Supplementary Materials: Mean Dimension Reduction and Testing for Nonparametric Tensor Response Regression

Chung Eun Lee, Xin Zhang, and Lexin Li

Baruch College, Florida State University, and University of California at Berkeley

The supplementary materials contain the proofs of Proposition 1, Theorems 1 to 5, and additional simulation results.

A Proofs

A.1 Proof of Proposition 1

First, we show the first and third assertions in Proposition 1. Let α be a vector that satisfies $\mathbb{E}(\mathcal{Y} \times_{(k)} \alpha^{\mathrm{T}} \mid X) = \mathbb{E}(\mathcal{Y} \times_{(k)} \alpha^{\mathrm{T}})$ a.s., which is further equivalent to $\mathbb{E}(\alpha^{\mathrm{T}}\mathcal{Y}_{(k)} \mid X) = \alpha^{\mathrm{T}}\mu_{(k)}$ a.s. For $k = 1, \ldots, m$, we first show that α is in the null space of span$(M^{(k)}(\mathcal{Y} \mid X))$. Note that

$$M^{(k)}(\mathcal{Y} \mid X)\alpha = -\mathbb{E}\left[\mathbb{E}\left\{(\mathcal{Y}_{(k)} - \mu_{(k)})(\alpha^{\mathrm{T}}\mathcal{Y}'_{(k)} - \alpha^{\mathrm{T}}\mu_{(k)})^{\mathrm{T}} \mid X, X'\right\} \|X - X'\|\right] = \mathbf{0},$$

and $\alpha^{\mathrm{T}}M^{(k)}(\mathcal{Y} \mid X)\alpha = 0$ which implies $M^{(k)}(\mathcal{Y} \mid X)$ is singular and α is in its null space.

We next show that, if α is in the null space of span$(M^{(k)}(\mathcal{Y} \mid X))$, then $\mathbb{E}(\alpha^{\mathrm{T}}\mathcal{Y}_{(k)} \mid X) = \alpha^{\mathrm{T}}\mu_{(k)}$ a.s. Since $M^{(k)}(\mathcal{Y} \mid X)$ is positive semidefinite and α is in the null space, we have $\alpha^{\mathrm{T}}M^{(k)}(\mathcal{Y} \mid X)\alpha = \mathrm{MDD}^2(\alpha^{\mathrm{T}}\mathcal{Y}_{(k)} \mid X) = 0$. This is equivalent to $\mathbb{E}(\alpha^{\mathrm{T}}\mathcal{Y}_{(k)} \mid X) = \alpha^{\mathrm{T}}\mu_{(k)}$ a.s., and $\mathbb{E}(\mathcal{Y} \times_{(k)} \alpha^{\mathrm{T}} \mid X) = \mathbb{E}(\mathcal{Y} \times_{(k)} \alpha^{\mathrm{T}})$ a.s., which holds for all $k = 1, \ldots, m$. Thus, we have $\mathbb{E}(\mathcal{Y} \times_{(k)} Q_k \mid X) = \mathbb{E}(\mathcal{Y} \times_{(k)} Q_k)$ a.s. for all $k = 1, \ldots, m$, where $Q_k = \beta_{k,0}\beta_{k,0}^{\mathrm{T}}$, $\beta_{k,0}$ is a basis that constructs $\mathcal{E}_k^{\perp}$. Also, we obtain that $d_k = \mathrm{rank}\{M^{(k)}(\mathcal{Y} \mid X)\}$ for all $k = 1, \ldots, m$.

Next, we show the second assertion in Proposition 1. Let $\alpha \in \operatorname{span}\left\{M^{(k)}\left(\mathcal{Y} \mid X\right)\right\}^{\perp} = \mathcal{E}_k^{\perp}$, i.e, $\alpha^{\mathrm{T}} M^{(k)}(\mathcal{Y} \mid X)\alpha = 0$. By the first assertion in Proposition 1, this is equivalent to

$$\mathbb{E}(\mathcal{Y} \times_{(k)} \alpha^{\mathrm{T}} \mid X) = \mathbb{E}(\mathcal{Y} \times_{(k)} \alpha^{\mathrm{T}}) \text{ a.s.}$$

Note that this is further identical to

$$\mathbb{E}(\alpha^{\mathrm{T}} \mathcal{Y}_{(k),j} \mid X) = \mathbb{E}(\alpha^{\mathrm{T}} \mathcal{Y}_{(k),j}) \text{ a.s., for all } j = 1, \ldots, \prod_{l \neq k} r_l \tag{S1}$$

Moreover, (S1) is equivalent to the following by Theorem 1 in Lee and Shao (2018).

$$\alpha \in \operatorname{span}\left\{M\left(\mathcal{Y}_{(k),j} \mid X\right)\right\}^{\perp} \text{ for all } j = 1, \ldots, \prod_{l \neq k} r_l$$

Therefore, we have

$$\mathcal{E}_k = \operatorname{span}\left\{M^{(k)}(\mathcal{Y} \mid X)\right\} = \sum_j \operatorname{span}\left\{M\left(\mathcal{Y}_{(k),j} \mid X\right)\right\},$$

which holds for any $k = 1, \ldots, m$. This completes the proof of Proposition 1. □

A.2 Proof of Theorem 1

For any $k = 1, \ldots, m$, we first show that

$$\left\|\widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X)\right\|_2^2 = O_p(n^{-1}). \tag{S2}$$

For any $k = 1, \ldots, m$, we have

$$\left\|\widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X)\right\|_2 \leq \left\|\widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X)\right\|_F.$$

Recall that

$$\widehat{M}^{(k)}(\mathcal{Y} \mid X) = \left[\left(\widehat{M}^{(k)}(\mathcal{Y} \mid X)\right)_{hl}\right]_{h,l=1}^{r_k} = \frac{-1}{n^2} \sum_{s,t} \left\{(\mathcal{Y}_s)_{(k)} - \overline{\mathcal{Y}}_{(k)}\right\}\left\{(\mathcal{Y}_t)_{(k)} - \overline{\mathcal{Y}}_{(k)}\right\}^{\mathrm{T}} \|X_s - X_t\|.$$

