Biometrika (2016), *pp*. 1–11
© 2016 Biometrika Trust
Printed in Great Britain
Advance Access publication on xx xx xxxx

Supplementary material for 'Kernel-based covariate functional balancing for observational studies'

BY RAYMOND K. W. WONG
Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.
raywong@stat.tamu.edu

AND KWUN CHUEN GARY CHAN
Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A.
kcgchan@u.washington.edu

S1. PROOFS AND TECHNICAL RESULTS

S1·1. *Proof of Proposition* 1

Proof of Proposition 1. Consider any $v_1, v_2 \in \mathbb{R}^r$, and $t \in [0,1]$. For $\beta \in \mathbb{R}^r$,

$$\begin{aligned}\beta^{\mathrm{T}}[\{tv_1 + (1-t)v_2\}\{tv_1 + (1-t)v_2\}^{\mathrm{T}} + B]\beta &= [\{tv_1 + (1-t)v_2\}^{\mathrm{T}}\beta]^2 + \beta^{\mathrm{T}}B\beta \\ &= \{tv_1^{\mathrm{T}}\beta + (1-t)v_2^{\mathrm{T}}\beta\}^2 + \beta^{\mathrm{T}}B\beta \\ &\le t(v_1^{\mathrm{T}}\beta)^2 + (1-t)(v_2^{\mathrm{T}}\beta)^2 + \beta^{\mathrm{T}}B\beta \\ &= t\beta^{\mathrm{T}}(v_1v_1^{\mathrm{T}} + B)\beta + (1-t)\beta^{\mathrm{T}}(v_2v_2^{\mathrm{T}} + B)\beta\end{aligned}$$

Therefore, $\sigma_{\max}[\{tv_1+(1-t)v_2\}\{tv_1+(1-t)v_2\}^{\mathrm{T}}+B] \le t\sigma_{\max}(v_1v_1^{\mathrm{T}}+B)+(1-t)\sigma_{\max}(v_2v_2^{\mathrm{T}}+B)$.□

S1·2. *Proof of Theorems* 1 *and* 2

We begin with several definitions that will be used throughout the theoretical development. Write $w^* = (w_1^*, \ldots, w_N^*)^{\mathrm{T}} = [\{\pi(X_1)\}^{-1}, \ldots, \{\pi(X_N)\}^{-1}]^{\mathrm{T}}$ and

$$F_{N,\lambda_1,\lambda_2}(w) = \sup_{u \in \widetilde{\mathcal{H}}_N} \left\{ S_N(w,u) - \lambda_1 \|u\|_{\mathcal{H}}^2 \right\} + \lambda_2 V_N(w).$$

Obviously $w^* \ge 1$. Due to the definition of the proposed estimator, we have $F_{N,\lambda_1,\lambda_2}(\widehat{w}) \le F_{N,\lambda_1,\lambda_2}(w^*)$. This implies that for any $f \in \widetilde{\mathcal{H}}_N$,

$$S_N(\widehat{w}, f) - \lambda_1 \|f\|_{\mathcal{H}}^2 + \lambda_2 V_N(\widehat{w}) \le S_N(w^*, u^*) - \lambda_1 \|u^*\|_{\mathcal{H}}^2 + \lambda_2 V_N(w^*), \tag{S1}$$

where $u^* = \operatorname{argmin}_{u \in \widetilde{\mathcal{H}}_N} \{S_N(w^*, u) - \lambda_1 \|u\|_{\mathcal{H}}^2\}$ and its existence is shown in §2·3. Since $S_N(\widehat{w}, u) = 0$ for any $u \in \mathcal{H}$ such that $\|u\|_N = 0$, (S1) also implies that, for any $u \in \mathcal{H}$,

$$S_N(\widehat{w}, u) - \lambda_1 \|u\|_{\mathcal{H}}^2 + \lambda_2 V_N(\widehat{w}) \|u\|_N^2 \le \left\{ S_N(w^*, u^*) - \lambda_1 \|u^*\|_{\mathcal{H}}^2 + \lambda_2 V_N(w^*) \right\} \|u\|_N^2. \tag{S2}$$

In below, we adopt several choices of $f \in \widetilde{\mathcal{H}}_N$ in (S1) and $u \in \mathcal{H}$ in (S2) to obtain various results. For instance, one obvious candidate of $f \in \widetilde{\mathcal{H}}$ is the constant function z where $z(x) \equiv 1$. On the other hand, the control of $S_N(w^*, u^*)$ is given in the following Lemma S1, whose proof is given in §S1·4, so as to control the right-hand side of (S1) and (S2).

LEMMA S1. *Suppose Assumptions* 1 *and* 2 *hold. Let* $w^* = (w_1^*, \ldots, w_N^*)^{\mathrm{T}} = [\{\pi(X_1)\}^{-1}, \ldots, \{\pi(X_N)\}^{-1}]^{\mathrm{T}}$. *There exists a constant* $c \geq 0$ *such that for all* $T \geq c$,

$$\mathrm{pr}\left\{\sup_{u\in\widetilde{\mathcal{H}}_N} \frac{NS_N(w^*,u)}{\|u\|_{\mathcal{H}}^{d/\ell}} \geq T^2\right\} \leq c\exp\left(-\frac{T^2}{c^2}\right).$$

Moreover, by central limit theorem, $V_N(w^*) = V + O_p(N^{-1/2})$ where $V = E\{\pi(X_1)^{-1}\}$. To prove Theorem 1, it suffices to establish the following two lemmas (Lemmas S2 and S3). The proof is given in §S1·4.

LEMMA S2. *Suppose Assumptions* 1 *and* 2 *hold. If* $\lambda_1 \asymp N^{-1}$ *and* $\lambda_2 \asymp N^{-1}$, *we have* $S_N(\widehat{w}, z) = O_p(N^{-1})$ *and* $V_N(\widehat{w}) = O_p(1)$. *Moreover, there exists a constant* $W > 0$ *such that* $E\{V_N(\widehat{w})\} \leq W$.

*Proof of Lemma S*2. Taking f as z (constant function of value 1) in (S1), we obtain a basic inequality:

$$S_N(\widehat{w}, z) + \lambda_1\|u^*\|_{\mathcal{H}}^2 + \lambda_2 V_N(\widehat{w}) \leq S_N(w^*, u^*) + \lambda_1\|z\|_{\mathcal{H}}^2 + \lambda_2 V_N(w^*). \tag{S3}$$

By Lemma S1, there exists a constant c such that for all $T \geq c$, $\mathrm{pr}\{S_N(w^*, u^*) \leq T^2 N^{-1}\|u^*\|_{\mathcal{H}}^{d/\ell}\} \geq 1 - c\exp(-T^2/c^2)$.

