Robust Causal Learning for the Estimation of Average Treatment Effects-Supplementary

1 Proofs

We present the theoretical proofs of Theorems and Corollaries given in the main paper.

Proof of Theorem ??.
Given the nuisance parameters $\varrho = (\mathscr{g}^i, a_i)$ and the true nuisance parameters $\rho = (g^i, \pi^i)$, we find out the RCL score $\psi^i(W, \vartheta, \varrho)$ w.r.t. the nuisance parameters $\varrho = (\mathscr{g}^i, a_i)$ which can be used to construct the estimators of the causal parameter $\theta^i := \mathbb{E}\left[g^i(\mathbf{Z})\right]$. We try an ansatz of $\psi^i(W, \vartheta, \varrho)$ such that

$$\psi^i(W, \vartheta, \varrho) = \vartheta - \mathscr{g}^i(\mathbf{Z}) - (Y^i - \mathscr{g}^i(\mathbf{Z}))A(D, \mathbf{Z}; a_i), \tag{1}$$

where

$$\begin{aligned} A(D, \mathbf{Z}; a_i) = &\bar{b}_r \left[\mathbf{1}_{\{D=d^i\}} - a_i(\mathbf{Z})\right]^r \\ &+ \sum_{q=1}^{k-1} b_q \big(\left[\mathbf{1}_{\{D=d^i\}} - a_i(\mathbf{Z})\right]^q - \mathbb{E}\left[(\nu^i)^q \mid \mathbf{Z}\right]\big). \end{aligned} \tag{2}$$

Here, the coefficients $b_1, \cdots, b_{k-1}, \bar{b}_r$ depend on $\mathbf{Z}$ and ν^i only. Using the ansatz, we notice that $\psi^i(W, \vartheta, \varrho)$ satisfies condition 1, i.e., $\mathbb{E}\left[\psi^i(W, \vartheta, \varrho) \mid_{\vartheta=\theta^i,\ \varrho=\rho}\right] = 0$. Indeed, we have

$$\begin{aligned} &\mathbb{E}\left[\psi^i(W, \vartheta, \varrho) \mid_{\vartheta=\theta^i,\ \varrho=\rho}\right] \\ =&\mathbb{E}\left[\theta^i - g^i(\mathbf{Z}) - (Y^i - g^i(\mathbf{Z}))A(D, \mathbf{Z}; \pi^i)\right] \\ =&\mathbb{E}\left[\theta^i - g^i(\mathbf{Z})\right] - \mathbb{E}\left[(Y^i - g^i(\mathbf{Z}))A(D, \mathbf{Z}; \pi^i)\right] \\ =&-\mathbb{E}\left[\xi^i \times A(D, \mathbf{Z}; \pi^i)\right] \\ =&-\mathbb{E}\left[\mathbb{E}\left[\xi^i \times A(D, \mathbf{Z}; \pi^i) \mid D, \mathbf{Z}\right]\right] \\ =&-\mathbb{E}\left[A(D, \mathbf{Z}; \pi^i)\mathbb{E}\left[\xi^i \mid D, \mathbf{Z}\right]\right] \\ =&-\mathbb{E}\left[A(D, \mathbf{Z}; \pi^i)\mathbb{E}\left[\xi^i \mid \mathbf{Z}\right]\right] = 0. \end{aligned}$$

The second last equality comes from the fact that $A(D, \mathbf{Z}; \pi^i)$ is a function of $(D, \mathbf{Z})$. The last equality comes from the fact that $(\xi^i \perp\!\!\!\perp D) \mid \mathbf{Z}$. Now, we aim to find out the coefficients of $b_1, \cdots, b_{k-1}, \bar{b}_r$ such that the score (1) satisfies the k^{th} score. Indeed, we need to have $\mathbb{E}\left[\partial^{\alpha_1}_{\mathscr{g}^i} \partial^{\alpha_2}_{a_i} \psi^i(W, \vartheta, \varrho) \mid_{\vartheta=\theta^i,\ \varrho=\rho} \mid \mathbf{Z}\right] = 0$ for all α_1 and α_2 which are non-negative integers such that $1 \le \alpha_1 + \alpha_2 \le k$. Since $\partial^{\alpha_1}_{\mathscr{g}^i} \partial^{\alpha_2}_{a_i} \psi^i(W, \vartheta, \varrho) = 0$ when $\alpha_1 \ge 2$, we only need to solve the coefficients $b_1, \cdots, b_{k-1}, \bar{b}_r$ from

$$\begin{cases} 0 = \mathbb{E}\left[\partial^k_{a_i} \psi^i(W, \vartheta, \varrho) \mid_{\vartheta=\theta^i,\ \varrho=\rho} \mid \mathbf{Z}\right], & \text{(3a)} \\ 0 = \mathbb{E}\left[\partial^1_{\mathscr{g}^i} \partial^q_{a_i} \psi^i(W, \vartheta, \varrho) \mid_{\vartheta=\theta^i,\ \varrho=\rho} \mid \mathbf{Z}\right] & \text{(3b)} \end{cases}$$

$\forall q = 0, \cdots, k-1$. However, (3a) always holds since

$$\begin{aligned} &\mathbb{E}\left[\partial^k_{a_i} \psi^i(W, \vartheta, \varrho) \mid_{\vartheta=\theta^i,\ \varrho=\rho} \mid \mathbf{Z}\right] \\ =&\mathbb{E}\left[(Y^i - g^i(\mathbf{Z})) \times \partial^k_{a_i} A(D, \mathbf{Z}; a_i) \mid_{a_i=\pi^i} \mid \mathbf{Z}\right] \\ =&\mathbb{E}\left[\mathbb{E}\left[(Y^i - g^i(\mathbf{Z})) \times \partial^k_{a_i} A(D, \mathbf{Z}; a_i) \mid_{a_i=\pi^i} \mid D, \mathbf{Z}\right] \mid \mathbf{Z}\right] \\ =&\mathbb{E}\left[\partial^k_{a_i} A(D, \mathbf{Z}; a_i) \mid_{a_i=\pi^i} \mathbb{E}\left[(Y^i - g^i(\mathbf{Z})) \mid D, \mathbf{Z}\right] \mid \mathbf{Z}\right] \\ =&\mathbb{E}\left[\partial^k_{a_i} A(D, \mathbf{Z}; a_i) \mid_{a_i=\pi^i} \mathbb{E}\left[\xi^i \mid \mathbf{Z}\right] \mid \mathbf{Z}\right] = 0. \end{aligned}$$

Consequently, we need to find out the coefficients $b_1, \cdots, b_{k-1}, \bar{b}_r$ from

$$\mathbb{E}\left[\partial^1_{\mathscr{g}^i} \partial^q_{a_i} \psi^i(W, \vartheta, \varrho) \mid_{\vartheta=\theta^i,\ \varrho=\rho} \mid \mathbf{Z}\right] = 0 \tag{3b}$$

$\forall q = 0, \cdots, k-1$. From (3b), there are k equations and we need to solve the k unknowns $b_1, \cdots, b_{k-1}, \bar{b}_r$ from the k equations. Generally, the k unknowns could be solved uniquely.

To start with, we compute $\partial^1_{\mathscr{g}^i} \partial^q_{a_i} \psi^i(W, \vartheta, \varrho)$ for $q = 0, \cdots, k-1$. Note that

$$\partial^1_{\mathscr{g}^i} \partial^q_{a_i} \psi^i(W, \vartheta, \varrho) = -1 + A(D, \mathbf{Z}; a_i)$$

when $q = 0$ and

$$\begin{aligned} \partial^1_{\mathscr{g}^i} \partial^q_{a_i} \psi^i(W, \vartheta, \varrho) = &\bar{b}_r \frac{r!(-1)^q [\mathbf{1}_{\{D=d^i\}} - a_i(\mathbf{Z})]^{r-q}}{(r-q)!} \\ &+ \sum_{u=q}^{k-1} b_u \frac{u!(-1)^q [\mathbf{1}_{\{D=d^i\}} - a_i(\mathbf{Z})]^{u-q}}{(u-q)!} \end{aligned}$$

when $1 \le q \le k-1$. Consequently, we need to solve for $b_1, \cdots, b_{k-1}$ and $\bar{b}_r$ simultaneously from

$$1 = \mathbb{E}\left[A(D, \mathbf{Z}; \pi^i) \mid \mathbf{Z}\right] \tag{4a}$$

and

$$\begin{aligned} 0 = &\mathbb{E}\left[\bar{b}_r \frac{r!(-1)^q [\mathbf{1}_{\{D=d^i\}} - \pi^i(\mathbf{Z})]^{r-q}}{(r-q)!} \mid \mathbf{Z}\right] \\ &+ \mathbb{E}\left[\sum_{u=q}^{k-1} b_u \frac{u!(-1)^q [\mathbf{1}_{\{D=d^i\}} - \pi^i(\mathbf{Z})]^{u-q}}{(u-q)!} \mid \mathbf{Z}\right]. \end{aligned} \tag{4b}$$

