# A Robust and Nonparametric Two-Sample Test in High Dimensions

Tao Qiu, Wangli Xu and Liping Zhu

*Renmin University of China and ZheJiang Gongshang University*

## Supplementary Material

This supplement contains the proofs of (2.11)-(2.13), the proofs of Theorem 1-Theorem 3 and additional simulation results.

## S1 Appendix A: Proof of 2.14

Recall the definition of $E\{\varphi_{11}^2(\mathbf{x}_1,\mathbf{x}_2)\}$ in (2.4). We have,

$$\begin{aligned} E\{\varphi_{11}^2(\mathbf{x}_1,\mathbf{x}_2)\} &= \sum_{k,l=1}^{p} E\big[\{I(X_{1k}\le Z_{1k})-F_k(Z_{1k})\}\{I(X_{2k}\le Z_{1k})-F_k(Z_{1k})\} \\ &\quad \{I(X_{1l}\le Z_{2l})-F_l(Z_{2l})\}\{I(X_{2l}\le Z_{2l})-F_l(Z_{2l})\}\big]. \end{aligned}$$

Hence, simple calculation yields

$$E\{\varphi_{11}^2(\mathbf{x}_1,\mathbf{x}_2)\} = \sum_{k,l=1}^{p} E\{h^2(Z_{1k},Z_{2l})\} = \sum_{k,l=1}^{p} H_{k,l}H_{l,k} = \mathrm{tr}(\mathbf{H}^2).$$

By the definition of $h(Z_{1k},Z_{2l})$ and $H_{k,l}$ in (2.10), $0 \le H_{k,l} \le 1$. In particular, $H_{k,l}=0$ implies that $X_{1k}$ and $X_{1l}$ are uncorrelated. Following similar arguments, we have

$$E\{\omega_{11}^2(\mathbf{x}_1, \mathbf{x}_2, \mathbf{z})\} = \sum_{k,l=1}^{p} E\{h^2(Z_{1k}, Z_{1l})\} = O\{\mathrm{tr}(\mathbf{H}^2)\}.$$

Therefore, (2.11) follows immediately. □

## S2 Appendix B: Proof of 2.15

We consider (2.12) under the two cases for $\mathbf{x} = (X_1, \ldots, X_p)$ : (i) correlated structure; (ii) banded dependence structure;

*Correlated structure:* In this case, we suppose $X_k$ and $X_l$ are correlated for $1 \leq k \neq l \leq p$. This means $0 \leq H_{k,l} \leq 1$. Recall that $h(Z_{1k}, Z_{2l}) \stackrel{\text{def}}{=} \mathrm{cov}\{I(X_{1k} \leq Z_{1k}), I(X_{1l} \leq Z_{2l}) \mid Z_{1k}, Z_{2l}\}$. By Hölder inequality and algebraic calculation, we have

$$\begin{aligned} E\{\varphi_{11}^4(\mathbf{x}_1, \mathbf{x}_2)\} &\leq \sum_{k,l,s,t=1}^{p} E\{h^2(Z_{1k}, Z_{2l})h^2(Z_{3s}, Z_{4t})\} \\ &\leq \sum_{k,l,s,t=1}^{p} H_{kl}H_{lk}H_{st}H_{ts}. \end{aligned}$$

Thus, $E\{\varphi_{11}^4(\mathbf{x}_1, \mathbf{x}_2)\} = O\{\mathrm{tr}^2(\mathbf{H}^2)\}$.

*Banded dependence structure:* We first introduce the definition of banded dependence structure for $\mathbf{x} = (X_1, \ldots, X_p)$ which was given by Zhang et al. (2018). Suppose for an unknown permutation $\pi : \{1, 2, ..., p\} \rightarrow$

$\{1, 2, ..., p\}$ :

$$X_{\pi(i)} = f_{\pi(i)}(\epsilon_i, \epsilon_{i+1}, \cdots, \epsilon_{i+L}),$$

for $1 \le i \le p$ and $L \ge 0$, where $\{\epsilon_i\}$ are independent random variables, and $f_{\pi(i)}$ is measurable function such that $X_{\pi(i)}$ is well defined. Apparently, $X_k$ and $X_l$ are independent if $|k-l| > L$. Hence,

$$\mathrm{tr}(\mathbf{H}^2) = \sum_{|k-l|\le L} E\{g(X_{1k}, X_{2k})g(X_{1l}, X_{2l})\}.$$

Define $g(X_{1k}, X_{2k}) \stackrel{\text{def}}{=} E\big\{U_k(X_{1k}, Z_{1k})U_k(X_{2k}, Z_{1k}) \mid X_{1k}, X_{2k}\big\}$. By simple algebraic calculation, we have

$$E\{\varphi_{11}^4(\mathbf{x}_1, \mathbf{x}_2)\} = \sum_{k,l,s,t=1}^{p} E\{g(X_{1k}, X_{2k})g(X_{1l}, X_{2l})g(X_{1s}, X_{2s})g(X_{1t}, X_{2t})\}.$$

By Hölder inequality, $E\big[E^2\{\varphi_{11}^2\ (\mathbf{x}_1, \mathbf{x}_2) \mid \mathbf{x}_1\}\big] \le E\{\varphi_{11}^4(\mathbf{x}_1, \mathbf{x}_2)\}$ and

$$\begin{aligned} E\{\varphi_{11}^4(\mathbf{x}_1, \mathbf{x}_2)\} &\le c_1 \sum_{k=1}^{p} \sum_{l,s,t=k-3L}^{k+3L} E\{g(X_{1k}, X_{2k})g(X_{1l}, X_{2l})g(X_{1s}, X_{2s})g(X_{1t}, X_{2t})\} \\ &+ c_2 \left[\sum_{|k-l|\le L} E\{g(X_{1k}, X_{2k})g(X_{1l}, X_{2l})\}\right]^2 \end{aligned}$$

for some positive constant $c_1$ and $c_2$. Therefore,

$$E\{\varphi_{11}^4(\mathbf{x}_1, \mathbf{x}_2)\} \le \max(c_1, c_2)\left[p(6L+1)^3 \max_{1\le k\le p} E\{g^4(X_{1k}, X_{2k})\} + \mathrm{tr}^2(\mathbf{H}^2)\right].$$

As a consequence, (2.12) is implied by the fact that

$$\left[p(L+1)^3 \max_{1\le k\le p} E\{g^4(X_{1k}, X_{2k})\}\right] = O\{\mathrm{tr}^2(\mathbf{H}^2)\}.$$

In particular, if $L = o(p^{1/3})$, (2.12) is satisfied. □

## S3 Appendix C: Proof of 2.16

Recall that $h(Z_{1k}, Z_{2l}) \stackrel{\text{def}}{=} \text{cov}\{I(X_{1k} \leq Z_{1k}), I(X_{1l} \leq Z_{2l}) \mid Z_{1k}, Z_{2l}\}$. It follows that $E[E^2\{\varphi_{11}(\mathbf{x}_1, \mathbf{x}_2)\varphi_{11}(\mathbf{x}_1, \mathbf{x}_3) \mid (\mathbf{x}_2, \mathbf{x}_3)\}] = \sum_{k,l,s,t=1}^{p} E\big\{h(Z_{1k}, Z_{2l}) h(Z_{2l}, Z_{4t})h(Z_{4t}, Z_{3s})h(Z_{3s}, Z_{1k})\big\}$. We consider (2.13) under the two cases for $\mathbf{x} = (X_1, \ldots, X_p)$ : (i) correlated structure; (ii) banded dependence, which are defined in Appendix B.

*Correlated structure:* By simple algebraic calculation, we have

$$
\begin{aligned}
& E[E^2\{\varphi_{11}(\mathbf{x}_1, \mathbf{x}_2)\varphi_{11}(\mathbf{x}_1, \mathbf{x}_3) \mid (\mathbf{x}_2, \mathbf{x}_3)\}] \\
\leq & \sum_{k,l,s,t=1}^{p} [E\{h^2(Z_{1k}, Z_{2l})h^2(Z_{4t}, Z_{3s})\}E\{h^2(Z_{2l}, Z_{4t})h^2(Z_{3s}, Z_{1k})\}]^{1/2} \\
\leq & \sum_{k,l,s,t=1}^{p} H_{kl}H_{lt}H_{ts}H_{sk}.
\end{aligned}
$$

Thus, $E[E^2\{\varphi_{11}(\mathbf{x}_1, \mathbf{x}_2)\varphi_{11}(\mathbf{x}_1, \mathbf{x}_3) \mid (\mathbf{x}_2, \mathbf{x}_3)\}] = O\{\text{tr}(\mathbf{H}^4)\}$.