Let $\widetilde{E}_{N,1}$, $\widetilde{E}_{N,2}$ and $\widetilde{E}_{N,3}$ be the events that $S_N(w^*, u^*)$ is the largest in right-hand side of (S3), that $\lambda_1\|z\|_{\mathcal{H}}^2$ is the largest in right-hand side of (S3), and that $\lambda_2 V_N(w^*)$ is the largest in right-hand side of (S3), respectively. Note that they are not necessarily disjoint. We write $E_{N,1} = \widetilde{E}_{N,1}$, $E_{N,2} = \widetilde{E}_{N,2} \setminus \widetilde{E}_{N,1}$ and $E_{N,3} = \widetilde{E}_{N,3} \setminus (\widetilde{E}_{N,1} \cup \widetilde{E}_{N,2})$. Therefore $\{E_1, E_2, E_3\}$ forms a partition of the sample space. We can further divide the event $E_{N,1}$ into two disjoint events, $E_{N,1,T} = E_{N,1} \cap \{S_N(w^*, u^*) \leq T^2N^{-1}\|u^*\|_{\mathcal{H}}^{d/\ell}\}$ and $\check{E}_{N,1,T} = E_{N,1} \cap \{S_N(w^*, u^*) > T^2N^{-1}\|u^*\|_{\mathcal{H}}^{d/\ell}\}$. Note that $\{E_{N,1,T} \cup \check{E}_{N,1,T} \cup E_{N,2} \cup E_{N,3}\}$ forms a partition of the sample space. We analyze (S3) on these events.

Case (i): On $E_{N,1,T}$, (S3) leads to $S_N(\widehat{w}, z) \leq T^{4d/(2\ell-d)}\lambda_1^{-d/(2\ell-d)}N^{-2\ell/(2\ell-d)}$, $\|u^*\|_{\mathcal{H}} \leq T^{2\ell/(2\ell-d)}\lambda_1^{-\ell/(2\ell-d)}N^{-\ell/(2\ell-d)}$ and $\lambda_2 V_N(\widehat{w}) \leq T^{4d/(2\ell-d)}\lambda_1^{-d/(2\ell-d)}N^{-2\ell/(2\ell-d)}$.

Case (ii): On $E_{N,2}$, (S3) leads to $S_N(\widehat{w}, z) \leq 3\lambda_1\|z\|_{\mathcal{H}}^2$, $\|u^*\|_{\mathcal{H}} \leq 3\|z\|_{\mathcal{H}}$ and $\lambda_2 V_N(\widehat{w}) \leq 3\lambda_1\|z\|_{\mathcal{H}}^2$.

Case (iii): On $E_{N,3}$, (S3) leads to $S_N(\widehat{w}, z) \leq 3\lambda_2 V_N(w^*)$, $\lambda_1\|u^*\|_{\mathcal{H}}^2 \leq 3\lambda_2 V_N(w^*)$ and $\lambda_2 V_N(\widehat{w}) \leq 3\lambda_2 V_N(w^*)$.

Now, we focus on $S_N(\widehat{w}, z)$:

$$\begin{aligned}
&\mathrm{pr}\left[S_N(\widehat{w}, z) \leq \max\left\{T^{4d/(2\ell-d)}\lambda_1^{-d/(2\ell-d)}N^{-2\ell/(2\ell-d)},\, 3\lambda_1\|z\|_{\mathcal{H}}^2,\, 3\lambda_2 V_N(w^*)\right\}\right] \\
&= \sum_{i=1}^{3}\mathrm{pr}\left[S_N(\widehat{w}, z) \leq \max\left\{T^{4d/(2\ell-d)}\lambda_1^{-d/(2\ell-d)}N^{-2\ell/(2\ell-d)},\, 3\lambda_1\|z\|_{\mathcal{H}}^2,\, 3\lambda_2 V_N(w^*)\right\} \cap E_{N,i}\right] \\
&\geq \mathrm{pr}(E_{N,1,T}) + \mathrm{pr}\left[\left\{S_N(\widehat{w}, z) \leq T^{4d/(2\ell-d)}\lambda_1^{-d/(2\ell-d)}N^{-2\ell/(2\ell-d)}\right\} \cap \check{E}_{N,1,T}\right] + \mathrm{pr}(E_{N,2}) + \mathrm{pr}(E_{N,3}) \qquad \text{(S4)} \\
&= 1 - \mathrm{pr}\left[\left\{S_N(\widehat{w}, z) > T^{4d/(2\ell-d)}\lambda_1^{-d/(2\ell-d)}N^{-2\ell/(2\ell-d)}\right\} \cap \check{E}_{N,1,T}\right]
\end{aligned}$$

$$\geq 1-\mathrm{pr}(\check{E}_{N,1,T}) \geq 1-\mathrm{pr}\left\{S_N(w^*,u^*) > T^2N^{-1}\|u^*\|_{\mathcal{H}}^{d/\ell}\right\} = \mathrm{pr}\left\{S_N(w^*,u^*) \leq T^2N^{-1}\|u^*\|_{\mathcal{H}}^{d/\ell}\right\}$$
$$\geq 1-c\exp(-T^2/c^2),$$

for all $T \geq c$, where (S4) follows from the above analyses of Cases (i), (ii) and (iii). We can show that $N^{1/2}\{V_N(w^*)-V\}$ converges to $N(0,\sigma_V^2)$ in distribution by central limit theorem, where $V = E[\{\pi(X_1)\}^{-1}] < \infty$ and $\sigma_V^2 = \{1-\pi(X_1)\}/\pi(X_1)^3 < \infty$. Therefore $S_N(\widehat{w},z) = O_p(N^{-1})$ under the condition that $\lambda_1 \asymp N^{-1}$ and $\lambda_2 = O(N^{-1/2})$. If $\lambda_1 \asymp N^{-1}$ and $\lambda_2 \asymp N^{-1}$, similar arguments lead to $V_N(\widehat{w}) = O_p(1)$.

Now, we focus on $E\{V_N(\widehat{w})\}$ under $\lambda_2 > 0$. The requirements that $\lambda_1 \asymp N^{-1}$ and $\lambda_2 \asymp N^{-1}$ imply $\widetilde{B}_1N^{-1} \leq \lambda_1 \leq B_1N^{-1}$ and $\widetilde{B}_2N^{-1} \leq \lambda_2 \leq B_2N^{-1}$ for some positive constants $\widetilde{B}_1$, B_1, $\widetilde{B}_2$ and B_2. We derive bounds for each term in the following decomposition of $E\{V_N(\widehat{w})\}$:

$$E\{V_N(\widehat{w}) \mid E_{N,1}\}\mathrm{pr}(E_{N,1}) + E\{V_N(\widehat{w}) \mid E_{N,2}\}\mathrm{pr}(E_{N,2}) + E\{V_N(\widehat{w}) \mid E_{N,3}\}\mathrm{pr}(E_{N,3}).$$