From (4a), we have

$$\begin{aligned}1 = &\bar{b}_r \mathbb{E}\left[\left(\mathbf{1}_{\{D=d^i\}} - \pi^i(\mathbf{Z})\right)^r \mid \mathbf{Z}\right] \\ &+ \sum_{q=1}^{k-1} b_q \mathbb{E}\left[\left(\mathbf{1}_{\{D=d^i\}} - \pi^i(\mathbf{Z})\right)^q \mid \mathbf{Z}\right] \\ &- \sum_{q=1}^{k-1} b_q \mathbb{E}\left[\mathbb{E}\left[(\nu^i)^q \mid \mathbf{Z}\right] \mid \mathbf{Z}\right]\end{aligned}$$

Since $\mathbb{E}\left[\left(\mathbf{1}_{\{D=d^i\}} - \pi^i(\mathbf{Z})\right)^q \mid \mathbf{Z}\right] = \mathbb{E}\left[(\nu^i)^q \mid \mathbf{Z}\right]$ and $\mathbb{E}\left[\mathbb{E}\left[(\nu^i)^q \mid \mathbf{Z}\right] \mid \mathbf{Z}\right] = \mathbb{E}\left[(\nu^i)^q \mid \mathbf{Z}\right]$, we understand that $\mathbb{E}\left[\left(\mathbf{1}_{\{D=d^i\}} - \pi^i(\mathbf{Z})\right)^q \mid \mathbf{Z}\right] - \mathbb{E}\left[\mathbb{E}\left[(\nu^i)^q \mid \mathbf{Z}\right] \mid \mathbf{Z}\right] = 0$. The above equality can therefore be reduced as

$$\begin{aligned}&\bar{b}_r \mathbb{E}\left[\left(\mathbf{1}_{\{D=d^i\}} - \pi^i(\mathbf{Z})\right)^r \mid \mathbf{Z}\right] = 1 \\ &\Rightarrow \bar{b}_r \mathbb{E}\left[(\nu^i)^r \mid \mathbf{Z}\right] = 1.\end{aligned}$$

Hence, we can solve for $\bar{b}_r$ such that

$$\bar{b}_r = \frac{1}{\mathbb{E}\left[(\nu^i)^r \mid \mathbf{Z}\right]},$$

It remains to find out $b_1, \cdots, b_{k-1}$ from (4b). Indeed, we can simplify (4b) as

$$\begin{aligned}&\bar{b}_r \mathbb{E}\left[\frac{r!(\nu^i)^{r-q}}{(r-q)!} \mid \mathbf{Z}\right] + \sum_{u=q}^{k-1} b_u \mathbb{E}\left[\frac{u!(\nu^i)^{u-q}}{(u-q)!} \mid \mathbf{Z}\right] = 0 \\ &\Rightarrow \bar{b}_r \binom{r}{q} \mathbb{E}\left[(\nu^i)^{r-q} \mid \mathbf{Z}\right] + \sum_{u=q}^{k-1} b_u \binom{u}{q} \mathbb{E}\left[(\nu^i)^{u-q} \mid \mathbf{Z}\right] = 0 \end{aligned} \tag{5}$$

$\forall 1 \leq q \leq k-1$. Now, we solve $b_1, \cdots, b_{k-1}$. We start with finding out b_{k-1}, followed by $b_{k-2}, b_{k-3}, \cdots, b_1$. When $q = k-1$, (5) becomes

$$\begin{aligned}0 = &\bar{b}_r \binom{r}{k-1} \mathbb{E}\left[(\nu^i)^{r-k+1} \mid \mathbf{Z}\right] \\ &+ b_{k-1} \binom{k-1}{k-1} \mathbb{E}\left[(\nu^i)^0 \mid \mathbf{Z}\right] \\ \Rightarrow b_{k-1} = &-\bar{b}_r \binom{r}{k-1} \mathbb{E}\left[(\nu^i)^{r-k+1} \mid \mathbf{Z}\right].\end{aligned}$$

Now, when $q = k-2$, (5) becomes

$$\begin{aligned}0 = &\bar{b}_r \binom{r}{k-2} \mathbb{E}\left[(\nu^i)^{r-k+2} \mid \mathbf{Z}\right] \\ &+ b_{k-1} \binom{k-1}{k-2} \mathbb{E}\left[(\nu^i)^{(k-1)-(k-2)} \mid \mathbf{Z}\right] \\ &+ b_{k-2} \mathbb{E}\left[(\nu^i)^0 \mid \mathbf{Z}\right] \\ \Rightarrow b_{k-2} = &-b_{k-1} \binom{k-1}{k-2} \mathbb{E}\left[(\nu^i)^1 \mid \mathbf{Z}\right] \\ &- \bar{b}_r \binom{r}{k-2} \mathbb{E}\left[(\nu^i)^{r-k+2} \mid \mathbf{Z}\right].\end{aligned}$$

Iteratively, supposing $b_{q+1}, \cdots, b_{k-1}$ are known and we want to find out what b_q is, we have to solve it from

$$\begin{aligned}0 = &b_q \mathbb{E}\left[(\nu^i)^0 \mid \mathbf{Z}\right] \\ &+ \bar{b}_r \binom{r}{q} \mathbb{E}\left[(\nu^i)^{r-q} \mid \mathbf{Z}\right] \\ &+ \sum_{u=q+1}^{k-1} b_u \binom{u}{q} \mathbb{E}\left[(\nu^i)^{u-q} \mid \mathbf{Z}\right].\end{aligned}$$

We can obtain b_q from the above equation, which gives

$$\begin{aligned}b_q = &-\sum_{u=q+1}^{k-1} b_u \binom{u}{q} \mathbb{E}\left[(\nu^i)^{u-q} \mid \mathbf{Z}\right] \\ &- \bar{b}_r \binom{r}{q} \mathbb{E}\left[(\nu^i)^{r-q} \mid \mathbf{Z}\right] \\ \Rightarrow b_q = &-\sum_{u=1}^{k-1-q} b_{q+u} \binom{q+u}{q} \mathbb{E}\left[(\nu^i)^u \mid \mathbf{Z}\right] \\ &- \bar{b}_r \binom{r}{q} \mathbb{E}\left[(\nu^i)^{r-q} \mid \mathbf{Z}\right].\end{aligned}$$

The proof is completed. □

Proof of Corollary ??. We have discussed the way to obtain the estimator in the main paper. To facilitate our following proofs, we first define some notations. We denote $Y^{i;F}$ as the factual outcome if one receives d^i, i.e., $Y^{i;F} = Y^F = Y^i$ if $D = d^i$. We denote $Y^{i;CF}$ as the counterfactual outcome if one is not treated with d^i, i.e., $Y^{i;CF} = Y^i$ if $D \neq d^i$. Thus, for m^{th} individual, $\xi_m^{i;F} = Y_m^{i;F} - g^i(\mathbf{Z}_m)$ if $m \in \mathscr{I}$ and $\xi_m^{i;CF} = Y_m^{i;CF} - g^i(\mathbf{Z}_m)$ if $m \in \mathscr{I}^c$. □

First, we introduce proposition 1.

Proposition 1. *Given the covariates* $\mathbf{Z}$*, the random variable* $\xi_m^{i;F}$ *and* $\xi_{\bar{m}}^{i;CF}$ *are independent and identically distributed, i.e.,* $\xi_m^{i;F} \overset{d}{=} \xi_{\bar{m}}^{i;CF} \overset{d}{=} \xi^i$ *and* $\xi_m^{i;F} \perp\!\!\!\perp \xi_{\bar{m}}^{i;CF}$.

Proof.

