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 $\beta \in \mathbb{R}^r$,

$$\beta^\top[(tv_1 + (1-t)v_2)\{tv_1 + (1-t)v_2\}^\top + B]\beta = [(tv_1 + (1-t)v_2)^\top\beta]^2 + \beta^\top B \beta$$

$$= (tv_1^\top \beta + (1-t)v_2^\top \beta)^2 + \beta^\top B \beta$$

$$\leq t(v_1^\top \beta)^2 + (1-t)(v_2^\top \beta)^2 + \beta^\top B \beta$$

$$= t\beta^\top (v_1 v_1^\top + B)\beta + (1-t)\beta^\top (v_2 v_2^\top + B)\beta$$

Therefore, $\sigma_{\max}[(tv_1 + (1-t)v_2)\{tv_1 + (1-t)v_2\}^\top + B] \leq t\sigma_{\max}(v_1 v_1^\top + B) + (1-t)\sigma_{\max}(v_2 v_2^\top + B)$.

S1.2. Proof of Theorems 1 and 2
We begin with several definitions that will be used throughout the theoretical development. Write $w^* = (w_1^*, \ldots, w_N^*)^\top = [(\pi(X_1)^{-1})\ldots, (\pi(X_N)^{-1})]$, and

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

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

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

where $u^* = \arg\min_{u \in \tilde{\mathcal{H}}_N} \{ S_N(w^*, u) - \lambda_1 \|u\|_{\mathcal{H}}^2 \}$ and its existence is shown in §2.3. Since $S_N(\bar{w}, u) = 0$ for any $u \in \mathcal{H}$ such that $\|u\|_{\mathcal{H}} = 0$, (S1) also implies that, for any $u \in \mathcal{H}$,

$$S_N(\bar{w}, u) - \lambda_1 \|u\|_{\mathcal{H}}^2 + \lambda_2 V_N(\bar{w}) \|u\|_{\mathcal{H}}^2 \leq \left\{ S_N(w^*, u^*) - \lambda_1 \|u^*\|_{\mathcal{H}}^2 + \lambda_2 V_N(w^*) \right\} \|u\|_{\mathcal{H}}^2. \quad (S2)$$

In below, we adopt several choices of $f \in \tilde{\mathcal{H}}_N$ in (S1) and $u \in \mathcal{H}$ in (S2) to obtain various results. For instance, one obvious candidate of $f \in \tilde{\mathcal{H}}_N$ is the constant function $z$ where $z(x) = 1$. 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 \( w^* = (w_1^*, \ldots, w_N^*)^T = [(\pi(X_1))^{-1}, \ldots, (\pi(X_N))^{-1}]^T \). There exists a constant \( c \geq 0 \) such that for all \( T \geq c \),

\[
\text{pr} \left\{ \sup_{u \in \mathcal{H}_N} \frac{N S_N(w^*, u)}{\|u\|_{\mathcal{H}}^{d/\ell}} \geq T^2 \right\} \leq c \exp \left( -\frac{T^2}{c^2} \right).
\]

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 \( \lambda_1 \propto N^{-1} \) and \( \lambda_2 \propto N^{-1} \), we have \( S_N(\bar{w}, z) = O_p(N^{-1}) \) and \( V_N(\bar{w}) = O_p(1) \). Moreover, there exists a constant \( W > 0 \) such that \( E[V_N(\bar{w})] \leq W \).

Proof of Lemma S2. Taking \( f \) as \( z \) (constant function of value 1) in (S1), we obtain a basic inequality:

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

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

Let \( \tilde{E}_{N,1}, \tilde{E}_{N,2} \) and \( \tilde{E}_{N,3} \) 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 \( E_{N,1} = \tilde{E}_{N,1} \), \( E_{N,2} = \tilde{E}_{N,2} \setminus \tilde{E}_{N,1} \) and \( E_{N,3} = \tilde{E}_{N,3} \setminus (\tilde{E}_{N,1} \cup \tilde{E}_{N,2}) \). 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, \( E_{N,1,T} = E_{N,1} \cap \{S_N(w^*, u^*) \leq T^2 N^{-1}\|u^*\|_{\mathcal{H}}^d \} \) and \( \tilde{E}_{N,1,T} = E_{N,1} \cap \{S_N(w^*, u^*) > T^2 N^{-1}\|u^*\|_{\mathcal{H}}^d \} \). Note that \( E_{N,1,T} \cup \tilde{E}_{N,1,T} \cup E_{N,2} \cup E_{N,3} \) forms a partition of the sample space. We analyze (S3) on these events.

