Hall and Jin (2011) (HC2) at 5% nominal significance for the Gaussian data and non-sparse covariances in Model $2^{(\mathrm{II})}$ when $n = m = 80$ and $p = 120, 360$.

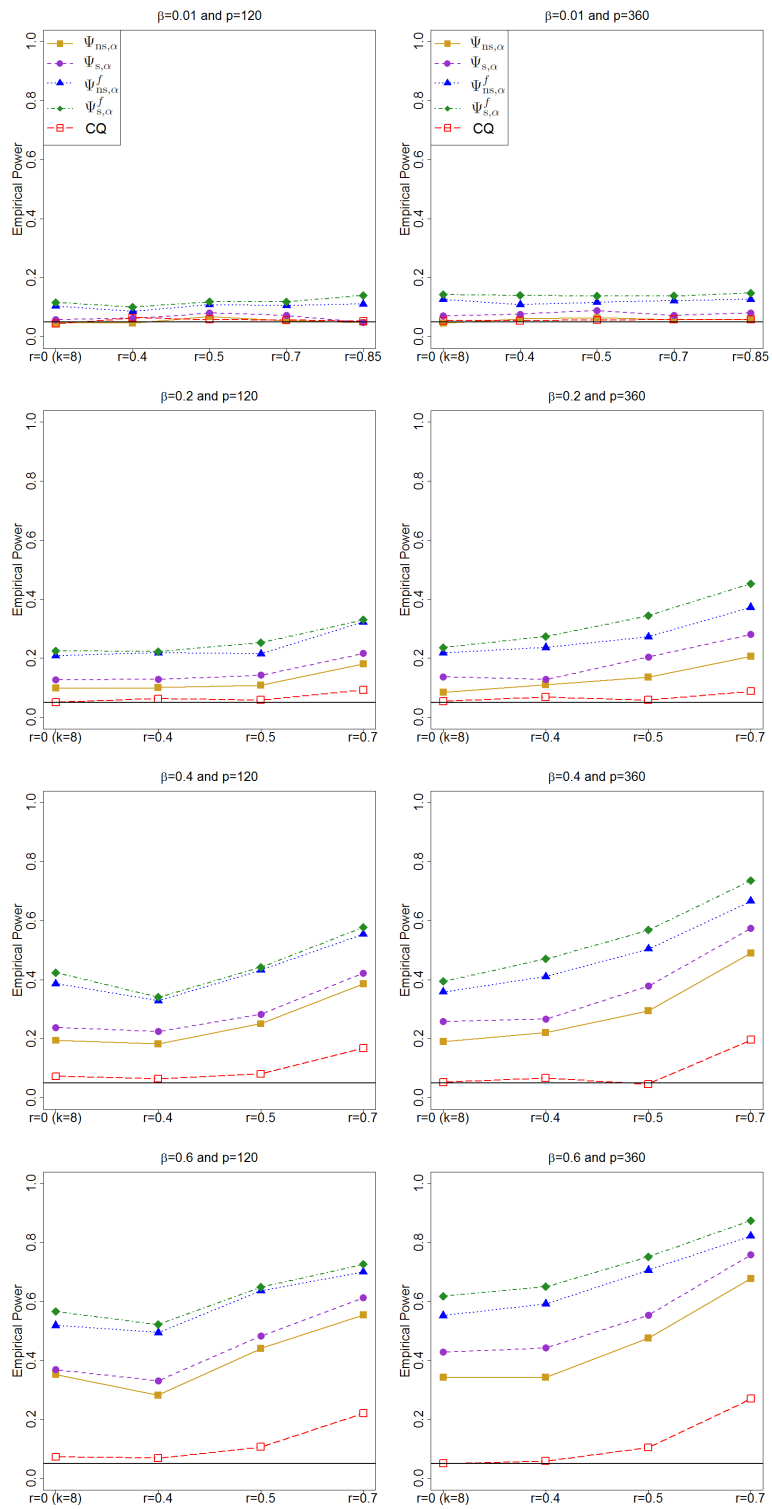


Figure S9: Empirical powers of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^{f}_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^{f}_{\mathrm{s},\alpha}$) against alternatives with different levels of the signal strength (β) and sparsity ($1-r$) for the two-sample problem (1.2), along of those of the test by Chen and Qin (2010) (CQ) at 5% nominal significance for the autoregressive process model, Model $3^{(\mathrm{II})}$, with t-distributed innovations when $n = m = 80$ and $p = 120, 360$.