$\xi_m^{i;F} \mid D_m = d^i, \mathbf{Z}_m \overset{\triangle}{=} \xi_m^i \mid D_m = d^i, \mathbf{Z}_m$;
Using ignorability assumption, we have
$\mathbb{E}\left[(\xi_m^i)^r \mid D_m = d^i, \mathbf{Z}_m\right] = \mathbb{E}\left[(\xi_m^i)^r \mid \mathbf{Z}_m\right] = \mathbb{E}\left[(\xi^i)^r \mid \mathbf{Z}\right]$.
$\xi_{\bar{m}}^{i;F} \mid D_{\bar{m}} = d^i, \mathbf{Z}_{\bar{m}} \overset{\triangle}{=} \xi_{\bar{m}}^i \mid D_{\bar{m}} = d^i, \mathbf{Z}_{\bar{m}}$;
Using ignorability assumption, we have
$\mathbb{E}\left[(\xi_{\bar{m}}^i)^r \mid D_{\bar{m}} = d^i, \mathbf{Z}_{\bar{m}}\right] = \mathbb{E}\left[(\xi_{\bar{m}}^i)^r \mid \mathbf{Z}_{\bar{m}}\right] = \mathbb{E}\left[(\xi^i)^r \mid \mathbf{Z}\right]$.

According to the moment generating function, we can conclude $\xi_m^{i;F} \overset{d}{=} \xi_{\bar{m}}^{i;CF} \overset{d}{=} \xi^i \mid \mathbf{Z}$. Using the SUTVA assumption, we have $\xi_m^{i;F} \perp\!\!\!\perp \xi_{\bar{m}}^{i;CF}$. □

For notational simplicity, let our RCL estimator $\hat{\theta}^i_{RCL}$ be $\hat{\theta}^i_N$.

$$\hat{\theta}^i_N = \underbrace{\frac{1}{N}\sum_{m=1}^{N}\hat{g}^i(\mathbf{Z}_m)}_{(a)} + \underbrace{\frac{1}{N}\sum_{m\in\mathscr{I}}(Y^{i;F}_m - \hat{g}^i(\mathbf{Z}_m))\hat{A}^i_m}_{(b)} + \underbrace{\frac{1}{R}\sum_{u=1}^{R}\left[\frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;F}_{m,u}\hat{A}^i_m\right]}_{(c)}, \tag{6}$$

In the remaining sequel, we investigate the consistency of our RCL estimators provided that $\hat{g}^i(\cdot)$ and $\hat{\pi}^i(\cdot)$ are some good estimates of $g^i(\cdot)$ and $\pi^i(\cdot)$. The assumptions on the $\hat{g}^i(\cdot)$ and $\hat{\pi}^i(\cdot)$ are given in [Mackey et al., 2018]. Before that, we define two quantities $\hat{\tilde{\theta}}^i_N$ and $\hat{\bar{\theta}}^i_N$. They are

$$\begin{aligned}\hat{\tilde{\theta}}^i_N &= \frac{1}{N}\sum_{m=1}^{N}\hat{g}^i(\mathbf{Z}_m) + \frac{1}{N}\sum_{m\in\mathscr{I}}(Y^{i;F}_m - \hat{g}^i(\mathbf{Z}_m))\hat{A}^i_m \\ &\quad + \frac{1}{N}\sum_{m\in\mathscr{I}^c}(Y^{i;CF}_m - \hat{g}^i(\mathbf{Z}_m))\hat{A}^i_m \\ &= \frac{1}{N}\sum_{m=1}^{N}\hat{g}^i(\mathbf{Z}_m) + \frac{1}{N}\sum_{m\in\mathscr{I}}(Y^{i;F}_m - \hat{g}^i(\mathbf{Z}_m))\hat{A}^i_m \\ &\quad + \frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;CF}_m\hat{A}^i_m, \qquad (7) \\ \hat{\bar{\theta}}^i_N &= \frac{1}{N}\sum_{m=1}^{N}\hat{g}^i(\mathbf{Z}_m) + \frac{1}{N}\sum_{m\in\mathscr{I}}(Y^{i;F}_m - \hat{g}^i(\mathbf{Z}_m))\hat{A}^i_m \\ &\quad + \frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;F}_m\hat{A}^i_m. \qquad (8)\end{aligned}$$

We also define

$$\begin{aligned}\kappa^{i;F}_N &= \frac{1}{N}\sum_{m\in\mathscr{I}^c}\xi^{i;F}_m A^i_m,\ \hat{\kappa}^{i;F}_N = \frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;F}_m\hat{A}^i_m, \\ \kappa^{i;CF}_N &= \frac{1}{N}\sum_{m\in\mathscr{I}^c}\xi^{i;CF}_m A^i_m,\ \hat{\kappa}^{i;CF}_N = \frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;CF}_m\hat{A}^i_m, \\ \hat{\kappa}^{i;F}_{R,N} &= \frac{1}{R}\sum_{u=1}^{R}[\frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;F}_{m,u}\hat{A}^i_m], \\ \kappa^{i;F}_{R,N} &= \frac{1}{R}\sum_{u=1}^{R}[\frac{1}{N}\sum_{m\in\mathscr{I}^c}\xi^{i;F}_{m,u}A^i_m].\end{aligned}$$

Then (7) and (8) can be rewritten as

$$\begin{aligned}\hat{\tilde{\theta}}^i_N &= \frac{1}{N}\sum_{m=1}^{N}\hat{g}^i(\mathbf{Z}_m) \\ &\quad + \frac{1}{N}\sum_{m\in\mathscr{I}}(Y^{i;F}_m - \hat{g}^i(\mathbf{Z}_m))\hat{A}^i_m + \hat{\kappa}^{i;CF}_N, \qquad (7) \\ \hat{\bar{\theta}}^i_N &= \frac{1}{N}\sum_{m=1}^{N}\hat{g}^i(\mathbf{Z}_m) \\ &\quad + \frac{1}{N}\sum_{m\in\mathscr{I}}(Y^{i;F}_m - \hat{g}^i(\mathbf{Z}_m))\hat{A}^i_m + \hat{\kappa}^{i;F}_N. \qquad (8)\end{aligned}$$

for simplicity.

In addition, we have to use two lemmas and two propositions to study the consistency of $\hat{\theta}^i_N$. We state them with the proofs.

Lemma 2. *Given two sequences of random variables $(X_N)^\infty_{N=1}$ and $(Y_N)^\infty_{N=1}$ such that $X_N \stackrel{d}{=} Y_N$. If $X_N \stackrel{p}{\to} c$ for some constant c, then $Y_N \stackrel{p}{\to} c$.*

Proof. Let $f_{X_N}(\cdot)$ and $f_{Y_N}(\cdot)$ be the density functions of the random variables X_N and Y_N respectively. Since $X_N \stackrel{d}{=} Y_N$, $f_{X_N}(\cdot) = f_{Y_N}(\cdot)$. Hence, $\mathbb{P}\{|X_N - c| \geq \epsilon\} = \int_{|z-c|\geq\epsilon} f_{X_N}(z)dz = \int_{|z-c|\geq\epsilon} f_{Y_N}(z)dz = \mathbb{P}\{|Y_N - c| \geq \epsilon\}$. Consequently, $X_N \stackrel{p}{\to} c$ implies $Y_N \stackrel{p}{\to} c$. □

Lemma 3. *Given random variables X, Y, E, Z. If $(X \stackrel{d}{=} Y) \mid Z$, $(X \perp\!\!\!\perp E) \mid Z$, $(Y \perp\!\!\!\perp E) \mid Z$, then $Xh(E,Z) \stackrel{d}{=} Yh(E,Z)$ for any function h.*

