ONLINE APPENDIX

NENG-CHIEH CHANG

S1. MONTE CARLO

ML estimation (repeated cross sections): Let N be the sample size and p the dimension of control variables, $X_i \sim N(0, I_{p\times p})$. Also, let $\gamma_0 = (1, 1/2, 1/3, 1/4, 1/5, 0, ..., 0) \in \mathbb{R}^p$. D is generated by the propensity score

$$P(D=1 \mid X) = \frac{1}{1+\exp(-X'\gamma_0)} \text{(Logistic)}.$$

The potential outcomes are generated by $Y_i^0(0) = 1 + \varepsilon_1, Y_i^0(1) = Y_i^0(0) + 1 + \varepsilon_2, Y_i^1(1) = \theta_0 + Y_i^0(1) + \varepsilon_3$, where $\beta_0 = \gamma_0 + 0.5$ and $\theta_0 = 3$, and all error terms follow $N(0, 0.1)$. Define $Y_i(0) = Y_i^0(0)$ and $Y_i(1) = Y_i^0(1)(1-D_i) + Y_i^1(1)D_i$. Let T_i follow a Bernoulli distribution with parameter 0.5. Researchers observe $\{Y_i, T_i, D_i, X_i\}$ for $i = 1, ..., N$, where $Y_i = Y_i(0) + T_i(Y_i(1) - Y_i(0))$.

ML estimation (multilevel treatments): Suppose there are two levels of treatment so that $W \in \{0, 1, 2\}$. Let N be the sample size and p the dimension of control variables, $X_i \sim N(0, I_{p\times p})$. Also, let $\gamma_0 \in\in \mathbb{R}^p$, such that $\gamma_0 = (1, 1/2, 1/3, 1/4, 1/5, 0, ..., 0)$ and

$$(P(W=0), P(W=1), P(W=2)) = (0.3, 0.3, 0.4)$$

.The potential outcomes are generated by $Y_i^0(0) = X'\beta_0 + \varepsilon_1, Y_i^0(1) = Y_i^0(0) + 1 + \varepsilon_2, Y_i^1(1) = \theta_{10} + Y_i^0(1) + \varepsilon_3, Y_i^2(1) = \theta_{20} + Y_i^0(1) + \varepsilon_4$, where $\beta_0 = \gamma_0 + 0.5$ and $\theta_{10} = 3$ and $\theta_{20} = 6$, and all error terms follow $N(0, 0.1)$. Researchers observe $\{Y_i(0), Y_i(1), W_i, X_i\}$ for $i = 1, ..., N$, where $Y_i(0) = Y_i^0(0)$ and $Y_i(1) = Y_i^0(1)I(W_i=0) + Y_i^1(1)I(W_i=1) + Y_i^2(1)I(W_i=2)$. I focus on the estimation of the second-level ATT θ_{20}.

S2. PROOFS OF THE NEYMAN-ORTHOGONAL SCORES

Proof of Lemma 3.1. ***Repeated outcomes:***
The Gateaux derivative of (3.1) in the direction $\eta_1 - \eta_{10} = (g - g_0, \ell_1 - \ell_{10})$ is

$$\begin{aligned}
\partial_{\eta_1} E_P[\psi_1(W, \theta_0, p_0, \eta_{10})] =& E_P\left[\frac{(D-1)(Y(1) - Y(0) - \ell_{10}(X))}{p_0(1-g_0(X))^2}(g(X) - g_0(X))\right] \\
&- E_P\left[\frac{D - g_0(X)}{p_0(1-g_0(X))}(\ell_1(X) - \ell_{10}(X))\right] \\
=& - E_P\left[\frac{g(X) - g_0(X)}{p_0(1-g_0(X))}E[Y(1) - Y(0) - \ell_{10}(X) \mid X, D=0]\right] \\
&- E_P\left[\frac{(\ell_1(X) - \ell_{10}(X))}{p_0(1-g_0(X))}E_P[D - g_0(X) \mid X]\right] \\
=& - E_P\left[\frac{g(X) - g_0(X)}{p_0(1-g_0(X))}(\ell_{10}(X) - \ell_{10}(X))\right] - 0 \\
=& 0,
\end{aligned}$$

where the second inequality follows from the law of iterated expectations and the third from the definition of $\ell_{10}(X)$ and $E_P[D - g_0(X)|X] = 0$.

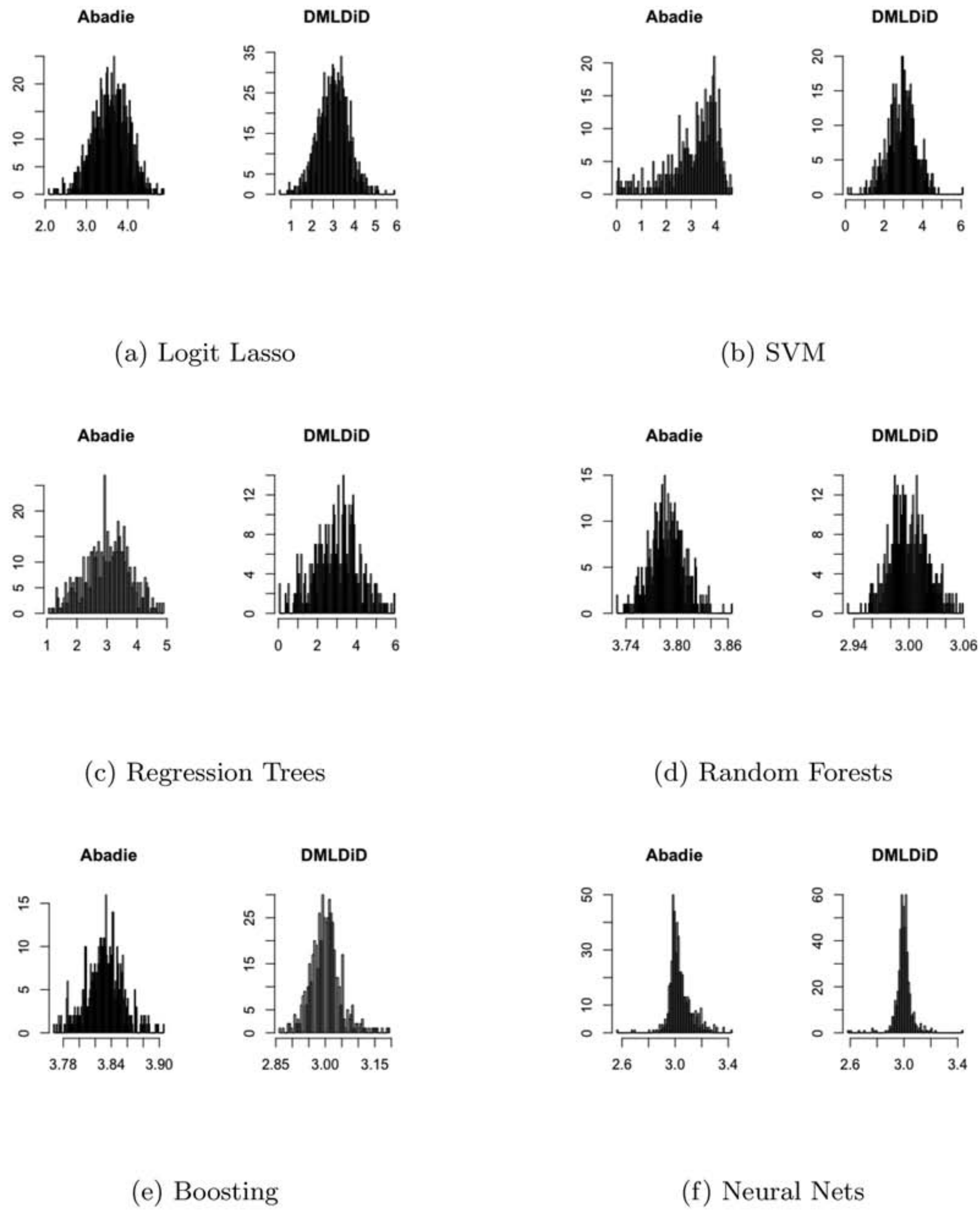

Figure S1. Repeated cross sections.

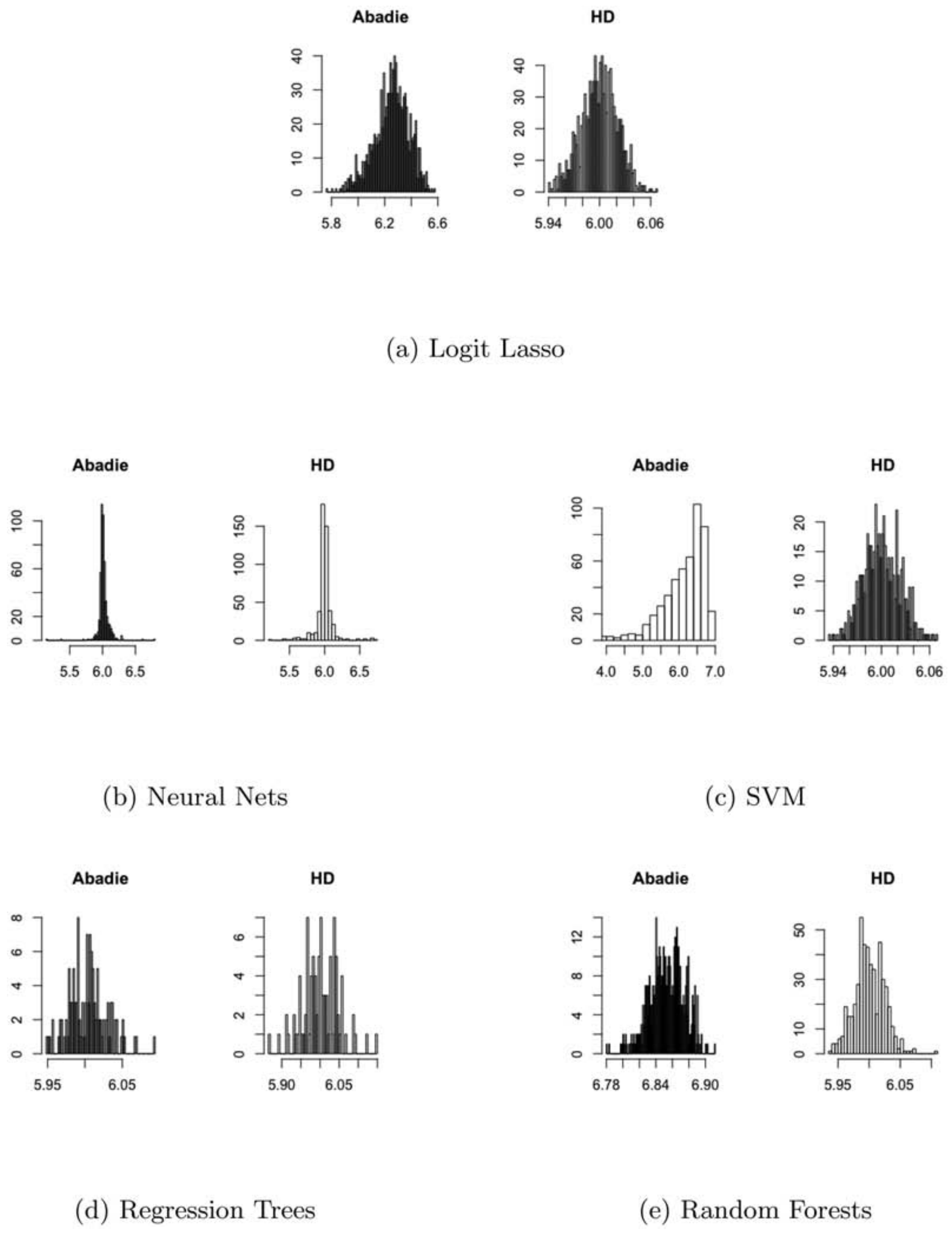

Figure S2. Multilevel treatment.