The first term: Fix $\widetilde{c} = \max\{c, 2\widetilde{B}_1^{-d/(2\ell-d)}\widetilde{B}_2^{-1}, 2\}$ and $a > 0$ such that $\widetilde{c} > \widetilde{B}_1^{-d/(2\ell-d)}\widetilde{B}_2^{-1}\widetilde{c}^{4da/(2\ell-d)}$ and $4da/(2\ell-d) < 1$. That means, a is a constant fulfilling $\min[(2\ell-d)\log\{\widetilde{c}\widetilde{B}_1^{d/(2\ell-d)}\widetilde{B}_2\}/4d\log\widetilde{c}, (2\ell-d)/4d] > a > 0$.

$$\begin{aligned}
E\{V_N(\widehat{w}) \mid E_{N,1}\}\mathrm{pr}(E_{N,1}) &= \int_0^\infty \mathrm{pr}\{V_N(\widehat{w}) > t \mid E_{N,1}\}\mathrm{pr}(E_{N,1})dt \\
&= \int_0^\infty \mathrm{pr}[\{V_N(\widehat{w}) > t\} \cap E_{N,1}]dt \\
&\leq \widetilde{c} + \int_{\widetilde{c}}^\infty \mathrm{pr}[\{V_N(\widehat{w}) > t\} \cap E_{N,1,t^a}]dt + \int_{\widetilde{c}}^\infty \mathrm{pr}[\{V_N(\widehat{w}) > t\} \cap \check{E}_{N,1,t^a})]dt
\end{aligned}$$

Due to the construction of a,

$$\int_{\widetilde{c}}^\infty \mathrm{pr}[\{V_N(\widehat{w}) > t\} \cap E_{N,1,t^a}]dt \leq \int_{\widetilde{c}}^\infty \mathrm{pr}\{t^{4da/(2\ell-d)}\widetilde{B}_1^{-d/(2\ell-d)}\widetilde{B}_2^{-1} \geq V_N(\widehat{w}) > t\}dt = 0.$$

Now, we look at the last term

$$\int_{\widetilde{c}}^\infty \mathrm{pr}[\{V_N(\widehat{w}) > t\} \cap \check{E}_{N,1,t^a}]dt \leq \int_{\widetilde{c}}^\infty \mathrm{pr}(\check{E}_{N,1,t^a})dt \leq \int_{\widetilde{c}}^\infty c\exp\left(\frac{-t^{2a}}{c^2}\right)dt = -c\frac{c^{1/a}\Gamma(1/2a, t^{2a}/c^2)}{2a}\Bigg|_{\widetilde{c}}^\infty,$$

which is bounded due to the fact that $\Gamma(s,x)/(x^{s-1}e^{-x}) \to 1$ as $x \to \infty$. Therefore $E\{V_N(\widehat{w}) \mid E_{N,1}\}\mathrm{pr}(E_{N,1}) < \infty$.

The second term: As shown in Case (ii):

$$E\{V_N(\widehat{w}) \mid E_{N,2}\}\mathrm{pr}(E_{N,2}) \leq 3B_1\widetilde{B}_2^{-1}\|z\|_{\mathcal{H}}^2 < \infty$$

The third term: As shown in Case (iii):

$$E\{V_N(\widehat{w}) \mid E_{N,3}\}\mathrm{pr}(E_{N,3}) \leq 3B_2E\{V_N(w^*) \mid E_{N,3}\}\mathrm{pr}(E_{N,3}) \leq 3B_2E\{V_N(w^*)\} = B_2V < \infty.$$

Combining the above results, there exists a constant $W > 0$ such that $E\{V_N(\widehat{w})\} \leq W$. □

LEMMA S3. *Suppose Assumptions* 1-3 *hold. If* $\lambda_1 \asymp N^{-1}$ *and* $\lambda_2 = O(N^{-1})$*, then* $S_N(\widehat{w},m) = O_p(N^{-1})\|m\|_N^2$*. Further, if* $\lambda_2 \asymp N^{-1}$*, there exists a constant* $S^2 > 0$ *such that* $E\{NS_N(\widehat{w},m)\} \leq S^2$.

*Proof of Lemma S*3. Rearranging the terms in (S2), we obtain the basic inequality:

$$S_N(\widehat{w},m)+\lambda_1\|u^*\|_{\mathcal{H}}^2\|m\|_N^2+\lambda_2 V_N(\widehat{w})\|m\|_N^2 \le S_N(w^*,u^*)\|m\|_N^2+\lambda_1\|m\|_{\mathcal{H}}^2+\lambda_2 V_N(w^*)\|m\|_N^2. \tag{S5}$$

By Lemma S1, we have $S_N(w^*,u^*)=O_p(N^{-1})\|u^*\|_{\mathcal{H}}^{d/\ell}$. Now we compare different scenarios of (S5):

Case (i): Suppose that $S_N(w^*,u^*)\|m\|_N^2$ is the largest in right-hand side of (S5). If $\|m\|_N\neq 0$, we have $\|u^*\|_{\mathcal{H}}\le \lambda_1^{-\ell/(2\ell-d)}O_p\{N^{-\ell/(2\ell-d)}\}$ and therefore $S_N(\widehat{w},m)\le \lambda_1^{-d/(2\ell-d)}O_p\{N^{-2\ell/(2\ell-d)}\}\|m\|_N^2$. As if $\|m\|_N=0$, we have $S_N(\widehat{w},m)=0\le \lambda_1^{-d/(2\ell-d)}O_p\{N^{-2\ell/(2\ell-d)}\}\|m\|_N^2$.

Case (ii): Suppose that $\lambda_1\|m\|_{\mathcal{H}}^2$ is the largest in right-hand side of (S5). We obtain $S_N(\widehat{w},m)\le 3\lambda_1\|m\|_{\mathcal{H}}^2$.

Case (iii): Suppose that $\lambda_2 V_N(w^*)\|m\|_N^2$ is the largest in right-hand side of (S5). We obtain $S_N(\widehat{w},m)\le 3\lambda_2\{V+O_p(N^{-1/2})\}\|m\|_N^2$.

Due to Lemma S7 in §S1·4, $\|m\|_N\le R\|m\|_{\mathcal{H}}<\infty$. Overall, we have

$$S_N(\widehat{w},m)=O_p\left[\max\left\{\lambda_1^{-d/(2\ell-d)}N^{-2\ell/(2\ell-d)}\|m\|_N^2,\ \lambda_1\|m\|_{\mathcal{H}}^2,\ \lambda_2\|m\|_N^2\right\}\right].$$

Since $S_N(\widehat{w},m)=0$ if $\|m\|_N^2=0$, we have $S_N(\widehat{w},m)=O_p(N^{-1})\|m\|_N^2$ due to the conditions of λ_1 and λ_2.