Proof. Define $f_Z(z)$ as the density function of Z, $f_{X|Z}(x|z)$ is the conditional density function of $X|Z$, $f_{Y|Z}(y|z)$ is the conditional density function of $Y|Z$, $f_{E|Z}(e|z)$ is the conditional density function of $E|Z$, $f_{X,E|Z}(x,e|z)$ is the conditional joint density function of $X,E|Z$, and $f_{Y,E|Z}(y,e|z)$ is the conditional joint density function of $Y,E|Z$. For a measurable set $\mathscr{A}$, we have

$$\begin{aligned}&\mathbb{P}\{Xh(E,Z)\in\mathscr{A}\} \\ =&\int_{\Omega_Z}\mathbb{P}\{Xh(E,Z)\in\mathscr{A}|Z=z\}f_Z(z)dz \\ =&\int_{\Omega_Z}\mathbb{P}\{Xh(E,z)\in\mathscr{A}|Z=z\}f_Z(z)dz \\ =&\int_{\Omega_Z}\left\{\iint_{\Omega_X\times\Omega_E}\mathbf{1}_{\{xh(e,z)\in\mathscr{A}\}}f_{X,E|Z}(x,e|z)dxde\right\}f_Z(z)dz \\ \stackrel{*}{=}&\int_{\Omega_Z}\left\{\iint_{\Omega_X\times\Omega_E}\mathbf{1}_{\{xh(e,z)\in\mathscr{A}\}}f_{X|Z}(x|z)f_{E|Z}(e|z)dxde\right\}f_Z(z)dz \\ \stackrel{\triangle}{=}&\int_{\Omega_Z}\left\{\iint_{\Omega_Y\times\Omega_E}\mathbf{1}_{\{yh(e,z)\in\mathscr{A}\}}f_{Y|Z}(y|z)f_{E|Z}(e|z)dyde\right\}f_Z(z)dz \\ \stackrel{\square}{=}&\int_{\Omega_Z}\left\{\iint_{\Omega_Y\times\Omega_E}\mathbf{1}_{\{yh(e,z)\in\mathscr{A}\}}f_{Y,E|Z}(y,e|z)dyde\right\}f_Z(z)dz \\ =&\mathbb{P}\{Yh(E,Z)\in\mathscr{A}\}\end{aligned}$$

$*$ holds since $(X \perp\!\!\!\perp E) \mid Z$, $\triangle$ holds since $(X \overset{d}{=} Y) \mid Z$, and $\square$ holds since $(Y \perp\!\!\!\perp E) \mid Z$. $\square$

Proposition 4. *Given two i.i.d. sequences $(\xi_m^{i;F})$ and $(\xi_m^{i;CF})$. Suppose $(\xi_m^{i;F} \overset{d}{=} \xi_{\bar{m}}^{i;CF}) \mid \mathbf{Z}$ and $(\xi_m^{i;F} \perp\!\!\!\perp \xi_{\bar{m}}^{i;CF}) \mid \mathbf{Z}$ for any m and $\bar{m}$. Furthermore, suppose $\mathbb{E}\left[(\xi_m^{i;F})^2 \mid \mathbf{Z}\right]$ and $(A_m^i)^2$ exist such that $\mathbb{E}\left[(A_m^i)^2 \mathbb{E}\left[(\xi_m^{i;F})^2 \mid \mathbf{Z}\right]\right]$ is finite for all m. We have $\kappa_N^{i;CF} - \kappa_N^{i;F} \xrightarrow{p} 0$ when $N \to \infty$.*

Proof. $\forall \epsilon > 0$, we consider $\mathbb{P}\left\{\left|\kappa_N^{i;CF} - \kappa_N^{i;F}\right| \geq \epsilon\right\}$. Indeed, we have

$$\mathbb{P}\left\{\left|\kappa_N^{i;CF} - \kappa_N^{i;F}\right| \geq \epsilon\right\} \leq \frac{\mathbb{E}\left[\left(\kappa_N^{i;CF} - \kappa_N^{i;F}\right)^2\right]}{\epsilon^2}$$
$$= \frac{\frac{1}{N^2}\mathbb{E}\left[\left(\sum_{m \in \mathscr{I}^c} (\xi_m^{i;CF} - \xi_m^{i;F}) A_m^i\right)^2\right]}{\epsilon^2}.$$

Denoting $\xi_m^{i;CF} - \xi_m^{i;F}$ as Ξ_m^i, we have

$$\begin{aligned}
&\mathbb{E}[(\sum_{m \in \mathscr{I}^c} (\xi_m^{i;CF} - \xi_m^{i;F}) A_m^i)^2] = \mathbb{E}[(\sum_{m \in \mathscr{I}^c} \Xi_m^i A_m^i)^2] \\
=&\mathbb{E}[\sum_{m,\bar{m} \in \mathscr{I}^c} \Xi_m^i A_m^i \Xi_{\bar{m}}^i A_{\bar{m}}^i] = \sum_{m,\bar{m} \in \mathscr{I}^c} \mathbb{E}\left[\Xi_m^i A_m^i \Xi_{\bar{m}}^i A_{\bar{m}}^i\right] \\
=& \sum_{m \in \mathscr{I}^c} \mathbb{E}\left[(A_m^i)^2 \mathbb{E}\left[(\Xi_m^i)^2 \mid D, \mathbf{Z}\right]\right] \\
&+ \sum_{\substack{m,\bar{m} \in \mathscr{I}^c \\ m \neq \bar{m}}} \mathbb{E}\left[A_m^i A_{\bar{m}}^i \mathbb{E}\left[\Xi_m^i \Xi_{\bar{m}}^i \mid D, \mathbf{Z}\right]\right] \\
=& \sum_{m \in \mathscr{I}^c} \mathbb{E}\left[(A_m^i)^2 \mathbb{E}\left[(\Xi_m^i)^2 \mid D, \mathbf{Z}\right]\right] \\
&+ \sum_{\substack{m,\bar{m} \in \mathscr{I}^c \\ m \neq \bar{m}}} \mathbb{E}\left[A_m^i A_{\bar{m}}^i \mathbb{E}\left[\Xi_m^i \mid D, \mathbf{Z}\right] \mathbb{E}\left[\Xi_{\bar{m}}^i \mid D, \mathbf{Z}\right]\right] \\
=& \sum_{m \in \mathscr{I}^c} \mathbb{E}\left[(A_m^i)^2 \mathbb{E}\left[(\Xi_m^i)^2 \mid \mathbf{Z}\right]\right] \\
=&2 \sum_{m \in \mathscr{I}^c} \mathbb{E}\left[(A_m^i)^2 \mathbb{E}\left[(\xi_m^{i;F})^2 \mid \mathbf{Z}\right]\right] \\
\leq& 2N \mathbb{E}\left[(A^i)^2 \mathbb{E}\left[(\xi^{i;F})^2 \mid \mathbf{Z}\right]\right].
\end{aligned}$$

The last equality follows from

$$\begin{aligned}
&\mathbb{E}[(\Xi_m^i)^2 \mid \mathbf{Z}] = \mathbb{E}[(\xi_m^{i;CF} - \xi_m^{i;F})^2 \mid \mathbf{Z}] \\
=&\mathbb{E}[(\xi_m^{i;CF})^2 \mid \mathbf{Z}] - 2\mathbb{E}[\xi_m^{i;F} \xi_m^{i;CF} \mid \mathbf{Z}] + \mathbb{E}[(\xi_m^{i;F})^2 \mid \mathbf{Z}] \\
=&\mathbb{E}[(\xi_m^{i;CF})^2 \mid \mathbf{Z}] - 2\mathbb{E}[\xi_m^{i;F} \mid \mathbf{Z}]\mathbb{E}[\xi_m^{i;CF} \mid \mathbf{Z}] + \mathbb{E}[(\xi_m^{i;F})^2 \mid \mathbf{Z}] \\
=&\mathbb{E}[(\xi_m^{i;CF})^2 \mid \mathbf{Z}] + \mathbb{E}[(\xi_m^{i;F})^2 \mid \mathbf{Z}] = 2\mathbb{E}[(\xi_m^{i;F})^2 \mid \mathbf{Z}].
\end{aligned}$$

As a consequence, we have

$$\begin{aligned}
\mathbb{P}\left\{\left|\kappa_N^{i;CF} - \kappa_N^{i;F}\right| \geq \epsilon\right\} &\leq \frac{2N\mathbb{E}\left[(A^i)^2 \mathbb{E}\left[(\xi^{i;F})^2 \mid \mathbf{Z}\right]\right]}{N^2 \epsilon^2} \\
&= \frac{2\mathbb{E}\left[(A^i)^2 \mathbb{E}\left[(\xi^{i;F})^2 \mid \mathbf{Z}\right]\right]}{N \epsilon^2} \to 0
\end{aligned}$$

when $N \to \infty$. As a result, we have $\kappa_N^{i;CF} - \kappa_N^{i;F} \xrightarrow{p} 0$. The proof is completed. $\square$

Proposition 5. *Suppose that, conditioning on $\mathbf{Z}$, $\xi_{m,u}^{i;F}$ are i.i.d. of $\xi_m^{i;F}$ and $\xi_{m,u}^{i;F}$ are i.i.d. of $\xi_{m,\bar{u}}^{i;F}$ $\forall u$, $\bar{u} \in \{1, 2, \cdots, R\}$. We have*

$$\kappa_N^{i;F} - \kappa_{R,N}^{i;F} \xrightarrow{p} 0 \quad \text{when} \quad N \to \infty.$$