For the ease of notation, we denote the h-th row of the s-th observation of $\mathcal{Y}_{(k)}$ as $(\mathcal{Y}_s)_{h\cdot}$, the sample mean of the h-th row of $\mathcal{Y}_{(k)}$ as $\overline{\mathcal{Y}}_h$, and $\mu_h = \mathbb{E}\{(\mathcal{Y}_s)_{h\cdot}\}$. For any (h, l), we can rewrite $\left(\widehat{M}^{(k)}(\mathcal{Y} \mid X)\right)_{hl}$ as

$$\left(\widehat{M}^{(k)}(\mathcal{Y} \mid X)\right)_{hl} = \frac{(n-1)}{n}\left\{(\mathcal{U}_{n1})_{hl} + (\mathcal{U}_{n2})_{hl} + (\mathcal{U}_{n3})_{hl} + (\mathcal{U}_{n4})_{hl}\right\},$$

where

$$\begin{aligned}
(\mathcal{U}_{n1})_{hl} &= \frac{-1}{\binom{n}{2}}\sum_{s<t}\{(\mathcal{Y}_s)_{h\cdot} - \mu_h\}\{(\mathcal{Y}_t)_{l\cdot} - \mu_l\}^{\mathrm{T}}\|X_s - X_t\|, \\
(\mathcal{U}_{n2})_{hl} &= \frac{-1}{\binom{n}{2}}\sum_{s<t}\{(\mathcal{Y}_s)_{h\cdot} - \mu_h\}(\mu_l - \overline{\mathcal{Y}}_l)^{\mathrm{T}}\|X_s - X_t\|, \\
(\mathcal{U}_{n3})_{hl} &= \frac{-1}{\binom{n}{2}}\sum_{s<t}(\mu_h - \overline{\mathcal{Y}}_h)\{(\mathcal{Y}_t)_{l\cdot} - \mu_l\}^{\mathrm{T}}\|X_s - X_t\|, \\
(\mathcal{U}_{n4})_{hl} &= \frac{-1}{\binom{n}{2}}\sum_{s<t}(\mu_h - \overline{\mathcal{Y}}_h)(\mu_l - \overline{\mathcal{Y}}_l)^{\mathrm{T}}\|X_s - X_t\|.
\end{aligned}$$

Note that $(\mathcal{U}_{n1})_{hl}$ is a second-order U-statistic which has a form of

$$(\mathcal{U}_{n1})_{hl} = \frac{1}{\binom{n}{2}}\sum_{s<t}\widetilde{H}(\mathcal{Z}_s, \mathcal{Z}_t), \;\; \widetilde{H}(\mathcal{Z}_s, \mathcal{Z}_t) = \frac{-1}{2!}\sum_{(q,r)}^{(s,t)}\{(\mathcal{Y}_q)_{h\cdot} - \mu_h\}\{(\mathcal{Y}_r)_{l\cdot} - \mu_l\}^{\mathrm{T}}\|X_q - X_r\|,$$

where $\mathcal{Z}_s = (\mathcal{Y}_s, X_s)$ and $\sum_{(q,r)}^{(s,t)}$ denotes the summation over all permutations of the 2-tuple of indices (s, t).

Under the assumptions that $\mathbb{E}(\|X - \mu_X\|^2\|\mathcal{Y}_{(k)} - \mu_{(k)}\|_F^2) < \infty$ and $\mathbb{E}(\|X\|^2 + \|\mathcal{Y}\|_F^2) < \infty$, we have $\mathbb{E}\{\widetilde{H}(\mathcal{Z}, \mathcal{Z}')^2\} < \infty$ and notice that $\mathbb{E}\{(\mathcal{U}_{n1})_{hl}\} = \left(M^{(k)}(\mathcal{Y} \mid X)\right)_{hl}$. Then by applying Lemma 5.2.1.A (page 183) in Serfling (1980) to $(\mathcal{U}_{n1})_{hl}$, we obtain

$$\left|(\mathcal{U}_{n1})_{hl} - \left(M^{(k)}(\mathcal{Y} \mid X)\right)_{hl}\right|^2 = O_p(n^{-1}).$$

Also, due to the fact that $\|\mu_h - \overline{\mathcal{Y}}_h\|^2 = O_p(n^{-1})$ for $h = 1, \ldots, r_k$, we have $(\mathcal{U}_{nj})_{hl} = O_p(n^{-1/2}),\; j = 2, 3, 4,\; h, l = 1, \ldots, r_k$. Hence, we have

$$
\begin{aligned}
&\left\|\widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X)\right\|_F^2 \\
\leq\ & \sum_{h,l=1}^{r_k} \left|\left(\widehat{M}^{(k)}(\mathcal{Y} \mid X)\right)_{hl} - \left(M^{(k)}(\mathcal{Y} \mid X)\right)_{hl}\right|^2 = O_p(n^{-1}) \qquad \text{(S3)}
\end{aligned}
$$

which implies (S2).

Recall that $\{\lambda_j^{(k)}, \gamma_j^{(k)}\}_{j=1}^{r_k}$ are the eigenvalues and eigenvectors of $\mathrm{M}^{(k)}(\mathcal{Y} \mid X)$ and $\{\widehat{\lambda}_j^{(k)}, \widehat{\gamma}_j^{(k)}\}_{j=1}^{r_k}$ are the sample counterparts. For the ease of notation, we shall drop the index (k) and denote $\{\lambda_j, \gamma_j\}_{j=1}^{r_k}$ as the eigenvalues and eigenvectors of $M^{(k)}(\mathcal{Y} \mid X)$ and $\{\widehat{\lambda}_j, \widehat{\gamma}_j\}_{j=1}^{r_k}$ as the sample counterparts. Next, we use Lemma A.1. in Kneip and Utikal (2001) to show that

$$
\|\widehat{\gamma}_j - \gamma_j\| = O_p(n^{-1/2}) \text{ for } j = 1, \ldots, d_k.
$$

By applying part (b) of Lemma A.1. in Kneip and Utikal (2001), we have

$$
\widehat{\gamma}_j - \gamma_j = -\left\{\sum_{h \neq j} \frac{1}{\lambda_h - \lambda_j} \gamma_h \gamma_h^{\mathrm{T}}\right\} \left(\widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X)\right) \gamma_j + R_1,
$$

where

$$
\|R_1\|_2 \leq \frac{6\|\widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X)\|_2^2}{min_{\lambda \in \{\lambda_1, \ldots, \lambda_{r_k}\}, \lambda \neq \lambda_j} |\lambda - \lambda_j|^2} = O_p(n^{-1}).
$$

