Difference-in-Differences with Compositional Changes: Supplemental Appendix

Pedro H. C. Sant'Anna* Qi Xu†

November 14, 2025

This supplemental appendix contains auxiliary lemmas, proofs of the main theorems, and additional results presented in the main text.

A Notations and Definitions

Hereafter, we use the abbreviations CLT, CMT, LIE, and LLN to represent the central limit theorem, continuous mapping theorem, law of iterated expectations, and law of large numbers, respectively. Let $(s.o.)$ stands for smaller/higher order terms. Let $\mathbf{1}_k = (1, \ldots, 1)^\top$ denote the k-dimensional vector of ones. Define $f_X(x) = f_{X_c|X_d}(x_c|x_d) \cdot \mathbb{P}(X_d = x_d)$, $\mathbb{N}_n = \{1, 2, ..., n\}$, and $\iota(d,t) = \mathbb{1}\{d = 1, t = 0\} + 2 \cdot \mathbb{1}\{d = 0, t = 1\} + 3 \cdot \mathbb{1}\{d = 0, t = 0\}$. We denote the ATT by τ, i.e.,

$$ATT = \tau = \mathbb{E}\left[Y_1(1)\,|D = 1, T = 1\right] - \mathbb{E}\left[Y_1(0)\,|D = 1, T = 1\right].$$

We write $f \in L_2(\mathcal{U})$ if $\int_{\mathcal{U}} f^2 d\mu < \infty$, and denote by $\|f\|_{L_2}$ and $\|f\|_\infty$ the L_2- and *sup-norm* of f, respectively. The *covering number* $N(\varepsilon, \mathcal{F}, L_2(Q))$ is the minimal number of $L_2(Q)$-balls of radius ε needed to cover the function class $\mathcal{F}$. The *uniform covering numbers* is defined as

$$\sup_Q N(\varepsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)),$$

for a given envelop function F of the class $\mathcal{F}$,where the supremum is taken over all finitely-discrete probability measures Q for which $\mathcal{F}$ is not identically zero (and hence $\|F\|_Q^2 = \int F^2 dQ > 0$). The *uniform entropy integral* is defined as

$$J(\delta, \mathcal{F}, L_2) = \int_0^\delta \sqrt{\log \sup_Q N(\varepsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q))}\, d\varepsilon.$$

For a measurable function $f : \mathcal{X} \to \mathbb{R}$, define the empirical process

$$\mathbb{G}_n f = \sqrt{n}\big(\mathbb{P}_n f - \mathbb{P} f\big) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\right),$$

where $\mathbb{P}_n$ is the empirical distribution and $\mathbb{P}$ is the true distribution.

*Emory University. E-mail: pedro.santanna@emory.edu

†University of Florida. E-mail: qixu@ufl.edu.

A class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$ is called *P-Donsker* if the sequence of processes $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ converges in distribution to a tight limit process in the space $\ell^\infty(\mathcal{F})$.

Consider the sequence of empirical processes $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ indexed by a class of functions $\mathcal{F}$ equipped with a semimetric $\rho_{\mathcal{F}}$. We say that this sequence is *stochastically equicontinuous* if, for every $\eta > 0$ and $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}^* \left(\sup_{\rho_{\mathcal{F}}(f_1,f_2)<\delta} \left|\mathbb{G}_n f_1 - \mathbb{G}_n f_2\right| > \eta \right) < \varepsilon,$$

where $\mathbb{P}^*$ denotes outer probability. An equivalent characterization of stochastic equicontinuity, stated in Andrews (1994), is that the process $\{G_n f : f \in \mathcal{F}\}$ is stochastically equicontinuous at f_0 if, for every sequence of random elements $\{f_n : n \geqslant 1\}$ satisfying $\rho_{\mathcal{F}}(f_n, f_0) = o_p(1)$, we have $\mathbb{G}_n \{f_n - f_0\} = o_p(1)$.

B Proofs for main results in the text

Let

$$\tau_{or} = \mathbb{E}\left[Y \middle| D = 1, T = 1\right] - \mathbb{E}\left[m_{1,0}(X) + m_{0,1}(X) - m_{0,0}(X) \middle| D = 1, T = 1\right],$$

where $m_{d,t}(x) = E[Y|D = d, T = t, X = x]$, and

$$\tau_{ipw} = \mathbb{E}\left[\left(w_{1,1}(D,T) - w_{1,0}(D,T,X) - w_{0,1}(D,T,X) + w_{0,0}(D,T,X)\right) Y\right],$$

where, for $(d,t) \in \mathcal{S}_-$,

$$\begin{aligned} w_{1,1}(D,T) &= \frac{DT}{\mathbb{E}[DT]}, \\ w_{d,t}(D,T,X) &= \frac{\mathbb{1}\{D=d,T=t\}p(1,1,X)}{p(d,t,X)} \Bigg/ \mathbb{E}\left[\frac{\mathbb{1}\{D=d,T=t\}p(1,1,X)}{p(d,t,X)}\right], \end{aligned}$$

and $p(d,t,x) = \mathbb{P}\left(D = d, T = t | X = x\right)$ is a so-called generalized propensity score.

