Supplement to "Conditional Mean and Quantile Dependence Testing in High Dimension"

Xianyang Zhang, Shun Yao, and Xiaofeng Shao*

Texas A&M University and University of Illinois at Urbana-Champaign

This version: March 21, 2017

1 Technical Appendix

Throughout the technical appendix, we let $c, c', c'', C, C', C'', c_i, C_i$ be generic constants which can be different from line to line.

1.1 Unbiasedness of $MDD_n(\mathcal{Y}|\mathcal{X})^2$

By Lemma 1 of Park et al. (2014), we have $n(n-3)(\widetilde{A}\cdot\widetilde{B}) = \sum_{i\neq j}\widetilde{A}_{ij}\widetilde{B}_{ij} = \sum_{i\neq j}A_{ij}\widetilde{\widetilde{B}}_{ij}$. Using the fact that $B_{ii}=0$, it can be verified that $\widetilde{\widetilde{B}}_{ij}=\widetilde{B}_{ij}$. By the definition of $\mathcal{U}$-centering, we have

$$\begin{aligned}\sum_{i\neq j}A_{ij}\widetilde{B}_{ij} =&\sum_{i\neq j}A_{ij}\left(B_{ij}-\frac{1}{n-2}\sum_{l=1}^{n}B_{il}-\frac{1}{n-2}\sum_{k=1}^{n}B_{kj}+\frac{1}{(n-1)(n-2)}\sum_{k,l=1}^{n}B_{kl}\right)\\ =&\sum_{i\neq j}A_{ij}B_{ij}-\frac{1}{n-2}\sum_{i\neq j}A_{ij}\sum_{l=1}^{n}B_{il}-\frac{1}{n-2}\sum_{i\neq j}A_{ij}\sum_{k=1}^{n}B_{kj}\\ &+\frac{1}{(n-1)(n-2)}\sum_{i,j=1}^{n}A_{ij}\sum_{k,l=1}^{n}B_{kl}\\ =&\mathrm{tr}(AB)+\frac{\mathbf{1}_p^TA\mathbf{1}_p\mathbf{1}_p^TB\mathbf{1}_p}{(n-1)(n-2)}-\frac{2\mathbf{1}_p^TAB\mathbf{1}_p}{(n-2)},\end{aligned}$$

*Xianyang Zhang is Assistant Professor in the Department of Statistics, Texas A&M University, College Station, TX 77843, USA. E-mail: zhangxiany@stat.tamu.edu. Shun Yao is a Ph.D. candidate and Xiaofeng Shao is Professor in the Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA. E-mails: shunyao2,xshao@illinois.edu. Zhang's research is partially supported by NSF DMS-1607320. Shao's research is partially supported by NSF-DMS1104545, NSF-DMS1407037 and DMS-1607489. We would like to thank Dr. Ping-Shou Zhong for providing the data set and the R codes used for the tests in the models with fractional design. We are also grateful to Dr. Min Qian for providing the matlab code for the methods developed in McKeague and Qian (2015).

where $\mathbf{1}_p \in \mathbb{R}^p$ is the vector of all ones. Thus we have

$$MDD_n(\mathcal{Y}|\mathcal{X})^2 = \frac{1}{n(n-3)}\left(\mathrm{tr}(AB) + \frac{\mathbf{1}_p^T A\mathbf{1}_p\mathbf{1}_p^T B\mathbf{1}_p}{(n-1)(n-2)} - \frac{2\mathbf{1}_p^T AB\mathbf{1}_p}{(n-2)}\right).$$

Let $(n)_k = n!/(n-k)!$ and I_k^n be the collections of k-tuples of indices (chosen from $\{1,2,\ldots,n\}$) such that each index occurs exactly once. It can be shown that

$$(n)_2^{-1}\mathbb{E}\left[\sum_{(i,j)\in I_2^n} A_{ij}B_{ij}\right] = (n)_2^{-1}\mathbb{E}[\mathrm{tr}(AB)] = \mathbb{E}[K(\mathcal{X},\mathcal{X}')L(\mathcal{Y},\mathcal{Y}')],$$

$$(n)_4^{-1}\mathbb{E}\left[\sum_{(i,j,q,r)\in I_4^n} A_{ij}B_{qr}\right] = (n)_4^{-1}\mathbb{E}[\mathbf{1}_p^T A\mathbf{1}_p\mathbf{1}_p^T B\mathbf{1}_p - 4\mathbf{1}_p^T AB\mathbf{1}_p + 2\mathrm{tr}(AB)] = \mathbb{E}[K(\mathcal{X},\mathcal{X}')]\mathbb{E}[L(\mathcal{Y},\mathcal{Y}')],$$

$$(n)_3^{-1}\mathbb{E}\left[\sum_{(i,j,r)\in I_3^n} A_{ij}B_{ir}\right] = (n)_3^{-1}\mathbb{E}[\mathbf{1}_p^T AB\mathbf{1}_p - \mathrm{tr}(AB)] = \mathbb{E}[K(\mathcal{X},\mathcal{X}')L(\mathcal{Y},\mathcal{Y}'')],$$

$$MDD_n(\mathcal{Y}|\mathcal{X})^2 = (n)_2^{-1}\sum_{(i,j)\in I_2^n} A_{ij}B_{ij} + (n)_4^{-1}\sum_{(i,j,q,r)\in I_4^n} A_{ij}B_{qr} - 2(n)_3^{-1}\sum_{(i,j,r)\in I_3^n} A_{ij}B_{ir},$$

which implies the unbiasedness of $MDD_n(\mathcal{Y}|\mathcal{X})^2$. The above derivation also indicates that

$$MDD_n(\mathcal{Y}|\mathcal{X})^2 =(n)_4^{-1}\sum_{(i,j,q,r)\in I_4^n}(A_{ij}B_{qr} + A_{ij}B_{ij} - 2A_{ij}B_{ir}) = \frac{1}{\binom{n}{4}}\sum_{i<j<q<r} h(\mathcal{Z}_i,\mathcal{Z}_j,\mathcal{Z}_q,\mathcal{Z}_r),$$

where

$$\begin{aligned} h(\mathcal{Z}_i,\mathcal{Z}_j,\mathcal{Z}_q,\mathcal{Z}_r) =& \frac{1}{4!}\sum_{(s,t,u,v)}^{(i,j,q,r)}(A_{st}B_{uv} + A_{st}B_{st} - 2A_{st}B_{su}) \\ =& \frac{1}{6}\sum_{s<t,u<v}^{(i,j,q,r)}(A_{st}B_{uv} + A_{st}B_{st}) - \frac{1}{12}\sum_{(s,t,u)}^{(i,j,q,r)} A_{st}B_{su}, \end{aligned} \tag{1}$$

with $\mathcal{Z}_i = (\mathcal{X}_i,\mathcal{Y}_i)$, and the summation is over all permutation of the 4-tuples of indices (i,j,q,r). Therefore $MDD_n(\mathcal{Y}|\mathcal{X})^2$ is a U-statistic of order four. From the above arguments, we have

$$\sum_{k=1}^p MDD_n(Y|x_k)^2 = \frac{1}{\binom{n}{4}}\sum_{i<j<q<r}\sum_{k=1}^p \mathfrak{h}_k(Z_{ik},Z_{jk},Z_{qk},Z_{rk}),$$

where $Z_{ik} = (x_{ik}, Y_i)$ and $\mathfrak{h}_k$ is defined in a similar way as h by replacing A_{st} with $A_{st}(k)$. Thus $\sum_{k=1}^p MDD_n(Y|x_k)^2$ is also a fourth order U-statistic with the kernel $\sum_{k=1}^p \mathfrak{h}_k$.

1.2 Hoeffding decomposition

Recall the definition of h in (1). Define $h_c(w_1,\ldots,w_c) = \mathbb{E}h(w_1,\ldots,w_c,\mathcal{Z}_{c+1},\cdots,\mathcal{Z}_4)$, where $\mathcal{Z}_i = (\mathcal{X}_i,\mathcal{Y}_i) \overset{D}{=} (\mathcal{X},\mathcal{Y})$ and $c = 1,2,3,4$. The symbol "$\overset{D}{=}$" here means "equal in distribution". Let $w = (x, y)$, $w' = (x', y')$, $w'' = (x'', y'')$ and $w''' = (x''', y''')$, where $x, x', x'', x''' \in \mathbb{R}^q$ and $y, y', y'', y''' \in \mathbb{R}$. Further let $\mathcal{Z}' = (\mathcal{X}', \mathcal{Y}')$, $\mathcal{Z}'' = (\mathcal{X}'', \mathcal{Y}'')$ and $\mathcal{Z}''' = (\mathcal{X}''', \mathcal{Y}''')$ be independent copies of $\mathcal{Z} = (\mathcal{X}, \mathcal{Y})$. With some abuse of notation, we define $U(x, x') = \mathbb{E}[K(x, \mathcal{X}')] + \mathbb{E}[K(\mathcal{X}, x')] - K(x, x') - \mathbb{E}[K(\mathcal{X}, \mathcal{X}')]$ and $V(y, y') = (y - \mu_{\mathcal{Y}})(y' - \mu_{\mathcal{Y}})$ with $\mu_{\mathcal{Y}} = \mathbb{E}\mathcal{Y}$. In the following derivations, $(\mathcal{X}_i, \mathcal{Y}_i)$ is allowed to be an triangular array and the arguments are still valid for general kernels U_n and V_n which can vary with n. Direct calculation shows that

$$h_1(w) = \frac{1}{2}\Big\{\mathbb{E}[U(x, \mathcal{X})V(y, \mathcal{Y})] + MDD(\mathcal{Y}|\mathcal{X})^2\Big\},$$

and

$$\begin{aligned} h_2(w, w') =& \frac{1}{6}\Big\{U(x, x')V(y, y') + MDD(\mathcal{Y}|\mathcal{X})^2 + \mathbb{E}[U(x, \mathcal{X})V(y, \mathcal{Y})] + \mathbb{E}[U(x', \mathcal{X})V(y', \mathcal{Y})] \\ & + \mathbb{E}[(U(x, \mathcal{X}) - U(x', \mathcal{X}))(V(y, \mathcal{Y}) - V(y', \mathcal{Y}))]\Big\}. \end{aligned}$$

Moreover, we have

$$\begin{aligned} h_3(w, w', w'') =& \frac{1}{12}\Big\{(2U(x, x') - U(x', x'') - U(x, x''))V(y, y') \\ & + (2U(x, x'') - U(x, x') - U(x', x''))V(y, y'') \\ & + (2U(x', x'') - U(x, x') - U(x, x''))V(y', y'') \\ & + \mathbb{E}[(2U(x, \mathcal{X}) - U(x', \mathcal{X}) - U(x'', \mathcal{X}))V(y, \mathcal{Y})] \\ & + \mathbb{E}[(2U(x', \mathcal{X}) - U(x, \mathcal{X}) - U(x'', \mathcal{X}))V(y', \mathcal{Y})] \\ & + \mathbb{E}[(2U(x'', \mathcal{X}) - U(x, \mathcal{X}) - U(x', \mathcal{X}))V(y'', \mathcal{Y})]\Big\}. \end{aligned}$$

Using similar calculation as that for $h_3(w, w', w'')$, we obtain

$$\begin{aligned} & h_4(w, w', w'', w''') \\ =& \frac{1}{12}\Big\{(2U(x, x') + 2U(x'', x''') - U(x, x'') - U(x, x''') - U(x', x'') - U(x', x'''))(V(y, y') + V(y'', y''')) \\ & + (2U(x, x'') + 2U(x', x''') - U(x, x') - U(x, x''') - U(x'', x') - U(x'', x'''))(V(y, y'') + V(y', y''')) \\ & + (2U(x, x''') + 2U(x'', x') - U(x, x'') - U(x, x') - U(x''', x'') - U(x''', x'))(V(y, y''') + V(y', y''))\Big\}. \end{aligned}$$

Analysis under the null hypothesis When $MDD(\mathcal{Y}|\mathcal{X})^2 = 0$, it can be verified that $h_1(w) =$

0. In this case, we have $h_2(w,w') = U(x,x')V(y,y')/6$ and

$$\begin{aligned} h_3(w,w',w'') =& \frac{1}{12}\Big\{(2U(x,x') - U(x',x'') - U(x,x''))V(y,y') \\ &+ (2U(x,x'') - U(x,x') - U(x',x''))V(y,y'') \\ &+ (2U(x',x'') - U(x,x') - U(x,x''))V(y',y'')\Big\}. \end{aligned}$$

It is not hard to verify that under the null

$$\mathrm{var}(h_2(\mathcal{Z},\mathcal{Z}')) = \frac{1}{36}\mathbb{E}[U^2(\mathcal{X},\mathcal{X}')V^2(\mathcal{Y},\mathcal{Y}')] = \frac{1}{36}\xi^2,$$

and

$$\begin{aligned} \mathrm{var}(h_3(\mathcal{Z},\mathcal{Z}',\mathcal{Z}'')) =& \frac{3}{144}\mathrm{var}\{(2U(\mathcal{X},\mathcal{X}') - U(\mathcal{X}',\mathcal{X}'') - U(\mathcal{X},\mathcal{X}''))V(\mathcal{Y},\mathcal{Y}')\} \\ =& \frac{3}{144}\Big\{4\xi^2 + 2\mathbb{E}[U(\mathcal{X},\mathcal{X}'')^2V(\mathcal{Y},\mathcal{Y}')^2] + 2\mathbb{E}[U(\mathcal{X},\mathcal{X}'')U(\mathcal{X}',\mathcal{X}'')V(\mathcal{Y},\mathcal{Y}')^2]\Big\}, \end{aligned}$$

where $\xi^2 = \mathrm{var}(U(\mathcal{X},\mathcal{X}')V(\mathcal{Y},\mathcal{Y}'))$. Furthermore, careful calculation yields that

$$\begin{aligned} \mathrm{var}(h_4(\mathcal{Z},\mathcal{Z}',\mathcal{Z}'',\mathcal{Z}''')) =& \frac{6}{144}\mathbb{E}\Big\{V(\mathcal{Y},\mathcal{Y}')^2[U(\mathcal{X},\mathcal{X}'') + U(\mathcal{X}',\mathcal{X}''') + U(\mathcal{X}',\mathcal{X}'') \\ &+ U(\mathcal{X},\mathcal{X}''') - 2U(\mathcal{X},\mathcal{X}') - 2U(\mathcal{X}'',\mathcal{X}''')]^2\Big\} \\ =& \frac{1}{6}\{\mathbb{E}[V(\mathcal{Y},\mathcal{Y}')^2U(\mathcal{X},\mathcal{X}'')U(\mathcal{X}',\mathcal{X}'')] + \mathbb{E}[V(\mathcal{Y},\mathcal{Y}')^2U(\mathcal{X},\mathcal{X}'')^2] \\ &+ \mathbb{E}[V(\mathcal{Y},\mathcal{Y}')^2]\mathbb{E}[U(\mathcal{X},\mathcal{X}')^2] + \xi^2\}. \end{aligned}$$

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \mathbb{E}[U(\mathcal{X},\mathcal{X}'')U(\mathcal{X}',\mathcal{X}'')V(\mathcal{Y},\mathcal{Y}')^2] \le& \{\mathbb{E}[U(\mathcal{X},\mathcal{X}'')^2V(\mathcal{Y},\mathcal{Y}')^2]\}^{1/2}\{\mathbb{E}[U(\mathcal{X}',\mathcal{X}'')^2V(\mathcal{Y},\mathcal{Y}')^2]\}^{1/2} \\ =& \mathbb{E}[U(\mathcal{X},\mathcal{X}'')^2V(\mathcal{Y},\mathcal{Y}')^2]. \end{aligned}$$

Under the assumption that

$$\begin{aligned} \frac{\mathbb{E}[U(\mathcal{X},\mathcal{X}'')^2V(\mathcal{Y},\mathcal{Y}')^2]}{\xi^2} &= o(n), \\ \frac{\mathbb{E}[V(\mathcal{Y},\mathcal{Y}')^2]\mathbb{E}[U(\mathcal{X},\mathcal{X}')^2]}{\xi^2} &= o(n^2), \end{aligned}$$

we have

$$MDD_n(\mathcal{Y}|\mathcal{X})^2 = \frac{1}{\binom{n}{2}}\sum_{1\le i<j\le n} U(\mathcal{X}_i,\mathcal{X}_j)V(\mathcal{Y}_i,\mathcal{Y}_j) + \mathcal{R}_n, \tag{2}$$

where $\mathcal{R}_n$ is the remainder term which is asymptotically negligible [see Serfling (1980)].

Analysis under local alternatives We consider the case where $MDD(\mathcal{Y}|\mathcal{X})^2$ is nonzero, i.e., the conditional mean of $\mathcal{Y}$ may depend on $\mathcal{X}$. Recall that $\mathcal{L}(x, y) = \mathbb{E}[U(x, \mathcal{X})V(y, \mathcal{Y})]$. Under the assumption that

$$\text{var}(\mathcal{L}(\mathcal{X}, \mathcal{Y})) = o(n^{-1}\xi^2), \quad \text{var}(\mathcal{L}(\mathcal{X}, \mathcal{Y}')) = o(\xi^2), \tag{3}$$

we get

$$\text{var}(h_1) = o(n^{-1}\xi^2), \quad \text{var}(h_2) = \frac{\xi^2}{36}(1 + o(1)).$$

Moreover, we have

$$\text{var}(h_3(\mathcal{Z}, \mathcal{Z}', \mathcal{Z}'')) \le C\Big\{\xi^2 + \mathbb{E}[U(\mathcal{X}, \mathcal{X}'')^2 V(\mathcal{Y}, \mathcal{Y}')^2]\Big\},$$

and

$$\text{var}(h_4(\mathcal{Z}, \mathcal{Z}', \mathcal{Z}'', \mathcal{Z}''')) \le C'\{\mathbb{E}[V(\mathcal{Y}, \mathcal{Y}')^2 U(\mathcal{X}, \mathcal{X}'')^2] + \mathbb{E}[V(\mathcal{Y}, \mathcal{Y}')^2]\mathbb{E}[U(\mathcal{X}, \mathcal{X}')^2] + \xi^2\}.$$

Thus under Assumption (3),

$$MDD_n(\mathcal{Y}|\mathcal{X})^2 - MDD(\mathcal{Y}|\mathcal{X})^2 = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \{U(\mathcal{X}_i, \mathcal{X}_j)V(\mathcal{Y}_i, \mathcal{Y}_j) - \mathbb{E}[U(\mathcal{X}_i, \mathcal{X}_j)V(\mathcal{Y}_i, \mathcal{Y}_j)]\} + \mathcal{R}_n. \tag{4}$$

Applying the above arguments to $\sum_{k=1}^p MDD_n(Y|x_k)^2$, we deduce that

$$\begin{aligned}
&\sum_{k=1}^p \{MDD_n(Y|x_k)^2 - MDD(Y|x_k)^2\} \\
=&\frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \{\widetilde{U}(X_i, X_j)V(Y_i, Y_j) - \mathbb{E}[\widetilde{U}(X_i, X_j)V(Y_i, Y_j)]\} + \sum_{k=1}^p R_{k,n},
\end{aligned}$$

where the kernel $\widetilde{U}$ is changing with (n, p). The remainder term $\sum_{k=1}^p R_{k,n}$ is asymptotically negligible provided that

$$\begin{aligned}
&\frac{\mathbb{E}[\widetilde{U}(X, X'')^2 V(Y, Y')^2]}{\mathcal{S}^2} = o(n), \\
&\frac{\mathbb{E}[\widetilde{U}(X, X')^2]\mathbb{E}[V(Y, Y')^2]}{\mathcal{S}^2} = o(n^2), \\
&\text{var}(\widetilde{\mathcal{L}}(X, Y)) = o(n^{-1}\mathcal{S}^2), \quad \text{var}(\widetilde{\mathcal{L}}(X, Y')) = o(\mathcal{S}^2),
\end{aligned}$$

where $\widetilde{\mathcal{L}}(x, y) = \mathbb{E}[\widetilde{U}(x, X)V(y, Y)]$.

