A Supplemental Note to "Semiparametric Models with Single-Index Nuisance Parameters"

Kyungchul Song[1]

Department of Economics, University of British Columbia

August 3, 2012

This note contains the technical proofs of the results in the paper titled "Semiparametric Models with Single-Index Nuisance Parameters."

1 Bahadur Representation of Sample Linear Functionals of SNN Estimators

In this section, we present a Bahadur representation of sample linear functionals of SNN estimators that is uniform over function spaces. (The proofs are found at the end of the paper.) In a different context, Stute and Zhu (2005) obtained a related result that is not uniform.

Suppose that we are given a random sample $\{(S_i, W_i, X_i)\}_{i=1}^n$ drawn from the distribution of a random vector $(S, W, X) \in \mathbf{R}^{d_S+1+d_X}$. Let $\mathcal{S}_S, \mathcal{S}_X$ and $\mathcal{S}_W$ be the supports of S, X, and W respectively. Let Λ be a class of $\mathbf{R}$-valued functions on $\mathbf{R}^{d_X}$ with a generic element denoted by λ. We also let Φ and Ψ be classes of real functions on $\mathbf{R}$ and $\mathbf{R}^{d_S}$ with generic elements φ and ψ and let $\tilde{\varphi}$ and $\tilde{\psi}$ be their envelopes. Let $L_p(P)$, $p \geq 1$, be the space of L_p-bounded functions: $||f||_p \equiv \{\int |f(x)|^p P(dx)\}^{1/p} < \infty$, and for a space of functions $\mathcal{F} \subset L_p(P)$ for $p \geq 1$, let $N_{[]}(\varepsilon, \mathcal{F}, ||\cdot||_p)$ denote the bracketing number of $\mathcal{F}$ with respect to the norm $||\cdot||_p$, i.e., the smallest number r such that there exist $f_1, \cdots, f_r$ and $\Delta_1, \cdots, \Delta_r \in L_p(P)$ such that

[1] Address correspondence to: Kyungchul Song, Department of Economics, University of British Columbia, 997-1873 East Mall, Vancouver, BC, V6T 1Z1, Canada.

$||\Delta_i||_p < \varepsilon$ and for all $f \in \mathcal{F}$, there exists $1 \le i \le r$ with $||f_i - f||_p < \Delta_i/2$. Similarly, we define $N_{[]}(\varepsilon, \mathcal{F}, ||\cdot||_\infty)$ to be the bracketing number of $\mathcal{F}$ with respect to the sup norm $||\cdot||_\infty$, where for any real map f on $\mathbf{R}^{d_X}$, we define $||f||_\infty = \sup_{z \in \mathbf{R}^{d_X}} |f(z)|$. For any norm $||\cdot||$ which is equal to $||\cdot||_p$ or $||\cdot||_\infty$, we define $N(\varepsilon, \mathcal{F}, ||\cdot||)$ to be the covering number of $\mathcal{F}$, i.e., the smallest number of ε-balls that cover $\mathcal{F}$. Letting $F_\lambda(\cdot)$ be the CDF of $\lambda(X)$, we denote $U_\lambda \equiv F_\lambda(\lambda(X))$. Define $g_{\varphi,\lambda}(u) \equiv \mathbf{E}[\varphi(W)|U_\lambda = u]$ and $g_{\psi,\lambda}(u) \equiv \mathbf{E}[\psi(S)|U_\lambda = u]$.

Let $U_{n,\lambda,i} \equiv \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} 1\{\lambda(X_j) \le \lambda(X_i)\}$ and consider the estimator:

$$\hat{g}_{\varphi,\lambda,i}(u) \equiv \frac{1}{(n-1)\hat{f}_{\lambda,i}(u)} \sum_{j=1,j\neq i}^{n} \varphi(W_j) K_h\left(U_{n,\lambda,j} - u\right),$$

where $\hat{f}_{\lambda,i}(u) \equiv (n-1)^{-1}\sum_{j=1,j\neq i}^{n} K_h(U_{n,\lambda,j} - u)$. The semiparametric process of focus takes the following form:

$$\nu_n(\lambda, \varphi, \psi) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(S_i)\left\{\hat{g}_{\varphi,\lambda,i}(U_{n,\lambda,i}) - g_\varphi(U_{\lambda,i})\right\},$$

with $(\lambda, \varphi, \psi) \in \Lambda \times \Phi \times \Psi$.

The main focus in this section is on establishing an asymptotic linear representation of $\nu_n(\lambda, \varphi, \psi)$. The critical element in the proof is to bound the size of the class of conditional expectation functions $\mathcal{G} \equiv \{g_{\varphi,\lambda}(\cdot) : (\varphi, \lambda) \in \Phi \times \Lambda\}$. We begin with the following lemma that establishes the bracketing entropy bound for $\mathcal{G}$ with respect to $||\cdot||_q$, $q \ge 1$.

Lemma B1 : *Suppose that the density f_λ of $\lambda(X)$ is bounded uniformly over $\lambda \in \Lambda$. Furthermore, assume that there exists an envelope $\tilde{\varphi}$ for Φ such that $G_\Phi \equiv \sup_{x \in \mathcal{S}_X} \mathbf{E}[\tilde{\varphi}(W)|X = x] < \infty$, and that for some $C_L > 0$,*

$$\sup_{\varphi \in \Phi} \sup_{\lambda \in \Lambda} |g_{\varphi,\lambda}(u_1) - g_{\varphi,\lambda}(u_2)| \le C_L |u_1 - u_2|, \text{ for all } u_1, u_2 \in [0,1].$$

Then for all $\varepsilon > 0$, $q \geq 1$, *and* $p \geq 1$,

$$N_{[]}(C_{\Phi}\varepsilon^{1/(q+1)}, \mathcal{G}, ||\cdot||_q) \leq N_{[]}(\varepsilon, \Phi, ||\cdot||_p) \cdot N_{[]}(\varepsilon, \Lambda, ||\cdot||_\infty),$$

where $C_\Phi \equiv 1 + 8C_\Lambda G_\Phi + C_L + G_\Phi/2$ *and* $C_\Lambda \equiv \sup_{\lambda\in\Lambda}\sup_{v\in\mathbf{R}} f_\lambda(v)$.

Proof of Lemma B1 : For each $\eta \in (0,1]$, we define $g_{\varphi,\lambda,\eta}(u) = \mathbf{E}\left[\varphi(W)Q_\eta(U_\lambda - u)\right]$, where $Q_\eta(u) = Q(u/\eta)/\eta$, and $Q(u) = 1\{u \in [-1/2, 1/2]\}$. For each $\eta \in (0,1]$, we let $\mathcal{G}_\eta = \{g_{\varphi,\lambda,\eta}(\cdot) : (\varphi,\lambda) \in \Phi\times\Lambda\}$. First, we show that for all $\eta \in (0,1]$, and for all $\varepsilon > 0$ and $p \geq 1$,

$$N_{[]}((1 + 8C_\Lambda G_\Phi)\cdot\varepsilon/\eta, \mathcal{G}_\eta, ||\cdot||_\infty) \leq N_{[]}(\varepsilon, \Phi, ||\cdot||_p) \cdot N_{[]}(\varepsilon, \Lambda, ||\cdot||_\infty). \tag{1}$$

Fix $\eta \in (0,1]$, $\varepsilon > 0$, and $p \geq 1$, and choose two sets of brackets $\{\varphi_j, \Delta_{\Phi,j}\}_{j=1}^{N_\Phi}$ and $\{\lambda_j, \Delta_{\Lambda,j}\}_{j=1}^{N_\Lambda}$ that constitute ε-brackets for Φ and Λ, with respect to $||\cdot||_p$ and $||\cdot||_\infty$ respectively, where $N_\Phi = N_{[]}(\varepsilon, \Phi, ||\cdot||_p)$ and $N_\Lambda = N_{[]}(\varepsilon, \Lambda, ||\cdot||_\infty)$. Define $g_{jk,\eta}(u) = \mathbf{E}\left[\varphi_j(W)Q_\eta(U_{\lambda_k} - u)\right]$. For any $g_{\varphi,\lambda,\eta} \in \mathcal{G}_\eta$, we can choose the pairs $(\varphi_j, \Delta_{\Phi,j})$ and $(\lambda_k, \Delta_{\Lambda,k})$ such that

$$\begin{aligned} |\varphi(w) - \varphi_j(w)| &\leq \Delta_{\Phi,j}(w), \text{ for all } w \in \mathcal{S}_W \text{ and} \\ |\lambda(x) - \lambda_k(x)| &\leq \Delta_{\Lambda,k}(x), \text{ for all } x \in \mathcal{S}_X \end{aligned}$$

and $||\Delta_{\Phi,j}||_p \leq \varepsilon$ and $||\Delta_{\Lambda,k}||_\infty \leq \varepsilon$. Note that for $u \in [0,1]$,

$$\begin{aligned} |g_{\varphi,\lambda,\eta}(u) - g_{jk,\eta}(u)| &\leq \mathbf{E}\left[\Delta_{\Phi,j}(W)\cdot|Q_\eta(U_\lambda - u)|\right] \\ &\quad + \mathbf{E}\left[\tilde{\varphi}(W)\cdot|Q_\eta(U_\lambda - u) - Q_\eta(U_{\lambda_k} - u)|\right]. \end{aligned}$$