Proof.
Write

$$\begin{aligned}
\kappa_{R,N}^{i;F} &= \frac{1}{R}\sum_{u=1}^{R}[\frac{1}{N}\sum_{m \in \mathscr{I}^c} \xi_{m,u}^{i;F} A_m^i] \\
&= \frac{1}{N}\sum_{m \in \mathscr{I}^c}\left(\frac{1}{R}\sum_{u=1}^{R}\xi_{m,u}^{i;F}\right) A_m^i = \frac{1}{N}\sum_{m \in \mathscr{I}^c} \mathscr{E}_m^i A_m^i
\end{aligned}$$

$\forall \epsilon > 0$, we have

$$\mathbb{P}\left\{\left|\kappa_N^{i;F} - \kappa_{R,N}^{i;F}\right| \geq \epsilon\right\} \leq \frac{\mathbb{E}\left[\left(\frac{1}{N}\sum_{m \in \mathscr{I}^c}\left[\mathscr{E}_m^i - \xi_m^{i;F}\right] A_m^i\right)^2\right]}{\epsilon^2}.$$

Considering $\mathbb{E}\left[\left(\frac{1}{N}\sum_{m \in \mathscr{I}^c}\left[\mathscr{E}_m^i - \xi_m^{i;F}\right] A_m^i\right)^2\right]$, we have

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{N}\sum_{m \in \mathscr{I}^c}\left[\mathscr{E}_m^i - \xi_m^{i;F}\right] A_m^i\right)^2\right] \\
=&\frac{1}{N^2}\sum_{m,\bar{m} \in \mathscr{I}^c}\mathbb{E}\left[\left(\mathscr{E}_m^i - \xi_m^{i;F}\right)\left(\mathscr{E}_{\bar{m}}^i - \xi_{\bar{m}}^{i;F}\right) A_{\bar{m}}^i A_m^i\right] \\
=&\frac{1}{N^2}\sum_{m \in \mathscr{I}^c}\mathbb{E}\left[\left(\mathscr{E}_m^i - \xi_m^{i;F}\right)^2 (A_m^i)^2\right] \\
&+ \frac{1}{N^2}\sum_{\substack{m,\bar{m} \in \mathscr{I}^c \\ m \neq \bar{m}}}\mathbb{E}\left[\left(\mathscr{E}_m^i - \xi_m^{i;F}\right)\left(\mathscr{E}_{\bar{m}}^i - \xi_{\bar{m}}^{i;F}\right) A_{\bar{m}}^i A_m^i\right] \\
=&\frac{1}{N^2}\sum_{m \in \mathscr{I}^c}\mathbb{E}\left[\left(\mathscr{E}_m^i - \xi_m^{i;F}\right)^2 (A_m^i)^2\right] \\
&+ \frac{1}{N^2}\sum_{\substack{m,\bar{m} \in \mathscr{I}^c \\ m \neq \bar{m}}}\mathbb{E}\left[A_{\bar{m}}^i A_m^i \mathbb{E}\left[\left(\mathscr{E}_m^i - \xi_m^{i;F}\right)\left(\mathscr{E}_{\bar{m}}^i - \xi_{\bar{m}}^{i;F}\right) \mid D, \mathbf{Z}\right]\right] \\
=&\frac{1}{N^2}\sum_{m \in \mathscr{I}^c}\mathbb{E}\left[\left(\mathscr{E}_m^i - \xi_m^{i;F}\right)^2 (A_m^i)^2\right] \\
&+ \frac{1}{N^2}\sum_{\substack{m,\bar{m} \in \mathscr{I}^c \\ m \neq \bar{m}}}\mathbb{E}\left[A_{\bar{m}}^i A_m^i \mathbb{E}\left[\left(\mathscr{E}_m^i - \xi_m^{i;F}\right) \mid \mathbf{Z}\right] \mathbb{E}\left[\left(\mathscr{E}_{\bar{m}}^i - \xi_{\bar{m}}^{i;F}\right) \mid \mathbf{Z}\right]\right] \\
=&\frac{1}{N^2}\sum_{m \in \mathscr{I}^c}\mathbb{E}\left[\left(\mathscr{E}_m^i - \xi_m^{i;F}\right)^2 (A_m^i)^2\right].
\end{aligned}$$

The last equality in the above derivation follows from the fact that, conditioning on $\mathbf{Z}$, $\xi_{m,u}^{i;F}$ are i.i.d. of $\xi_m^{i;F}$ for any $u \in \{1, 2, \cdots, R\}$. Indeed, we have $\mathbb{E}\left[\xi_{m,u}^{i;F} \mid \mathbf{Z}\right] = \mathbb{E}\left[\xi_u^{i;F} \mid \mathbf{Z}\right]$ for any m and $u \in$

$\{1, 2, \cdots, R\}$. Consequently, we have

$$
\begin{aligned}
&\mathbb{E}\left[\left(\mathscr{E}_m^i - \xi_m^{i;F}\right) \mid \mathbf{Z}\right] = \mathbb{E}\left[\left(\frac{1}{R}\sum_{u=1}^{R}\xi_{m,u}^{i;F} - \xi_m^{i;F}\right) \mid \mathbf{Z}\right] \\
=&\frac{1}{R}\sum_{u=1}^{R}\mathbb{E}\left[\xi_{m,u}^{i;F} \mid \mathbf{Z}\right] - \mathbb{E}\left[\xi_m^{i;F} \mid \mathbf{Z}\right] \\
=&\frac{1}{R}\sum_{u=1}^{R}\mathbb{E}\left[\xi_{m}^{i;F} \mid \mathbf{Z}\right] - \mathbb{E}\left[\xi_m^{i;F} \mid \mathbf{Z}\right] \\
=&\mathbb{E}\left[\xi_{m}^{i;F} \mid \mathbf{Z}\right] - \mathbb{E}\left[\xi_m^{i;F} \mid \mathbf{Z}\right] = 0.
\end{aligned}
$$

In addition, we simplify the quantity $\mathbb{E}\left[\left(\mathscr{E}_m^i - \xi_m^{i;F}\right)^2\right]$. Note that $\mathscr{E}_m^i - \xi_m^{i;F} = \frac{1}{R}\sum_{u=1}^{R}[\xi_{m,u}^{i;F} - \xi_m^{i;F}]$. We therefore have

$$
\begin{aligned}
&\mathbb{E}\left[\left(\sum_{u=1}^{R}\left[\xi_{m,u}^{i;F} - \xi_m^{i;F}\right]\right)^2\right] \\
=&\sum_{u,\bar{u}=1}^{R}\mathbb{E}\left[\left(\xi_{m,u}^{i;F} - \xi_m^{i;F}\right)\left(\xi_{m,\bar{u}}^{i;F} - \xi_m^{i;F}\right)\right] \\
=&\sum_{u,\bar{u}=1}^{R}\left\{\mathbb{E}\left[\xi_{m,u}^{i;F}\xi_{m,\bar{u}}^{i;F}\right] - \mathbb{E}\left[\xi_{m,u}^{i;F}\xi_m^{i;F}\right]\right. \\
&\qquad \left. - \mathbb{E}\left[\xi_m^{i;F}\xi_{m,\bar{u}}^{i;F}\right] + \mathbb{E}\left[\xi_m^{i;F}\xi_m^{i;F}\right]\right\} \\
=&\sum_{u=1}^{R}\mathbb{E}\left[(\xi_{m,u}^{i;F})^2\right] - 2R\sum_{u=1}^{R}\mathbb{E}\left[\xi_{m,u}^{i;F}\xi_m^{i;F}\right] + R^2\mathbb{E}\left[(\xi_m^{i;F})^2\right] \\
&+ \sum_{\substack{u,\bar{u}=1 \\ u\neq\bar{u}}}^{R}\mathbb{E}\left[\xi_{m,u}^{i;F}\xi_{m,\bar{u}}^{i;F}\right] = \left[R^2 + R\right]\mathbb{E}\left[\left(\xi_m^{i;F}\right)^2\right].
\end{aligned}
$$

