Supplementary materials for "Robust inference of conditional average treatment effects using dimension reduction"

Ming-Yueh Huang

Institute of Statistical Science, Academia Sinica

Shu Yang

Department of Statistics, North Carolina State University

1. Additional Notation and Regularity Conditions

Let $(\cdot)^{\otimes}$ denote the Kronecker power of a vector and let $\|\cdot\|$ represent the Frobenius norm of a matrix. Denote $f_{B^{\mathrm{T}}X}(u)$ as the marginal density of $B^{\mathrm{T}}X$,

$$f^{[m]}(x,u;B) = \partial_u^m[\mathbb{E}\{(X_l - x_l)^{\otimes m} \mid B^{\mathrm{T}}X = u\}f_{B^{\mathrm{T}}X}(u)],$$

$$E_a^{[m]}(x,u;B) = \partial_u^m[\mathbb{P}(A = a \mid B^{\mathrm{T}}X = u)\mathbb{E}\{(X_l - x_l)^{\otimes m} \mid B^{\mathrm{T}}X = u\}f_{B^{\mathrm{T}}X}(u)],$$

$$F_a^{[m]}(x,u;B) = \partial_u^m[\mathbb{E}\{Y1(A = a) \mid B^{\mathrm{T}}X = u\}\mathbb{E}\{(X_l - x_l)^{\otimes m} \mid B^{\mathrm{T}}X = u\}f_{B^{\mathrm{T}}X}(u)],$$

$$G^{[m]}(x,u;B) = \partial_u^m[\mathbb{E}(Z \mid B^{\mathrm{T}}X = u)\mathbb{E}\{(X_l - x_l)^{\otimes m} \mid B^{\mathrm{T}}X = u\}f_{B^{\mathrm{T}}X}(u)], \ \ (a = 0,1, \ \ m = 0,1,2),$$

where $Z = (2A-1)\{Y - \mu_{1-A}(B_{1-A}^{\mathrm{T}}X; B_{1-A})\}$. We will show that

$$\partial_{\mathrm{vecl}(B)}^m \widehat{\mu}_a(B^{\mathrm{T}}x;B) \to \mu^{[m]}(x;B) = \sum_{\ell=0}^{m}\binom{m}{\ell}F_a^{[\ell]}(x,B^{\mathrm{T}}x;B)E_{a,\mathrm{inv}}^{[m-\ell]}(x,B^{\mathrm{T}}x;B),$$

and

$$\partial_{\mathrm{vecl}(B)}^m \widehat{\tau}(B^{\mathrm{T}}x;B) \to \tau^{[m]}(x;B) = \sum_{\ell=0}^{m}\binom{m}{\ell}G^{[\ell]}(x,B^{\mathrm{T}}x;B)f_{\mathrm{inv}}^{[m-\ell]}(x,B^{\mathrm{T}}x;B),$$

uniformly as $n \to \infty$, where

$$f_{\mathrm{inv}}^{[0]}(x,u;B) = 1/f_{B^{\mathrm{T}}X}(u), \qquad E_{a,\mathrm{inv}}^{[0]}(x,u;B) = 1/E_a^{[0]}(x,u;B),$$

$$f_{\mathrm{inv}}^{[1]}(x,u;B) = -\frac{f^{[1]}(x,u;B)}{f_{B^{\mathrm{T}}X}^2(u)}, \qquad f_{\mathrm{inv}}^{[2]}(x,u;B) = \frac{2\{f^{[1]}(x,u;B)\}^2}{f_{B^{\mathrm{T}}X}^3(u)} - \frac{f^{[2]}(x,u;B)}{f_{B^{\mathrm{T}}X}^2(u)},$$

$$E_{a,\mathrm{inv}}^{[1]}(x,u;B) = -\frac{E_a^{[1]}(x,u;B)}{\{E_a^{[0]}(x,u;B)\}^2}, \qquad E_{a,\mathrm{inv}}^{[2]}(x,u;B) = \frac{2\{E_a^{[1]}(x,u;B)\}^2}{E_a^{[0]}(x,u;B)} - \frac{E_a^{[2]}(x,u;B)}{\{E_a^{[0]}(x,u;B)\}^2}.$$