1.3 Asymptotic normality under the null and alternatives

We shall establish the asymptotic normality for $\breve{T}_n$ using the Central Limit Theorem for martingale difference sequences. We first restrict our analysis under H_0'. Let

$$S_r := \sum_{j=2}^{r}\sum_{i=1}^{j-1}\widetilde{U}(X_i,X_j)V(Y_i,Y_j) = \sum_{j=2}^{r}\sum_{i=1}^{j-1}H(Z_i,Z_j).$$

Define the filtration $\mathcal{F}_r = \sigma\{Z_1, Z_2, \ldots, Z_r\}$ with $Z_i = (X_i, Y_i)$. It is not hard to see that S_r is adaptive to $\mathcal{F}_r$ and S_r is a mean-zero martingale sequence, i.e., $\mathbb{E}[S_r] = 0$ and

$$\mathbb{E}[S_{r'}|\mathcal{F}_r] = S_r + \sum_{j=r+1}^{r'}\sum_{i=1}^{j-1}\mathbb{E}[\mathbb{E}[\widetilde{U}(X_i,X_j)V(Y_i,Y_j)|\mathcal{F}_r, X_i, X_j]|\mathcal{F}_r] = S_r$$

for $r' \geq r$. Thus by verifying the following two conditions [see e.g. Lemmas 2-3 of Chen and Qin (2010)], we can establish the asymptotic normality for $\breve{T}_n$ by Corollary 3.1 of Hall and Heyde (1980). Specifically, define $\mathcal{W}_j = \sum_{i=1}^{j-1} H(Z_i,Z_j)$. We need to show that

$$\sum_{j=1}^{n} B^{-2}\mathbb{E}[\mathcal{W}_j^2\mathbf{I}\{|\mathcal{W}_j| > \epsilon B\}|\mathcal{F}_{j-1}] \to^p 0, \tag{5}$$

for B such that

$$\sum_{j=1}^{n}\mathbb{E}[\mathcal{W}_j^2|\mathcal{F}_{j-1}]/B^2 \to^p C > 0. \tag{6}$$

We shall first prove that

$$\frac{2}{n(n-1)\mathcal{S}^2}\sum_{j=1}^{n}\mathbb{E}[\mathcal{W}_j^2|\mathcal{F}_{j-1}] \to^p 1, \tag{7}$$

i.e., $B^2 = n(n-1)\mathcal{S}^2/2$ and $C = 1$ in (6). Notice that

$$\mathbb{E}[\mathcal{W}_j^2|\mathcal{F}_{j-1}] = \mathbb{E}\left[\sum_{i,k=1}^{j-1} H(Z_i,Z_j)H(Z_k,Z_j)\middle|\mathcal{F}_{j-1}\right] = \sum_{i,k=1}^{j-1} G(Z_i,Z_k),$$

and

$$\begin{aligned}\frac{2}{n(n-1)}\sum_{j=2}^{n}\mathbb{E}[\mathcal{W}_j^2] =& \frac{2}{n(n-1)}\sum_{j=2}^{n}\mathbb{E}\left[\sum_{i,k=1}^{j-1}(Y_i-\mu)(Y_k-\mu)(Y_j-\mu)^2\widetilde{U}(X_i,X_j)\widetilde{U}(X_k,X_j)\right]\\ =& \frac{2}{n(n-1)}\sum_{j=2}^{n}\mathbb{E}\left[\sum_{i=1}^{j-1}(Y_i-\mu)^2(Y_j-\mu)^2\widetilde{U}(X_i,X_j)^2\right]\\ =& \mathbb{E}\left[V(Y,Y')^2\widetilde{U}(X,X')^2\right] = E[H(Z,Z')^2] = \mathcal{S}^2.\end{aligned}$$

Define the following quantities

$$\mathcal{D}_1 = \mathbb{E}[H(Z,Z'')^2 H(Z',Z'')^2] - (\mathbb{E}[H(Z,Z')^2])^2 = \mathrm{var}(G(Z,Z)),$$
$$\mathcal{D}_2 = \mathbb{E}[H(Z,Z')H(Z',Z'')H(Z'',Z''')H(Z''',Z)] = E[G(Z,Z')^2].$$

We have for $j \geq j'$

$$\begin{aligned}\mathrm{cov}\left(\mathbb{E}[\mathcal{W}_j^2|\mathcal{F}_{j-1}], \mathbb{E}[\mathcal{W}_{j'}^2|\mathcal{F}_{j'-1}]\right) =& \sum_{i,k=1}^{j-1}\sum_{i',k'=1}^{j'-1} \mathrm{cov}(G(Z_i,Z_k), G(Z_{i'},Z_{k'})) \\ =&(j'-1)\mathcal{D}_1 + 2(j'-1)(j'-2)\mathcal{D}_2.\end{aligned}$$

Under the assumption that

$$\frac{\mathbb{E}[G(Z,Z')^2]}{\{\mathbb{E}[H(Z,Z')^2]\}^2} \to 0, \quad \frac{\mathbb{E}[H(Z,Z'')^2 H(Z',Z'')^2]}{n\{\mathbb{E}[H(Z,Z')^2]\}^2} \to 0,$$

we have

$$\frac{4}{n^2(n-1)^2}\sum_{j,j'=2}^{n} \mathrm{cov}\left(\mathbb{E}[\mathcal{W}_j^2|\mathcal{F}_{j-1}], \mathbb{E}[\mathcal{W}_{j'}^2|\mathcal{F}_{j'-1}]\right) = O(\mathcal{D}_1/n + \mathcal{D}_2) = o(\mathcal{S}^4),$$

which ensures (7). To show (5), we note that

$$\sum_{j=1}^{n} B^{-2}\mathbb{E}[\mathcal{W}_j^2 \mathbf{I}\{|\mathcal{W}_j| > \epsilon B\}|\mathcal{F}_{j-1}] \leq B^{-2-s}\epsilon^{-s}\sum_{j=1}^{n}\mathbb{E}[|\mathcal{W}_j|^{2+s}|\mathcal{F}_{j-1}], \tag{8}$$

for some $s > 0$. With $s = 2$, we prove that

$$B^{-4}\sum_{j=1}^{n}\mathbb{E}[|\mathcal{W}_j|^4|\mathcal{F}_{j-1}] \to^p 0,$$

where $B^2 = n(n-1)\mathcal{S}^2/2$. To this end, it suffices to show that

$$B^{-4}\sum_{j=1}^{n}\mathbb{E}[|\mathcal{W}_j|^4] \to 0.$$

Under the assumption

$$\frac{\mathbb{E}[H(Z,Z')^4]/n + \mathbb{E}[H(Z,Z'')^2 H(Z',Z'')^2]}{n\{\mathbb{E}[H(Z,Z')^2]\}^2} \to 0,$$

we have

$$\sum_{j=2}^{n}\mathbb{E}[|\mathcal{W}_j|^4] = \sum_{j=2}^{n}\sum_{i_1,i_2,i_3,i_4=1}^{j-1}\mathbb{E}[H(Z_{i_1},Z_j)H(Z_{i_2},Z_j)H(Z_{i_3},Z_j)H(Z_{i_4},Z_j)]$$
$$= \frac{n(n-1)}{2}\mathbb{E}[H(Z,Z')^4] + 3\sum_{j=2}^{n}\sum_{i_1\neq i_2}\mathbb{E}[H(Z_{i_1},Z_j)^2H(Z_{i_2},Z_j)^2]$$
$$= \frac{n(n-1)}{2}\mathbb{E}[H(Z,Z')^4] + 3\sum_{j=2}^{n}(j-1)(j-2)\mathbb{E}[H(Z,Z'')^2H(Z',Z'')^2] = o(B^4),$$

which implies (5).

We next extend the above arguments to local alternatives. Under the assumption that $\text{var}(\widetilde{\mathcal{L}}(X,Y)) = o(n^{-1}\mathcal{S}^2)$, we have

$$\frac{1}{\sqrt{\binom{n}{2}}\mathcal{S}}\sum_{1\le i<j\le n}\Big(\widetilde{U}(X_i,X_j)V(Y_i,Y_j) - \mathbb{E}[\widetilde{U}(X_i,X_j)V(Y_i,Y_j)]\Big)$$
$$= \frac{1}{\sqrt{\binom{n}{2}}\mathcal{S}}\sum_{1\le i<j\le n}H^*(Z_i,Z_j) + \frac{1}{\sqrt{\binom{n}{2}}\mathcal{S}}\sum_{1\le i<j\le n}\Big(\widetilde{\mathcal{L}}(X_i,Y_i)+\widetilde{\mathcal{L}}(X_j,Y_j) - 2\mathbb{E}[\widetilde{U}(X_i,X_j)V(Y_i,Y_j)]\Big)$$
$$= \frac{1}{\sqrt{\binom{n}{2}}\mathcal{S}}\sum_{1\le i<j\le n}H^*(Z_i,Z_j) + o_p(1).$$

Using similar arguments by replacing H with H^*, we can show that

$$\frac{1}{\sqrt{\binom{n}{2}}\mathcal{S}}\sum_{1\le i<j\le n}H^*(Z_i,Z_j) \to^d N(0,1),$$

which implies that

$$\frac{1}{\sqrt{\binom{n}{2}}\mathcal{S}}\sum_{1\le i<j\le n}\Big(\widetilde{U}(X_i,X_j)V(Y_i,Y_j) - \mathbb{E}[\widetilde{U}(X_i,X_j)V(Y_i,Y_j)]\Big) \to^d N(0,1).$$

The proof is thus completed.

1.4 Interpretation of Condition (13) based on Fourier embedding

In this subsection, we provide more discussions on Condition (13) in the paper. Consider a generic random variable $\mathcal{Z} = (\mathcal{X},\mathcal{Y})$ and a random sample $\{\mathcal{Z}_i\}_{i=1}^n$ from the distribution of $\mathcal{Z}$. Let $\imath = \sqrt{-1}$ and $c_q = \pi^{(1+q)/2}/\Gamma((1+q)/2)$. For $a_1, a_2 \in \mathbb{R}^q$, define the inner product $< a_1, a_2 >= a_1^T a_2$. Denote by $\bar{a}$ the conjugate of a complex number a. First note that an alternative expression for MDD is given by

$$MDD(\mathcal{Y}|\mathcal{X})^2 = \frac{1}{c_q}\int_{\mathbb{R}^q}\frac{|\mathbb{E}[\mathcal{Y}e^{\imath<t,\mathcal{X}>}] - \mu_{\mathcal{Y}}f_{\mathcal{X}}(t)|}{|t|_q^{1+q}}dt,$$

where $f_{\mathcal{X}}(t) = \mathbb{E}[e^{\imath<t,\mathcal{X}>}]$ and $\mu_{\mathcal{Y}} = \mathbb{E}[\mathcal{Y}]$. Recall that $U(x,x') = \mathbb{E}[K(x,\mathcal{X}')] + \mathbb{E}[K(\mathcal{X},x')] - K(x,x') - \mathbb{E}[K(\mathcal{X},\mathcal{X}')]$ and $V(y,y') = (y-\mu_{\mathcal{Y}})(y'-\mu_{\mathcal{Y}})$ with $\mu_{\mathcal{Y}} = \mathbb{E}\mathcal{Y}$ and $\mathcal{Z}' = (\mathcal{X}',\mathcal{Y}') \stackrel{D}{=} (\mathcal{X},\mathcal{Y})$. By the analysis in Sections 1.1 and 1.2, we know $MDD(\mathcal{Y}|\mathcal{X})^2$ is a fourth-order U-statistic, whose major term is given by

$$\mathcal{V}_n = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} U(\mathcal{X}_i,\mathcal{X}_j)V(\mathcal{Y}_i,\mathcal{Y}_j). \tag{9}$$

By Lemma 1 of Székely et al. (2007), we have

$$\begin{aligned} U(x,x') &= \int_{\mathbb{R}^q} \mathbb{E}(e^{\imath<t,\mathcal{X}>} - e^{\imath<t,x>})(\overline{e^{\imath<t,\mathcal{X}'>} - e^{\imath<t,x'>}})w_q(t)dt \\ &= \int_{\mathbb{R}^q} (f_{\mathcal{X}}(t) - e^{\imath<t,x>})(\overline{f_{\mathcal{X}}(t)} - e^{-\imath<t,x'>})w_q(t)dt, \end{aligned}$$

where $w_q(t) = 1/(c_q|t|_q^{1+q})$. Let $K(z,z') = U(x,x')V(y,y')$ with $z=(x,y)$ and $z'=(x',y')$. Thus we have

$$\begin{aligned} K(z,z') &= \int_{\mathbb{R}^q} (f_{\mathcal{X}}(t) - e^{\imath<t,x>})(\mu_{\mathcal{Y}} - y)(\overline{f_{\mathcal{X}}(t)} - e^{-\imath<t,x'>})(\mu_{\mathcal{Y}} - y')w_q(t)dt \\ &= \int_{\mathbb{R}^q} G(x,y;t)\overline{G(x',y';t)}w_q(t)dt, \end{aligned}$$

where $G(x,y;t) = (f_{\mathcal{X}}(t) - e^{\imath<t,x>})(\mu_{\mathcal{Y}} - y)$. Note that $\mathbb{E}[G(\mathcal{X},\mathcal{Y};t)] = \mathbb{E}[\mathcal{Y}e^{\imath<t,\mathcal{X}>}] - \mu_{\mathcal{Y}}f_{\mathcal{X}}(t)$, and thus

$$\mathbb{E}[K(\mathcal{Z},\mathcal{Z}')] = \int_{\mathbb{R}^q} |\mathbb{E}[\mathcal{Y}e^{\imath<t,\mathcal{X}>}] - \mu_{\mathcal{Y}}f_{\mathcal{X}}(t)|^2 w_q(t)dt = MDD(\mathcal{Y}|\mathcal{X})^2.$$

Define the operator $A_K g(z) = \mathbb{E}[K(\mathcal{Z},z)g(\mathcal{Z})]$ for $g : \mathbb{R}^{q+1} \to \mathbb{R}$. The eigenfunction h of A_K satisfies that

$$\mathbb{E}[K(\mathcal{Z},z)h(\mathcal{Z})] = \int_{\mathbb{R}^q} \mathbb{E}[G(\mathcal{X},\mathcal{Y};t)h(\mathcal{X},\mathcal{Y})]\overline{G(x,y;t)}w_q(t)dt = \lambda h(x,y),$$

which implies that h has the form of

$$h(x,y) = \int_{\mathbb{R}^q} \eta(t)\overline{G(x,y;t)}w_q(t)dt.$$

Plugging back into the above equation, we obtain,

$$\begin{aligned} &\int_{\mathbb{R}^q} \mathbb{E}[G(\mathcal{X},\mathcal{Y};t)h(\mathcal{X},\mathcal{Y})]\overline{G(x,y;t)}w_q(t)dt \\ =&\int_{\mathbb{R}^q} \left(\int_{\mathbb{R}^q} \eta(t')\mathbb{E}[G(\mathcal{X},\mathcal{Y};t)\overline{G(\mathcal{X},\mathcal{Y};t')}]w_q(t')dt'\right) \overline{G(x,y;t)}w_q(t)dt \\ =&\lambda\int_{\mathbb{R}^q} \eta(t)\overline{G(x,y;t)}w_q(t)dt. \end{aligned}$$

The above equation holds provided that

$$\lambda\eta(t) = \int_{\mathbb{R}^q} \eta(t')\mathbb{E}[G(\mathcal{X},\mathcal{Y};t)\overline{G(\mathcal{X},\mathcal{Y};t')}]w_q(t')dt' := \varphi(\eta)(t), \tag{10}$$

where $\varphi(\cdot)$ is the corresponding operator defined in (10). Thus η is the eigenfunction of φ and λ is the corresponding eigenvalue. When $\mathcal{X}$ and $(\mathcal{Y}-\mu_{\mathcal{Y}})^2$ are independent, we have

$$\mathbb{E}[G(\mathcal{X},\mathcal{Y};t)\overline{G(\mathcal{X},\mathcal{Y};t')}] = \text{var}(\mathcal{Y})(f_{\mathcal{X}}(t-t') - f_{\mathcal{X}}(t)f_{\mathcal{X}}(-t')).$$

Suppose that

$$f_{\mathcal{X}}(t-t') - f_{\mathcal{X}}(t)f_{\mathcal{X}}(-t') = \sum_{j=1}^{+\infty} \lambda_{\mathcal{X},j}\beta_{\mathcal{X},j}(t)\overline{\beta_{\mathcal{X},j}(t')},$$

where $\int_{\mathbb{R}^q} \beta_{\mathcal{X},i}(t)\overline{\beta_{\mathcal{X},j}(t)}w_q(t)dt = \mathbf{1}\{i=j\}$. In this case, Condition (13) reduces to

$$\frac{(\sum_{i=1}^{+\infty}\lambda_{\mathcal{X},i}^t)^{2/t}}{\sum_{i=1}^{+\infty}\lambda_{\mathcal{X},i}^2} \to 0, \tag{11}$$

for some $t > 2$.

Recall that $H(Z_i,Z_k) = V(Y_i,Y_k)\widetilde{U}(X_i,X_k) = V(Y_i,Y_k)\sum_{j=1}^p U_j(x_{ij},x_{kj})$. Below we focus on the case where $K(\cdot,\cdot) = H(\cdot,\cdot)$ and $\mathcal{Z} = Z$. Define $G_j(u_j,y;t) = (f_{x_j}(t) - e^{\imath t u_j})(\mu - y)$ for $u_j, y, t \in \mathbb{R}$. Then it can be shown that

$$H(z,z') = \int_{\mathbb{R}} \sum_{j=1}^p G_j(u_j,y;t)\overline{G_j(u_j',y';t)}w_1(t)dt,$$

where $z = (u_1,\ldots,u_p,y)^T$ and $z' = (u_1',\ldots,u_p',y')^T$. The eigenfunction of A_H satisfies that

$$\mathbb{E}[H(Z,z)h(Z)] = \int_{\mathbb{R}} \sum_{j=1}^p \mathbb{E}[G_j(x_j,Y;t)h(Z)]\overline{G_j(u_j,y;t)}w_1(t)dt = \lambda h(z).$$

Thus we must have $h(z) = \int_{\mathbb{R}}\sum_{j=1}^p \eta_j(t)\overline{G_j(u_j,y;t)}w_1(t)dt$. It implies that

$$\begin{aligned}&\int_{\mathbb{R}} \sum_{j=1}^p \left(\int_{\mathbb{R}} \sum_{i=1}^p \eta_i(t')\mathbb{E}[G_j(x_j,Y;t)\overline{G_i(x_i,Y;t')}]w_1(t')dt'\right)\overline{G_j(u_j,y;t)}w_1(t)dt \\ =&\lambda\int_{\mathbb{R}}\sum_{j=1}^p \eta_j(t)\overline{G_j(u_j,y;t)}w_1(t)dt.\end{aligned} \tag{12}$$

Let $\varphi_{ji}(g)(t) = \int_{\mathbb{R}} g(t')\mathbb{E}[G_j(x_j,Y;t)\overline{G_i(x_i,Y;t')}]w_1(t')dt'$. When $X = (x_1,\ldots,x_p)^T$ and $(Y-\mu)^2$ are independent, we have

$$\mathbb{E}[G_j(x_j,Y;t)\overline{G_i(x_i,Y;t')}] = \text{var}(Y)(f_{x_j,x_i}(t,-t') - f_{x_j}(t)f_{x_i}(-t')),$$

in which case $\varphi_{ji}(g)(t) = \text{var}(Y) \int_{\mathbb{R}} g(t')(f_{x_j,x_i}(t,-t') - f_{x_j}(t)f_{x_i}(-t'))w_1(t')dt'$. Notice that (12) holds provided that

$$\sum_{i=1}^{p} \varphi_{ji}(\eta_i) = \lambda \eta_j, \tag{13}$$

for all $1 \leq j \leq p$. Let $L^2(w_1)$ be the space of functions such that $\int_{\mathbb{R}} f(t)^2 w_1(t)dt < \infty$ for $f \in L^2(w_1)$. For $g = (g_1, \ldots, g_p)^T$ with $g_j \in L^2(w_1)$, define the operator

$$\Phi(g) = \left(\sum_{j=1}^{p} \varphi_{1j}(g_j), \ldots, \sum_{j=1}^{p} \varphi_{pj}(g_j) \right).$$

Therefore, λ is the eigenvalue associated with Φ. Define $Tr(\cdot)$ as the nuclear norm for a Hermitian operator. Then with $t = 4$, the condition needed to ensure the asymptotic normality [Hall (1984)] is

$$\frac{Tr^{1/2}(\Phi^4)}{Tr(\Phi^2)} \to 0.$$

Remark **1.1**. We remark that the above analysis can be extended to a more general class of dependence measures such as the distance covariance in Székely et al. (2007).

1.5 Conditions in Theorem 2.1

We study the following conditions imposed in Theorem 2.1 [see (8), (10) and (11) therein],

$$\frac{\mathbb{E}[G(Z,Z')^2]}{\mathcal{S}^4} \to 0, \tag{14}$$

$$\frac{\mathbb{E}[H(Z,Z')^4]}{n\mathcal{S}^4} \to 0, \tag{15}$$

$$\frac{\mathbb{E}[\widetilde{U}(X,X'')^2 V(Y,Y')^2]}{\mathcal{S}^2} = o(n), \tag{16}$$

$$\frac{\mathbb{E}[\widetilde{U}(X,X')^2]\mathbb{E}[V(Y,Y')^2]}{\mathcal{S}^2} = o(n^2). \tag{17}$$

Note that (15) is slightly stronger than the second part of (8) in the paper. To proceed, we summarize the general conditions we shall study in this subsection:

$$0 < c \leq \text{var}(Y|X) = E[(Y - E[Y|X])^2|X] \leq \mathbb{E}[(Y - \mathbb{E}[Y|X])^4|X]^{1/2} \leq C < +\infty, \tag{18}$$

$$\frac{\mathbb{E}[\widetilde{U}(X,X')\widetilde{U}(X',X'')\widetilde{U}(X'',X''')\widetilde{U}(X''',X)]}{(\sum_{j,j'=1}^{p} dcov(x_j,x_{j'})^2)^2} \to 0, \tag{19}$$

$$\frac{\mathbb{E}[\widetilde{U}(X,X')^4]}{n(\sum_{j,j'=1}^{p} dcov(x_j,x_{j'})^2)^2} \to 0, \tag{20}$$

where (18) holds almost surely for some constants c and C.