We justify the last equality. The last equality follows from the fact that, conditioning on $\mathbf{Z}$, $\xi_{m,u}^{i;F}$ are i.i.d. of $\xi_m^{i;F}$ and $\xi_{m,u}^{i;F}$ are i.i.d. of $\xi_{m,\bar{u}}^{i;F}$ for any u, $\bar{u} \in \{1, 2, \cdots, R\}$. Indeed, under the given fact, we have

$$
\begin{aligned}
&\mathbb{E}\left[\xi_{m,u}^{i;F}\xi_m^{i;F}\right] = \mathbb{E}\left[\mathbb{E}\left[\xi_{m,u}^{i;F}\xi_m^{i;F} \mid D, \mathbf{Z}\right]\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[\xi_{m,u}^{i;F}D, \mid \mathbf{Z}\right]\mathbb{E}\left[\xi_m^{i;F} \mid D, \mathbf{Z}\right]\right] = 0
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left[\xi_{m,u}^{i;F}\xi_{m,\bar{u}}^{i;F}\right] = \mathbb{E}\left[\mathbb{E}\left[\xi_{m,u}^{i;F}\xi_{m,\bar{u}}^{i;F} \mid D, \mathbf{Z}\right]\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[\xi_{m,u}^{i;F} \mid D, \mathbf{Z}\right]\mathbb{E}\left[\xi_{m,\bar{u}}^{i;F} \mid D, \mathbf{Z}\right]\right] = 0.
\end{aligned}
$$

Consequently, we have

$$
\mathbb{E}\left[\left(\mathscr{E}_m^i - \xi_m^{i;F}\right)^2\right] = \left(1 + \frac{1}{R}\right)\mathbb{E}\left[\left(\xi_m^{i;F}\right)^2\right].
$$

Thus, we have

$$
\begin{aligned}
\mathbb{P}\left\{\left|\kappa_N^{i;F} - \kappa_{R,N}^{i;F}\right| \geq \epsilon\right\} &\leq \frac{\frac{1}{N^2}\sum_{m\in\mathscr{I}^c}\left(1 + \frac{1}{R}\right)\mathbb{E}\left[\left(\xi_m^{i;F}\right)^2\right]}{\epsilon^2} \\
&\leq \frac{\left(1 + \frac{1}{R}\right)\mathbb{E}\left[\left(\xi^{i;F}\right)^2\right]}{N\epsilon^2}.
\end{aligned} \tag{9}
$$

We notice that no matter we set $R \to \infty$ followed by $N \to \infty$ or vice versa, or we fix R but let $N \to \infty$, we see that $\mathbb{P}\left\{\left|\kappa_N^{i;F} - \kappa_{R,N}^{i;F}\right| \geq \epsilon\right\} \to 0$. $\square$

Now, we are ready to investigate if the estimator $\hat{\theta}_N^i$ is a consistent estimator of θ^i. Our goal is to show that $\mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\theta}_N^i - \theta^i\right| \geq \epsilon\right\} \xrightarrow{p} 0$.

Proof. $\forall\epsilon > 0$, we have

$$
\begin{aligned}
&\mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\theta}_N^i - \theta^i\right| \geq \epsilon\right\} \\
=&\mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\theta}_N^i - \hat{\hat{\theta}}_N^i + \hat{\hat{\theta}}_N^i - \theta^i\right| \geq \epsilon\right\} \\
\leq&\mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\theta}_N^i - \hat{\hat{\theta}}_N^i\right| \geq \frac{\epsilon}{2}\right\} + \mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\hat{\theta}}_N^i - \theta^i\right| \geq \frac{\epsilon}{2}\right\}.
\end{aligned}
$$

Since $\left(\xi^{i;F} \overset{d}{=} \xi^{i;CF}\right) \mid \mathbf{Z}$ and $\left(\hat{\xi}^{i;F} \overset{d}{=} \hat{\xi}^{i;CF}\right) \mid \mathbf{Z}$, we have $\hat{\hat{\theta}}_N^i \overset{d}{=} \hat{\tilde{\theta}}_N^i$ by Lemma 3. Moreover, we know that $\mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\tilde{\theta}}_N^i - \theta^i\right| \geq \frac{\epsilon}{2}\right\} \xrightarrow{p} 0$ under the assumptions given in [Mackey et al., 2018]. Together with the fact that $\hat{\hat{\theta}}_N^i \overset{d}{=} \hat{\tilde{\theta}}_N^i$, we have $\mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\hat{\theta}}_N^i - \theta^i\right| \geq \frac{\epsilon}{2}\right\} \xrightarrow{p} 0$ by Lemma 2. We turn to consider the quantity $\mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\theta}_N^i - \hat{\hat{\theta}}_N^i\right| \geq \frac{\epsilon}{2}\right\}$, and we aim to show that $\mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\theta}_N^i - \hat{\hat{\theta}}_N^i\right| \geq \frac{\epsilon}{2}\right\} \xrightarrow{p} 0$. Notice that

$$
\begin{aligned}
&\mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\theta}_N^i - \hat{\hat{\theta}}_N^i\right| \geq \frac{\epsilon}{2}\right\} = \mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\kappa}_{R,N}^{i;F} - \hat{\kappa}_N^{i;F}\right| \geq \frac{\epsilon}{2}\right\} \\
\leq&\underbrace{\mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\kappa}_{R,N}^{i;F} - \kappa_{R,N}^{i;F}\right| \geq \frac{\epsilon}{8}\right\}}_{(a)} + \underbrace{\mathbb{P}_{\hat{\rho}}\left\{\left|\kappa_{R,N}^{i;F} - \kappa_N^{i;F}\right| \geq \frac{\epsilon}{8}\right\}}_{(b)} \\
&+ \underbrace{\mathbb{P}_{\hat{\rho}}\left\{\left|\kappa_N^{i;F} - \kappa_N^{i;CF}\right| \geq \frac{\epsilon}{8}\right\}}_{(c)} + \underbrace{\mathbb{P}_{\hat{\rho}}\left\{\left|\kappa_N^{i;CF} - \hat{\kappa}_N^{i;F}\right| \geq \frac{\epsilon}{8}\right\}}_{(d)}.
\end{aligned} \tag{10}
$$

Note that $\left|\kappa_{R,N}^{i;F} - \kappa_N^{i;F}\right|$ and $\left|\kappa_N^{i;F} - \kappa_N^{i;CF}\right|$ do not incorporate any terms related to the estimated function $\hat{\rho}$. From Proposition 4 and Proposition 5, we conclude that (10b) and (10c) converge to 0 in probability respectively. It remains to show the convergence of $\mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\kappa}_{R,N}^{i;F} - \kappa_{R,N}^{i;F}\right| \geq \frac{\epsilon}{8}\right\}$ and $\mathbb{P}_{\hat{\rho}}\left\{\left|\kappa_N^{i;CF} - \hat{\kappa}_N^{i;F}\right| \geq \frac{\epsilon}{8}\right\}$.

Consider (10d) first. Since

$$\begin{aligned}
&\left|\kappa_N^{i;CF} - \hat{\kappa}_N^{i;F}\right| = \left|\frac{1}{N}\sum_{m\in\mathscr{I}^c}\xi_m^{i;CF}A_m^i - \frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}_m^{i;F}\hat{A}_m^i\right| \\
\leq &\underbrace{\left|\frac{1}{N}\sum_{m\in\mathscr{I}^c}(\xi_m^{i;CF}A_m^i - \xi_m^{i;F}A_m^i)\right|}_{\Gamma_1} \\
&+\underbrace{\left|\frac{1}{N}\sum_{m\in\mathscr{I}^c}(\xi_m^{i;F}A_m^i - \hat{\xi}_m^{i;F}A_m^i)\right|}_{\Gamma_2} + \underbrace{\left|\frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}_m^{i;F}(A_m^i - \hat{A}_m^i)\right|}_{\Gamma_3},
\end{aligned}$$

(10d) is bounded above by

$$\underbrace{\mathbb{P}_{\hat{\rho}}\left\{\Gamma_1 \geq \frac{\epsilon}{24}\right\}}_{(a)} + \underbrace{\mathbb{P}_{\hat{\rho}}\left\{\Gamma_2 \geq \frac{\epsilon}{24}\right\}}_{(b)} + \underbrace{\mathbb{P}_{\hat{\rho}}\left\{\Gamma_3 \geq \frac{\epsilon}{24}\right\}}_{(c)}. \tag{11}$$

(11a) converges to 0 in probability due to Proposition 4. We study the quantities (11b) and (11c).