Repeated cross sections:

Similar to the proof of repeated outcomes, the Gateaux derivative of (3.2) in the direction $\eta_2 - \eta_{20} = (g - g_0, \ell_2 - \ell_{20})$ is

$$\begin{aligned}\partial_{\eta_2} E_P\left[\psi_2\left(W, \theta_0, p_0, \lambda_0, \eta_{20}\right)\right] =& E_P\left[\frac{(D-1)\left(\left(T-\lambda_0\right) Y-\ell_{20}(X)\right)}{p_0^{\prime}\left(1-g_0(X)\right)^2}\left(g(X)-g_0(X)\right)\right] \\ & -E_P\left[\frac{D-g_0(X)}{p_0^{\prime}\left(1-g_0(X)\right)}\left(\ell_2(X)-\ell_{20}(X)\right)\right] \\ =& -E_P\left[\frac{g(X)-g_0(X)}{p_0^{\prime}\left(1-g_0(X)\right)}\left(\ell_{20}(X)-\ell_{20}(X)\right)\right] \\ & -E_P\left[\frac{\ell_2(X)-\ell_{20}(X)}{p \lambda(1-\lambda)(1-g(X))} E_P\left[D-g_0(X) \mid X\right]\right] \\ =& 0,\end{aligned}$$

where $p_0^{\prime} \equiv p_0 \lambda_0\left(1-\lambda_0\right)$.

Multilevel treatment:

Let $\Delta_w = g_w - g_{w0}$, $\Delta_z = g_z - g_{z0}$, and $\Delta_{\ell 3} = \ell_3 - \ell_{30}$. The Gateaux derivative of (3.3) in the direction $\eta_w - \eta_{w0} = (g_w - g_{w0}, g_z - g_{z0}, \ell_3 - \ell_{30})$ is

$$\begin{aligned}\partial_{\eta_w} E_P\left[\psi_w\left(W, \theta_0, p_{w0}, \eta_{w0}\right)\right] =& E_P\left[\frac{I(W=0) g_{w0}(X)}{p_{w0} g_{z0}(X)^2}\left(Y(1)-Y(0)-\ell_{30}\right) \Delta_w\right] \\ & -E_P\left[\frac{I(W=0)}{p_{w0} g_{z0}(X)}\left(Y(1)-Y(0)-\ell_{30}\right) \Delta_z\right] \\ & +E_P\left[\frac{I(W=0) g_{w0}(X)-I(W=w) g_{z0}(X)}{p_{w0} g_{z0}(X)} \Delta_{\ell 3}\right] \\ =& 0\end{aligned}$$

by the law of iterated expectation on each term. □

S3. ADDITIONAL PROOFS

Proof of Theorem 3.1. The proof proceeds in five steps. In Step 1, I show the main result using the auxiliary results (A.1) through (A.4). In Steps 2 through 5, I prove the auxiliary results.

$$\sup_{\eta_1 \in \mathcal{T}_N}\left(E\left[\| \psi_1\left(W, \theta_0, p_0, \eta_1\right)-\psi_1\left(W, \theta_0, p_0, \eta_{10}\right) \|^2\right]\right)^{1/2} \leq \varepsilon_N, \tag{A.1}$$

$$\sup_{r \in(0,1), \eta_1 \in \mathcal{T}_N} \| \partial_r^2 E\left[\psi_1\left(W, \theta_0, p_0, \eta_{10}+r\left(\eta_1-\eta_{10}\right)\right)\right] \| \leq\left(\varepsilon_N\right)^2, \tag{A.2}$$

$$\sup_{\eta_1 \in \mathcal{T}_N}\left(E_P\left[\| \partial_p \psi_1\left(W, \theta_0, p_0, \eta_1\right)-\partial_p \psi_1\left(W, \theta_0, p_0, \eta_{10}\right) \|^2\right]\right)^{1/2} \leq \varepsilon_N, \tag{A.3}$$

$$\sup_{p \in \mathcal{P}_N, \eta_1 \in \mathcal{T}_N}\left(E_P\left[\| \partial_p^2 \psi_1\left(W, \theta_0, p, \eta_1\right)-\partial_p^2 \psi_1\left(W, \theta_0, p_0, \eta_{10}\right) \|^2\right]\right)^{1/2} \leq \varepsilon_N, \tag{A.4}$$

where $\mathcal{T}_N$ is the set of all $\eta_1 = (g, \ell_1)$ consisting of P-square-integrable functions g and ℓ_1, such that

$$\| \eta_1-\eta_{10} \|_{P,2} \leq \varepsilon_N,$$

$$\parallel g - 1/2 \parallel_{P,\infty} \leq 1/2 - \kappa,$$

$$\parallel g - g_0 \parallel_{P,2}^2 + \parallel g - g_0 \parallel_{P,2} \times \parallel \ell_1 - \ell_{10} \parallel_{P,2} \leq (\varepsilon_N)^2,$$

and $\mathcal{P}_N$ is the set of all $p > 0$, such that $|p - p_0| \leq N^{-1/2}$. Then, by Assumption (3.1) and $|\hat{p}_k - p_0| = O_P\left(N^{-1/2}\right)$, we have $\hat{\eta}_{1k} \in \mathcal{T}_N$ and $\hat{p}_k \in \mathcal{P}_N$ with probability $1 - o(1)$.

Step 1. Observe that we have the decomposition

$$\begin{aligned}
\sqrt{N}\left(\tilde{\theta} - \theta_0\right) =& \sqrt{N}\left(\frac{1}{K}\sum_{k=1}^{K}\tilde{\theta}_k - \theta_0\right) \\
=& \sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\psi_1\left(W, \theta_0, \hat{p}_k, \hat{\eta}_{1k}\right)\right] \\
=& \sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\psi_1\left(W, \theta_0, p_0, \hat{\eta}_{1k}\right)\right] \\
&+ \underbrace{\sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\partial_p\psi_1\left(W, \theta_0, p_0, \hat{\eta}_{1k}\right)\right]\left(\hat{p}_k - p_0\right)}_{a} \\
&+ \underbrace{\sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\partial_p^2\psi_1\left(W, \theta_0, \bar{p}_k, \hat{\eta}_{1k}\right)\right]\left(\hat{p}_k - p_0\right)^2}_{b},
\end{aligned}$$

where $\bar{p}_k \in (\hat{p}_k, p_0)$. By the triangle inequality, the expectation in term (a) satisfies

$$|\mathbb{E}_{n,k}\left[\partial_p\psi_1\left(W, \theta_0, p_0, \hat{\eta}_{1k}\right)\right] - E_P\left[\partial_p\psi_1\left(W, \theta_0, p_0, \eta_{10}\right)\right]| \leq J_{1,k} + J_{2,k},$$

where

$$J_{1,k} = |\mathbb{E}_{n,k}\left[\partial_p\psi_1\left(W, \theta_0, p_0, \hat{\eta}_{1k}\right)\right] - \mathbb{E}_{n,k}\left[\partial_p\psi_1\left(W, \theta_0, p_0, \eta_{10}\right)\right]|,$$

$$J_{2,k} = |\mathbb{E}_{n,k}\left[\partial_p\psi_1\left(W, \theta_0, p_0, \eta_{10}\right)\right] - E_p\left[\partial_p\psi_1\left(W, \theta_0, p_0, \eta_{10}\right)\right]|.$$

The goal is to show that $J_{1,k} = o_p(1)$ and $J_{2,k} = o_p(1)$. To bound $J_{2,k}$, we have $E_P[J_{2,k}] = 0$ and

$$\begin{aligned}
E_P\left[J_{2,k}^2\right] \leq& n^{-1}E_P\left[\left(\partial_p\psi_1\left(W, \theta_0, p_0, \eta_{10}\right)^2\right)\right] \\
=& n^{-1}E_P\left[\frac{1}{p_0^4}\frac{U^2V_1^2}{(1-g_0)^2}\right] \\
\leq& n^{-1}\left(\frac{C^2}{p_0^4\kappa^2}\right),
\end{aligned}$$

where the last inequality follows from Assumption (3.1). By Chebyshev's inequality, $J_{2,k} = O_P(n^{-1/2}) = o_P(1)$. Next, we bound $J_{1,k}$. Conditional on the auxiliary sample I_k^c, $\hat{\eta}_{1k}$ can be treated as fixed. Under the event that $\hat{\eta}_{1k} \in \mathcal{T}_N$, we have

$$\begin{aligned} E_P\left[J_{1,k}^2 \mid (W_i)_{i\in I_k^c}\right] =& E_P\left[\| \partial_p \psi_1 (W, \theta_0, p_0, \hat{\eta}_{1k}) - \partial_p \psi_1 (W, \theta_0, p_0, \eta_{10}) \|^2 | (W_i)_{i\in I_k^c}\right] \\ \leq& \sup_{\eta_1 \in \mathcal{T}_N} E_P\left[\| \partial_p \psi_1 (W, \theta_0, p_0, \eta_1) - \partial_p \psi_1 (W, \theta_0, p_0, \eta_{10}) \|^2\right] \\ =& \varepsilon_N^2 \end{aligned}$$

by (A.3). Because conditional convergence implies unconditional convergence (Lemma A.1), $J_{1,k} = O_P(\varepsilon_N) = o_P(1)$. Together, we have

$$\mathbb{E}_{n,k}\left[\partial_p \psi_1 (W, \theta_0, p_0, \hat{\eta}_{1k})\right] \xrightarrow{p} E_p\left[\partial_p \psi_1 (W, \theta_0, p_0, \eta_{10})\right] = G_{1p0}.$$

By the triangle inequality again, the expectation in term (b) satisfies

$$|\mathbb{E}_{n,k}\left[\partial_p^2 \psi_1 (W, \theta_0, \bar{p}_k, \hat{\eta}_{1k})\right] - E_p\left[\partial_p^2 \psi_1 (W, \theta_0, p_0, \eta_{10})\right]| \leq J_{3,k} + J_{4,k},$$

where

$$\begin{aligned} J_{3,k} &= |\mathbb{E}_{n,k}\left[\partial_p^2 \psi_1 (W, \theta_0, \bar{p}_k, \hat{\eta}_{1k})\right] - \mathbb{E}_{n,k}\left[\partial_p^2 \psi_1 (W, \theta_0, p_0, \eta_{10})\right]|, \\ J_{4,k} &= |\mathbb{E}_{n,k}\left[\partial_p^2 \psi_1 (W, \theta_0, p_0, \eta_{10})\right] - E_p\left[\partial_p^2 \psi_1 (W, \theta_0, p_0, \eta_{10})\right]|. \end{aligned}$$

To bound $J_{4,k}$, we have

$$\begin{aligned} E_P\left[J_{4,k}^2\right] \leq& n^{-1} E_P\left[\left(\partial_p^2 \psi_1 (W, \theta_0, p_0, \eta_{10})^2\right)\right] \\ =& n^{-1} E_P\left[\frac{4}{p_0^6} \frac{U^2 V_1^2}{(1-g_0)^2}\right] \\ \leq& n^{-1}\left(\frac{4C^2}{p_0^6 \kappa^2}\right), \end{aligned}$$

where the last inequality follows from the regularity conditions. By Chebyshev's inequality, $J_{4,k} = O_P(n^{-1/2}) = o_P(1)$. Conditional on I_k^c, both $\bar{p}_k$ and $\hat{\eta}_{1k}$ can be treated as fixed. Under the event that $\hat{p}_k \in \mathcal{P}_N$ (thus $\bar{p}_k \in \mathcal{P}_N$) and $\hat{\eta}_{1k} \in \mathcal{T}_N$, we have

$$\begin{aligned} E_P\left[J_{3,k}^2 \mid (W_i)_{i\in I_k^c}\right] =& E_P\left[\| \partial_p^2 \psi_1 (W, \theta_0, \bar{p}_k, \hat{\eta}_{1k}) - \partial_p^2 \psi_1 (W, \theta_0, p_0, \eta_{10}) \|^2 | (W_i)_{i\in I_k^c}\right] \\ \leq& \sup_{p\in\mathcal{P}_N, \eta_1 \in \mathcal{T}_N} E_P\left[\| \partial_p \psi_1 (W, \theta_0, p, \eta_1) - \partial_p \psi_1 (W, \theta_0, p_0, \eta_{10}) \|^2\right] \\ \leq& \varepsilon_N^2 \end{aligned}$$

by (A.4). By Lemma A.1 again, $J_{3,k} = O_P(\varepsilon_N) = o_P(1)$. Hence,

$$\mathbb{E}_{n,k}\left[\partial_p^2 \psi_1 (W, \theta_0, \bar{p}_k, \hat{\eta}_{1k})\right] \xrightarrow{p} E_P\left[\partial_p^2 \psi_1 (W, \theta_0, \bar{p}_k, \hat{\eta}_{1k})\right].$$

Combine the above results with $\hat{p}_k - p_0 = \mathbb{E}_{n,k}[D - p_0]$ and $(\hat{p}_k - p_0)^2 = O_P(N^{-1})$, and the decomposition of $\tilde{\theta}$ becomes

$$
\begin{aligned}
\sqrt{N}\left(\tilde{\theta}-\theta_0\right) &= \sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\psi_1\left(W,\theta_0,p_0,\hat{\eta}_{1k}\right)\right] \\
&+ \left[\sqrt{N}\frac{1}{K}\sum_{k=1}^{K}G_{1p0}\mathbb{E}_{n,k}\left[\left(D-p_0\right)\right]+o_p\left(1\right)\right]+O_p\left(N^{-1/2}\right) \\
&= \sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\psi_1\left(W,\theta_0,p_0,\hat{\eta}_{1k}\right)+G_{1p0}\left(D-p_0\right)\right]+o_P\left(1\right) \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\psi_1\left(W_i,\theta_0,p_0,\eta_{10}\right)+G_{1p0}\left(D_i-p_0\right)\right]+\sqrt{N}R_N+o_P\left(1\right),
\end{aligned}
$$

where

$$
\begin{aligned}
R_N =& \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\psi_1\left(W,\theta_0,p_0,\hat{\eta}_{1k}\right)+G_{1p0}\left(D-p_0\right)\right] \\
&- \frac{1}{N}\sum_{i=1}^{N}\left[\psi_1\left(W_i,\theta_0,p_0,\eta_{10}\right)+G_{1p0}\left(D_i-p_0\right)\right] \\
=& \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\psi_1\left(W,\theta_0,p_0,\hat{\eta}_{1k}\right)\right]-\frac{1}{N}\sum_{i=1}^{N}\psi_1\left(W_i,\theta_0,p_0,\eta_{10}\right).
\end{aligned}
$$

It remains to show that $\sqrt{N}R_N = o_P(1)$.