Next, suppose $\lambda_2\asymp N^{-1}$. Based on the exponential inequality in Lemma S1, one could apply a similar argument of Lemma S2, and show that there exists a constant $\widetilde{S}^2>0$ such $E\{N^2\widetilde{S}_N^2(\widehat{w},m)\}\le\widetilde{S}^2$ where

$$\widetilde{S}_N(\widehat{w},m)=\begin{cases}S_N(\widehat{w},m/\|m\|_N), & \text{if } \|m\|_N\neq 0,\\ 0, & \text{if } \|m\|_N=0.\end{cases}$$

Moreover,

$$\begin{aligned}E\{NS_N(\widehat{w},m)\}=E\left\{N\widetilde{S}_N(\widehat{w},m)\|m\|_N^2\right\}&\le\frac{1}{2}\left[E\{N^2\widetilde{S}_N^2(\widehat{w},\widetilde{m})\}+E(\|m\|_N^4)\right]\\&\le\frac{1}{2}\left\{\widetilde{S}^2+\frac{\int m^4dP}{N}+\frac{N-1}{N}\left(\int m^2dP\right)^2\right\}\\&\le\frac{1}{2}\left\{\widetilde{S}^2+\int m^4dP+\left(\int m^2dP\right)^2\right\}.\end{aligned}$$

Due to Lemma S7 in §S1·4, $\int m^2dP<\infty$ and $\int m^4dP<\infty$. □

Proof of Theorem 2. Recall the decomposition:

$$\frac{1}{N}\sum_{i=1}^N T_i\widehat{w}_iY_i=\frac{1}{N}\sum_{i=1}^N(T_i\widehat{w}_i-1)m(X_i)+\frac{1}{N}\sum_{i=1}^N T_i\widehat{w}_i\varepsilon_i+\left[\frac{1}{N}\sum_{i=1}^N m(X_i)-E\{Y(1)\}\right]+E\{Y(1)\}.$$

Due to Lemma S7 in §S1·4, $\|m\|_2^2 = \int m^2 dP < \infty$. Since $X_1, \ldots, X_N$ are i.i.d., we can show that $\|m\|_N = \int m^2 dP + o_p(1)$. Therefore, the first term can be controlled:

$$\left| \frac{1}{N} \sum_{i=1}^{N} (T_i \widehat{w}_i - 1) m(X_i) \right| = S_N(\widehat{w}, m)^{1/2} = O_p(N^{-1/2}) \|m\|_2 + o_p(N^{-1/2}),$$

due to Theorem 1. Moreover, $E\{N S_N(\widehat{w}, m)\} < \infty$ due to Theorem 1. As for the second term, we write $\widehat{\delta_i} = T_i \widehat{w}_i$. Under Assumption 4, we have $E(\varepsilon_i \mid \widehat{\delta_1}, \ldots, \widehat{\delta_N}) = 0$. Therefore,

$$\operatorname{var}\left(\frac{1}{N} \sum_{i=1}^{N} T_i \widehat{w}_i \varepsilon_i \right) = E\left\{ \operatorname{var}\left(\frac{1}{N} \sum_{i=1}^{N} \widehat{\delta_i} \varepsilon_i \,\middle|\, \widehat{\delta_1}, \ldots, \widehat{\delta_N} \right) \right\} \le \frac{\sigma^2}{N} E\{V_N(\widehat{w})\} \le \frac{\sigma^2 W}{N},$$

due to Theorem 1. Therefore, $N^{-1} \sum_{i=1}^{N} T_i \widehat{w}_i \varepsilon_i = O_p(N^{-1/2})$. The above derivation also implies that

$$E\left\{ N^{-1/2} \sum_{i=1}^{N} (T_i \widehat{w}_i - 1) \varepsilon_i \right\}^2 < \infty.$$

Finally, by central limit theorem, we have

$$\left[N^{-1} \sum_{i=1}^{N} m(X_i) - E\{Y(1)\} \right] = O_p(N^{-1/2}),$$

due to Assumption 4. Also,

$$E\left[N^{-1/2} \sum_{i=1}^{N} m(X_i) - E\{Y(1)\} \right]^2 < \infty.$$

Therefore, $\sum_{i=1}^{N} T_i \widehat{w}_i Y_i / N - E\{Y(1)\} = O_p(N^{-1/2})$ and $N^{1/2}[\sum_{i=1}^{N} T_i \widehat{w}_i Y_i / N - E\{Y(1)\}]$ has bounded variance. □

S1·3. *Proof of Theorem* 3

LEMMA S4. *Suppose Assumptions* 1 *and* 2 *hold. Assume* $\lambda_1 = O(N^{-1})$ *and* $\lambda_1^{-1} = o\{\lambda_2^{(2\ell-d)/d} N^{2\ell/d}\}$. *We have* $V_N(\widetilde{w}) \le V\{1 + o_p(1)\}$ *where* $V = E[\{\pi(X_1)\}^{-1}]$. *Moreover, there exists a constant* $W' > 0$ *such that* $E\{V_N(\widetilde{w})\} \le W'$.

*Proof of Lemma S*4. Taking f as z (constant function of value 1) in (S1), we obtain the basic inequality:

$$S_N(\widetilde{w}, z) + \lambda_1 \|u^*\|_{\mathcal{H}}^2 + \lambda_2 V_N(\widetilde{w}) \le S_N(w^*, u^*) + \lambda_1 \|z\|_{\mathcal{H}}^2 + \lambda_2 V_N(w^*), \tag{S6}$$

for all large N such that $1 \le BN^{1/3}$. By Lemma S1, $S_N(w^*, u^*) = O_p(N^{-1}) \|u^*\|_{\mathcal{H}}^{d/\ell}$. Moreover, it is easy to show that $V_N(w^*) = V + O_p(N^{-1/2})$. Due to the condition of λ_1 and λ_2, we have $\lambda_1^{-d/(2\ell-d)} N^{-2\ell/(2\ell-d)} = o(\lambda_2)$ which implies $(\lambda_1 N)^{-d/(2\ell-d)} = o(\lambda_1^{-1} \lambda_2)$. As $\lambda_1 N = O(1)$, therefore $\lambda_1^{-1} \lambda_2 \to \infty$.

$$\lambda_1 \|z\|_{\mathcal{H}}^2 + \lambda_2 V_N(\widetilde{w}) = \lambda_2 \{o(1) + V + O_p(N^{-1/2})\} = \lambda_2 V\{1 + o_p(1)\}.$$

Now, we come back to (S6). Let $\mathcal{A}$ be the event that $S_N(w^*, u^*) \le \lambda_1 \|z\|_{\mathcal{H}}^2 + \lambda_2 V_N(\widetilde{w})$. On the event $\mathcal{A}^c$, from (S6), we obtain $\|u^*\|_{\mathcal{H}} \le \lambda_1^{-\ell/(2\ell-d)} O_p\{N^{-\ell/(2\ell-d)}\}$ which implies