Under (18) and the null hypothesis (i.e. $E[Y|X] = \mu$),

$$\begin{aligned}\mathcal{S}^2 = \mathbb{E}[H(Z,Z')^2] =& \mathbb{E}[\mathbb{E}[(Y-\mu)^2(Y'-\mu)^2\widetilde{U}(X,X')^2|X,X']] \\ \geq & c^2\mathbb{E}[\widetilde{U}(X,X')^2] = c^2\sum_{j,j'=1}^{p} dcov(x_j,x_{j'})^2.\end{aligned}$$

Also note that

$$\begin{aligned}\mathbb{E}[\widetilde{U}(X,X'')^2V(Y,Y')^2] \leq & \mathbb{E}[\mathbb{E}[\widetilde{U}(X,X'')^2V(Y,Y')^2|X,X']] \\ \leq & C^2\mathbb{E}[\widetilde{U}(X,X'')^2] \leq C^2\sum_{j,j'=1}^{p} dcov(x_j,x_{j'})^2.\end{aligned}$$

Thus (16) and (17) hold under (18) and the null hypothesis. Using the conditioning argument, we have

$$\begin{aligned}&\mathbb{E}[G(Z,Z')^2] \leq C^4\mathbb{E}[\widetilde{U}(X,X'')\widetilde{U}(X',X'')\widetilde{U}(X,X''')\widetilde{U}(X',X''')], \\ &\mathbb{E}[H(Z,Z')^4] \leq C^4\mathbb{E}[\widetilde{U}(X,X')^4].\end{aligned}$$

Thus (19)-(20) imply (14)-(15) under (18). To sum up, (18)-(20) together imply (14)-(17).

Using the fact that

$$U_j(u_j,u_j') = \int_{\mathbb{R}}(f_j(t)-e^{\imath t u_j})(f_j(-t)-e^{-\imath t u_j'})w(t)dt, \quad \imath = \sqrt{-1},$$

we have

$$\begin{aligned}&\mathbb{E}[\widetilde{U}(X,X'')\widetilde{U}(X',X'')\widetilde{U}(X,X''')\widetilde{U}(X',X''')] \\ =&\sum_{j,k,l,m=1}^{p}\int\{f_{jl}(t_1,t_2)-f_j(t_1)f_l(t_2)\}\{f_{lm}(-t_2,-t_4)-f_l(-t_2)f_m(-t_4)\} \\ &\{f_{mk}(t_4,t_3)-f_m(t_4)f_k(t_3)\}\{f_{kj}(-t_3,-t_1)-f_k(-t_3)f_j(-t_1)\} \\ &w(t_1)w(t_2)w(t_3)w(t_4)dt_1dt_2dt_3dt_4,\end{aligned}$$

where f_{jk} and f_j denote the (joint) characteristic functions for (X_j, X_k) and X_j respectively. Similarly, we have

$$\begin{aligned}&\mathbb{E}[\widetilde{U}(X,X')^4] \\ =&\sum_{j,k,l,m=1}^{p}\int\left|\mathbb{E}[\{f_j(t_1)-e^{\imath t_1X_j}\}\{f_k(t_2)-e^{\imath t_2X_k}\}\{f_l(t_3)-e^{\imath t_3X_l}\}\{f_m(t_4)-e^{\imath t_4X_m}\}]\right|^2 \\ &w(t_1)w(t_2)w(t_3)w(t_4)dt_1dt_2dt_3dt_4,\end{aligned}$$

where

$$
\begin{aligned}
&\mathbb{E}[\{f_j(t_1) - e^{\imath t_1 X_j}\}\{f_k(t_2) - e^{\imath t_2 X_k}\}\{f_l(t_3) - e^{\imath t_3 X_l}\}\{f_m(t_4) - e^{\imath t_4 X_m}\}] \\
=&\mathbb{E}[\{f_j(t_1)f_k(t_2) - e^{\imath t_1 X_j} f_k(t_2) - f_j(t_1)e^{\imath t_2 X_k} + e^{\imath t_1 X_j + \imath t_2 X_k}\} \\
&\{f_l(t_3)f_m(t_4) - e^{\imath t_3 X_l} f_m(t_4) - f_l(t_3)e^{\imath t_4 X_m} + e^{\imath t_3 X_l + \imath t_4 X_m}\}] \\
=&-3f_j(t_1)f_k(t_2)f_l(t_3)f_m(t_4) + f_{jl}(t_1,t_3)f_k(t_2)f_m(t_4) + f_{jk}(t_1,t_2)f_l(t_3)f_m(t_4) \\
&+ f_{jm}(t_1,t_4)f_k(t_2)f_l(t_3) + f_{km}(t_2,t_4)f_j(t_1)f_l(t_3) + f_{kl}(t_2,t_3)f_j(t_1)f_m(t_4) \\
&+ f_{lm}(t_3,t_4)f_j(t_1)f_k(t_2) - f_{jlm}(t_1,t_3,t_4)f_k(t_2) - f_{klm}(t_2,t_3,t_4)f_j(t_1) - f_{jkl}(t_1,t_2,t_3)f_m(t_4) \\
&- f_{jkm}(t_1,t_2,t_4)f_l(t_3) + f_{jklm}(t_1,t_2,t_3,t_4).
\end{aligned}
\tag{21}
$$

Hence (19)-(20) can be translated into conditions on the joint characteristic functions.

Banded dependence structure: For the ease of notation, we assume that $\pi(i) = i$. The result can be extended to the case where $\pi(\cdot)$ is an arbitrary permutation of $\{1, 2 \dots, p\}$. In this case, we know that X_i and X_j are independent provided that $|i - j| > L$. By Lemma 1.1 below, we have

$$
\begin{aligned}
&\mathbb{E}[\widetilde{U}(X,X')\widetilde{U}(X',X'')\widetilde{U}(X'',X''')\widetilde{U}(X''',X)] \\
=&\sum_{j=1}^{p}\sum_{k=j-L}^{j+L}\sum_{l=k-L}^{k+L}\sum_{m=j-L}^{j+L} \mathbb{E}[U_j(x_j,x_j')U_k(x_k',x_k'')U_l(x_l'',x_l''')U_m(x_m''',x_m)] \\
\leq& C\sum_{j=1}^{p}\sum_{k=j-L}^{j+L}\sum_{l=k-L}^{k+L}\sum_{m=j-L}^{j+L} \mathbb{E}[|x_j-\mu_j|]\mathbb{E}[|x_k-\mu_k|]\mathbb{E}[|x_l-\mu_l|]\mathbb{E}[|x_m-\mu_m|] \\
\leq& C'p(L+1)^3 \max_{1\leq j\leq p}(E[|x_j-\mu_j|])^4.
\end{aligned}
$$

By the Hölder's inequality and Lemma 1.1, we obtain

$$
\begin{aligned}
\mathbb{E}[\widetilde{U}(X,X')^4] \leq& C\sum_{j=1}^{p}\sum_{k,l,m=j-3L}^{j+3L} |\mathbb{E}[U_j(x_j,x_j')U_k(x_k,x_k')U_l(x_l,x_l')U_m(x_m,x_m')]| \\
&+ C'\left\{\sum_{|j-k|\leq L} dcov(x_j,x_k)^2\right\}^2 \\
\leq& Cp(6L+1)^3 \max_{1\leq j\leq p}\mathbb{E}[U_j(x_j,x_j')^4] \\
&+ C'\left\{\sum_{|j-k|\leq L} dcov(x_j,x_k)^2\right\}^2 \\
\leq& C''\left[p(L+1)^3 \max_{1\leq j\leq p}\mathrm{var}(x_j)^2 + \left\{\sum_{|j-k|\leq L} dcov(x_j,x_k)^2\right\}^2\right].
\end{aligned}
$$

Therefore, (19) and (20) are implied by

$$\frac{p(L+1)^3 \max\left\{n^{-1}\max_{1\le j\le p}\operatorname{var}(x_j)^2, \max_{1\le j\le p}(\mathbb{E}[|x_j-\mu_j|])^4\right\}}{(\sum_{|j-k|\le L} dcov(x_j,x_k)^2)^2} \to 0. \tag{22}$$

In particular, if $L = o(p^{1/3})$ and $\max_{1\le j\le p}\operatorname{var}(x_j)/\min_{1\le j\le p} dcov(x_j,x_j)^2$ is bounded from above, (22) is satisfied.

Lemma **1.1**. *We have*

$$\mathbb{E}[U_j(x_j,x_j')^4] \le C var(x_j)^2,$$
$$|\mathbb{E}[U_j(x_j,x_j')U_k(x_k',x_k'')U_l(x_l'',x_l''')U_m(x_m''',x_m)]| \le C'\mathbb{E}[|x_j-\mu_j|]\mathbb{E}[|x_k-\mu_k|]\mathbb{E}[|x_l-\mu_l|]\mathbb{E}[|x_m-\mu_m|],$$

for some constant $C, C' > 0$.

Proof. By the triangle inequality, we have

$$\left|\mathbb{E}[K(x_j,x')]-K(x,x')\right| \le \mathbb{E}[K(x,x_j')]$$

for $x, x' \in \mathbb{R}$. Thus we have

$$|U_j(x,x')| \le \max\{|\mathbb{E}[K(x_j,x_j')]-2\mathbb{E}[K(x,x_j')]|, \mathbb{E}[K(x_j,x_j')]\} := a_j(x),$$

which implies that

$$\mathbb{E}[U_j(x_j,x_j')^4] \le \mathbb{E}[a_j(x_j)^2]\mathbb{E}[a_j(x_j')^2]$$

as $a_j(x_j)$ and $a_j(x_j')$ are independent. Some simple algebra shows that

$$\mathbb{E}[a_j(x_j)^2] \le C\operatorname{var}(x_j)^2.$$

Therefore, we have $\mathbb{E}[U_j(x_j,x_j')^4] \le C^2\operatorname{var}(x_j)^2$. Similarly, we get

$$\begin{aligned}&|\mathbb{E}[U_j(x_j,x_j')U_k(x_k',x_k'')U_l(x_l'',x_l''')U_m(x_m''',x_m)]|\\ \le&\mathbb{E}[a_j(x_j)]\mathbb{E}[a_k(x_k')]\mathbb{E}[a_l(x_l'')]\mathbb{E}[a_m(x_m''')]\\ \le&C\mathbb{E}[|x_j-\mu_j|]\mathbb{E}[|x_k-\mu_k|]\mathbb{E}[|x_l-\mu_l|]\mathbb{E}[|x_m-\mu_m|].\end{aligned}$$

◇

Conditions under Gaussianity: Under the Gaussian assumption, we verify (19) and (20) in the following two steps.

Step 1: We have

$$
\begin{aligned}
&V(\sigma_{jl}, \sigma_{lm}, \sigma_{mk}, \sigma_{kj}) \\
:=&\mathbb{E}[U_j(x_j, x_j'')U_k(x_k', x_k'')U_l(x_l, x_l''')U_m(x_m', x_m''')] \\
=&\int \{f_{jl}(t_1, t_2) - f_j(t_1)f_l(t_2)\}\{f_{lm}(-t_2, -t_4) - f_l(-t_2)f_m(-t_4)\} \\
&\{f_{mk}(t_4, t_3) - f_m(t_4)f_k(t_3)\}\{f_{kj}(-t_3, -t_1) - f_k(-t_3)f_j(-t_1)\} \\
&w(t_1)w(t_2)w(t_3)w(t_4)dt_1dt_2dt_3dt_4 \\
=&\int e^{-(t_1^2+t_2^2+t_3^2+t_4^2)}(1 - e^{-\sigma_{jl}t_1t_2})(1 - e^{-\sigma_{lm}t_2t_4})(1 - e^{-\sigma_{mk}t_3t_4})(1 - e^{-\sigma_{kj}t_1t_3}) \\
&w(t_1)w(t_2)w(t_3)w(t_4)dt_1dt_2dt_3dt_4.
\end{aligned}
$$

Using the power series expansion, we have

$$
\begin{aligned}
&(1 - e^{-\sigma_{jl}t_1t_2})(1 - e^{-\sigma_{lm}t_2t_4})(1 - e^{-\sigma_{mk}t_3t_4})(1 - e^{-\sigma_{kj}t_1t_3}) \\
=&\sum_{i_1,i_2,i_3,i_4=1}^{\infty} \frac{(-1)^{i_1+i_2+i_3+i_4}}{i_1!i_2!i_3!i_4!}\sigma_{jl}^{i_1}\sigma_{lm}^{i_2}\sigma_{mk}^{i_3}\sigma_{kj}^{i_4}(t_1t_2)^{i_1}(t_2t_4)^{i_2}(t_3t_4)^{i_3}(t_1t_3)^{i_4},
\end{aligned}
$$

which implies that

$$
\begin{aligned}
&V(\sigma_{jl}, \sigma_{lm}, \sigma_{mk}, \sigma_{kj}) \\
=&C_1\int \sum_{i_1,i_2,i_3,i_4=1}^{\infty} \frac{(-1)^{i_1+i_2+i_3+i_4}}{i_1!i_2!i_3!i_4!}\sigma_{jl}^{i_1}\sigma_{lm}^{i_2}\sigma_{mk}^{i_3}\sigma_{kj}^{i_4}e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} \\
&(t_1t_2)^{i_1-1}(t_2t_4)^{i_2-1}(t_3t_4)^{i_3-1}(t_1t_3)^{i_4-1}dt_1dt_2dt_3dt_4 \\
=&C_1\sigma_{jl}\sigma_{lm}\sigma_{mk}\sigma_{kj}\sum_{i_1,i_2,i_3,i_4=0}^{\infty} \frac{(-1)^{i_1+i_2+i_3+i_4}}{(i_1+1)!(i_2+1)!(i_3+1)!(i_4+1)!}\sigma_{jl}^{i_1}\sigma_{lm}^{i_2}\sigma_{mk}^{i_3}\sigma_{kj}^{i_4}\int e^{-(t_1^2+t_2^2+t_3^2+t_4^2)} \\
&(t_1t_2)^{i_1}(t_2t_4)^{i_2}(t_3t_4)^{i_3}(t_1t_3)^{i_4}dt_1dt_2dt_3dt_4.
\end{aligned}
$$

Note that $\int e^{-(t_1^2+t_2^2+t_3^2+t_4^2)}(t_1t_2)^{i_1}(t_2t_4)^{i_2}(t_3t_4)^{i_3}(t_1t_3)^{i_4}dt_1dt_2dt_3dt_4$ is nonzero if and only if i_1, i_2, i_3, i_4 are all even or odd numbers simultaneously. Thus using the fact that $|\sigma_{kl}| \leq 1$, we have

$$
\begin{aligned}
&V(\sigma_{jl},\sigma_{lm},\sigma_{mk},\sigma_{kj})\\
=&C_1\sigma_{jl}\sigma_{lm}\sigma_{mk}\sigma_{kj}\sum_{i_1,i_2,i_3,i_4\text{ are all even or odd}}\frac{1}{(i_1+1)!(i_2+1)!(i_3+1)!(i_4+1)!}\sigma_{jl}^{i_1}\sigma_{lm}^{i_2}\sigma_{mk}^{i_3}\sigma_{kj}^{i_4}\\
&\int e^{-(t_1^2+t_2^2+t_3^2+t_4^2)}(t_1t_2)^{i_1}(t_2t_4)^{i_2}(t_3t_4)^{i_3}(t_1t_3)^{i_4}dt_1dt_2dt_3dt_4\\
\leq&C_1|\sigma_{jl}\sigma_{lm}\sigma_{mk}\sigma_{kj}|\sum_{i_1,i_2,i_3,i_4\text{ are all even or odd}}\frac{1}{(i_1+1)!(i_2+1)!(i_3+1)!(i_4+1)!}\int e^{-(t_1^2+t_2^2+t_3^2+t_4^2)}\\
&(t_1t_2)^{i_1}(t_2t_4)^{i_2}(t_3t_4)^{i_3}(t_1t_3)^{i_4}dt_1dt_2dt_3dt_4\\
=&|\sigma_{jl}\sigma_{lm}\sigma_{mk}\sigma_{kj}|V(1,1,1,1),
\end{aligned}
$$

where $V(1,1,1,1)=\mathbb{E}[U_j(x_j,x_j'')U_k(x_j',x_j'')U_l(x_j,x_j''')U_m(x_j',x_j''')]\leq\mathbb{E}[U_j(x_j,x_j')^4]<\infty$ with $x_j,x_j',x_j'',x_j'''\sim^{i.i.d}N(0,1)$. On the other hand, in view of the proof of Theorem 7 of Székely et al. (2007), we have

$$
dcov(x_j,x_{j'})^2=\frac{\sigma_{j,j'}^2}{\pi^2}\sum_{k=1}^{+\infty}\frac{2^{2k}-2}{(2k)!}\sigma_{j,j'}^{2(k-1)}\int_{\mathbb{R}^2}e^{-t^2-s^2}(ts)^{2(k-1)}dtds\geq\frac{\sigma_{j,j'}^2}{\pi},
$$

which implies that $\sum_{j,j'=1}^p dcov(x_j,x_{j'})^2\geq\pi^{-1}\sum_{j,j'=1}^p\sigma_{j,j'}^2=\pi^{-1}\mathrm{Tr}(\Sigma^2)$. Thus Condition (19) is implied by the assumption $\frac{\sum_{j,k,l,m=1}^p|\sigma_{jk}\sigma_{kl}\sigma_{lm}\sigma_{mj}|}{\mathrm{Tr}^2(\Sigma^2)}\to 0$.

Step 2: Using (21) and the power series expansion, we deduce after laborious calculations that

$$
\begin{aligned}
&\mathbb{E}[U_j(x_j,x_j')U_k(x_k,x_k')U_l(x_l,x_l')U_m(x_m,x_m')]\\
=&\int e^{-(t_1^2+t_2^2+t_3^2+t_4^2)}U_{jklm}(t_1,t_2,t_3,t_4)^2w(t_1)w(t_2)w(t_3)w(t_4)dt_1dt_2dt_3dt_4,
\end{aligned}
$$

where

$$
U_{jklm}(t_1,t_2,t_3,t_4)=\sum_{u=2}^{+\infty}\sum_{\mathbf{s}}\frac{(-1)^u}{\mathbf{s}!}t_1^{s_1+s_2+s_3}t_2^{s_1+s_4+s_5}t_3^{s_2+s_4+s_6}t_4^{s_3+s_5+s_6}\sigma_{jk}^{s_1}\sigma_{jl}^{s_2}\sigma_{jm}^{s_3}\sigma_{kl}^{s_4}\sigma_{km}^{s_5}\sigma_{lm}^{s_6}
$$

with $\sum_{\mathbf{s}}$ denoting the summation over all $\mathbf{s}=(s_1,s_2,s_3,s_4,s_5,s_6)$ such that $s_1+s_2+s_3+s_4+s_5+s_6=u, s_1+s_2+s_3\geq 1, s_1+s_4+s_5\geq 1, s_2+s_4+s_6\geq 1$ and $s_3+s_5+s_6\geq 1$, and