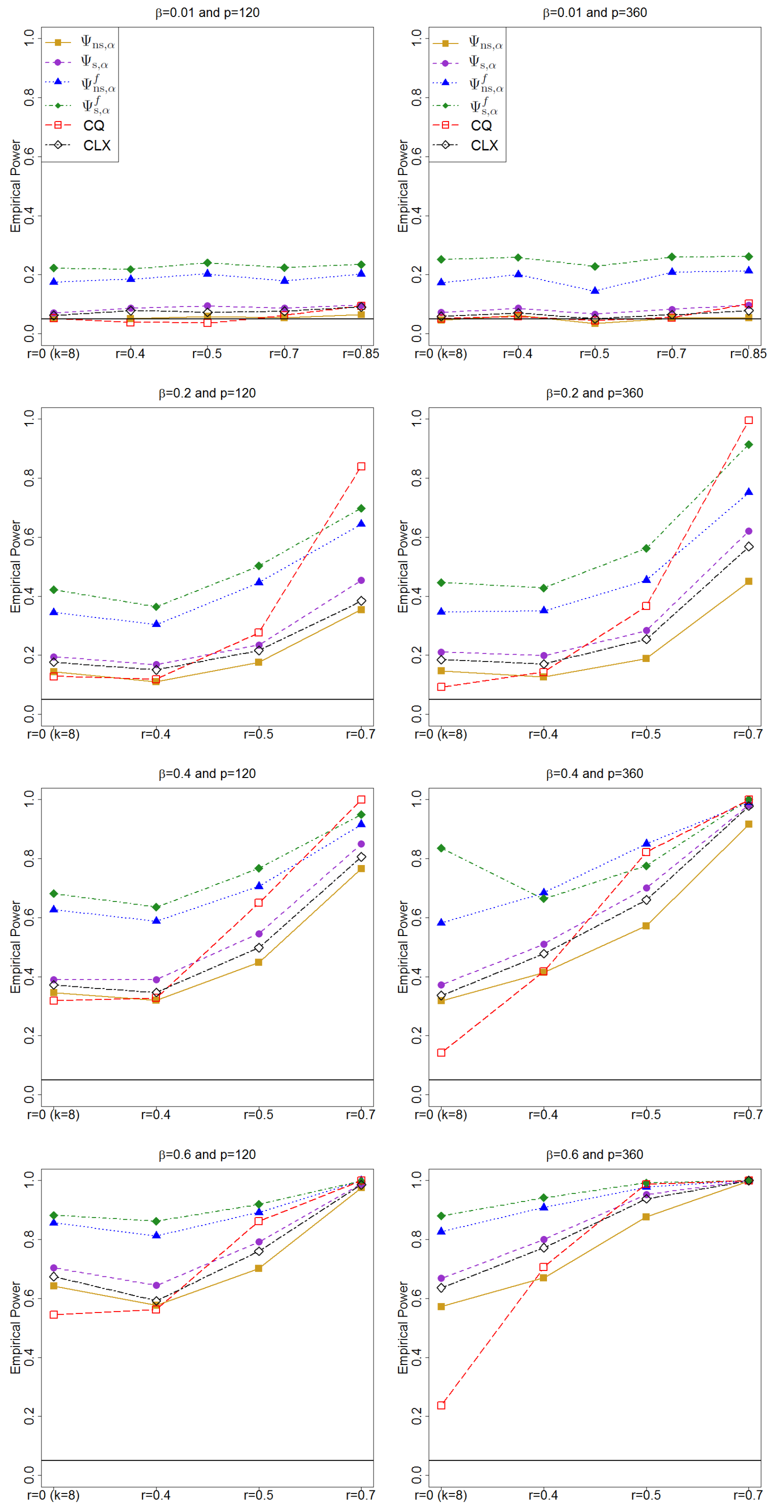


Figure S10: Empirical powers of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^{f}_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^{f}_{\mathrm{s},\alpha}$) against alternatives with different levels of the signal strength (β) and sparsity ($1-r$) for the two-sample problem (1.2), along of those of the tests by Chen and Qin (2010) (CQ) and Cai, Liu and Xia (2014) (CLX) at 5% nominal significance for the Gaussian data and long range dependence covariances in Model $4^{(\mathrm{II})}$ when $n = m = 80$ and $p = 120, 360$.