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

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

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

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

\[
\text{pr} \left[ S_N(\bar{w}, z) \leq \max \left\{ T^{4d/(2\ell-d)} \lambda_1^{-d/(2\ell-d)} N^{-2\ell/(2\ell-d)}, 3\lambda_1\|z\|_{\mathcal{H}}^2, 3\lambda_2 V_N(w^*) \right\} \right]
\]

\[
= \sum_{i=1}^{3} \text{pr} \left[ S_N(\bar{w}, z) \leq \max \left\{ T^{4d/(2\ell-d)} \lambda_1^{-d/(2\ell-d)} N^{-2\ell/(2\ell-d)}, 3\lambda_1\|z\|_{\mathcal{H}}^2, 3\lambda_2 V_N(w^*) \right\} \cap E_{N,i} \right]
\]

\[
\geq \text{pr}(E_{N,1,T}) + \text{pr} \left[ \left\{ S_N(\bar{w}, z) \leq T^{4d/(2\ell-d)} \lambda_1^{-d/(2\ell-d)} N^{-2\ell/(2\ell-d)} \right\} \cap \tilde{E}_{N,1,T} \right] + \text{pr}(E_{N,2}) + \text{pr}(E_{N,3}) \tag{S4}
\]

\[
= 1 - \text{pr} \left[ S_N(\bar{w}, z) > T^{4d/(2\ell-d)} \lambda_1^{-d/(2\ell-d)} N^{-2\ell/(2\ell-d)} \right] \cap E_{N,1,T} \]
Suppose Assumptions 1-3 hold. If \( \lambda_1 = N^{-1} \) and \( \lambda_2 = O(N^{-1}) \), then \( S_N(\bar{w}, m) = O_p(N^{-1}) \|m\|_{\mathcal{H}}^2 \). Further, if \( \lambda_2 \asymp N^{-1} \), there exists a constant \( S^2 > 0 \) such that \( E[NS_N(\bar{w}, m)] \leq S^2 \).
Proof of Lemma S3. Rearranging the terms in (S2), we obtain the basic inequality:

\[ S_N(\bar{w}, m) + \lambda_1 ||u^*||_H^2 ||m||_N^2 + \lambda_2 V_N(\bar{w}) ||m||_N^2 \leq S_N(w^*, u^*) ||m||_N^2 + \lambda_1 ||m||_H^2 + \lambda_2 V_N(w^*) ||m||_N^2. \] (S5)

By Lemma S1, we have \( S_N(w^*, u^*) = O_p(N^{-1}) ||u^*||_H^d \). 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 \( ||u^*||_H \leq \lambda_1^{-\ell/(2\ell-d)} O_p(N^{-\ell/(2\ell-d)}) \) and therefore

\[ S_N(\bar{w}, m) \leq \lambda_1^{-d/(2\ell-d)} O_p(N^{-2\ell/(2\ell-d)}) ||m||_N^2. \]

If \( ||m||_N = 0 \), we have \( S_N(\bar{w}, m) = 0 \leq \lambda_1^{-d/(2\ell-d)} O_p(N^{-2\ell/(2\ell-d)}) ||m||_N^2. \)

Case (ii): Suppose that \( \lambda_1 ||m||_H^2 \) is the largest in right-hand side of (S5). We obtain \( S_N(\bar{w}, m) \leq 3\lambda_1 ||m||_H^2. \)

Case (iii): Suppose that \( \lambda_2 V_N(w^*) ||m||_N^2 \) is the largest in right-hand side of (S5). We obtain

\[ S_N(\bar{w}, m) \leq 3\lambda_2 \lambda_2 V_N(\bar{w}) ||m||_N^2. \]

Due to Lemma S7 in §S1-4, \( ||m||_N \leq R ||m||_H < \infty \). Overall, we have

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

Since \( S_N(\bar{w}, m) = 0 \) if \( ||m||_N^2 = 0 \), we have \( S_N(\bar{w}, m) = O_p(N^{-1}) ||m||_N^2 \) due to the conditions of \( \lambda_1 \) and \( \lambda_2 \).