Certainly, $\mathbf{E}\left[\Delta_{\Phi,j}(W)\cdot|Q_\eta(U_\lambda - u)|\right] \leq \mathbf{E}\left[\Delta_{\Phi,j}(W)\right]/\eta$. As for the second term, let $\Delta_{\lambda,\lambda_k} =$

$|U_\lambda - U_{\lambda_k}|$ and bound it by

$$\begin{aligned} & \frac{1}{\eta} G_\Phi \cdot P\left\{\eta/2 - \Delta_{\lambda,\lambda_k} \leq |U_\lambda - u| \leq \eta/2 + \Delta_{\lambda,\lambda_k}\right\} \\ \leq \ & \frac{8}{\eta} C_\Lambda G_\Phi ||\Delta_{\Lambda,k}||_\infty, \end{aligned}$$

because $\Delta_{\lambda,\lambda_k} \leq 4C_\Lambda ||\Delta_{\Lambda,k}||_\infty$. Take $\Delta_{jk,\eta}(u) \equiv \mathbf{E}\left[\Delta_{\Phi,j}(W)\right]/\eta + 8C_\Lambda G_\Phi ||\Delta_{\Lambda,k}||_\infty/\eta$ (a constant function), so that $||\Delta_{jk,\eta}||_\infty \leq (1 + 8C_\Lambda G_\Phi)\varepsilon/\eta$. Hence take $\{g_{jk,\eta}, \Delta_{jk,\eta}\}_{j,k=1}^{N_\Phi, N_\Lambda}$ to be $(1 + 8C_\Lambda G_\Phi)\varepsilon/\eta$-brackets of $\mathcal{G}_\eta$, affirming (1).

We turn to $\mathcal{G}$. For any $(\varphi, \lambda) \in \Phi \times \Lambda$, we obtain that for $u \in [0,1]$,

$$\begin{aligned} |g_{\varphi,\lambda,\eta}(u) - g_{\varphi,\lambda}(u)| & \leq \left|\mathbf{E}\left[\{g_{\varphi,\lambda}(U_\lambda) - g_{\varphi,\lambda}(u)\} \cdot Q_\eta(U_\lambda - u)\right]\right| \\ & \quad + \left|\mathbf{E}\left[g_{\varphi,\lambda}(u)\{Q_\eta(U_\lambda - u) - 1\}\right]\right| \\ & \leq C_L \eta + G_\Phi \cdot \left|\mathbf{E}\left[Q_\eta(U_\lambda - u) - 1\right]\right| \leq C_L \eta + G_\Phi \cdot b_\eta(u), \end{aligned}$$

where $b_\eta(u) \equiv \frac{1}{2}\left(1\left\{u \in (1 - \eta/2, 1]\right\} + 1\left\{u \in [0, \eta/2)\right\}\right)$. The second inequality follows by change of variables applied to the leading term. The last inequality follows because for all $\eta \in (0,1]$,

$$\begin{aligned} |1 - \mathbf{E}Q_\eta(U_\lambda - u)| & = \left|1 - \frac{1}{\eta}\int_0^1 1\{|v - u| \leq \eta/2\} dv\right| \\ & = \left|1 - \frac{1}{\eta}\int_{[u-\eta/2, u+\eta/2]\cap[0,1]} dv\right| \leq b_\eta(u). \end{aligned}$$

Fix $\varepsilon > 0$ and $q \geq 1$. We select $\eta = \varepsilon^{q/(q+1)}$ and take $((1 + 8C_\Lambda G_\Phi)\varepsilon/\eta)$-brackets $\{g_{jk,\eta}, \Delta_{jk,\eta}\}_{j,k=1}^{N_\Phi, N_\Lambda}$ that cover $\mathcal{G}_\eta$ (with respect to $||\cdot||_\infty$) with $N_\Phi = N_{[]}(\varepsilon, \Phi, ||\cdot||_p)$ and $N_\Lambda = N_{[]}(\varepsilon, \Lambda, ||\cdot||_\infty)$. We define

$$\tilde{\Delta}_{jk,\eta}(u) \equiv \Delta_{jk,\eta}(u) + C_L \eta + G_\Phi \cdot b_\eta(u).$$

Then, certainly,

$$
\begin{aligned}
||\tilde{\Delta}_{jk}||_q &\leq (1+8C_\Lambda G_\Phi)\varepsilon/\eta + C_L\eta + G_\Phi \cdot ||b_\eta||_q \\
&\leq (1+8C_\Lambda G_\Phi)\varepsilon/\eta + (C_L + G_\Phi/2)\eta^{1/q}.
\end{aligned}
$$

Therefore, the set $\{g_{jk,\eta}, \tilde{\Delta}_{jk,\eta}\}_{j,k=1}^{N_\Phi,N_\Lambda}$ forms the set of $C_\Phi\varepsilon^{1/(q+1)}$-brackets for $\mathcal{G}$ with respect to $||\cdot||_q$. This gives the desired entropy bound for $\mathcal{G}$. ■

We are prepared to present the uniform Bahadur representation of $\nu_n(\lambda,\varphi,\psi)$. Let $\Lambda_n \equiv \{\lambda \in \Lambda : ||\lambda - \lambda_0||_\infty \leq c_n\}$, where $0 < c_n n^{1/4} \to 0$. We let $X = [X_1^\top, X_2^\top]^\top$, where X_1 is a continuous random vector and X_2 is a discrete random vector taking values in a finite set $\{x_1, \cdots, x_M\}$. We make the following assumptions.

Assumption B1 : (i) For some $C > 0$, $p \geq q > 4$, $b_\Psi \in (0, q/(q-1))$, and $b_\Phi \in (0, q/\{(q+1)(q-1)\})$,

$$
\log N_{[]}(\varepsilon, \Phi, ||\cdot||_p) < C\varepsilon^{-b_\Phi} \text{ and } \log N_{[]}(\varepsilon, \Psi, ||\cdot||_p) < C\varepsilon^{-b_\Psi}, \text{ for each } \varepsilon > 0,
$$

and $\mathbf{E}[\tilde{\varphi}(W)^p] + \mathbf{E}[\tilde{\psi}(S)^p] + \sup_{x\in\mathcal{S}_X} \mathbf{E}[\tilde{\varphi}(W)|X=x] + \sup_{x\in\mathcal{S}_X} \mathbf{E}[\tilde{\psi}(W)|X=x] < \infty$.
(ii) (a) For $q > 4$ in (i) and for some $b_\Lambda \in (0, q/\{(q+1)(q-1)\})$ and $C > 0$, $\log N_{[]}(\varepsilon, \Lambda, ||\cdot||_\infty) \leq C\varepsilon^{-b_\Lambda}$, for each $\varepsilon > 0$.
(b) For all $\lambda \in \Lambda$, the density $f_\lambda(\cdot)$ of $\lambda(X)$ is bounded uniformly over $\lambda \in \Lambda$ and bounded away from zero on the interior of its support uniformly over $\lambda \in \Lambda$.

Assumption B2 : (i) $K(\cdot)$ is symmetric, nonnegative, compact supported, twice continuously differentiable with bounded derivatives, and $\int K(t)dt = 1$.
(ii) $n^{1/2}h^3 + n^{-1/2}h^{-2}(-\log h) \to 0$ as $n \to \infty$.

Assumption B3 : $\mathbf{E}[\varphi(W)|U_\lambda = \cdot]$ is twice continuously differentiable with derivatives bounded uniformly over $(\lambda,\varphi) \in B(\lambda_0;\varepsilon) \times \Phi$ with some $\varepsilon > 0$.

The following lemma offers a uniform representation of ν_n.

Lemma B2 : *Suppose that* Assumptions B1-B4 *hold. Then,*

$$\sup_{(\lambda,\varphi,\psi)\in\Lambda_n\times\Phi\times\Psi}\left|\nu_n(\lambda,\varphi,\psi)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_{\psi,\lambda}(U_{\lambda,i})\{\varphi(W_i)-g_{\varphi,\lambda}(U_{\lambda,i})\}\right|=o_P(1).$$

Proof of Lemma B2: To make the flow of the arguments more visible, the proof proceeds by making certain claims and proving them at the end of the proof. Without loss of generality, assume that the support of K is contained in $[-1,1]$. Throughout the proofs, the notation $\mathbf{E}_i$ indicates conditional expectation given (W_i, S_i, X_i).