$S_N(\widetilde{w},u^*) \le \lambda_1^{-d/(2\ell-d)} O_p\{N^{-2\ell/(2\ell-d)}\} = o_p(\lambda_2)$ due to the conditions of λ_1 and λ_2. Notice that, on $\mathcal{A}^c$, we also have $S_N(w^*,u^*) > \lambda_1\|z\|_{\mathcal{H}}^2 + \lambda_2 V_N(\widetilde{w}) = \lambda_2 V\{1+o_p(1)\}$. This implies that $\mathrm{pr}(\mathcal{A}^c) \to 0$ as $N \to \infty$. Therefore we only have to focus on $\mathcal{A}$. From (S6), we obtain $\lambda_1\|u^*\|_{\mathcal{H}}^2 \le 2\lambda_2\{V + o_p(1)\}$. In this case, $\|u^*\|_{\mathcal{H}}^2 \le 2\lambda_1^{-1}\lambda_2 V\{1+o_p(1)\} = O_p(\lambda_1^{-1}\lambda_2)$. Therefore $S_N(w^*,u^*) = O_p(N^{-1})\|u^*\|_{\mathcal{H}}^{d/\ell} = O_p\{N^{-1}(\lambda_1^{-1}\lambda_2)^{d/(2\ell)}\} = o_p(\lambda_2)$. This implies that the right-hand side of (S6) is $\lambda_2 V\{1+o_p(1)\}$. Hence $\lambda_2 V_N(\widetilde{w}) \le \lambda_2 V\{1+o_p(1)\}$. Finally, using a similar but simpler argument in Lemma S2, one can show that there exists a constant $W' > 0$ such that $E\{V_N(\widetilde{w})\} \le W'$. □

LEMMA S5. *Suppose Assumptions* 1 *and* 2 *hold. Let* $h = m - \widehat{m} \in \mathcal{H}$ *such that* $\|h\|_N = o_p(1)$ *and* $\|h\|_{\mathcal{H}} = O_p(1)$. *Further, assume* $\lambda_1 = o(N^{-1})$, $\lambda_1^{-1}\|h\|_N^{2(2\ell-d)/d} = o_p(N)$ *and* $\lambda_2\|h\|_N^2 = o_p(N^{-1})$. *Then* $S_N(\widetilde{w},h) = o_p(N^{-1})$. *Moreover, there exists a constant* $S' > 0$ *such that* $E\{NS_N(\widetilde{w},h)\} \le S'$.

*Proof of Lemma S*5. Rearranging the terms in (S2), we obtain the basic inequality:

$$S_N(\widetilde{w},h) + \lambda_1\|u^*\|_{\mathcal{H}}^2\|h\|_N^2 + \lambda_2 V_N(\widetilde{w})\|h\|_N^2 \le S_N(w^*,u^*)\|h\|_N^2 + \lambda_1\|h\|_{\mathcal{H}}^2 + \lambda_2 V_N(w^*)\|h\|_N^2, \quad \text{(S7)}$$

for all large N such that $C \le BN^{1/3}$. The rest of the proof is similar to the proof of Lemma S3 but with different conditions of λ_1 and λ_2. □

Proof of Theorem 3. Recall the decomposition:

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^N T_i\widetilde{w}_i\{Y_i - \widehat{m}(X_i)\} + \frac{1}{N}\sum_{i=1}^N \widehat{m}(X_i) \\ &= \frac{1}{N}\sum_{i=1}^N (T_i\widetilde{w}_i - 1)h(X_i) + \frac{1}{N}\sum_{i=1}^N T_i\widetilde{w}_i\varepsilon_i + \left[\frac{1}{N}\sum_{i=1}^N m(X_i) - E\{Y(1)\}\right] + E\{Y(1)\}\,. \end{aligned}$$

Note that the assumed conditions imply the conditions of Lemmas S4 and S5. By Lemma S5, the first term of the decomposition is $o_p(N^{-1/2})$. By dominated convergence theorem, with Skorohod Representation Theorem to extend its result to weakly convergent sequence of random variables, we have $\mathrm{var}\{N^{-1/2}\sum_{i=1}^N (T_i\widetilde{w}_i - 1)h(X_i)\} \le E\{NS_N(\widetilde{w},h)\} \to 0$ using Lemma S5. Write

$$Z_N = N^{1/2}\left(\frac{1}{N}\sum_{i=1}^N T_i\widetilde{w}_i\varepsilon_i + \left[\frac{1}{N}\sum_{i=1}^N m(X_i) - E\{Y(1)\}\right]\right).$$

It is obvious that $\mathrm{var}(Z_N) = \mathrm{var}\{m(X_1)\} + \sigma^2 E V_N(\widetilde{w})$. By Lemma S4, we have $\limsup_N E V_N(\widetilde{w}) \le E\{V + o_p(1)\} = V$ using dominated convergence theorem.

Now, since the first term of the decomposition is $o_p(N^{-1/2})$, we focus on Z_N. We will utilize Theorem S1, by setting $\tau^2 = \mathrm{var}\{m(X_1)\}$, $g^2(\mathcal{D}_N) = \sigma^2 V_N(\widetilde{w})$, and

$$A_j = \frac{m(X_j) - E\{Y(1)\}}{[N\mathrm{var}\{m(X_1)\}]^{1/2}}, \quad \mathcal{B}_j = \{X_j, T_j\}, \quad C_j = \frac{T_j\widetilde{w}_j\varepsilon_j}{(\sigma^2\sum_{i=1}^N T_i\widetilde{w}_i^2)^{1/2}}, \quad (j = 1,\dots,N).$$

Write $\mathcal{D}_N = \{A_1,\dots,A_N,\mathcal{B}_1,\dots,\mathcal{B}_N\}$. By the definition of $\widetilde{w}_i$ $(i : T_i = 1)$, $1 \le \widetilde{w}_i \le BN^{1/3}$ for all i. Therefore, $(\sum_{i=1}^N T_i\widetilde{w}_i^2)^{-1} = O_p(N)$ and $\max_i|\widetilde{w}_i| = o_p(N^{1/2})$. Moreover, $\max_i E|\varepsilon_i|^3 < \infty$ by assumption. Hence

$$0 \le E\left(\sum_{i=1}^{N} |C_i|^3 \mid \mathcal{D}_N\right) = \frac{(\max_i E|\varepsilon_i|^3)\sum_{j=1}^{N} T_j \widetilde{w}_j^3}{\sigma^3(\sum_{i=1}^{N} T_i \widetilde{w}_i^2)^{3/2}} \le \frac{(\max_i E|\varepsilon_i|^3)\max_i |\widetilde{w}_i|}{\sigma^3(\sum_{i=1}^{N} T_i \widetilde{w}_i^2)^{1/2}} = o_p(1).$$

By Lemma S4, we have $E\{g^2(\mathcal{D}_N)\} \le M$ and $g^2(\mathcal{D}_N) \le M + o_p(1)$, by taking $M = \sigma^2 \max\{W', V\}$. Write

$$Z_n = \tau \sum_{j=1}^{n} A_j + g(\mathcal{D}_N)\sum_{j=1}^{n} C_j = N^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N} T_i \widetilde{w}_i \varepsilon_i + \left[\frac{1}{N}\sum_{i=1}^{N} m(X_i) - E\{Y(1)\}\right]\right)$$

$$Z_n^* = \tau F + g(\mathcal{D}_N)\sum_{j=1}^{n} \mathrm{var}(C_j \mid \mathcal{D}_N)^{1/2} G_j = [\mathrm{var}\{m(X_1)\}]^{1/2} F + \sigma N^{-1/2}\sum_{j=1}^{N} T_j \widetilde{w}_j G_j$$

where $F, G_1, \ldots, G_N$ are i.i.d. standard normal random variables independent of $C_1, \ldots, C_N$ and $\mathcal{D}_N$. Let ϕ_N and ϕ_N^* be the corresponding characteristic function of Z_N and Z_N^* respectively. Applying Theorem S1, we have $|\phi_N(t) - \phi_N^*(t)| \to 0$ for every $t \in \mathbb{R}$ and ϕ_N^* is twice differentiable.