(11b) can be further bounded. If N^c is the size of $\mathfrak{D}_i^c \cap \mathscr{I}$, then we have

$$\begin{aligned}
\Gamma_2 = &\left|\frac{1}{N}\sum_{m\in\mathscr{I}^c}(\xi_m^{i;F}A_m^i - \hat{\xi}_m^{i;F}A_m^i)\right| \\
\leq &\underbrace{\left|\frac{1}{N}\sum_{m\in\mathscr{I}^c}\xi_m^{i;F}A_m^i - \frac{N^c}{N}\mathbb{E}_{\hat{\rho}}\left[\xi^{i;F}A^i\right]\right|}_{\Gamma_{2;1}} \\
&+\underbrace{\left|\frac{N^c}{N}\mathbb{E}_{\hat{\rho}}\left[\xi^{i;F}A^i\right] - \frac{N^c}{N}\mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}A^i\right]\right|}_{\Gamma_{2;2}} \\
&+\underbrace{\left|\frac{N^c}{N}\mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}A^i\right] - \frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}_m^{i;F}A_m^i\right|}_{\Gamma_{2;3}},
\end{aligned}$$

we see that (11b) can be further bounded by

$$\underbrace{\mathbb{P}_{\hat{\rho}}\left\{\Gamma_{2;1} \geq \frac{\epsilon}{72}\right\}}_{(a)} + \underbrace{\mathbb{P}_{\hat{\rho}}\left\{\Gamma_{2;2} \geq \frac{\epsilon}{72}\right\}}_{(b)} + \underbrace{\mathbb{P}_{\hat{\rho}}\left\{\Gamma_{2;3} \geq \frac{\epsilon}{72}\right\}}_{(c)}. \tag{12}$$

We investigate if (12a), (12b), and (12c) converge to 0 in probability. We consider (12a) first. Recall the assumptions that $(\xi^{i;F} \perp\!\!\!\perp D) \mid \mathbf{Z}$, $\left(\xi^{i;F} \overset{d}{=} \xi^{i;CF}\right) \mid \mathbf{Z}$ and $(\xi^{i;CF} \perp\!\!\!\perp D) \mid \mathbf{Z}$, we have $\frac{1}{N}\sum_{m\in\mathscr{I}^c}\xi_m^{i;F}A_m^i \overset{d}{=} \frac{1}{N}\sum_{m\in\mathscr{I}^c}\xi_m^{i;CF}A_m^i$ by Lemma 3. Since $\Gamma_{2;1} = \frac{N^c}{N}\left|\frac{1}{N^c}\sum_{m\in\mathscr{I}^c}\xi_m^{i;F}A_m^i - \mathbb{E}_{\hat{\rho}}\left[\xi^{i;F}A^i\right]\right|$, we have

$$\begin{aligned}
&\mathbb{P}_{\hat{\rho}}\left\{\Gamma_{2;1} \geq \frac{\epsilon}{72}\right\} \\
=&\mathbb{P}_{\hat{\rho}}\left\{\frac{N^c}{N}\left|\frac{1}{N^c}\sum_{m\in\mathscr{I}^c}\xi_m^{i;F}A_m^i - \mathbb{E}_{\hat{\rho}}\left[\xi^{i;F}A^i\right]\right| \geq \frac{\epsilon}{72}\right\} \\
=&\mathbb{P}_{\hat{\rho}}\left\{\left|\frac{1}{N^c}\sum_{m\in\mathscr{I}^c}\xi_m^{i;F}A_m^i - \mathbb{E}_{\hat{\rho}}\left[\xi^{i;F}A^i\right]\right| \geq \frac{\epsilon}{72}\cdot\frac{N}{N^c}\right\} \\
\leq&\mathbb{P}_{\hat{\rho}}\left\{\left|\frac{1}{N^c}\sum_{m\in\mathscr{I}^c}\xi_m^{i;F}A_m^i - \mathbb{E}_{\hat{\rho}}\left[\xi^{i;F}A^i\right]\right| \geq \frac{\epsilon}{72}\right\} \\
\leq&\frac{\mathbb{E}_{\hat{\rho}}\left[\left|\frac{1}{N^c}\sum_{m\in\mathscr{I}^c}\xi_m^{i;F}A_m^i - \mathbb{E}_{\hat{\rho}}\left[\xi^{i;F}A^i\right]\right|^2\right]}{\left(\frac{\epsilon}{72}\right)^2}.
\end{aligned}$$

Consider $\mathbb{E}_{\hat{\rho}}\left[\left|\frac{1}{N^c}\sum_{m\in\mathscr{I}^c}\xi_m^{i;F}A_m^i - \mathbb{E}_{\hat{\rho}}\left[\xi^{i;F}A^i\right]\right|^2\right]$. Note that

$$\begin{aligned}
&\mathbb{E}_{\hat{\rho}}\left[\left|\frac{1}{N^c}\sum_{m\in\mathscr{I}^c}\xi_m^{i;F}A_m^i - \mathbb{E}_{\hat{\rho}}\left[\xi^{i;F}A^i\right]\right|^2\right] \\
=&\frac{1}{(N^c)^2}\sum_{m\in\mathscr{I}^c}\mathbb{E}_{\hat{\rho}}\left[\left|\xi_m^{i;F}A_m^i - \mathbb{E}_{\hat{\rho}}\left[\xi^{i;F}A^i\right]\right|^2\right] \\
&+\frac{1}{(N^c)^2}\sum_{\substack{m,\bar{m}\in\mathscr{I}^c\\ m\neq\bar{m}}}\mathbb{E}_{\hat{\rho}}\left[(\xi_m^{i;F}A_m^i - \mathbb{E}_{\hat{\rho}}[\xi^{i;F}A^i])(\xi_{\bar{m}}^{i;F}A_{\bar{m}}^i - \mathbb{E}_{\hat{\rho}}[\xi^{i;F}A^i])\right] \\
=&\frac{1}{(N^c)^2}\sum_{m\in\mathscr{I}^c}\mathbb{E}_{\hat{\rho}}\left[\left(A_m^i\right)^2\mathbb{E}_{\hat{\rho}}\left[\left(\xi_m^{i;F}\right)^2 \mid D, \mathbf{Z}\right]\right] \\
&+\frac{1}{(N^c)^2}\sum_{\substack{m,\bar{m}\in\mathscr{I}^c\\ m\neq\bar{m}}}\mathbb{E}_{\hat{\rho}}\left[A_m^iA_{\bar{m}}^i\mathbb{E}_{\hat{\rho}}\left[\xi_m^{i;F} \mid D, \mathbf{Z}\right]\mathbb{E}_{\hat{\rho}}\left[\xi_{\bar{m}}^{i;F} \mid D, \mathbf{Z}\right]\right] \\
=&\frac{1}{(N^c)^2}\sum_{m\in\mathscr{I}^c}\mathbb{E}_{\hat{\rho}}\left[\left(A_m^i\right)^2\mathbb{E}_{\hat{\rho}}\left[\left(\xi_m^{i;F}\right)^2 \mid \mathbf{Z}\right]\right] \\
=&\frac{1}{N^c}\mathbb{E}_{\hat{\rho}}\left[\left(A^i\right)^2\mathbb{E}_{\hat{\rho}}\left[\left(\xi^{i;F}\right)^2 \mid \mathbf{Z}\right]\right].
\end{aligned}$$

Since A^i and $\xi^{i;F}$ do not include the estimated nuisance parameters, $\mathbb{E}_{\hat{\rho}}\left[\left(A^i\right)^2\mathbb{E}_{\hat{\rho}}\left[\left(\xi^{i;F}\right)^2 \mid \mathbf{Z}\right]\right]$ is a constant. Moreover, note that $N^c \to \infty$ when $N \to \infty$, we have

$$\mathbb{P}_{\hat{\rho}}\left\{\Gamma_{2;1} \geq \frac{\epsilon}{72}\right\} \leq \frac{72^2\ \mathbb{E}_{\hat{\rho}}\left[\left(A^i\right)^2\mathbb{E}_{\hat{\rho}}\left[\left(\xi^{i;F}\right)^2 \mid \mathbf{Z}\right]\right]}{\epsilon^2 N^c} \xrightarrow{p} 0.$$