$\mathbf{s}! = \prod_{i=1}^6 s_i!$. Therefore

$$\begin{aligned}
&\mathbb{E}[U_j(x_j,x_j')U_k(x_k,x_k')U_l(x_l,x_l')U_m(x_m,x_m')]\\
=&\sum_{u,v=2}^{+\infty}\sum_{\mathbf{s}}\sum_{\mathbf{r}}\frac{(-1)^{u+v}}{\mathbf{s}!\mathbf{r}!}\sigma_{jk}^{s_1+r_1}\sigma_{jl}^{s_2+r_2}\sigma_{jm}^{s_3+r_3}\sigma_{kl}^{s_4+r_4}\sigma_{km}^{s_5+r_5}\sigma_{lm}^{s_6+r_6}\int e^{-(t_1^2+t_2^2+t_3^2+t_4^2)}\\
&t_1^{s_1+s_2+s_3+r_1+r_2+r_3}t_2^{s_1+s_4+s_5+r_1+r_4+r_5}t_3^{s_2+s_4+s_6+r_2+r_4+r_6}t_4^{s_3+s_5+s_6+r_3+r_5+r_6}\\
&w(t_1)w(t_2)w(t_3)w(t_4)dt_1dt_2dt_3dt_4\\
=&\sum_{h=4}^{+\infty}(-1)^h\sum_{u+v=h;u,v\geq 2}\sum_{\mathbf{s}}\sum_{\mathbf{v}}\frac{1}{\mathbf{s}!\mathbf{r}!}\sigma_{jk}^{s_1+r_1}\sigma_{jl}^{s_2+r_2}\sigma_{jm}^{s_3+r_3}\sigma_{kl}^{s_4+r_4}\sigma_{km}^{s_5+r_5}\sigma_{lm}^{s_6+r_6}a_{\mathbf{s},\mathbf{r}},
\end{aligned}$$

where

$$\begin{aligned}
a_{\mathbf{s},\mathbf{r}}=&\int e^{-(t_1^2+t_2^2+t_3^2+t_4^2)}t_1^{s_1+s_2+s_3+r_1+r_2+r_3}t_2^{s_1+s_4+s_5+r_1+r_4+r_5}t_3^{s_2+s_4+s_6+r_2+r_4+r_6}t_4^{s_3+s_5+s_6+r_3+r_5+r_6}\\
&w(t_1)w(t_2)w(t_3)w(t_4)dt_1dt_2dt_3dt_4,
\end{aligned}$$

and $s_1+s_2+s_3+r_1+r_2+r_3$, $s_1+s_4+s_5+r_1+r_4+r_5$, $s_2+s_4+s_6+r_2+r_4+r_6$, $s_3+s_5+s_6+r_3+r_5+r_6$ are all even numbers. Under the above constraint, the integration inside the curly bracket is always nonnegative. For $h\geq 4$ and $\mathbf{s},\mathbf{r}$ satisfying the above constraint, the term $|\sigma_{jk}^{s_1+r_1}\sigma_{jl}^{s_2+r_2}\sigma_{jm}^{s_3+r_3}\sigma_{kl}^{s_4+r_4}\sigma_{km}^{s_5+r_5}|$ is bounded by one of the following terms

$$\begin{aligned}
&\sigma_{jk}^2\sigma_{lm}^2,\sigma_{jl}^2\sigma_{km}^2,\sigma_{jm}^2\sigma_{kl}^2,\sigma_{jk}^2\sigma_{jl}^2\sigma_{jm}^2,\sigma_{kj}^2\sigma_{kl}^2\sigma_{km}^2,\sigma_{lj}^2\sigma_{lk}^2\sigma_{lm}^2,\sigma_{mj}^2\sigma_{mk}^2\sigma_{ml}^2,\\
&|\sigma_{jk}\sigma_{kl}\sigma_{lm}\sigma_{mj}|,|\sigma_{jk}\sigma_{km}\sigma_{ml}\sigma_{lj}|,|\sigma_{jl}\sigma_{lk}\sigma_{km}\sigma_{mj}|.
\end{aligned}$$

On the other hand, as $\mathbb{E}[U_j(x_j,x_j')^4]=\sum_{h=4}^{+\infty}(-1)^h\sum_{u+v=h;u,v\geq 2}\sum_{\mathbf{s}}\sum_{\mathbf{v}}a_{\mathbf{s},\mathbf{r}}/(\mathbf{s}!\mathbf{r}!)<\infty$, we must have $\sum_{u+v=h;u,v\geq 2}\sum_{\mathbf{s}}\sum_{\mathbf{v}}a_{\mathbf{s},\mathbf{r}}/(\mathbf{s}!\mathbf{r}!)\to 0$ as $k\to+\infty$. Thus we get

$$\begin{aligned}
&\mathbb{E}[U_j(x_j,x_j')U_k(x_k,x_k')U_l(x_l,x_l')U_m(x_m,x_m')]\\
\leq&(\sigma_{jk}^2\sigma_{lm}^2+\sigma_{jl}^2\sigma_{km}^2+\sigma_{jm}^2\sigma_{kl}^2+|\sigma_{jk}\sigma_{kl}\sigma_{lm}\sigma_{mj}|+|\sigma_{jk}\sigma_{km}\sigma_{ml}\sigma_{lj}|+|\sigma_{jl}\sigma_{lk}\sigma_{km}\sigma_{mj}|)\\
&\times\left(\sum_{h=4}^{+\infty}c^{h-4}\sum_{u+v=h;u,v\geq 2}\sum_{\mathbf{s}}\sum_{\mathbf{v}}\frac{1}{\mathbf{s}!\mathbf{r}!}a_{\mathbf{s},\mathbf{r}}\right)\\
&+(\sigma_{jk}^2\sigma_{jl}^2\sigma_{jm}^2+\sigma_{kj}^2\sigma_{kl}^2\sigma_{km}^2+\sigma_{lj}^2\sigma_{lk}^2\sigma_{lm}^2+\sigma_{mj}^2\sigma_{mk}^2\sigma_{ml}^2)\\
&\times\left(\sum_{h=6}^{+\infty}c^{h-6}\sum_{u+v=h;u,v\geq 2}\sum_{\mathbf{s}}\sum_{\mathbf{v}}\frac{1}{\mathbf{s}!\mathbf{r}!}a_{\mathbf{s},\mathbf{r}}\right)\\
\leq&C_1(\sigma_{jk}^2\sigma_{lm}^2+\sigma_{jl}^2\sigma_{km}^2+\sigma_{jm}^2\sigma_{kl}^2+|\sigma_{jk}\sigma_{kl}\sigma_{lm}\sigma_{mj}|+|\sigma_{jk}\sigma_{km}\sigma_{ml}\sigma_{lj}|+|\sigma_{jl}\sigma_{lk}\sigma_{km}\sigma_{mj}|\\
&+\sigma_{jk}^2\sigma_{jl}^2\sigma_{jm}^2+\sigma_{kj}^2\sigma_{kl}^2\sigma_{km}^2+\sigma_{lj}^2\sigma_{lk}^2\sigma_{lm}^2+\sigma_{mj}^2\sigma_{mk}^2\sigma_{ml}^2).
\end{aligned}$$

It follows that

$$
\begin{aligned}
&\mathbb{E}[\widetilde{U}(X,X')^4] \\
\leq &C_1 \sum_{j,k,l,m=1}^{p} (\sigma_{jk}^2\sigma_{lm}^2 + \sigma_{jl}^2\sigma_{km}^2 + \sigma_{jm}^2\sigma_{kl}^2 + |\sigma_{jk}\sigma_{kl}\sigma_{lm}\sigma_{mj}| + |\sigma_{jk}\sigma_{km}\sigma_{ml}\sigma_{lj}| + |\sigma_{jl}\sigma_{lk}\sigma_{km}\sigma_{mj}| \\
&+ \sigma_{jk}^2\sigma_{jl}^2\sigma_{jm}^2 + \sigma_{kj}^2\sigma_{kl}^2\sigma_{km}^2 + \sigma_{lj}^2\sigma_{lk}^2\sigma_{lm}^2 + \sigma_{mj}^2\sigma_{mk}^2\sigma_{ml}^2) \\
\leq &C_2 \left(\mathrm{Tr}^2(\Sigma^2) + \mathrm{Tr}(\breve{\Sigma}^4) + \sum_{j=1}^{p}(\sum_{k=1}^{p}\sigma_{jk}^2)^3 \right),
\end{aligned}
$$

where $\breve{\Sigma} = (|\sigma_{jk}|)$. Therefore, (20) is satisfied under (15) in the paper.

1.6 Consistency of wild Bootstrap and variance estimator

In this subsection, we prove Theorem 2.2 (variance ratio consistency) and Theorem 2.3 (the bootstrap consistency) under Conditions (8), (10) and (11) in the main paper. In particular, the variance ratio consistency is an intermediate step in the proof of bootstrap consistency, see the details in Step 6 below.

Define $d_{kl} = \sum_{j=1}^{p} \widetilde{A}_{kl}(j)\widetilde{B}_{kl}$ and note that $MDD_n^*(Y|x_j)^2 = \frac{2}{n(n-1)}\sum_{k<l} d_{kl}e_k e_l$. Denote by $\mathbb{E}^*$ and $\mathrm{cov}^*/\mathrm{var}^*$ the expectation and covariance/variance in the bootstrap world. Then we have $\mathbb{E}^*[MDD_n^*(Y|x_j)^2] = 0$ for any $1 \leq j \leq p$, and

$$
\mathrm{var}^*\left(\sqrt{\binom{n}{2}}\sum_{j=1}^{p} MDD_n^*(Y|x_j)^2\right) = \frac{2}{n(n-1)}\sum_{k<l} d_{kl}^2 = \hat{\mathcal{S}}^2.
$$

Define the event

$$
\mathcal{A} = \left\{ \frac{\sum_{j\geq j'}\left(\sum_{i=1}^{j'-1} d_{ij}d_{ij'}\right)^2}{(\sum_{k<l} d_{kl}^2)^2} \to 0 \right\}. \tag{23}
$$

Our proof involves the following four steps:

1. Conditional on $\mathcal{A}$, we show
$$
\sqrt{\binom{n}{2}}\frac{\sum_{j=1}^{p} MDD_n^*(Y|x_j)^2}{\hat{\mathcal{S}}} \to^{d^*} N(0,1),
$$
where $\to^{d^*}$ denotes convergence in distribution with respect to $\{e_i\}$.

2. Conditional on $\mathcal{A}$, we show
$$
\frac{\hat{\mathcal{S}}^*}{\hat{\mathcal{S}}} \to^{p^*} 1,
$$
where $\to^{p^*}$ denotes convergence in probability with respect to $\{e_i\}$.

3. Under (18)-(20), we prove that

$$\frac{\sum_{j\geq j'}\left(\sum_{i=1}^{j'-1} d_{ij}d_{ij'}\right)^2}{(\sum_{k<l} d_{kl}^2)^2} \to^p 0.$$

4. Finally, combining the above results, we obtain the convergence in distribution in probability, i.e.,

$$\sqrt{\binom{n}{2}}\frac{\sum_{j=1}^p MDD_n^*(Y|x_j)^2}{\hat{S}^*} \to^{\mathcal{D}^*} N(0,1).$$

For clarity, we present the proofs in the following 6 steps.

Step 1: The basic idea is to apply the martingale CLT to the quadratic form $\sum_{j=1}^p MDD_n^*(Y|x_j)^2$. The arguments are similar to those in Section 1.3. Define $S_r^* = \sum_{j=2}^r \sum_{i=1}^{j-1} d_{ij}e_ie_j$. It is straightforward to see that S_r^* is a mean-zero martingale with respect to the filtration $\mathcal{F}_r^* = \sigma\{e_1, e_2, \ldots, e_r\}$. We establish the asymptotic normality by Corollary 3.1 of Hall and Heyde (1980). Define $W_j^* = \sum_{i=1}^{j-1} d_{ij}e_ie_j$ and note that $\mathbb{E}^*[W_j^{*2}|\mathcal{F}_{j-1}^*] = \sum_{i,k=1}^{j-1} d_{ij}d_{kj}e_ie_k$, and $2\sum_{j=1}^n \mathbb{E}^*[W_j^{*2}]/\{n(n-1)\} = \hat{\mathcal{S}}^2$. Direct calculation yields that for $j \geq j'$,

$$\begin{aligned}
&\text{cov}^*(\mathbb{E}[W_j^{*2}|\mathcal{F}_{j-1}^*], \mathbb{E}[W_{j'}^{*2}|\mathcal{F}_{j'-1}^*]) \\
=&\sum_{i,k=1}^{j-1}\sum_{i',k'=1}^{j'-1} d_{ij}d_{kj}d_{i'j'}d_{k'j'}\text{cov}(e_ie_k, e_{i'}e_{k'}) \\
=&\sum_{i,k,i',k'=1}^{j'-1} d_{ij}d_{kj}d_{i'j'}d_{k'j'}\text{cov}(e_ie_k, e_{i'}e_{k'}) \\
=&2\sum_{i=1}^{j'-1} d_{ij}^2d_{ij'}^2 + 2\sum_{1\leq i\neq k\leq j'-1} d_{ij}d_{ij'}d_{kj}d_{kj'} = 2\left(\sum_{i=1}^{j'-1} d_{ij}d_{ij'}\right)^2.
\end{aligned}$$

It implies that on $\mathcal{A}$

$$\begin{aligned}
&\frac{4}{n^2(n-1)^2}\sum_{j,j'=2}^n \text{cov}^*(\mathbb{E}[W_j^{*2}|\mathcal{F}_{j-1}^*], \mathbb{E}[W_{j'}^{*2}|\mathcal{F}_{j'-1}^*]) \\
=&\frac{8}{n^2(n-1)^2}\sum_{j,j'=2}^n\left(\sum_{i=1}^{\min\{j,j'\}-1} d_{ij}d_{ij'}\right)^2 = o(\hat{\mathcal{S}}^4).
\end{aligned}$$

We also note that on $\mathcal{A}$

$$\begin{aligned}\sum_{j=2}^{n}\mathbb{E}^*[|W_j^*|^4] =&3\sum_{j=2}^{n}\sum_{i_1,i_2,i_3,i_4=1}^{j-1} d_{i_1j}d_{i_2j}d_{i_3j}d_{i_4j}\mathbb{E}[e_{i_1}e_{i_2}e_{i_3}e_{i_4}]\\ =&9\sum_{j=2}^{n}\sum_{i=1}^{j-1}d_{ij}^4+9\sum_{j=2}^{n}\sum_{1\le i_1\neq i_2\le j-1}d_{i_1j}^2d_{i_2j}^2=9\sum_{j=2}^{n}\left(\sum_{i=1}^{j-1}d_{ij}^2\right)^2\\ =&o\left(\left(\sum_{k<l}d_{kl}^2\right)^2\right).\end{aligned}$$

Therefore by the martingale CLT (see Hall and Heyde (1980)), we have on $\mathcal{A}$,

$$\sqrt{\binom{n}{2}}\frac{\sum_{j=1}^{p}MDD_n^*(Y|x_j)^2}{\hat{\mathcal{S}}}\to^{d^*}N(0,1).$$

Step 2: To show the ratio consistency for the bootstrap variance estimator, we note that conditional on the sample, $\hat{S}^{*2}$ is an unbiased estimator for $\hat{\mathcal{S}}^2$. Therefore, it suffices to show that conditional on $\mathcal{A}$, $\frac{\text{var}^*(\hat{S}^{*2})}{\hat{\mathcal{S}}^4}\to 0$. Simple algebra shows that

$$\begin{aligned}\frac{\text{var}^*(\hat{S}^{*2})}{\hat{\mathcal{S}}^4}=&\frac{1}{\hat{\mathcal{S}}^4}\text{var}^*\left(\frac{1}{\binom{n}{2}}\sum_{1\le k<l\le n}d_{kl}^2e_k^2e_l^2\right)\\ =&\frac{4}{n^2(n-1)^2\hat{\mathcal{S}}^4}\left\{\sum_{k<l}d_{kl}^4\text{cov}(e_k^2e_l^2,e_k^2e_l^2)+\sum_{k<l\neq l'}d_{kl}^2d_{kl'}^2\text{cov}(e_k^2e_l^2,e_k^2e_{l'}^2)\right.\\ &\left.+\sum_{k\neq k'<l}d_{kl}^2d_{k'l}^2\text{cov}(e_k^2e_l^2,e_{k'}^2e_l^2)+2\sum_{k<l<l'}d_{kl}^2d_{ll'}^2\text{cov}(e_k^2e_l^2,e_l^2e_{l'}^2)\right\}\\ \le&\frac{C}{(\sum_{k<l}d_{kl}^2)^2}\left\{\sum_{k<l}d_{kl}^4+\sum_{k\neq l\neq l'}d_{kl}^2d_{ll'}^2\right\},\end{aligned}$$

for some constant $C>0$. On $\mathcal{A}$, we have

$$\frac{\sum_{k<l}d_{kl}^4}{(\sum_{k<l}d_{kl}^2)^2}\to 0,\quad \frac{\sum_{k\neq l\neq l'}d_{kl}^2d_{kl'}^2}{(\sum_{k<l}d_{kl}^2)^2}\to 0,$$

which indicates that

$$\frac{\hat{S}^*}{\hat{\mathcal{S}}}\to^{p^*}1.$$

Step 3: To deal with the $\mathcal{U}$-centered version of A_{ij} (here we omit the superscript), we write

$\widetilde{A}_{ij}$ in the following form,

$$\begin{aligned}\widetilde{A}_{ij} =& \frac{n-3}{n-1}A_{ij} - \frac{n-3}{(n-1)(n-2)}\sum_{l\notin\{i,j\}}A_{il} - \frac{n-3}{(n-1)(n-2)}\sum_{k\notin\{i,j\}}A_{kj} \\ &+ \frac{1}{(n-1)(n-2)}\sum_{k,l\notin\{i,j\},k\neq l}A_{kl} \\ =& \frac{n-3}{n-1}\bar{A}_{ij} - \frac{n-3}{(n-1)(n-2)}\sum_{l\notin\{i,j\}}\bar{A}_{il} - \frac{n-3}{(n-1)(n-2)}\sum_{k\notin\{i,j\}}\bar{A}_{kj} \\ &+ \frac{2}{(n-1)(n-2)}\sum_{k,l\notin\{i,j\},k<l}\bar{A}_{kl} \\ :=& I_{n,1} - I_{n,2} - I_{n,3} + I_{n,4}\end{aligned}$$

where $\bar{A}_{ij} = A_{ij} - \mathbb{E}[A_{il}|X_i] - \mathbb{E}[A_{kj}|X_j] + \mathbb{E}[A_{kl}]$ is the double centered version of A_{ij}. Recall that $\bar{A}_{ij}^{(k)} = -U_k(x_{ik}, x_{jk})$. A useful property is that the four terms $I_{n,k}, 1 \le k \le 4$ are uncorrelated with each other.

Remark **1.2**. This expression provides an alternative way of justifying the unbiasedness of the $\mathcal{U}$-centered estimator. For example, $\mathbb{E}[\widetilde{A}_{ij}^2]$ is simply equal to the sum of the variances of the above four terms (since they are uncorrelated by construction), which is equal to

$$\left\{\frac{(n-3)^2}{(n-1)^2} + \frac{2(n-3)^2}{(n-1)^2(n-2)} + \frac{2(n-3)}{(n-1)^2(n-2)}\right\}\mathbb{E}[\bar{A}_{ij}^2] = \frac{n-3}{n-1}\mathbb{E}[\bar{A}_{ij}^2].$$

Hence $\sum_{i\neq j}\widetilde{A}_{ij}^2/\{n(n-3)\}$ is an unbiased estimator for $\mathbb{E}[\bar{A}_{ij}^2]$.

Step 4: In this step, we prove that

$$\frac{\sum_{j\ge j'}\left(\sum_{i=1}^{j'-1} d_{ij}d_{ij'}\right)^2}{(\sum_{k<l} d_{kl}^2)^2} \to^p 0. \tag{24}$$

The above condition can be viewed as the sample version of Condition (8) in the paper. In Step 6 below, we show that $\hat{\mathcal{S}}^2$ is ratio-consistent under the null, i.e.,

$$\frac{\hat{\mathcal{S}}^2}{\mathcal{S}^2} \to^p 1.$$

To simplify the calculation, we shall assume (18). However, we emphasize that it is not essential and can be relaxed. Using the expression in Step 3, we can show $\mathbb{E}[\widetilde{B}_{ij}^4|X_i, X_j] \le C_1^4$ for some constant $C_1 > 0$. Using this fact, (18) and again the expression in Step 3 (see more details in

Step 5 below), we get

$$\begin{aligned}\mathbb{E}[d_{ij}^4] =& \sum_{j_1,j_2,j_3,j_4=1}^{p} \mathbb{E}[\widetilde{A}_{ij}^{(j_1)}\widetilde{A}_{ij}^{(j_2)}\widetilde{A}_{ij}^{(j_3)}\widetilde{A}_{ij}^{(j_4)}\widetilde{B}_{ij}^4] \\ \leq& C_1^4 \sum_{j_1,j_2,j_3,j_4=1}^{p} \mathbb{E}[\widetilde{A}_{ij}^{(j_1)}\widetilde{A}_{ij}^{(j_2)}\widetilde{A}_{ij}^{(j_3)}\widetilde{A}_{ij}^{(j_4)}] = C_1'\mathbb{E}[\widetilde{U}(X,X')^4].\end{aligned}$$

Similarly, for $i < j' < j$, we have

$$\begin{aligned}\mathbb{E}[d_{ij}^2 d_{ij'}^2] \leq& C_1^4 \sum_{j_1,j_2,j_3,j_4=1}^{p} \mathbb{E}[\widetilde{A}_{ij}^{(j_1)}\widetilde{A}_{ij}^{(j_2)}\widetilde{A}_{ij'}^{(j_3)}\widetilde{A}_{ij'}^{(j_4)}] \\ =& C_2'\mathbb{E}[\widetilde{U}(X,X')^2\widetilde{U}(X,X'')^2] + C_2''n^{-1}\mathbb{E}[\widetilde{U}(X,X')^4],\end{aligned}$$

and for $i \neq i' < j' < j$,

$$\begin{aligned}\mathbb{E}[d_{ij}d_{ij'}d_{i'j}d_{i'j'}] \leq& C_1^4 \sum_{j_1,j_2,j_3,j_4=1}^{p} \mathbb{E}[\widetilde{A}_{ij}^{(j_1)}\widetilde{A}_{ij'}^{(j_2)}\widetilde{A}_{i'j}^{(j_3)}\widetilde{A}_{i'j'}^{(j_4)}] \\ =& C_3'\mathbb{E}[\widetilde{U}(X,X')\widetilde{U}(X',X'')\widetilde{U}(X'',X''')\widetilde{U}(X''',X)] + C_3''n^{-1}\mathbb{E}[\widetilde{U}(X,X')^4].\end{aligned}$$

Therefore, by (19)-(20), we have,

$$\begin{aligned}&\frac{\mathbb{E}\sum_{j>j'}\left(\sum_{i=1}^{j'-1} d_{ij}d_{ij'}\right)^2}{n^4\mathcal{S}^4} \\ \leq& C_1\left(\frac{\mathbb{E}[\widetilde{U}(X,X')\widetilde{U}(X',X'')\widetilde{U}(X'',X''')\widetilde{U}(X''',X)]}{\mathcal{S}^4} + \frac{\mathbb{E}[\widetilde{U}(X,X')^4]}{n\mathcal{S}^4}\right) \to 0,\end{aligned}$$

which implies (24) by the Markov inequality.