According to the notation, we can define the corresponding score vectors and information matrices of $\textsc{cv}_a(d,B,h)$ and $\textsc{cv}(d,B,h)$:

$$S_a(B) = -1(A=a)\{Y - \mu_a(B^{\mathrm{T}}X;B)\}\mu^{[1]}(X;B),$$

$$V_a(B) = \mathbb{E}(1(A=a)[\{\mu^{[1]}(X;B)\}^{\otimes 2} - \{Y - \mu_a(B^{\mathrm{T}}X;B)\}\mu^{[2]}(X;B)]),$$

$$S(B) = -\{Z - \mathbb{E}(Z \mid B^{\mathrm{T}}X)\}\tau^{[1]}(X;B),$$

$$V(B) = \mathbb{E}[\{\tau^{[1]}(X;B)\}^{\otimes 2} - \{Z - \mathbb{E}(Z \mid B^{\mathrm{T}}X)\}\tau^{[2]}(X;B)].$$

In addition, let $B_{d,a}$ be the minimizer of $b_a^2(B) = \mathbb{E}[\{\mu_a(B^{\mathrm{T}}X;B) - \mu(X)\}^2]$ and let $B_{d,\tau}$ be the minimizer of $b_\tau^2(B) = \mathbb{E}[\{\mathbb{E}(Z \mid B^{\mathrm{T}}X) - \tau(X)\}^2]$ over all $p \times d$ matrices B. Then, $b_a^2(B) \to b_a^2(B_{d,a})$ implies $B \to B_{d,a}$ for $\mathrm{span}(B) \not\supseteq \mathrm{span}(B_a)$, and $b_\tau^2(B) \to b_\tau^2(B_{d,\tau})$ implies $B \to B_{d,\tau}$ for $\mathrm{span}(B) \not\supseteq \mathrm{span}(B_\tau)$. The following regularity conditions are imposed for our theorems:

A1 $\partial_u^{q+m}\mathbb{E}\{(X_l - x_l)^{\otimes m} \mid B^{\mathrm{T}}X = u\}$, $\partial_u^{q+2} f_{B^{\mathrm{T}}X}(u)$, $\partial_u^{q+2}\mathbb{P}(A = a \mid B^{\mathrm{T}}X = u)$, $\partial_u^{q+2}\mathbb{E}\{Y1A = a \mid B^{\mathrm{T}}X = u\}$, and $\partial_u^{q+2}\mathbb{E}(Z \mid B^{\mathrm{T}}X = u)$ ($a = 0, 1,\ m = 1, 2$), are Lipschitz continuous in u with the Lipschitz constants being independent of (x, B).

A2 $\inf_{(x,B)} f_{B^{\mathrm{T}}X}(B^{\mathrm{T}}x) > 0$ and $\inf_{(x,B)} \mathbb{P}(A = a \mid B^{\mathrm{T}}X = B^{\mathrm{T}}x) > 0$ ($a = 0, 1$).

A3 For each working dimension $d > 0$, h falls in the interval $H_{\delta,n} = [h_l n^{-\delta}, h_u n^{-\delta}]$ for some positive constants h_l and h_u and $\delta \in (1/(4q), 1/\max\{2d+2, d+4\})$. In particular, this requires $q > \max(d/2+1, 2)$.

A4 $\inf_{\{B: d < d_a\}} b_a^2(B) > 0$ and $b_a^2(B) = 0$ if and only if $B = B_a$ when $d = d_a$ ($a = 0, 1$).

A5 $V_a(B_{d,a})$ is non-singular for $d \ge d_a$ ($a = 0, 1$).

A6 For each working dimension d, $q_a > qd_a/d$ ($a = 0, 1$).

A7 $\inf_{\{B:d<d_\tau\}} b_\tau^2(B) > 0$ and $b_\tau^2(B) = 0$ if and only if $B = B_\tau$ when $d = d_\tau$.