*Banded dependence structure:* Direct calculation shows that

$$
\begin{aligned}
& E[E^2\{\varphi_{11}(\mathbf{x}_1, \mathbf{x}_2)\varphi_{11}(\mathbf{x}_1, \mathbf{x}_3) \mid (\mathbf{x}_2, \mathbf{x}_3)\}] \\
= & \sum_{k=1}^{p} \sum_{l=k-L}^{k+L} \sum_{t=l-L}^{l+L} \sum_{s=t-L}^{t+L} E\big\{h(Z_{1k}, Z_{2l})h(Z_{2l}, Z_{4t})h(Z_{4t}, Z_{3s})h(Z_{3s}, Z_{1k})\big\} \\
\leq & c_3\{p(L+1)^3\}
\end{aligned}
$$

for some positive constant $c_3$. Therefore, (2.13) is implied by $\{p(L+1)^3\} = o\{\text{tr}^2(\mathbf{H}^2)\}$. In particular, if $L = o(p^{1/3})$, (2.13) is satisfied. □

## S4 Appendix D: Proof of Theorem 1

Recall the definition of $\widehat{Q}$ in (2.1). It can be verified that $\widehat{Q} = \widehat{T}_1 + \widehat{T}_2 + D_0$, where $\widehat{T}_1, \widehat{T}_2$, and $D_0$ are presented in (2.2), (2.6), and (2.7), respectively. Under the null hypothesis, it can be verified that $\widehat{T}_2 = D_0 = 0$. For $\widehat{T}_1$, it is of a complicated form in that it is a $U$-statistic estimate of three random samples. To simplify subsequent illustrations, we define

$$\begin{aligned} \widehat{T}_{11} &\overset{\text{def}}{=} \{m(m-1)(m+n)\}^{-1} \sum_{i \neq j}^{m} \sum_{r=1}^{m+n} \omega_{11}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r), \\ \widehat{T}_{12} &\overset{\text{def}}{=} \{n(n-1)(m+n)\}^{-1} \sum_{i \neq j}^{n} \sum_{r=1}^{m+n} \omega_{12}(\mathbf{y}_i, \mathbf{y}_j, \mathbf{z}_r), \qquad \text{(S4.1)} \\ \text{and } \widehat{T}_{13} &\overset{\text{def}}{=} -2\{mn(m+n)\}^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{m+n} \omega_{13}(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r). \end{aligned}$$

It follows that $\widehat{T}_1 = \widehat{T}_{11} + \widehat{T}_{12} + \widehat{T}_{13}$. Next, we consider the Hoeffding decomposition for $\widehat{T}_{1k}s$. According to the definition of $\widehat{T}_{11}$ in (S4.1), we know $\widehat{T}_{11}$ is a two sample U-statistic with degrees (2,1) and kernel $\omega_{11}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r)$. Direct calculation shows that

$$\begin{aligned} E\{\omega_{11}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r) \mid \mathbf{x}_i\} &= E\{\omega_{11}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r) \mid \mathbf{z}_r\} = 0, \\ E\{\omega_{11}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r) \mid (\mathbf{x}_i, \mathbf{z}_r)\} &= 0, \text{ and } E\{\omega_{11}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r) \mid (\mathbf{x}_i, \mathbf{x}_j)\} = \varphi_{11}(\mathbf{x}_i, \mathbf{x}_j), \end{aligned}$$

where $\varphi_{11}(\mathbf{x}_i, \mathbf{x}_j)$ is defined in (2.4). Furthermore, we apply H-decomposition (Lee (1990), section 1.6) to $\widehat{T}_{11}$ :

$$\begin{aligned}\widehat{T}_{11} &= \{m(m-1)\}^{-1} \sum_{i \neq j}^{m} \varphi_{11}(\mathbf{x}_i, \mathbf{x}_j) \\ &\quad + \{m(m-1)(m+n)\}^{-1} \sum_{i \neq j}^{m} \sum_{r=1}^{m+n} \big\{\omega_{11}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r) - \varphi_{11}(\mathbf{x}_i, \mathbf{x}_j)\big\}.\end{aligned}$$

When $p \to \infty$, unlike the fixed dimension cases, $\mathrm{var}\{\varphi_{11}(\mathbf{x}_i, \mathbf{x}_j)\}$ and $\mathrm{var}\{\omega_{11}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r)\}$ may no longer be bounded and can diverge. For high-dimensional data, under Condition (C2),

$$\{m(m-1)(m+n)\}^{-1} \sum_{i \neq j}^{m} \sum_{r=1}^{m+n} \big\{\omega_{11}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r) - \varphi_{11}(\mathbf{x}_i, \mathbf{x}_j)\big\}$$

is the remainder term which is asymptotically negligible.

For $\widehat{T}_{12}$ in (S4.1), similar to the deviation for $\widehat{T}_{11}$, we have, under Condition (C2),

$$\begin{aligned}\widehat{T}_{12} &= \{n(n-1)\}^{-1} \sum_{i \neq j}^{n} \varphi_{12}(\mathbf{y}_i, \mathbf{y}_j) \\ &\quad + \{n(n-1)(m+n)\}^{-1} \sum_{i \neq j}^{n} \sum_{r=1}^{m+n} \big\{\omega_{11}(\mathbf{y}_i, \mathbf{y}_j, \mathbf{z}_r) - \varphi_{12}(\mathbf{y}_i, \mathbf{y}_j)\big\}.\end{aligned}$$

Here, $\{n(n-1)(m+n)\}^{-1} \sum_{i \neq j}^{n} \sum_{r=1}^{m+n} \big\{\omega_{11}(\mathbf{y}_i, \mathbf{y}_j, \mathbf{z}_r) - \varphi_{12}(\mathbf{y}_i, \mathbf{y}_j)\big\}$ is the remainder term which is asymptotically negligible.

Next, we discuss the decomposition of $\widehat{T}_{13}$ in (S4.1). $\widehat{T}_{13}$ is a three sample U-statistic with degrees (1,1,1) and kernel $\omega_{13}(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r)$. Direct calculation shows that

$$E\{\omega_{13}(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r) \mid \mathbf{x}_i\} = E\{\omega_{13}(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r) \mid \mathbf{y}_j\} = E\{\omega_{13}(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r) \mid \mathbf{z}_r\} = 0,$$

$$E\{\omega_{13}(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r) \mid (\mathbf{x}_i, \mathbf{z}_r)\} = E\{\omega_{13}(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r) \mid (\mathbf{y}_j, \mathbf{z}_r)\} = 0, \text{and}$$

$$E\{\omega_{13}(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r) \mid (\mathbf{x}_i, \mathbf{x}_j)\} = \varphi_{13}(\mathbf{x}_i, \mathbf{y}_j).$$

Furthermore, we can apply H-decomposition to $\widehat{T}_{13}$ and obtain

$$\begin{aligned}\widehat{T}_{13} &= -2(mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \varphi_{13}(\mathbf{x}_i, \mathbf{y}_j) \\ &\quad -2\{mn(m+n)\}^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{m+n} \big\{\omega_{13}(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r) - \varphi_{13}(\mathbf{x}_i, \mathbf{y}_j)\big\}.\end{aligned}$$

Using similar arguments to deal with the remainder term of $\widehat{T}_{11}$, we can show that, under Condition (C2),

$$-2\{mn(m+n)\}^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{m+n} \big\{\omega_{13}(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r) - \varphi_{13}(\mathbf{x}_i, \mathbf{y}_j)\big\}$$

is asymptotically negligible.