Let $\hat{\rho}_{\varphi,\lambda,i}(t)\equiv(n-1)^{-1}\sum_{j=1,j\neq i}^{n}K_h(U_{n,\lambda,j}-t)\varphi(W_j)$, $\xi_{1n}(u)\equiv\int_{[-u/h,(1-u)/h]\cap[-1,1]}K(v)dv$, and

$$\Delta_i^{\varphi,\psi}(\lambda)\equiv g_{\psi,\lambda}(U_{\lambda,i})\{\varphi(W_i)-g_{\varphi,\lambda}(U_{\lambda,i})\}.$$

We write $\hat{g}_{\varphi,\lambda,i}(U_{n,\lambda,i})-g_{\varphi,\lambda}(U_{\lambda,i})$ as

$$\begin{aligned}R_{1i}(\lambda,\varphi) &\equiv \frac{\hat{\rho}_{\varphi,\lambda,i}(U_{n,\lambda,i})-g_{\varphi,\lambda}(U_{\lambda,i})\hat{f}_{\lambda,i}(U_{\lambda,i})}{\xi_{1n}(U_{\lambda,i})}\\ &\quad+\frac{[\hat{\rho}_{\varphi,\lambda,i}(U_{n,\lambda,i})-g_{\varphi,\lambda}(U_{\lambda,i})\hat{f}_{\lambda,i}(U_{\lambda,i})](\xi_{1n}(U_{\lambda,i})-\hat{f}_{\lambda,i}(U_{n,\lambda,i}))}{\hat{f}_{\lambda,i}(U_{n,\lambda,i})\xi_{1n}(U_{\lambda,i})}\\ &= R_{1i}^{A}(\lambda,\varphi)+R_{1i}^{B}(\lambda,\varphi).\end{aligned}$$

Put $\pi\equiv(\lambda,\varphi,\psi)$ and $\Pi_n\equiv\Lambda_n\times\Phi\times\Psi$, and write

$$\nu_n(\pi)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(S_i)R_{1i}^{A}(\lambda,\varphi)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(S_i)R_{1i}^{B}(\lambda,\varphi)\equiv r_{1n}^{A}(\pi)+r_{1n}^{B}(\pi),\ \pi\in\Pi_n.$$

First, we show the following:

C1: $\sup_{\pi\in\Pi_n}\left|r_{1n}^{B}(\pi)\right|=o_P(1)$.

We turn to $r_{1n}^A(\pi)$, which we write as

$$\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K_{ij}^{\lambda}+\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}\{K_{n,ij}^{\lambda}-K_{ij}^{\lambda}\}$$
$$\equiv \quad R_{1n}(\pi)+R_{2n}(\pi),\ \text{say},$$

where $\psi_{n,\lambda,i}\equiv\psi(S_i)/\xi_{1n}(U_{\lambda,i})$, $\Delta_{\varphi,\lambda,ij}\equiv\varphi(W_j)-g_{\varphi,\lambda}(U_{\lambda,i})$, $K_{n,ij}^{\lambda}\equiv K_h(U_{n,\lambda,j}-U_{n,\lambda,i})$ and $K_{ij}^{\lambda}\equiv K_h(U_{\lambda,j}-U_{\lambda,i})$. We will now show that

$$\sup_{\pi\in\Pi_n}|R_{2n}(\pi)|\rightarrow_P 0. \tag{2}$$

Let $\delta_i^{\lambda}\equiv U_{n,\lambda,i}-U_{\lambda,i}$ and $d_{\lambda,ji}\equiv\delta_j^{\lambda}-\delta_i^{\lambda}$ and write $R_{2n}(\pi)$ as

$$\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K_{h,ij}'d_{\lambda,ji}+\frac{1}{2(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}d_{\lambda,ji}^2K_{h,ij}''$$
$$= \quad A_{1n}(\pi)+A_{2n}(\pi),\ \text{say},$$

where $K_{h,ij}'\equiv h^{-2}\partial K(t)/\partial t$ at $t=(U_{\lambda,i}-U_{\lambda,j})/h$ and

$$K_{h,ij}''\equiv h^{-3}\partial^2K(t)/\partial t^2|_{t=t_{ij}}$$

with $t_{ij}\equiv\{(1-a_{ij})(U_{\lambda,i}-U_{\lambda,j})+a_{ij}(U_{n,\lambda,i}-U_{n,\lambda,j})\}/h$, for some $a_{ij}\in[0,1]$. Later we will show the following:

C2: $\sup_{\pi\in\Pi_n}|A_{2n}(\pi)|=o_P(1)$.

We turn to $A_{1n}(\pi)$ which we write as

$$\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K_{h,ij}'\delta_j^{\lambda} \tag{3}$$
$$-\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K_{h,ij}'\delta_i^{\lambda}$$
$$= \quad B_{1n}(\pi)+B_{2n}(\pi),\ \text{say}.$$

Write $B_{1n}(\pi)$ as

$$\frac{1}{n-1}\sum_{j=1}^{n}\left[\frac{1}{\sqrt{n}}\sum_{i=1,i\neq j}^{n}\left\{\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K'_{h,ij}-\mathbf{E}_j\left[\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K'_{h,ij}\right]\right\}\right](U_{n,\lambda,j}-U_{\lambda,j})$$
$$+\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\mathbf{E}_j\left[\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K'_{h,ij}\right](U_{n,\lambda,j}-U_{\lambda,j})=C_{1n}(\pi)+C_{2n}(\pi),\text{ say.}$$

As for $C_{1n}(\pi)$, we show the following later.

C3: $\sup_{\pi\in\Pi_n}|C_{1n}(\pi)|=o_P(1)$.

We deduce a similar result for $B_{2n}(\pi)$, so that we write

$$\begin{aligned}A_{1n}(\pi) &= \frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\mathbf{E}_j\left[\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K'_{h,ij}\right](U_{n,\lambda,j}-U_{\lambda,j}) \\ &\quad -\frac{1}{(n-1)\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1,i\neq j}^{n}\mathbf{E}_i\left[\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K'_{h,ij}\right](U_{n,\lambda,i}-U_{\lambda,i})+o_P(1) \\ &= D_{1n}(\pi)-D_{2n}(\pi)+o_P(1),\text{ say.}\end{aligned} \tag{4}$$

Now, we show that $D_{1n}(\pi)$ and $D_{2n}(\pi)$ cancel out asymptotically. As for $D_{1n}(\pi)$, using Hoeffding's decomposition and taking care of the degenerate U-process (e.g. see C3 and its proof below),

$$\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\int_0^1\mathbf{E}\left[\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K'_{h,ij}\right](1\{U_{\lambda,i}\leq u_1\}-u_1)\,du_1+o_P(1),$$

uniformly over $\pi\in\Pi_n$. Similarly, as for $D_{2n}(\pi)$, we can write it as

$$\frac{1}{(n-1)\sqrt{n}}\sum_{j=1}^{n}\sum_{i=1,i\neq j}^{n}\int_0^1\mathbf{E}\left[\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K'_{h,ij}\right](1\{U_{\lambda,j}\leq u_1\}-u_1)\,du_1+o_P(1),$$

uniformly over $\pi\in\Pi_n$. Note that $\mathbf{E}\left[\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K'_{h,ij}\right]$ does not depend on a particular choice of the pair (i,j) as long as $i\neq j$. Hence $D_{1n}(\pi)=D_{2n}(\pi)+o_P(1)$, uniformly over $\pi\in\Pi_n$, and that $\sup_{\pi\in\Pi_n}|A_{1n}(\pi)|=o_P(1)$, which, together with (C2), completes the proof of (2).

It remains to show that

$$\sup_{\pi\in\Pi_n}\left|R_{1n}(\pi)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_{\psi,\lambda}(U_{\lambda,i})\{\varphi(W_i)-g_{\varphi,\lambda}(U_{\lambda,i})\}\right|=o_P(1). \tag{5}$$

We define $q_{n,ij}^{\pi}\equiv\psi_{n,\lambda,i}\Delta_{\varphi,\lambda,ij}K_{ij}^{\lambda}$ and write $R_{1n}(\pi)$ as

$$\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}q_{n,ij}^{\pi}. \tag{6}$$

Let $\rho_{n,ij}^{\pi}\equiv q_{n,ij}^{\pi}-\mathbf{E}_i[q_{n,ij}^{\pi}]-\mathbf{E}_j[q_{n,ij}^{\pi}]+\mathbf{E}[q_{n,ij}^{\pi}]$ and define

$$u_n(\pi)\equiv\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\rho_{n,ij}^{\pi}.$$

Then, $\{u_n(\pi):\pi\in\Pi_n\}$ is a degenerate U-process on Π_n. We write (6) as

$$\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left\{\mathbf{E}_i[q_{n,ij}^{\pi}]+\mathbf{E}_j[q_{n,ij}^{\pi}]-\mathbf{E}[q_{n,ij}^{\pi}]\right\}+u_n(\pi). \tag{7}$$

We will later show the following two claims.

C4: $\sup_{\pi\in\Pi_n}|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\mathbf{E}_i[q_{n,ij}^{\pi}]-\mathbf{E}[q_{n,ij}^{\pi}]\}|=o_P(1)$.

C5: $\sup_{\pi\in\Pi_n}|u_n(\pi)|=o_P(1)$.

We conclude from these claims that uniformly over $\pi\in\Pi_n$,

$$\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}q_{n,ij}^{\pi}=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{E}_j[q_{n,ij}^{\pi}]+o_P(1).$$

Then the proof of Lemma B2 is completed by showing the following.