A8 $V(B_{d,\tau})$ is non-singular for $d \geq d_\tau$.

A9 $h_\tau \to 0$ and $nh_\tau^{d_\tau} \to \infty$.

A10 For each working dimension d, $q_\tau > qd_\tau/d$.

Conditions A1–A2 are the smoothness and boundedness conditions for the population functions to ensure the uniform convergence of kernel estimators, which are commonly assumed in nonparametric smoothing methods. Moreover, to remove the remainder terms in the approximation of $\mathrm{CV}(d, B, h)$ and $\mathrm{CV}(d, B, h)$ to their target functions, the constraints for the orders of kernel functions and the bandwidths are drawn in Conditions A3 and A6. These conditions ensure the $n^{1/2}$-consistency of the estimated central mean subspaces, and our proposed data-driven bandwidths can automatically satisfy these conditions. Conditions A4–A5 and A7–A8 ensure the identifiability of B_a $(a = 0, 1)$ and B_τ, respectively, which are the base of our proposed semiparametric framework. The requirements of h_τ and q_τ used in $\widehat{\tau}(\widehat{B}^{\mathrm{T}}x; \widehat{B})$ are given in Condition A9–A10. All these conditions are analogues to assumptions in Huang and Chiang (2017) but modified for estimating central mean subspaces.

2. Preliminary Lemmas

The proofs of the main theorems rely on the following lemma:

Lemma 1. *Suppose that Assumption 1 and Conditions A1–A6 are satisfied. Then,*

$$\widehat{\tau}(u; B) - \mathbb{E}(Z \mid B^{\mathrm{T}}X = u) = \frac{1}{n}\sum_{i=1}^{n}[Z_i - \mathbb{E}(Z \mid B^{\mathrm{T}}X = u) + \{1 - \pi(X_i)\}\varepsilon_{1,i} - \pi(X_i)\varepsilon_{0,i}]\omega_{h,i}(u; B) + r_n(u; B),$$

where $\varepsilon_{a,i} = \{Y_i - \mu_a(X_i)\}1(A_i = a)$, $(a = 0, 1)$, $\omega_{h,i}(u; B) = \mathcal{K}_{q,h}(B^{\mathrm{T}}X_i - u)/\sum_{j=1}^{n}\mathcal{K}_{q,h}(B^{\mathrm{T}}X_j - u)$, *and* $\sup_{(u,B)} |r_n(u, B)| = o_{\mathbb{P}}[h^q + \{\log n/(nh^d)\}^{1/2}]$.

Proof. First note that

$$\begin{aligned}\widehat{\tau}(u; B) - \mathbb{E}(Z \mid B^{\mathrm{T}}X = u) &= \frac{1}{n}\{\widehat{D}_i - \mathbb{E}(Z \mid B^{\mathrm{T}}X = u)\}\omega_{h,i}(u; B) \\ &= \frac{1}{n}\{Z_i - \mathbb{E}(Z \mid B^{\mathrm{T}}X = u)\}\omega_{h,i}(u; B) + \frac{1}{n}\{\widehat{D}_i - Z_i\}\omega_{h,i}(u; B).\end{aligned}$$

Further,

$$
\begin{aligned}
&\frac{1}{n}(\widehat{D}_i - Z_i)\omega_{h,i}(u;B) \\
=\ &\frac{1}{n}\sum_{i=1}^{n}[(1-A_i)\{\widehat{\mu}_1(\widehat{B}_1^{\mathrm{T}}X_i;\widehat{B}_1) - \mu_1(X_i)\} - A_i\{\widehat{\mu}_0(\widehat{B}_0^{\mathrm{T}}X_i;\widehat{B}_0) - \mu_0(X_i)\}]\omega_{h,i}(u;B) \\
=\ &\frac{1}{n}\sum_{i=1}^{n}(1-A_i)\{\widehat{\mu}_1(B_1^{\mathrm{T}}X_i;B_1) - \mu_1(X_i)\}\omega_{h,i}(u;B) \\
&- \frac{1}{n}\sum_{i=1}^{n}A_i\{\widehat{\mu}_0(B_0^{\mathrm{T}}X_i;B_0) - \mu_0(X_i)\}\omega_{h,i}(u;B) + O_{\mathbb{P}}(n^{-1/2}) \\
\triangleq\ &I_1 + I_2 + O_{\mathbb{P}}(n^{-1/2}), \qquad (21)
\end{aligned}
$$