By using (S2),

$$
\begin{aligned}
&\left\| -\left\{\sum_{h \neq j} \frac{1}{\lambda_h - \lambda_j} \gamma_h \gamma_h^{\mathrm{T}}\right\} \left(\widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X)\right) \gamma_j \right\|^2 \\
=\ & \sum_{h \neq j} \frac{\left\{\gamma_h^{\mathrm{T}} \left(\widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X)\right) \gamma_j\right\}^2}{(\lambda_h - \lambda_j)^2} = O_p(n^{-1}).
\end{aligned}
$$

Hence, we have

$$
\|\widehat{\gamma}_j - \gamma_j\| = O_p(n^{-1/2}) \text{ for } j = 1, \ldots, d_k. \qquad \text{(S4)}
$$

Note that (S4) holds for any $k = 1, \ldots, m$. This completes the proof of Theorem 1. □

A.3 Proof of Theorem 2

We adopt similar arguments as the proof of Theorem 1 to prove Theorem 2. For any $k = 1, \ldots, m$, we have

$$\begin{aligned} &\left\| \widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X) \right\|_2^2 \leq \left\| \widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X) \right\|_F^2 \\ \leq &\sum_{h,l=1}^{r_k} \left| \left(\widehat{M}^{(k)}(\mathcal{Y} \mid X) \right)_{hl} - \left(M^{(k)}(\mathcal{Y} \mid X) \right)_{hl} \right|^2, \end{aligned} \tag{S5}$$

where $\left(\widehat{M}^{(k)}(\mathcal{Y} \mid X) \right)_{hl} = \frac{(n-1)}{n} \{(\mathcal{U}_{n1})_{hl} + (\mathcal{U}_{n2})_{hl} + (\mathcal{U}_{n3})_{hl} + (\mathcal{U}_{n4})_{hl}\}$ and $\{(\mathcal{U}_{ni})_{hl}\}_{i=1}^4$ are as defined in the proof of Theorem 1.

Under the assumptions that $\mathbb{E}(\|X - \mu_X\|^2 \|\mathcal{Y}_{(k)} - \mu_{(k)}\|_F^2) < \infty$ and $\mathbb{E}(\|X\|^2 + \|\mathcal{Y}\|_F^2) < \infty$, the following remains valid after applying Lemma 5.2.1.A (page 183) in Serfling (1980) to $(\mathcal{U}_{n1})_{hl}$.

$$\left| (\mathcal{U}_{n1})_{hl} - \left(M^{(k)}(\mathcal{Y} \mid X) \right)_{hl} \right|^2 = O_p(n^{-1}). \tag{S6}$$

Note that the dominant term of $(\mathcal{U}_{n2})_{hl}$ is a third-order U-statistics, i.e.,

$$(\mathcal{U}_{n2})_{hl} = O(n^{-3}) \sum_{s<t<w} \overline{H}(\mathcal{Z}_s, \mathcal{Z}_t, \mathcal{Z}_w) + O_p(n^{-1}),$$

where $\overline{H}(\mathcal{Z}_s, \mathcal{Z}_t, \mathcal{Z}_w) = \frac{-1}{3!} \sum_{(q,r,e)}^{(s,t,w)} \{(\mathcal{Y}_q)_{h\cdot} - \mu_h\}\{(\mathcal{Y}_r)_{l\cdot} - \mu_l\}^{\mathrm{T}} \|X_q - X_e\|$. By using the similar arguments in (S6) under the assumption that $\mathbb{E}(\|X - \mu_X\|^2 \|\mathcal{Y} - \mu\|_F^2) < \infty$, $\mathbb{E}(\|X - \mu_X\| \|\mathcal{Y} - \mu\|_F^2) < \infty$ and $\mathbb{E}(\|X\|^2 + \|\mathcal{Y}\|_F^2) < \infty$, we have $(\mathcal{U}_{n2})_{hl} = O_p(n^{-1/2})$. Similarly, we obtain $(\mathcal{U}_{nj})_{hl} = O_p(n^{-1/2}),\ j = 3, 4,\ h, l = 1, \ldots, r_k$. Thus, by (S5), we have

$$\begin{aligned} &\left\| \widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X) \right\|_2^2 \\ \leq &\sum_{h,l=1}^{r_k} \left| \left(\widehat{M}^{(k)}(\mathcal{Y} \mid X) \right)_{hl} - \left(M^{(k)}(\mathcal{Y} \mid X) \right)_{hl} \right|^2 = O_p(r_k^2 n^{-1}). \end{aligned} \tag{S7}$$

By Lemma A.1. in Kneip and Utikal (2001), we have

$$\widehat{\gamma}_j - \gamma_j = -\left\{\sum_{h\neq j} \frac{1}{\lambda_h - \lambda_j}\gamma_h\gamma_h^{\mathrm{T}}\right\}\left(\widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X)\right)\gamma_j + R_1,$$

where R_1 is the remainder term such that $\|R_1\|_2 \to^p 0$ under the assumption that $r_k^2 n^{-1} \to 0$ as $n \to \infty$. Note that the first term in the above equation is equivalent to

$$\begin{aligned}
&\left\|-\left\{\sum_{h\neq j} \frac{1}{\lambda_h - \lambda_j}\gamma_h\gamma_h^{\mathrm{T}}\right\}\left(\widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X)\right)\gamma_j\right\|^2 \\
= &\sum_{h\neq j} \frac{\left\{\gamma_h^{\mathrm{T}}\left(\widehat{M}^{(k)}(\mathcal{Y} \mid X) - M^{(k)}(\mathcal{Y} \mid X)\right)\gamma_j\right\}^2}{(\lambda_h - \lambda_j)^2} = O_p(r_k^2 n^{-1}).
\end{aligned}$$

Therefore, we have

$$\|\widehat{\gamma}_j - \gamma_j\| \to^p 0 \text{ for } j = 1, \ldots, d_k. \tag{S8}$$

Note that (S8) holds for any $k = 1, \ldots, m$. This completes the proof of Theorem 2. □

A.4 Proof of Theorem 3

Note that (S3) holds for $k = 1, \ldots, m$. Then with the conditions that $c_{1n} \to 0$, $c_{2n} \to 0$, $c_{1n}c_{2n}n \to \infty$, $0 < \tau < 1$, and $r_k,\ k = 1, \ldots, m$ are fixed, we apply Theorem 2.1 (i) in Zhu, Guo, Wang, and Zhu (2020) and obtain the consistency result for $k = 1, \ldots, m$. This completes the proof of Theorem 3. □

