8. Web Supplement

A function $h(\cdot)$ with domain $[0,1]^d$ is said to belong to a Hölder ball $H(\beta, C)$, with Hölder exponent $\beta > 0$ and radius $C > 0$, if and only if $h(\cdot)$ is uniformly bounded by C, all partial derivatives of $h(\cdot)$ up to order $\lfloor \beta \rfloor$ exist and are bounded, and all partial derivatives $\nabla^{\lfloor \beta \rfloor}$ of order $\lfloor \beta \rfloor$ satisfy

$$\sup_{x, x+\delta x \in [0,1]^d} \left| \nabla^{\lfloor \beta \rfloor} h(x+\delta x) - \nabla^{\lfloor \beta \rfloor} h(x) \right| \leq C ||\delta x||^{\beta - \lfloor \beta \rfloor}.$$

We note that the uniform $L_p, 2 < p < \infty$ and L_∞ rates of convergence for estimation of a marginal density or conditional expectation $h(\cdot) \in H(\beta, C)$ are $O\left(n^{-\frac{\beta}{2\beta+d}}\right)$ and $O\left(\left(\frac{n}{\log n}\right)^{-\frac{\beta}{2\beta+d}}\right)$ respectively. We refer to an estimator attaining these rates as rate optimal. Throughout the paper, we derive the properties of our estimators conditional on the data in the training sample.

We define the following assumptions which have been used throughout the paper.

$(A.1)$: Each $h \in \{b, p, f\}$ belongs to a Hölder class of smooth functions $H(\beta_h, C_h)$ with known Hölder exponent β_h and radius C_h.

$(A.2.1)$: $|b(X)| \leq c_1$, $p(X) \leq c_2$ wp 1 for some constants $0 < c_1, c_2 < \infty$,

$(A.2.2)$: $var(Y|X) = \sigma_Y^2(X) \leq c_3$ wp 1 for some constant $0 < c_3 < \infty$.

$(A.2.3)$: there exist $\delta_f, c_4 > 0$ such that $\delta_f < f(X) < c_4$ wp 1.

$(A.3.1)$: As can be easily achieved with standard nonparametric estimators, we suppose $\widehat{p}(\cdot)$ and $\widehat{b}(\cdot)$ are rate optimal nonparametric estimates of $p(\cdot)$ and $b(\cdot)$ with L_2 convergence rate of respective order $(N-n)^{-\frac{\beta_p}{2\beta_p+d}}$ and $(N-n)^{-\frac{\beta_b}{2\beta_b+d}}$ in probability.

$(A.3.2)$: $\widehat{f}(\cdot)$ converge to $f(\cdot)$ wrt the L_p and L_∞ norm at the optimal rates

$(N-n)^{-\frac{\beta_f}{2\beta_f+d}}\ (\frac{N-n}{log(N-n)})^{-\frac{\beta_f}{2\beta_f+d}}$ in probability.

$(A.3.3)$: $(\widehat{b}(\cdot),\widehat{p}(\cdot))$ are uniformly bounded in sup norm. Further, $\widehat{f}(\cdot)$ and $1/\widehat{f}(\cdot)$ are bounded in sup norm.

Definition 9. *C_n is a honest (i.e. conservative, uniform) asymptotic $(1-\alpha)$ confidence set for $\psi(\theta)$ if*

$$\liminf_n \inf_\theta \left\{ \Pr_\theta \left[\psi\left(\theta\right)\in\mathcal{C}_n\right]-\left(1-\alpha\right)\right\}\geq 0$$

Proof of Theorem 1. Recall that

$$\widehat{\psi}_{2,k}=\psi_1\left(\widehat{\theta}\right)-\frac{1}{n\left(n-1\right)}\sum_{i\neq j}\widehat{\epsilon}_i K_{\widehat{f}_X,k}\left(X_i,X_j\right)\widehat{\Delta}_j.$$