because of $\|\mathrm{vecl}(\widehat{B}_a - B_a)\| = O_{\mathbb{P}}(n^{-1/2})$ by Theorem 1. Now let $\kappa_{a,h,i}(u) = \mathcal{K}_{q_a,h}(B_a^{\mathrm{T}}X_i - u)/\sum_{j=1}^{n}1(A_j = a)\mathcal{K}_{q_a,h}(B_a^{\mathrm{T}}X_j - u)$. Then, we decompose I_1 into

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(1-A_i)\widehat{\mu}_1(B_1^{\mathrm{T}}X_i;B_1) - \mu_1(X_i)\}\omega_{h,i}(u;B) \\
=\ &\frac{1}{n}\sum_{i=1}^{n}\{1-\pi(X_i)\}\omega_{h,i}(u;B)\sum_{j=1}^{n}\{Y_j - \mu_1(X_i)\}1(A_j = 1)\kappa_{1,h_1,j}(B_1^{\mathrm{T}}X_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\{\pi(X_i) - A_i\}\omega_{h,i}(u;B)\sum_{j=1}^{n}\{Y_j - \mu_1(X_i)\}1(A_j = 1)\kappa_{1,h_1,j}(B_1^{\mathrm{T}}X_i) \\
=\ &\frac{1}{n}\sum_{i=1}^{n}\{1-\pi(X_i)\}\varepsilon_{1,i}\omega_{h,i}(u;B) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\{1-\pi(X_i)\}\left\{\sum_{j=1}^{n}\varepsilon_{1,j}\kappa_{1,h_1,j}(B_1^{\mathrm{T}}X_i) - \varepsilon_{1,i}\right\}\omega_{h,i}(u;B) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\{1-\pi(X_i)\}\left[\sum_{j=1}^{n}\{\mu_1(X_j) - \mu_1(X_i)\}1(A_j = 1)\kappa_{1,h_1,j}(B_1^{\mathrm{T}}X_i)\right]\omega_{h,i}(u;B) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\{\pi(X_i) - A_i\}\omega_{h,i}(u;B)\sum_{j=1}^{n}\{Y_j - \mu_1(X_i)\}1(A_j = 1)\kappa_{1,h_1,j}(B_1^{\mathrm{T}}X_i) \\
\triangleq\ &J_0 + J_1 + J_2 + J_3. \qquad (22)
\end{aligned}
$$

To bound J_1, we re-write it as

$$
J_1 = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{1,i}\left\{\sum_{j=1}^{n}\{1-\pi(X_j)\}\omega_{h,j}(u;B)\kappa_{1,h_1,j}(B_1^{\mathrm{T}}X_i) - \{1-\pi(X_i)\}\omega_{h,i}(u;B)\right\}.
$$

Since $\mathbb{E}(\varepsilon_{1,i} \mid X_i) = 0$, we can show that J_1 is a degenerate U-process indexed by (u,B). An application of Theorem 6 in Nolan and Pollard (1987) ensures that $\mathbb{E}(\sup_{(u,B)}|J_1|) \le C/(n^2h_1^{d_1}h^d)$. Thus, by selecting h_1 in an optimal rate $O\{n^{-1/(2q_1+d_1)}\}$ and coupled with Conditions A3 and A6, we have

$$
\sup_{(u,B)}|J_1| = o_{\mathbb{P}}\left\{h^q + \left(\frac{\log n}{nh^d}\right)^{1/2}\right\}. \qquad (23)
$$