C6: $\sup_{\pi\in\Pi_n}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\mathbf{E}_j[q_{n,ij}^{\pi}]-g_{\psi,\lambda}(U_{\lambda,j})\{\varphi(W_j)-g_{\varphi,\lambda}(U_{\lambda,j})\}\right)\right|=o_P(1)$.

Proof of C1: From the proof of Lemma A3 of Song (2009) (by replacing λ with $F_\lambda\circ\lambda$

there), it follows that

$$\max_{1\leq i\leq n}\sup_{\lambda\in\Lambda}\sup_{x\in\mathbf{R}^{d_X}}|F_{n,\lambda,i}(\lambda(x))-F_\lambda(\lambda(x))|=O_P(n^{-1/2}), \tag{8}$$

where $F_{n,\lambda,i}(\bar{\lambda}) \equiv \frac{1}{n-1}\sum_{j=1,j\neq i}^{n}1\{\lambda(X_j)\leq\bar{\lambda}\}$. We bound $\max_{1\leq i\leq n}\sup_{\lambda\in\Lambda}|\hat{f}_{\lambda,i}(U_{n,\lambda,i})-\xi_{1n}(U_{\lambda,i})|$ by

$$\begin{aligned}&\max_{1\leq i\leq n}\sup_{\lambda\in\Lambda}\sup_{u\in[0,1]}\left|\hat{f}_{\lambda,i}(u)-\xi_{1n}(u)\right|+\max_{1\leq i\leq n}\sup_{\lambda\in\Lambda}|\xi_{1n}(U_{n,\lambda,i})-\xi_{1n}(U_{\lambda,i})| \\ =\;& O_P(n^{-1/2}h^{-1}\sqrt{-\log h})+O_P(n^{-1/2}h^{-1})=O_P(n^{-1/2}h^{-1}\sqrt{-\log h}),\end{aligned} \tag{9}$$

using (29) of Song (2009). Hence, uniformly over $1\leq i\leq n$ and over $\lambda\in\Lambda$,

$$\begin{aligned}&|\hat{\rho}_{\varphi,\lambda,i}(U_{n,\lambda,i})-\hat{f}_{\lambda,i}(U_{n,\lambda,i})g_{\varphi,\lambda}(U_{\lambda,i})| \\ \leq\;& |\hat{\rho}_{\varphi,\lambda,i}(U_{n,\lambda,i})-\xi_{1n}(U_{\lambda,i})g_{\varphi,\lambda}(U_{\lambda,i})|+|g_{\varphi,\lambda}(U_{\lambda,i})||\xi_{1n}(U_{\lambda,i})-\hat{f}_{\lambda,i}(U_{n,\lambda,i})| \\ =\;& |\hat{\rho}_{\varphi,\lambda,i}(U_{n,\lambda,i})-\xi_{1n}(U_{\lambda,i})g_{\varphi,\lambda}(U_{\lambda,i})|+O_P(n^{-1/2}h^{-1}\sqrt{-\log h}).\end{aligned}$$

As for the leading term, we apply (23) of Song (2009) and Assumption B3 to deduce that uniformly over $1\leq i\leq n$ and over $(\lambda,\varphi)\in\Lambda\times\Phi$,

$$\begin{aligned}|\hat{\rho}_{\varphi,\lambda,i}(U_{n,\lambda,i})-\xi_{1n}(U_{\lambda,i})g_{\varphi,\lambda}(U_{\lambda,i})|(1-1_{n,\lambda,i}) &= O_P(h)\text{ and} \\ |\hat{\rho}_{\varphi,\lambda,i}(U_{n,\lambda,i})-\xi_{1n}(U_{\lambda,i})g_{\varphi,\lambda}(U_{\lambda,i})|1_{n,\lambda,i} &= O_P(h^2+n^{-1/2}h^{-1}\sqrt{-\log h}),\end{aligned}$$

where $1_{n,\lambda,i}=1\{|1-U_{n,\lambda,i}|>h\}$. Also, observe that (e.g. see arguments after (29) of Song (2009))

$$\xi_{1n}(u)\geq 1/2\text{ for all }u\in[0,1]. \tag{10}$$

Therefore, we bound $\left|r_{1n}^{B}(\pi)\right|$ by

$$\frac{C_1}{n}\sum_{i=1}^{n}\left|[\hat{\rho}_{\varphi,\lambda,i}(U_{n,\lambda,i})-g_{\varphi,\lambda}(U_{\lambda,i})\xi_{1n}(U_{\lambda,i})](\xi_{1n}(U_{\lambda,i})-\hat{f}_{\lambda,i}(U_{n,\lambda,i}))\right|1_{n,\lambda,i}$$
$$+\frac{C_2}{n}\sum_{i=1}^{n}\left|[\hat{\rho}_{\varphi,\lambda,i}(U_{n,\lambda,i})-g_{\varphi,\lambda}(U_{\lambda,i})\xi_{1n}(U_{\lambda,i})](\xi_{1n}(U_{\lambda,i})-\hat{f}_{\lambda,i}(U_{n,\lambda,i}))\right|(1-1_{n,\lambda,i}),$$

for some $C_1, C_2 > 0$, uniformly over $\pi \in \Pi_n$ with large probability. From (9), the first term is equal to $O_P(n^{-1}h^{-2}(-\log h)) = o_P(n^{-1/2})$ uniformly over $1 \leq i \leq n$. As for the second term, it is bounded with large probability by

$$Cn^{-1/2}h^{-1}\sqrt{-\log h}\cdot\frac{1}{n}\sum_{i=1}^{n}1\left\{|1-U_{0,i}|\leq Ch\right\}=O_P(n^{-1/2}\sqrt{-\log h})=o_P(n^{-1/2}),$$

for some $C > 0$ with large probability. Hence (C1) is established.

Proof of C2: Let $\tilde{\Delta}_{ij} \equiv \tilde{\varphi}(W_i) + \sup_{x\in\mathcal{S}_X}\mathbf{E}[\tilde{\varphi}(W_j)|X_j = x]$. Since $\max_{1\leq i,j\leq n}\sup_{\lambda\in\Lambda} d_{\lambda,ji}^2 = O_P(n^{-1})$ by (8), with large probability, we bound $|A_{2n}(\pi)|$ by the absolute value of

$$\frac{C}{(n-1)n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left(\tilde{\psi}_i\tilde{\Delta}_{ij}\left|K''_{h,ij}\right|-\mathbf{E}\left[\tilde{\psi}_i\tilde{\Delta}_{ij}\left|K''_{h,ij}\right|\right]\right)+\frac{\sqrt{n}}{n}\mathbf{E}\left[\tilde{\psi}_i\tilde{\Delta}_{ij}\left|K''_{h,ij}\right|\right]. \quad (11)$$

Using the standard U statistics theory, we bound the leading term by

$$\left|\frac{C}{n\sqrt{n}}\sum_{i=1}^{n}\left(\mathbf{E}_i\left[\tilde{\psi}_i\tilde{\Delta}_{ij}\left|K''_{h,ij}\right|\right]-\mathbf{E}\left[\tilde{\psi}_i\tilde{\Delta}_{ij}\left|K''_{h,ij}\right|\right]\right)\right|$$
$$+\left|\frac{C}{\sqrt{n}}\sum_{j=1}^{n}\left(\mathbf{E}_j\left[\tilde{\psi}_i\tilde{\Delta}_{ij}\left|K''_{h,ij}\right|\right]-\mathbf{E}\left[\tilde{\psi}_i\tilde{\Delta}_{ij}\left|K''_{h,ij}\right|\right]\right)\right|+o_P(1).$$

The expected value of the sum above is bounded by

$$\frac{C}{n}\sqrt{Var\left(\mathbf{E}_i\left[\tilde{\psi}_i\tilde{\Delta}_{ij}\left|K''_{h,ij}\right|\right]\right)}+\frac{C}{n}\sqrt{Var\left(\mathbf{E}_j\left[\tilde{\psi}_i\tilde{\Delta}_{ij}\left|K''_{h,ij}\right|\right]\right)}=O(n^{-1}h^{-5/2})=o(1),$$

so that the leading term of (11) is $o(1)$. On the other hand, $\frac{\sqrt{n}}{n}\mathbf{E}\left[\tilde{\psi}_i\tilde{\Delta}_{ij}|K''_{h,ij}|\right] = O(n^{-1/2}h^{-2}) =$

$o(1)$.