A.5 Proof of Theorem 4

Based on the results in Székeley and Rizzo (2014), $\mathrm{trace}\left\{\alpha_k^{\mathrm{T}}\widetilde{M}^{(k)}(\mathcal{Y} \mid X)\alpha_k\right\}$ is an unbiased estimator of $\mathrm{trace}\left\{\alpha_k^{\mathrm{T}}M^{(k)}(\mathcal{Y} \mid X)\alpha_k\right\}$, and it is a fourth-order U-statistic, which has the form of

$$\mathrm{trace}\left\{\alpha_k^{\mathrm{T}}\widetilde{M}^{(k)}(\mathcal{Y} \mid X)\alpha_k\right\} = \sum_{h<l<q<r} H(\mathcal{Z}_h, \mathcal{Z}_l, \mathcal{Z}_q, \mathcal{Z}_r),$$

$$H(\mathcal{Z}_h, \mathcal{Z}_l, \mathcal{Z}_q, \mathcal{Z}_r) = \frac{1}{4!} \sum_{(s,t,u,v)}^{(h,l,q,r)} \left\{ \sum_{i=1}^{q_k} \alpha_{k,i}^{\mathrm{T}} (a_{st}b_{uv} + a_{st}b_{st} - a_{st}b_{su} - a_{st}b_{tv}) \alpha_{k,i} \right\},$$

where $\alpha_{k,i}$ is the i-th column of α_k, $\sum_{(s,t,u,v)}^{(h,l,q,r)}$ denotes the summation over all permutations of the 4-tuple of indices (h, l, q, r). By the calculations following Zhang, Yao, and Shao (2018) and under the null hypothesis, we have $\mathbb{E}\left\{H(\mathcal{Z}, \mathcal{Z}^{'}, \mathcal{Z}^{''}, \mathcal{Z}^{'''}) \mid \mathcal{Z} = z\right\} = 0$, and $\mathbb{E}\left\{H(\mathcal{Z}, \mathcal{Z}^{'}, \mathcal{Z}^{''}, \mathcal{Z}^{'''}) \mid \mathcal{Z} = z, \mathcal{Z}^{'} = z^{'}\right\} = U(x, x^{'})V(y, y^{'})/6$, for $z = (x, y)$ and $z^{'} = (x^{'}, y^{'})$. Under the null and the assumptions that $\mathbb{E}(\|X\|^2 + \|\mathcal{Y}\|_F^2) < \infty$ and $\mathbb{E}(\|X - \mu_X\|^2 \|\mathcal{Y} - \mu_{\mathcal{Y}}\|_F^2) < \infty$, we have $\mathbb{E}[H(\mathcal{Z}_h, \mathcal{Z}_l, \mathcal{Z}_q, \mathcal{Z}_r)^2] < \infty$. Therefore, we obtain

$$T_n \to^D \sum_{l=1}^{\infty} \nu_l (G_l^2 - 1),$$

by applying Theorem 5.5.2 in Serfling (1980).

On the other hand, when the null hypothesis does not hold, we apply Hoeffding decomposition and have

$$\begin{aligned} &\operatorname{trace}\left\{\alpha_k^{\mathrm{T}} \widetilde{M}^{(k)}(\mathcal{Y} \mid X)\alpha_k\right\} - \operatorname{trace}\left\{\alpha_k^{\mathrm{T}} M^{(k)}(\mathcal{Y} \mid X)\alpha_k\right\} \\ &= \frac{2}{n} \sum_{h=1}^{n} \left[K(\mathcal{Z}_h) - \operatorname{trace}\left\{\alpha_k^{\mathrm{T}} M^{(k)}(\mathcal{Y} \mid X)\alpha_k\right\}\right] + R_2, \end{aligned}$$

where R_2 is asymptotically negligible. Since we have $\mathbb{E}[H(\mathcal{Z}_h, \mathcal{Z}_l, \mathcal{Z}_q, \mathcal{Z}_r)^2] < \infty$ under the assumptions that $\mathbb{E}(\|X\|^2 + \|\mathcal{Y}\|_F^2) < \infty$ and $\mathbb{E}(\|X - \mu_X\|^2 \|\mathcal{Y} - \mu_{\mathcal{Y}}\|_F^2) < \infty$, we apply Theorem 5.5.1 in Serfling (1980) and obtain

$$\sqrt{n}\left[n^{-1}T_n - \operatorname{trace}\left\{\alpha_k^{\top} M^{(k)}(\mathcal{Y} \mid X)\alpha_k\right\}\right] \to^D N(0, 4\sigma^2),$$

where $\sigma^2 = \operatorname{var}(K(\mathcal{Z}))$. This completes the proof of Theorem 4. □

A.6 Proof of Theorem 5

We first recall the definition of the bootstrap order defined in Chang and Park (2003) and Li, Hsiao, and Zinn (2003). Let T_n^* be a bootstrap statistic that depends on the random sample

$(\mathcal{Z}_j)_{j=1}^n$. Denote $T_n^* = o_p^*(1)$ a.s. if $\mathbb{P}^*(|T_n^*| > \epsilon) \to 0$ a.s., for any $\epsilon > 0$, where $\mathbb{P}^*$ is the conditional probability given $(\mathcal{Z}_j)_{j=1}^n$. Furthermore, denote $T_n^* = O_p^*(1)$ a.s. if for every $\epsilon > 0$, there exists a constant $M > 0$, such that for large n, $\mathbb{P}^*(|T_n^*| > M) < \epsilon$ a.s.

Recall that $J(\mathcal{Z}_h, \mathcal{Z}_g) = U(X_h, X_g)V(\mathcal{Y}_h, \mathcal{Y}_g)$. We first show that

$$\frac{1}{(n-3)} \sum_{h \neq g} \widetilde{A}_{hg} \left(\sum_i \alpha_{k,i}^{\mathrm{T}} \widetilde{B}_{hg} \alpha_{k,i} \right) \eta_h \eta_g = \frac{1}{(n-3)} \sum_{h \neq g} J(\mathcal{Z}_h, \mathcal{Z}_g) \eta_h \eta_g + o_p^*(1) \text{ a.s.} \quad \text{(S9)}$$

We next show the following, which implies (S9).

$$\begin{aligned} & \mathrm{var}^* \left[\frac{1}{(n-3)} \sum_{h \neq g} \left\{ \widetilde{A}_{hg} \left(\sum_i \alpha_{k,i}^{\mathrm{T}} \widetilde{B}_{hg} \alpha_{k,i} \right) - J(\mathcal{Z}_h, \mathcal{Z}_g) \right\} \eta_h \eta_g \right] \\ = & \frac{1}{(n-3)^2} \sum_{h \neq g} \left\{ \widetilde{A}_{hg} \left(\sum_i \alpha_{k,i}^{\mathrm{T}} \widetilde{B}_{hg} \alpha_{k,i} \right) - J(\mathcal{Z}_h, \mathcal{Z}_g) \right\}^2 \to^{a.s.} 0, \end{aligned} \quad \text{(S10)}$$

where var^* denotes the conditional variance given $\{\mathcal{Z}_h, \mathcal{Z}_g\}$.