Step 5: We provide some details on the calculation in Step 4 above. The key idea is to use the alternative expression given in Step 3. Let

$$c_n = \frac{(n-3)^4}{(n-1)^4} + \frac{2(n-3)^4}{(n-1)^4(n-2)^3} + \frac{2(n-3)}{(n-1)^4(n-2)^3},$$

such that $c_n \to 1$. We begin by considering

$$\begin{aligned}\mathbb{E}[d_{ij}^4] \leq& C_1^4 \sum_{j_1,j_2,j_3,j_4=1}^{p} \mathbb{E}[\widetilde{A}_{ij}^{(j_1)}\widetilde{A}_{ij}^{(j_2)}\widetilde{A}_{ij}^{(j_3)}\widetilde{A}_{ij}^{(j_4)}] \\ =& c_n C_1^4 \sum_{j_1,j_2,j_3,j_4=1}^{p} \mathbb{E}[\bar{A}_{ij}^{(j_1)}\bar{A}_{ij}^{(j_2)}\bar{A}_{ij}^{(j_3)}\bar{A}_{ij}^{(j_4)}] + \mathcal{R}_n \\ =& c_n C_1^4\mathbb{E}[\widetilde{U}(X,X')^4] + \mathcal{R}_n,\end{aligned}$$

where $\mathcal{R}_n = O(n^{-1}\mathbb{E}[\widetilde{U}(X,X')^4])$ is the smaller order term. Some typical terms in $\mathcal{R}_n$ include

$$O(n^{-4}) \sum_{j_1,j_2,j_3,j_4=1}^{p} \mathbb{E}[\bar{A}_{ij}^{(j_1)} \bar{A}_{jl}^{(j_2)} \bar{A}_{lm}^{(j_3)} \bar{A}_{mi}^{(j_4)}] = O\left(n^{-4}\mathbb{E}[\widetilde{U}(X,X')^4]\right),$$
$$O(n^{-4}) \sum_{j_1,j_2,j_3,j_4=1}^{p} \mathbb{E}[\bar{A}_{ij}^{(j_1)} \bar{A}_{ij}^{(j_2)} \bar{A}_{kl}^{(j_3)} \bar{A}_{kl}^{(j_4)}] = O\left(n^{-4}\mathbb{E}[\widetilde{U}(X,X')^4]\right), \quad \{i,j\}\cap\{k,l\} = \emptyset,$$
$$O(n^{-2}) \sum_{j_1,j_2,j_3,j_4=1}^{p} \mathbb{E}[\bar{A}_{ij}^{(j_1)} \bar{A}_{ij}^{(j_2)} \bar{A}_{im}^{(j_3)} \bar{A}_{im}^{(j_4)}] = O\left(n^{-2}\mathbb{E}[\widetilde{U}(X,X')^4]\right), \quad m \in \{i,j\},$$

where we have used the Hölder's inequality. The calculation for $\mathbb{E}[d_{ij}^2 d_{ij'}^2]$ and $\mathbb{E}[d_{ij}d_{ij'}d_{i'j}d_{i'j'}]$ is similar and hence is skipped.

Step 6: To finish the proof, we show the ratio consistency $\frac{\hat{\mathcal{S}}^2}{\mathcal{S}^2} \to^p 1$. By the Markov inequality, we only need to show that

$$\frac{\mathrm{var}(\sum_{k<l} d_{kl}^2)}{n^4\mathcal{S}^4} + \frac{|\mathbb{E}[d_{kl}^2] - \mathcal{S}^2|^2}{\mathcal{S}^4} \to 0.$$

Using the expression in Step 3, we have,

$$\mathbb{E}[d_{kl}^2] = \sum_{j,j'=1}^{p} \mathbb{E}[\widetilde{A}_{kl}^{(j)} \widetilde{A}_{kl}^{(j')} \widetilde{B}_{kl}^2] = (1+O(n^{-1}))\mathcal{S}^2 + C_1 n^{-1}\mathbb{E}[\widetilde{U}(X,X')^4]^{1/2},$$

which implies $\frac{|\mathbb{E}[d_{kl}^2]-\mathcal{S}^2|^2}{\mathcal{S}^4} \to 0$. On the other hand, we note that

$$\begin{aligned}
&\frac{\mathrm{var}(\sum_{k<l} d_{kl}^2)}{n^4\mathcal{S}^4} \\
=&\frac{\sum_{k<l}\sum_{k'<l'} \mathrm{cov}(d_{kl}^2, d_{k'l'}^2)}{n^4\mathcal{S}^4} \\
=&\frac{2\sum_{k<l<l'} \mathrm{cov}(d_{kl}^2, d_{l'l}^2) + \sum_{k<l} \mathrm{var}(d_{kl}^2)}{n^4\mathcal{S}^4} + \frac{\sum_{k<l,k'<l',\{k,l\}\cap\{k',l'\}=\emptyset} \mathrm{cov}(d_{kl}^2, d_{k'l'}^2)}{n^4\mathcal{S}^4} \\
\leq&\frac{2\sum_{k<l<l'} \mathbb{E}[d_{kl}^2 d_{l'l}^2] + \sum_{k<l} \mathbb{E}[d_{kl}^4]}{n^4\mathcal{S}^4} + \frac{\sum_{k<l,k'<l',\{k,l\}\cap\{k',l'\}=\emptyset} \mathrm{cov}(d_{kl}^2, d_{k'l'}^2)}{n^4\mathcal{S}^4}.
\end{aligned} \tag{25}$$

Following similar arguments in Step 4, we can show that the first term in (25) converges to zero. To deal with the second term in (25), we again use the expression in Step 3. Using the fact that $\mathrm{cov}(\bar{A}_{kl}^{(j_1)} \bar{A}_{kl}^{(j_2)} \bar{B}_{kl}^2, \bar{A}_{k'l'}^{(j_3)} \bar{A}_{k'l'}^{(j_4)} \bar{B}_{k'l'}^2) = 0$ for any $1 \leq j_1, j_2, j_3, j_4 \leq p$ and $\{k,l\}\cap\{k',l'\} = \emptyset$, and the Hölder's inequality, we know $\sum_{k<l,k'<l',\{k,l\}\cap\{k',l'\}=\emptyset} \mathrm{cov}(d_{kl}^2, d_{k'l'}^2)$ is bounded by $C_1 n^3 \mathbb{E}[\widetilde{U}(X,X')^4]$ for some sufficiently large constant C_1. Therefore by (20), we have

$$\frac{\sum_{k<l,k'<l',\{k,l\}\cap\{k',l'\}=\emptyset} \mathrm{cov}(d_{kl}^2, d_{k'l'}^2)}{n^4\mathcal{S}^4} \leq \frac{C_2\mathbb{E}[\widetilde{U}(X,X')^4]}{n(\sum_{j,j'=1}^{p} dcov(x_j, x_{j'})^2)^2} \to 0.$$

This completes the proof.

1.7 Characterization of local alternative models

We provide some discussions on the local alternative models. Let $\imath = \sqrt{-1}$ and $w(t) = 1/(\pi t^2)$. Define $G_j(u_j, y; t) = (f_{x_j}(t) - e^{\imath t u_j})(\mu - y)$ for $u_j, y, t \in \mathbb{R}$. By Lemma 1 of Székely et al. (2007), it can be shown that

$$H(z, z') = \int_{\mathbb{R}} \sum_{j=1}^{p} G_j(u_j, y; t)\overline{G_j(u'_j, y'; t)}w(t)dt, \tag{26}$$

where $z = (u_1, \dots, u_p, y)^T$ and $z' = (u'_1, \dots, u'_p, y')^T$. Recall that $\widetilde{\mathcal{L}}(x, y) = \mathbb{E}[\widetilde{U}(x, \mathcal{X})V(y, \mathcal{Y})]$. To have a better understanding of our local alternative model, we shall show that the two conditions

$$\operatorname{var}(\widetilde{\mathcal{L}}(X, Y)) = o(n^{-1}\mathcal{S}^2), \tag{27}$$

$$\operatorname{var}(\widetilde{\mathcal{L}}(X, Y')) = o(\mathcal{S}^2), \tag{28}$$

have a similar interpretation as equation (4.2) in Zhong and Chen (2011). Using representation (26), we have

$$\widetilde{\mathcal{L}}(x, y) = \int_{\mathbb{R}} \sum_{j=1}^{p} G_j(x_j, y; t)\overline{\mathbb{E}[G_j(x'_j, Y'; t)]}w(t)dt.$$

Thus (27) can be re-expressed as

$$\begin{aligned}\operatorname{var}(\widetilde{\mathcal{L}}(X, Y)) = \int_{\mathbb{R}}\int_{\mathbb{R}} \sum_{j,k=1}^{p} &\mathbb{E}\{(G_j(x_j, Y; t_1) - \mathbb{E}G_j(x_j, Y; t_1))\overline{(G_k(x_k, Y; t_2) - \mathbb{E}G_k(x_k, Y; t_2))}\} \\ &\mathbb{E}[G_k(x'_k, Y'; t_2)]\overline{\mathbb{E}[G_j(x'_j, Y'; t_1)]}w(t_1)w(t_2)dt_1dt_2 = O(n^{-1}\mathcal{S}^2),\end{aligned}$$

and (28) can be written as

$$\begin{aligned}\operatorname{var}(\widetilde{\mathcal{L}}(X, Y')) &= \int_{\mathbb{R}}\int_{\mathbb{R}} \sum_{j,k=1}^{p} \mathbb{E}\{G_j(x_j, Y'; t_1)\overline{G_k(x_k, Y'; t_2)}\} \\ &\quad \mathbb{E}[G_k(x'_k, Y'; t_2)]\overline{\mathbb{E}[G_j(x'_j, Y'; t_1)]}w(t_1)w(t_2)dt_1dt_2 \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \sum_{j,k=1}^{p} \mathbb{E}\{(f_{x_j}(t_1) - e^{\imath t_1 u_j})(f_{x_k}(-t_2) - e^{-\imath t_2 u_k})\}\operatorname{var}(Y) \\ &\quad \mathbb{E}[G_k(x'_k, Y'; t_2)]\overline{\mathbb{E}[G_j(x'_j, Y'; t_1)]}w(t_1)w(t_2)dt_1dt_2.\end{aligned}$$

By the definition $H^*(Z, Z') = H(Z, Z') - \widetilde{\mathcal{L}}(X, Y) - \widetilde{\mathcal{L}}(X', Y') + \mathbb{E}[\widetilde{U}(X, X')V(Y, Y')]$ and $\mathcal{S}^2 = O(\operatorname{var}(H^*(Z, Z')))$. Also note that H^* can be re-expressed as

$$H^*(Z, Z') = \int_{\mathbb{R}} \sum_{j=1}^{p}\{G_j(x_j, Y; t) - \mathbb{E}G_j(x_j, Y; t)\}\overline{\{G_j(x'_j, Y'; t) - \mathbb{E}[G_j(x'_j, Y'; t)]\}}w(t)dt. \tag{29}$$

We can define an operator Φ^* in a similar fashion as Φ by replacing $G_j(x_j, Y; t)$ with its demeaned version $G_j(x_j, Y; t) - \mathbb{E}[G_j(x_j, Y; t)]$. Denote by Φ_0 with $\mathbb{E}[G_j(x_j, Y; t)\overline{G_i(x_i, Y; t')}]$ being replaced by $\mathrm{var}(Y)(f_{x_j,x_i}(t, -t') - f_{x_j}(t)f_{x_i}(-t'))$ in Φ. Note that $\Phi = \Phi_0$ when X and Y are independent. Let $h(t) = (h_1(t), \dots, h_p(t))$ with $h_k(t) = E[G_k(x_k, Y; t)]$. For $g(t) = (g_1(t), \dots, g_p(t))$ and $\widetilde{g}(t) = (\widetilde{g}_1(t), \dots, \widetilde{g}_p(t))$, define the inner product,

$$< g, \widetilde{g} >_w = \sum_{j=1}^{p} \int_{\mathbb{R}} w(t) g_j(t) \overline{\widetilde{g}_j(t)} dt.$$

Thus Conditions (27)-(28) become

$$< \Phi^*(h), h >_w = O(n^{-1}Tr((\Phi^*)^2)), \tag{30}$$

$$< \Phi_0(h), h >_w = O(Tr((\Phi^*)^2)). \tag{31}$$

Notice that $h(t) = \mathbf{0}_{p\times 1}$ for all t under the null. Conditions (30)-(31) quantify the distance between the alternatives and the null hypothesis. The characterization of the local alternative model is somewhat abstract but this is sensible because MDD targets at very broad alternatives, i.e., arbitrary type of conditional mean dependence.

1.8 Asymptotic analysis for conditional quantile dependence testing

In this subsection, we prove Proposition 3.1. We first state the following lemma whose proof is given in Shao and Zhang (2014).

Lemma **1.2**. *Suppose the distribution function of Y satisfies Assumption 3.1, then there exist $\epsilon_0 > 0$ and $c > 0$ such that for $\epsilon \in (0, \epsilon_0)$,*

$$P\left(\frac{1}{n}\sum_{k=1}^{n}\left|\hat{W}_k - W_k\right| > \epsilon\right) \le 3\exp(-2nc\epsilon^2).$$

Let $B_{Q,kl} = |W_k - W_l|^2/2$ and $B^*_{Q,kl}$ be its $\mathcal{U}$-centered version. Note that

$$MDD_n(\hat{W}|x_j)^2 - MDD_n(W|x_j)^2 = \frac{1}{n(n-3)}\sum_{k\neq l}\widetilde{A}_{kl}(j)(\hat{B}^*_{Q,kl} - B^*_{Q,kl}).$$

Thus we need to show that

$$\frac{1}{\mathcal{S}_Q}\sum_{k\neq l}(\hat{B}^*_{Q,kl} - B^*_{Q,kl})\sum_{j=1}^{p}\widetilde{A}_{kl}(j) = o_p(n).$$

Direct calculation yields that

$$\begin{aligned}\hat{B}^*_{Q,kl} =& \frac{-1}{n-2}(\hat{W}_k^2+\hat{W}_l^2) - \hat{W}_k\hat{W}_l + \frac{1}{(n-1)(n-2)}\sum_{j=1}^n \hat{W}_j^2 \\ &+ \frac{1}{n-2}(\hat{W}_k+\hat{W}_l)\sum_{j=1}^n \hat{W}_j - \frac{1}{(n-1)(n-2)}\left(\sum_{j=1}^n \hat{W}_j\right)^2,\end{aligned}$$

and a similar decomposition for $B^*_{Q,kl}$. Hence we have

$$\begin{aligned}&\frac{1}{\mathcal{S}_Q}\sum_{k\neq l}(\hat{B}^*_{Q,kl}-B^*_{Q,kl})\sum_{j=1}^p \widetilde{A}_{kl}(j) \\ =&\frac{-2}{\mathcal{S}_Q(n-2)}\sum_{k\neq l}(\hat{W}_k^2-W_k^2)\sum_{j=1}^p \widetilde{A}_{kl}(j) + \frac{1}{\mathcal{S}_Q(n-1)(n-2)}\sum_{i=1}^n(\hat{W}_i^2-W_i^2)\sum_{k\neq l}\sum_{j=1}^p \widetilde{A}_{kl}(j) \\ &- \frac{1}{\mathcal{S}_Q}\sum_{k\neq l}(\hat{W}_k\hat{W}_l - W_kW_l)\sum_{j=1}^p \widetilde{A}_{kl}(j) + \frac{2}{(n-2)S_Q}\sum_{k\neq l}\left(\hat{W}_k\sum_{i=1}^n \hat{W}_i - W_k\sum_{i=1}^n W_i\right)\sum_{j=1}^p \widetilde{A}_{kl}(j) \\ &- \frac{1}{(n-1)(n-2)S_Q}\left\{\left(\sum_{i=1}^n \hat{W}_i\right)^2 - \left(\sum_{i=1}^n W_i\right)^2\right\}\sum_{k\neq l}\sum_{j=1}^p \widetilde{A}_{kl}(j) \\ =&J_{1,n}+J_{2,n}+J_{3,n}+J_{4,n}+J_{5,n},\end{aligned}$$

where $J_{i,n}$ with $1\leq i\leq 5$ are defined implicity. Notice that

$$\begin{aligned}|J_{1,n}|^2 \leq& \frac{4}{\mathcal{S}_Q^2(n-2)^2}\sum_{k\neq l}(\hat{W}_k^2-W_k^2)^2\sum_{k\neq l}\left(\sum_{j=1}^p \widetilde{A}_{kl}(j)\right)^2 \\ \leq& \frac{C(n-1)}{\mathcal{S}_Q^2(n-2)^2}\sum_{k=1}^n|\hat{W}_k-W_k|\sum_{k\neq l}\left(\sum_{j=1}^p \widetilde{A}_{kl}(j)\right)^2,\end{aligned}$$

and

$$\begin{aligned}|J_{2,n}| \leq& \frac{C'}{\mathcal{S}_Q(n-1)(n-2)}\sum_{i=1}^n|\hat{W}_i-W_i|\left|\sum_{k\neq l}\sum_{j=1}^p \widetilde{A}_{kl}(j)\right| \\ \leq& \frac{C'n}{\mathcal{S}_Q(n-1)(n-2)}\sum_{i=1}^n|\hat{W}_i-W_i|\left\{\sum_{k\neq l}\left(\sum_{j=1}^p \widetilde{A}_{kl}(j)\right)^2\right\}^{1/2},\end{aligned}$$

for some constants $C, C' > 0$. By Lemma 1.2, we need to show that

$$\frac{1}{n^2\mathcal{S}_Q^2}\mathbb{E}\sum_{k\neq l}\left(\sum_{j=1}^p \widetilde{A}_{kl}(j)\right)^2 = O(1).$$

By Székely and Rizzoz (2014) and (11) in the paper, we have

$$\mathbb{E}\left[\sum_{k\neq l}\left(\sum_{j=1}^{p}\widetilde{A}_{kl}(j)\right)^2\right]=\sum_{j,j'=1}^{p}\mathbb{E}\left(\sum_{k\neq l}\widetilde{A}_{kl}(j)\widetilde{A}_{kl}(j')\right)=n(n-3)\sum_{j,j'=1}^{p}dcov(x_j,x_{j'})^2=O(n^2\mathcal{S}_Q^2).$$

Then we have $J_{1,n}=o_p(n)$ and $J_{2,n}=o_p(n)$. Recall that $Q_\tau=Q_\tau(Y)$ is the τth quantile of Y. For $J_{3,n}$, we note that

$$\begin{aligned}
J_{3,n}=&-\frac{1}{\mathcal{S}_Q}\sum_{k\neq l}(\hat{W}_k\hat{W}_l-W_kW_l)\sum_{j=1}^{p}\widetilde{A}_{kl}(j)\\
=&-\frac{1}{\mathcal{S}_Q}\sum_{k\neq l}(\hat{W}_k-W_k)\hat{W}_l\sum_{j=1}^{p}\widetilde{A}_{kl}(j)-\frac{1}{\mathcal{S}_Q}\sum_{k\neq l}(\hat{W}_l-W_l)W_k\sum_{j=1}^{p}\widetilde{A}_{kl}(j)\\
=&-\frac{1}{\mathcal{S}_Q}\sum_{k\neq l}\mathbf{1}\{Q_\tau<Y_k\leq\hat{Q}_\tau\}\hat{W}_l\sum_{j=1}^{p}\widetilde{A}_{kl}(j)+\frac{1}{\mathcal{S}_Q}\sum_{k\neq l}\mathbf{1}\{\hat{Q}_\tau<Y_k\leq Q_\tau\}\hat{W}_l\sum_{j=1}^{p}\widetilde{A}_{kl}(j)\\
&-\frac{1}{\mathcal{S}_Q}\sum_{k\neq l}\mathbf{1}\{Q_\tau<Y_l\leq\hat{Q}_\tau\}W_k\sum_{j=1}^{p}\widetilde{A}_{kl}(j)+\frac{1}{\mathcal{S}_Q}\sum_{k\neq l}\mathbf{1}\{\hat{Q}_\tau<Y_l\leq Q_\tau\}W_k\sum_{j=1}^{p}\widetilde{A}_{kl}(j)\\
=&I_{1,n}+I_{2,n}+I_{3,n}+I_{4,n}.
\end{aligned}$$