This part is essentially identical to Step 3 in the proof of Theorem 3.1 in Chernozhukov et al. (2018). I reproduce it here for the reader's convenience. Because K is a fixed integer, which is independent of N, it suffices to show that for any $k \in [K]$,

$$
\mathbb{E}_{n,k}\left[\psi_1\left(W,\theta_0,p_0,\hat{\eta}_{1k}\right)\right]-\frac{1}{n}\sum_{i\in I_k}\psi_1\left(W_i,\theta_0,p_0,\eta_{10}\right)=o_P\left(N^{-1/2}\right).
$$

Define the empirical process notation:

$$
\mathbb{G}_{n,k}\left[\phi\left(W\right)\right]=\frac{1}{\sqrt{n}}\sum_{i\in I_k}\left(\phi\left(W_i\right)-\int\phi\left(w\right)dP\right),
$$

where ϕ is any P-integrable function on $\mathcal{W}$. By the triangle inequality, we have

$$
\parallel\mathbb{E}_{n,k}\left[\psi_1\left(W,\theta_0,p_0,\hat{\eta}_{1k}\right)\right]-\frac{1}{n}\sum_{i\in I_k}\psi_1\left(W_i,\theta_0,p_0,\eta_{10}\right)\parallel\leq\frac{I_{1,k}+I_{2,k}}{\sqrt{n}},
$$

where

$$
\begin{aligned}
I_{1,k} &\equiv\parallel\mathbb{G}_{n,k}\left[\psi_1\left(W,\theta_0,p_0,\hat{\eta}_{1k}\right)\right]-\mathbb{G}_{n,k}\left[\psi_1\left(W,\theta_0,p_0,\eta_{10}\right)\right]\parallel, \\
I_{2,k} &\equiv\sqrt{n}\parallel E_P\left[\psi_1\left(W,\theta_0,p_0,\hat{\eta}_{1k}\right)\mid\left(W_i\right)_{i\in I_k^c}\right]-E_P\left[\psi_1\left(W,\theta_0,p_0,\eta_{10}\right)\right]\parallel.
\end{aligned}
$$

To bound $I_{1,k}$, note that conditional on $(W_i)_{i \in I_k^c}$, the estimator $\hat{\eta}_{1k}$ is nonstochastic. Under the event that $\hat{\eta}_{1k} \in \mathcal{T}_N$, we have

$$\begin{aligned} E_P\left[I_{1,k}^2 \mid (W_i)_{i \in I_k^c}\right] &= E_P\left[\| \psi_1(W, \theta_0, p_0, \hat{\eta}_{1k}) - \psi_1(W, \theta_0, p_0, \eta_{10}) \|^2 \mid (W_i)_{i \in I_k^c}\right] \\ &\leq \sup_{\eta_1 \in \mathcal{T}_N} E_P\left[\| \psi_1(W, \theta_0, p_0, \eta_1) - \psi_1(W, \theta_0, p_0, \eta_{10}) \|^2 \mid (W_i)_{i \in I_k^c}\right] \\ &= \sup_{\eta_1 \in \mathcal{T}_N} E_P\left[\| \psi_1(W, \theta_0, p_0, \eta_1) - \psi_1(W, \theta_0, p_0, \eta_{10}) \|^2\right] \\ &= (\varepsilon_N)^2 \end{aligned}$$

by (A.1). Hence, $I_{1,k} = O_P(\varepsilon_N)$ by Lemma A.1. To bound $I_{2,k}$, define the following function

$$f_k(r) = E_P\left[\psi_1(W, \theta_0, p_0, \eta_{10} + r(\hat{\eta}_{1k} - \eta_{10})) \mid (W_i)_{i \in I_k^c}\right] - E\left[\psi_1(W, \theta_0, p_0, \eta_{10})\right]$$

for $r \in [0, 1)$. By Taylor series expansion, we have

$$f_k(1) = f_k(0) + f_k'(0) + f_k''(\tilde{r})/2, \text{ for some } \tilde{r} \in (0, 1).$$

Note that $f_k(0) = 0$ because $E\left[\psi_1(W, \theta_0, p_0, \eta_{10}) \mid (W_i)_{i \in I_k^c}\right] = E\left[\psi_1(W, \theta_0, p_0, \eta_{10})\right]$. Furthermore, on the event $\hat{\eta}_{1k} \in \mathcal{T}_N$,

$$\| f_k'(0) \| = \| \partial_{\eta_1} E_P \psi_1(W, \theta_0, p_0, \eta_{10})[\hat{\eta}_{1k} - \eta_{10}] \| = 0$$

by the orthogonality of ψ_1. Also, on the event $\hat{\eta}_{1k} \in \mathcal{T}_N$,

$$\| f_k''(\tilde{r}) \| \leq \sup_{r \in (0,1)} \| f_k''(r) \| \leq (\varepsilon_N)^2$$

by (A.2). Thus,

$$I_{2,k} = \sqrt{n} \, \| f_k(1) \| = O_P\left(\sqrt{n}(\varepsilon_N)^2\right).$$

Together with the result on $I_{1,k}$, we have

$$\begin{aligned} \mathbb{E}_{n,k}\left[\psi_1(W, \theta_0, p_0, \hat{\eta}_{1k})\right] - \frac{1}{n}\sum_{i \in I_k} \psi_1(W_i, \theta_0, p_0, \eta_{10}) &\leq \frac{I_{1,k} + I_{2,k}}{\sqrt{n}} \\ &= O_P\left(n^{-1/2}\varepsilon_N + (\varepsilon_N)^2\right) \\ &= o_P\left(N^{-1/2}\right) \end{aligned}$$

by $n = O(N)$ and $\varepsilon_N = o(N^{-1/4})$. Hence, $\sqrt{N} R_N = o_P(1)$.

Step 2. In this step, I present the proof of (A.1). We have the following decomposition:

$$\begin{aligned} \psi_1(W, \theta_0, p_0, \eta_1) - \psi_1(W, \theta_0, p_0, \eta_{10}) &= \frac{D - g(X)}{p_0(1 - g(X))}(Y(1) - Y(0) - \ell_1(X)) \\ &\quad - \frac{D - g_0(X)}{p_0(1 - g_0(X))}(Y(1) - Y(0) - \ell_{10}(X)) \\ &= \frac{U + g_0(X) - g(X)}{p_0(1 - g(X))}(V_1 + \ell_{10}(X) - \ell_1(X)) \\ &\quad - \frac{U V_1}{p_0(1 - g_0(X))}. \end{aligned}$$

Thus, we have

$$\psi_1(W,\theta_0,p_0,\eta_1)-\psi_1(W,\theta_0,p_0,\eta_{10}) = \frac{UV_1}{p_0(1-g(X))}+\frac{U(\ell_{10}(X)-\ell_1(X))}{p_0(1-g(X))} + \frac{(g_0(X)-g(X))V_1}{p_0(1-g(X))}-\frac{UV_1}{p_0(1-g_0(X))} + \frac{(g_0(X)-g(X))(\ell_{10}(X)-\ell_1(X))}{p_0(1-g(X))}.$$

Given $\kappa \le g_0(X) \le 1-\kappa$ and $\kappa \le g(X) \le 1-\kappa$,

$$\| \psi_1(W,\theta_0,p_0,\eta_1)-\psi_1(W,\theta_0,p_0,\eta_{10}) \|_{P,2} \le \frac{1}{p_0\kappa^2} \| UV_1(1-g_0(X)) + U(\ell_{10}(X)-\ell_1(X))(1-g_0(X)) + V_1(g_0(X)-g(X))(1-g_0(X)) + (g_0-g)(\ell_{10}-\ell_1)(1-g_0(X)) - UV_1(1-g(X)) \|_{P,2}.$$

By $\kappa \le g_0(X) \le 1-\kappa$ and $\kappa \le g(X) \le 1-\kappa$ again, we can obtain

$$\| \psi_1(W,\theta_0,p_0,\eta_1)-\psi_1(W,\theta_0,p_0,\eta_{10}) \|_{P,2} \le \frac{1-\kappa}{p_0\kappa^2} \| UV_1 + U(\ell_{10}(X)-\ell_1(X)) + V_1(g_0(X)-g(X)) + (g_0(X)-g(X))(\ell_{10}(X)-\ell_1(X)) - UV_1 \|_{P,2}.$$

Thus, by $E_P[U^2|X] \le C$ and $E_P\left[V_1^2 \mid X\right] \le C$,

$$\begin{aligned}\| \psi_1(W,\theta_0,p_0,\eta_1)-\psi_1(W,\theta_0,p_0,\eta_{10}) \|_{P,2} &\le \frac{(1-\kappa)\sqrt{C}}{p_0\kappa^2} \| \ell_{10}-\ell_1 \|_{P,2} \\ &+ \frac{(1-\kappa)\sqrt{C}}{p_0\kappa^2} \| g_0-g \|_{P,2} \\ &+ \frac{(1-\kappa)}{p_0\kappa^2} \| g_0-g \|_{P,2} \| \ell_{10}-\ell_1 \|_{P,2} \\ &\le O\left(\varepsilon_N+\varepsilon_N+(\varepsilon_N)^2\right) \\ &= O(\varepsilon_N).\end{aligned}$$

Step 3. In this step, I present the proof of (A.2). Define

$$f(r) = E_P[\psi_1(W,\theta_0,p_0,\eta_{10}+r(\eta_1-\eta_{10}))].$$

Then, its second-order derivative is

$$\begin{aligned}\partial_r^2 f(r) =& \frac{2}{p_0} E_P\left[\frac{(D-1)(g-g_0)^2}{(1-g_0-r(g-g_0))^3}(Y(1)-Y(0)-\ell_{10}-r(\ell_1-\ell_{10}))\right] \\ &- \frac{2}{p_0} E_P\left[\frac{D-1}{(1-g_0-r(g-g_0))^2}(\ell_1-\ell_{10})(g-g_0)\right].\end{aligned}$$

It follows that

$$| \partial_r^2 f(r) | \le O\left(\| (g-g_0) \|_{P,2}^2 + \| (g-g_0) \|_{P,2} \times \| (\ell_1-\ell_{10}) \|_{P,2}\right) \le (\varepsilon_N)^2.$$

Step 4. Notice that

$$\begin{aligned}\partial_p \psi_1 (W, \theta, p, \eta_1) &= - \frac{1}{p} \frac{D - g(X)}{1 - g(X)} (Y(1) - Y(0) - \ell_1 (X)) \\ &= - \frac{1}{p} (\psi_1 (W, \theta, p, \eta_1) + \theta),\end{aligned}$$

and then, we have

$$\begin{aligned}&\| \partial_p \psi_1 (W, \theta_0, p_0, \eta_1) - \partial_p \psi_1 (W, \theta_0, p_0, \eta_{10}) \|_{P,2} \\ &= \frac{1}{p_0} \| \psi_1 (W, \theta_0, p_0, \eta_1) - \psi_1 (W, \theta_0, p_0, \eta_{10}) \|_{P,2} \\ &= O(\varepsilon_N)\end{aligned}$$

by Step 2.