S1·4. *Proof of Lemma S*1

LEMMA S6. *For $d/\ell < 2$, there exists a constant A such that the uniform entropy $H_\infty(\xi, \{u \in \mathcal{H} : \|u\|_{\mathcal{H}} \le 1\}) \le A\xi^{-d/\ell}$ for $\xi > 0$, where the uniform entropy is defined in Definition 2.3 of van de Geer (2000).*

*Proof of Lemma S*6. This is shown by Birman & Solomyak (1967) and the fact that $\mathcal{H}$ is a subspace of the Sobolev space $\mathcal{W}^{\ell,2}([0,1]^d)$. □

LEMMA S7. *There exists a constant R such that* $\sup_{\{u\in\mathcal{H}:\|u\|_{\mathcal{H}}\le 1\}} \|u\|_\infty \le R$.

*Proof of Lemma S*7. This is due to Lemma 2.1 of Lin (2000) and norm equivalence. □

*Proof of Lemma S*1. Let $\delta_i = T_i w_i^* - 1$. Note that the conditional expectation $E(\delta_i \mid X_i) = 0$. We will focus on the empirical process $\{N^{-1/2}\sum_{i=1}^{N} \delta_i u(X_i) : u \in \widetilde{\mathcal{H}}\}$. Due to Assumption 1, $0 < w_i^* \le C$ for all $i = 1, \ldots, N$. Therefore, $\delta_i (i = 1, \ldots, N)$ are uniformly sub-Gaussian: there exist constants K and σ_0^2, independent of X_i $(i = 1, \ldots, N)$, such that

$$\max_{i=1,\ldots,N} K^2\left[E\left(e^{|\delta_i|^2/K^2} \mid \{X_i\}_{i=1}^N\right) - 1\right] \le \sigma_0^2.$$

For instance, take $K = \max\{|C-1|, 1\}$ and $\sigma_0^2 = K^2(e-1)$, we have

$$K^2\left[E\left(e^{|\delta_i|^2/K^2} \mid \{X_i\}_{i=1}^N\right) - 1\right] \le K^2(e-1) = \sigma_0^2.$$

To derive the modulus of continuity of the aforementioned empirical process, we need upper bound on the entropy results related to $\mathcal{H}$ supplied by Lemma S6. Namely, under Assumption 2, there exists a constant A such that $H_\infty(\xi, \{u \in \mathcal{H} : \|u\|_{\mathcal{H}} \le 1\}) \le A\xi^{-d/\ell}$ for $\xi > 0$. Due to Lemma S7, there exists a constant R, independent of X_i $(i = 1, \ldots, N)$, such that $\sup_{\{u\in\mathcal{H}:\|u\|_{\mathcal{H}}\le 1\}} \|u\|_\infty \le R$. Now, we apply Lemma 8.4 of van de Geer (2000). For some constant c depending on A, d, ℓ,

R, K and σ_0, we have for all $T \geq c$,

$$\mathrm{pr}\left\{ \sup_{u \in \widetilde{\mathcal{H}}_N} \frac{|N^{-1/2}\sum_{i=1}^N \delta_i u(X_i)|}{\|u\|_{\mathcal{H}}^{d/2\ell}} \geq T \;\middle|\; X_1,\ldots,X_N \right\} \leq c \exp\left(-\frac{T^2}{c^2}\right).$$

Note that the constant c is independent of $\{X_i\}$ which leads to the unconditional probability inequality that, for all $T \geq c$,

$$\mathrm{pr}\left\{ \sup_{u \in \widetilde{\mathcal{H}}_N} \frac{|N^{-1/2}\sum_{i=1}^N \delta_i u(X_i)|}{\|u\|_{\mathcal{H}}^{d/2\ell}} \geq T \right\} \leq c \exp\left(-\frac{T^2}{c^2}\right).$$

This implies the desired result. □

S1·5. *Partially Conditional Central Limit Theorem*

THEOREM S1. *Let $(A_1,\mathcal{B}_1),\ldots,(A_n,\mathcal{B}_n)$ be independent and identically distributed where $A_1,\ldots,A_n$ are random variables and $\mathcal{B}_1,\ldots,\mathcal{B}_n$ are sets of random variables. Let $\{C_1,\ldots,C_n\}$ be another set of random variables. Write $\mathcal{D}_n = \{A_1,\ldots,A_n,\mathcal{B}_1,\ldots,\mathcal{B}_n\}$. Assume these variables satisfy*

$$E(A_j)=0, \quad E(C_j \mid \mathcal{D}_n)=0, \qquad (j=1,\ldots,n),$$
$$\sum_{j=1}^n \mathrm{var}(A_j)=1, \quad \sum_{j=1}^n \mathrm{var}(C_j \mid \mathcal{D}_n)=1,$$

and there exists $\delta > 0$ such that $\sum_{j=1}^n E\left(|C_j|^{2+\delta} \mid \mathcal{D}_n\right) \to 0$ in probability. Moreover, $C_1,\ldots,C_n$ are conditionally independent given $\mathcal{D}_n$. Let g be a (non-random) function mapping from the support of $\mathcal{D}_n$ to $\mathbb{R}^+$ such that there exists a constant $M > 0$ such that $Eg^2(\mathcal{D}_n) \leq M$ and $g^2(\mathcal{D}_n) \leq M + o_p(1)$. For any positive real number τ, consider two random variables:

$$Z_n = \tau \sum_{j=1}^n A_j + g(\mathcal{D}_n)\sum_{j=1}^n C_j \quad \text{and} \quad Z_n^* = \tau F + g(\mathcal{D}_n)\sum_{j=1}^n \{\mathrm{var}(C_j \mid \mathcal{D}_n)\}^{1/2} G_j,$$

where $F, G_1,\ldots,G_n$ are i.i.d. standard normal random variables independent of $C_1,\ldots,C_n$ and $\mathcal{D}_n$. Let ϕ_n and ϕ_n^* be the corresponding characteristic function of Z_n and Z_n^* respectively. Then $|\phi_n(t) - \phi_n^*(n)| \to 0$ for every $t \in \mathbb{R}$. Moreover, $E(Z_n^{*2}) = \tau^2 + E\{g^2(\mathcal{D}_n)\} \leq \tau^2 + M$ and therefore ϕ_n^* is twice differentiable.