Next, suppose \( \lambda_2 \approx N^{-1} \). Based on the exponential inequality in Lemma S1, one could apply a similar argument of Lemma S2, and show that there exists a constant \( \tilde{S}^2 > 0 \) such \( E[N^2 S_N^2(\bar{w}, m)] \leq \tilde{S}^2 \) where

\[ \tilde{S}_N^2(\bar{w}, m) = \begin{cases} S_N(\bar{w}, m) ||m||_N, & \text{if } ||m||_N \neq 0, \\ 0, & \text{if } ||m||_N = 0. \end{cases} \]

Moreover,

\[ E[N S_N(\bar{w}, m)] = E[N \tilde{S}_N(\bar{w}, m) ||m||_N^2] \leq \frac{1}{2} \left[ E[N^2 \tilde{S}_N^2(\bar{w}, m)] + E(||m||_N^4) \right] \leq \frac{1}{2} \left( \frac{m^4 dP}{N} + \frac{N - 1}{N} \left( \int m^2 dP \right)^2 \right) \leq \frac{1}{2} \left( \frac{m^4 dP}{N} + \left( \int m^2 dP \right)^2 \right). \]

Due to Lemma S7 in §S1-4, \( \int m^2 dP < \infty \) and \( \int m^4 dP < \infty \). \( \square \)

Proof of Theorem 2. Recall the decomposition:

\[ \frac{1}{N} \sum_{i=1}^{N} T_i \bar{w}_i Y_i = \frac{1}{N} \sum_{i=1}^{N} (T_i \bar{w}_i - 1) m(X_i) + \frac{1}{N} \sum_{i=1}^{N} T_i \bar{w}_i \epsilon_i + \left[ \frac{1}{N} \sum_{i=1}^{N} m(X_i) - E(Y(1)) \right] + E(Y(1)). \]
Due to Lemma S7 in §S1.4, \( \|m\|_N^2 = \int m^2 dP < \infty \). 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:

\[
\left| \frac{1}{N} \sum_{i=1}^{N} (T_i \tilde{w}_i - 1)m(X_i) \right| = S_N(\tilde{w}, m)^{1/2} = O_p(N^{-1/2})\|m\|_2 + o_p(N^{-1/2}),
\]
due to Theorem 1. Moreover, \( E[N S_N(\tilde{w}, m)] < \infty \) due to Theorem 1. As for the second term, we write \( \tilde{\delta}_i = T_i \tilde{w}_i \). Under Assumption 4, we have \( E(\varepsilon_i \mid \tilde{\delta}_1, \ldots, \tilde{\delta}_N) = 0 \). Therefore,

\[
\operatorname{var} \left( \frac{1}{N} \sum_{i=1}^{N} T_i \tilde{w}_i \varepsilon_i \right) = E \left( \operatorname{var} \left( \frac{1}{N} \sum_{i=1}^{N} \delta_i \varepsilon_i \right) \mid \tilde{\delta}_1, \ldots, \tilde{\delta}_N \right) \leq \frac{\sigma^2}{N} E[V_N(\tilde{w})] \leq \frac{\sigma^2 W}{N},
\]
due to Theorem 1. Therefore, \( N^{-1} \sum_{i=1}^{N} T_i \tilde{w}_i \varepsilon_i = O_p(N^{-1/2}) \). The above derivation also implies that

\[
E N^{-1/2} \sum_{i=1}^{N} (T_i \tilde{w}_i - 1) \varepsilon_i \to \infty.
\]
Finally, by central limit theorem, we have

\[
N^{-1} \sum_{i=1}^{N} m(X_i) - E[Y(1)] = O_p(N^{-1/2}),
\]
due to Assumption 4. Also,

\[
E N^{-1/2} \sum_{i=1}^{N} m(X_i) - E[Y(1)]^2 < \infty.
\]
Therefore, \( \sum_{i=1}^{N} T_i \tilde{w}_i Y_i / N - E[Y(1)] = O_p(N^{-1/2}) \) and \( N^{1/2}[\sum_{i=1}^{N} T_i \tilde{w}_i Y_i / N - E[Y(1)]] \) has bounded variance.