Step 5. Notice that

$$\begin{aligned}\partial_p^2 \psi_1 (W, \theta, p, \eta_1) &= \frac{2}{p^3} \frac{D - g(X)}{1 - g(X)} (Y(1) - Y(0) - \ell_1 (X)) \\ &= \frac{2}{p^2} (\psi_1 (W, \theta, p, \eta_1) + \theta),\end{aligned}$$

and then, we have

$$\begin{aligned}\partial_p^2 \psi_1 (W, \theta_0, p, \eta_1) - \partial_p^2 \psi_1 (W, \theta_0, p_0, \eta_{10}) =& \partial_p^2 \psi_1 (W, \theta_0, p_0, \eta_1) - \partial_p^2 \psi_1 (W, \theta_0, p_0, \eta_{10}) \\ &+ \partial_{p^3}^3 \psi_1 (W, \theta_0, \bar{p}, \eta_1) (p - p_0) \\ =& \frac{2}{p_0^2} (\psi_1 (W, \theta_0, p_0, \eta_1) - \psi_1 (W, \theta_0, p_0, \eta_{10})) \\ &- \frac{6}{\bar{p}^4} \frac{(D - g(X))(Y(1) - Y(0) - \ell_1 (X))}{1 - g(X)} \\ &\times (p - p_0),\end{aligned}$$

where $\bar{p} \in (p, p_0)$. Thus, $\| \partial_p^2 \psi_1 (W, \theta_0, p, \eta_1) - \partial_p^2 \psi_1 (W, \theta_0, p_0, \eta_{10}) \|_{P,2}$ is bounded by

$$\begin{aligned}&\frac{2}{p_0^2} \| \psi_1 (W, \theta_0, p_0, \eta_1) - \psi_1 (W, \theta_0, p_0, \eta_{10}) \|_{P,2} \\ &+ \| \frac{6}{\bar{p}^4} \frac{D - g(X)}{1 - g(X)} (Y(1) - Y(0) - \ell_1 (X)) \|_{P,2} \times | p - p_0 | .\end{aligned}$$

The term in the second line is bounded by

$$\begin{aligned}\frac{6}{\bar{p}^4\kappa} \parallel (U+g_0-g)(V_1+\ell_{10}-\ell_1) \parallel_{P,2} \leq & \frac{6}{\bar{p}^4\kappa} \parallel UV_1 \parallel_{P,2} + \frac{6}{\bar{p}^4\kappa} \parallel U(\ell_{10}-\ell_1) \parallel_{P,2} \\ & + \frac{6}{\bar{p}^4\kappa} \parallel V_1(g_0-g) \parallel_{P,2} \\ & + \frac{6}{\bar{p}^4\kappa} \parallel g_0-g \parallel_{P,2} \parallel \ell_{10}-\ell_1 \parallel_{P,2} \\ \leq & \frac{6}{\bar{p}^4\kappa}\left(C+\sqrt{C} \parallel \ell_{10}-\ell_1 \parallel_{P,2}\right) \\ & + \frac{6}{\bar{p}^4\kappa}\sqrt{C} \parallel g_0-g \parallel_{P,2} \\ & + \frac{6}{\bar{p}^4\kappa} \parallel g_0-g \parallel_{P,2} \parallel \ell_{10}-\ell_1 \parallel_{P,2} \\ = & O(1)\end{aligned}$$

by $\|UV_1\|_{P,2} \leq \|UV_1\|_{P,4} \leq C$, $E_P[U^2|X] \leq C$, $E_P\left[V_1^2 \mid X\right] \leq C$, and the conditions on the rates of convergence. Together with Step 2, we obtain

$$\begin{aligned}\parallel \partial_p^2\psi_1(W,\theta_0,p,\eta_1) - \partial_p^2\psi_1(W,\theta_0,p_0,\eta_{10}) \parallel_{P,2} \leq & O(\varepsilon_N) + O(1) \times O\left(N^{-1/2}\right) \\ = & O(\varepsilon_N),\end{aligned}$$

where I assume that ε_N converges to zero no faster than $N^{-1/2}$.

Repeated cross sections:

In Step 1, I show the main result with the following auxiliary results:

$$\sup_{\eta_2\in\mathcal{T}_N}\left(E\left[\parallel \psi_2(W,\theta_0,p_0,\lambda_0,\eta_2) - \psi_2(W,\theta_0,p_0,\lambda_0,\eta_{20}) \parallel^2\right]\right)^{1/2} \leq \varepsilon_N, \tag{A.5}$$

$$\sup_{r\in(0,1),\eta_2\in\mathcal{T}_N} \parallel \partial_r^2 E\left[\psi_2(W,\theta_0,p_0,\lambda_0,\eta_{20}+r(\eta_2-\eta_{20}))\right] \parallel \leq (\varepsilon_N)^2. \tag{A.6}$$

$$\sup_{\eta_2\in\mathcal{T}_N}\left(E_P\left[\parallel \partial_p\psi_2(W,\theta_0,p_0,\lambda_0,\eta_2) - \partial_p\psi_2(W,\theta_0,p_0,\lambda_0,\eta_{20}) \parallel^2\right]\right)^{1/2} \leq \varepsilon_N, \tag{A.7}$$

$$\sup_{\eta_2\in\mathcal{T}_N}\left(E_P\left[\parallel \partial_\lambda\psi_2(W,\theta_0,p_0,\lambda_0,\eta_2) - \partial_\lambda\psi_2(W,\theta_0,p_0,\lambda_0,\eta_{20}) \parallel^2\right]\right)^{1/2} \leq \varepsilon_N, \tag{A.8}$$

$$\sup_{p\in\mathcal{P}_N,\eta_2\in\mathcal{T}_N}\left(E_P\left[\parallel \partial_p^2\psi_2(W,\theta_0,p,\lambda_0,\eta_2) - \partial_p^2\psi_2(W,\theta_0,p_0,\lambda_0,\eta_{20}) \parallel^2\right]\right)^{1/2} \leq \varepsilon_N, \tag{A.9}$$

$$\sup_{p\in\mathcal{P}_N,\lambda\in\Lambda_N,\eta_2\in\mathcal{T}_N}\left(E_P\left[\parallel \partial_\lambda^2\psi_2(W,\theta_0,p,\lambda,\eta_2) - \partial_\lambda^2\psi_2(W,\theta_0,p_0,\lambda_0,\eta_{20}) \parallel^2\right]\right)^{1/2} \leq \varepsilon_N, \tag{A.10}$$

$$\sup_{p\in\mathcal{P}_N,\eta_2\in\mathcal{T}_N}\left(E_P\left[\parallel \partial_\lambda\partial_p\psi_2(W,\theta_0,p,\lambda_0,\eta_2) - \partial_\lambda\partial_p\psi_2(W,\theta_0,p_0,\lambda_0,\eta_{20}) \parallel^2\right]\right)^{1/2} \leq \varepsilon_N, \tag{A.11}$$

where $\mathcal{T}_N$ is the set of all $\eta_2 = (g, \ell_2)$ consisting of P-square-integrable functions g and ℓ_2, such that

$$\parallel \eta_2 - \eta_{20} \parallel_{P,2} \leq \varepsilon_N,$$

$$\parallel g - 1/2 \parallel_{P,\infty} \leq 1/2 - \kappa,$$

$$\parallel (g-g_0) \parallel_{P,2}^2 + \parallel (g-g_0) \parallel_{P,2} \times \parallel (\ell_2-\ell_{20}) \parallel_{P,2} \leq (\varepsilon_N)^2,$$

and $\mathcal{P}_N$ and Λ_N are the sets consisting of all $p > 0$ and $\lambda > 0$, such that $|p - p_0| \leq N^{-1/2}$ and $|\lambda - \lambda_0| \leq N^{-1/2}$, respectively. By the regularity condition (3.2), $|\hat{p}_k - p_0| = O_P\left(N^{-1/2}\right)$, and $|\hat{\lambda}_k - \lambda_0| = O_P\left(N^{-1/2}\right)$, we have $\hat{\eta}_{2k} \in \mathcal{T}_N$, $\hat{p}_k \in \mathcal{P}_N$, and $\hat{\lambda}_k \in \Lambda_N$ with probability $1 - o(1)$.

In Steps 2 through 4, I show the above auxiliary results.

Step 1. Notice that

$$\begin{aligned}
\sqrt{N}\left(\tilde{\theta} - \theta_0\right) =& \sqrt{N}\left(\frac{1}{K}\sum_{k=1}^{K}\tilde{\theta}_k - \theta_0\right) \\
=& \sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\psi_2\left(W, \theta_0, \hat{p}_k, \hat{\lambda}_k, \hat{\eta}_{2k}\right)\right] \\
=& \sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\psi_2\left(W, \theta_0, p_0, \lambda_0, \hat{\eta}_{2k}\right)\right] \\
&+ \sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\partial_p\psi_2\left(W, \theta_0, p_0, \lambda_0, \hat{\eta}_{2k}\right)\right]\left(\hat{p}_k - p_0\right) \\
&+ \sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\partial_\lambda\psi_2\left(W, \theta_0, p_0, \lambda_0, \hat{\eta}_{2k}\right)\right]\left(\hat{\lambda}_k - \lambda_0\right) + o_P(1),
\end{aligned}$$

where the term $o_P(1)$, by the same arguments for the term b in repeated outcomes and the auxiliary results (A.9) through (A.11), contains the second-order terms

$$\sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\partial_p^2\psi_2\left(W, \theta_0, \bar{p}_k, \lambda_0, \hat{\eta}_{2k}\right)\right]\left(\hat{p}_k - p_0\right)^2,$$

$$\sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\partial_\lambda^2\psi_2\left(W, \theta_0, \hat{p}_k, \bar{\lambda}_k, \hat{\eta}_{2k}\right)\right]\left(\hat{\lambda}_k - \lambda_0\right)^2,$$

$$\sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\partial_\lambda\partial_p\psi_2\left(W, \theta_0, \bar{p}_k, \lambda_0, \hat{\eta}_{2k}\right)\right]\left(\hat{\lambda}_k - \lambda_0\right)\left(\hat{p}_k - p_0\right),$$

where $\bar{p}_k \in (\hat{p}_k, p_0)$ and $\bar{\lambda}_k \in \left(\hat{\lambda}_k, \lambda_0\right)$. On the other hand, by the same arguments for the term a in repeated outcomes and the auxiliary results (A.7) and (A.8), we have

$$\mathbb{E}_{n,k}\left[\partial_p\psi_2\left(W, \theta_0, p_0, \lambda_0, \hat{\eta}_{2k}\right)\right] \xrightarrow{p} E_p\left[\partial_p\psi_2\left(W, \theta_0, p_0, \lambda_0, \eta_{20}\right)\right] = G_{2p0},$$

$$\mathbb{E}_{n,k}\left[\partial_\lambda\psi_2\left(W, \theta_0, p_0, \lambda_0, \hat{\eta}_{2k}\right)\right] \xrightarrow{p} E_p\left[\partial_\lambda\psi_2\left(W, \theta_0, p_0, \lambda_0, \eta_{20}\right)\right] = G_{2\lambda 0}.$$

Hence, because $\hat{p}_k - p_0 = \mathbb{E}_{n,k}[D - p_0]$ and $\hat{\lambda}_k - \lambda_0 = \mathbb{E}_{n,k}[T - \lambda_0]$, we have

$$\begin{aligned}
\sqrt{N}\left(\tilde{\theta} - \theta_0\right) =& \sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\psi_2\left(W, \theta_0, p_0, \lambda_0, \hat{\eta}_{2k}\right)\right] \\
=& \sqrt{N}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\psi_2\left(W, \theta_0, p_0, \lambda_0, \hat{\eta}_{1k}\right) + G_{2p0}\left(D - p_0\right) + G_{2\lambda 0}\left(T - \lambda_0\right)\right] \\
&+ o_P(1) \\
=& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\psi_2\left(W_i, \theta_0, p_0, \lambda_0, \eta_{20}\right) + G_{2p0}\left(D_i - p_0\right) + G_{2\lambda 0}\left(T_i - \lambda_0\right)\right] \\
&+ \sqrt{N}R_N' + o_P(1),
\end{aligned}$$

where

$$
\begin{aligned}
R_N' =& \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\psi_2\left(W,\theta_0,p_0,\lambda_0,\hat{\eta}_{2k}\right)+G_{2p0}\left(D-p_0\right)+G_{2\lambda 0}\left(T-\lambda_0\right)\right] \\
&-\frac{1}{N}\sum_{i=1}^{N}\left[\psi_2\left(W_i,\theta_0,p_0,\lambda_0,\eta_{20}\right)+G_{2p0}\left(D_i-p_0\right)+G_{2\lambda 0}\left(T_i-\lambda_0\right)\right] \\
=& \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\psi_2\left(W,\theta_0,p_0,\lambda_0,\hat{\eta}_{2k}\right)\right]-\frac{1}{N}\sum_{i=1}^{N}\psi_2\left(W_i,\theta_0,p_0,\lambda_0,\eta_{10}\right).
\end{aligned}
$$

Using (A.5) and (A.6) and the same arguments as for Step 1 in repeated outcomes, one can show that $\sqrt{N}R_N' = o_P(1)$. Hence, it remains to prove the auxiliary results (A.5) through (A.11).