Then

$$\begin{aligned}
&BI\left(\widehat{\psi}_{2,k},\theta\right)\\
&=E_\theta\left[\widehat{\psi}_{2,k}\right]-\psi\left(\theta\right)\\
&=BI\left(\psi_1\left(\widehat{\theta}\right),\theta\right)-E_\theta\left[\frac{1}{n\left(n-1\right)}\sum_{i\neq j}\widehat{\epsilon}_i K_{\widehat{f}_X,k}\left(X_i,X_j\right)\widehat{\Delta}_j\right]\\
&=E_\theta\left[\left(b\left(X_i\right)-\widehat{b}\left(X_i\right)\right)\left(p\left(X_i\right)-\widehat{p}\left(X_i\right)\right)\right]\\
&-E_\theta\left[\left(b\left(X_i\right)-\widehat{b}\left(X_i\right)\right)K_{\widehat{f}_X,k}\left(X_i,X_j\right)\left(p\left(X_j\right)-\widehat{p}\left(X_j\right)\right)\right]\\
&=E_\theta\left[\left(b\left(X_i\right)-\widehat{b}\left(X_i\right)\right)\left(p\left(X_i\right)-\widehat{p}\left(X_i\right)\right)\right]\\
&-E_\theta\left[\left(b\left(X_i\right)-\widehat{b}\left(X_i\right)\right)K_{f_X,k}\left(X_i,X_j\right)\left(p\left(X_j\right)-\widehat{p}\left(X_j\right)\right)\right]\\
&+E_\theta\left[\left(b\left(X_i\right)-\widehat{b}\left(X_i\right)\right)\left\{K_{f_X,k}\left(X_i,X_j\right)-K_{\widehat{f}_X,k}\left(X_i,X_j\right)\right\}\left(p\left(X_j\right)-\widehat{p}\left(X_j\right)\right)\right]\\
&=E_\theta\left[\left(b\left(X_i\right)-\widehat{b}\left(X_i\right)\right)\left\{K_{f_X,\infty}\left(X_i,X_j\right)-K_{f_X,k}\left(X_i,X_j\right)\right\}\left(p\left(X_j\right)-\widehat{p}\left(X_j\right)\right)\right]\\
&+E_\theta\left[\left(b\left(X_i\right)-\widehat{b}\left(X_i\right)\right)\left\{K_{f_X,k}\left(X_i,X_j\right)-K_{\widehat{f}_X,k}\left(X_i,X_j\right)\right\}\left(p\left(X_j\right)-\widehat{p}\left(X_j\right)\right)\right].
\end{aligned}$$

The first term is non-zero because of the truncated at k approximation $K_{f_X,k}(X_i,X_j)$ to $K_{f_X,\infty}(X_i,X_j)$. The second term is non-zero because we estimated $K_{f_X,k}(X_i,X_j)$ by $K_{\widehat{f}_X,k}(X_i,X_j)$. We use $TB_k(\theta)$ and $EB_2(\theta)$ to indicate these two terms respectively. By the definition of $K_{f_X,k}(X_i,X_j)$,

$$
\begin{aligned}
&E_\theta\left[\left(b(X_i)-\widehat{b}(X_i)\right)K_{f_X,k}(X_i,X_j)\left(p(X_j)-\widehat{p}(X_j)\right)\right]\\
&=E_\theta\left[\left(b(X_i)-\widehat{b}(X_i)\right)\Pi_\theta\left[\left(b(X)-\widehat{b}(X)\right)|\overline{\phi}_k(X_i)\right]\right]\\
&=E_\theta\left\{\Pi_\theta\left[\left(b(X)-\widehat{b}(X)\right)|\overline{\phi}_k(X)\right]\Pi_\theta\left[\left(b(X)-\widehat{b}(X)\right)|\overline{\phi}_k(X)\right]\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
&TB_k(\theta)\\
&=E_\theta\left[\left(b(X)-\widehat{b}(X)\right)\left(p(X)-\widehat{p}(X)\right)\right]\\
&-E_\theta\left\{\Pi_\theta\left[\left(b(X)-\widehat{b}(X)\right)|\overline{\phi}_k(X)\right]\Pi_\theta\left[\left(b(X)-\widehat{b}(X)\right)|\overline{\phi}_k(X)\right]\right\}\\
&=E_\theta\left[\begin{array}{c}\left(\Pi_\theta^\perp\left[\left(b(X)-\widehat{b}(X)\right)|\overline{\phi}_k(X)\right]\right)\\ \times\left(\Pi_\theta^\perp\left[\left(p(X)-\widehat{p}(X)\right)|\overline{\phi}_k(X)\right]\right)\end{array}\right],
\end{aligned}
$$

where $\Pi_\theta\left[h(X)|\overline{\phi}_k(X)\right]$ and $\Pi_\theta^\perp\left[h(X)|\overline{\phi}_k(X)\right]$ are, respectively, the $L_2(F_X)$ projection of $h(X)$ on the k dimensional linear subspace $lin\left\{\overline{\phi}_k(X)\right\}$ spanned by the components of the vector $\overline{\phi}_k(X)$ and the projection on the ortho-complement of this subspace.