For the ease of notation, write $B_{hg} = \left(\sum_i \alpha_{k,i}^{\mathrm{T}} \widetilde{B}_{hg} \alpha_{k,i} \right)$, $U_{hg} = U(X_h, X_g)$, and $V_{hg} = V(\mathcal{Y}_h, \mathcal{Y}_g)$. We have that

$$\begin{aligned} & O(n^{-2}) \sum_{h \neq g} (\widetilde{A}_{hg} B_{hg} - U_{hg} V_{hg})^2 \\ \leq & O(n^{-2}) \sum_{h \neq g} (\widetilde{A}_{hg} - U_{hg})^2 (B_{hg} - V_{hg})^2 + O(n^{-2}) \sum_{h \neq g} U_{hg}^2 (B_{hg} - V_{hg})^2 \\ + & O(n^{-2}) \sum_{h \neq g} (\widetilde{A}_{hg} - U_{hg})^2 V_{hg}^2 \\ \leq & O(n^{-2}) \left\{ \sum_{h \neq g} (\widetilde{A}_{hg} - U_{hg})^4 \right\}^{1/2} \left\{ \sum_{h \neq g} (B_{hg} - V_{hg})^4 \right\}^{1/2} \\ + & O(n^{-2}) \left(\sum_{h \neq g} U_{hg}^4 \right)^{1/2} \left\{ \sum_{h \neq g} (B_{hg} - V_{hg})^4 \right\}^{1/2} \\ + & O(n^{-2}) \left(\sum_{h \neq g} V_{hg}^4 \right)^{1/2} \left\{ \sum_{h \neq g} (\widetilde{A}_{hg} - U_{hg})^4 \right\}^{1/2}. \end{aligned}$$

Note that $\mathbb{E}(\|\mathcal{Y}\|_F^8 + \|X\|^4) < \infty$ and α_k is a fixed matrix. Thus, we have

$$\frac{1}{n^2}\sum_{h\neq g} U_{hg}^4 \rightarrow^{a.s.} \mathbb{E}(U_{12}^4), \quad \frac{1}{n^2}\sum_{h\neq g} V_{hg}^4 \rightarrow^{a.s.} \mathbb{E}(V_{12}^4).$$

We can further show that

$$\frac{1}{n^2}\sum_{h\neq g}(\widetilde{A}_{hg} - U_{hg})^4 \rightarrow^{a.s.} 0 \tag{S11}$$

$$\frac{1}{n^2}\sum_{h\neq g}(B_{hg} - V_{hg})^4 \rightarrow^{a.s.} 0 \tag{S12}$$

We first prove (S12). After some direct calculations, we have

$$\begin{aligned}\frac{1}{n^2}\sum_{h\neq g}(B_{hg} - V_{hg})^4 \leq \frac{C}{n^2}\sum_{h\neq g}\Bigg[\left\{\frac{1}{n}\sum_{s=1}^{n}\sum_{i}\alpha_{k,i}^{\mathrm{T}}(b_{hs} - \mathbb{E}(b_{hs} \mid \mathcal{Y}_h))\alpha_{k,i}\right\}^4 \\ + \left\{\frac{1}{n^2}\sum_{s\neq t}\sum_{i}\alpha_{k,i}^{\mathrm{T}}(b_{st} - \mathbb{E}(b_{st}))\alpha_{k,i}\right\}^4\Bigg] + o_{a.s.}(1),\end{aligned}$$

where $C > 0$ is a constant. By the assumption that $\mathbb{E}(\|\mathcal{Y}\|_F^8) < \infty$ and the strong law of large numbers, we have

$$\begin{aligned}&\frac{1}{n}\sum_{h=1}^{n}\left\{\frac{1}{n}\sum_{s=1}^{n}\sum_{i}\alpha_{k,i}^{\mathrm{T}}(b_{hs} - \mathbb{E}(b_{hs} \mid \mathcal{Y}_h))\alpha_{k,i}\right\}^4 \\ = \ &\frac{1}{n^5}\sum_{h=1}^{n}\sum_{s_1,s_2,s_3,s_4=1}^{n}\prod_{j=1}^{4}\left\{\sum_{i}\alpha_{k,i}^{\mathrm{T}}(b_{hs_j} - \mathbb{E}(b_{hs_j} \mid \mathcal{Y}_h))\alpha_{k,i}\right\} \rightarrow^{a.s.} 0.\end{aligned}$$

Following similar arguments, we obtain

$$\frac{1}{n^2}\sum_{h\neq g}\left\{\frac{1}{n^2}\sum_{s\neq t}\sum_{i}\alpha_{k,i}^{\mathrm{T}}(b_{st} - \mathbb{E}(b_{st}))\alpha_{k,i}\right\}^4 \rightarrow^{a.s.} 0.$$

Similarly, we can show (S11), and we obtain (S9).

Next, we show that

$$\frac{1}{(n-3)}\sum_{h\neq g} J(\mathcal{Z}_h, \mathcal{Z}_g)\eta_h\eta_g \rightarrow^{D^*} \sum_{l=1}^{\infty}\nu_l(G_l^2 - 1) \text{ a.s.} \tag{S13}$$

Note that $\mathbb{E}\{J(\mathcal{Z},\mathcal{Z}^{'})^2\} < \infty$ under the assumption $\mathbb{E}(\|X-\mu_X\|^2\|\mathcal{Y}-\mu\|_F^2) < \infty$ and $\mathbb{E}(\|X\|^4 + \|\mathcal{Y}\|_F^8) < \infty$. Therefore, we apply Dunford and Schwartz (1963, p108, Exercise 56) to $J(\mathcal{Z},\mathcal{Z}^{'})$ and obtain

$$J(\mathcal{Z},\mathcal{Z}^{'}) = \sum_{l=1}^{\infty} \nu_l \phi_l(\mathcal{Z})\phi_l(\mathcal{Z}^{'}),$$

where $(\nu_l,\phi_l)_{l=1}^{+\infty}$ is a sequence of eigenvalues and eigenfunctions of J. Define

$$n\mathcal{U}_n^* = \frac{1}{(n-3)}\sum_{h\neq g} J(\mathcal{Z}_h,\mathcal{Z}_g)\eta_h\eta_g, \;\; n\mathcal{U}_n^{(K)*} = \frac{1}{(n-3)}\sum_{h\neq g} J^{(K)}(\mathcal{Z}_h,\mathcal{Z}_g)\eta_h\eta_g,$$

where $J^{(K)}(\mathcal{Z},\mathcal{Z}^{'}) = \sum_{l=1}^{K}\nu_l\phi_l(\mathcal{Z})\phi_l(\mathcal{Z}^{'})$.