Step 2. Recall that $p_0' = p_0\lambda_0\left(1-\lambda_0\right)$. For (A.5), notice that

$$
\begin{aligned}
\psi_2\left(W,\theta_0,p_0,\lambda_0,\eta_2\right)-\psi_2\left(W,\theta_0,p_0,\lambda_0,\eta_{20}\right) =& \frac{D-g(X)}{p_0'\left(1-g(X)\right)}\left(\left(T-\lambda_0\right)Y-\ell_2(X)\right) \\
&-\frac{D-g_0(X)}{p_0'\left(1-g_0(X)\right)}\left(\left(T-\lambda_0\right)Y-\ell_{20}(X)\right) \\
=& \frac{U+g_0(X)-g(X)}{p_0'\left(1-g(X)\right)}\left(V_2+\ell_{20}(X)-\ell_2(X)\right) \\
&-\frac{UV_2}{p_0'\left(1-g_0(X)\right)}.
\end{aligned}
$$

The decomposition becomes

$$
\begin{aligned}
\psi_2\left(W,\theta_0,p_0,\lambda_0,\eta_2\right)-\psi_2\left(W,\theta_0,p_0,\lambda_0,\eta_{20}\right) =& \frac{UV_2}{p_0'\left(1-g(X)\right)}+\frac{U\left(\ell_{20}(X)-\ell_2(X)\right)}{p_0'\left(1-g(X)\right)} \\
&+\frac{\left(g_0(X)-g(X)\right)V_2}{p_0'\left(1-g(X)\right)} \\
&+\frac{\left(g_0(X)-g(X)\right)\left(\ell_{20}(X)-\ell_2(X)\right)}{p_0'\left(1-g(X)\right)} \\
&-\frac{UV_2}{p_0'\left(1-g_0(X)\right)}.
\end{aligned}
$$

Given that $\kappa \le g_0(X) \le 1-\kappa$, $\kappa \le g(X) \le 1-\kappa$, we have

$$
\begin{aligned}
\|\ \psi_2\left(W,\theta_0,p_0,\lambda_0,\eta_2\right)-\psi_2\left(W,\theta_0,p_0,\lambda_0,\eta_{20}\right)\|_{P,2} \le& \frac{1}{p_0'\kappa^2}\ \|\ UV_2\left(1-g_0(X)\right) \\
&+U\left(\ell_{20}(X)-\ell_2(X)\right)\left(1-g_0(X)\right) \\
&+V_2\left(g_0(X)-g(X)\right)\left(1-g_0(X)\right) \\
&+\left(g_0-g\right)\left(\ell_{20}-\ell_2\right)\left(1-g_0(X)\right) \\
&-UV_2\left(1-g(X)\right)\|_{P,2}\ .
\end{aligned}
$$

By $\kappa \leq g_0(X) \leq 1-\kappa$, $\kappa \leq g(X) \leq 1-\kappa$ again, we obtain

$$
\begin{aligned}
\| \psi_2(W,\theta_0,p_0,\lambda_0,\eta_2) - \psi_2(W,\theta_0,p_0,\lambda_0,\eta_{20}) \|_{P,2} \leq \frac{1-\kappa}{p_0'\kappa^2} \| & UV_2 \\
& + U(\ell_{20}(X) - \ell_2(X)) \\
& + V_2(g_0(X) - g(X)) \\
& + (g_0 - g)(\ell_{20} - \ell_2) \\
& - UV_2 \|_{P,2} .
\end{aligned}
$$

Given $E_P[U^2|X] \leq C$, $E_P\left[V_2^2 \mid X\right] \leq C$, and the conditions on the rates of convergence,

$$
\begin{aligned}
\| \psi_2(W,\theta_0,p_0,\lambda_0,\eta_2) - \psi_2(W,\theta_0,p_0,\lambda_0,\eta_{20}) \|_{P,2} \leq & \frac{(1-\kappa)\sqrt{C}}{p_0'\kappa^2} \| \ell_{20}(X) - \ell_2(X) \|_{P,2} \\
& + \frac{(1-\kappa)\sqrt{C}}{p_0'\kappa^2} \| g_0(X) - g(X) \|_{P,2} \\
& + \frac{(1-\kappa)}{p_0'\kappa^2} \| g_0 - g \|_{P,2} \| \ell_{20} - \ell_2 \|_{P,2} \\
\leq & O\left(\varepsilon_N + \varepsilon_N + (\varepsilon_N)^2\right) \\
= & O(\varepsilon_N) .
\end{aligned}
$$

For (A.6), let $f(r) = E_P[\psi_2(W,\theta_0,p_0,\lambda_0,\eta_{20} + r(\eta_2 - \eta_{20}))]$. Then, the second-order derivative is

$$
\begin{aligned}
\partial_r^2 f(r) = & \frac{2}{p_0'} E_P\left[\frac{(D-1)(g-g_0)^2}{(1-g_0-r(g-g_0))^3}((T-\lambda_0)Y - \ell_{20} - r(\ell_2 - \ell_{20}))\right] \\
& - \frac{2}{p_0'} E_P\left[\frac{D-1}{(1-g_0-r(g-g_0))^2}(\ell_2 - \ell_{20})(g-g_0)\right].
\end{aligned}
$$

It follows that

$$
| \partial_r^2 f(r) | \leq O\left(\| (g-g_0) \|_{P,2}^2 + \| (g-g_0) \|_{P,2} \times \| (\ell_2 - \ell_{20}) \|_{P,2}\right) \leq (\varepsilon_N)^2 .
$$

Step 3. For (A.7), notice that

$$
\begin{aligned}
\partial_p \psi_2(W,\theta,p,\lambda,\eta_2) = & - \frac{1}{p^2\lambda(1-\lambda)} \frac{D-g(X)}{1-g(X)}((T-\lambda)Y - \ell_2(X)) \\
= & - \frac{1}{p}(\psi_2(W,\theta,p,\lambda,\eta_2) + \theta) ,
\end{aligned}
$$

and then, we have

$$
\begin{aligned}
\| \partial_p\psi_2(W,\theta_0,p_0,\lambda_0,\eta_2) - \partial_p\psi_2(W,\theta_0,p_0,\lambda_0,\eta_{20}) \|_{P,2} = & \frac{1}{p_0} \| \psi_2(W,\theta_0,p_0,\lambda_0,\eta_2) \\
& - \psi_2(W,\theta_0,p_0,\lambda_0,\eta_{20}) \|_{P,2} \\
= & O(\varepsilon_N)
\end{aligned}
$$

by the proof of (A.5).

For (A.8), notice that

$$
\begin{aligned}
\partial_\lambda \psi_2(W,\theta,p,\lambda,\eta_2) = & - \frac{1-2\lambda}{\lambda^2(1-\lambda)^2} \frac{D-g(X)}{p(1-g(X))}((T-\lambda)Y - \ell_2(X)) \\
& - \frac{Y}{p\lambda(1-\lambda)} \frac{D-g(X)}{1-g(X)} .
\end{aligned}
$$

Define $\partial_\lambda \psi_{20} \equiv \partial_\lambda \psi_2(W, \theta_0, p_0, \lambda_0, \eta_{20})$. Then,

$$\begin{aligned}
\| \partial_\lambda \psi_2(W, \theta_0, p_0, \lambda_0, \eta_2) - \partial_\lambda \psi_{20} \|_{P,2} = & \| \psi_2(W, \theta_0, p_0, \lambda_0, \eta_2) - \psi_2(W, \theta_0, p_0, \lambda_0, \eta_{20}) \|_{P,2} \\
& \times \frac{| 1 - 2\lambda_0 |}{\lambda_0 (1 - \lambda_0)} \\
& + \| \frac{Y}{p_0'} \left(\frac{D - g(X)}{1 - g(X)} - \frac{D - g_0(X)}{1 - g_0(X)} \right) \|_{P,2} \\
= & O(\varepsilon_N) + \| \frac{Y}{p_0'} \left(\frac{D - g(X)}{1 - g(X)} - \frac{D - g_0(X)}{1 - g_0(X)} \right) \|_{P,2} \\
\leq & O(\varepsilon_N) + \frac{1}{p_0' \kappa^2} \| Y (g - g_0)(D - 1) \|_{P,2} \\
\leq & O(\varepsilon_N) + \frac{\sqrt{C}}{p_0' \kappa^2} \| g - g_0 \|_{P,2} \\
= & O(\varepsilon_N),
\end{aligned}$$

by (A.5) and $E_P[Y^2|X] \leq C$.

Step 4. For (A.9), notice that we have

$$\partial_p^2 \psi_2 (W, \theta, p, \lambda, \eta_2) = \frac{2}{p^3 \lambda (1 - \lambda)} \frac{D - g(X)}{1 - g(X)} ((T - \lambda) Y - \ell_2(X)).$$

Define $\partial_p^2 \psi_{20} \equiv \partial_p^2 \psi_2 (W, \theta_0, p_0, \lambda_0, \eta_{20})$. Then, we have

$$\begin{aligned}
\partial_p^2 \psi_2 (W, \theta_0, p, \lambda_0, \eta_2) - \partial_p^2 \psi_{20} = & \partial_p^2 \psi_2 (W, \theta_0, p_0, \lambda_0, \eta_2) - \partial_p^2 \psi_{20} \\
& + \partial_p^3 \psi_2 (W, \theta_0, \bar{p}, \lambda_0, \eta_2)(p - p_0) \\
= & \frac{2}{p^2} (\psi_2 (W, \theta_0, p_0, \lambda_0, \eta_2) - \psi_2 (W, \theta_0, p_0, \lambda_0, \eta_{20})) \\
& + \partial_p^3 \psi_2 (W, \theta_0, \bar{p}, \lambda_0, \eta_2)(p - p_0),
\end{aligned}$$

where $\bar{p} \in (p, p_0)$. Hence, we have

$$\begin{aligned}
\| \partial_p^2 \psi_2 (W, \theta_0, p, \lambda_0, \eta_2) - \partial_p^2 \psi_{20} \|_{P,2} \leq & \frac{2}{p^2} \| \psi_2 (W, \theta_0, p_0, \lambda_0, \eta_2) - \psi_2 (W, \theta_0, p_0, \lambda_0, \eta_{20}) \| \\
& + \| \frac{D - g(X)}{1 - g(X)} ((T - \lambda_0) Y - \ell_2(X)) \|_{P,2} \\
& \times \frac{6}{\bar{p}^4 \lambda_0 (1 - \lambda_0)} | p - p_0 | .
\end{aligned}$$

By (A.5), we have $\|\psi_2(W, \theta_0, p_0, \lambda_0, \eta_2) - \psi_2(W, \theta_0, p_0, \lambda_0, \eta_{20})\|_{P,2} = O(\varepsilon_N)$. The term in the second line is bounded by

$$\begin{aligned}
\frac{1}{\kappa} \parallel (U + g_0 - g)(V_2 + \ell_{20} - \ell_2) \parallel_{P,2} \leq & \frac{1}{\kappa} \parallel UV_2 \parallel_{P,2} + \frac{1}{\kappa} \parallel U(\ell_{20} - \ell_2) \parallel_{P,2} \\
& + \frac{1}{\kappa} \parallel V_2 (g_0 - g) \parallel_{P,2} \\
& + \frac{1}{\kappa} \parallel g_0 - g \parallel_{P,2} \parallel \ell_{20} - \ell_2 \parallel_{P,2} \\
\leq & \frac{1}{\kappa} \left(C + \sqrt{C} \parallel \ell_{20} - \ell_2 \parallel_{P,2} + \sqrt{C} \parallel g_0 - g \parallel_{P,2} \right) \\
& + \frac{1}{\kappa} \parallel g_0 - g \parallel_{P,2} \parallel \ell_{20} - \ell_2 \parallel_{P,2} \\
= & O(1)
\end{aligned}$$

by $\|UV_2\|_{P,2} \leq \|UV_2\|_{P,4} \leq C$, $E_P[U^2|X] \leq C$, and $E_P\left[V_2^2 \mid X\right] \leq C$. Thus, we obtain

$$\begin{aligned}
\parallel \partial_p^2 \psi_2(W, \theta_0, p, \lambda_0, \eta_2) - \partial_p^2 \psi_{20} \parallel_{P,2} \leq & O(\varepsilon_N) + O(1) \times O\left(N^{-1/2}\right) \\
= & O(\varepsilon_N),
\end{aligned}$$

where I assume that ε_N converges to zero no faster than $N^{-1/2}$.