Let $\mathcal{A}_\delta=\{\hat{Q}_\tau-Q_\tau>\delta\}\cup\{\hat{Q}_\tau-Q_\tau\leq-\delta\}$. For any $\epsilon>0$, choose $0<\delta<\delta_0$ such that $P(\mathcal{A}_\delta)\leq\epsilon/2$ for large enough n. We claim that that for small enough δ, and $C_i=Q_\tau$ or $\hat{Q}_\tau$,

$$\begin{aligned}
&\mathbb{E}[\mathbf{1}\{Y_{j_1}\leq C_1,Y_{j_2}\leq C_2,\ldots,Y_{j_k}\leq C_k,-\delta<\hat{Q}_\tau-Q_\tau\leq\delta\}|X_1,\ldots,X_n]\\
=&\mathbb{E}[\mathbf{1}\{Y_{j_1}\leq C_1,Y_{j_2}\leq C_2,\ldots,Y_{j_k}\leq C_k,-\delta<\hat{Q}_\tau-Q_\tau\leq\delta\}],
\end{aligned}\tag{32}$$

where $k = 2, 3, 4$. The proof of (32) is given at the end. By (32), we have for any $\varepsilon > 0$,

$$
\begin{aligned}
&P(|I_{1,n}/n| > \varepsilon) \\
\leq &P(|I_{1,n}/n| > \varepsilon, \mathcal{A}_\delta^c) + P(\mathcal{A}_\delta) \leq P(|I_{1,n}/n| > \varepsilon, \mathcal{A}_\delta^c) + \epsilon/2 \\
\leq &P\left(\left| \frac{1}{n\mathcal{S}_Q} \sum_{k \neq l} \mathbf{1}\{Q_\tau < Y_k \leq \hat{Q}_\tau, \mathcal{A}_\delta^c\} \hat{W}_l \sum_{j=1}^{p} \widetilde{A}_{kl}(j) \right| > \varepsilon \right) + \epsilon/2 \\
\leq &\frac{1}{\varepsilon^2 n^2 \mathcal{S}_Q^2} \mathbb{E} \left| \sum_{k \neq l} \mathbf{1}\{Q_\tau < Y_k \leq \hat{Q}_\tau, \mathcal{A}_\delta^c\} \hat{W}_l \sum_{j=1}^{p} \widetilde{A}_{kl}(j) \right|^2 + \epsilon/2 \\
\leq &\frac{1}{\varepsilon^2 n^2 \mathcal{S}_Q^2} \sum_{\{k,l\} \neq \{k',l'\}, k \neq l, k' \neq l'} \mathbb{E}[\mathbf{1}\{Q_\tau < Y_k \leq \hat{Q}_\tau, Q_\tau < Y_{k'} \leq \hat{Q}_\tau, \mathcal{A}_\delta^c\}] \\
&\times \hat{W}_l \hat{W}_{l'} \sum_{j,j'=1}^{p} \mathbb{E}[\widetilde{A}_{kl}(j) \widetilde{A}_{k'l'}(j')] + \frac{C\delta}{\varepsilon^2 n^2 \mathcal{S}_Q^2} \mathbb{E} \sum_{k \neq l} \left(\sum_{j=1}^{p} \widetilde{A}_{kl}(j) \right)^2 + \epsilon/2, \\
\leq &\frac{C\delta}{\varepsilon^2 n^2 \mathcal{S}_Q^2} \sum_{\{k,l\} \neq \{k',l'\}, k \neq l, k' \neq l'} \left| \sum_{j,j'=1}^{p} \mathbb{E} \widetilde{A}_{kl}(j) \widetilde{A}_{k'l'}(j') \right| + \frac{C\delta}{\varepsilon^2 n^2 \mathcal{S}_Q^2} \mathbb{E} \sum_{k \neq l} \left(\sum_{j=1}^{p} \widetilde{A}_{kl}(j) \right)^2 + \epsilon/2,
\end{aligned}
$$

for some $C > 0$, where we have used the fact that

$$
\begin{aligned}
&\mathbb{E}[\mathbf{1}\{Q_\tau < Y_k \leq \hat{Q}_\tau, Q_\tau < Y_{k'} \leq \hat{Q}_\tau, \mathcal{A}_\delta^c\} \hat{W}_l \hat{W}_{l'}] \\
\leq &C\mathbb{E}[\mathbf{1}\{Q_\tau < Y_k \leq Q_\tau + \delta\}] \leq C G_2(\delta_0)\delta.
\end{aligned}
$$

Define the following quantities

$$
d_1(j) = \mathbb{E}[|x_j - x_j'|], \quad d_2(j,j') = \mathbb{E}[|x_j - x_j'||x_{j'} - x_{j'}''|], \quad d_3(j,j') = \mathbb{E}[|x_j - x_j'||x_{j'} - x_{j'}'|].
$$

We have for $\{k, l\} \cap \{k', l'\} = \emptyset$ and any $1 \leq j, j' \leq p$,

$$
\begin{aligned}
&\mathbb{E}[\widetilde{A}_{kl}(j) \widetilde{A}_{k'l'}(j')] \\
= &d_1(j) d_1(j') - \frac{4}{n-2} \left\{ 2 d_2(j,j') + (n-3) d_1(j) d_1(j') \right\} \\
&- \frac{n}{(n-1)(n-2)^2} \left\{ (n-2)(n-3) d_1(j) d_1(j') + 2 d_3(j,j') + 4(n-2) d_2(j,j') \right\} \\
&+ \frac{4}{(n-2)^2} \left\{ (n-2)(n-3) d_1(j) d_1(j') + 3(n-2) d_2(j,j') + d_3(j,j') \right\} \\
= &\frac{2}{(n-1)(n-2)} (d_1(j) d_1(j') - 2 d_2(j,j') + d_3(j,j')) = \frac{2}{(n-1)(n-2)} dcov(x_j, x_{j'})^2.
\end{aligned}
$$

Notice that in the summation over $\{k, l\} \neq \{k', l'\}, k \neq l, k' \neq l'$, we have $O(n^4)$ such terms.

Using similar calculation, we have for $l \neq l'$, $l \neq k$ and $l' \neq k$,

$$
\begin{aligned}
&\mathbb{E}[\widetilde{A}_{kl}(j)\widetilde{A}_{kl'}(j')] \\
=&d_2(j,j') - \frac{n-3}{(n-2)^2}\left\{(n-2)d_2(j,j') + d_3(j,j')\right\} - \frac{1}{n-2}\left\{2d_2(j,j') + (n-3)d_1(j)d_1(j')\right\} \\
&+ \frac{3}{(n-2)^2}\left\{(n-2)(n-3)d_1(j)d_1(j') + 3(n-2)d_2(j,j') + d_3(j,j')\right\} \\
&- \frac{n}{(n-1)(n-2)^2}\left\{(n-2)(n-3)d_1(j)d_1(j') + 2d_3(j,j') + 4(n-2)d_2(j,j')\right\} \\
&- \frac{1}{n-2}\left\{(n-3)d_1(j)d_1(j') + 2d_2(j,j')\right\} \\
=&-\frac{(n-3)}{(n-1)(n-2)}(d_1(j)d_1(j') - 2d_2(j,j') + d_3(j,j')) = -\frac{(n-3)}{(n-1)(n-2)}dcov(x_j, x_{j'})^2.
\end{aligned}
$$

And we have $O(n^3)$ such terms in the summation over $\{k,l\} \neq \{k',l'\}, k \neq l, k' \neq l'$. Thus we deduce that

$$
\sum_{\{k,l\}\neq\{k',l'\},k\neq l,k'\neq l'} \sum_{j,j'=1}^{p} \left|\mathbb{E}\widetilde{A}_{kl}(j)\widetilde{A}_{k'l'}(j')\right| \leq Cn^2 \sum_{j,j'=1}^{p} dcov(x_j, x_{j'})^2.
$$

By (11) in the paper, W is independent of X and thus

$$
\frac{\sum_{j,j'=1}^{p} dcov(x_j, x_{j'})^2}{S_Q^2} = \tau^{-2}(1-\tau)^{-2}.
$$

Hence we can choose a small enough δ such that

$$
P(|I_{1,n}/n| > \varepsilon) \leq \epsilon,
$$

which suggests that $I_{1,n} = o_p(n)$. Similarly we have $I_{j,n} = o_p(n)$ for $2 \leq j \leq 4$, which implies that $J_{3,n} = o_p(n)$. Using similar arguments, we can show that $J_{4,n} = o_p(n)$. Finally note that by Lemma 1.2,

$$
J_{5,n}/n \leq o_p(1)\frac{1}{nS_Q}\left|\sum_{k\neq l}\sum_{j=1}^{p}\widetilde{A}_{kl}(j)\right|.
$$

Because

$$
\mathbb{E}\frac{1}{n^2S_Q^2}\left|\sum_{k\neq l}\sum_{j=1}^{p}\widetilde{A}_{kl}(j)\right|^2 = O(1),
$$

we get $J_{5,n} = o_p(n)$.

Finally we prove (32). Note that $\hat{Q}_\tau = \inf\{y : F_n(y) \geq \tau\}$, where F_n is the empirical distribution function based on $\{Y_i\}_{i=1}^n$. Thus $-\delta < \hat{Q}_\tau - Q_\tau \leq \delta$ is equivalent to

$$F_n(Q_\tau + \delta) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{Y_i \leq Q_\tau + \delta\} \geq \tau,$$
$$F_n(Q_\tau - \delta) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{Y_i \leq Q_\tau - \delta\} < \tau.$$

We see that the event $\mathbf{1}\{Y_{j_1} \leq C_1, Y_{j_2} \leq C_2, \ldots, Y_{j_k} \leq C_k, -\delta < \hat{Q}_\tau - Q_\tau \leq \delta\}$ only depends on $\mathbf{1}\{Y_{j_i} \leq C_i\}$ for $1 \leq i \leq k$, and $\mathbf{1}\{Y_i \leq Q_\tau \pm \delta\}$. Moreover, given that $-\delta < \hat{Q}_\tau - Q_\tau \leq \delta$, the value of $\hat{Q}_\tau$ is determined by $\mathbf{1}\{Y_i \leq Q_\tau + a\}$ for $-\delta \leq a \leq \delta$ and $1 \leq i \leq n$. Hence the event $\mathbf{1}\{Y_{j_1} \leq C_1, Y_{j_2} \leq C_2, \ldots, Y_{j_k} \leq C_k, -\delta < \hat{Q}_\tau - Q_\tau \leq \delta\}$ is determined by $\mathbf{1}\{Y_i \leq Q_\tau + a\}$ for $-\delta \leq a \leq \delta$ and $1 \leq i \leq n$. Therefore, (32) holds under Assumption 3.2. The proof for Proposition 3.1 is thus completed.

2 Extension to factorial designs

As motivated by the study in Zhong and Chen (2011), we propose an extension of our MDD-based test to the situation where the observation (X_i, Y_i) is not a simple random sample but has a factorial design structure, as often the case in microarray study. Following Zhong and Chen (2011), we shall focus on the two way factorial designs with two factors A and B, where A has I levels and B has J levels. In the latter paper, they assumed the observations (X_{ijk}, Y_{ijk}) in the ith level of A and jth level of B satisfy a linear model:

$$\mathbb{E}(Y_{ijk}|X_{ijk}) = \mu_{ij} + X_{ijk}^T\beta, \quad i = 1, \ldots, I; \quad j = 1, \ldots, J,$$

for $k = 1, 2, \ldots, n_{ij}$, where $X_{ijk} = (x_{ijk,1}, \ldots, x_{ijk,p})^T$, n_{ij} denotes the number of observations in cell (i, j), p is the dimension of the covariates, and μ_{ij} denotes the fixed or random effect corresponding to the cell (i, j). In Zhong and Chen (2011), they tested the hypothesis $\tilde{H}_0 : \beta = \beta_0$ versus $\tilde{H}_a : \beta \neq \beta_0$. In particular, when $\beta_0 = 0$, the null corresponds to the conditional mean independence of Y given X in each of the cell (i, j) for the factorial design regardless of the nuisance parameters μ_{ij}. In this section, we generalize the test in Section 2.2 of the main paper to the factorial design case.

We introduce some notation first. Let $(X_{ij}, Y_{ij}) \overset{D}{=} (X_{ijk}, Y_{ijk})$ for $k = 1, 2, ..., n_{ij}$, where $X_{ij} = (x_{ij,1}, ..., x_{ij,p})^T$. One way to formulate the test is to consider the following hypothesis

$$H_0'' : \mathbb{E}[Y_{ij}|x_{ij,h}] = \mathbb{E}[Y_{ij}] \quad \text{almost surely},$$

for all $1 \leq h \leq p$, $1 \leq i \leq I$, and $1 \leq j \leq J$ versus the alternative that

$$H_a'' : P\left(\mathbb{E}[Y_{ij}|x_{ij,h}] \neq \mathbb{E}[Y_{ij}]\right) > 0$$

for some $1 \leq h \leq p$, $1 \leq i \leq I$, and $1 \leq j \leq J$.

Throughout the discussions, we assume independence across different cells, i.e.,

$$(Y_{ij}, X_{ij}) \text{ is independent of } (Y_{i'j'}, X_{i'j'}), \tag{33}$$

for $(i,j) \neq (i',j')$. Let $MDD_{n_{ij}}(Y_{ij}|x_{ij,h})^2$ be the unbiased estimator for $MDD(Y_{ij}|x_{ij,h})^2$ based on the sample $(Y_{ijk}, X_{ijk})_{k=1}^{n_{ij}}$. By the Hoeffding decomposition, we have that under the H_0'',

$$\begin{aligned}
&MDD_{n_{ij}}(Y_{ij}|x_{ij,h})^2 \\
&= \frac{1}{\binom{n_{ij}}{2}} \sum_{1\le k<l\le n_{ij}} U_{ij,h}(x_{ijk,h}, x_{ijl,h}) V_{ij}(Y_{ijk}, Y_{ijl}) + R_{ij,h,n_{ij}}
\end{aligned}$$

where $R_{ij,h,n_{ij}}$ is an asymptotically negligible remainder term. Define

$$S_{ij}^2 = \sum_{h,h'=1}^{p} \sigma_{h,h'}^{ij}, \quad \sigma_{h,h'}^{ij} = E[U_{ij,h}(x_{ij,h}, x'_{ij,h}) U_{ij,h'}(x_{ij,h'}, x'_{ij,h'}) V_{ij}^2(Y_{ij}, Y'_{ij})],$$

where (Y'_{ij}, X'_{ij}) is an independent copy of (Y_{ij}, X_{ij}). A natural estimator for $\sigma_{h,h'}^{ij}$ is given by

$$\hat{\sigma}_{h,h'}^{ij} = \frac{1}{\binom{n_{ij}}{2}} \sum_{1\le k<l\le n_{ij}} \widetilde{A}_{kl}^{ij}(h) \widetilde{A}_{kl}^{ij}(h') (\widetilde{B}_{kl}^{ij})^2$$

with $\widetilde{A}_{kl}^{ij}(h)$ and $\widetilde{B}_{kl}^{ij}$ being the $\mathcal{U}$-centered versions of $A_{kl}^{ij}(h) = |x_{ijk,h} - x_{ijl,h}|$ and $\widetilde{B}_{kl}^{ij} = |Y_{ijk} - Y_{ijl}|^2/2$ respectively. Denote $\hat{\Sigma}_{ij} = (\hat{\sigma}_{h,h'}^{ij})_{h,h'=1}^{p}$ and $\hat{\mathcal{S}}_{ij}^2 = \mathbf{1}_p' \hat{\Sigma}_{ij} \mathbf{1}_p$. The test statistics we considered for the case of factorial designs are

$$\breve{T}_{F,n} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} \sqrt{\binom{n_{ij}}{2}} \sum_{h=1}^{p} MDD_{n_{ij}}(Y_{ij}|x_{ij,h})^2}{\sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} \mathcal{S}_{ij}^2}},$$

and

$$T_{F,n} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} \sqrt{\binom{n_{ij}}{2}} \sum_{h=1}^{p} MDD_{n_{ij}}(Y_{ij}|x_{ij,h})^2}{\sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} \hat{\mathcal{S}}_{ij}^2}}.$$

Similarly, we can define $Z'_{ij} = (X'_{ij}, Y'_{ij})$, $Z''_{ij} = (X''_{ij}, Y''_{ij})$ and $Z'''_{ij} = (X'''_{ij}, Y'''_{ij})$ to be independent copies of $Z_{ij} = (X_{ij}, Y_{ij})$. Recall the definitions $\widetilde{U}_{ij}(X_{ij}, X'_{ij}) = \sum_{h=1}^{p} \widetilde{U}_{ij,h}(x_{ij,h}, x'_{ij,h})$, $H_{ij}(Z_{ij}, Z'_{ij}) = \widetilde{U}_{ij}(X_{ij}, X'_{ij}) V_{ij}(Y_{ij}, Y'_{ij})$ and $G_{ij}(Z_{ij}, Z'_{ij}) = \mathbb{E}[H_{ij}(Z_{ij}, Z''_{ij}) H_{ij}(Z'_{ij}, Z''_{ij})|(Z_{ij}, Z'_{ij})]$. For each $(i,j) \in \{1,...,I\} \times \{1,...,J\}$, we impose the following conditions:

$$\frac{\mathbb{E}[G_{ij}(Z_{ij},Z'_{ij})^2]}{\{\mathbb{E}[H_{ij}(Z_{ij},Z'_{ij})^2]\}^2} \to 0,$$
$$\frac{\mathbb{E}[H_{ij}(Z_{ij},Z'_{ij})^4]/n + \mathbb{E}[H_{ij}(Z_{ij},Z''_{ij})^2 H_{ij}(Z'_{ij},Z''_{ij})^2]}{n\{\mathbb{E}[H_{ij}(Z_{ij},Z'_{ij})^2]\}^2} \to 0, \tag{34}$$

and

$$\frac{\mathbb{E}[\widetilde{U}_{ij}(X_{ij},X''_{ij})^2 V_{ij}(Y_{ij},Y'_{ij})^2]}{\mathcal{S}_{ij}^2} = O(n), \tag{35}$$

$$\frac{\mathbb{E}[\widetilde{U}_{ij}(X_{ij},X'_{ij})^2]\mathbb{E}[V_{ij}(Y_{ij},Y'_{ij})^2]}{\mathcal{S}_{ij}^2} = O(n^2). \tag{36}$$

Theorem **2.1**. *Under the assumption (33)-(36), and the null hypothesis H_0'', we have*

$$\breve{T}_{F,n} \to^d N(0,1).$$

Theorem 2.1 is readily attained by slightly modifying the proof of Theorem 2.1 of the main paper. We impose conditions on each cell, which generalize those in Theorem 2.1 in the paper to the case of factorial designs.

3 Additional Simulation Results

We conduct additional simulations to assess the finite sample performance of the proposed MDD-based tests.

3.1 Conditional mean independence

Example **3.1**. We consider the simple linear model:

$$Y_i = X_i^T\beta + \epsilon_i, \quad i = 1,2,\dots,n,$$

where $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ is a p-dimensional vector of covariates, $\beta = (\beta_1, \dots, \beta_p)^T$ is the regression coefficient and ϵ_i is the error that is independent of X_i. The covariates are generated from the following model:

$$x_{ij} = (\varsigma_{ij} + \varsigma_{i0})/\sqrt{2}, \quad j = 1,2,...,p, \tag{37}$$

where $(\varsigma_{i0}, \varsigma_{i1}, \dots, \varsigma_{ip})^T \overset{i.i.d.}{\sim} N(0, I_{p+1})$. Hence the covariates are strongly correlated with the pairwise correlation equal to 0.5.

Under the null, $\beta = \mathbf{0}_{p\times 1}$; under both the sparse and non-sparse alternative, we fix $|\beta|_p = 0.06$ as in Example 4.1 of the paper. We set $n = 100$, $p = 50, 100, 200$, and also consider three configurations for the error ϵ: $N(0,1)$, t_3, and $\chi_1^2 - 1$. Table 1 presents the empirical sizes and powers of the proposed test and the ZC test for the cases of $p < n$, $p = n$ and $p > n$. The empirical sizes of both tests are reasonably close to the 10% nominal level for three different error distributions. At the 5% significance level, however, we see both tests have slightly inflated rejection probabilities under the null, which is similar to what we observe in Example 4.1. For the empirical powers, our test is highly comparable to ZC test under the simple linear model.

EXAMPLE **3.2**. This example specifies another non-linear relationship between Y and X. The model is given by

$$Y_i = \frac{p}{\sqrt{\sum_{j=1}^{p} \beta_j x_{ij}^2}} + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{38}$$

where the covariates $X_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$ are generated according to (37). Again we consider three configurations for the error, namely $N(0,1)$, t_3, and $\chi^2 - 1$ respectively.

Under the non-sparse alternative, we set $\beta_j = 1$ for $j = 1, 2, ..., n/2$. For sparse alternative, $\beta_j = 1$ for $j = 1, 2, ..., 5$. The configurations for p and n are the same as Example 3.1.