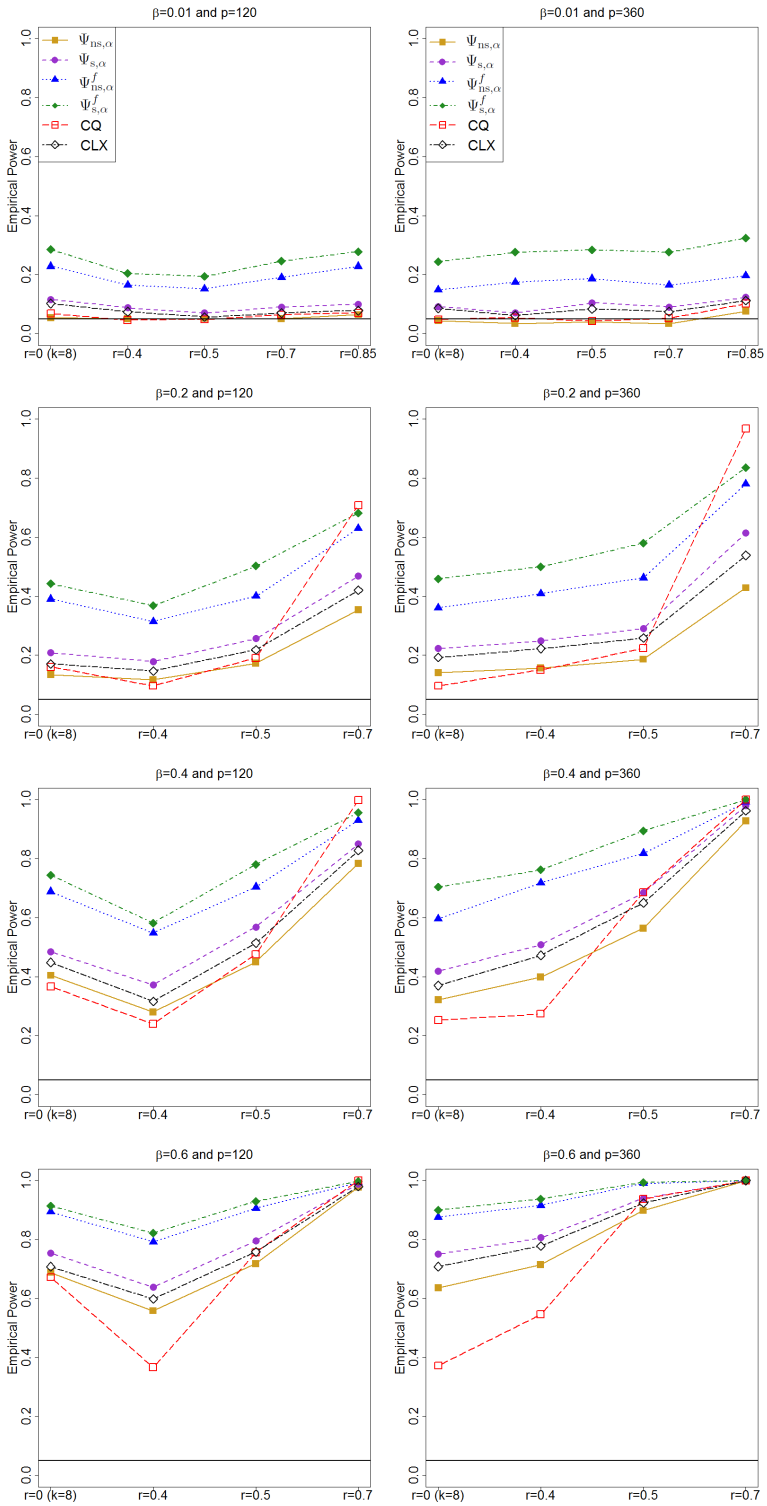
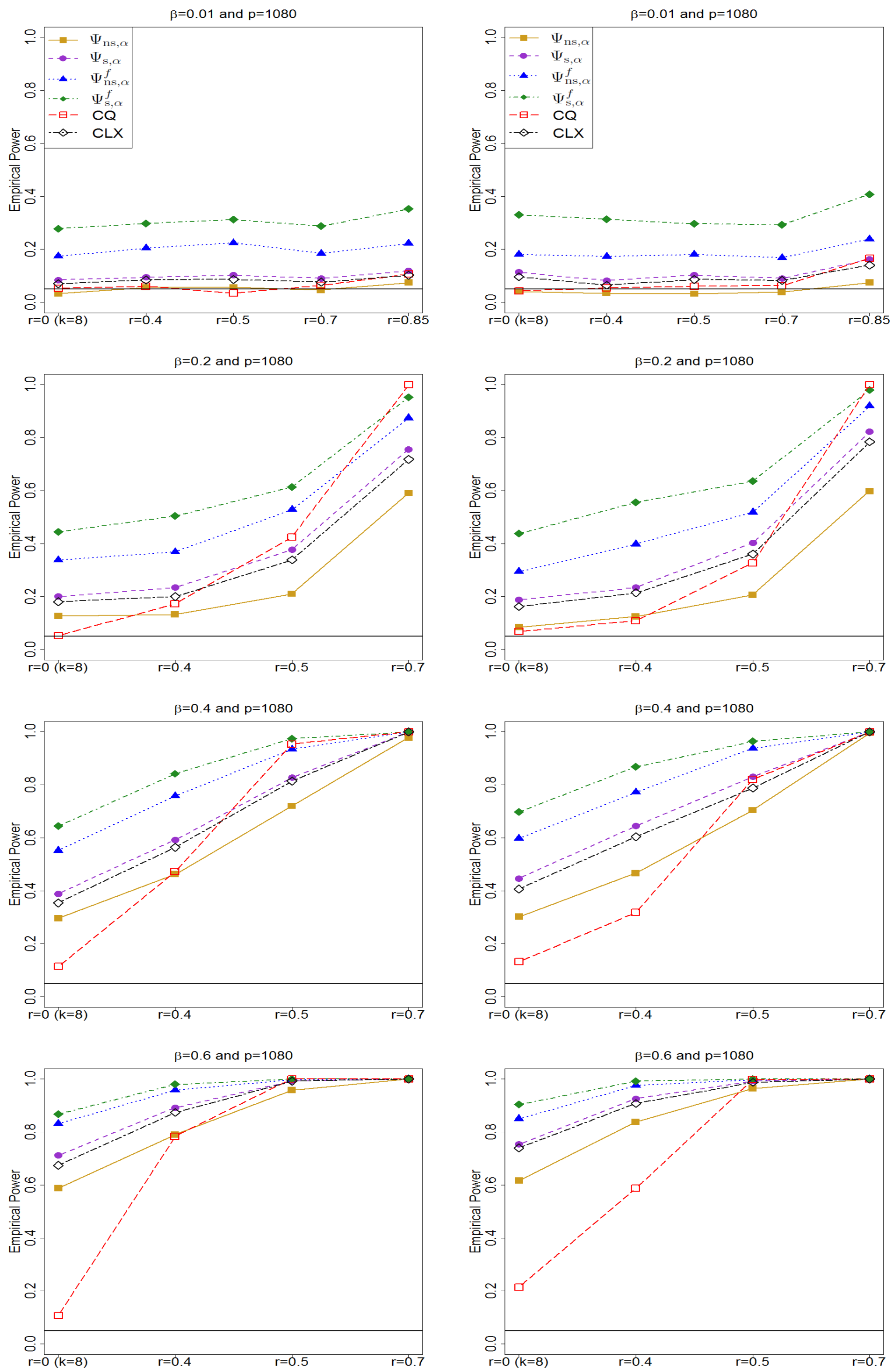