Combining the above results for $\widehat{T}_{11}, \widehat{T}_{13}$ and, $\widehat{T}_{13}$, we have,

$$\begin{aligned}\widehat{T}_1 &= \Big\{\{m(m-1)\}^{-1} \sum_{i \neq j}^{m} \varphi_{11}(\mathbf{x}_i, \mathbf{x}_j) + \{n(n-1)\}^{-1} \sum_{i \neq j}^{n} \varphi_{12}(\mathbf{y}_i, \mathbf{y}_j) \\ &\quad -\{mn\}^{-1} 2 \sum_{i=1}^{m} \sum_{j=1}^{n} \varphi_{13}(\mathbf{x}_i, \mathbf{y}_j)\Big\}\{1 + o_p(1)\} \\ &\stackrel{\text{def}}{=} \widehat{T}_{1,1}\left\{1 + o_p(1)\right\}.\end{aligned}$$

It thus follows that

$$\begin{aligned}\mathrm{var}(\widehat{T}_{1,1}) &= 2\{m(m-1)\}^{-1}E\{\varphi_{11}^2(\mathbf{x_1},\mathbf{x_2})\}+2\{n(n-1)\}^{-1}E\{\varphi_{12}^2(\mathbf{y_1},\mathbf{y_2})\}\\ &\quad +4(mn)^{-1}E\{\varphi_{13}^2(\mathbf{x_1},\mathbf{y_1})\}.\end{aligned}$$

Therefore, under Condition (C2),

$$\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-1/2}\widehat{Q}=\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-1/2}\widehat{T}_{1,1}+o_p(1).$$

Below, we establish the asymptotic normality for $\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-1/2}\widehat{Q}$ using the central limit theorem for martingale difference sequences. We first consider the limit behavior under the null hypothesis. Let $\mathbf{w}_i=\mathbf{x}_i$ for $i=1,\cdots,m$ and $\mathbf{w}_{j+m}=\mathbf{y}_j$ for $j=1,\cdots,n$. We further define

$$\phi_{ij}=\begin{cases}\{m(m-1)\}^{-1}\varphi_{11}(\mathbf{w}_i,\mathbf{w}_j) & \text{if } i,j\in\{1,2,\cdots,m\},\\ (-mn)^{-1}\varphi_{13}(\mathbf{w}_i,\mathbf{w}_j) & \text{if } i\in\{1,2,\cdots,m\}\\ & \text{and } j\in\{m+1,m+2,\cdots,m+n\},\\ \{n(n-1)\}^{-1}\varphi_{12}(\mathbf{w}_i,\mathbf{w}_j) & \text{if } i,j\in\{m+1,m+2,\cdots,m+n\}.\end{cases} \tag{S4.2}$$

for $i\neq j$. Denote $W_j=\sum_{i=1}^{j-1}\phi_{ij}$ for $j=2,\ 3,\ \cdots,\ m+n$, $\mathcal{S}_r=\sum_{j=2}^{r}W_j$ and $\mathcal{F}_r=\sigma\{\mathbf{w}_1,\mathbf{w}_2,\cdots\mathbf{w}_r\}$. Here, $\mathcal{F}_r$ is the $\sigma$ algebra generated by $\{\mathbf{w}_1,\mathbf{w}_2,\cdots\mathbf{w}_r\}$. We have $\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-1/2}\widehat{Q}=2\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-1/2}\sum_{j=2}^{m+n}W_j+o_p(1)$. Thus, by verifying the following Lemma 1-3, we can establish the asymptotic normality for $\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-1/2}\widehat{Q}$ by Corollary 3.1 of Hall and Heyde (1980).

Lemma 1. For each $m+n, \{\mathcal{S}_r, \mathcal{F}_r\}_{r=1}^{m+n}$ is the sequence of zero mean and a square integrable martingale .

*Proof of Lemma 1.* Following similar arguments to prove Lemma 1 in Chen and Qin (2010), we can show that $\mathcal{S}_r$ is adaptive to $\mathcal{F}_r$, and $\mathcal{S}_r$ is a mean-zero martingale sequence, i.e., $E(\mathcal{S}_r) = 0$ and $E(\mathcal{S}_{r'} \mid \mathcal{F}_r) = \sum_{j=1}^{r'} E(W_j \mid \mathcal{F}_r) = \sum_{j=1}^{r} W_j = \mathcal{S}_r$ for $r' > r$. This completes the proof of Lemma 1.

Lemma 2. Under Condition (C3) and (C4),

$$\big\{\mathrm{var}(\widehat{T}_{1,1})\big\}^{-1} \sum_{j=2}^{m+n} E(W_j^2 \mid \mathcal{F}_{j-1}) \xrightarrow{p} 1/4.$$

*Proof of Lemma 2.* Recall the definition of $\phi_{ij}$ in (S4.2). We have

$$E(W_j^2 \mid \mathcal{F}_{j-1}) = E(\sum_{i,t=1}^{j-1} \phi_{ij}\phi_{tj} \mid \mathcal{F}_{j-1}) \text{ and } \sum_{j=2}^{m+n} E(W_j^2) = \sum_{j=2}^{m+n} E\big(\sum_{i,t=1}^{j-1} \phi_{ij}\phi_{tj}\big).$$

For notation convenience, we define

$$\begin{aligned} s_1^2 &\stackrel{\text{def}}{=} \sum_{j=2}^{m}\sum_{i=t}^{j-1} E(\phi_{ij}\phi_{tj}),\ s_2^2 \stackrel{\text{def}}{=} \sum_{j=m+1}^{m+n}\sum_{i=t}^{m} E(\phi_{ij}\phi_{tj}),\ \text{and} \\ s_3^2 &\stackrel{\text{def}}{=} \sum_{j=m+1}^{m+n}\sum_{i=t=m+1}^{m+n} E(\phi_{ij}\phi_{tj}). \end{aligned}$$

Direct calculation shows that $E(\phi_{ij}\phi_{tj}) = 0$ if $i \neq t$. Hence, we have $\sum_{j=2}^{m+n} E(W_j^2) = s_1^2 + s_2^2 + s_3^2$. By the definition of $s_k^2 s$, simple calculation yields

$$\begin{aligned} s_1^2 &= \{2m(m-1)\}^{-1} E\{\varphi_{11}^2(\mathbf{x}_1, \mathbf{x}_2)\},\ s_2^2 = (mn)^{-1} E\{\varphi_{13}^2(\mathbf{x}_1, \mathbf{y}_1)\},\ \text{and} \\ s_3^2 &= \{2n(n-1)\}^{-1} E\{\varphi_{12}^2(\mathbf{y}_1, \mathbf{y}_2)\}. \end{aligned}$$

The above analysis indicates that

$$E\left\{\sum_{j=2}^{m+n} E(W_j^2 \mid \mathcal{F}_{j-1})\right\} = \sum_{j=2}^{m+n} E(W_j^2) = \left\{\mathrm{var}(\widehat{T}_{1,1})\right\}/4. \tag{S4.3}$$

Next, we consider the variance of $\sum\limits_{j=2}^{m+n} E(W_j^2 \mid \mathcal{F}_{j-1})$. Following the derivation for (S4.3), we can obtain that $\mathrm{var}\{\sum\limits_{j=2}^{m+n} E(W_j^2 \mid \mathcal{F}_{j-1})\} = 2S_1^2 + S_2^2$, where

$$S_1^2 \overset{\text{def}}{=} \sum_{2\le j'<j}^{m+n} \sum_{i,t=1}^{j-1} \sum_{i',t'=1}^{j'-1} \mathrm{cov}\Big(E(\phi_{ij}\phi_{tj} \mid \mathcal{F}_{j-1}), E(\phi_{i'j'}\phi_{k'j'} \mid \mathcal{F}_{j'-1})\Big) \tag{S4.4}$$

$$\text{and} \quad S_2^2 \overset{\text{def}}{=} \sum_{j=2}^{m+n} \sum_{i,t=1}^{j-1} \sum_{i',t'=1}^{j-1} \mathrm{cov}\Big(E(\phi_{ij}\phi_{tj} \mid \mathcal{F}_{j-1}), E(\phi_{i'j}\phi_{t'j} \mid \mathcal{F}_{j'-1})\Big). \tag{S4.5}$$

For $S_1^2$ in (S4.4), we can obtain $S_1^2 = S_{11}^2 + S_{12}^2 + S_{13}^2$, where

$$\begin{aligned}
S_{11}^2 &\overset{\text{def}}{=} \sum_{2\le j'<j}^{m} \sum_{i,t=1}^{j-1} \sum_{i',t'=1}^{j'-1} \mathrm{cov}\Big(E(\phi_{ij}\phi_{tj} \mid \mathcal{F}_{j-1}), E(\phi_{i'j'}\phi_{t'j'} \mid \mathcal{F}_{j'-1})\Big), \\
S_{12}^2 &\overset{\text{def}}{=} \sum_{m\le j'<j}^{m+n} \sum_{i,t=1}^{j-1} \sum_{i',t'=1}^{j'-1} \mathrm{cov}\Big(E(\phi_{ij}\phi_{tj} \mid \mathcal{F}_{j-1}), E(\phi_{i'j'}\phi_{t'j'} \mid \mathcal{F}_{j'-1})\Big), \\
\text{and } S_{13}^2 &\overset{\text{def}}{=} \sum_{j=m+1}^{m+n} \sum_{j'=1}^{m} \sum_{i,t=1}^{j-1} \sum_{i',t'=1}^{j'-1} \mathrm{cov}\Big(E(\phi_{ij}\phi_{tj} \mid \mathcal{F}_{j-1}), E(\phi_{i'j'}\phi_{t'j'} \mid \mathcal{F}_{j'-1})\Big).
\end{aligned}$$