Table 2 summarizes the empirical powers for Example 3.2. Similar to Example 4.2 of the paper, ZC test, which is designed for linear model, exhibits very little power under model mis-specification. The powers for ZC test are even below the nominal level under sparse alternatives. We notice that our proposed test has gradually decreasing powers as dimension increases under both sparse and non-sparse alternatives, which might be explained by the fact that the covariates enter into the denominator of the model and the signals are weakened by increasing the denominator in (38) as the dimension increases. Overall, the power performance is reasonably good for our proposed model-free test.

3.2 Conditional quantile independence

EXAMPLE **3.3**. This example considers a simple linear model with heteroscedesticity:

$$Y_i = X_i^T \beta + \left(1 + X_i^T \beta\right)^2 \epsilon_i, \quad i = 1, 2, \ldots, n,$$

where $X_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$ is a p dimensional vector and ϵ_i is the error independent of X_i. All the other configurations are the same as Example 3.1.

Table 3 shows the empirical sizes and powers for different configurations in Example 3.3. The sizes are generally precise at 10% level and slightly inflated at 5% level, and they do not seem to depend on the error distribution much. The empirical powers apparently depend on the error distribution. For $N(0,1)$, t_3 and Cauchy(0,1) error, they have higher powers at $\tau = 0.5$, 0.75 than that at $\tau = 0.25$. The powers are almost 1 in the non-sparse alternatives and suffer big reduction under sparse alternatives. In addition, the powers at $\tau = 0.75$ under sparse case are higher than that at $\tau = 0.5$. In contrast, the powers for $\chi_1^2 - 1$ error configuration perform the best at $\tau = 0.25$ and 0.75, while suffer a great power loss at $\tau = 0.5$. Moreover, the powers at $\tau = 0.25$ under sparse alternatives are higher than that at $\tau = 0.75$. These phenomenon should be related to the error distribution. Notice that normal, student's t and Cauchy distributions are all symmetric; while $\chi_1^2 - 1$ is skewed. The performance of MDD-based test for conditional mean is very similar to that for Example 4.4.

As we observe some size distortion in Table 4 for Example 4.4 of the paper, We further apply the wild bootstrap to approximate the finite sample distribution of $T_{Q,n}$ in the same way as the studentized bootstrap statistic T_n^* used in the conditional mean dependence testing. According to Table 4, the bootstrap helps to reduce the size distortion substantially and the size corresponding to the bootstrap approximation is fairly accurate. Similar results are observed for 10% level and are not presented. We also tried the wild bootstrap for the case of $\tau = 0.5$, and nonzero βs, which falls under our null hypothesis. There is some reduction in size distortion in this case (results not shown) but the size of the bootstrap based test is still quite inflated for large p (say, $p = 100, 200$). This again points to the importance of our local quantile independence assumption (Assumption 3.2), which is violated in this case and seems crucial to the validity of our normal and bootstrap approximation.

3.3 Conditional mean independence under factorial designs

In this subsection, we evaluate the finite sample performance of the proposed test for conditional mean independence under factorial designs in comparison with the ZC test.

EXAMPLE **3.4**. Consider the factorial design with non-linear models in each cell,

$$Y_{ijk} = \mu_{ij} + \sqrt{\sum_{h=1}^{p} \beta_h x_{ijk,h}^2} + \epsilon_{ijk}, \quad k = 1, 2, \ldots, n_{ij},$$

where $(\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}) = (1, 3, 3, 4)$, $n_{11} = n_{12} = n_{21} = n_{22} = n$ and $\epsilon_{ijk} \sim^{i.i.d} N(0, 4)$. We consider two distributions for the covariates X_{ijk} in each cell:
Case 1: X_{ijk} is generated independently from the moving average model as in Example 4.1 of the paper.

$$x_{ijk,h} = \alpha_{ij1} z_{ijk,h} + \alpha_{ij2} z_{ijk,(h+1)} + \cdots + \alpha_{ijT_{ij}} z_{ijk,(h+T_{ij}-1)} + \mu_{ij,h},$$

for $i = 1, 2$, $j = 1, 2$, and $k = 1, \ldots, n_{ij}$. By choosing $(T_{11}, T_{12}, T_{21}, T_{22}) = (10, 15, 20, 25)$, the dependence structure in each cell is different from the others.
Case 2: $x_{ijk,h} = (\varsigma_{ijk,h} + \varsigma_{ijk,0})/\sqrt{2}$, where $(\varsigma_{ijk,0}, \varsigma_{ijk,1}, \ldots, \varsigma_{ijk,p})^T \overset{i.i.d}{\sim} N(0, I_{p+1})$ for $i = 1, 2$, $j = 1, 2$, $k = 1, \ldots, n_{ij}$, and $h = 1, \ldots, p$.

We consider $n_{ij} = 30, 50, 70$, and $p = 100, 150, 200$. For the sparse and non-sparse alternatives, we keep $|\beta|_p = 0.06$.

Table 5 summarizes the results for both cases. There is slight size distortion at 5% significance level. Both tests have satisfactory power in Case 1. The ZC test, however, suffers from a significant power loss in Case 2 in part due to model mis-specification. In contrast, we observe that our proposed test retains the high power under the non-sparse alternative and the power increases rapidly as sample size increases. Although our test exhibits less power under sparse alternatives, it still outperforms ZC test in power by a noticeable amount.

EXAMPLE **3.5**. This example is based on the two factor balanced design with two levels for each factor which has been considered in Zhong and Chen (2011),

$$Y_{ijk} = \mu_{ij} + X_{ijk}^T\beta + \epsilon_{ijk}, \quad k = 1, 2, \ldots, n_{ij},$$

where $(\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}) = (1, 3, 3, 4)$, and $n_{11} = n_{12} = n_{21} = n_{22}$. Within each cell, X_{ijk} is generated independently from the following moving average model,

$$x_{ijk,h} = \alpha_{ij1}z_{ijk,h} + \alpha_{ij2}z_{ijk,(h+1)} + ... + \alpha_{ijT_{ij}}z_{ijk,(h+T_{ij}-1)} + \mu_{ij,h},$$

for $i = 1, 2,\ j = 1, 2,$ and $k = 1, \ldots, n_{ij}$. By choosing $(T_{11}, T_{12}, T_{21}, T_{22}) = (10, 15, 20, 25)$, the dependence structure in each cell is different from each other. We also consider two error distributions for ϵ_i: $N(0, 4)$ and centralized gamma distribution with shape parameter 1 and scale parameter 0.5. Consider $n_{ij} = 30, 50, 70$, and $p = 100, 150, 200$. For both the sparse and non-sparse alternatives, we keep $|\beta|_p = 0.06$.

The simulation results are presented in Table 6. The size phenomenon is similar to what we observed in other examples. The powers for both tests in the non-sparse and sparse alternatives are quite high in most cases. We observe that the two tests are very much comparable in terms of size and power under the simple linear model with factorial design.

3.4 Further comparison with existing methods

In this subsection, we further compare our test for conditional mean independence with some recent ones developed by McKeague and Qian (2015) and the discussants for the latter paper. McKeague and Qian (2015) proposed an adaptive re-sampling test (MQ test, hereafter) for detecting the existence of significant predictors in a linear regression model. They test $H_{10} : \text{cov}(Y, x_j) = 0,\ j = 1, \cdots, p$ versus $H_{11} : \text{cov}(Y, x_j) \neq 0,$ for at least one $j = 1, \cdots, p$. Under finite second moment assumptions for Y and X, our null hypothesis H_0' implies their H_{10} but there are situations where their null hypothesis holds but Y is marginally conditionally mean dependent on x_j for some j, i.e., H_a' holds; see below for an example. MQ's test is based on the estimated marginal regression coefficient of the selected predictor, which has the strongest marginal correlation with the response. It was shown in MQ that their test's limiting distribution over the parameter space has a discontinuity at the origin, which causes the inconsistency of naive bootstrap. MQ proposed a modified bootstrap method that is adaptive to the nonregular behavior of their test statistic by introducing a thresholding parameter, the choice of which is based on another level of bootstrap. While their method is well suited to identify the most significant predictor, it is computationally expensive due to the use of double bootstrap, and its applicability to large p case is theoretically unknown and computationally prohibitive.

In the discussions of McKeague and Qian (2015), Zhang and Laber (2015) and Chatterjee and Lahiri (2015) have proposed alternative tests that overcome some limitation of MQ test either computationally or methodologically. In particular, Zhang and Laber (2015) proposed to use the largest marginal t-statistic in magnitude to achieve scale-invariance, and use a parametric bootstrap procedure to obtain the critical values. Their test can be implemented much faster than MQ test. They also proposed an adaptive parametric bootstrap procedure that can be adaptive to unknown level of sparsity. The second test by Zhang and Laber (2015) is, however, a bit more complex to implement and computationally more expensive and it does not seem to lead to a substantial power gain as seen from Table 2 in their paper, so we decided not to include it into our comparison. It is worth mentioning that the simpler test procedure proposed in Zhang and Laber (2015) is based on direction estimation of the asymptotic variance in Theorem 1 of McKeague and Qian (2015), which requires stronger independence (rather than uncorrelatedness) conditions of the regression error with covariates. Chatterjee and Lahiri (2015) proposed a L_2 type test statistic via aggregating the marginal t-statistic using L_2 norm, and the limiting null distribution can be well approximated by the naive bootstrap. Hence their test does not suffer the non-continuity issue, does not require the selection of a tuning parameter, and is computationally simple.

Following the suggestion of associate editor and a referee, we shall compare the performance of our test (MDD), with MQ test, ZL test (Zhang and Laber's parametric bootstrap test corresponding to $\widehat{\xi}_n$ in their paper), CL test (Chatterjee and Lahiri 2015). We do not aim to provide a comprehensive simulation study to compare these four tests as the latter three tests have been partially compared in Zhang and Lahiri (2015), and Chatterjee and Lahiri (2015). Also our MDD-based test targets a different null hypothesis so in some situations our test is not directly comparable to the other three. This point will be highlighted via a numerical example below.

Example **3.6**. *Consider the following linear model,*

$$Y_i = X_i^T \beta + \epsilon_i, \quad i = 1, ..., n$$

where $X_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$ *is generated from multivariate Gaussian with zero means and covariance* $\Sigma = (\sigma_{ij})_{i,j=1}^p$*, where* $\sigma_{ij} = \rho^{|i-j|}$*; we consider four cases* $\rho = 0.1, 0.5, 0.8, -0.5$ *respectively. Here* $\beta = (\beta_1, ..., \beta_p)^T$ *is the regression coefficient and* ϵ_i *is the error that is independent of* X_i *and also generated from i.i.d* $N(0, 1)$.

Example **3.7**. *Consider the following two models,*

i) $Y_i = X_i^T \beta + \epsilon_i$, $X_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$ *is generated from i.i.d N(0,1);*

ii) $Y_i = g(X_i)^T \beta + \epsilon_i$, *where* $g(x) = (x_1^2, ..., x_p^2)^T$, X_i *is the same as in case i);*

for $i = 1, 2, ..., n$. *Here* $\beta = (\beta_1, ..., \beta_p)^T$ *is the regression coefficient and* ϵ_i *is the error that is independent of* X_i *and also generated from i.i.d* $N(0, 1)$.

Following the setting in Chatterjee and Lahiri (2015), we consider three scenarios: under H_0, $\beta = \mathbf{0}_{p\times 1}$; under non-sparse H_a, $\beta = p^{-1/2} \cdot c \cdot \mathbf{1}_{p\times 1}$; under sparse H_a, $\beta = (c, 0, 0, ..., 0)$. We choose $c = 0.2$ for Example 3.6 and $c = 0.5$ for Example 3.7; we fix sample size $n = 200$ and $p = 10, 50, 150$. For MQ test, we use fixed thresholds $\lambda_n = (2, 4, 4.5, 5, 10)$ as the use of double bootstrap to select thresholding parameter is too expensive for $p = 50, 150$ in our simulations. It is worth noting that in Example 3.7(ii), $cov(Y, x_i) = 0$ for $i = 1, ..., p$, but $\mathbb{E}(Y|x_i) \neq \mathbb{E}(Y)$. Therefore, this model falls under the alternative hypothesis for our proposed test, but under the null for ZL, CL and MQ tests.

From Table 7, we observe that all tests (mdd, ZL, CL and MQ with $\lambda_n = 5, 10$) have quite reasonable sizes for different ρs; mdd and CL tests have higher power under the dense alternative since they are both L_2 type statistics, while the other two L_∞ tests: ZL and MQ test perform better under sparse alternatives. The power generally increases as the correlation ρ increases and the positive dependence within the covariates seems to enhance the power under both dense and sparse alternatives. When $\rho = -0.5$, none of the tests have good power under the dense alternative. Notice that MQ's performance is fairly sensitive to the threshold value used, thus a double bootstrap approach will hopefully provide more stable size/power at the expense of heavy computation. Overall, the performance of mdd is on a par with that of CL, and the performance of ZL and MQ with thresholding parameter $\lambda_n = 5, 10$ are comparable.

Table 8 further compares the four tests when the covariates are independent. It appears that the size accuracy is similar to that reported for Example 3.6. For Case i), the powers of mdd and CL tests are highest under dense alternatives as expected, whereas the powers of ZL and MQ are higher under sparse alternatives. It can be seen that the power of mdd test under dense and sparse alternatives are roughly the same for both i) and ii), which can be explained by the fact that for independent covariates, the power of our MDD-based test is proportional to the signal strength $|\beta|$, which is fixed at $|\beta| = 0.5$ here. Under the non-linear model in case ii), the power of mdd test is satisfactory for small p and it decreases as the dimension increases. By contrast, the ZL and CL tests exhibit certain size inflation when $p = 10$. Again MQ's size is sensitive to the thresholding value used and seems accurate for $\lambda_n = 5, 10$. This example highlights the fact that our mdd test targets marginal conditional mean dependence, whereas MQ, ZL and CL are developed for detecting marginal uncorrelatedness. Overall, the results demonstrate that our proposed test is highly competitive in the case when mdd can be compared to MQ test and variations (see Example 3.6), and when they are not directly comparable (see Example 3.7 ii), the mdd test is able to detect marginal conditional mean dependence and perform reasonably well.

4 Data illustration

We apply the proposed test to the data set described in Lkhagvadorj et al. (2009) and Zhong and Chen (2011). The data is from clinical outcomes in a randomized factorial design experiment, where 24 six-month-old Yorkshire gilts from a line selected for high feed efficiency were used. All the gilts were genotyped for the melanocortin-4 receptor gene (MC4R) variant at position 298; and 12 gilt homozygous for N298 and 12 gilt homozygous for D298. Two feed treatments are randomly assigned to each group of the MC4R genotype. One is ad libitum a crude protein standard swine diet, the other one is fasting diet, which leads to decreased body weight, backfat, and serum urea concentration and increased serum non-esterified fatty acid. The genotype and feed treatments are the two factors in the randomized complete factorial design. A total of six pigs were used for each combination of genotype and feed treatment across the four blocks. The goal of this study is to identify conditional mean dependence of triiodothyronine (T_3) measurements on gene sets, as described in Zhong and Chen (2011), where T_3 is a vital thyroid hormone that increases the metabolic rate, protein synthesis, and stimulates breakdown of cholesterol.

The gene sets mentioned above are defined by the Gene Ontology (GO term). The dataset includes the gene expression values for 24,123 genes in the gilts' liver tissues. These genes are then classified into different gene sets (GO term) according to their biological functions among three categories: cellular component, molecular function and biological process. The dataset included 6176 GO terms in total, where each of them contains some of the genes from the 24,123 genes collected. Our response is the T_3 measurements in the blood. We aim to find the gene sets (GO terms) that have an impact on the T_3 measurements in terms of conditional mean after accounting for the two factors we have in the design.

We use i, j, k to denote the indices for feed treatment, MC4R genotype and observations, respectively. Our response is Y_{ijk}, the T_3 measurement for the kth pig in ith feed treatment and jth genotype. We want to test whether the lth GO term, denoted as X_{ijk}^l, contributes to the conditional mean of the T_3 measurement or not. The following four different factorial designs have been considered in Zhong and Chen (2011),

$$\begin{aligned}
\text{Design I:} \quad & \mathbb{E}(Y_k) = \alpha + (X_k^l)^T\beta, \quad k = 1, ..., 24; \\
\text{Design II:} \quad & \mathbb{E}(Y_{ik}) = \alpha + \mu_i + (X_{ik}^l)^T\beta, \quad k = 1, ..., 12; \\
\text{Design III:} \quad & \mathbb{E}(Y_{jk}) = \alpha + \gamma_j + (X_{jk}^l)^T\beta, \quad k = 1, ..., 12; \\
\text{Design IV:} \quad & \mathbb{E}(Y_{ijk}) = \alpha + \mu_i + \gamma_j + (\mu\gamma)_{ij} + (X_{ijk}^l)^T\beta, \quad k = 1, ..., 6;
\end{aligned}$$

for $i = 1, 2$, $j = 1, 2$, and $l = 1, .., 6176$, which correspond to 6176 total gene sets (GO terms). To avoid the linear model assumption, we test instead the following hypothesis for each of the designs above

$$\begin{aligned}
\text{Design I:} \quad & \mathbb{E}(Y_k|X_k^l) = \mathbb{E}(Y_k), \quad k = 1, ..., 24; \\
\text{Design II:} \quad & \mathbb{E}(Y_{ik}|X_{ik}^l) = \mathbb{E}(Y_{ik}), \quad k = 1, ..., 12; \\
\text{Design III:} \quad & \mathbb{E}(Y_{jk}|X_{jk}^l) = \mathbb{E}(Y_{jk}), \quad k = 1, ..., 12; \\
\text{Design IV:} \quad & \mathbb{E}(Y_{ijk}|X_{ijk}^l) = \mathbb{E}(Y_{ijk}), \quad k = 1, ..., 6;
\end{aligned}$$

for $i = 1, 2$, $j = 1, 2$, and $l = 1, .., 6176$. Note that the dimensions of the GO terms X_{ijk}^l range from 1 to 5158. The MDD-based test is implemented for the GO terms with dimension $p_l \geq 5$. The remaining GO terms are tested using a simple F-test.

We draw the histograms of p-values for all the GO terms as shown in Figure 1. From the histograms we can observe that the p-values from Design I and Design III are similar; while the results from Design II has lower portion of small p-values than the other three designs; Design IV has more large p-values relatively to others.

We then use the Benjamini–Hochberg step-up procedure to control the false discovery rate at level $\alpha = 0.05$. Namely, for m hypothesis, we find the largest integer k such that $P_{(k)} \leq \frac{k}{m}\alpha$, where $P_{(k)}$ is the ordered p-values from the total m hypothesis tests. Then we reject the null hypothesis for all $H_{(i)}$ for $i = 1, \ldots, k$, where $H_{(i)}$ is the null hypothesis corresponds to the ordered p-value $P_{(i)}$. After controlling the false discovery rate for the p-values, we find 5 gene sets that are signified as significant under all four designs. They are GO:0032012, GO:0005086, GO:0043536, GO:0005161 and GO:0045095. Compared with Zhong and Chen (2011), gene sets GO:0032012, GO:0005086 are among the three significant gene sets they found using their method. To demonstrate the non-linear dependence of the GO terms we found, we choose GO:0043536, which contains 5 genes, and present the scatter plot of the response versus the five gene expression values and also fit a local polynomial regression line (LOESS) with degree 2 in Figure 2. From the plot, we do observe some non-linear dependence between the response and the five genes.

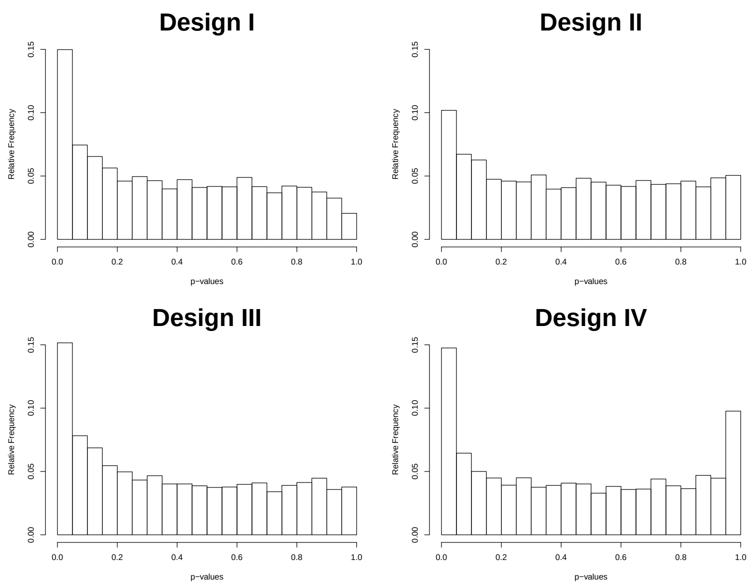

Figure 1: Histograms of the p-values on all gene sets

Our null considers the conditional mean independence of the T_3 measurement with the gene sets, which includes the null hypothesis in Zhong and Chen (2011) under the linear model assumption. Thus the above finding seems reasonable since we expect to detect more significant gene sets as our test is more powerful than ZC test when the gene set contributes to the mean of T_3 in a non-linear fashion as shown in our simulations.