**S1.3. Proof of Theorem 3**

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

**Proof of Lemma S4.** Taking \( f \) as \( z \) (constant function of value 1) in (S1), we obtain the basic inequality:

\[
S_N(\tilde{w}, z) + \lambda_1 \|u^*\|_H^2 + \lambda_2 V_N(\tilde{w}) \leq S_N(w^*, u^*) + \lambda_1 \|z\|_H^2 + \lambda_2 V_N(w^*),
\]
for all large \( N \) such that \( 1 \leq BN^{1/3} \). By Lemma S1, \( S_N(w^*, u^*) = O_p(N^{-1})\|u^*\|_H^{d/\ell} \). Moreover, it is easy to show that \( V_N(w^*) = V + O_p(N^{-1/2}) \). Due to the condition of \( \lambda_1 \) and \( \lambda_2 \), we have \( \lambda_1^{-d/(2\ell-d)} N^{-2\ell/(2\ell-d)} = o(\lambda_2) \) which implies \( (\lambda_1 N)^{-d/(2\ell-d)} = o(\lambda_2^{-1}) \). As \( \lambda_1 N = O(1) \), therefore \( \lambda_1^{-1} \lambda_2 \to 0 \).

\[
\lambda_1 \|z\|_H^2 + \lambda_2 V_N(\tilde{w}) = \lambda_2 [o(1) + V + O_p(N^{-1/2})] = \lambda_2 V [1 + o_p(1)].
\]
Now, we come back to (S6). Let \( \mathcal{A} \) be the event that \( S_N(w^*, u^*) \leq \lambda_1 \|z\|_H^2 + \lambda_2 V_N(\tilde{w}) \). On the event \( \mathcal{A}^c \), from (S6), we obtain \( \|u^*\|_H \leq \lambda_1^{-\ell/(2\ell-d)} \lambda_2 \), which implies
Suppose Assumptions 1 and 2 hold. Let $h = m - \widetilde{m} \in \mathcal{H}$ such that $||h||_{\mathcal{H}} = o_p(1)$ and $||h||_{\mathcal{H}} = O_p(N^{1/2})$. Further, assume $\lambda_1 = o(N^{-1})$, $\lambda_1^{-1}||h||_{\mathcal{H}}^{2(2\ell-d)/d} = o_p(N)$ and $\lambda_2 ||h||_{\mathcal{H}}^2 = o_p(N)$. Then $S_N(\widetilde{w}, h) = o_p(N^{-1})$. Moreover, there exists a constant $S' > 0$ such that $E[N S_N(\widetilde{w}, h)] \leq S'$.

**Proof of Lemma S5.** Rearranging the terms in (S2), we obtain the basic inequality:

$$S_N(\widetilde{w}, u^*) + \lambda_1 ||u^*||_{\mathcal{H}}^2 ||h||_{\mathcal{H}}^2 + \lambda_2 V_N(\widetilde{w}) ||h||_{\mathcal{H}}^2 \leq S_N(w^*, u^*) ||h||_{\mathcal{H}}^2 + \lambda_1 ||h||_{\mathcal{H}}^2 + \lambda_2 V_N(w^*) ||h||_{\mathcal{H}}^2$$

(S7)

for all large $N$ such that $C \leq BN^{1/3}$. The rest of the proof is similar to the proof of Lemma S3 but with different conditions of $\lambda_1$ and $\lambda_2$. 

**Proof of Theorem 3.** Recall the decomposition:

$$\frac{1}{N} \sum_{i=1}^{N} T_i \bar{w}_i (Y_i - \bar{m}(X_i)) + \frac{1}{N} \sum_{i=1}^{N} \bar{m}(X_i)$$

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

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 $\text{var}\{N^{-1/2} \sum_{i=1}^{N} (T_i \bar{w}_i - 1) \bar{h}(X_i)\} \leq E[N S_N(\widetilde{w}, h)] \rightarrow 0$ using Lemma S5. Write

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

It is obvious that $\text{var}(Z_N) = \text{var}\{m(X_1)\} + \sigma^2 E V_N(\widetilde{w})$. By Lemma S4, we have $\lim sup_N E V_N(\widetilde{w}) \leq E[V + o_p(1)] = V$ 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 $\tau^2 = \text{var}\{m(X_1)\}$, $g^2(D_N) = \sigma^2 V_N(\widetilde{w})$, and