In order to simplify $S_1^2$, we need to deal with $S_{1k}^2s$, respectively. Firstly, we consider the term of $S_{11}^2$. Define

$$\begin{aligned}
S_{111}^2 &\overset{\text{def}}{=} E\{\varphi_{11}^2(\mathbf{w}_i, \mathbf{w}_j)\varphi_{11}^2(\mathbf{w}_i, \mathbf{w}_{j'})\} - \{E\varphi_{11}^2(\mathbf{w}_i, \mathbf{w}_j)\}^2, \text{ and} \\
S_{112}^2 &\overset{\text{def}}{=} E\{\varphi_{11}(\mathbf{w}_i, \mathbf{w}_j)\varphi_{11}(\mathbf{w}_i, \mathbf{w}_{j'})\varphi_{11}(\mathbf{w}_{i'}, \mathbf{w}_j)\varphi_{11}(\mathbf{w}_{i'}, \mathbf{w}_{j'})\}.
\end{aligned}$$

We have

$$\begin{aligned} S_{11}^2 &= O(m^{-8})\sum_{j=3}^{m}\sum_{j'=2}^{j-1}\{(j'-1)S_{111}^2+2(j'-1)(j'-2)S_{112}^2\} \\ &= O(m^{-5}S_{111}^2)+O(m^{-4}S_{112}^2). \end{aligned} \tag{S4.6}$$

Let

$$\begin{aligned} S_{121}^2 &\stackrel{\text{def}}{=} \sum_{k=2}^{3}\left[E\{\varphi_{1k}^2(\mathbf{w}_i,\mathbf{w}_j)\varphi_{1k}^2(\mathbf{w}_i,\mathbf{w}_{j'})\}-\{E\varphi_{1k}^2(\mathbf{w}_i,\mathbf{w}_j)\}^2\right], \\ S_{122}^2 &\stackrel{\text{def}}{=} \sum_{k=2}^{3}E\left[E^2\{\varphi_{1k}(\mathbf{w}_i,\mathbf{w}_j)\varphi_{1k}(\mathbf{w}_i,\mathbf{w}_{j'})\mid(\mathbf{w}_j,\mathbf{w}_{j'})\}\right] \\ &+ E\left[\prod_{k=2}^{3}E\{\varphi_{1k}(\mathbf{w}_i,\mathbf{w}_j)\varphi_{1k}(\mathbf{w}_i,\mathbf{w}_{j'})\mid(\mathbf{w}_j,\mathbf{w}_{j'})\}\right], \\ S_{131}^2 &\stackrel{\text{def}}{=} \operatorname{cov}\{\varphi_{11}^2(\mathbf{w}_i,\mathbf{w}_j),\varphi_{13}^2(\mathbf{w}_i,\mathbf{w}_{j'})\}, \\ \text{and } S_{132}^2 &\stackrel{\text{def}}{=} E\{\varphi_{11}(\mathbf{w}_i,\mathbf{w}_j)\varphi_{11}(\mathbf{w}_{i'},\mathbf{w}_j)\varphi_{13}(\mathbf{w}_i,\mathbf{w}_{j'})\varphi_{13}(\mathbf{w}_{i'},\mathbf{w}_{j'})\}. \end{aligned}$$

Similar to the derivation for $S_{11}^2$, we have

$$S_{12}^2 = O(m^{-5}S_{121}^2)+O(m^{-4}S_{122}^2),\ \text{and } S_{13}^2 = O(m^{-5}S_{131}^2)+O(m^{-4}S_{132}^2). \tag{S4.7}$$

As a result of (S4.6) and (S4.7),

$$S_1 = O(m^{-5})(S_{111}^2+S_{121}^2+S_{131}^2)+O(m^{-4})(S_{112}^2+S_{122}^2+S_{132}^2). \tag{S4.8}$$

Furthermore, by the Hölder inequality, we obtain

$$\begin{aligned} S_{111}^2 + S_{121}^2 + S_{131}^2 &\leq \sum_{t=1}^{3} E\{\varphi_{1t}^2(\mathbf{w}_i, \mathbf{w}_j)\varphi_{1t}^2(\mathbf{w}_i, \mathbf{w}_{j'})\} \\ &+ E\{\varphi_{11}^4(\mathbf{w}_i, \mathbf{w}_j)\} + E\{\varphi_{13}^4(\mathbf{w}_i, \mathbf{w}_j)\}, \\ \text{and } S_{112}^2 + S_{122}^2 + S_{132}^2 &\leq 2\sum_{t=1}^{3} E[E^2\{\varphi_{1t}(\mathbf{w}_i, \mathbf{w}_j)\varphi_{1t}(\mathbf{w}_i, \mathbf{w}_{j'})|(\mathbf{w}_j, \mathbf{w}_{j'})\}]. \end{aligned}$$

Consequently, $S_1^2 = O(\nu_2/m^4 + \nu_3/m^5)$. Under Conditions (C3)-(C4), we obtain that $S_1^2 = o\left[\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^2\right]$. Following the derivation for $S_1^2$, we can verify that $S_2^2 = o\left[\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^2\right]$. Since $\mathrm{var}\Big\{\sum_{j=2}^{m+n} E(W_j^2 \mid \mathcal{F}_{j-1})\Big\} = 2S_1^2 + S_2^2$, we have

$$\mathrm{var}\Big\{\sum_{j=2}^{m+n} E(W_j^2 \mid \mathcal{F}_{j-1})\Big\} = o\left[\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^2\right]. \tag{S4.9}$$

Combine (S4.3) and (S4.9), we have

$$\begin{aligned} \left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-2} E\Big\{\sum_{j=2}^{m+n} E(W_j^2 \mid \mathcal{F}_{j-1})\Big\} = \left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-2} \sum_{j=2}^{m+n} E(W_j^2) = \frac{1}{4} \\ \text{and } \left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-2} \mathrm{var}\Big\{\sum_{j=2}^{m+n} E(W_j^2 \mid \mathcal{F}_{j-1})\Big\} = o(1). \end{aligned}$$

This completes the proof of Lemma 2.

Lemma 3. Under Condition C4,

$$\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-1} \sum_{j=2}^{m+n} E\{W_j^2 I(\mid W_j \mid) > \varepsilon\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-1/2} \mid \mathcal{F}_{j-1}\} \xrightarrow{p} 0.$$

*Proof of Lemma 3.* It is easy to see that

$$\begin{aligned} & \left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-1}\sum_{j=2}^{m+n}E\{W_j^2 I(\mid W_j \mid) > \varepsilon\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-1/2} \mid \mathcal{F}_{j-1}\} \\ \overset{p}{\leq} \quad & \left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^{-q/2}\varepsilon^{2-q}\sum_{j=2}^{m+n}E(W_j^q \mid \mathcal{F}_{j-1}), \end{aligned}$$

for some $q > 2$. Let $q = 4$, we can prove Lemma 3 as long as

$$E\Big\{\sum_{j=2}^{m+n}E(W_j^4 \mid \mathcal{F}_{j-1})\Big\} = o\left[\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^2\right].$$

Let $P_1 \overset{\mathrm{def}}{=} \sum_{j=2}^{m+n}\sum_{i\neq t}^{j-1}E(\phi_{ij}\phi_{tj})^2$ and $P_2 \overset{\mathrm{def}}{=} \sum_{j=2}^{m+n}\sum_{i=1}^{j-1}E(\phi_{ij})^4$. We have

$$E\Big\{\sum_{j=2}^{m+n}E(W_j^4 \mid \mathcal{F}_{j-1})\Big\} = \sum_{j=2}^{m+n}E(W_j^4) = \sum_{j=2}^{m+n}E\Big(\sum_{i=2}^{j-1}\phi_{ij}\Big)^4 = 3P_1 + p_2.$$

By Lemma 3 in Chen and Qin (2010), it suffices to show $3P_1 + P_2 = o\left[\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^2\right]$. Hölder inequality yields that

$$\begin{aligned} P_1 &= \sum_{j=2}^{m}\sum_{i\neq t}^{j-1}E(\phi_{ij}\phi_{tj})^2 + \sum_{j=m}^{m+n}\sum_{i\neq t}^{j-1}E(\phi_{ij}\phi_{tj})^2 \\ &\leq O(m^{-5})\left[\sum_{k=1}^{3}E\{\varphi_{1k}^2(\mathbf{w}_i,\mathbf{w}_j)\varphi_{1k}^2(\mathbf{w}_{i'},\mathbf{w}_j)\} + \sum_{k=2}^{3}E\{\varphi_{1k}^4(\mathbf{w}_i,\mathbf{w}_j)\}\right]. \end{aligned}$$

and $P_2 = O(m^{-6})\sum_{t=1}^{3}E\{\varphi_{1t}^4(\mathbf{w}_i,\mathbf{w}_j)\}$. This, together with (C3)-(C4), entails that

$$E\Big\{\sum_{j=2}^{m+n}E(W_j^4 \mid \mathcal{F}_{j-1})\Big\} = o\left[\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}^2\right].$$

Thus Lemma 3 follows.