The proof of eq. (6) is straightforward and is omitted. □

Proof of Theorem 2. We first prove eq. (7). Under our assumptions, the following holds uniformly for $\theta \in \Theta$.

$$
\begin{aligned}
& [TB_k(\theta)]^2 \\
& = \left\{ E_\theta \left(\Pi_\theta^\perp \left[(p(X) - \widehat{p}(X)) \,|\bar{\phi}_k(X) \right] \Pi_\theta^\perp \left[\left(b(X) - \widehat{b}(X) \right) |\bar{\phi}_k(X) \right] \right) \right\}^2 \\
& \leq E_\theta \left[\left(\Pi_\theta^\perp \left[(p(X) - \widehat{p}(X)) \,|\bar{\phi}_k(X) \right] \right)^2 \right] E_\theta \left[\left(\Pi_\theta^\perp \left[\left(b(X) - \widehat{b}(X) \right) |\bar{\phi}_k(X) \right] \right)^2 \right]
\end{aligned}
$$

by Cauchy Shwartz. Further, for any $h \in \{b, p\}$, as a result of the optimal approximation property of $\{z_l(X), l = 1, 2, ...\}$,

$$
\begin{aligned}
& E_\theta \left[\left\{ \Pi_\theta^\perp \left[\left(h(X) - \widehat{h}(X) \right) |\bar{\phi}_k(X) \right] \right\}^2 \right] \\
& = E_\theta \left[\left\{ \Pi_\theta^\perp \left[\left(h(X) - \widehat{h}(X) \right) |\bar{z}_k(X) \right] \right\}^2 \right] \\
& = inf_{\varsigma_l} \int_{R^d} \left(h(X) - \widehat{h}(X) - \sum_{l=1}^{k} \varsigma_l z_l(X) \right)^2 f(X)\, dX \\
& \leq ||f(\cdot)||_\infty \, inf_{\varsigma_l} \int_{R^d} \left(h(X) - \widehat{h}(X) - \sum_{l=1}^{k} \varsigma_l z_l(X) \right)^2 dX \\
& = O_p\left(k^{-2\beta_h/d} \right).
\end{aligned}
$$

In summary, $TB_k = \sup_{\theta \in \Theta} TB_k(\theta) = O_p\left(k^{-(\beta_b + \beta_p)/d} \right)$.

Next, we prove eq. (8). Actually to simplify the proof we only prove eq. (8) up to a log term. A true proof would repeatedly use Hölder's inequality to allow us to replace the L_∞ norm below by a L_p norm with p finite.

$$
\begin{aligned}
&EB_2(\theta) \\
&= \left\{ \begin{array}{c} E_\theta\left[\left(b(X)-\widehat{b}(X)\right)\bar{\phi}_k(X)^T\right]\times \\ \left\{\left(E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right)^{-1}-I_{k\times k}\right\} \\ \times E_\theta\left[\bar{\phi}_k(X)\left(p(X)-\widehat{p}(X)\right)\right] \end{array} \right\} \\
&= -\left\{ \begin{array}{c} E_\theta\left[\left(b(X)-\widehat{b}(X)\right)\bar{\phi}_k(X)^T\right]\left(E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right)^{-1} \\ \times\left\{E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]-E_{\widehat{\theta}}\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right\} \\ \times E_\theta\left[\bar{\phi}_k(X)\left(p(X)-\widehat{p}(X)\right)\right] \end{array} \right\} \\
&= -\left\{ \begin{array}{c} E_\theta\left[\left(b(X)-\widehat{b}(X)\right)\bar{\phi}_k(X)^T\right]\left(E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right)^{-1} \\ \times\left\{E_{\widehat{\theta}}\left[\left(\frac{f(X)}{\widehat{f}(X)}-1\right)\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right\} \\ \times E_\theta\left[\bar{\phi}_k(X)\left(p(X)-\widehat{p}(X)\right)\right] \end{array} \right\} \\
&= \\
&= -\left\{ \begin{array}{c} E_\theta\left[\left(b(X)-\widehat{b}(X)\right)\bar{\phi}_k(X)\right]\left\{E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right\}^{-1} \\ \times\left\{E_{\widehat{\theta}}\left[\left(\frac{f(X)}{\widehat{f}(X)}-1\right)\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right\} \\ \times E_{\widehat{\theta}}\left[\frac{f(X)}{\widehat{f}(X)}\bar{\phi}_k(X)\left(p(X)-\widehat{p}(X)\right)\right] \end{array} \right\} \\
&= -E_{\widehat{\theta}}\left\{ \begin{array}{c} \left\{\left(\frac{f(X)}{\widehat{f}(X)}-1\right)\Pi_\theta\left[\left(b(X)-\widehat{b}(X)\right)|\bar{\phi}_k(X)\right]\right\} \\ \times\Pi_{\widehat{\theta}}\left[\left(\frac{f(X)}{\widehat{f}(X)}\left(p(X)-\widehat{p}(X)\right)\right)|\bar{\phi}_k(X)\right] \end{array} \right\},
\end{aligned}
$$

where $\Pi_\theta\left[\cdot|\bar{\phi}_k(X)\right]$ and $\Pi_{\widehat{\theta}}\left[\cdot|\bar{\phi}_k(X)\right]$ denote the projection under $F(O;\theta)$ and $F\left(O;\widehat{\theta}\right)$ respectively onto the subspace $lin\left\{\bar{\phi}_k(X)\right\} = lin\left\{\overline{z}_k(X)\right\}.$ By Cauchy Shwartz and projection operators having operator norm equal to 1, we obtain the following.