For (A.10), notice that we have

$$\begin{aligned}
\partial_\lambda^2 \psi_2(W, \theta, p, \lambda, \eta_2) = & \frac{c_1}{p\lambda^3 (1-\lambda)^3} \frac{D - g(X)}{1 - g(X)} ((T - \lambda) Y - \ell_2(X)) \\
& + \frac{2 - 4\lambda}{p\lambda^2 (1-\lambda)^2} \frac{D - g(X)}{1 - g(X)} Y,
\end{aligned}$$

where c_1 is a constant depending on λ. Define $\partial_\lambda^2 \psi_{20} \equiv \partial_\lambda^2 \psi_2(W, \theta_0, p_0, \lambda_0, \eta_{20})$. Then, we have

$$\begin{aligned}
\partial_\lambda^2 \psi_2(W, \theta_0, p, \lambda, \eta_2) - \partial_\lambda^2 \psi_{20} = & \partial_\lambda^2 \psi_2(W, \theta_0, p_0, \lambda_0, \eta_2) - \partial_\lambda^2 \psi_{20} \\
& + \partial_\lambda^2 \partial_p \psi_2(W, \theta_0, \bar{p}, \lambda, \eta_2)(p - p_0) \\
& + \partial_\lambda^3 \psi_2\left(W, \theta_0, p_0, \bar{\lambda}, \eta_2\right)(\lambda - \lambda_0) \\
= & \frac{c_1}{\lambda_0^2 (1-\lambda_0)^2} (\psi_2(W, \theta_0, p_0, \lambda_0, \eta_2) - \psi_2(W, \theta_0, p_0, \lambda_0, \eta_{20})) \\
& + \frac{2 - 4\lambda_0}{p_0 \lambda_0^2 (1-\lambda_0)^2} \left(\frac{D - g(X)}{1 - g(X)} - \frac{D - g_0(X)}{1 - g_0(X)} \right) Y \\
& + \partial_\lambda^2 \partial_p \psi_2(W, \theta_0, \bar{p}, \lambda, \eta_2)(p - p_0) \\
& + \partial_\lambda^3 \psi_2\left(W, \theta_0, p_0, \bar{\lambda}, \eta_2\right)(\lambda - \lambda_0),
\end{aligned}$$

where $\bar{p} \in (p, p_0)$ and $\bar{\lambda} \in (\lambda, \lambda_0)$. By the triangle inequality, we have

$$\begin{aligned}
\| \partial_\lambda^2 \psi_2 (W, \theta_0, p, \lambda, \eta_2) - \partial_\lambda^2 \psi_{20} \|_{P,2} \leq & \frac{| c_1 |}{\lambda^2 (1-\lambda)^2} \times \\
& \| \psi_2 (W, \theta_0, p_0, \lambda_0, \eta_2) - \psi_2 (W, \theta_0, p_0, \lambda_0, \eta_{20}) \|_{P,2} \\
& + \frac{| 2 - 4\lambda_0 | \, Y}{p_0 \lambda_0^2 (1-\lambda_0)^2} \left\| \left(\frac{D - g(X)}{1 - g(X)} - \frac{D - g_0(X)}{1 - g_0(X)} \right) \right\|_{P,2} \\
& + \| \partial_\lambda^2 \partial_p \psi_2 (W, \theta_0, \bar{p}, \lambda, \eta_2) \|_{P,2} | \, p - p_0 \, | \\
& + \| \partial_\lambda^3 \psi_2 \left(W, \theta_0, p_0, \bar{\lambda}, \eta_2\right) \|_{P,2} | \, \lambda - \lambda_0 \, | \, .
\end{aligned}$$

The norm term is the second line is bounded by

$$\begin{aligned}
\frac{1}{\kappa^2} \| Y (D-1)(g - g_0) \|_{P,2} \leq & \frac{\sqrt{C}}{\kappa^2} \| g - g_0 \|_{P,2} \\
= & O(\varepsilon_N),
\end{aligned}$$

by $E_P[Y^2|X] \leq C$ and $D \in \{0, 1\}$. The two high-order terms are bounded by

$$\begin{aligned}
\| \partial_\lambda^2 \partial_p \psi_2 (W, \theta_0, \bar{p}, \lambda, \eta_2) \|_{P,2} \leq & \frac{| c_1 |}{\bar{p}^2 \lambda^3 (1-\lambda)^3} \left\| \frac{D - g(X)}{1 - g(X)} ((T - \lambda) Y - \ell_2(X)) \right\|_{P,2} \\
& + \frac{| 2 - 4\lambda |}{\bar{p} \lambda^2 (1-\lambda)^2} \left\| \frac{D - g(X)}{1 - g(X)} Y \right\|_{P,2}
\end{aligned}$$

and

$$\begin{aligned}
\| \partial_\lambda^3 \psi_2 \left(W, \theta_0, p_0, \bar{\lambda}, \eta_2\right) \|_{P,2} \leq & \frac{| c_2 |}{p_0 \bar{\lambda}^4 \left(1 - \bar{\lambda}\right)^4} \left\| \frac{D - g(X)}{1 - g(X)} \left(\left(T - \bar{\lambda}\right) Y - \ell_2(X)\right) \right\|_{P,2} \\
& + \frac{| c_3 |}{p_0 \bar{\lambda}^3 \left(1 - \bar{\lambda}\right)^3} \left\| \frac{D - g(X)}{1 - g(X)} \times Y \right\|_{P,2},
\end{aligned}$$

where c_2 and c_3 are constants depending on λ. Using the same arguments in (A.9), one can show that

$$\begin{aligned}
& \left\| \frac{D - g(X)}{1 - g(X)} ((T - \lambda) Y - \ell_2(X)) \right\|_{P,2} \leq O(1), \\
& \left\| \frac{D - g(X)}{1 - g(X)} \left(\left(T - \bar{\lambda}\right) Y - \ell_2(X)\right) \right\|_{P,2} \leq O(1).
\end{aligned}$$

Also, we have

$$\begin{aligned}
\left\| \frac{D - g(X)}{1 - g(X)} \times Y \right\|_{P,2} = & \left\| \frac{U + g_0(X) - g(X)}{1 - g(X)} \times Y \right\|_{P,2} \\
\leq & \frac{1}{\kappa} \left(\| UY \|_{P,2} + \| (g_0 - g) Y \|_{P,2} \right) \\
\leq & \frac{1}{\kappa} \left(C + \sqrt{C} \| g_0 - g \|_{P,2} \right) \\
= & O(1)
\end{aligned}$$

by $\|UY\|_{P,2} \leq C$ and $E_P[Y^2|X] \leq C$.

Finally, we obtain

$$\begin{aligned}
\| \partial_\lambda^2 \psi_2 (W, \theta_0, p, \lambda, \eta_2) - \partial_\lambda^2 \psi_{20} \|_{P,2} \leq & O(\varepsilon_N) + O(\varepsilon_N) + O(1) \, O\left(N^{-1/2}\right) \\
& + O(1) \, O\left(N^{-1/2}\right) \\
= & O(\varepsilon_N),
\end{aligned}$$

where I assume that ε_N converges to zero no faster than $N^{-1/2}$.

For (A.11), notice that the derivative is

$$\partial_\lambda \partial_p \psi_2 (W, \theta, p, \lambda, \eta_2) = \frac{1-2\lambda}{p^2 \lambda^2 (1-\lambda)^2} \frac{D-g(X)}{1-g(X)} ((T-\lambda) Y - \ell_2 (X)) + \frac{Y}{p^2 \lambda (1-\lambda)} \frac{D-g(X)}{1-g(X)}.$$

Define $\partial_\lambda \partial_p \psi_{20} \equiv \partial_\lambda \partial_p \psi_2(W, \theta_0, p_0, \lambda_0, \eta_{20})$. Then, we have

$$\partial_\lambda \partial_p \psi_2 (W, \theta_0, p, \lambda_0, \eta_2) - \partial_\lambda \partial_p \psi_{20} = \partial_\lambda \partial_p \psi_2 (W, \theta_0, p_0, \lambda_0, \eta_2) - \partial_\lambda \partial_p \psi_{20} + \partial_\lambda \partial_p^2 \psi_2 (W, \theta_0, \bar{p}, \lambda_0, \eta_2)(p - p_0),$$

where $\bar{p} \in (p, p_0)$. By the triangle inequality, we obtain

$$\| \partial_\lambda \partial_p \psi_2 (W, \theta_0, p, \lambda_0, \eta_2) - \partial_\lambda \partial_p \psi_{20} \|_{P,2} \leq \frac{1}{p} \| \partial_\lambda \psi_2 (W, \theta_0, p_0, \lambda_0, \eta_2) - \partial_\lambda \psi_{20} \|_{P,2} + \| \partial_\lambda \partial_p^2 \psi_2 (W, \theta_0, \bar{p}, \lambda_0, \eta_2) \|_{P,2} | p - p_0 | .$$

Using the same arguments in (A.9) and (A.10), one can show that the high-order term is bounded by

$$\begin{aligned} \| \partial_\lambda \partial_p^2 \psi_2 (W, \theta_0, \bar{p}, \lambda_0, \eta_2) \|_{P,2} \leq & \left\| \frac{2-4\lambda_0}{\bar{p}^3 \lambda_0^2 (1-\lambda_0)^2} \frac{D-g(X)}{1-g(X)} ((T-\lambda_0) Y - \ell_2 (X)) \right\|_{P,2} \\ & + \left\| \frac{2Y}{\bar{p}^3 \lambda_0 (1-\lambda_0)} \frac{D-g(X)}{1-g(X)} \right\|_{P,2} \\ \leq & O(1). \end{aligned}$$

Together with (A.8), we obtain

$$\begin{aligned} \| \partial_\lambda \partial_p \psi_2 (W, \theta_0, p, \lambda_0, \eta_2) - \partial_\lambda \partial_p \psi_{20} \|_{P,2} \leq & O(\varepsilon_N) + O(1) O\left(N^{-1/2}\right) \\ = & O(\varepsilon_N), \end{aligned}$$

where I assume that ε_N converges to zero no faster than $N^{-1/2}$. □

Proof of Theorem 3.2. In Step 1, I show the main result using the auxiliary results

$$\sup_{p \in \mathcal{P}_N, \eta_1 \in \mathcal{T}_N} \left(E_P \left[\| \bar{\psi}_1 (W, \theta_0, p, \eta_1) - \bar{\psi}_1 (W, \theta_0, p_0, \eta_{10}) \|^2 \right] \right)^{1/2} \leq \varepsilon_N, \quad \text{(A.12)}$$

$$\left(E_P \left[\bar{\psi}_1 (W, \theta_0, p_0, \eta_{10})^4 \right] \right)^{1/4} \leq C_1, \quad \text{(A.13)}$$

where $\mathcal{P}_N$ and $\mathcal{T}_N$ are specified in the proof of Theorem 3.1, C_1 is a constant, and

$$\bar{\psi}_1 (W, \theta, p, \eta_1) \equiv \frac{1}{p} \frac{D-g(X)}{1-g(X)} (Y(1) - Y(0) - \ell_1 (X)) - \frac{D\theta}{p}.$$

In fact, we have $E_P \left[\left(\bar{\psi}_1 (W, \theta_0, p_0, \eta_{10}) \right)^2 \right] = \Sigma_{10}$. In Step 2, I show the auxiliary results (A.12) and (A.13).

Step 1. Notice that

$$\begin{aligned} \hat{\Sigma}_1 = & \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}_{n,k} \left[\left(\psi_1 \left(W, \tilde{\theta}, \hat{p}_k, \hat{\eta}_{1k} \right) + \hat{G}_{1p} (D - \hat{p}_k) \right)^2 \right] \\ = & \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}_{n,k} \left[\left(\frac{1}{\hat{p}_k} \frac{D - \hat{g}_k (X)}{1 - \hat{g}_k (X)} \left(Y(1) - Y(0) - \hat{\ell}_{1k} (X) \right) - \frac{D\tilde{\theta}}{\hat{p}_k} \right)^2 \right] \\ = & \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}_{n,k} \left[\bar{\psi}_1 \left(W, \tilde{\theta}, \hat{p}_k, \hat{\eta}_{1k} \right)^2 \right], \end{aligned}$$

where the second equality follows from $\hat{G}_{1p} = -\tilde{\theta}/\hat{p}_k$.