Proof of C3: Let $K'(u) \equiv (\partial K(u)/\partial u)1\{u \in (0,1)\}$. Then $K'(\cdot/h)$ is uniformly bounded and bounded variation. Let $\mathcal{K}_\Lambda \equiv \{K'((u-\sigma(\cdot))/h) : (\sigma, u) \in \mathcal{I} \times [0,1]\}$, where $\mathcal{I} \equiv \{(F_\lambda \circ \lambda)(x) : \lambda \in \Lambda\}$. Also, let $\mathcal{W}_\Lambda \equiv \{\xi_{1n}(\sigma(\cdot)) : \sigma \in \mathcal{I}\}$. Take $p \geq 4$ as in Assumption B1. By Lemma A1 of Song (2009) and Assumptions B1(ii) and B3(iv),

$$\begin{aligned} \log N_{[]}(\varepsilon, \mathcal{K}_\Lambda, ||\cdot||_p) &\leq \log N(C\varepsilon, \mathcal{I}, ||\cdot||_\infty) + C/\varepsilon \leq C\varepsilon^{-b_\Lambda} \text{ and} \\ \log N_{[]}(\varepsilon, \mathcal{W}_\Lambda, ||\cdot||_p) &\leq \log N(C\varepsilon, \mathcal{I}, ||\cdot||_\infty) + C/\varepsilon \leq C\varepsilon^{-b_\Lambda}, \end{aligned} \tag{12}$$

for some $C > 0$. Let $\xi_{\pi,u}(S_i, X_i) \equiv \psi_{n,\lambda,i} K'((u - U_{\lambda,i})/h)$. We bound $\sup_{\pi \in \Pi_n} |C_{1n}(\pi)| \leq h^{-2} G_{1n} \cdot V_{1n} + h^{-2} G_{2n} \cdot V_{2n}$, where

$$\begin{aligned} V_{1n} &\equiv \sup_{(\pi,u) \in \Pi_n \times [0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\xi_{\pi,u}(S_i, X_i) - \mathbf{E}\xi_{\pi,u}(S_i, X_i) \right) \right| \text{ and} \\ V_{2n} &\equiv \sup_{(\pi,u) \in \Pi_n \times [0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\xi_{\pi,u}(S_i, X_i) g_{\varphi,\lambda}(U_{\lambda,i}) - \mathbf{E}\xi_{\pi,u}(S_i, X_i) g_{\varphi,\lambda}(U_{\lambda,i}) \right) \right|, \end{aligned}$$

where $G_{1n} \equiv \sup_{\lambda \in \Lambda_n} \frac{1}{n} \sum_{j=1}^n \tilde{\varphi}(W_j) |\delta_j^\lambda|$ and $G_{2n} \equiv \sup_{\lambda \in \Lambda_n} \frac{1}{n} \sum_{j=1}^n |\delta_j^\lambda|$. Define $\mathcal{F}_1 \equiv \{\xi_{\pi,u}(\cdot,\cdot) : (\pi, u) \in \Pi_n \times [0,1]\}$ and $\mathcal{F}_2 \equiv \{\xi_{\pi,u}(\cdot,\cdot)(g_{\varphi,\lambda} \circ \sigma_\lambda)(\cdot) : (\pi, u) \in \Pi_n \times [0,1]\}$. Then, $\mathcal{F}_1 \subset (\Psi/\mathcal{W}_\Lambda) \cdot \mathcal{K}_{1,\Lambda}$ and $\mathcal{F}_2 \subset (\Psi/\mathcal{W}_\Lambda) \cdot \mathcal{K}_{1,\Lambda} \cdot \mathcal{H}$, where

$$\mathcal{H} \equiv \{(g_{\varphi,\lambda} \circ \sigma_\lambda)(\cdot) : (\varphi, \lambda) \in \Phi \times \Lambda\}.$$

By Assumption B1(i) and Lemma B1, $\log N_{[]}(\varepsilon, \mathcal{H}, ||\cdot||_q) \leq C\varepsilon^{-(q+1)\{b_\Phi \vee b_\Lambda\}}$. Combining this with (12), the entropy bound for Ψ in Assumption B1(i), (12), and the fact that $\sup_{x \in \mathcal{S}_X} \mathbf{E}\left[\tilde{\psi}(S)|X = x\right] < \infty$ and $\sup_{x \in \mathcal{S}_X} \mathbf{E}\left[\tilde{\varphi}(W)|X = x\right] < \infty$, we find that

$$\log N_{[]}(\varepsilon, \mathcal{F}_1, ||\cdot||_2) \leq C\varepsilon^{-\{b_\Psi \vee b_\Phi \vee b_\Lambda\}}, \text{ and } \log N_{[]}(\varepsilon, \mathcal{F}_2, ||\cdot||_2) \leq C\varepsilon^{-(b_\Psi \vee [(q+1)\{b_\Phi \vee b_\Lambda\}])}.$$

From (10), we take an envelope of $\mathcal{F}_1$ as $\bar{\xi}_1(s,x) = 2\tilde{\psi}(s)||K'||_\infty$ and an envelope of $\mathcal{F}_2$ as $\bar{\xi}_2(s,x) = 2\tilde{\psi}(s)||K'||_\infty \cdot \sup_{x\in\mathcal{S}_X} \mathbf{E}\left[\tilde{\varphi}(W)|X=x\right]$. Certainly, $||\bar{\xi}_1||_p < \infty$ and $||\bar{\xi}_2||_p < \infty$ by Assumption B1(i). Using the maximal inequality of Pollard (1989)[2] and using the fact that $b_\Psi \vee [(q+1)\{b_\Phi \vee b_\Lambda\}] < 2$ (Assumptions B1(i) and B1(ii)(a)), we find that

$$V_{1n} = O_P(1) \text{ and } V_{2n} = O_P(1).$$

By the fact that $\max_{1\le j\le n}\sup_{\lambda\in\Lambda_n}|\delta_j^\lambda| = O_P(n^{-1/2})$, we also deduce that $G_{1n} = O_P(n^{-1/2})$ and $G_{2n} = O_P(n^{-1/2})$. The desired result follows because $O_P(n^{-1/2}h^{-2}) = o_P(1)$.

Proof of C4: Observe that $\mathbf{E}_i[q_{n,ij}^\pi]$ is equal to $\zeta_\pi(S_i,X_i)$, where for $\pi\in\Pi_n$,

$$\zeta_\pi(S_i,X_i) = \frac{\psi(S_i)}{\xi_{1n}(U_{\lambda,i})}\int_0^1 \{g_{\varphi,\lambda}(u) - g_{\varphi,\lambda}(U_{\lambda,i})\}K_h(u-U_{\lambda,i})du.$$

Define $\mathcal{F}_3 = \{\zeta_\pi(\cdot,\cdot) : \pi\in\Pi_n\}$. Take $p\ge 4$ as in Assumption B1. Then, similarly as in the proof of (C3), we can show that

$$\log N_{[]}(\varepsilon,\mathcal{F}_3,||\cdot||_q) \le C\varepsilon^{-(b_\Psi\vee[(q+1)\{b_\Phi\vee b_\Lambda\}])} \tag{13}$$

for some $C>0$. With $\sigma_\lambda(x) \equiv (F_\lambda\circ\lambda)(x)$, we take

$$\bar{\zeta}(s,x) = 2\tilde{\psi}(s)\sup_{(\varphi,\lambda)\in\Phi\times\Lambda}\left|\int_0^1 \{g_{\varphi,\lambda}(u) - g_{\varphi,\lambda}(\sigma_\lambda(x))\}K_h(u-\sigma_\lambda(x))du\right|$$

as an envelope of $\mathcal{F}_3$. Observe that $\mathbf{E}[\bar{\zeta}(S_i,X_i)^2]$ is bounded by

$$4\sup_{x\in\mathcal{S}_X}\mathbf{E}\left[\tilde{\psi}^2(S)|X=x\right]\cdot\int_0^1\sup_{(\varphi,\lambda)\in\Phi\times\Lambda}\left[\int_0^1\{g_{\varphi,\lambda}(t_2)-g_{\varphi,\lambda}(t_1)\}K_h(t_2-t_1)dt_2\right]^2 dt_1. \tag{14}$$

[2]The result is replicated in Theorem A.2 of van der Vaart (1996).

By change of variables, the integral inside the bracket becomes

$$\int_{[-t_1/h,(1-t_1)/h]\cap[-1,1]} \{g_{\varphi,\lambda}(t_1+ht_2)-g_{\varphi,\lambda}(t_1)\}K(t_2)dt_2.$$

After tedious algebra (using the symmetry of K), we can show that the outside integral in (14) is $O(h^3)$. Therefore, $||\bar{\zeta}||_2 = O(h^{3/2})$ as $n\to\infty$. Applying this and the maximal inequality of Pollard (1989),

$$\mathbf{E}\left[\sup_{\pi\in\Pi_n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{E}_i[q_{n,ij}^{\pi}]-\mathbf{E}[q_{n,ij}^{\pi}])\right|\right] \le C\int_0^{Ch^{3/2}}\sqrt{1+\log N_{[]}(\varepsilon,\mathcal{F}_3,||\cdot||_2)}d\varepsilon$$

for some $C>0$. By (13), the last term is of order $O(h^{(3/2)\times\{1-(b_\Psi\vee[(q+1)\{b_\Phi\vee b_\Lambda\}])/2\}}\sqrt{-\log h})$. Since $b_\Psi\vee[(q+1)\{b_\Phi\vee b_\Lambda\}]<2$, we obtain the wanted result.