We next prove that, for any $\epsilon > 0$,

$$\lim_{K\to\infty}\text{limsup}_{n\to\infty}\mathbb{P}^*(|n\mathcal{U}_n^* - n\mathcal{U}_n^{(K)*}| > \epsilon) = 0 \text{ a.s.} \tag{S14}$$

Let

$$\frac{1}{n(n-3)}\sum_{h\neq g} J^*(\mathcal{Z}_h,\mathcal{Z}_g) := \frac{1}{n(n-3)}\sum_{h\neq g}\left\{J(\mathcal{Z}_h,\mathcal{Z}_g) - J^{(K)}(\mathcal{Z}_h,\mathcal{Z}_g)\right\}^2.$$

Since $\mathbb{E}\{J(\mathcal{Z},\mathcal{Z}^{'})^2\} = \sum_{l=1}^{\infty}\nu_l^2 < \infty$, we have

$$\mathbb{E}\{|J^*(\mathcal{Z},\mathcal{Z}^{'})|\} = \mathbb{E}\left[\left\{J(\mathcal{Z},\mathcal{Z}^{'}) - J^{(K)}(\mathcal{Z},\mathcal{Z}^{'})\right\}^2\right] = \sum_{l=K+1}^{\infty}\nu_l^2 < \infty.$$

We apply the strong law of large numbers and obtain that

$$\mathbb{E}^*\left\{|n\mathcal{U}_n^* - n\mathcal{U}_n^{(K)*}|^2\right\} = \frac{1}{(n-3)^2}\sum_{h\neq g}\left\{\sum_{l=K+1}^{\infty}\nu_l\phi_l(\mathcal{Z}_h)\phi_l(\mathcal{Z}_g)\right\}^2 \to^{a.s.} \mathbb{E}\left[\left\{\sum_{l=K+1}^{\infty}\nu_l\phi_l(\mathcal{Z})\phi_l(\mathcal{Z}^{'})\right\}^2\right].$$

Henceforth, (S14) holds by Markov inequality and the fact that

$$\lim_{K\to\infty}\mathbb{E}\left[\left\{J(\mathcal{Z},\mathcal{Z}^{'}) - J^{(K)}(\mathcal{Z},\mathcal{Z}^{'})\right\}^2\right] = \lim_{K\to\infty}\sum_{l=K+1}^{\infty}\nu_l^2 = 0.$$

Next, we show that for any K,

$$n\mathcal{U}_n^{(K)*} \to^{D*} \sum_{l=1}^{K} \nu_l (G_l^2 - 1) \text{ a.s.} \tag{S15}$$

Note that $n\mathcal{U}_n^{(K)*}$ is identical to

$$n\mathcal{U}_n^{(K)*} = \frac{1}{n}\sum_h \sum_g \left\{ \sum_{l=1}^{K} \nu_l \phi_l(\mathcal{Z}_h)\phi_l(\mathcal{Z}_g)\eta_h\eta_g \right\} - \frac{1}{n}\sum_h \sum_{l=1}^{K} \nu_l \{\phi_l(\mathcal{Z}_h)\eta_h\}^2 + o_p^*(1) \text{ a.s.} \tag{S16}$$

By Lemma 2 in the supplement of Lee, Zhang, and Shao (2020), and $\mathbb{E}\{J(\mathcal{Z},\mathcal{Z}')^4\} < \infty$ under the assumptions that $\mathbb{E}(\|X\|^4 + \|\mathcal{Y}\|_F^8)$ and $\mathbb{E}(\|X - \mu_X\|^4 \|\mathcal{Y} - \mu\|_F^4) < \infty$, we have $\mathbb{E}\{\phi_l(\mathcal{Z}_h)^4\} < \infty$. This further implies that $\mathbb{E}\{\sum_{h=1}^{\infty} \phi_l(\mathcal{Z}_h)^4/h^2\} < \infty$, where $\phi_l(\cdot)$ corresponds to $\nu_l \neq 0$.

Define the set

$$\mathcal{A}_l := \left\{ w \in \Omega : \sum_{h=1}^{\infty} \frac{\phi_l(\mathcal{Z}_h(w))^4}{h^2} < \infty \text{ and } \frac{1}{n}\sum_{h=1}^{n} \phi_l(\mathcal{Z}_h(w))^b \to \mathbb{E}\{\phi_l(\mathcal{Z}_h)^b\} \text{ for } b = 2, 4 \right\}.$$

Then $\mathbb{P}(\cap_{l=1}^{(K)} \mathcal{A}_l) = 1$, where $\cap_{l=1}^{(K)}$ is the intersection of indices where eigenvalues $(\nu_l)_{l=1}^{K}$ are nonzero. After conditioning on $\{\mathcal{Z}_h(w)\}$ with $w \in \cap_{l=1}^{(K)} \mathcal{A}_l$ and applying Corollary 7.4.1 of Resnick (2005), we have $\frac{1}{n}\sum_{h=1}^{n}(\eta_h^2 - 1)\phi_l(\mathcal{Z}_h)^2 \to^{a.s.} 0$. Since $\sum_{h=1}^{n} \phi_l(\mathcal{Z}_h)^2/n \to 1$, we have $\frac{1}{n}\sum_{h=1}^{n} \eta_h^2 \phi_l(\mathcal{Z}_h)^2 \to^{a.s.} 1$ and this further implies that $\frac{1}{n}\sum_{h=1}^{n}\sum_{l=1}^{K} \nu_l\{\phi_l(\mathcal{Z}_h)\eta_h\}^2 \to^{a.s.} \sum_{l=1}^{K} \nu_l$.