$$
\begin{aligned}
&\left|EB_2\right|^2 \\
&\leq \left(E_{\widehat{\theta}} \left[\begin{array}{c} \left|\frac{f(X)}{\widehat{f}(X)} - 1\right| \left|\Pi_\theta \left[\left(b\left(X\right) - \widehat{b}\left(X\right)\right) |\bar{\phi}_k\right]\right| \\ \times \left|\Pi_{\widehat{\theta}} \left[\frac{f(X)}{\widehat{f}(X)} \left(p\left(X\right) - \widehat{p}\left(X\right)\right) |\bar{\phi}_k\right]\right| \end{array} \right] \right)^2 \\
&\leq \left\| \frac{f\left(\cdot\right)}{\widehat{f}\left(\cdot\right)} - 1 \right\|_\infty^2 E_{\widehat{\theta}} \left[\left(\Pi_\theta \left[\left(b\left(X\right) - \widehat{b}\left(X\right)\right) |\bar{\phi}_k\right]\right)^2 \right] \\
&\times E_{\widehat{\theta}} \left[\left(\Pi_{\widehat{\theta}} \left[\frac{f\left(X\right)}{\widehat{f}\left(X\right)} \left(p\left(X\right) - \widehat{p}\left(X\right)\right) |\bar{\phi}_k\right]\right)^2 \right] \\
&\leq \left\| \frac{f\left(\cdot\right)}{\widehat{f}\left(\cdot\right)} - 1 \right\|_\infty^2 E_{\widehat{\theta}} \left[\left(\frac{f\left(X\right)}{\widehat{f}\left(X\right)} \left(p\left(X\right) - \widehat{p}\left(X\right)\right)\right)^2 \right] \\
&\times E_\theta \left[\left| \frac{\widehat{f}\left(X\right)}{f\left(X\right)} \right| \left(\Pi_\theta \left[\left(b\left(X\right) - \widehat{b}\left(X\right)\right) |\bar{\phi}_k\right]\right)^2 \right].
\end{aligned}
$$

Further,

$$
\begin{aligned}
&E_{\widehat{\theta}} \left[\left(\frac{f\left(X\right)}{\widehat{f}\left(X\right)} \left(p\left(X\right) - \widehat{p}\left(X\right)\right)\right)^2 \right] \\
&= E_\theta \left[\frac{f\left(X\right)}{\widehat{f}\left(X\right)} \left(p\left(X\right) - \widehat{p}\left(X\right)\right)^2 \right] \\
&\leq \left\| \frac{f}{\widehat{f}} \right\|_\infty E_\theta \left[\left(p\left(X\right) - \widehat{p}\left(X\right)\right)^2 \right]
\end{aligned}
$$

and

$$
\begin{aligned}
&E_\theta \left[\left| \frac{\widehat{f}\left(X\right)}{f\left(X\right)} \right| \left(\Pi_\theta \left[\left(b\left(X\right) - \widehat{b}\left(X\right)\right) |\bar{\phi}_k\right]\right)^2 \right] \\
&\leq \left\| \frac{\widehat{f}}{f} \right\|_\infty E_\theta \left[\left(\Pi_\theta \left[\left(b\left(X\right) - \widehat{b}\left(X\right)\right) |\bar{\phi}_k\right]\right)^2 \right] \\
&\leq \left\| \frac{\widehat{f}}{f} \right\|_\infty E_\theta \left[\left(b\left(X\right) - \widehat{b}\left(X\right)\right)^2 \right].
\end{aligned}
$$

In summary,

$$
\begin{aligned}
&\left|EB_2\right|^2 \\
&\leq \left|\left|\frac{f-\widehat{f}}{\widehat{f}}\right|\right|_\infty^2 \left|\left|\frac{f}{\widehat{f}}\right|\right|_\infty \left|\left|\frac{\widehat{f}}{f}\right|\right|_\infty \times \\
&E_\theta\left[\left(p\left(X\right)-\widehat{p}\left(X\right)\right)^2\right] E_\theta\left[\left(b\left(X\right)-\widehat{b}\left(X\right)\right)^2\right] \\
&= O_p\left(\left(\frac{\log n}{n}\right)^{\frac{2\beta_f}{d+2\beta_f}} n^{-\left(\frac{2\beta_b}{d+2\beta_b}+\frac{2\beta_p}{d+2\beta_p}\right)}\right)
\end{aligned}
$$

by condition $(A.3)$. □

Proof of Theorem 3. To examine the variance of $\widehat{\psi}_{2,k}$, we re-write $\widehat{\psi}_{2,k}$ as below. Note that $\widehat{b}\left(\cdot\right), \widehat{p}\left(\cdot\right)$, and $\widehat{f}\left(\cdot\right)$ are regarded as known functions:

$$
\begin{aligned}
\widehat{\psi}_{2,k} &= \psi_1\left(\widehat{\theta}\right) - \frac{1}{n\left(n-1\right)}\sum_{i\neq j} \widehat{\epsilon}_i K_{\widehat{f}_X,k}\left(X_i, X_j\right)\widehat{\Delta}_j \\
&= \frac{1}{n}\sum_{i=1}^n \left\{\begin{array}{c} \widehat{\epsilon}_i\widehat{\Delta}_i - \widehat{\epsilon}_i\bar{\phi}_k\left(X_i\right)^T E_\theta\left[\bar{\phi}_k\left(X\right)\widehat{\Delta}\right] \\ -E_\theta\left[\widehat{\epsilon}\bar{\phi}_k\left(X\right)^T\right]\widehat{\Delta}_i\bar{\phi}_k\left(X_i\right) + E_\theta\left[\widehat{\epsilon}\bar{\phi}_k\left(X\right)^T\right] E_\theta\left[\bar{\phi}_k\left(X\right)\widehat{\Delta}\right] \end{array}\right\} \\
&- \frac{1}{n\left(n-1\right)}\sum_{i\neq j}\left\{\begin{array}{c} \widehat{\epsilon}_i\bar{\phi}_k\left(X_i\right)^T\bar{\phi}_k\left(X_j\right)\widehat{\Delta}_j - \widehat{\epsilon}_i\bar{\phi}_k\left(X_i\right)^T E_\theta\left[\bar{\phi}_k\left(X\right)\widehat{\Delta}\right] \\ -E_\theta\left[\widehat{\epsilon}\bar{\phi}_k\left(X\right)^T\right]\bar{\phi}_k\left(X_j\right)\widehat{\Delta}_j \\ +E_\theta\left[\widehat{\epsilon}\bar{\phi}_k\left(X\right)^T\right] E_\theta\left[\bar{\phi}_k\left(X\right)\widehat{\Delta}\right] \end{array}\right\} \\
&= \frac{1}{n}\sum_{i=1}^n U_{1,i}^{\widehat{\psi}_{2,k}} + \frac{1}{n\left(n-1\right)}\sum_{i\neq j} U_{2,2,i_1,i_2}^{\widehat{\psi}_{2,k}}
\end{aligned}
$$

where

$$
U_{1,i}^{\widehat{\psi}_{2,k}} \equiv \left\{\begin{array}{c} \widehat{\epsilon}_i\widehat{\Delta}_i - \widehat{\epsilon}_i\bar{\phi}_k\left(X_i\right)^T E_\theta\left[\bar{\phi}_k\left(X\right)\widehat{\Delta}\right] - E_\theta\left[\widehat{\epsilon}\bar{\phi}_k\left(X\right)^T\right]\widehat{\Delta}_i\bar{\phi}_k\left(X_i\right) \\ +E_\theta\left[\widehat{\epsilon}\bar{\phi}_k\left(X\right)^T\right] E_\theta\left[\bar{\phi}_k\left(X\right)\widehat{\Delta}\right] \end{array}\right\}
$$

and

$$U_{2,2,i_1,i_2}^{\widehat{\psi}_{2,k}} \equiv -\left\{ \begin{array}{c} \widehat{\epsilon}_i \bar{\phi}_k (X_i)^T \bar{\phi}_k (X_j) \widehat{\Delta}_j - \widehat{\epsilon}_i \bar{\phi}_k (X_i)^T E_\theta \left[\bar{\phi}_k (X) \widehat{\Delta} \right] - \\ E_\theta \left[\widehat{\epsilon} \bar{\phi}_k (X)^T \right] \bar{\phi}_k (X_j) \widehat{\Delta}_j + E_\theta \left[\widehat{\epsilon} \bar{\phi}_k (X)^T \right] E_\theta \left[\bar{\phi}_k (X) \widehat{\Delta} \right] \end{array} \right\}.$$

It can be shown that $U_{1,i}^{\widehat{\psi}_{2,k}}$ and $U_{2,2,i_1,i_2}^{\widehat{\psi}_{2,k}}$ are uncorrelated wrt $F(\cdot;\theta)$. Thus,

$$var_\theta \left[\widehat{\psi}_{2,k} \right] = \frac{1}{n} var_\theta \left[U_{1,i}^{\widehat{\psi}_{2,k}} \right] + \frac{1}{2n(n-1)} E_\theta \left[\left\{ U_{2,2,i_1,i_2}^{\widehat{\psi}_{2,k}} + U_{2,2,i_2,i_1}^{\widehat{\psi}_{2,k}} \right\}^2 \right]$$

Next, we prove the two equations below hold. Then Theorem 3 holds as a direct consequence.

$$var_\theta \left[U_{1,i}^{\widehat{\psi}_{2,k}} \right] \asymp 1 \text{ wp } 1 \tag{16}$$

and

$$var_\theta \left[U_{2,2,i_1,i_2}^{\widehat{\psi}_{2,k}} \right] \asymp k \text{ wp } 1 \tag{17}$$