Second, similar to the proofs in Huang and Chiang (2017), standard arguments in kernel smoothing estimation show that

$$\sup_i |\sum_{j=1}^n \{\mu_1(X_j) - \mu_1(X_i)\} 1(A_j = 1)\kappa_{1,h_1,j}(B_1^{\mathrm{T}} X_i)|$$
$$= O_{\mathbb{P}} \left\{ h_1^{q_1} + \left(\frac{\log n}{n h_1^{d_1}} \right)^{1/2} \right\} = O_{\mathbb{P}}\{n^{-q_1/(2q_1+d_1)}\}$$

by selecting h_1 in an optimal rate $O\{n^{-1/(2q_1+d_1)}\}$. Under Conditions A3 and A6, one can further show that this rate is $o_{\mathbb{P}}[h^q + \{\log n/(nh^d)\}^{1/2}]$. Thus, we have

$$\sup_{(u,B)} |J_2| = o_{\mathbb{P}} \left\{ h^q + \left(\frac{\log n}{nh^d} \right)^{1/2} \right\}. \tag{24}$$

Finally, note that J_3 is also a degenerate U-process indexed by (u, B). Thus, by the same argument for J_1, we can show that

$$\sup_{(u,B)} |J_3| = o_{\mathbb{P}} \left\{ h^q + \left(\frac{\log n}{nh^d} \right)^{1/2} \right\}. \tag{25}$$

By substituting (23)–(25) into (22), we then have

$$\sup_{(u,B)} |I_1 - \frac{1}{n}\sum_{i=1}^n (1 - A_i)\varepsilon_{1,i}\omega_{h,i}(u; B)| = o_{\mathbb{P}} \left\{ h^q + \left(\frac{\log n}{nh^d} \right)^{1/2} \right\}. \tag{26}$$

Following the same arguments above, we can also show that

$$\sup_{(u,B)} |I_2 - \frac{1}{n}\sum_{i=1}^n A_i\varepsilon_{0,i}\omega_{h,i}(u; B)| = o_{\mathbb{P}} \left\{ h^q + \left(\frac{\log n}{nh^d} \right)^{1/2} \right\}. \tag{27}$$

Substituting (26)–(27) into (21) completes the proof. □

Now we derive the independent and identically distributed representations of $\widehat{\tau}(B^{\mathrm{T}}x; B) - \tau^{[0]}(x; B)$ and $\partial_{\mathrm{vecl}(B)}\widehat{\tau}(B^{\mathrm{T}}x; B) - \tau^{[1]}(x; B)$.

Lemma 2. *Suppose that Assumption 1 and Conditions A1–A6 are satisfied. Then,*

$$\sup_{(x,B)} |\widehat{\tau}(B^{\mathrm{T}}x; B) - \tau^{[0]}(x; B) - \frac{1}{n}\sum_{i=1}^n \eta_{h,i}^{[0]}(x; B)| = o_{\mathbb{P}} \left(h^{2q} + \frac{\log n}{nh^d} \right), \tag{28}$$

$$\sup_{(x,B)} \|\partial_{\mathrm{vecl}(B)}\widehat{\tau}(B^{\mathrm{T}}x; B) - \tau^{[1]}(x; B) - \frac{1}{n}\sum_{i=1}^n \eta_{h,i}^{[1]}(x; B)\| = o_{\mathbb{P}} \left(h^{2q} + \frac{\log n}{nh^{d+1}} \right), \tag{29}$$

where

$$\eta_{h,i}^{[0]}(x;B) = \frac{\xi_i(x;B)}{f_{B^{\mathrm{T}}X}(B^{\mathrm{T}}x)}\mathcal{K}_{q,h}(B^{\mathrm{T}}X_i - B^{\mathrm{T}}x),$$

$$\eta_{h,i}^{[1]}(x;B) = \frac{\xi_i(x;B)}{f_{B^{\mathrm{T}}X}(B^{\mathrm{T}}x)}\partial_{\mathrm{vecl}(B)}\mathcal{K}_{q,h}(B^{\mathrm{T}}X_i - B^{\mathrm{T}}x) - \tau^{[1]}(x;B)\mathcal{K}_{q,h}(B^{\mathrm{T}}X_i - B^{\mathrm{T}}x) - \frac{f^{[1]}(x,B^{\mathrm{T}}x;B)}{f_{B^{\mathrm{T}}X}(B^{\mathrm{T}}x)}\eta_{h,i}^{[0]}(x;B),$$

and $\xi_i(x;B) = Z_i - \mathbb{E}(Z \mid B^{\mathrm{T}}X = B^{\mathrm{T}}x)$.