Because K is fixed, and is independent of N, it suffices to show that for each $k \in [k]$,

$$I_k \equiv |\mathbb{E}_{n,k}\left[\bar{\psi}_1\left(W, \tilde{\theta}, \hat{p}_k, \hat{\eta}_{1k}\right)^2\right] - E_P\left[\bar{\psi}_1\left(W, \theta_0, p_0, \eta_{10}\right)^2\right]| = o_P(1).$$

By the triangle inequality, we have

$$I_k \leq I_{3,k} + I_{4,k},$$

where

$$I_{3,k} \equiv |\mathbb{E}_{n,k}\left[\bar{\psi}_1\left(W, \tilde{\theta}, \hat{p}_k, \hat{\eta}_{1k}\right)^2\right] - \mathbb{E}_{n,k}\left[\bar{\psi}_1\left(W, \theta_0, p_0, \eta_{10}\right)^2\right]|,$$

$$I_{4,k} \equiv |\mathbb{E}_{n,k}\left[\bar{\psi}_1\left(W, \theta_0, p_0, \eta_{10}\right)^2\right] - E_P\left[\bar{\psi}_1\left(W, \theta_0, p_0, \eta_{10}\right)^2\right]|.$$

To bound $I_{4,\,k}$, we have

$$\begin{aligned} E_P\left[I_{4,k}^2\right] &\leq n^{-1} E_P\left[\bar{\psi}_1\left(W, \theta_0, p_0, \eta_{10}\right)^4\right] \\ &\leq n^{-1} C_1^4, \end{aligned}$$

where the last inequality follows from (A.13). Then, we have $I_{4,\,k} = O_P(n^{1/2})$.

Next, we bound $I_{3,\,k}$. This part is essentially identical to the proof of Theorem 3.2 in Chernozhukov et al. (2018). I reproduce it here for the reader's convenience. Observe that for any number a and δa,

$$|(a + \delta a)^2 - a^2| \leq 2(\delta a)(a + \delta a).$$

Denote $\psi_i = \bar{\psi}_1\left(W_i, \theta_0, p_0, \eta_{10}\right)$ and $\hat{\psi}_i = \bar{\psi}_1\left(W_i, \tilde{\theta}, \hat{p}_k, \hat{\eta}_{1k}\right)$, and $a \equiv \psi_i$, $a + \delta a \equiv \hat{\psi}_i$. Then,

$$\begin{aligned} I_{3,k} &= \left| \frac{1}{n}\sum_{i \in I_k} \left(\hat{\psi}_i\right)^2 - \left(\psi_i\right)^2 \right| \leq \frac{1}{n}\sum_{i \in I_k} \left| \left(\hat{\psi}_i\right)^2 - \left(\psi_i\right)^2 \right| \\ &\leq \frac{2}{n}\sum_{i \in I_k} |\hat{\psi}_i - \psi_i| \times \left(|\psi_i| + |\hat{\psi}_i - \psi_i|\right) \\ &\leq \left(\frac{2}{n}\sum_{i \in I_k} |\hat{\psi}_i - \psi_i|^2\right)^{1/2} \left(\frac{2}{n}\sum_{i \in I_k} \left(|\psi_i| + |\hat{\psi}_i - \psi_i|\right)^2\right)^{1/2} \\ &\leq \left(\frac{2}{n}\sum_{i \in I_k} |\hat{\psi}_i - \psi_i|^2\right)^{1/2} \left[\left(\frac{2}{n}\sum_{i \in I_k} |\psi_i|^2\right)^{1/2} + \left(\frac{2}{n}\sum_{i \in I_k} |\hat{\psi}_i - \psi_i|^2\right)^{1/2}\right]. \end{aligned}$$

Thus,

$$I_{3,k}^2 \lesssim S_N \times \left(\frac{1}{n}\sum_{i \in I_k} \| \bar{\psi}_1\left(W_i, \theta_0, p_0, \eta_{10}\right) \|^2 + S_N\right),$$

where

$$S_N \equiv \frac{1}{n}\sum_{i \in I_k} \| \bar{\psi}_1\left(W_i, \tilde{\theta}, \hat{p}_k, \hat{\eta}_{1k}\right) - \bar{\psi}_1\left(W_i, \theta_0, p_0, \eta_{10}\right) \|^2.$$

Because $\frac{1}{n}\sum_{i\in I_k}\parallel\bar{\psi}_1(W_i,\theta_0,p_0,\eta_0)\parallel^2=O_P(1)$, it suffices to bound S_N. We have the decomposition

$$\begin{aligned}S_N=&\frac{1}{n}\sum_{i\in I_k}\parallel\bar{\psi}_1(W_i,\theta_0,\hat{p}_k,\hat{\eta}_{1k})+\partial_\theta\bar{\psi}_1\left(W_i,\bar{\theta},\hat{p}_k,\hat{\eta}_{1k}\right)\left(\tilde{\theta}-\theta_0\right)-\bar{\psi}_1(W_i,\theta_0,p_0,\eta_{10})\parallel^2\\=&\frac{1}{n}\sum_{i\in I_k}\parallel\bar{\psi}_1(W_i,\theta_0,\hat{p}_k,\hat{\eta}_{1k})+\frac{D_i}{\hat{p}_k}\left(\tilde{\theta}-\theta_0\right)-\bar{\psi}_1(W_i,\theta_0,p_0,\eta_{10})\parallel^2\\\leq&\frac{1}{n}\sum_{i\in I_k}\parallel\frac{D_i}{\hat{p}_k}\left(\tilde{\theta}-\theta_0\right)\parallel^2+\frac{1}{n}\sum_{i\in I_k}\parallel\bar{\psi}_1(W_i,\theta_0,\hat{p}_k,\hat{\eta}_{1k})-\bar{\psi}_1(W_i,\theta_0,p_0,\eta_{10})\parallel^2,\end{aligned}$$

where $\bar{\theta}\in\left(\tilde{\theta}-\theta_0\right)$. The first term is bounded by

$$\begin{aligned}\frac{1}{n}\sum_{i\in I_k}\parallel\frac{D_i}{\hat{p}_k}\left(\tilde{\theta}-\theta_0\right)\parallel^2\leq&\left(\frac{1}{n}\sum_{i\in I_k}\left(\frac{D_i}{\hat{p}_k}\right)^2\right)\parallel\tilde{\theta}-\theta_0\parallel^2\\=&\left(\frac{1}{n}\sum_{i\in I_k}\left(\frac{D_i}{p_0}\right)^2+o_P(1)\right)\parallel\tilde{\theta}-\theta_0\parallel^2\\=&O_P(1)\times O_P\left(N^{-1}\right).\end{aligned}$$

Also, notice that conditional on $(W_i)_{i\in I_k^c}$, both $\hat{p}_k$ and $\hat{\eta}_{1k}$ can be treated as fixed. Under the event that $\hat{p}_k\in\mathcal{P}_N$ and $\hat{\eta}_{1k}\in\mathcal{T}_N$, we have

$$\begin{aligned}&E_P\left[\parallel\bar{\psi}_1(W_i,\theta_0,\hat{p}_k,\hat{\eta}_{1k})-\bar{\psi}_1(W_i,\theta_0,p_0,\eta_{10})\parallel^2|\,(W_i)_{i\in I_k^c}\right]\\\leq&\sup_{p\in\mathcal{P}_N,\eta_1\in\mathcal{T}_N}E_P\left[\parallel\bar{\psi}_1(W_i,\theta_0,p,\eta_1)-\bar{\psi}_1(W_i,\theta_0,p_0,\eta_{10})\parallel^2\right]=(\varepsilon_N)^2\end{aligned}$$

by (A.12). It follows that $S_N=O_P(N^{-1}+(\varepsilon_N)^2)$. Therefore, we obtain

$$I_k=O_P\left(N^{-1/2}\right)+O_P\left(N^{-1/2}+\varepsilon_N\right)=o_P(1).$$

Hence, $\hat{\Sigma}_1\xrightarrow{p}\Sigma_{10}$.

Step 2. It remains to prove (A.12) and (A.13). By Taylor series expansion,

$$\begin{aligned}\bar{\psi}_1(W,\theta_0,p,\eta_1)-\bar{\psi}_1(W,\theta_0,p_0,\eta_{10})=&\bar{\psi}_1(W,\theta_0,p_0,\eta_1)-\bar{\psi}_1(W,\theta_0,p_0,\eta_{10})\\&+\partial_p\psi_1(W,\theta_0,\bar{p},\eta_1)(p-p_0)\\=&\psi_1(W,\theta_0,p_0,\eta_1)-\psi_1(W,\theta_0,p_0,\eta_{10})\\&+\partial_p\psi_1(W,\theta_0,\bar{p},\eta_1)(p-p_0),\end{aligned}$$

where $\bar{p}\in(p,p_0)$. Then, we have

$$\begin{aligned}\parallel\bar{\psi}_1(W,\theta_0,p,\eta_1)-\bar{\psi}_1(W,\theta_0,p_0,\eta_{10})\parallel_{P,2}\leq&\parallel\psi_1(W,\theta_0,p_0,\eta_1)-\psi_1(W,\theta_0,p_0,\eta_{10})\parallel_{P,2}\\&+\parallel\frac{1}{\bar{p}^2}\frac{D-g(X)}{1-g(X)}(Y(1)-Y(0)-\ell_1(X))\\&+\frac{D\theta_0}{\bar{p}^2}\parallel_{P,2}\times|\,p-p_0\,|.\end{aligned}$$

By (A.1), we have $\|\psi_1(W, \theta_0, p_0, \eta_1) - \psi_1(W, \theta_0, p_0, \eta_{10})\|_{P,2} = O(\varepsilon_N)$. The term in the second line is bounded by

$$\begin{aligned}
&\| \frac{1}{\bar{p}^2} \frac{U + g_0 - g}{1 - g} (U + \ell_{10} - \ell_1) \|_{P,2} + \| \frac{D\theta_0}{\bar{p}^2} \|_{P,2} \\
&\leq \frac{1}{\bar{p}^2\kappa} \| UV_1 \|_{P,2} + \frac{1}{\bar{p}^2\kappa} \| U(\ell_{10} - \ell_1) \|_{P,2} \\
&+ \frac{1}{\bar{p}^2\kappa} \| V_1 (g_0 - g) \|_{P,2} + \frac{1}{\bar{p}^2} | \theta_0 | \\
&+ \frac{1}{\bar{p}^2\kappa} \| g_0 - g_1 \|_{P,2} \| \ell_{10} - \ell_1 \|_{P,2} \\
&\leq \frac{1}{\bar{p}^2\kappa} \left(C + \sqrt{C} \| \ell_{10} - \ell_1 \|_{P,2} + \sqrt{C} \| g_0 - g \|_{P,2} \right) \\
&+ \frac{C}{\bar{p}^2 p_0 \kappa} + \frac{1}{\bar{p}^2\kappa} \| g_0 - g_1 \|_{P,2} \| \ell_{10} - \ell_1 \|_{P,2} \\
&= O(1),
\end{aligned}$$

where I use $\|UV_1\|_{P,2} \leq \|UV_1\|_{P,4} \leq C$, $E_P[U^2|X] \leq C$, $E_P\left[V_1^2 \mid X\right] \leq C$, and

$$\begin{aligned}
| \theta_0 | &= | E_P\left[\frac{Y(1) - Y(0)}{p_0} \frac{D - g_0(X)}{1 - g_0(X)} \right] | \\
&\leq \frac{1}{p_0\kappa} | E_P[(Y(1) - Y(0)) U] | \\
&= \frac{1}{p_0\kappa} | E_P[(\ell_{10}(X) + V_1) U] | \\
&= \frac{1}{p_0\kappa} | E_P[UV_1] | \\
&\leq \frac{C}{p_0\kappa}
\end{aligned}$$

by $|E_P[UV_1]| \leq \|UV_1\|_{P,4} \leq C$. Thus, we obtain

$$\begin{aligned}
\| \bar{\psi}_1(W, \theta_0, p, \eta_1) - \bar{\psi}_1(W, \theta_0, p_0, \eta_{10}) \|_{P,2} &\leq O(\varepsilon_N) + O(1)\, O\left(N^{-1/2}\right) \\
&= O(\varepsilon_N),
\end{aligned}$$

where I assume that ε_N converges to zero no faster than $N^{-1/2}$.