Now, we consider (12b). Indeed, we have

$$
\begin{aligned}
&\mathbb{P}_{\hat{\rho}}\left\{\Gamma_{2;2} \geq \frac{\epsilon}{72}\right\} \\
=&\mathbb{P}_{\hat{\rho}}\left\{\left|\mathbb{E}_{\hat{\rho}}\left[\xi^{i;F}A^i\right] - \mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}A^i\right]\right| \geq \frac{\epsilon}{72}\cdot\frac{N}{N^c}\right\} \\
\leq&\mathbb{P}_{\hat{\rho}}\left\{\left|\mathbb{E}_{\hat{\rho}}\left[(\xi^{i;F}-\hat{\xi}^{i;F})A^i\right]\right| \geq \frac{\epsilon}{72}\right\} \\
\leq&\frac{72^2\left\{\mathbb{E}_{\hat{\rho}}\left[(\xi^{i;F}-\hat{\xi}^{i;F})A^i\right]\right\}^2}{\epsilon^2} \\
\leq&\frac{72^2\ \left\{\mathbb{E}_{\hat{\rho}}\left[(\xi^{i;F}-\hat{\xi}^{i;F})^{4q}\right]\right\}^{\frac{1}{2q}}\left\{\mathbb{E}_{\hat{\rho}}\left[(A^i)^{\frac{4q}{4q-1}}\right]\right\}^{2-\frac{1}{2q}}}{\epsilon^2} \xrightarrow{p} 0.
\end{aligned}
$$

Here, the last inequality follows from the Hölders inequality, while the convergence holds $\forall q \in \{1,2,\ldots,k\}$ according to Assumption 1.5 of [Mackey et al., 2018]. Finally, we consider (12c). We can rewrite $\Gamma_{2;3}$ as

$$
\Gamma_{2;3} = \frac{N^c}{N}\left|\mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}A^i\right] - \frac{1}{N^c}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;F}_m A^i_m\right|.
$$

Now, we have

$$
\begin{aligned}
&\mathbb{P}_{\hat{\rho}}\left\{\Gamma_{2;3} \geq \frac{\epsilon}{72}\right\} \\
\leq&\mathbb{P}_{\hat{\rho}}\left\{\left|\mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}A^i\right] - \frac{1}{N^c}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;F}_m A^i_m\right| \geq \frac{\epsilon}{72}\right\} \\
\leq&\frac{72^2\ \mathbb{E}_{\hat{\rho}}\left[\left\{\sum_{m\in\mathscr{I}^c}\left(\mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}A^i\right] - \hat{\xi}^{i;F}_m A^i_m\right)\right\}^2\right]}{\epsilon^2\left(N^c\right)^2} \\
=&\frac{72^2\sum_{m\in\mathscr{I}^c}\mathbb{E}_{\hat{\rho}}\left[\left(\mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}A^i\right] - \hat{\xi}^{i;F}_m A^i_m\right)^2\right]}{\epsilon^2\left(N^c\right)^2} \\
&+\frac{72^2\sum_{\substack{m,\bar{m}\in\mathscr{I}^c\\ m\neq\bar{m}}}\mathbb{E}_{\hat{\rho}}[(\hat{\xi}^{i;F}_m-\xi^{i;F}_m)A^i_m]\mathbb{E}_{\hat{\rho}}[(\hat{\xi}^{i;F}_{\bar{m}}-\xi^{i;F}_{\bar{m}})A^i_{\bar{m}}]}{\epsilon^2\left(N^c\right)^2} \\
&-2\frac{72^2\sum_{\substack{m,\bar{m}\in\mathscr{I}^c\\ m<\bar{m}}}\mathbb{E}_{\hat{\rho}}[(\hat{\xi}^{i;F}_m-\xi^{i;F}_m)A^i_m]\mathbb{E}_{\hat{\rho}}[(\hat{\xi}^{i;F}-\xi^{i;F})A^i]}{\epsilon^2\left(N^c\right)^2} \\
&+\frac{72^2\sum_{\substack{m,\bar{m}\in\mathscr{I}^c\\ m\neq\bar{m}}}\left\{\mathbb{E}_{\hat{\rho}}\left[\left(\hat{\xi}^{i;F}-\xi^{i;F}\right)A^i\right]\right\}^2}{\epsilon^2\left(N^c\right)^2}.
\end{aligned}
$$

Using Assumption 1.5 of [Mackey et al., 2018], we can conclude that $\mathbb{P}_{\hat{\rho}}\left\{\Gamma_{2;3} \geq \frac{\epsilon}{72}\right\} \xrightarrow{p} 0$. Next, we come to bound (11c). Again, we denote N^c as the size of $\mathfrak{D}_i^c \cap \mathscr{I}$.

Since

$$
\begin{aligned}
\Gamma_3 =& \left|\frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;F}_m A^i_m - \frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;F}_m \hat{A}^i_m\right| \\
\leq&\left|\frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;F}_m A^i_m - \frac{N^c}{N}\mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}A^i\right]\right| \\
&+\left|\frac{N^c}{N}\mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}A^i\right] - \frac{N^c}{N}\mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}\hat{A}^i\right]\right| \\
&+\left|\frac{N^c}{N}\mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}\hat{A}^i\right] - \frac{1}{N}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;F}_m \hat{A}^i_m\right| \\
=&\frac{N^c}{N}\underbrace{\left|\frac{1}{N^c}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;F}_m A^i_m - \mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}A^i\right]\right|}_{\Gamma_{3;1}} \\
&+\frac{N^c}{N}\underbrace{\left|\mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}A^i\right] - \mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}\hat{A}^i\right]\right|}_{\Gamma_{3;2}} \\
&+\frac{N^c}{N}\underbrace{\left|\mathbb{E}_{\hat{\rho}}\left[\hat{\xi}^{i;F}\hat{A}^i\right] - \frac{1}{N^c}\sum_{m\in\mathscr{I}^c}\hat{\xi}^{i;F}_m \hat{A}^i_m\right|}_{\Gamma_{3;3}},
\end{aligned}
$$

we see that (11c) can be further bounded by

$$
\underbrace{\mathbb{P}_{\hat{\rho}}\left\{\Gamma_{3;1} \geq \frac{\epsilon}{72}\right\}}_{(a)} + \underbrace{\mathbb{P}_{\hat{\rho}}\left\{\Gamma_{3;2} \geq \frac{\epsilon}{72}\right\}}_{(b)} + \underbrace{\mathbb{P}_{\hat{\rho}}\left\{\Gamma_{3;3} \geq \frac{\epsilon}{72}\right\}}_{(c)}. \tag{13}
$$

Similarly, we can prove that (13a) and (13c) converge to 0 in probability when $N \to \infty$ using the arguments in proving that (12a) and (12c) converge to 0. As a result, the quantity (10d) converges to 0 in probability when $N \to \infty$.

Lastly, we turn to consider the quantity (10a). In fact, we have

$$
\begin{aligned}
&\mathbb{P}_{\hat{\rho}}\left\{\left|\hat{\kappa}^{i;F}_{R,N} - \kappa^{i;F}_{R,N}\right| \geq \frac{\epsilon}{8}\right\} \\
\leq&\mathbb{P}_{\hat{\rho}}\left\{\left|\frac{1}{N}\sum_{m\in\mathscr{I}^c}\frac{1}{R}\sum_{u=1}^{R}\left(\hat{\xi}^{i;F}_{m,u}\right)(\hat{A}^i_m - A^i_m)\right| \geq \frac{\epsilon}{16}\right\} \quad (14a) \\
&+\mathbb{P}_{\hat{\rho}}\left\{\left|\frac{1}{N}\sum_{m\in\mathscr{I}^c}A^i_m\frac{1}{R}\sum_{u=1}^{R}\left[(\hat{\xi}^{i;F}_{m,u} - \xi^{i;F}_{m,u})\right]\right| \geq \frac{\epsilon}{16}\right\}. \quad (14b)
\end{aligned}
$$

We can argue that (14a) converges to 0 in probability as $N \to \infty$ using similar arguments when we prove that (11b) converges to 0 in probability. Simultaneously, we can argue (14b) converges to 0 in probability as $N \to \infty$ using similar arguments when we prove that (11c) converges to 0 in probability. Consequently, we have $\hat{\kappa}^{i;F}_{R,N} - \kappa^{i;F}_{R,N}$ converges to 0 in probability.

The proof is completed. □

References

[Mackey et al., 2018] Mackey, L., Syrgkanis, V., and Zadik, I. (2018). Orthogonal machine learning: Power and limitations. In *International Conference on Machine Learning*, pages 3375–3383. PMLR.