Eq. (16) is obvious as

$$var_\theta \left[\widehat{\epsilon}_i \widehat{\Delta}_i \right] \asymp 1 \text{ wp } 1,$$

$$\begin{aligned} & \left[\widehat{\epsilon} \overline{\phi}_k (X)^T \right]_i E_\theta \left[\widehat{\Delta} \overline{\phi}_k (X) \right] \\ &= \left[\widehat{\epsilon} \overline{\phi}_k (X)^T \right]_i E_\theta \left[(p(X) - \widehat{p}(X)) \overline{\phi}_k (X) \right] \\ &= E_{\widehat{\theta}} \left[(p(X) - \widehat{p}(X)) \frac{f(X)}{\widehat{f}(X)} \overline{\phi}_k (X)^T \right] \left[\overline{\phi}_k (X) \widehat{\epsilon} \right]_i \\ &= \left\{ \Pi_{\widehat{\theta}} \left[(p(X) - \widehat{p}(X)) \frac{f(X)}{\widehat{f}(X)} | \overline{\phi}_k (X) \right] \widehat{\epsilon} \right\}_i \\ &= O_p \left(\| P - \widehat{P} \| \right) = o_p(1), \end{aligned}$$

and similarly,

$$E_\theta \left[\widehat{\epsilon} \overline{\phi}_k (X)^T \right] E \left[\widehat{\Delta} \overline{\phi}_k (X) \right] \ll E_\theta \left[\widehat{\epsilon} \overline{\phi}_k (X)^T \right] \left[\widehat{\Delta} \overline{\phi}_k (X) \right]_i = o_p(1).$$

To prove eq. (17), we first show that, conditional on $\widehat{\theta}$

$$E_\theta\left[\left(\left[\widehat{\Delta}\overline{\phi}_k\left(X\right)^T\right]_i\left[\overline{\phi}_k\left(X\right)\widehat{\epsilon}\right]_j\right)^2\right]\asymp k$$

Specifically, by conditions $(A.1)-(A.3)$,

$$\begin{aligned}&E_\theta\left[\left(\left[\widehat{\Delta}\overline{\phi}_k\left(X\right)^T\right]_i\left[\overline{\phi}_k\left(X\right)\widehat{\epsilon}\right]_j\right)^2\right]\\&=E_\theta\left\{E_\theta\left[\widehat{\Delta}^2|X_i\right]E_\theta\left[\widehat{\epsilon}^2|X_j\right]\overline{\phi}_k\left(X_i\right)^T\overline{\phi}_k\left(X_j\right)\overline{\phi}_k\left(X_j\right)^T\overline{\phi}_k\left(X_i\right)\right\}\\&\asymp E_\theta\left[\overline{\phi}_k\left(X_i\right)^T\overline{\phi}_k\left(X_j\right)\overline{\phi}_k\left(X_j\right)^T\overline{\phi}_k\left(X_i\right)\right]\\&\asymp E_{\widehat{\theta}}\left[\overline{\phi}_k\left(X_i\right)^T\overline{\phi}_k\left(X_j\right)\overline{\phi}_k\left(X_j\right)^T\overline{\phi}_k\left(X_i\right)\right]=k\quad\text{wp }1.\end{aligned}$$

Further,

$$\begin{aligned}&E_\theta\left[\left(E_\theta\left[\widehat{\epsilon}\overline{\phi}_k\left(X\right)^T\right]\left[\overline{\phi}_k\left(X\right)\widehat{\Delta}\right]_j\right)^2\right]\\&=E_\theta\left[\left(\left[E_\theta\left[\left(\widehat{b}\left(X\right)-b\left(X\right)\right)\overline{\phi}_k\left(X\right)^T\right]\overline{\phi}_k\left(X\right)\widehat{\Delta}\right]_j\right)^2\right]\\&=O_p\left(\|b\left(\cdot\right)-\widehat{b}\left(\cdot\right)\|^2\right)=o_p\left(1\right).\end{aligned}$$

Similarly,

$$\begin{aligned}&E_\theta\left[\left(E_\theta\left[\widehat{\Delta}\overline{\phi}_k\left(X\right)^T\right]\left[\overline{\phi}_k\left(X\right)\widehat{\epsilon}\right]_j\right)^2\right]\\&=O_p\left(\|p\left(\cdot\right)-\widehat{p}\left(\cdot\right)\|^2\right)=o_p\left(1\right),\end{aligned}$$

and

$$\begin{aligned}&E_\theta\left\{\left[\widehat{\Delta}\overline{\phi}_k\left(X\right)^T\right]_i\left[\overline{\phi}_k\left(X\right)\widehat{\epsilon}\right]_j\right\}\\&=E_\theta\left[\left(P-\widehat{P}\right)\overline{\phi}_k\left(X\right)^T\right]E_\theta\left[\overline{\phi}_k\left(X\right)\left(B-\widehat{B}\right)\right]\\&=o_p\left(1\right).\end{aligned}$$