Proof of C5: Let us define $\tilde{\mathcal{J}}_n = \{hq_n^{\pi}(\cdot,\cdot):\pi\in\Pi\}$, where $q_n^{\pi}(Z_i,Z_j)=q_{n,ij}^{\pi}$, $Z_i=(S_i,W_i,X_i)$, and $q_{n,ij}^{\pi}$ is defined prior to (6). Take $q>4$ as in Assumption B1. Using similar arguments as in (C3), we can show that for some $C>0$, $\log N_{[]}(\varepsilon,\tilde{\mathcal{J}}_n,||\cdot||_q)\le C\varepsilon^{-(b_\Psi\vee[(q+1)\{b_\Phi\vee b_\Lambda\}])}$ for all $\varepsilon>0$. By Assumption B1, $(b_\Psi\vee[(q+1)\{b_\Phi\vee b_\Lambda\}])(1-1/q)<1$. Then, from the proof of C3,

$$\int_0^1\left\{\log N_{[]}(\varepsilon,\tilde{\mathcal{J}}_n,||\cdot||_q)\right\}^{(1-1/q)}d\varepsilon \le C\int_0^1 \varepsilon^{-(b_\Psi\vee(q+1)\{b_\Phi\vee b_\Lambda\})(1-1/q)}d\varepsilon<\infty$$

for some $C>0$. Furthermore, as in the proof of C3, we can take an envelope of $\tilde{\mathcal{J}}_n$ that is L_q-bounded. By Theorem 1 of Turki-Moalla (1998), p.878, for some small $\varepsilon>0$,

$$h\sup_{\pi\in\Pi_n}|u_{1n}(\pi)| = O_P(n^{1/2-(1-1/q)+\varepsilon}) = O_P(n^{-1/2+1/q+\varepsilon}).$$

Therefore, $\sup_{\pi\in\Pi_n}|u_{1n}(\pi)| = O_P(n^{-1/2+1/q+\varepsilon}h^{-1}) = o_P(1)$ by taking small $\varepsilon>0$ and using Assumption B2(ii) and the fact that $q>4$. Hence the proof is complete.

Proof of C6: For $i \neq j$, we write

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\mathbf{E}_j [q^{\pi}_{n,ij}] - g_{\psi,\lambda}(U_{\lambda,j}) \{ \varphi(W_j) - g_{\varphi,\lambda}(U_{\lambda,j}) \} \right) \\
= \; & \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\mathbf{E}_j \left[\left(q^{\pi}_{n,ij} - \frac{1}{\xi_{1n}(U_{\lambda,i})} g_{\psi,\lambda}(U_{\lambda,j}) \{ \varphi(W_j) - g_{\varphi,\lambda}(U_{\lambda,j}) \} \right) \right] \right) \\
& - \mathbf{E} \left[\left(\frac{\xi_{1n}(U_{\lambda,i}) - 1}{\xi_{1n}(U_{\lambda,i})} \right) \right] \times \frac{1}{\sqrt{n}} \sum_{j=1}^{n} g_{\psi,\lambda}(U_{\lambda,j}) \{ \varphi(W_j) - g_{\varphi,\lambda}(U_{\lambda,j}) \} \\
\equiv \; & E_{1n}(\pi) - E_{2n}(\pi), \text{ say.}
\end{aligned}$$

We focus on E_{1n} first. Note that

$$\begin{aligned}
& \mathbf{E} \left[\sup_{\pi \in \Pi_n} \left(\mathbf{E}_j [q^{\pi}_{n,ij}] - \frac{g_{\psi,\lambda}(U_{\lambda,j}) \{ \varphi(W_j) - g_{\varphi,\lambda}(U_{\lambda,j}) \}}{\xi_{1n}(U_{\lambda,i})} \right)^2 \right] \qquad (15) \\
= \; & \int \sup_{\pi \in \Pi_n} \left\{ \int_0^1 A_{n,\pi}(t_1, t_2, y) dt_1 \right\}^2 dF_{Y,\lambda}(y, t_2),
\end{aligned}$$

where $\int \cdot dF_{Y,\lambda}$ denotes the integration with respect to the joint distribution of $(W_i, U_{\lambda,i})$ and

$$A_{n,\pi}(t_1, t_2, y) = \frac{1}{\xi_{1n}(t_1)} \left(g_{\psi,\lambda}(t_1) \{ \varphi(y) - g_{\varphi,\lambda}(t_1) \} K_h(t_1 - t_2) - g_{\psi,\lambda}(t_2) \{ \varphi(y) - g_{\varphi,\lambda}(t_2) \} \right).$$

After some tedious algebra, we can show that the last term in (15) is $O(h^3)$ (see the proof of C4). Therefore, $\sup_{\pi \in \Pi_n} |E_{1n}(\pi)| = o_P(1)$.

We turn to E_{2n}. It is not hard to see that for all $\lambda \in \Lambda_n$,

$$\mathbf{E} \left[\left(\frac{\xi_{1n}(U_{\lambda,i}) - 1}{\xi_{1n}(U_{\lambda,i})} \right) \right] = \mathbf{E} \left[\left(\frac{\xi_{1n}(U_{0,i}) - 1}{\xi_{1n}(U_{0,i})} \right) \right] = o(1), \text{ as } h \to 0.$$

The first equality follows because $U_{\lambda,i}$ follows a uniform distribution on $[0,1]$. Furthermore, we have

$$\sup_{\pi \in \Pi_n} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(g_{\psi,\lambda}(U_{\lambda,j}) \{ \varphi(W_j) - g_{\varphi,\lambda}(U_{\lambda,j}) \} \right) \right| = O_P(1),$$

using bracketing entropy conditions for Ψ, Φ, and Λ_n, Lemma B1, and the maximal inequality of Pollard (1989). Therefore, we find that $\sup_{\pi\in\Pi_n}|E_{2n}(\pi)| = o_P(1)$. ■

Let $D_i \in \{0,1\}$ be a binary random variable and for $d \in \{0,1\}$, define $g_{\varphi,\lambda,d}(u) \equiv \mathbf{E}[\varphi(W_i)|U_{\lambda,i} = u, D_i = d]$ and $g_{\psi,\lambda,d}(u) \equiv \mathbf{E}[\psi(S_i)|U_{\lambda,i} = u, D_i = d]$. Consider the estimator:

$$\hat{g}_{\varphi,\lambda,d}(U_{n,\lambda,i}) \equiv \frac{1}{(n-1)\hat{f}_{\lambda,d}(U_{n,\lambda,i})} \sum_{j=1,j\neq i}^{n} \varphi(W_j)1\{D_j = d\}K_h\left(U_{n,\lambda,j} - U_{n,\lambda,i}\right),$$

where $\hat{f}_{\lambda,d}(U_{n,\lambda,i}) \equiv (n-1)^{-1}\sum_{j=1,j\neq i}^{n} 1\{D_j = d\}K_h(U_{n,\lambda,j} - U_{n,\lambda,i})$. Similarly as before, we define

$$\nu_{n,d}(\lambda,\varphi,\psi) \equiv \frac{\sqrt{n}}{\sum_{i=1}^{n} D_i} \sum_{i=1}^{n} \psi(S_i)D_i\left\{\hat{g}_{\varphi,\lambda,d}(U_{n,\lambda,i}) - g_{\varphi,\lambda,d}(U_{\lambda,i})\right\},$$

with $(\lambda,\varphi,\psi) \in \Lambda\times\Phi\times\Psi$. The following lemma presents variants of Lemma B2.

Lemma B3 **:** *Suppose that* Assumptions B1-B3 *hold, and let* $P_d \equiv P\{D = d\}$, *and* $\varepsilon_{\varphi,\lambda,d,i} \equiv \varphi(W_i) - g_{\varphi,\lambda,d}(U_{\lambda,i})$, $d \in \{0,1\}$.

(i) *If there exists* $\varepsilon > 0$ *such that* $P\{D_i = 1|U_{\lambda,i} = u\} \geq \varepsilon$ *for all* $(u,\lambda) \in [0,1]\times\Lambda$, *then*

$$\sup_{(\lambda,\varphi,\psi)\in\Lambda_n\times\Phi\times\Psi}\left|\nu_{n,1}(\lambda,\varphi,\psi) - \frac{1}{\sqrt{n}P_1}\sum_{i=1}^{n} D_i g_{\psi,\lambda,1}(U_{\lambda,i})\varepsilon_{\varphi,\lambda,1,i}\right| = o_P(1).$$

(ii) *If there exists* $\varepsilon > 0$ *such that* $P\{D_i = 1|U_{\lambda,i} = u\} \in [\varepsilon, 1-\varepsilon]$ *for all* $(u,\lambda) \in [0,1]\times\Lambda$, *then*

$$\sup_{(\lambda,\varphi,\psi)\in\Lambda_n\times\Phi\times\Psi}\left|\nu_{n,0}(\lambda,\varphi,\psi) - \frac{1}{\sqrt{n}P_1}\sum_{i=1}^{n} \frac{(1-D_i)P(U_{\lambda,i})g_{\psi,\lambda,1}(U_{\lambda,i})}{1-P(U_{\lambda,i})}\varepsilon_{\varphi,\lambda,0,i}\right| = o_P(1).$$