From (S16), note that the first term is equivalent to $\sum_{l=1}^{K} \nu_l \left\{\frac{1}{n^{1/2}}\sum_{h=1}^{n} \eta_h \phi_l(\mathcal{Z}_h)\right\}^2$ and

$$\begin{aligned} \text{cov}^* \left\{ \frac{1}{n^{1/2}} \sum_{h=1}^{n} \eta_h \phi_s(\mathcal{Z}_h), \frac{1}{n^{1/2}} \sum_{g=1}^{n} \eta_g \phi_t(\mathcal{Z}_g) \right\} &= \frac{1}{n}\sum_{h=1}^{n} \phi_s(\mathcal{Z}_h)\phi_t(\mathcal{Z}_h) \\ &\to^{a.s.} \mathbb{E}\{\phi_s(\mathcal{Z})\phi_t(\mathcal{Z})\} = \mathbb{I}\{s = t\}. \end{aligned} \tag{S17}$$

Similarly, define the set

$$\mathcal{B}_l := \left\{ w \in \Omega : \frac{1}{n} \max_{1 \le h \le n} \phi_l(\mathcal{Z}_h(w))^2 \to 0 \right\}.$$

By Lemma 1 in the supplement of Lee et al. (2020), and the fact that $\mathbb{E}\{\phi_l(\mathcal{Z})^2\} < \infty$ for $l = 1, 2, \ldots, K$, we have $\mathbb{P}(\cap_{l=1}^{(K)} \mathcal{B}_l) = 1$, which implies that $\mathbb{P}\{\cap_{l=1}^{(K)}(\mathcal{A}_l \cap \mathcal{B}_l)\} = 1$. After conditioning on $\{\mathcal{Z}_h(w)\}$ with $w \in \cap_{l=1}^{(K)}(\mathcal{A}_l \cap \mathcal{B}_l)$, we obtain

$$\frac{\max_{1\le h\le n} \mathrm{var}^*\{\eta_h \phi_l(\mathcal{Z}_h)\}}{\sum_{g=1}^{n} \mathrm{var}^*\{\eta_g \phi_l(\mathcal{Z}_g)\}} = \frac{\frac{1}{n}\max_{1\le h\le n} \phi_l(\mathcal{Z}_h)^2}{\frac{1}{n}\sum_{g=1}^{n} \phi_l(\mathcal{Z}_g)^2} \to 0. \tag{S18}$$

Hence, we can apply Theorem D.19 in Greene (2007) and the Cramer-Wold device, and have

$$\left(\frac{1}{n^{1/2}}\sum_{h=1}^{n} \eta_h \phi_{(1)}(\mathcal{Z}_h), \ldots, \frac{1}{n^{1/2}}\sum_{h=1}^{n} \eta_h \phi_{(K)}(\mathcal{Z}_h)\right) \to^{D} N(0, I_{(K)}),$$

for almost every realization of $\{\mathcal{Z}_h\}$, where $((1), \ldots, (K))$ are indices that correspond to nonzero eigenvalues $(\nu_l)_{l=1}^{K}$, $I_{(K)}$ is the $(K) \times (K)$ identity matrix. Therefore, we obtain (S15).

Finally, since (S14) and (S15) are both satisfied, we apply Theorem 2 in Dehling, Durieu, and Volny (2009) and obtain

$$\frac{1}{(n-3)}\sum_{h\neq g} J(\mathcal{Z}_h, \mathcal{Z}_g)\eta_h\eta_g \to^{D^*} \sum_{l=1}^{\infty} \nu_l (G_l^2 - 1) \text{ a.s.}$$

Also, with (S9), we conclude that

$$T_n^* \to^{D*} \sum_{l=1}^{\infty} \nu_l (G_l^2 - 1) \text{ a.s.},$$

which completes the proof of Theorem 5. □

B Additional Simulations

B.1 Assessing the form of $f(X)$

To further investigate the idea of using $\mathrm{trace}\left\{M^{(k)}(\mathcal{Y} \mid f(X))\right\}$ to assess the form of $f(X)$, we carry out a simulation study, where $f(X)$ takes the form of a polynomial function with a varying order,

$$\text{Linear model}: \quad \mathcal{Y} = \mathcal{B} \times_{(m+1)} X + 0.1\epsilon,$$

$$\text{Quadratic model}: \quad \mathcal{Y} = \mathcal{B} \times_{(m+1)} X^2 + 0.1\epsilon,$$

$$\text{Cubic model}: \quad \mathcal{Y} = \mathcal{B} \times_{(m+1)} X^3 + 0.1\epsilon,$$

$$\text{Quartic model}: \quad \mathcal{Y} = \mathcal{B} \times_{(m+1)} X^4 + 0.1\epsilon,$$

where $m = 2$, $\mathcal{B} = [\![\Theta; \beta_1, \beta_2, I]\!]$, the entries of $\Theta \in \mathbb{R}^{d_1 \times d_2 \times p}$ are randomly generated from Uniform$(0, 1)$, $\beta_1 \in \mathbb{R}^{r_1 \times d_1}$ and $\beta_2 \in \mathbb{R}^{r_2 \times d_2}$ are randomly generated from Uniform$(-1, 1)$ and orthogonalized. The predictors $X \in \mathbb{R}^p$ are generated from a standard normal distribution, and the error $\text{vec}(\varepsilon)$ is generated from Uniform$(-1, 1)$. We set $r_1 = r_2 = 100$, $(d_1, d_2) = (5, 5)$, $p = 5$, and the sample size $n = 100$.

Table S1: Estimation of the form of $f(X)$. Reported is the percentage of times of that the function $f(X)$ is correctly selected and incorrectly selected out of 100 data replications.

	Linear model	Quadratic model	Cubic model	Quartic model
$\widehat{f} = f$	1.00	0.74	0.60	1.00
$\widehat{f} \neq f$	0.00	0.26	0.40	0.00

For each model, we compute $\left\{\text{trace}\left\{\widetilde{M}^{(k)}\left(\mathcal{Y} \mid f_i(X)\right)\right\}\right\}_{i=1}^{4}$, with $k = 1$, $f_i(X) = C_i X^i, i = 1, 2, 3, 4$, where $C_i \in \mathbb{R}^{p \times p}$ is a matrix so that $\text{var}(f_i(X)) = I$. We select the function form that produces the largest value of $\left\{\text{trace}\left\{\widetilde{M}^{(k)}\left(\mathcal{Y} \mid f_i(X)\right)\right\}\right\}$. Table S1 reports the percentage of times that the function f is correctly selected and incorrectly selected based on 100 data replications. We see that this approach manages to select the correct form of $f(X)$ in the majority of times.