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

Write $D_N = \{A_1, \ldots, A_N, B_1, \ldots, B_N\}$. By the definition of $\widetilde{w}_i (i : T_i = 1), 1 \leq \widetilde{w}_i \leq BN^{1/3}$ for all $i$. Therefore, $\sum_{i=1}^{N} T_i \widetilde{w}_i^2 = O_p(N)$ and $\max_i |\widetilde{w}_i| = o_p(N^{1/2})$. Moreover, $\max_i E[|\varepsilon_i|^3] < \infty$ by
assumption. Hence

$$0 \leq E\left( \sum_{i=1}^{N} |C_i|^3 \mid D_N \right) = \frac{(\max_{j} E[|\varepsilon_j|^3] \sum_{j=1}^{N} T_j \overline{w}_j^3)}{\sigma^3 \left( \sum_{i=1}^{N} T_i \overline{w}_i^3 \right)^{3/2}} \leq \frac{(\max_{j} E[|\varepsilon_j|^3] \max_{i} |\overline{w}_i|)}{\sigma^3 \left( \sum_{i=1}^{N} T_i \overline{w}_i^2 \right)^{1/2}} = o_p(1).$$

By Lemma S4, we have $E[g^2(D_N)] \leq M$ and $g^2(D_N) \leq M + o_p(1)$, by taking $M = \sigma^2 \max\{W^*, V\}$. Write

$$Z_n = \tau \sum_{j=1}^{n} A_j + g(D_N) \sum_{j=1}^{n} C_j = N^{1/2} \left( \frac{1}{N} \sum_{j=1}^{N} T_j \overline{w}_j \varepsilon_j \right) + \frac{1}{N} \sum_{j=1}^{N} m(X_j) - E\{Y(1)\}$$

where $F, G_1, \ldots, G_N$ are i.i.d. standard normal random variables independent of $C_1, \ldots, C_N$ and $D_N$. Let $\phi_N$ and $\phi_N^*$ 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 $\phi_N^*$ is twice differentiable.

**S1.4. Proof of Lemma S1**

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

**Proof of Lemma S6.** This is shown by Birman & Solomyak (1967) and the fact that $\mathcal{H}$ is a subspace of the Sobolev space $\mathcal{W}^{0,2}([0,1]^d)$.

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

**Proof of Lemma S7.** This is due to Lemma 2.1 of Lin (2000) and norm equivalence.

**Proof of Lemma S1.** Let $\delta_i = T_i \overline{w}_i^* - 1$. Note that the conditional expectation $E(\delta_i \mid X_i) = 0$. We will focus on the empirical process $\{N^{-1/2} \sum_{i=1}^{N} \delta_i u(X_i) : u \in \tilde{\mathcal{H}}\}$. Due to Assumption 1, $0 < \overline{w}_i^* \leq C$ for all $i = 1, \ldots, N$. Therefore, $\delta_i (i = 1, \ldots, N)$ are uniformly sub-Gaussian: there exist constants $K$ and $\sigma_0^2$ independent of $X_i (i = 1, \ldots, N)$, such that