For (A.13),

$$\begin{aligned}
\| \bar{\psi}_1(W, \theta_0, p_0, \eta_{10}) \|_{P,4} &= \| \frac{1}{p_0} \frac{UV_1}{1 - g_0} - \frac{D\theta_0}{p_0} \|_{P,4} \\
&\leq \| \frac{1}{p_0} \frac{UV_1}{1 - g_0} \|_{P,4} + \| \frac{D\theta_0}{p_0} \|_{P,4} \\
&\leq \frac{1}{p_0\kappa} \| UV_1 \|_{P,4} + \frac{1}{p_0} | \theta_0 | \\
&\leq \frac{C}{p_0\kappa} + \frac{C}{p_0^2\kappa}
\end{aligned}$$

because $\|UV_1\|_{P,4} \leq C$.

Repeated cross sections:

In Step 1, I show the main result with the auxiliary results:

$$\sup_{p\in\mathcal{P}_N,\lambda\in\Lambda_N,\eta_2\in\mathcal{T}_N}\left(E_P[\| \bar{\psi}_2\left(W,\theta_0,p,\lambda,G_{2\lambda 0},\eta_2\right)-\bar{\psi}_2\left(W,\theta_0,p_0,\lambda_0,G_{2\lambda 0},\eta_{20}\right)\|^2]\right)^2\leq\varepsilon_N, \quad (A.14)$$

$$\left(E_P\left[\bar{\psi}_2\left(W,\theta_0,p_0,\lambda_0,G_{2\lambda 0},\eta_{20}\right)^4\right]\right)^{1/4}\leq C_2, \quad (A.15)$$

where $(P_N,\Lambda_N,\mathcal{T}_N)$ are specified in the proof of Theorem 3.1, C_2 is a constant, and

$$\bar{\psi}_2\left(W,\theta,p,\lambda,G_{2\lambda},\eta_2\right)\equiv\frac{1}{\lambda\left(1-\lambda\right)p}\frac{D-g\left(X\right)}{1-g\left(X\right)}\left(\left(T-\lambda\right)Y-\ell_2\left(X\right)\right)-\frac{D\theta}{p}+G_{2\lambda}\left(T-\lambda\right).$$

In fact, we have $E_P\left[\left(\bar{\psi}_2\left(W,\theta_0,p_0,\lambda_0,G_{2\lambda_0},\eta_{20}\right)\right)^2\right]=\Sigma_{20}$. In Step 2, I prove (A.14) and (A.15).

Step 1. Notice that

$$\begin{aligned}\hat{\Sigma}_2=&\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\left(\psi_2\left(W,\tilde{\theta},\hat{p}_k,\hat{\eta}_{1k}\right)+\hat{G}_{2p}\left(D-\hat{p}_k\right)+\hat{G}_{2\lambda}\left(T-\hat{\lambda}_k\right)\right)^2\right]\\=&\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{n,k}\left[\bar{\psi}_2\left(W,\tilde{\theta},\hat{p}_k,\hat{\lambda}_k,\hat{G}_{2\lambda},\hat{\eta}_{2k}\right)^2\right],\end{aligned}$$

where the second inequality follows from $\hat{G}_{2p}=-\tilde{\theta}/\hat{p}_k$.

Because *K* is fixed, which is independent of *N*, it suffices to show that

$$J_k\equiv|\mathbb{E}_{n,k}\left[\bar{\psi}_2\left(W,\tilde{\theta},\hat{p}_k,\hat{\lambda}_k,\hat{G}_{2\lambda},\hat{\eta}_{2k}\right)^2\right]-E_P\left[\bar{\psi}_2\left(W,\theta_0,p_0,\lambda_0,G_{2\lambda 0},\eta_{20}\right)^2\right]|=o_P\left(1\right).$$

By the triangle inequality, we have

$$J_k\leq J_{5,k}+J_{6,k},$$

where

$$J_{5,k}\equiv|\mathbb{E}_{n,k}\left[\bar{\psi}_2\left(W,\tilde{\theta},\hat{p}_k,\hat{\lambda}_k,\hat{G}_{2\lambda},\hat{\eta}_{2k}\right)^2\right]-\mathbb{E}_{n,k}\left[\bar{\psi}_2\left(W,\theta_0,p_0,\lambda_0,G_{2\lambda 0},\eta_{20}\right)^2\right]|,$$

$$J_{6,k}\equiv|\mathbb{E}_{n,k}\left[\bar{\psi}_2\left(W,\theta_0,p_0,\lambda_0,G_{2\lambda 0},\eta_{20}\right)^2\right]-E_P\left[\bar{\psi}_2\left(W,\theta_0,p_0,\lambda_0,G_{2\lambda 0},\eta_{20}\right)^2\right]|.$$

Using the same arguments for $I_{3,k}$ and $I_{4,k}$ in the proof of repeated outcomes and the conditions (A.14) and (A.15), we can show $J_{5,k}=o_P(1)$ and $J_{6,k}=o_P(1)$. Hence, $\hat{\Sigma}_2\xrightarrow{p}\Sigma_{20}$.

Step 2. It remains to show (A.14) and (A.15). Define $\bar{\psi}_{20}\equiv\bar{\psi}_2\left(W,\theta_0,p_0,\lambda_0,G_{2\lambda 0},\eta_{20}\right)$. By the triangle inequality and

$$\bar{\psi}_2\left(W,\theta_0,p_0,\lambda_0,G_{2\lambda 0},\eta_2\right)-\bar{\psi}_{20}=\psi_2\left(W,\theta_0,p_0,\lambda_0,\eta_2\right)-\psi_2\left(W,\theta_0,p_0,\lambda_0,\eta_{20}\right),$$

we have

$$\begin{aligned}\|\bar{\psi}_2\left(W,\theta_0,p,\lambda,G_{2\lambda 0},\eta_2\right)-\bar{\psi}_{20}\|_{P,2}\leq&\|\psi_2\left(W,\theta_0,p_0,\lambda_0,\eta_2\right)-\psi_2\left(W,\theta_0,p_0,\lambda_0,\eta_{20}\right)\|_{P,2}\\&+\|\partial_\lambda\bar{\psi}_2\left(W_i,\theta_0,p_0,\bar{\lambda},G_{2\lambda 0},\eta_2\right)\|_{P,2}|\lambda-\lambda_0|\\&+\|\partial_p\bar{\psi}_2\left(W_i,\theta_0,\bar{p},\lambda,G_{2\lambda 0},\eta_2\right)\|_{P,2}|p-p_0|,\end{aligned}$$

where $\bar{p} \in (p, p_0)$ and $\bar{\lambda} \in (\lambda, \lambda_0)$. The term in the second line is bounded by

$$\begin{aligned}\| \partial_\lambda \bar{\psi}_2 \left(W_i, \theta_0, p_0, \bar{\lambda}, G_{2\lambda 0}, \eta_2\right) \|_{P,2} \leq & \frac{| 1 - 2\bar{\lambda} |}{p_0 \bar{\lambda}^2 \left(1 - \bar{\lambda}\right)^2} \| \frac{D - g(X)}{1 - g(X)}((T - \bar{\lambda})Y - \ell_2(X)) \|_{P,2} \\ & + \frac{1}{p_0 \bar{\lambda} \left(1 - \bar{\lambda}\right)} \| \frac{D - g(X)}{1 - g(X)} \times Y \|_{P,2} + | G_{2\lambda 0} | \\ \leq & O(1)\end{aligned}$$

by the same arguments as in (A.9) through (A.11), and

$$\begin{aligned}| G_{2\lambda 0} | = & | E_P \left[- \frac{1 - 2\lambda_0}{\lambda_0^2 (1 - \lambda_0)^2 p_0} \frac{D - g_0}{1 - g_0} ((T - \lambda_0) Y - \ell_{20}) - \frac{Y}{\lambda_0 (1 - \lambda_0) p_0} \frac{D - g_0}{1 - g_0} \right] | \\ \leq & \frac{| 1 - 2\lambda_0 |}{\lambda_0^2 (1 - \lambda_0)^2 p_0 \kappa} | E_P [U V_2] | + \frac{1}{\lambda_0 (1 - \lambda_0) p_0 \kappa} | E_P [Y U] | \\ \leq & \frac{| 1 - 2\lambda_0 |}{\lambda_0^2 (1 - \lambda_0)^2 p_0 \kappa} C + \frac{1}{\lambda_0 (1 - \lambda_0) p_0 \kappa} C \\ = & O(1)\end{aligned}$$

because $|E_P[UV_2]| \leq \| UV_2 \|_{P,4} \leq C$ and $|E_P[YU]| \leq C$. Also, we have

$$\begin{aligned}\| \partial_p \bar{\psi}_2 (W_i, \theta_0, \bar{p}, \lambda, G_{2\lambda 0}, \eta_2) \|_{P,2} \leq & \frac{1}{\lambda (1 - \lambda) \bar{p}^2} \| \frac{D - g(X)}{1 - g(X)} ((T - \lambda) Y - \ell_2 (X)) \|_{P,2} \\ & + \| \frac{D\theta_0}{\bar{p}^2} \|_{P,2} \\ \leq & O(1)\end{aligned}$$

by the same arguments as in (A.9) through (A.11), and

$$\begin{aligned}| \theta_0 | = & | E_P \left[\frac{D - g_0 (X)}{p_0' (1 - g_0 (X))} (T - \lambda_0) Y \right] | \\ \leq & \frac{1}{p_0' \kappa} | E_P [(T - \lambda_0) Y U] | \\ = & \frac{1}{p_0 \kappa} | E_P [(\ell_{20} (X) + V_2) U] | \\ = & \frac{1}{p_0 \kappa} | E_P [U V_2] | \\ \leq & \frac{C}{p_0 \kappa}\end{aligned}$$

because $|E_P[UV_2]| \leq \| UV_2 \|_{P,4} \leq C$. Together with (A.5), we have

$$\begin{aligned}\| \bar{\psi}_2 (W, \theta_0, p, \lambda, G_{2\lambda 0}, \eta_2) - \bar{\psi}_{20} \|_{P,2} \leq & O(\varepsilon_N) + O(1) O\left(N^{-1/2}\right) + O(1) O\left(N^{-1/2}\right) \\ = & O(\varepsilon_N),\end{aligned}$$

where I assume that ε_N converges to zero no faster than $N^{-1/2}$.

For (A.15), we have

$$\| \bar{\psi}_2 (W, \theta_0, p_0, \lambda_0, G_{2\lambda 0}, \eta_{20}) \|_{P,4} = \| \frac{1}{\lambda_0 (1-\lambda_0) p_0} \frac{UV_2}{1-g_0} - \frac{D\theta_0}{p_0} + G_{2\lambda 0} (T - \lambda_0) \|_{P,4}$$

$$\leq \frac{1}{\lambda_0 (1-\lambda_0) p_0 \kappa} \| UV_2 \|_{P,4} + \frac{1}{p_0} | \theta_0 | + | G_{2\lambda 0} |$$

$$\leq O(1)$$

because $\|UV_2\|_{P,4} \leq C$. □ □

LEMMA A.1 (CONDITIONAL CONVERGENCE IMPLIES UNCONDITIONAL). *Let $\{X_m\}$ and $\{Y_m\}$ be sequences of random vectors. (i) If for $\epsilon_m \to 0$, $Pr(\| X_m \| > \epsilon_m \mid Y_m) \xrightarrow{p} 0$, then $Pr(\|X_m\| > \epsilon_m) \to 0$. This occurs if $E\left[\| X_m \|^q / \epsilon_m^q \mid Y_m\right] \xrightarrow{p} 0$ for some $q \geq 1$, by Markov's inequality. (ii) Let $\{A_m\}$ be a sequence of positive constants. If $\|X_m\| = O_P(A_m)$ conditional on Y_m, namely, that for any $\ell_m \to \infty$, $Pr(\| X_m \| > \ell_m A_m \mid Y_m) \xrightarrow{p} 0$, then $\|X_m\| = O_P(A_m)$ unconditionally, namely, that for any $\ell_m \to \infty$, $Pr(\| X_m \| > \ell_m A_m) \to 0$.*

PROOF: This lemma is the Lemma 6.1 in Chernozhukov et al. (2018).

Abadie, A. (2005). Semiparametric difference-in-differences estimators, *Review of Economic Studies*, *72*, 1–19.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters, *Econometrics Journal*, *21*, C1–C68.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing, *Handbook of Econometrics*, *4*, 2111–245.