Proof of Lemma B3: Write

$$\nu_{n,1}(\lambda,\varphi,\psi) = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} D_i}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(S_i)D_i\left\{\frac{\hat{g}_{\varphi,\lambda,i}^{[1]}(U_{n,\lambda,i})}{\hat{g}_{\lambda,i}^{[1]}(U_{n,\lambda,i})} - \frac{g_{\varphi,\lambda}^{[1]}(U_{\lambda,i})}{g_{\lambda}^{[1]}(U_{\lambda,i})}\right\},$$

where $g^{[d]}_{\varphi,\lambda}(u) = \mathbf{E}[\varphi(W_i)1\{D_i = d\}|U_{\lambda,i} = u]$, $g^{[d]}_{\lambda}(u) = P\{D_i = d|U_{\lambda,i} = u\}$,

$$\begin{aligned}
\hat{g}^{[d]}_{\varphi,\lambda,i}(u) &= \frac{1}{(n-1)\hat{f}_{\lambda,d}(u)} \sum_{j=1, j\neq i}^{n} \varphi(W_j)1\{D_j = d\}K_h\left(U_{n,\lambda,j} - u\right), \text{ and} \\
\hat{g}^{[d]}_{\lambda,i}(u) &= \frac{1}{(n-1)\hat{f}_{\lambda,d}(u)} \sum_{j=1, j\neq i}^{n} 1\{D_j = d\}K_h\left(U_{n,\lambda,j} - u\right).
\end{aligned}$$

Using the arguments in the proof of Lemma B2, we can write

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(S_i)D_i\left\{\frac{\hat{g}^{[1]}_{\varphi,\lambda,i}(U_{n,\lambda,i})}{\hat{g}^{[1]}_{\lambda,i}(U_{n,\lambda,i})} - \frac{g^{[1]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}\right\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\psi(S_i)D_i}{g^{[1]}_{\lambda}(U_{\lambda,i})}\left\{\hat{g}^{[1]}_{\varphi,\lambda,i}(U_{n,\lambda,i}) - g^{[1]}_{\varphi,\lambda}(U_{\lambda,i})\right\} \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(S_i)D_i\frac{g^{[1]}_{\varphi,\lambda}(U_{\lambda,i})}{(g^{[1]}_{\lambda}(U_{\lambda,i}))^2}\left\{g^{[1]}_{\lambda}(U_{\lambda,i}) - \hat{g}^{[1]}_{\lambda,i}(U_{n,\lambda,i})\right\} + o_P(1).
\end{aligned}$$

By applying Lemma B2 to both terms, we obtain that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(S_i)D_i\left\{\frac{\hat{g}^{[1]}_{\varphi,\lambda,i}(U_{n,\lambda,i})}{\hat{g}^{[1]}_{\lambda,i}(U_{n,\lambda,i})} - \frac{g^{[1]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}\right\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{g^{[1]}_{\psi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}\left\{D_i\varphi(W_i) - g^{[1]}_{\varphi,\lambda}(U_{\lambda,i})\right\} \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{g^{[1]}_{\psi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}\frac{g^{[1]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}\left\{g^{[1]}_{\lambda}(U_{\lambda,i}) - D_i\right\} + o_P(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{g^{[1]}_{\psi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}D_i\left\{\varphi(W_i) - \frac{g^{[1]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}\right\} + o_P(1).
\end{aligned}$$

Note that

$$
\begin{aligned}
& \mathbf{E}\left[\frac{g^{[1]}_{\psi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}D_i\left\{\varphi(W_i)-\frac{g^{[1]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}\right\}\right] \\
= & \mathbf{E}\left[\mathbf{E}\left[g^{[1]}_{\psi,\lambda}(U_{\lambda,i})\left\{\varphi(W_i)-\frac{g^{[1]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}\right\}|U_{\lambda,i},D_i=1\right]\right] \\
= & \mathbf{E}\left[g^{[1]}_{\psi,\lambda}(U_{\lambda,i})\mathbf{E}\left[\varphi(W_i)-\frac{g^{[1]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}|U_{\lambda,i},D_i=1\right]\right]=0.
\end{aligned}
$$

Therefore, using similar arguments in the proof of Lemma B2, we conclude that uniformly over $(\lambda,\varphi,\psi)\in\Lambda_n\times\Phi\times\Psi$,

$$
\nu_{n,1}(\lambda,\varphi,\psi)=\frac{1}{P_1}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{g^{[1]}_{\psi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}D_i\left\{\varphi(W_i)-\frac{g^{[1]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[1]}_{\lambda}(U_{\lambda,i})}\right\}+o_P(1).
$$

Consider the second statement. Write

$$
\nu_{n,0}(\lambda,\varphi,\psi)=\frac{1}{\frac{1}{n}\sum_{i=1}^{n}D_i}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(S_i)D_i\left\{\frac{\hat{g}^{[0]}_{\varphi,\lambda,i}(U_{n,\lambda,i})}{\hat{g}^{[0]}_{\lambda,i}(U_{n,\lambda,i})}-\frac{g^{[0]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}\right\}.
$$

Using the arguments in the proof of Lemma B2, we can write

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(S_i)D_i\left\{\frac{\hat{g}^{[0]}_{\varphi,\lambda,i}(U_{n,\lambda,i})}{\hat{g}^{[0]}_{\lambda,i}(U_{n,\lambda,i})}-\frac{g^{[0]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}\right\} \\
= & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\psi(S_i)D_i}{g^{[0]}_{\lambda}(U_{\lambda,i})}\left\{\hat{g}^{[0]}_{\varphi,\lambda,i}(U_{n,\lambda,i})-g^{[0]}_{\varphi,\lambda}(U_{\lambda,i})\right\} \\
& +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(S_i)D_i\frac{g^{[0]}_{\varphi,\lambda}(U_{\lambda,i})}{(g^{[0]}_{\lambda}(U_{\lambda,i}))^2}\left\{g^{[0]}_{\lambda}(U_{\lambda,i})-\hat{g}^{[0]}_{\lambda,i}(U_{n,\lambda,i})\right\}+o_P(1) \\
= & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{g^{[1]}_{\psi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}\left\{(1-D_i)\varphi(W_i)-g^{[0]}_{\varphi,\lambda}(U_{\lambda,i})\right\} \\
& +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{g^{[1]}_{\psi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}\frac{g^{[0]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}\left\{g^{[0]}_{\lambda}(U_{\lambda,i})-(1-D_i)\right\}+o_P(1) \\
= & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{g^{[1]}_{\psi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}(1-D_i)\left\{\varphi(W_i)-\frac{g^{[0]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}\right\}+o_P(1).
\end{aligned}
$$

Note that $\mathbf{E}\left[\frac{g^{[1]}_{\psi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}(1-D_i)\left\{\varphi(W_i)-\frac{g^{[0]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}\right\}\right]$ is equal to

$$\begin{aligned}
& \mathbf{E}\left[\mathbf{E}\left[\frac{g^{[1]}_{\psi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}\left\{\varphi(W_i)-\frac{g^{[0]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}\right\}|U_{\lambda,i},D_i=0\right]\right] \\
= & \ \mathbf{E}\left[\frac{g^{[1]}_{\psi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}\mathbf{E}\left[\varphi(W_i)-\frac{g^{[0]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}|U_{\lambda,i},D_i=0\right]\right]=0.
\end{aligned}$$

From similar arguments in the proof of Lemma B2, uniformly over $(\lambda,\varphi,\psi)\in\Lambda_n\times\Phi\times\Psi$,

$$\nu_{n,0}(\lambda,\varphi,\psi)=\frac{1}{P_1}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{g^{[1]}_{\psi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}(1-D_i)\left\{\varphi(W_i)-\frac{g^{[0]}_{\varphi,\lambda}(U_{\lambda,i})}{g^{[0]}_{\lambda}(U_{\lambda,i})}\right\}+o_P(1).$$

■

2 Derivation of Asymptotic Covariance Matrices

2.1 Example 1: Sample Selection Model with Conditional Median Restrictions

Let $u_{Z,i}=Z_i-\mathbf{E}\left[Z_i|U_{0,i},D_i=1\right]$ and $u_{Y,i}=Y_i-\mathbf{E}\left[Y_i|U_{0,i},D_i=1\right]$. Let $\tilde{S}_{ZZ}$, $\tilde{S}_{ZY}$, $\tilde{\mu}_Z(u)$ and $\tilde{\mu}_Y(u)$ be $\hat{S}_{ZZ}$, $\hat{S}_{ZY}$, $\hat{\mu}_Z(u)$ and $\hat{\mu}_Y(u)$ except that $\hat{\theta}$ is replaced by θ_0. Also, let $\tilde{U}_i$ be $\hat{U}_i$ except that $\hat{\theta}$ is replaced by θ_0. Write $\tilde{S}_{ZZ}-S_{ZZ}=B_{1n}+B_{2n}$, where

$$\begin{aligned}
B_{1n} &\equiv \frac{1}{\sum_{i=1}^n D_i}\sum_{i=1}^{n}D_i\left(\tilde{u}_{Z,i}\tilde{u}_{Z,i}^{\top}-u_{Z,i}u_{Z,i}^{\top}\right) \\
B_{2n} &\equiv \frac{1}{\sum_{i=1}^n D_i}\sum_{i=1}^{n}D_i\left(u_{Z,i}u_{Z,i}^{\top}-\mathbf{E}[u_{Z,i}u_{Z,i}^{\top}|D_i=1]\right),
\end{aligned}$$

and $\tilde{u}_{Z,i} = Z_i - \tilde{\mu}_Z(\tilde{U}_i)$ and $\tilde{u}_{Y,i} = Y_i - \tilde{\mu}_Y(\tilde{U}_i)$. Under regularity conditions, we can write B_{1n} as

$$\frac{1}{\sum_{i=1}^n D_i}\sum_{i=1}^n D_i u_{Z,i}\{\tilde{u}_{Z,i} - u_{Z,i}\}^\top + \frac{1}{\sum_{i=1}^n D_i}\sum_{i=1}^n D_i\{\tilde{u}_{Z,i} - u_{Z,i}\}u_{Z,i}^\top + o_P(1/\sqrt{n}).$$