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

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

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

To derive the modulus of continuity of the aforementioned empirical process, we need upper bound on the entropy results related to $\hat{\mathcal{H}}$ supplied by Lemma S6. Namely, under Assumption 2, there exists a constant $A$ such that $H_\infty(\xi; \{u \in \mathcal{H} : \|u\|_{\mathcal{H}} \leq 1\}) \leq A \xi^{-d/\ell}$ for $\xi > 0$. Due to Lemma S7, there exists a constant $R$, independent of $X_i (i = 1, \ldots, N)$, such that $\sup_{u \in \mathcal{H} : \|u\|_{\infty} \leq 1} \|u\|_{\infty} \leq R$. Now, we apply Lemma 8.4 of van de Geer (2000). For some constant $c$ depending on $A, d, \ell,$
We extend the arguments of Dvoretzky (1972) to our partially conditional setting. Let
\[ \phi_i \] be another set of random variables. Write \( D_n = \{A_1, \ldots, A_n\} \). Assume these variables satisfy
\[ E(A_j) = 0, \quad E(C_j \mid D_n) = 0, \quad (j = 1, \ldots, n), \]
\[ \sum_{j=1}^n \text{var}(A_j) = 1, \quad \sum_{j=1}^n \text{var}(C_j \mid D_n) = 1, \]
and there exists \( \delta > 0 \) such that \( \sum_{j=1}^n E\left(|C_j|^{2+\delta} \mid D_n\right) \rightarrow 0 \) in probability. Moreover, \( C_1, \ldots, C_n \) are conditionally independent given \( D_n \). Let \( g \) be a (non-random) function mapping from the support of \( D_n \) to \( \mathbb{R}^+ \) such that there exists a constant \( M > 0 \) such that \( E g^2(D_n) \leq M \) and \( g^2(D_n) \leq M + o_p(1) \). For any positive real number \( \tau \), consider two random variables:
\[ Z_n = \tau \sum_{j=1}^n A_j + g(D_n) \sum_{j=1}^n C_j \quad \text{and} \quad Z_n^* = \tau F + g(D_n) \sum_{j=1}^n \text{var}(C_j \mid D_n)^{1/2} G_j, \]
where \( F, G_1, \ldots, G_n \) are i.i.d. standard normal random variables independent of \( C_1, \ldots, C_n \) and \( D_n \). Let \( \phi_n \) and \( \phi_n^* \) be the corresponding characteristic function of \( Z_n \) and \( Z_n^* \) respectively. Then \( |\phi_n(t) - \phi_n^*(n)| \rightarrow 0 \) for every \( t \in \mathbb{R} \). Moreover, \( E(Z_n^*^2) = \tau^2 + E\{g^2(D_n)\} \leq \tau^2 + M \) and therefore \( \phi_n^* \) is twice differentiable.

**Proof of Theorem S1.** 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 \( D_n \). Write \( \sigma^2_{C_j}(D_n) = \text{var}(C_j \mid D_n) \) 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,
\[ \exp\left\{ it \sum_{j=1}^n A_j + g(D_n) \sum_{j=1}^n C_j \right\} - \exp\left\{ it \left( \tau \sum_{j=1}^n F_j + g(D_n) \sum_{j=1}^n \sigma_{C_j}(D_n) G_j \right) \right\} \]
Similarly,

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

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

Therefore,

\[
\left| E(Q_2) \right| \leq \sum_{k=1}^{n} \left| E \left( \exp\left( \tau \sum_{j=1}^{n} A_j \right) \right) \exp\left( \tau \sum_{j=1}^{n} \sigma_{C,j}(D_n)G_j \right) \right| \left| \left( \exp\left( \tau \sum_{j=1}^{n} C_j \right) - \exp\left( \tau \sum_{j=1}^{n} \sigma_{C,j}(D_n)G_j \right) \right) \right|
\]

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

\[
Q_1 = \exp\left( \tau \sum_{j=1}^{n} A_j \right) \left\{ \exp\left( \tau \tilde{C}_n \right) - \exp\left( \tau \tilde{G}_0 \right) \right\}
\]

\[
= \exp\left( \tau \sum_{j=1}^{n} A_j \right) \sum_{k=1}^{n} \left\{ \exp\left( \tau \tilde{C}_k + \tilde{G}_k \right) - \exp\left( \tau \tilde{C}_{k-1} + \tilde{G}_{k-1} \right) \right\}
\]

\[
= \exp\left( \tau \sum_{j=1}^{n} A_j \right) \sum_{k=1}^{n} \exp\left( \tau \tilde{C}_{k-1} + \tilde{G}_k \right) \left\{ \exp\left( \tau \tilde{G}_k \right) - \exp\left( \tau \tilde{G}_{k-1} \right) \right\}
\]

\[
\left| E(Q_1) \right| \leq \sum_{k=1}^{n} \left| E \left( \exp\left( \tau \sum_{j=1}^{n} A_j \right) \right) \exp\left( \tau \sum_{j=1}^{n} \sigma_{C,j}(D_n)G_j \right) \right| \left| \left( \exp\left( \tau \sum_{j=1}^{n} C_j \right) - \exp\left( \tau \sum_{j=1}^{n} \sigma_{C,j}(D_n)G_j \right) \right) \right|
\]