It follows from Corollary 3.1 of Hall and Heyde (1980) and Lemmas

1 - 3 that $\widehat{Q}/\big\{\mathrm{var}(\widehat{T}_{1,1})\big\}^{1/2} \stackrel{d}{\to} N(0,1)$ under $H_0$, as long as both $p$ and $\min(m,n)$ diverge to $\infty$. This completes the proof for Theorem 1. □

# S5 Appendix E: Proof of Theorem 2

It is obvious that $\widehat{\mathrm{var}}(\widehat{T}_{1,1})/\mathrm{var}(\widehat{T}_{1,1}) \stackrel{p}{\to} 1$ is equivalent to $\widehat{\sigma}^2_{1k}/\sigma^2_{1k} \stackrel{p}{\to} 1$ for $k=1,2,3$. It thus suffices to prove the consistency of $\widehat{\sigma}^2_{11}$ since the proofs for $\widehat{\sigma}^2_{12}$ and $\widehat{\sigma}^2_{13}$ are very similar. Below, we shall show $E(\widehat{\sigma}^2_{11}) = \sigma^2_{11}\{1+o(1)\}$ and $\mathrm{var}(\widehat{\sigma}^2_{11}) = o(\sigma^4_{11})$. For notational simplicity, we define

$$\lambda_1(\mathbf{x}_i,\mathbf{z}_r) \stackrel{\text{def}}{=} \sum_{k=1}^{p} U_k(X_{ik},Z_{rk})F_k(Z_{rk}),$$

$$\lambda_2(\mathbf{x}_j,\mathbf{z}_r) \stackrel{\text{def}}{=} \sum_{k=1}^{p} \{F_k(Z_{rk}) - \widehat{F}_{k(-i,-j)}(Z_{rk})\}U_k(X_{jk},Z_{rk}),$$

$$\text{and } \lambda_3(\mathbf{z}_r) \stackrel{\text{def}}{=} \sum_{k=1}^{p} F_k(Z_{rk})\{F_k(Z_{rk}) - \widehat{F}_{k(-i,-j)}(Z_{rk})\}.$$

Directly calculation shows that

$$\begin{aligned}\widehat{\sigma}^2_{11} &= \left\{4\binom{m}{2}\binom{m+n}{2}\right\}^{-1} \sum_{i\neq j}^{m}\sum_{r\neq s}^{m+n} \big[\{\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r) + \lambda_1(\mathbf{x}_i,\mathbf{z}_r) + \lambda_2(\mathbf{x}_j,\mathbf{z}_r) \\ &\quad +\lambda_3(\mathbf{z}_r)\} \times \{\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s) + \lambda_1(\mathbf{x}_j,\mathbf{z}_s) + \lambda_2(\mathbf{x}_i,\mathbf{z}_s) + \lambda_3(\mathbf{z}_s)\}\big] \\ &\stackrel{\text{def}}{=} \sum_{l=1}^{16} A_l,\end{aligned}$$

where $A_1 \stackrel{\text{def}}{=} \omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s)$, $A_2 \stackrel{\text{def}}{=} \omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\lambda_1(\mathbf{x}_j,\mathbf{z}_s)$, $A_3 \stackrel{\text{def}}{=} \omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\lambda_2(\mathbf{x}_i,\mathbf{z}_s)$, $A_4 \stackrel{\text{def}}{=} \omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\lambda_3(\mathbf{z}_s)$, $A_5 \stackrel{\text{def}}{=} \lambda_1(\mathbf{x}_i,\mathbf{z}_r)\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s)$, $A_6 \stackrel{\text{def}}{=} \lambda_1(\mathbf{x}_i,\mathbf{z}_r)\lambda_1(\mathbf{x}_j,\mathbf{z}_s)$, $A_7 \stackrel{\text{def}}{=} \lambda_1(\mathbf{x}_i,\mathbf{z}_r)\lambda_2(\mathbf{x}_i,\mathbf{z}_s)$, $A_8 \stackrel{\text{def}}{=} \lambda_1(\mathbf{x}_i,\mathbf{z}_r)\lambda_3(\mathbf{z}_s)$, $A_9 \stackrel{\text{def}}{=}$

$\lambda_2(\mathbf{x}_j,\mathbf{z}_r)\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s)$, $A_{10} \stackrel{\text{def}}{=} \lambda_2(\mathbf{x}_j,\mathbf{z}_r)\lambda_1(\mathbf{x}_j,\mathbf{z}_s)$, $A_{11} \stackrel{\text{def}}{=} \lambda_2(\mathbf{x}_j,\mathbf{z}_r)\lambda_2(\mathbf{x}_i,\mathbf{z}_s)$, $A_{12} \stackrel{\text{def}}{=} \lambda_2(\mathbf{x}_j,\mathbf{z}_r)\lambda_3(\mathbf{z}_s)$, $A_{13} \stackrel{\text{def}}{=} \lambda_3(\mathbf{z}_r)\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s)$, $A_{14} \stackrel{\text{def}}{=} \lambda_3(\mathbf{z}_r)\lambda_1(\mathbf{x}_j,\mathbf{z}_s)$, $A_{15} \stackrel{\text{def}}{=} \lambda_3(\mathbf{z}_r)\lambda_2(\mathbf{x}_i,\mathbf{z}_s)$, and $A_{16} \stackrel{\text{def}}{=} \lambda_3(\mathbf{z}_r)\lambda_3(\mathbf{z}_s)$. It is easy to verify that $E(A_1) = \sigma_{11}^2$, $E(A_i) = 0$ for $i = 2, ...15$ and $E(A_{16}) = o(\sigma_{11}^2)$.

Next, we show $\text{var}(A_i)$ for $i = 1, ...16$. Following similar arguments to prove Theorem 2 in Chen and Qin (2010), we can show that $\text{var}(A_i) = o(\sigma_{11}^4)$ for $i = 1, ...16$. For notational convenience, we only derive $\text{var}(A_1)$ since derivations for other $\text{var}(A_i)$ are similar. Denote

$B_1 = O(m^{-1})E[E^2\{\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s) \mid \mathbf{z}_r\}]$,

$B_2 = O(m^{-2})E[E^2\{\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s) \mid (\mathbf{z}_r,\mathbf{z}_s)\}]$,

$B_3 = O(m^{-1})E[E^2\{\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s) \mid \mathbf{x}_j\}]$,

$B_4 = O(m^{-2})E[E^2\{\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s) \mid (\mathbf{x}_i,\mathbf{z}_r)\}]$,

$B_5 = O(m^{-3})E[E^2\{\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s) \mid (\mathbf{x}_i,\mathbf{z}_r,\mathbf{z}_s)\}]$,

$B_6 = O(m^{-2})E[E^2\{\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s) \mid (\mathbf{x}_i,\mathbf{x}_j)\}]$,

$B_7 = O(m^{-3})E[E^2\{\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s) \mid (\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\}]$, and

$B_8 = O(m^{-4})E\{\omega_{11}^2(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\omega_{11}^2(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s)\}$. It is straightforward to obtain that

$$\begin{aligned}\text{var}(A_1) &= \left\{4\binom{m}{2}\binom{m+n}{2}\right\}^{-2}\sum_{i\neq j}^{m}\sum_{i'\neq j'}^{m}\sum_{r\neq s}^{m+n}\sum_{r'\neq s'}^{m+n} E\big[\{\omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r) \\ &\quad \omega_{11}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_s)\}\times\{\omega_{11}(\mathbf{x}_{i'},\mathbf{x}_{j'},\mathbf{z}_{r'})\omega_{11}(\mathbf{x}_{i'},\mathbf{x}_{j'},\mathbf{z}_{s'})\}\big] - \sigma_{11}^4 \\ &\stackrel{\text{def}}{=} \sum_{l=1}^{8} B_l + o(\sigma_{11}^4).\end{aligned}$$

Consider $B_1$. By Hölder inequality, we have,

$$B_1 \leq O[m^{-1/2}E\{\omega_{11}^2(\mathbf{x}_i,\mathbf{x}_j,\mathbf{z}_r)\}]^2.$$