*Proof of Theorem S*1. We extend the arguments of Dvoretzky (1972) to our partially conditional setting. Let $F_1,\ldots,F_n,G_1,\ldots,G_n$ be i.i.d. standard normal random variables independent of $C_1,\ldots,C_n$ and $\mathcal{D}_n$. Write $\sigma_{C,j}^2(\mathcal{D}_n) = \mathrm{var}(C_j \mid \mathcal{D}_n)$ for all $j=1,\ldots,n$. Throughout this proof, i represents the complex number such that $i^2=-1$. Let $t \in \mathbb{R}$. First,

$$\exp\left\{ it\left(\tau\sum_{j=1}^n A_j + g(\mathcal{D}_n)\sum_{j=1}^n C_j\right)\right\} - \exp\left[it\left\{\tau\sum_{j=1}^n n^{-1/2}F_j + g(\mathcal{D}_n)\sum_{j=1}^n \sigma_{C,j}(\mathcal{D}_n)G_j\right\}\right]$$

$$= \left(\exp\left[it\left\{\tau\sum_{j=1}^{n}A_j + g(\mathcal{D}_n)\sum_{j=1}^{n}C_j\right\}\right] - \exp\left[it\left\{\tau\sum_{j=1}^{n}A_j + g(\mathcal{D}_n)\sum_{j=1}^{n}\sigma_{C,j}(\mathcal{D}_n)G_j\right\}\right]\right)$$

$$+\left(\exp\left[it\left\{\tau\sum_{j=1}^{n}A_j + g(\mathcal{D}_n)\sum_{j=1}^{n}\sigma_{C,j}(\mathcal{D}_n)G_j\right\}\right] - \exp\left[it\left\{\tau\sum_{j=1}^{n}n^{-1/2}F_j + g(\mathcal{D}_n)\sum_{j=1}^{n}\sigma_{C,j}(\mathcal{D}_n)G_j\right\}\right]\right).$$

Denote the first bracket and the second bracket as Q_1 and Q_2 respectively. Write $\tilde{C}_k = g(\mathcal{D}_n)\sum_{j=1}^{k}C_j$ and $\tilde{G}_k = g(\mathcal{D}_n)\sum_{j=k+1}^{n}\sigma_{C,j}(\mathcal{D}_n)G_j$.

$$\begin{aligned}
Q_1 &= \exp\left(it\tau\sum_{j=1}^{n}A_j\right)\left\{\exp\left(it\tilde{C}_n\right) - \exp\left(it\tilde{G}_0\right)\right\} \\
&= \exp\left(it\tau\sum_{j=1}^{n}A_j\right)\sum_{k=1}^{n}\left[\exp\left\{it(\tilde{C}_k + \tilde{G}_k)\right\} - \exp\left\{it(\tilde{C}_{k-1} + \tilde{G}_{k-1})\right\}\right] \\
&= \exp\left(it\tau\sum_{j=1}^{n}A_j\right)\sum_{k=1}^{n}\exp\left\{it(\tilde{C}_{k-1} + \tilde{G}_k)\right\}\left[\exp\{itg(\mathcal{D}_n)C_k\} - \exp\{itg(\mathcal{D}_n)\sigma_{C,k}(\mathcal{D}_n)G_k\}\right]
\end{aligned}$$

Therefore,

$$\begin{aligned}
&|E(Q_1)| \\
&\leq \sum_{k=1}^{n}\left|E\left(\exp\left(it\tau\sum_{j=1}^{n}A_j\right)E\left[\exp\left\{it(\tilde{C}_{k-1} + \tilde{G}_k)\right\} \mid \mathcal{D}_n\right]\right.\right. \\
&\qquad\qquad \left.\left.\times E\left[\exp\{itg(\mathcal{D}_n)C_k\} - \exp\{itg(\mathcal{D}_n)\sigma_{C,k}(\mathcal{D}_n)G_k\} \mid \mathcal{D}_n\right]\right)\right| \\
&\leq \sum_{k=1}^{n}E\left(\left|\exp\left(it\tau\sum_{j=1}^{n}A_j\right)\right| E\left[\left|\exp\left\{it(\tilde{C}_{k-1} + \tilde{G}_k)\right\}\right| \mid \mathcal{D}_n\right]\right. \\
&\qquad\qquad \left.\times\left|E\left[\exp\{itg(\mathcal{D}_n)C_k\} - \exp\{itg(\mathcal{D}_n)\sigma_{C,k}(\mathcal{D}_n)G_k\} \mid \mathcal{D}_n\right]\right|\right) \\
&\leq \sum_{k=1}^{n}E\left|E\left[\exp\{itg(\mathcal{D}_n)C_k\} - \exp\{itg(\mathcal{D}_n)\sigma_{C,k}(\mathcal{D}_n)G_k\} \mid \mathcal{D}_n\right]\right| \qquad \text{(S8)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
|E(Q_2)| &\leq \left|E\left[\exp\left\{itg(\mathcal{D}_n)\sum_{j=1}^{n}\sigma_{C,k}(\mathcal{D}_n)G_j\right\}\left\{\exp\left(it\tau\sum_{j=1}^{n}A_j\right) - \exp\left(it\tau\sum_{j=1}^{n}n^{-1/2}F_j\right)\right\}\right]\right| \\
&\leq \left|E\left\{\exp\left(it\tau\sum_{j=1}^{n}A_j\right) - \exp\left(it\tau\sum_{j=1}^{n}n^{-1/2}F_j\right)\right\}\right|
\end{aligned}$$

$$\leq \sum_{k=1}^{n} \left| E\left\{\exp\left(it\tau A_k\right) - \exp\left(it\tau n^{-1/2} F_k\right)\right\}\right|, \tag{S9}$$

where the last inequality is due to a similar argument applied to Q_1.

Now, we focus on (S8). As

$$\left|\exp(it) - \sum_{k=0}^{K} \frac{(it)^k}{k!}\right| \leq 2\min\left\{\frac{|t|^{n+1}}{(n+1)!}, \frac{2|t|^n}{n!}\right\}.$$

For any $\varepsilon > 0$,

$$\begin{aligned}
&\sum_{k=1}^{n} \left| E\left[\exp\left\{itg(\mathcal{D}_n)C_k\right\} \mid \mathcal{D}_n\right] - 1 + \frac{1}{2}t^2 g^2(\mathcal{D}_n)\sigma_{C,k}^2(\mathcal{D}_n)\right| \\
&\leq \frac{1}{6}|t|^3 \sum_{k=1}^{n} E\left[|g(\mathcal{D}_n)C_k|^3 I\{|g(\mathcal{D}_n)C_k| \leq \varepsilon\} \mid \mathcal{D}_n\right] + t^2 \sum_{k=1}^{n} E\left[g^2(\mathcal{D}_n)C_k^2 I\{|g(\mathcal{D}_n)C_k| > \varepsilon\} \mid \mathcal{D}_n\right] \\
&\leq \frac{1}{6}\varepsilon^3|t|^3 + t^2 g^2(\mathcal{D}_n) \sum_{k=1}^{n} E\left[C_k^2 I\{|g(\mathcal{D}_n)C_k| > \varepsilon\} \mid \mathcal{D}_n\right].
\end{aligned}$$