Proof. First, (28) is a direct result of Lemma 1. As for (29), note that

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{D}_i\partial_{\mathrm{vecl}(B)}\mathcal{K}_{q,h}(B^{\mathrm{T}}X_i - B^{\mathrm{T}}x) - G^{[1]}(x,B^{\mathrm{T}}x;B) = \frac{1}{n}\sum_{i=1}^{n}\xi_i(x;B)\partial_{\mathrm{vecl}(B)}\mathcal{K}_{q,h}(B^{\mathrm{T}}X_i - B^{\mathrm{T}}x) + r_{1n}(x;B), \quad (210)$$

where $\sup_{(x,B)}|r_{1n}(x,B)| = o_{\mathbb{P}}[h^q + \{\log n/(nh^{d+1})\}^{1/2}]$, by paralleling the proof steps of Lemma 1. Now by using the Taylor expansion, we have

$$\begin{aligned}&\partial_{\mathrm{vecl}(B)}\widehat{\tau}(B^{\mathrm{T}}x;B) - \tau^{[1]}(x;B)\\ = &\frac{\sum_{i=1}^{n}\widehat{D}_i\partial_{\mathrm{vecl}(B)}\mathcal{K}_{q,h}(B^{\mathrm{T}}X_i - B^{\mathrm{T}}x)/n - \tau^{[0]}(x;B)\sum_{i=1}^{n}\partial_{\mathrm{vecl}(B)}\mathcal{K}_{q,h}(B^{\mathrm{T}}X_i - B^{\mathrm{T}}x)/n}{f_{B^{\mathrm{T}}X}(B^{\mathrm{T}}x)}\\ &- \frac{\tau^{[1]}(x;B)}{n}\sum_{i=1}^{n}\mathcal{K}_{q,h}(B^{\mathrm{T}}X_i - B^{\mathrm{T}}x) - \frac{f^{[1]}(x,B^{\mathrm{T}}x;B)}{f_{B^{\mathrm{T}}X}(B^{\mathrm{T}}x)}\{\widehat{\tau}(B^{\mathrm{T}}x;B) - \tau^{[0]}(x;B)\}\\ &+ r_{2n}(x;B), \end{aligned} \quad (211)$$

where

$$\begin{aligned} r_{2n}(x,B) \quad = \quad & O_{\mathbb{P}}\{|\widehat{\tau}(B^{\mathrm{T}}x;B) - \tau^{[0]}(x;B)|^2\\ &+\|\sum_{i=1}^{n}\widehat{D}_i\partial_{\mathrm{vecl}(B)}\mathcal{K}_{q,h}(B^{\mathrm{T}}X_i - B^{\mathrm{T}}x)/n - G^{[1]}(x,B^{\mathrm{T}}x;B)\|^2\}.\end{aligned}$$

Finally, substituting the result in Lemma 1 and (210) into (211) completes the proof. □

Corollary 1. *Suppose that Assumption 1 and Conditions A1–A6 are satisfied. Then,*

$$\sup_{(x,B)}|\widehat{\tau}(B^{\mathrm{T}}x;B) - \tau^{[0]}(x;B)| = O_{\mathbb{P}}\left\{h^q + \left(\frac{\log n}{nh^d}\right)^{1/2}\right\},$$

$$\sup_{(x,B)}\|\partial_{\mathrm{vecl}(B)}\widehat{\tau}(B^{\mathrm{T}}x;B) - \tau^{[1]}(x;B)\| = O_{\mathbb{P}}\left\{h^q + \left(\frac{\log n}{nh^{d+1}}\right)^{1/2}\right\}.$$