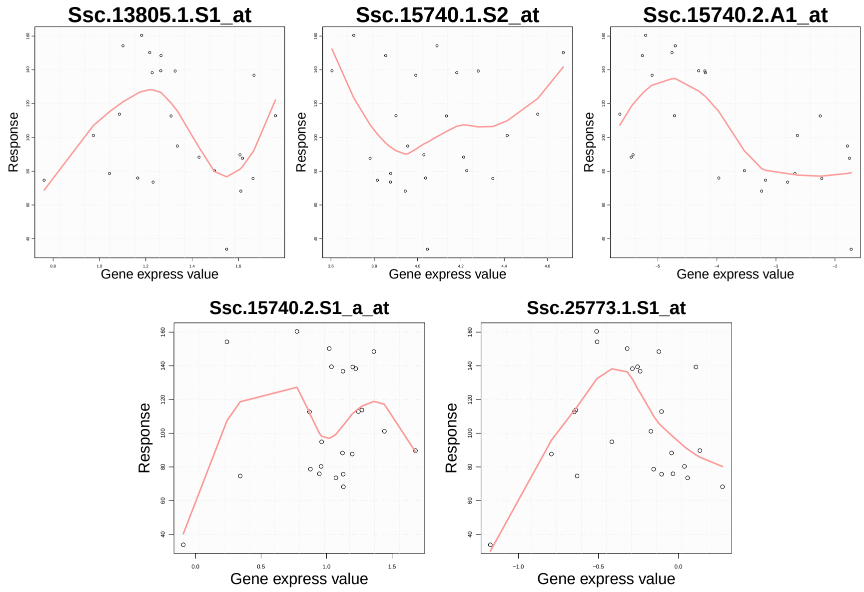

Figure 2: Scatter plots of the T_3 measurements versus five gene express values in the GO term 0043536; the red lines correspond to the LOESS fitting.

References

[1] Chatterjee, A. and Lahiri, S. N. (2015). Comment: An Adaptive Resampling Test for Detecting the Presence of Significant Predictors. *J. Amer. Statist. Assoc.* **110** 1434-1438.

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

[3] Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14** 1-16.

[4] Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application.* Academic Press, New York.

[5] Mckeague, Ian W. and Qian, Min (2015). An Adaptive Resampling Test for Detecting the Presence of Significant Predictors. *J. Amer. Statist. Assoc.* **110** 1422-1433.

[6] Park, T., Shao, X. and Yao, S. (2014). Partial martingale difference correlation. *Preprint.*

[7] Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* Wiley, New York.

[8] Shao, X. and Zhang, J. (2014). Martingale difference correlation and its use in high dimensional variable screening. *J. Amer. Statist. Assoc.* **109** 1302-1318.

[9] Székely, G. J. and Rizzo, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* To appear.

[10] Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007). Measuring and testing independence by correlation of distances. *Ann. Statist.* **35** 2769-2794.

[11] Zhang, Y. and Laber, Eric B. (2015). Comment: An Adaptive Resampling Test for Detecting the Presence of Significant Predictors. *J. Amer. Statist. Assoc.* **110** 1451-1454.

[12] Zhong, P. S. and Chen, S. X. (2011). Testing for high-dimensional regression coefficients with factorial designs. *J. Amer. Statist. Assoc.* **106** 260-274.

Table 1: Empirical sizes and powers of the MDD-based test for conditional mean independence and the ZC test at significance levels 5% and 10% for Example 3.1.

			mdd		ZC	
error	case	p	5%	10%	5%	10%
$N(0,1)$	H_0	50	0.078	0.112	0.075	0.098
		100	0.075	0.111	0.078	0.109
		200	0.065	0.091	0.062	0.089
	non-sparse H_a	50	0.575	0.635	0.595	0.640
		100	0.842	0.879	0.862	0.887
		200	0.982	0.989	0.986	0.992
	sparse H_a	50	0.189	0.243	0.199	0.248
		100	0.183	0.230	0.191	0.238
		200	0.186	0.224	0.187	0.224
t_3	H_0	50	0.084	0.110	0.080	0.104
		100	0.075	0.106	0.072	0.097
		200	0.082	0.112	0.063	0.091
	non-sparse H_a	50	0.619	0.656	0.620	0.669
		100	0.848	0.877	0.829	0.865
		200	0.949	0.956	0.954	0.957
	sparse H_a	50	0.234	0.278	0.226	0.275
		100	0.218	0.254	0.212	0.253
		200	0.226	0.269	0.215	0.262
$\chi_1^2 - 1$	H_0	50	0.083	0.114	0.083	0.110
		100	0.061	0.096	0.058	0.086
		200	0.058	0.084	0.047	0.078
	non-sparse H_a	50	0.746	0.791	0.744	0.788
		100	0.908	0.930	0.916	0.932
		200	0.978	0.988	0.978	0.987
	sparse H_a	50	0.272	0.319	0.264	0.317
		100	0.265	0.314	0.256	0.307
		200	0.250	0.313	0.246	0.297

Table 2: Empirical powers of the MDD-based test for conditional mean independence and the ZC test at significance levels 5% and 10% for Examples 3.2.

			mdd		ZC	
error	case	p	5%	10%	5%	10%
$N(0,1)$	non-sparse H_a	50	0.997	0.999	0.077	0.111
		100	0.986	1.000	0.092	0.126
		200	0.973	0.997	0.120	0.155
	sparse H_a	50	0.933	0.968	0.020	0.036
		100	0.874	0.933	0.020	0.040
		200	0.825	0.911	0.019	0.036
t_3	non-sparse H_a	50	0.789	0.888	0.074	0.109
		100	0.685	0.821	0.073	0.106
		200	0.593	0.750	0.095	0.119
	sparse H_a	50	0.930	0.968	0.019	0.037
		100	0.876	0.932	0.020	0.036
		200	0.823	0.910	0.019	0.035
$\chi_1^2 - 1$	non-sparse H_a	50	0.854	0.912	0.095	0.116
		100	0.777	0.874	0.095	0.122
		200	0.727	0.833	0.079	0.108
	sparse H_a	50	0.933	0.967	0.019	0.033
		100	0.874	0.929	0.021	0.036
		200	0.822	0.912	0.017	0.036

Table 3: Empirical sizes and powers of the MDD-based test for conditional quantile independence at significance levels 5% and 10% for Example 3.3.

τ	case	p	$N(0,1)$ 5%	$N(0,1)$ 10%	t_3 5%	t_3 10%	χ_1^2-1 5%	χ_1^2-1 10%	Cauchy(0,1) 5%	Cauchy(0,1) 10%
0.25	H_0	50	0.084	0.112	0.058	0.077	0.074	0.099	0.073	0.094
		100	0.073	0.104	0.067	0.094	0.071	0.099	0.069	0.101
		200	0.064	0.085	0.058	0.095	0.069	0.094	0.062	0.082
	non-sparse H_a	50	0.134	0.173	0.133	0.167	0.934	0.948	0.144	0.193
		100	0.148	0.192	0.164	0.199	0.952	0.965	0.225	0.271
		200	0.168	0.218	0.185	0.211	0.874	0.901	0.285	0.331
	sparse H_a	50	0.095	0.123	0.081	0.107	0.615	0.663	0.082	0.106
		100	0.091	0.126	0.081	0.101	0.612	0.670	0.082	0.120
		200	0.082	0.109	0.083	0.114	0.590	0.646	0.076	0.109
0.5	H_0	50	0.083	0.106	0.076	0.097	0.064	0.089	0.068	0.099
		100	0.072	0.100	0.063	0.087	0.062	0.091	0.057	0.092
		200	0.073	0.100	0.053	0.073	0.063	0.088	0.071	0.095
	non-sparse H_a	50	0.519	0.575	0.464	0.530	0.107	0.128	0.333	0.397
		100	0.764	0.819	0.722	0.774	0.122	0.154	0.568	0.636
		200	0.930	0.943	0.895	0.920	0.208	0.251	0.802	0.844
	sparse H_a	50	0.163	0.209	0.153	0.190	0.077	0.102	0.120	0.160
		100	0.169	0.206	0.141	0.180	0.069	0.092	0.124	0.162
		200	0.175	0.212	0.139	0.160	0.071	0.091	0.121	0.160
0.75	H_0	50	0.067	0.098	0.080	0.108	0.065	0.085	0.062	0.084
		100	0.057	0.085	0.078	0.098	0.073	0.095	0.076	0.111
		200	0.064	0.095	0.063	0.085	0.065	0.088	0.069	0.095
	non-sparse H_a	50	0.929	0.953	0.839	0.877	0.352	0.409	0.606	0.674
		100	0.995	0.998	0.967	0.977	0.596	0.649	0.850	0.886
		200	1.000	1.000	1.000	1.000	0.833	0.875	0.968	0.976
	sparse H_a	50	0.415	0.482	0.347	0.394	0.118	0.153	0.217	0.260
		100	0.418	0.473	0.342	0.392	0.131	0.162	0.213	0.255
		200	0.390	0.447	0.324	0.370	0.123	0.157	0.210	0.262
mean	H_0	50	0.071	0.103	0.069	0.101	0.067	0.098	0.052	0.097
		100	0.074	0.101	0.075	0.097	0.071	0.101	0.060	0.103
		200	0.067	0.094	0.078	0.105	0.066	0.088	0.054	0.096
	non-sparse H_a	50	0.414	0.466	0.260	0.306	0.206	0.280	0.075	0.117
		100	0.549	0.602	0.323	0.381	0.278	0.357	0.096	0.139
		200	0.605	0.667	0.408	0.464	0.390	0.487	0.115	0.159
	sparse H_a	50	0.181	0.233	0.136	0.180	0.106	0.140	0.056	0.106
		100	0.171	0.207	0.119	0.158	0.105	0.142	0.064	0.109
		200	0.178	0.211	0.138	0.171	0.100	0.141	0.059	0.101

Table 4: Size comparison for the proposed test using normal approximation (mdd) and the wild bootstrap approximation for Example 4.4 of the paper at the 5 % nominal level.

		$N(0,1)$		t_3		Cauchy(0,1)		$\chi_1^2 - 1$	
τ	p	mdd	Boot	mdd	Boot	mdd	Boot	mdd	Boot
	50	0.084	0.061	0.058	0.043	0.073	0.054	0.074	0.055
0.25	100	0.073	0.053	0.067	0.048	0.069	0.051	0.071	0.053
	200	0.064	0.043	0.058	0.042	0.062	0.035	0.069	0.045
	50	0.083	0.057	0.076	0.054	0.068	0.044	0.064	0.048
0.50	100	0.072	0.057	0.063	0.047	0.057	0.046	0.062	0.046
	200	0.073	0.054	0.053	0.040	0.071	0.051	0.063	0.045
	50	0.067	0.051	0.080	0.059	0.062	0.050	0.065	0.041
0.75	100	0.057	0.044	0.078	0.052	0.076	0.052	0.073	0.056
	200	0.064	0.050	0.063	0.042	0.069	0.051	0.065	0.047

Table 5: Empirical sizes and powers of the MDD-based test for conditional mean independence and the ZC test at significance levels 5% and 10% for Example 3.4.

			Case 1				Case 2			
			mdd		ZC		mdd		ZC	
	n	p	5%	10%	5%	10%	5%	10%	5%	10%
H_0	30	100	0.054	0.098	0.061	0.099	0.074	0.115	0.083	0.122
	30	150	0.056	0.105	0.061	0.117	0.060	0.102	0.067	0.114
	30	200	0.055	0.099	0.055	0.114	0.064	0.108	0.063	0.107
	50	100	0.064	0.104	0.057	0.103	0.062	0.098	0.067	0.102
	50	150	0.054	0.098	0.062	0.094	0.060	0.094	0.058	0.094
	50	200	0.067	0.110	0.062	0.112	0.071	0.108	0.070	0.105
	70	100	0.068	0.114	0.058	0.109	0.066	0.100	0.069	0.106
	70	150	0.075	0.109	0.073	0.112	0.065	0.102	0.067	0.106
	70	200	0.061	0.098	0.063	0.099	0.058	0.104	0.070	0.107
non-sparse H_a	30	100	0.726	0.813	0.744	0.815	0.412	0.546	0.196	0.237
	30	150	0.566	0.674	0.577	0.682	0.459	0.596	0.226	0.266
	30	200	0.452	0.575	0.464	0.582	0.508	0.645	0.227	0.266
	50	100	0.972	0.986	0.975	0.984	0.788	0.908	0.194	0.236
	50	150	0.889	0.934	0.891	0.927	0.871	0.959	0.187	0.233
	50	200	0.786	0.855	0.783	0.849	0.903	0.975	0.251	0.281
	70	100	0.999	0.999	0.989	0.997	0.991	0.998	0.189	0.230
	70	150	0.978	0.986	0.963	0.977	0.998	0.999	0.200	0.236
	70	200	0.948	0.971	0.918	0.954	1.000	1.000	0.210	0.242
sparse H_a	30	100	0.369	0.478	0.389	0.478	0.150	0.206	0.102	0.145
	30	150	0.283	0.396	0.301	0.395	0.128	0.185	0.091	0.127
	30	200	0.229	0.326	0.239	0.339	0.159	0.233	0.104	0.160
	50	100	0.694	0.782	0.677	0.756	0.194	0.267	0.097	0.145
	50	150	0.523	0.622	0.503	0.617	0.172	0.240	0.097	0.133
	50	200	0.413	0.535	0.404	0.512	0.188	0.260	0.093	0.139
	70	100	0.875	0.918	0.838	0.882	0.251	0.356	0.113	0.145
	70	150	0.720	0.805	0.658	0.750	0.224	0.340	0.108	0.145
	70	200	0.611	0.699	0.545	0.638	0.233	0.312	0.111	0.161

Table 6: Empirical sizes and powers of the MDD-based test for conditional mean independence and the ZC test at significance levels 5% and 10% for Example 3.5.

			Normal error				Gamma error			
			mdd		ZC		mdd		ZC	
	n	p	5%	10%	5%	10%	5%	10%	5%	10%
H_0	30	100	0.054	0.098	0.061	0.099	0.055	0.113	0.055	0.100
	30	150	0.056	0.105	0.061	0.117	0.055	0.100	0.047	0.093
	30	200	0.055	0.099	0.055	0.114	0.052	0.096	0.046	0.090
	50	100	0.064	0.104	0.057	0.103	0.064	0.108	0.053	0.105
	50	150	0.054	0.098	0.062	0.094	0.046	0.085	0.051	0.086
	50	200	0.067	0.110	0.062	0.112	0.057	0.098	0.052	0.092
	70	100	0.068	0.114	0.058	0.109	0.069	0.118	0.059	0.107
	70	150	0.075	0.109	0.073	0.112	0.044	0.090	0.041	0.094
	70	200	0.061	0.098	0.063	0.099	0.048	0.080	0.048	0.090
non-sparse H_a	30	100	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999
	30	150	0.996	1.000	0.998	1.000	0.989	0.992	0.993	0.995
	30	200	0.994	0.998	0.995	0.998	0.982	0.989	0.987	0.993
	50	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	150	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	70	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	70	150	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	70	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
sparse H_a	30	100	0.437	0.550	0.485	0.573	0.514	0.616	0.540	0.621
	30	150	0.341	0.452	0.382	0.478	0.323	0.417	0.331	0.434
	30	200	0.265	0.370	0.304	0.397	0.261	0.364	0.263	0.381
	50	100	0.819	0.877	0.790	0.849	0.817	0.875	0.795	0.865
	50	150	0.644	0.754	0.621	0.729	0.652	0.729	0.621	0.712
	50	200	0.532	0.640	0.505	0.608	0.573	0.688	0.561	0.658
	70	100	0.938	0.969	0.918	0.947	0.952	0.970	0.916	0.945
	70	150	0.847	0.905	0.792	0.850	0.864	0.907	0.805	0.865
	70	200	0.735	0.816	0.664	0.755	0.757	0.821	0.672	0.764

Table 7: Empirical sizes and powers from four tests for Example 3.6

$(\rho = 0.1)$	p	mdd	ZL	CL	MQ (λ_n)				
					2	4	4.5	5	10
H_0	10	0.045	0.035	0.030	0.335	0.060	0.040	0.040	0.040
	50	0.085	0.045	0.050	0.660	0.085	0.060	0.045	0.045
	150	0.020	0.085	0.015	0.835	0.195	0.100	0.080	0.075
non-sparse H_a	10	0.565	0.345	0.570	0.695	0.430	0.405	0.380	0.350
	50	0.250	0.105	0.220	0.665	0.200	0.150	0.120	0.095
	150	0.145	0.095	0.135	0.845	0.315	0.165	0.120	0.080
sparse H_a	10	0.415	0.490	0.450	0.830	0.540	0.510	0.500	0.495
	50	0.185	0.355	0.155	0.895	0.415	0.360	0.340	0.320
	150	0.075	0.300	0.060	0.880	0.490	0.330	0.295	0.230
$(\rho = 0.5)$	p	mdd	ZL	CL	MQ (λ_n)				
					2	4	4.5	5	10
H_0	10	0.040	0.020	0.035	0.285	0.045	0.035	0.030	0.030
	50	0.065	0.035	0.045	0.610	0.075	0.050	0.035	0.020
	150	0.025	0.025	0.020	0.805	0.145	0.070	0.050	0.040
non-sparse H_a	10	0.975	0.900	0.980	0.985	0.960	0.945	0.935	0.900
	50	0.905	0.405	0.905	0.915	0.770	0.615	0.510	0.425
	150	0.695	0.195	0.740	0.910	0.665	0.420	0.300	0.205
sparse H_a	10	0.485	0.545	0.455	0.815	0.595	0.565	0.545	0.540
	50	0.195	0.345	0.150	0.865	0.430	0.360	0.330	0.315
	150	0.050	0.265	0.050	0.890	0.455	0.350	0.270	0.225
$(\rho = 0.8)$	p	mdd	ZL	CL	MQ (λ_n)				
					2	4	4.5	5	10
H_0	10	0.075	0.030	0.035	0.175	0.085	0.060	0.045	0.045
	50	0.055	0.035	0.040	0.500	0.100	0.080	0.065	0.055
	150	0.045	0.020	0.025	0.705	0.070	0.040	0.030	0.030
non-sparse H_a	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	150	1.000	0.940	1.000	0.995	0.995	0.995	0.985	0.930
sparse H_0	10	0.520	0.625	0.485	0.770	0.660	0.640	0.610	0.595
	50	0.225	0.415	0.185	0.845	0.585	0.510	0.455	0.400
	150	0.115	0.285	0.080	0.860	0.450	0.330	0.290	0.245
$(\rho = -0.5)$	p	mdd	ZL	CL	MQ (λ_n)				
					2	4	4.5	5	10
H_0	10	0.060	0.035	0.045	0.370	0.055	0.055	0.055	0.055
	50	0.075	0.085	0.060	0.610	0.105	0.070	0.070	0.060
	150	0.025	0.080	0.030	0.845	0.205	0.145	0.115	0.095
non-sparse H_a	10	0.080	0.100	0.065	0.400	0.115	0.110	0.090	0.095
	50	0.080	0.075	0.055	0.590	0.120	0.090	0.075	0.075
	150	0.045	0.095	0.045	0.835	0.225	0.150	0.095	0.085
sparse H_0	10	0.540	0.570	0.510	0.845	0.625	0.590	0.580	0.580
	50	0.200	0.345	0.195	0.875	0.445	0.370	0.350	0.320
	150	0.125	0.310	0.070	0.930	0.490	0.335	0.290	0.255

Table 8: Empirical sizes and powers from four tests for Example 3.7

(i)	p	mdd	ZL	CL	MQ (λ_n)				
					2	4	4.5	5	10
H_0	10	0.055	0.080	0.065	0.360	0.105	0.085	0.070	0.070
	50	0.080	0.060	0.065	0.680	0.095	0.070	0.070	0.050
	150	0.060	0.075	0.035	0.865	0.245	0.130	0.090	0.070
non-sparse H_a	10	1.000	0.910	1.000	0.990	0.970	0.955	0.915	0.900
	50	0.940	0.370	0.975	0.905	0.700	0.550	0.470	0.375
	150	0.615	0.160	0.720	0.920	0.540	0.355	0.215	0.175
sparse H_a	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	0.910	1.000	0.945	1.000	1.000	1.000	1.000	1.000
	150	0.620	1.000	0.715	1.000	1.000	1.000	1.000	1.000

(ii)	p	mdd	ZL	CL	MQ (λ_n)				
					2	4	4.5	5	10
non-sparse H_a	10	0.925	0.080	0.095	0.370	0.095	0.065	0.050	0.050
	50	0.395	0.035	0.085	0.690	0.080	0.060	0.035	0.030
	150	0.240	0.065	0.065	0.850	0.265	0.125	0.080	0.070
sparse H_0	10	0.890	0.095	0.100	0.275	0.095	0.080	0.065	0.065
	50	0.385	0.060	0.055	0.595	0.085	0.050	0.040	0.020
	150	0.215	0.090	0.045	0.790	0.225	0.120	0.090	0.060