B.2 Sensitivity analysis

The proposed sparse TMDD approach involves two tuning parameters, the number of nonzero elements s for the sparse loadings, and the ridge regression penalty λ. We use the true s to calculate $\mathcal{D}$ or propose to choose the number of nonzero elements s using the BIC criterion of Sun and Li (2017) to compute TPR and FPR, while we fix the ridge penalty $\lambda = 10^{-6}$. We further carry out a sensitivity analysis regarding λ, and show that our method is relatively stable for a range of values of λ.

Table S2: Sparse dimension reduction estimation. Reported are the average and standard deviation of $\mathcal{D}(\beta_k, \widehat{\beta}_k)$, based on 1000 data replications for a range of values of λ.

			λ			
			10^{-2}	10^{-4}	10^{-6}	10^{-8}
Linear model	$\mathcal{D}(\beta_1, \widehat{\beta}_1)$	$n = 10$	0.900 (0.214)	0.902 (0.214)	0.902 (0.214)	0.902 (0.214)
		$n = 50$	0.329 (0.036)	0.332 (0.037)	0.332 (0.037)	0.332 (0.037)
		$n = 100$	0.233 (0.023)	0.236 (0.024)	0.236 (0.024)	0.236 (0.024)
	$\mathcal{D}(\beta_2, \widehat{\beta}_2)$	$n = 10$	0.907 (0.213)	0.909 (0.213)	0.909 (0.213)	0.909 (0.213)
		$n = 50$	0.331 (0.037)	0.334 (0.037)	0.334 (0.037)	0.334 (0.037)
		$n = 100$	0.233 (0.023)	0.235 (0.024)	0.235 (0.024)	0.235 (0.024)
Nonlinear model I	$\mathcal{D}(\beta_1, \widehat{\beta}_1)$	$n = 10$	0.494 (0.198)	0.495 (0.198)	0.495 (0.198)	0.496 (0.198)
		$n = 50$	0.207 (0.045)	0.209 (0.045)	0.209 (0.045)	0.209 (0.045)
		$n = 100$	0.152 (0.027)	0.155 (0.028)	0.155 (0.028)	0.155 (0.028)
	$\mathcal{D}(\beta_2, \widehat{\beta}_2)$	$n = 10$	0.498 (0.205)	0.500 (0.205)	0.500 (0.205)	0.500 (0.205)
		$n = 50$	0.208 (0.046)	0.210 (0.046)	0.210 (0.046)	0.210 (0.046)
		$n = 100$	0.154 (0.027)	0.156 (0.028)	0.156 (0.028)	0.156 (0.028)

Specifically, we adopt the linear model and the nonlinear model I in Section 4.2 of the paper, and we vary $\lambda = \{10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}\}$. Table S2 and S3 report the dimension reduction estimation and the selection results based on 1000 data replications, which shows that our method is not overly sensitive to the choice of λ in a certain range. This is consistent with the finding reported in Zou et al. (2006).

Table S3: Sparse dimension reduction selection. Reported are the average of the true positive rate (TPR), and false positive rate (FPR), based on 1000 data replications for a range of values of λ.

			λ							
			10^{-2}		10^{-4}		10^{-6}		10^{-8}	
			TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
Linear model	$\mathcal{D}(\beta_1, \widehat{\beta}_1)$	$n = 10$	0.684	0.121	0.685	0.122	0.685	0.122	0.675	0.113
		$n = 50$	0.911	0.088	0.913	0.087	0.913	0.087	0.912	0.086
		$n = 100$	0.939	0.061	0.939	0.061	0.939	0.061	0.939	0.061
	$\mathcal{D}(\beta_2, \widehat{\beta}_2)$	$n = 10$	0.679	0.102	0.682	0.106	0.682	0.105	0.693	0.114
		$n = 50$	0.914	0.083	0.916	0.083	0.916	0.083	0.917	0.083
		$n = 100$	0.941	0.059	0.940	0.060	0.940	0.060	0.940	0.060
Nonlinear model I	$\mathcal{D}(\beta_1, \widehat{\beta}_1)$	$n = 10$	0.866	0.117	0.866	0.117	0.866	0.117	0.864	0.115
		$n = 50$	0.946	0.054	0.945	0.055	0.945	0.055	0.945	0.055
		$n = 100$	0.959	0.041	0.959	0.041	0.959	0.041	0.959	0.042
	$\mathcal{D}(\beta_2, \widehat{\beta}_2)$	$n = 10$	0.870	0.109	0.870	0.110	0.870	0.110	0.872	0.111
		$n = 50$	0.948	0.052	0.948	0.052	0.948	0.052	0.948	0.052
		$n = 100$	0.960	0.040	0.960	0.041	0.960	0.041	0.960	0.040

References

Chang, Y. and Park, J. (2003) A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis*, **24**, 379-400.

Dehling, H., Durieu, O., and Volny, D. (2009). New techniques for empirical processes of dependent data. *Stochastic Processes and their Applications*, **119**, 3699-3718.

Dunford, N. and Schwartz, J. T. (1963). *Linear operators, spectral theory, self adjoint operators in Hilbert space, Part 2*, Wiley-Interscience.

Greene, W. (2007). *Econometric analysis*, Pearson Education India.

Kneip, A. and Utikal, K. J. (2001). Inference for density families using functional principal component analysis. *Journal of the American Statistical Association*, **96**, 519-542.

Lee, C. E., Zhang, X., and Shao, X. (2020). Testing conditional mean independence for functional data. *Biometrika*, **107**, 331-346.

Li, Q., Hsiao, C. and Zinn, J. (2003). Consistent specification tests for semiparametric/

nonparametric models based on series estimation methods. *Journal of Econometrics*, **112**, 295-325.

Park, T, Shao, X., and Yao, S. (2015). Partial martingale difference correlation. *Electronic Journal of Statistics*, **9**, 1492-1517.

Resnick, S. I. (2005). *A Probability path*, Springer Science & Business Media.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*, John Wiley & Sons.

Sun, W. W. and Li, L. (2017). Store: sparse tensor response regression and neuroimaging analysis. *Journal of Machine Learning Research*, **18**, 4908-4944.

Zhang, X., Yao, S. and Shao, X. (2018). Conditional mean and quantile dependence testing in high dimension, *The Annals of Statistics*, **46**, 219-246.

Zhu, X., Guo, X., Wang, T., and Zhu, L. (2020). Dimensionality determination: A thresholding double ridge ratio approach, *Computational Statistics and Data Analysis*, **146**, 106910.

Zou, H., Hastie, T., and Tibshirani, R. (2006). Sparse principal component analysis, *Journal of Computational and Graphical Statistics*, **15**, 265-286.