Using Lemma B3(i), we find that both the sums above are equal to $o_P(1/\sqrt{n})$. Following similar arguments for $\tilde{S}_{ZY} - S_{ZY}$, we conclude that

$$\begin{aligned}\tilde{S}_{ZZ} - S_{ZZ} &= \frac{1}{\sum_{i=1}^n D_i}\sum_{i=1}^n D_i\left(u_{Z,i}u_{Z,i}^\top - \mathbf{E}[u_{Z,i}u_{Z,i}^\top|D_i = 1]\right) + o_P(1/\sqrt{n}) \qquad (16)\\ \tilde{S}_{ZY} - S_{ZY} &= \frac{1}{\sum_{i=1}^n D_i}\sum_{i=1}^n D_i\left(u_{Z,i}u_{Y,i} - \mathbf{E}[u_{Z,i}u_{Y,i}|D_i = 1]\right) + o_P(1/\sqrt{n}).\end{aligned}$$

First write $\tilde{\beta} - \beta_0 = A_{1n} + A_{2n}$, where

$$A_{1n} \equiv \{\tilde{S}_{ZZ}^{-1} - S_{ZZ}^{-1}\}\tilde{S}_{ZY} \text{ and } A_{2n} \equiv S_{ZZ}^{-1}\{\tilde{S}_{ZY} - S_{ZY}\}.$$

From (16), $||[\tilde{S}_{ZY} \vdots \tilde{S}_{ZZ}] - [S_{ZY} \vdots S_{ZZ}]|| = O_P(1/\sqrt{n})$. Hence

$$\begin{aligned}A_{1n} &= \tilde{S}_{ZZ}^{-1}\{S_{ZZ} - \tilde{S}_{ZZ}\}S_{ZZ}^{-1}\tilde{S}_{ZY} = S_{ZZ}^{-1}\{S_{ZZ} - \tilde{S}_{ZZ}\}S_{ZZ}^{-1}\tilde{S}_{ZY} + o_P(1/\sqrt{n})\\ &= S_{ZZ}^{-1}\{S_{ZZ} - \tilde{S}_{ZZ}\}\beta_0 + o_P(1/\sqrt{n}),\end{aligned}$$

because $\beta_0 = S_{ZZ}^{-1}S_{ZY}$. Therefore, from (16), $\sqrt{n}\{\tilde{\beta} - \beta_0\}$ is equal to

$$\begin{aligned}&-S_{ZZ}^{-1}\left(\frac{1}{\sum_{i=1}^n D_i}\sum_{i=1}^n D_i\left(u_{Z,i}u_{Z,i}^\top - \mathbf{E}[u_{Z,i}u_{Z,i}^\top|D_i = 1]\right)\right)\beta_0 \qquad (17)\\ &+S_{ZZ}^{-1}\left(\frac{1}{\sum_{i=1}^n D_i}\sum_{i=1}^n D_i\left(u_{Z,i}u_{Y,i} - \mathbf{E}[u_{Z,i}u_{Y,i}|D_i = 1]\right)\right) + o_P(1/\sqrt{n}).\end{aligned}$$

As for the first term, observe that $u_{Z,i}^\top\beta_0 = u_{Y,i} - u_{v,i}$, where $u_{v,i} = v_i - \mathbf{E}\left[v_i|U_{0,i}, D_i = 1\right]$.

From this, we find that

$$
\begin{aligned}
& \left(u_{Z,i}u_{Z,i}^{\top} - \mathbf{E}[u_{Z,i}u_{Z,i}^{\top}|D_i = 1]\right)\beta_0 \\
= \ & \left(u_{Z,i}u_{Y,i} - \mathbf{E}[u_{Z,i}u_{Y,i}|D_i = 1]\right) - \left(u_{Z,i}u_{v,i} - \mathbf{E}[u_{Z,i}u_{v,i}|D_i = 1]\right).
\end{aligned}
$$

Plugging this into the first sum of (17), we find that $\tilde{\beta} - \beta_0 = S_{ZZ}^{-1}\xi_n + o_P(1/\sqrt{n})$, where

$$
\xi_n \equiv \frac{\sqrt{n}}{\sum_{i=1}^n D_i}\sum_{i=1}^n D_i \left(u_{Z,i}u_{v,i} - \mathbf{E}\left[u_{Z,i}u_{v,i}|D_i = 1\right]\right).
$$

Therefore the asymptotic variance is obtained through Assumption SS0 (i). ■

2.2 Example 2: Single-Index Matching Estimators of Treatment Effects on the Treated

Define $\tilde{\mu}(\tilde{U}_i)$ to be $\hat{\mu}(\hat{U}_i)$ except that $\hat{\theta}$ is replaced by θ_0. Write $\sqrt{n}(\tilde{\beta} - \beta_0)$ as

$$
\begin{aligned}
& \frac{1}{\frac{1}{n}\sum_{i=1}^n Z_i\sqrt{n}}\sum_{i=1}^n \left\{Z_i\left(Y_i - \mu_0(U_{0,i})\right) - \beta_0\right\} \\
& + \frac{1}{\frac{1}{n}\sum_{i=1}^n Z_i\sqrt{n}}\sum_{i=1}^n Z_i\left(\mu_0(U_{0,i}) - \tilde{\mu}(\tilde{U}_i)\right) \equiv A_{1n} + A_{2n}, \text{ say.}
\end{aligned}
$$

Let $P_d \equiv P\{Z = d\}$, $d \in \{0, 1\}$. As for A_{1n}, it is not hard to see that

$$
\begin{aligned}
A_{1n} &= \frac{1}{\frac{1}{n}\sum_{i=1}^n Z_i\sqrt{n}}\sum_{i=1}^n Z_i\left(Y_i - \mu_1(U_{0,i})\right) + \frac{1}{\frac{1}{n}\sum_{i=1}^n Z_i\sqrt{n}}\sum_{i=1}^n Z_i\left\{\mu_1(U_{0,i}) - \mu_0(U_{0,i}) - \beta_0\right\} \\
&= \frac{1}{P_1\sqrt{n}}\sum_{i=1}^n Z_i\left(Y_i - \mu_1(U_{0,i})\right) + \frac{1}{P_1\sqrt{n}}\sum_{i=1}^n Z_i\left(\mu_1(U_{0,i}) - \mu_0(U_{0,i}) - \beta_0\right) + o_P(1).
\end{aligned}
$$

As for A_{2n}, we apply Lemma B3(ii) to deduce that it is equal to

$$\frac{1}{P_1\sqrt{n}}\sum_{i=1}^{n}\frac{(1-Z_i)P(U_{0,i})}{1-P(U_{0,i})}\left(\mu_0(U_{0,i})-Y_i\right)+o_P(1).$$

Combining A_{1n} and A_{2n}, we find that $\sqrt{n}(\tilde{\beta}-\beta_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_i+o_P(1)$, where

$$\gamma_i=\frac{Z_i\varepsilon_{1,i}}{P_1}-\frac{(1-Z_i)P(U_{0,i})\varepsilon_{0,i}}{(1-P(U_{0,i}))P_1}+\frac{1}{P_1}Z_i\left(\mu_1(U_{0,i})-\mu_0(U_{0,i})-\beta_0\right),$$

and $\varepsilon_{d,i}=Y_i-\mu_d(U_{0,i})$, $d\in\{0,1\}$. Hence $V_{SM}=\mathbf{E}\gamma_i\gamma_i^{\top}$.

References

[1] Pollard, D., 1989. A maximal inequality for sums of independent processes under a bracketing condition. Unpublished manuscript.

[2] Song, K., 2009. Testing conditional independence using Rosenblatt transforms. Annals of Statistics 37, 4011-4045.

[3] Stute, W. and L. Zhu, 2005. Nonparametric checks for single-index models. Annals of Statistics 33, 1048-1083.

[4] Turki-Moalla, K., 1998. Rates of convergence and law of the iterated logarithm for U-processes. Journal of Theoretical Probability 11, 869-906.

[5] van der Vaart, A. W., 1996. New Donsker classes. Annals of Probability 24, 2128-2140.