Therefore,

$$\begin{aligned}
&var_\theta\left[U_{2,2,i,j}^{\widehat{\psi}_{2,k}}\right]=E\left[\left(U_{2,2,i,j}^{\widehat{\psi}_{2,k}}\right)^2\right]\\
&\asymp E_\theta\left[\left(\left[\widehat{\Delta}\overline{\phi}_k\left(X\right)^T\right]_i\left[\overline{\phi}_k\left(X\right)\widehat{\epsilon}\right]_j\right)^2\right]\\
&\asymp k\text{ wp }1
\end{aligned}$$

□

Proof of Theorem 6. By definition, the conditional bias $BI\left(\widehat{\psi}_{m,k},\theta\right)$ of $\widehat{\psi}_{m,k}$ equals $TB_k\left(\theta\right)+EB_m\left(\theta\right)$ with $TB_k\left(\theta\right)$ given in Theorem 1 and

$$EB_m\left(\theta\right)=EB_2\left(\theta\right)-\sum_{j=3}^{m}E[\mathbb{H}_{j,j}^{(k)}].$$

Next, we prove eq. (13) by induction. The validity of eq. (13) at $m=2$ has been proved in Theorem 1. Suppose it holds for any $k\leq m-1$, we will prove it also holds at $k=m$.

Recall

$$\begin{aligned}
&E\left[\mathbb{H}_{m,m}^{(k)}\right]\\
&=\left(-1\right)^m E_\theta\left[\begin{array}{c}\widehat{\epsilon}_{i_1}\bar{\phi}_k\left(X_{i_1}\right)^T\prod_{r=3}^{m}\left\{\left(\bar{\phi}_k\left(X_{i_r}\right)\bar{\phi}_k\left(X_{i_r}\right)^T-I_{k\times k}\right)\right\}\\ \times\bar{\phi}_k\left(X_{i_2}\right)\widehat{\Delta}_{i_2}\end{array}\right]\\
&=\left(-1\right)^m\left\{\begin{array}{c}E_\theta\left[\left(b\left(X\right)-\widehat{b}\left(X\right)\right)\bar{\phi}_k\left(X\right)^T\right]\times\\ \left(E_\theta\left[\bar{\phi}_k\left(X\right)\bar{\phi}_k\left(X\right)^T\right]-I_{k\times k}\right)^{m-2}\\ \times E_\theta\left[\bar{\phi}_k\left(X\right)\left(p\left(X\right)-\widehat{p}\left(X\right)\right)\right]\end{array}\right\}.
\end{aligned}$$

Then by definition and assumption,

$$
\begin{aligned}
& EB_m(\theta) \\
& = EB_{m-1}(\theta) - E_\theta\left[\mathbb{H}_{m,m}^{(k)}\right] \\
& = (-1)^{m-1}\left\{\begin{array}{c} E_\theta\left[\left(b(X)-\widehat{b}(X)\right)\bar{\phi}_k(X)^T\right]\times \\ \left\{\left(E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right)^{-1}-I_{k\times k}\right\}\times \\ \left\{E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]-I_{k\times k}\right\}^{m-3} \\ \times E_\theta\left[\bar{\phi}_k(X)(p(X)-\widehat{p}(X))\right] \end{array}\right\} \\
& -(-1)^m\left\{\begin{array}{c} E_\theta\left[\left(b(X)-\widehat{b}(X)\right)\bar{\phi}_k(X)^T\right]\times \\ \left(E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]-I_{k\times k}\right)^{m-2} \\ \times E_\theta\left[\bar{\phi}_k(X)(p(X)-\widehat{p}(X))\right] \end{array}\right\} \\
& = (-1)^{m-1}\left\{\begin{array}{c} E_\theta\left[\left(b(X)-\widehat{b}(X)\right)\bar{\phi}_k(X)^T\right]\times \\ \left\{\begin{array}{c}\left(E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right)^{-1}-I_{k\times k} \\ +E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]-I_{k\times k}\end{array}\right\}\times \\ \left\{E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]-I_{k\times k}\right\}^{m-3} \\ \times E_\theta\left[\bar{\phi}_k(X)(p(X)-\widehat{p}(X))\right] \end{array}\right\}
\end{aligned}
$$

Note that

$$
\begin{aligned}
& \left(E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right)^{-1}-I_{k\times k}+E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]-I_{k\times k} \\
& = -\left(E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right)^{-1}\left[E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]-I_{k\times k}\right] \\
& + E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]-I_{k\times k} \\
& = -\left(\begin{array}{c}\left[\left(E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right)^{-1}-I_{k\times k}\right] \\ \times\left(E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]-I_{k\times k}\right)\end{array}\right)
\end{aligned}
$$