3. Proofs of Theorems 2 and 3

3.1 Proof of Theorem 2

Proof. Let $\bar{\tau}^{-i}(B^{\mathrm{T}}X_i;B) = \sum_{j\neq i} Z_j \mathcal{K}_{q,h}(B^{\mathrm{T}}X_j - B^{\mathrm{T}}X_i)/\sum_{j\neq i}\mathcal{K}_{q,h}(B^{\mathrm{T}}X_j - B^{\mathrm{T}}X_i)$. We can decompose $\mathrm{CV}(d,B,h)$ into

$$\begin{aligned}
\mathrm{CV}(d,B,h) &= \frac{1}{n}\sum_{i=1}^{n}\{Z_i - \bar{\tau}^{-i}(B^{\mathrm{T}}X_i;B)\}^2 + \frac{1}{n}\sum_{i=1}^{n}(\widehat{D}_i - Z_i)^2 \\
&+ \frac{1}{n}\sum_{i=1}^{n}\{\widetilde{\tau}^{-i}(B^{\mathrm{T}}X_i;B) - \bar{\tau}^{-i}(B^{\mathrm{T}}X_i;B)\}^2 \\
&+ \frac{2}{n}\sum_{i=1}^{n}(\widehat{D}_i - Z_i)\{\widetilde{\tau}^{-i}(B^{\mathrm{T}}X_i;B) - \bar{\tau}^{-i}(B^{\mathrm{T}}X_i;B)\} \\
&+ \frac{2}{n}\sum_{i=1}^{n}(\widehat{D}_i - Z_i)\{Z_i - \tau(X_i)\} + \frac{2}{n}\sum_{i=1}^{n}(\widehat{D}_i - Z_i)\{\tau(X_i) - \bar{\tau}^{-i}(B^{\mathrm{T}}X_i;B)\} \\
&+ \frac{2}{n}\sum_{i=1}^{n}\{Z_i - \tau(X_i)\}\{\widetilde{\tau}^{-i}(B^{\mathrm{T}}X_i;B) - \bar{\tau}^{-i}(B^{\mathrm{T}}X_i;B)\} \\
&+ \frac{2}{n}\sum_{i=1}^{n}\{\tau(X_i) - \bar{\tau}^{-i}(B^{\mathrm{T}}X_i;B)\}\{\widetilde{\tau}^{-i}(B^{\mathrm{T}}X_i;B) - \bar{\tau}^{-i}(B^{\mathrm{T}}X_i;B)\} \\
&\triangleq SS_1 + SS_2 + SS_3 + SC_1 + SC_2 + SC_3 + SC_4 + SC_5.
\end{aligned}$$

Note that

$$\sup_i |\widehat{D}_i - Z_i| \leq \sum_{a=0}^{1}\sup_{(u,B)}|\widehat{\mu}_a(u;B) - \mu_a(u;B)| = o_{\mathbb{P}}\left\{h^q + \left(\frac{\log n}{nh^d}\right)^{1/2}\right\}, \tag{312}$$

$$\sup_{(i,B)}|\widetilde{\tau}^{-i}(B^{\mathrm{T}}X_i;B) - \bar{\tau}^{-i}(B^{\mathrm{T}}X_i;B)| \leq C\sum_{a=0}^{1}\sup_{(u,B)}|\widehat{\mu}_a(u;B) - \mu_a(u;B)| = o_{\mathbb{P}}\left\{h^q + \left(\frac{\log n}{nh^d}\right)^{1/2}\right\} \tag{313}$$

for some positive constant C, by using Conditions A1–A3, Condition A6, and standard arguments in kernel smoothing estimation.