Figure S11: Empirical powers of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^f_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^f_{\mathrm{s},\alpha}$) against alternatives with different levels of the signal strength (β) and sparsity ($1-r$) for the two-sample problem (1.2), along of those of the tests by Chen and Qin (2010) (CQ) and Cai, Liu and Xia (2014) (CLX) at 5% nominal significance for the moving average process with Gamma distributed innovations in Model $5^{(\mathrm{II})}$ when $n = m = 80$ and $p = 120, 360$.



(a) Model $3^{(\mathrm{II})}$ (b) Model $5^{(\mathrm{II})}$

Figure S12: Empirical powers of the proposed tests (non-studentized without screening $\Psi_{\mathrm{ns},\alpha}$, studentized without screening $\Psi_{\mathrm{s},\alpha}$, non-studentized with screening $\Psi^{f}_{\mathrm{ns},\alpha}$, and studenzied with screening $\Psi^{f}_{\mathrm{s},\alpha}$) against alternatives with different levels of the signal strength (β) and sparsity ($1-r$) for the two-sample problem (1.2) when $n = 80$ and $p = 1080$ at 5% nominal significance for the Gaussian data and long range dependence covariance matrices in Model $4^{(\mathrm{II})}$ (column (a)), and the moving average processes model with Gamma distributed innovations in Model $5^{(\mathrm{II})}$ (column (b)).