Lemma B.1 Under Assumptions 1 and 2, it follows that $\tau_{or} = \tau_{ipw} = \tau$.

Proof of Lemma B.1:
Outcome regression estimand: Using $m_{d,t}(\cdot) = \mathbb{E}[Y_t(d)|D = d, T = t, X = \cdot]$, $(d,t) \in \mathcal{S}_-$, we get

$$\begin{aligned} \tau_{or} &= \mathbb{E}[Y_1(1)|D=1,T=1] - \mathbb{E}[\mathbb{E}[Y_0(1)|D=1,T=0,X=x]|D=1,T=1] \\ &\quad + \sum_{t\in\{0,1\}} (-1)^t\, \mathbb{E}[\mathbb{E}[Y_t(0)|D=0,T=t,X=x]|D=1,T=1] \\ &= \mathbb{E}[Y_1(1)|D=1,T=1] - \mathbb{E}[\mathbb{E}[Y_0(0)|D=1,T=0,X=x]|D=1,T=1] \\ &\quad + \sum_{t\in\{0,1\}} (-1)^t\, \mathbb{E}[\mathbb{E}[Y_t(0)|D=0,T=t,X=x]|D=1,T=1] \\ &= \mathbb{E}[Y_1(1) - Y_1(0)|D=1,T=1] = \tau, \end{aligned}$$

where the second equality follows from Assumptions 2(ii) and the third holds under Assumptions 2(i).

Propensity score estimand: Let $p(1,1) = \mathbb{P}\left(D = 1, T = 1\right)$. Under the overlapping conditions in Assumption 2(iii), $w_{d,t}(d', t', x)$ are well defined for $(d,t) \in \mathcal{S}_-$, $(d',t') \in \{0,1\}^2$, and $x \in \mathcal{X}$ almost everywhere. Additionally,

$$
\begin{aligned}
\mathbb{E}[w_{d,t}(D,T,X)Y] &= \mathbb{E}\left[\frac{p(1,1,X)YI_{d,t}}{p(d,t,X)}\right] \Big/ \mathbb{E}\left[\frac{\mathbb{1}\{D=d,T=t\}p(1,1,X)}{p(d,t,X)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}[Y|D=d,T=t,X]\cdot\frac{I_{d,t}}{p(d,t,X)}\Big|X\right]\cdot\frac{p(1,1,X)}{p(1,1)}\right] \\
&= \mathbb{E}\left[\mathbb{E}[Y|D=d,T=t,X]\cdot\frac{p(1,1,X)}{p(1,1)}\right] \\
&= \mathbb{E}\left[\mathbb{E}[Y|D=d,T=t,X]|D=1,T=1\right] \\
&= \mathbb{E}\left[m_{d,t}(X)|D=1,T=1\right],
\end{aligned}
$$

for $(d,t)\in\mathcal{S}_-$. The second line follows by the LIE, the third equality is by the definition of propensity scores, and the next to last line is by Bayes' Law. Next, from $\mathbb{E}[w_{1,1}(D,T)Y]=\mathbb{E}[Y|D=1,T=1]$ and the same arguments for the OR estimand, we conclude that $\tau_{ipw}=\tau$. ■

Proof of Theorem 2.1:
We follow the steps in Hahn (1998) for the derivation of the efficient influence function. Let $f(y|d,t,x)=f(y|D=d,T=t,X=x)$ and $f(x|d,t)=f(x|D=d,T=t)$.

Step 1: characterize the tangent space of the statistical model. The observed likelihood is given as

$$
\begin{aligned}
f(y,d,t,x) =& f(y|1,1,x)^{dt}f(y|1,0,x)^{d(1-t)}f(y|0,1,x)^{(1-d)t}f(y|0,0,x)^{(1-d)(1-t)} \\
&\cdot p(1,1,x)^{dt}p(1,0,x)^{d(1-t)}p(0,1,x)^{(1-d)t}p(0,0,x)^{(1-d)(1-t)}\cdot f(x).
\end{aligned}
$$

Consider the regular sub-model parameterized by $\theta\geqslant 0$, with the true model indexed by $\theta_0=0$,

$$
\begin{aligned}
f_\theta(y,d,t,x) =& f_\theta(y|1,1,x)^{dt}f_\theta(y|1,0,x)^{d(1-t)}f_\theta(y|0,1,x)^{(1-d)t}f_\theta(y|0,0,x)^{(1-d)(1-t)} \\
&\cdot p_\theta(1,1,x)^{dt}p_\theta(1,0,x)^{d(1-t)}p_\theta(0,1,x)^{(1-d)t}p_\theta(0,0,x)^{(1-d)(1-t)} \\
&\cdot f_\theta(x).
\end{aligned}
$$