Thus,

$$EB_m(\theta)$$
$$= (-1)^m \left\{ \begin{array}{c} E_\theta\left[\left(b(X)-\widehat{b}(X)\right)\bar{\phi}_k(X)^T\right]\times \\ \left[\left(E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]\right)^{-1}-I_{k\times k}\right]\times \\ \left(E_\theta\left[\bar{\phi}_k(X)\bar{\phi}_k(X)^T\right]-I_{k\times k}\right)^{m-2} \\ \times E_\theta\left[\bar{\phi}_k(X)\left(p(X)-\widehat{p}(X)\right)\right] \end{array} \right\}$$

The order of $EB_m(\theta)$ can be derived similarly as in Theorem 2, and is omitted here but can be found in Robins et al. (2008). The proof of eq. 15 can be found in Robins et al. (2007). □

Theorem 10. *Define*

$$\widehat{\mathbb{W}}^2_{jj,k} = \binom{n}{j}^{-1} \mathbb{V}\left[\left(H^{(s)}_{jj,k}(O_{i_1}, O_{i_2}, ..., O_{i_m})\right)^2\right]$$

where $H^{(s)}_{jj,k}(O_{i_1}, O_{i_2}, ..., O_{i_m})$ *is the symmetric kernel of the mth-order U-statistic* $\mathbb{H}^{(k)}_{m,m}$ *defined in eq. (12). For instance,*

$$H^{(s)}_{22,k}(O_i, O_j) = \frac{\widehat{\epsilon}_{i_1}\overline{\phi}_k(X_{i_1})^T\overline{\phi}_k(X_{i_2})\widehat{\Delta}_{i_2} + \widehat{\Delta}_{i_1}\overline{\phi}_k(X_{i_1})^T\overline{\phi}_k(X_{i_2})\widehat{\epsilon}_{i_2}}{2},$$

so $H^{(s)}_{22,k}(O_i, O_j) = H^{(s)}_{22,k}(O_j, O_i)$*. For* $j \geq 2$*, define*

$$\widehat{\mathbb{W}}^2_{m,k} = \widehat{\mathbb{W}}^2_1 + \sum_{j=2}^{m} \widehat{\mathbb{W}}^2_{jj,k}.$$

i) Conditional on the training sample, we have

$$E_{\widehat{\theta}}\left[\widehat{\mathbb{W}}^2_{m,k}\right] = Var_{\widehat{\theta}}\left[\widehat{\psi}_{m,k}\right].$$

ii) Conditional on the training sample, if $k = k(n) >> n$,

$$\left\{\frac{1}{n}\left(\frac{k}{n}\right)^{m-1}\right\}^{-1/2}\left\{\widehat{\psi}_{m,k} - E\left[\widehat{\psi}_{m,k}\right]\right\}$$

converges uniformly for $\theta \in \Theta$ *to a normal distribution with finite variance as* $n \to \infty$. *The asymptotic variance is uniformly consistently estimated by*

$$\left\{\frac{1}{n}\max\left\{1, \left(\frac{k}{n}\right)^{m-1}\right\}\right\}^{-1}\widehat{\mathbb{W}}^2_{m,k}$$

Thus

$$\left\{\widehat{\psi}_{m,k} - E\left[\widehat{\psi}_{m,k}|\widehat{\theta}\right]\right\}/\widehat{\mathbb{W}}_{m,k}$$

converges in distribution to a standard normal distribution. iii) Define the interval $C_{m,k,\alpha} = \widehat{\psi}_{m,k} \pm z_{\alpha/2}\widehat{\mathbb{W}}_{m,k}$. *Let* $\widetilde{k}(m) >> k_*(m)$ *which is defined in Section 4.1.2. Then*

$$sup_{\theta\in\Theta}\left[\frac{E_\theta\left[\widehat{\psi}_{m,\widetilde{k}(m)}\right]}{\sqrt{Var_\theta\left[\widehat{\psi}_{m,\widetilde{k}(m)}\right]}}\right] = o_p(1)$$

and $\left\{\widehat{\psi}_{m,\widetilde{k}(m)} - \psi(\theta)\right\}/\widehat{\mathbb{W}}_{m,\widetilde{k}(m)}$ *converges uniformly in* $\theta \in \Theta$ *to a* $N(0,1)$. *Moreover,* $C_{m,\widetilde{k}(m),\alpha}$ *is a conservative uniform asymptotic* $(1-\alpha)$ *confidence interval for* $\psi(\theta)$.

Proof of Theorem 10. Part i) of the theorem is an easy calculation. Asymptotic normality in Part ii) is a corollary of Theorem 1.1 in Bhattacharya and Ghosh (1992) on the asymptotic distribution of degenerate U-statistics with kernels varying in n. Part iii) is immediate. □