This, together with Condition (C2), implies $B_1 = o(\sigma_{11}^4)$ immediately. Following similar arguments, we can show that $B_l = o(\sigma_{11}^4)$ for $l = 2, ...8$ under Conditions (C1)-(C4). The proof of Theorem 2 is completed. □

## S6 Appendix F: Proof of Theorem 3

Under local alternatives $H_1'$,

$$\max_{1\leq k\leq p} E\{F_k(Z_{rk}) - G_k(Z_{rk})\}^2 = o_p(n^{-1/2}).$$

By Hoeffding decomposition in Appendix D, we have, under conditions $\nu_1 = o\big\{\text{var}(\widehat{T}_{1,1})\big\}$ and $\nu_5 = o\big\{\text{var}(\widehat{T}_{1,1})\big\}$,

$$\big\{\widehat{\text{var}}(\widehat{T}_{1,1})\big\}^{-1/2}(\widehat{Q} - D_0) = \big\{\widehat{\text{var}}(\widehat{T}_{1,1})\big\}^{-1/2}(\widehat{T}_{1,1} + \widehat{T}_2) + o_p(1), \tag{S6.1}$$

where $\widehat{T}_2$ is asymptotically negligible. It follows that

$$\big\{\widehat{\text{var}}(\widehat{T}_{1,1})\big\}^{-1/2}(\widehat{Q} - D_0) = \big\{\widehat{\text{var}}(\widehat{T}_{1,1})\big\}^{-1/2}\widehat{T}_{1,1} + o_p(1).$$

Following the proof of Theorem 1 - 2, we can conclude that, under Condition (C1)-(C4), $\nu_5 = o\big\{\text{var}(\widehat{T}_{1,1})\big\}$, and $H_1'$, $(\widehat{Q} - D_0)/\big\{\widehat{\text{var}}(\widehat{T}_{1,1})\big\}^{1/2} \xrightarrow{d} N(0,1)$, as $p, \min(m,n) \to \infty$. This completes the proof of Theorem 3.

## S7 Appendix G: Additional Simulation Results

We conduct additional simulations to assess the finite sample performance of the proposed tests. Let $N(\mathbf{u}, \mathbf{\Sigma})$ stand for the multivariate normal distribution with location vector $\mathbf{u}$ and shape matrix $\mathbf{\Sigma}$, $\chi^2(1)$ stand for chi-squared distribution with one degree of freedom, and Exponential(1) stand for an exponential distribution with mean parameter 1.

**Example 1.** We consider two scenarios.

1. $X_{ik}$s are drawn independently from $N(1,1)$ distribution, and $Y_{jl}$s are generated independently from Exponential(1) distribution.

2. $X_{ik}$s and $Y_{jl}$s are drawn independently from $N(1,2)$ distribution, for $k = 1, \ldots, p$ and $l = 1, \ldots, [p/2]$, and $Y_{jl}$s are generated independently from $\chi^2(1)$ distribution, for $l = [p/2] + 1, \ldots, p$.

In both scenarios, the marginal mean and variance parameters of $\mathbf{x}$ are

Table 1: The empirical sizes and powers when $d = 2$.

| | QXZ | PTWZ | R | BMG | BG | BF | HT | H | MBG |
|---|---|---|---|---|---|---|---|---|---|
| $p$ | $(\delta, \sigma^2) = (0, 1.0)$ | | | | | | | | |
| 90 | 0.046 | 0.054 | 0.052 | 0.047 | 0.045 | 0.058 | 0.065 | 0.056 | 0.058 |
| 150 | 0.039 | 0.050 | 0.054 | 0.048 | 0.056 | 0.065 | 0.056 | 0.035 | 0.047 |
| 200 | 0.054 | 0.074 | 0.053 | 0.045 | 0.041 | 0.040 | 0.046 | 0.058 | 0.044 |
| 1000 | 0.037 | 0.049 | 0.059 | 0.050 | 0.037 | 0.043 | 0.005 | 0.044 | 0.044 |
| 1500 | 0.036 | 0.049 | 0.066 | 0.048 | 0.069 | 0.073 | 0.058 | 0.038 | 0.039 |
| $p$ | $(\delta, \sigma^2) = (0.25, 1.0)$ | | | | | | | | |
| 90 | 0.325 | 0.055 | 0.286 | 0.152 | 0.043 | 0.175 | 0.040 | 0.463 | 0.062 |
| 150 | 0.419 | 0.054 | 0.386 | 0.213 | 0.059 | 0.189 | 0.062 | 0.573 | 0.064 |
| 200 | 0.491 | 0.075 | 0.459 | 0.271 | 0.044 | 0.180 | 0.042 | 0.612 | 0.045 |
| 1000 | 0.743 | 0.051 | 0.949 | 0.715 | 0.039 | 0.160 | 0.062 | 0.804 | 0.046 |
| 1500 | 0.759 | 0.051 | 0.986 | 0.850 | 0.057 | 0.179 | 0.046 | 0.814 | 0.055 |
| $p$ | $(\delta, \sigma^2) = (0.15, 2.0)$ | | | | | | | | |
| 90 | 0.569 | 0.585 | 0.107 | 0.193 | 0.300 | 0.589 | 0.396 | 0.353 | 0.625 |
| 150 | 0.653 | 0.589 | 0.149 | 0.213 | 0.275 | 0.505 | 0.366 | 0.339 | 0.601 |
| 200 | 0.678 | 0.620 | 0.155 | 0.237 | 0.289 | 0.528 | 0.362 | 0.316 | 0.595 |
| 1000 | 0.769 | 0.590 | 0.282 | 0.440 | 0.278 | 0.544 | 0.336 | 0.345 | 0.502 |
| 1500 | 0.756 | 0.580 | 0.366 | 0.548 | 0.304 | 0.556 | 0.360 | 0.333 | 0.516 |
| $p$ | $(\delta, \sigma^2) = (0.0, 2.5)$ | | | | | | | | |
| 90 | 0.628 | 0.825 | 0.097 | 0.195 | 0.466 | 0.647 | 0.605 | 0.227 | 0.829 |
| 150 | 0.681 | 0.823 | 0.098 | 0.190 | 0.441 | 0.637 | 0.597 | 0.195 | 0.830 |
| 200 | 0.707 | 0.822 | 0.102 | 0.198 | 0.450 | 0.649 | 0.581 | 0.154 | 0.803 |
| 1000 | 0.786 | 0.829 | 0.118 | 0.216 | 0.438 | 0.675 | 0.575 | 0.058 | 0.752 |
| 1500 | 0.767 | 0.828 | 0.102 | 0.236 | 0.426 | 0.644 | 0.555 | 0.044 | 0.719 |

identical to those of $\mathbf{y}$. However, in the first scenario, $F_k \neq G_k$ for $k = 1, \ldots, p$, and in the second scenario, $F_k \neq G_k$ for $k = [p/2] + 1, \ldots, p$.

Table 4 shows that our proposed QXZ test is the most powerful, followed by the R and MBG tests, while all other tests are very insensitive to such distributional difference. The power performance of the MBG test is also very insensitive to the dimension $p$, while the R test improves as $p$ increases.

**Example 2.** In this example, we draw $\mathbf{x}_i$s independently from $N(\mathbf{u}_1, \mathbf{\Sigma}_1)$, and generate $\mathbf{y}_j$s independently from $\sim N(\mathbf{u}_2, \mathbf{\Sigma}_2)$. We consider three cases.