When $\mathrm{span}(B) \supseteq \mathrm{span}(B_\tau)$, Theorem 1 of Huang and Chiang (2017) implies that $SS_1 = \sigma_\tau^2 + O_{\mathbb{P}}\{h^{2q} + \log n/(nh^d)\}$, where $\sigma_\tau^2 = \mathbb{E}[\{Z - \tau(X)\}^2]$. From (312)–(313), $\sup_B |SS_3|$ and $\sup_B |SC_1|$ are of order $o_{\mathbb{P}}\{h^{2q} + \log n/(nh^d)\}$. Further, by using $\sup_{(x,B)}|\bar{\tau}(B^{\mathrm{T}}x;B) - \tau(x)| = O_{\mathbb{P}}[h^q + \{\log n/(nh^d)\}^{1/2}]$, $\sup_B |SC_3|$ and $\sup_B |SC_5|$ are also of order $o_{\mathbb{P}}\{h^{2q}+\log n/(nh^d)\}$. Now note that SC_4 can be expressed a U-process indexed by B asymptotically. By using the same proof steps for the cross term in Theorem 1 of Huang and Chiang (2017), one can immediately conclude that $\sup_B |SC_4| = o_{\mathbb{P}}\{h^{2q} + \log n/(nh^d)\}$. Combining the results above, we have $\mathrm{CV}(d,B,h) = SS_1 + SS_2 + SC_2 + o_p(SS_1)$ uniformly in B. When $\mathrm{span}(B) \nsupseteq \mathrm{span}(B_\tau)$, Theorem 1 of Huang and Chiang (2017) implies that

$SS_1 = \sigma_\tau^2 + b_\tau^2(B) + o_{\mathbb{P}}(1)$. By using (312)–(313) again, we have $\textsc{cv}(d, B, h) = SS_1 + SS_2 + SC_2 + o_{\mathbb{P}}(1)$ uniformly in B. Finally, since SS_2 and SC_2 are independent of B, the minimizer of $\textsc{cv}(d, B, h)$ has the same asymptotic distribution as the minimizer of SS_1. Thus, Theorem 2 is a direct result of Theorem 2 in Huang and Chiang (2017). □

3.2 Proof of Theorem 3

Proof. By using first-ordered Taylor expansion, we have

$$\begin{aligned}\widehat{\tau}(\widehat{B}^{\mathrm{T}}x; \widehat{B}) - \tau(x) &= \widehat{\tau}(\widehat{B}^{\mathrm{T}}x; \widehat{B}) - \widehat{\tau}(B_\tau^{\mathrm{T}}x; B_\tau) + \widehat{\tau}(B_\tau^{\mathrm{T}}x; B_\tau) - \tau(x) \\ &= \partial_{\mathrm{vecl}(B)}\widehat{\tau}(\bar{B}^{\mathrm{T}}x; \bar{B})\mathrm{vecl}(\widehat{B} - B_\tau) + \widehat{\tau}(B_\tau^{\mathrm{T}}x; B_\tau) - \tau(x),\end{aligned}$$

where $\bar{B}$ lies on the line segment between $\widehat{B}$ and B_τ. From Theorem 2, $\mathrm{vecl}(\widehat{B} - B_\tau) = O_{\mathbb{P}}(n^{-1/2})$. Coupled with Corollary 1 and continuous mapping theorem, $\partial_{\mathrm{vecl}(B)}\widehat{\tau}(\bar{B}^{\mathrm{T}}x; \bar{B}) = O_{\mathbb{P}}(1)$. Moreover, from (28), we have

$$(nh_\tau^{d_\tau})^{1/2}\{\widehat{\tau}(B_\tau^{\mathrm{T}}x; B_\tau) - \tau(x)\} - h_\tau^{q_\tau}\gamma(x) \to \mathrm{N}\{0, \sigma_\tau^2(x)\}$$

in distribution as $n \to \infty$. Combining the results above completes the proof of Theorem 3. □

References

Huang, M.-Y. and C.-T. Chiang (2017). An effective semiparametric estimation approach for the sufficient dimension reduction model. *Journal of the American Statistical Association 112*(519), 1296–1310.

Nolan, D. and D. Pollard (1987). U-processes: rates of convergence. *The Annals of Statistics 15*(2), 780–799.