Since $|g(\mathcal{D}_n)C_k| > \varepsilon$ implies $|g(\mathcal{D}_n)C_k/\varepsilon|^\delta > 1$, we have

$$\begin{aligned}
0 \leq g^2(\mathcal{D}_n) \sum_{k=1}^{n} E\left[C_k^2 I\{|g(\mathcal{D}_n)C_k| > \varepsilon\} \mid \mathcal{D}_n\right] &\leq \frac{g^{2+\delta}(\mathcal{D}_n)}{\varepsilon^\delta} \sum_{k=1}^{n} E\left[|C_k|^{2+\delta} I\{|g(\mathcal{D}_n)C_k| > \varepsilon\} \mid \mathcal{D}_n\right] \\
&\leq \frac{g^{2+\delta}(\mathcal{D}_n)}{\varepsilon^\delta} \sum_{k=1}^{n} E\left(|C_k|^{2+\delta} \mid \mathcal{D}_n\right),
\end{aligned}$$

where the rightmost expression converges to 0 in probability, since $\sum_{k=1}^{n} E(|C_k|^{2+\delta} \mid \mathcal{D}_n) \to 0$ in probability and $g^2(\mathcal{D}_n) \leq M + o_p(1)$. Moreover,

$$g^2(\mathcal{D}_n) \sum_{k=1}^{n} E\left[C_k^2 I\{|g(\mathcal{D}_n)C_k| > \varepsilon\} \mid \mathcal{D}_n\right] \leq g^2(\mathcal{D}_n),$$

where $Eg^2(\mathcal{D}_n) \leq M$. By dominated convergence theorem, with Skorohod Representation Theorem to extend its result to weakly convergent sequence of random variables, we have $E(g^2(\mathcal{D}_n)\sum_{k=1}^{n} E[C_k^2 I\{|g(\mathcal{D}_n)C_k| > \varepsilon\} \mid \mathcal{D}_n]) \to 0$. As $\varepsilon > 0$ is arbitrary,

$$E\sum_{k=1}^{n} \left| E\left[\exp\left\{itg(\mathcal{D}_n)C_k\right\} \mid \mathcal{D}_n\right] - 1 + \frac{1}{2}t^2 g^2(\mathcal{D}_n)\sigma_{C,k}^2(\mathcal{D}_n)\right| \to 0. \tag{S10}$$

Similarly, we have

$$\begin{aligned}
&\sum_{k=1}^{n}\left|E\left[\exp\{itg(\mathcal{D}_n)\sigma_{C,k}(\mathcal{D}_n)G_k\}\mid\mathcal{D}_n\right]-1+\frac{1}{2}t^2g^2(\mathcal{D}_n)\sigma_{C,k}^2(\mathcal{D}_n)\right|\\
&\le\frac{1}{6}\varepsilon^3|t|^3+t^2g^2(\mathcal{D}_n)\sum_{k=1}^{n}\sigma_{C,k}^2(\mathcal{D}_n)E\left[G_k^2I\{|g(\mathcal{D}_n)\sigma_{C,k}(\mathcal{D}_n)G_k|>\varepsilon\}\mid\mathcal{D}_n\right]\\
&\le\frac{1}{6}\varepsilon^3|t|^3+t^2\frac{g^{2+\delta}(\mathcal{D}_n)}{\varepsilon^\delta}E\left(|G_1|^{2+\delta}\right)\sum_{k=1}^{n}\sigma_{C,k}^{2+\delta}(\mathcal{D}_n)\\
&\le\frac{1}{6}\varepsilon^3|t|^3+t^2\frac{g^{2+\delta}(\mathcal{D}_n)}{\varepsilon^\delta}E\left(|G_1|^{2+\delta}\right)\sum_{k=1}^{n}E\left(|C_k|^{2+\delta}\mid\mathcal{D}_n\right)
\end{aligned}$$

where the last equality is due to Jensen's inequality as $(2+\delta)/2>1$. As G_1 is a standard normal random variable, $E|G_1|^{2+\delta}=\Gamma\{(3+\delta)/2\}\pi^{-1/2}$ where Γ is the Gamma function. Therefore $E|G_1|^{2+\delta}<\infty$. Hence, by a similar argument using dominated convergence theorem, we conclude

$$E\sum_{k=1}^{n}\left|E\left[\exp\{itg(\mathcal{D}_n)\sigma_{C,k}(\mathcal{D}_n)G_k\}\mid\mathcal{D}_n\right]-1+\frac{1}{2}t^2g^2(\mathcal{D}_n)\sigma_{C,k}^2(\mathcal{D}_n)\right|\to 0.$$

Combining with (S10), $|E(Q_1)|\to 0$. Similar but simpler argument can be used to control (S9) and conclude that $|E(Q_2)|\to 0$. As a result,

$$\left|E\exp\left[it\left\{\tau\sum_{j=1}^{n}A_j+g(\mathcal{D}_n)\sum_{j=1}^{n}C_j\right\}\right]-E\exp\left[it\left\{\tau\sum_{j=1}^{n}n^{-1/2}F_j+g(\mathcal{D}_n)\sum_{j=1}^{n}\sigma_{C,j}(\mathcal{D}_n)G_j\right\}\right]\right|\to 0,$$

for every t. Write $F=\sum_{j=1}^{n}n^{-1/2}F_j$ which is a standard normal random variable. Note that $E\{\tau F+g(\mathcal{D}_n)\sum_{j=1}^{n}\sigma_{C,j}(\mathcal{D}_n)G_j\}^2=\tau^2+E\{g^2(\mathcal{D}_n)\}\le\tau^2+M$, and therefore the second moment exists. Hence the characteristic function of $\tau F+g(\mathcal{D}_n)\sum_{j=1}^{n}\sigma_{C,j}(\mathcal{D}_n)G_j$ is at twice differentiable. Hence we obtain the desired result. Note that F is independent of both $G_1,\ldots,G_n$ and $\mathcal{D}_n$. Hence we obtain the desired result. □

REFERENCES

BIRMAN, M. S. & SOLOMYAK, M. Z. (1967). Piecewise-polynomial approximations of functions of the classes w_p^α. *Matematicheskii Sbornik* **115**, 331–355.

DVORETZKY, A. (1972). Asymptotic normality for sums of dependent random variables. In *Proc. 6th Berkeley Symp. Math. Statist. Probab* **2**, eds. L. M. Le Cam, J. Neyman and E. L. Scott, pp. 513–35. Berkeley: University of California Press.

LIN, Y. (2000). Tensor product space ANOVA models. *Ann. Statist.* **28**, 734–755.

VAN DE GEER, S. A. (2000). *Empirical Processes in M-estimation*. Cambridge: Cambridge University Press.

[*Received x x. Revised x x*]