Table 2: The empirical sizes and powers when $d = 3$.

| | QXZ | PTWZ | R | BMG | BG | BF | HT | H | MBG |
|---|---|---|---|---|---|---|---|---|---|
| $p$ | $(\delta, \sigma^2) = (0, 1.0)$ | | | | | | | | |
| 90 | 0.053 | 0.048 | 0.052 | 0.055 | 0.046 | 0.046 | 0.040 | 0.047 | 0.061 |
| 150 | 0.047 | 0.044 | 0.055 | 0.054 | 0.046 | 0.043 | 0.047 | 0.035 | 0.043 |
| 200 | 0.040 | 0.058 | 0.049 | 0.051 | 0.053 | 0.055 | 0.046 | 0.051 | 0.041 |
| 1000 | 0.048 | 0.060 | 0.052 | 0.065 | 0.058 | 0.052 | 0.062 | 0.035 | 0.044 |
| 1500 | 0.037 | 0.052 | 0.063 | 0.057 | 0.062 | 0.064 | 0.047 | 0.039 | 0.049 |
| $p$ | $(\delta, \sigma^2) = (0.25, 1.0)$ | | | | | | | | |
| 90 | 0.481 | 0.053 | 0.279 | 0.155 | 0.060 | 0.398 | 0.065 | 0.511 | 0.068 |
| 150 | 0.614 | 0.046 | 0.385 | 0.243 | 0.053 | 0.474 | 0.066 | 0.619 | 0.055 |
| 200 | 0.721 | 0.061 | 0.495 | 0.290 | 0.047 | 0.491 | 0.060 | 0.683 | 0.058 |
| 1000 | 0.982 | 0.064 | 0.963 | 0.806 | 0.065 | 0.581 | 0.046 | 0.860 | 0.038 |
| 1500 | 0.983 | 0.053 | 0.992 | 0.915 | 0.047 | 0.606 | 0.037 | 0.859 | 0.045 |
| $p$ | $(\delta, \sigma^2) = (0.15, 2.0)$ | | | | | | | | |
| 90 | 0.689 | 0.792 | 0.131 | 0.265 | 0.516 | 0.753 | 0.573 | 0.332 | 0.746 |
| 150 | 0.777 | 0.785 | 0.143 | 0.313 | 0.549 | 0.816 | 0.584 | 0.295 | 0.751 |
| 200 | 0.834 | 0.801 | 0.164 | 0.349 | 0.524 | 0.827 | 0.570 | 0.290 | 0.738 |
| 1000 | 0.950 | 0.804 | 0.307 | 0.568 | 0.541 | 0.840 | 0.552 | 0.252 | 0.660 |
| 1500 | 0.959 | 0.819 | 0.389 | 0.681 | 0.537 | 0.853 | 0.567 | 0.252 | 0.638 |
| $p$ | $(\delta, \sigma^2) = (0.0, 2.5)$ | | | | | | | | |
| 90 | 0.716 | 0.953 | 0.102 | 0.311 | 0.720 | 0.864 | 0.808 | 0.195 | 0.917 |
| 150 | 0.800 | 0.960 | 0.103 | 0.296 | 0.747 | 0.882 | 0.808 | 0.156 | 0.910 |
| 200 | 0.866 | 0.959 | 0.119 | 0.357 | 0.734 | 0.884 | 0.812 | 0.103 | 0.907 |
| 1000 | 0.937 | 0.962 | 0.103 | 0.389 | 0.759 | 0.913 | 0.810 | 0.029 | 0.871 |
| 1500 | 0.942 | 0.960 | 0.119 | 0.403 | 0.740 | 0.918 | 0.777 | 0.031 | 0.850 |

1. $\mathbf{u}_1 = \mathbf{0}_{p\times 1}$, $\mathbf{u}_2 = 2(\mathbf{1}_{1\times 5}, \mathbf{0}_{1\times(p-5)})^{\mathrm{T}}$, $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = (0.5^{|k-l|})_{p\times p}$.

2. $\mathbf{u}_1 = \mathbf{0}_{p\times 1}$, $\mathbf{u}_2 = 2(\mathbf{1}_{1\times 5}, \mathbf{0}_{1\times(p-5)})^{\mathrm{T}}$, $\mathbf{\Sigma}_1 = (0.5^{|k-l|})_{p\times p}$, $\mathbf{\Sigma}_2 = 2(0.2^{|k-l|})_{p\times p}$.

3. $\mathbf{u}_1 = \mathbf{u}_2 = \mathbf{0}_{p\times 1}$, $\mathbf{\Sigma}_1 = (0.5^{|k-l|})_{p\times p}$, and $\mathbf{\Sigma}_2 = 2(0.2^{|k-l|})_{p\times p}$.

In all three cases, both $\mathbf{x}$ and $\mathbf{y}$ follow normal distribution. However, the mean vectors are sparsely differently in cases 1 and 2, and the covariance matrices are different in cases 2 and 3. All three cases are expected to be very challenging to many two-sample tests.

The simulation results are summarized in Table 5. It can be clearly seen that, the empirical powers of all tests deteriorate as $p$ increases in

Table 3: The empirical sizes and powers when $d = 30$.

| | QXZ | PTWZ | R | BMG | BG | BF | HT | H | MBG |
|---|---|---|---|---|---|---|---|---|---|
| $p$ | $(\delta, \sigma^2) = (0, 1.0)$ | | | | | | | | |
| 90 | 0.043 | 0.053 | 0.064 | 0.051 | 0.053 | 0.061 | 0.041 | 0.046 | 0.058 |
| 150 | 0.054 | 0.041 | 0.056 | 0.067 | 0.040 | 0.053 | 0.044 | 0.050 | 0.033 |
| 200 | 0.069 | 0.052 | 0.059 | 0.054 | 0.046 | 0.043 | 0.054 | 0.037 | 0.044 |
| 1000 | 0.056 | 0.059 | 0.066 | 0.051 | 0.041 | 0.052 | 0.060 | 0.045 | 0.048 |
| 1500 | 0.049 | 0.059 | 0.049 | 0.054 | 0.047 | 0.053 | 0.056 | 0.040 | 0.055 |
| $p$ | $(\delta, \sigma^2) = (0.25, 1.0)$ | | | | | | | | |
| 90 | 0.772 | 0.070 | 0.286 | 0.244 | 0.077 | 0.914 | 0.078 | 0.568 | 0.067 |
| 150 | 0.952 | 0.054 | 0.444 | 0.343 | 0.071 | 0.993 | 0.073 | 0.695 | 0.079 |
| 200 | 0.983 | 0.064 | 0.541 | 0.405 | 0.052 | 1.000 | 0.064 | 0.742 | 0.067 |
| 1000 | 1.000 | 0.069 | 0.989 | 0.956 | 0.067 | 1.000 | 0.062 | 0.962 | 0.050 |
| 1500 | 1.000 | 0.071 | 0.998 | 0.995 | 0.060 | 1.000 | 0.067 | 0.975 | 0.046 |
| $p$ | $(\delta, \sigma^2) = (0.15, 2.0)$ | | | | | | | | |
| 90 | 0.972 | 1.000 | 0.140 | 0.733 | 1.000 | 1.000 | 1.000 | 0.030 | 1.000 |
| 150 | 0.996 | 1.000 | 0.158 | 0.871 | 1.000 | 1.000 | 1.000 | 0.003 | 1.000 |
| 200 | 1.000 | 1.000 | 0.165 | 0.932 | 1.000 | 1.000 | 1.000 | 0.003 | 1.000 |
| 1000 | 1.000 | 1.000 | 0.384 | 1.000 | 1.000 | 1.000 | 1.000 | 0.002 | 0.997 |
| 1500 | 1.000 | 1.000 | 0.470 | 1.000 | 1.000 | 1.000 | 1.000 | 0.001 | 0.993 |
| $p$ | $(\delta, \sigma^2) = (0.0, 2.5)$ | | | | | | | | |
| 90 | 0.995 | 1.000 | 0.116 | 0.927 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 150 | 1.000 | 1.000 | 0.112 | 0.970 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 200 | 1.000 | 1.000 | 0.117 | 0.988 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 1000 | 1.000 | 1.000 | 0.135 | 1.000 | 1.000 | 1.000 | 1.000 | 0.000 | 0.999 |
| 1500 | 1.000 | 1.000 | 0.144 | 1.000 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |

Table 4: The empirical powers of all tests in Example 1.

| $p$ | QXZ | PTWZ | R | BMG | BG | BF | HT | H | MBG |
|---|---|---|---|---|---|---|---|---|---|
| | scenario 1: $F_k \neq G_k$ for $k = 1, \ldots, p$. | | | | | | | | |
| 30 | 0.961 | 0.159 | 0.145 | 0.091 | 0.065 | 0.071 | 0.085 | 0.121 | 0.538 |
| 90 | 1.000 | 0.117 | 0.165 | 0.081 | 0.052 | 0.054 | 0.059 | 0.050 | 0.554 |
| 150 | 1.000 | 0.121 | 0.224 | 0.075 | 0.055 | 0.050 | 0.057 | 0.048 | 0.488 |
| 200 | 1.000 | 0.115 | 0.246 | 0.061 | 0.059 | 0.060 | 0.065 | 0.035 | 0.517 |
| 500 | 1.000 | 0.113 | 0.421 | 0.054 | 0.053 | 0.056 | 0.046 | 0.038 | 0.469 |
| 1000 | 1.000 | 0.109 | 0.628 | 0.047 | 0.053 | 0.048 | 0.046 | 0.034 | 0.509 |
| 1500 | 1.000 | 0.108 | 0.761 | 0.075 | 0.056 | 0.047 | 0.056 | 0.025 | 0.472 |
| 2000 | 1.000 | 0.107 | 0.850 | 0.059 | 0.063 | 0.056 | 0.069 | 0.050 | 0.459 |
| | scenario 2: $F_k \neq G_k$ for $k = [p/2] + 1, \ldots, p$. | | | | | | | | |
| 30 | 0.967 | 0.151 | 0.129 | 0.086 | 0.071 | 0.057 | 0.102 | 0.108 | 0.426 |
| 90 | 1.000 | 0.134 | 0.173 | 0.056 | 0.067 | 0.054 | 0.077 | 0.065 | 0.470 |
| 150 | 1.000 | 0.140 | 0.199 | 0.074 | 0.066 | 0.060 | 0.068 | 0.050 | 0.472 |
| 200 | 1.000 | 0.108 | 0.229 | 0.061 | 0.047 | 0.048 | 0.052 | 0.041 | 0.484 |
| 500 | 1.000 | 0.110 | 0.340 | 0.049 | 0.062 | 0.050 | 0.055 | 0.027 | 0.496 |
| 1000 | 1.000 | 0.120 | 0.502 | 0.066 | 0.059 | 0.049 | 0.060 | 0.041 | 0.482 |
| 1500 | 1.000 | 0.109 | 0.624 | 0.076 | 0.046 | 0.051 | 0.059 | 0.037 | 0.474 |
| 2000 | 1.000 | 0.088 | 0.689 | 0.081 | 0.040 | 0.056 | 0.042 | 0.033 | 0.486 |