Similarly,

\[
\left| E(Q_2) \right| \leq E \left\{ \exp\left( \tau \sum_{j=1}^{n} C_j \right) - \exp\left( \tau \sum_{j=1}^{n} \sigma_{C,j}(D_n)G_j \right) \right\}
\]

\[
\leq E \left\{ \exp\left( \tau \sum_{j=1}^{n} A_j \right) - \exp\left( \tau \sum_{j=1}^{n} \sigma_{C,j}(D_n)G_j \right) \right\}
\]
\[ \sum_{k=1}^{n} \left| E \{ \exp(i\tau A_k) - \exp(i\tau n^{-1/2} F_k) \} \right|, \]  
(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 \( \varepsilon > 0 \),

\[ \sum_{k=1}^{n} \left| E \left[ \exp\{itg(D_n)C_k\} \right] \right| \leq \frac{1}{6}|\varepsilon|^3 \sum_{k=1}^{n} E \left[ |g(D_n)C_k|^3 \right] \leq \frac{1}{6} |\varepsilon|^3 + t^2 g^2(D_n) \sum_{k=1}^{n} E \left[ |C_k|^2 \right]. \]

Since \( |g(D_n)C_k| > \varepsilon \) implies \( |g(D_n)C_k|/\varepsilon^\delta > 1 \), we have

\[ 0 \leq g^2(D_n) \sum_{k=1}^{n} E \left[ |C_k|^2 I[|g(D_n)C_k| > \varepsilon] \right] \leq \frac{g^2(D_n)}{\varepsilon^\delta} \sum_{k=1}^{n} E \left[ |C_k|^2 I[|g(D_n)C_k| > \varepsilon] \right] \leq \frac{g^2(D_n)}{\varepsilon^\delta} \sum_{k=1}^{n} E \left[ |C_k|^2 \right]. \]

where the rightmost expression converges to 0 in probability, since \( \sum_{k=1}^{n} E[|C_k|^{2+\delta}] \rightarrow 0 \) in probability and \( g^2(D_n) \leq M + o_p(1) \). Moreover,

\[ g^2(D_n) \sum_{k=1}^{n} E \left[ |C_k|^2 \right] \leq g^2(D_n), \]

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

\[ E \sum_{k=1}^{n} \left| E \{ itg(D_n)C_k \} \right| \rightarrow 0. \]  
(S10)
Similarly, we have

\[
\sum_{k=1}^{n} \left| E \left[ \exp \left\{ itg(D_n)\sigma_{C,k}(D_n)G_k \right\} \mid D_n \right] - 1 + \frac{1}{2} t^2 g^2(D_n)\sigma_{C,k}^2(D_n) \right| \leq \frac{1}{6} \varepsilon^3 |t|^3 + t^2 g^2(D_n) \sum_{k=1}^{n} \sigma_{C,k}^2(D_n) E \left[ |g(D_n)\sigma_{C,k}(D_n)G_k| > \varepsilon \right] |D_n| \leq \frac{1}{6} \varepsilon^3 |t|^3 + t^2 g^2(D_n) \sum_{k=1}^{n} \sigma_{C,k}^2(D_n) E \left[ |G_k|^2 + |D_n| > \varepsilon \right] |D_n| \leq \frac{1}{6} \varepsilon^3 |t|^3 + t^2 g^2(D_n) \sum_{k=1}^{n} E \left( |G_k|^2 + |D_n| \right)
\]

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 \(\Gamma\) is the Gamma function. Therefore \(E|G_1|^{2+\delta} < \infty\). Hence, by a similar argument using dominated convergence theorem, we conclude

\[
E \sum_{k=1}^{n} \left[ E \left[ \exp \left\{ itg(D_n)\sigma_{C,k}(D_n)G_k \right\} \mid D_n \right] - 1 + \frac{1}{2} t^2 g^2(D_n)\sigma_{C,k}^2(D_n) \right] \rightarrow 0.
\]

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

\[
\left| E \exp \left\{ it \left( \sum_{j=1}^{n} F_j \right) \sum_{j=1}^{n} C_j \right\} - E \exp \left\{ it \left( \sum_{j=1}^{n} n^{-1/2} F_j + g(D_n) \sum_{j=1}^{n} \sigma_{C,j}(D_n)G_j \right) \right\} \right| \rightarrow 0,
\]

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

\[\square\]

REFERENCES


[Received x x. Revised x x]