case 1 where the mean vectors are sparsely different. This is not surprising because the signals are more and more sparse as $p$ increases. Comparatively speaking, the BF test is the most insensitive to the dimension $p$, followed by the QXZ and H tests. Both the R and H tests, however, are very insensitive to the scale difference. They have the smallest power in case 2 and 3. Our proposed test is very powerful, particularly when $p$ is large.

Table 5: The empirical powers of all tests in Example 2.

| $p$ | QXZ | PTWZ | R | BMG | BG | BF | HT | H | MBG |
|---|---|---|---|---|---|---|---|---|---|
| | case 1: $\mathbf{u}_1 = \mathbf{0}_{p\times 1}$, $\mathbf{u}_2 = 2(\mathbf{1}_{1\times 5}, \mathbf{0}_{1\times(p-5)})^{\mathrm{T}}$, $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = (0.5^{\lvert k-l\rvert})$. | | | | | | | | |
| 30 | 1.000 | 1.000 | 0.992 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 90 | 1.000 | 1.000 | 0.813 | 0.991 | 1.000 | 1.000 | 1.000 | 1.000 | 0.976 |
| 150 | 1.000 | 0.999 | 0.678 | 0.968 | 0.995 | 1.000 | 0.989 | 0.998 | 0.915 |
| 200 | 1.000 | 0.489 | 0.568 | 0.946 | 0.975 | 1.000 | 0.955 | 0.993 | 0.805 |
| 500 | 0.979 | 0.998 | 0.329 | 0.677 | 0.572 | 1.000 | 0.633 | 0.927 | 0.402 |
| 1000 | 0.788 | 0.870 | 0.219 | 0.449 | 0.211 | 0.996 | 0.325 | 0.770 | 0.210 |
| 1500 | 0.629 | 0.313 | 0.169 | 0.314 | 0.136 | 0.981 | 0.216 | 0.609 | 0.104 |
| 2000 | 0.518 | 0.229 | 0.156 | 0.269 | 0.112 | 0.964 | 0.153 | 0.513 | 0.084 |
| | case 2: $\mathbf{u}_1 = \mathbf{0}$, $\mathbf{u}_2 = 2(\mathbf{1}_{1\times 5}, \mathbf{0}_{1\times(p-5)})^{\mathrm{T}}$, $\mathbf{\Sigma}_1 = (0.5^{\lvert k-l\rvert})$, $\mathbf{\Sigma}_2 = 2(0.2^{\lvert k-l\rvert})$. | | | | | | | | |
| 30 | 0.992 | 1.000 | 0.534 | 0.936 | 1.000 | 1.000 | 1.000 | 0.655 | 1.000 |
| 90 | 0.997 | 1.000 | 0.383 | 0.955 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 150 | 0.999 | 1.000 | 0.359 | 0.986 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 200 | 1.000 | 1.000 | 0.296 | 0.988 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 500 | 1.000 | 1.000 | 0.274 | 1.000 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 1000 | 1.000 | 1.000 | 0.258 | 1.000 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 1500 | 1.000 | 1.000 | 0.243 | 1.000 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 2000 | 1.000 | 1.000 | 0.234 | 1.000 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| | case 3: $\mathbf{u}_1 = \mathbf{u}_2 = \mathbf{0}$, $\mathbf{\Sigma}_1 = (0.5^{\lvert k-l\rvert})$, and $\mathbf{\Sigma}_2 = 2(0.2^{\lvert k-l\rvert})$. | | | | | | | | |
| 30 | 0.393 | 1.000 | 0.154 | 0.534 | 1.000 | 0.907 | 1.000 | 0.134 | 1.000 |
| 90 | 0.824 | 1.000 | 0.144 | 0.819 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 150 | 0.973 | 1.000 | 0.157 | 0.921 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 200 | 0.996 | 1.000 | 0.145 | 0.950 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 500 | 1.000 | 1.000 | 0.172 | 0.998 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 1000 | 1.000 | 1.000 | 0.172 | 1.000 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 1500 | 1.000 | 1.000 | 0.186 | 1.000 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |
| 2000 | 1.000 | 1.000 | 0.213 | 1.000 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 |

**Example 3.** We draw $\mathbf{x}_i$s independently from $N(\mathbf{0}_{p\times 1}, \mathbf{\Sigma}_1)$ with $\mathbf{\Sigma}_1 = (0.2^{\lvert k-l\rvert})_{p\times p}$, and $\mathbf{y}_i$s independently from $N(\mathbf{0}_{p\times 1}, \mathbf{\Sigma}_2)$ with $\mathbf{\Sigma}_2 = (0.7^{\lvert k-l\rvert})_{p\times p}$.

In this example, both the marginal means and the marginal variances of $\mathbf{x}$ and $\mathbf{y}$ are the same. Their correlation structures are however different.

The simulation results are charted in Table 6. It is not surprising to see that our proposed two-sample test has the smallest power, because our proposed test compares the marginal differences between $F_k$ and $G_k$, for $k = 1, \ldots, p$. The BG and BF tests suffer from similar issues. The MBG test is the most powerful, followed by the R and the BMG tests.

Table 6: The empirical powers of all tests in Example 3.

| $p$ | QXZ | PTWZ | R | BMG | BG | BF | HT | H | MBG |
|---|---|---|---|---|---|---|---|---|---|
| 30 | 0.036 | 0.196 | 0.495 | 0.451 | 0.079 | 0.106 | 0.151 | 0.449 | 0.940 |
| 90 | 0.042 | 0.169 | 0.534 | 0.563 | 0.053 | 0.085 | 0.092 | 0.396 | 0.924 |
| 150 | 0.048 | 0.166 | 0.589 | 0.539 | 0.054 | 0.068 | 0.078 | 0.391 | 0.878 |
| 200 | 0.050 | 0.163 | 0.585 | 0.553 | 0.054 | 0.088 | 0.091 | 0.390 | 0.874 |
| 500 | 0.051 | 0.150 | 0.604 | 0.573 | 0.057 | 0.077 | 0.082 | 0.354 | 0.830 |
| 1000 | 0.055 | 0.154 | 0.600 | 0.579 | 0.047 | 0.054 | 0.092 | 0.399 | 0.793 |
| 1500 | 0.058 | 0.162 | 0.578 | 0.559 | 0.052 | 0.064 | 0.072 | 0.376 | 0.823 |
| 2000 | 0.054 | 0.170 | 0.575 | 0.596 | 0.054 | 0.056 | 0.056 | 0.342 | 0.815 |

□

## References

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

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

Lee, A.J. (1990). *U-Statistics: Theory and Practice.* Textbooks and Mono-graphs. M. Dekker.

Zhang, X., Yao, S., and Shao, X. (2018). Conditional mean and quantile dependence testing in high dimension. *Annals of Statistics, to appear.*

Tao Qiu
School of Statistics, Renmin University of China, Beijing 100872, China.
E-mail: qiutao8505@126.com
Wangli Xu
School of Statistics, Renmin University of China, Beijing 100872, China.
E-mail: wlxu@ruc.edu.cn
Liping Zhu
Center for Applied Statistics, Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China.
Schoo of Statistics and Mathmatics, ZheJiang Gongshang University, Hangzhou 310018, China.
E-mail: zhu.liping@ruc.edu.cn