The score function of this sub-model is given by

$$
\begin{aligned}
s_\theta(y,d,t,x) =& dts_\theta(y|1,1,x)+d(1-t)s_\theta(y|1,0,x)+(1-d)ts_\theta(y|0,1,x)+(1-d)(1-t)s_\theta(y|0,0,x) \\
&+dt\frac{\dot{p}_\theta(1,1,x)}{p_\theta(1,1,x)}+d(1-t)\frac{\dot{p}_\theta(1,0,x)}{p_\theta(1,0,x)}+(1-d)t\frac{\dot{p}_\theta(0,1,x)}{p_\theta(0,1,x)}+(1-d)(1-t)\frac{\dot{p}_\theta(0,0,x)}{p_\theta(0,0,x)} \\
&+t_\theta(x),
\end{aligned}
$$

where $s_\theta(y|d,t,x)=\partial\log f_\theta(y|d,t,x)/\partial\theta$, $\dot{p}_\theta(d,t,x)=\partial p_\theta(d,t,x)/\partial\theta$, and $t_\theta(x)=\partial\log f_\theta(x)/\partial\theta$ for $(d,t)\in\mathcal{S}$. For notational simplicity, we suppress subscripts when $\theta=\theta_0$.

Now, the tangent space of this model is characterized by

$$
\begin{aligned}
\mathcal{T} =& \{dts_{11}(y,x)+d(1-t)s_{10}(y,x)+(1-d)ts_{01}(y,x)+(1-d)(1-t)s_{00}(y,x) \\
&+dtp_{11}(x)+d(1-t)p_{10}(x)+(1-d)tp_{01}(x)+(1-d)(1-t)p_{00}(x)+l(x)\},
\end{aligned}
$$

for any functions $\{s_{dt}(\cdot,\cdot),p_{dt}(\cdot)\}_{(d,t)\in\mathcal{S}}$, and $l(\cdot)$ such that, for $(d,t)\in\mathcal{S}$

$$
s_{dt}(\cdot,\cdot)\in L_2(\mathcal{Y}\times\mathcal{X}),\ \text{with}\ \int s_{dt}(y,x)f(y|d,t,x)dy=0,\ \forall x\in\mathcal{X}, \tag{B.1}
$$

$$p_{dt}(\cdot) \in L_2(\mathcal{X}), \text{ with } \sum_{(d,t)\in\mathcal{S}} p_{dt}(x)p(d,t,x) = 0, \ \forall x \in \mathcal{X}, \tag{B.2}$$

and

$$l(\cdot) \in L_2(\mathcal{X}), \text{ with } \int l(x)f(x)dx = 0. \tag{B.3}$$

In Step 2, we show that the target parameter associated with the parametric sub-model is *path-wise differentiable*, as defined in Newey (1990).

From Lemma B.1, we know the ATT can be identified by $\sum_{(d,t)\in\mathcal{S}}(-1)^{d+t}\,\mathbb{E}\left[\mathbb{E}[Y|D=d,T=t,X]|D=1,T=1\right]$ under Assumptions 1 and 2. For the parameterized sub-model, we define

$$\begin{aligned}\tau(\theta) = &\frac{(\iint p_\theta(1,1,x)yf_\theta(y|1,1,x)f_\theta(x)dydx - \iint p_\theta(1,1,x)yf_\theta(y|1,0,x)f_\theta(x)dydx)}{\int p_\theta(1,1,x)f_\theta(x)dx} \\ &- \frac{(\iint p_\theta(1,1,x)yf_\theta(y|0,1,x)f_\theta(x)dydx - \iint p_\theta(1,1,x)yf_\theta(y|0,0,x)f_\theta(x)dydx)}{\int p_\theta(1,1,x)f_\theta(x)dx}.\end{aligned} \tag{B.4}$$

Note that the derivative of $\tau(\theta)$ with respect to θ, evaluated at $\theta = 0$, is given by

$$\begin{aligned}\left.\frac{d\tau(\theta)}{d\theta}\right|_{\theta=0} = &\sum_{(d,t)\in\mathcal{S}} (-1)^{d+t}\frac{\iint yp(1,1,x)s(y|d,t,x)f(y|d,t,x)f(x)dydx}{p(1,1)} \\ &+ \frac{\int(\tau(x)-\tau)\dot{p}(1,1,x)f(x)dx}{p(1,1)} \\ &+ \frac{\int(\tau(x)-\tau)p(1,1,x)t(x)f(x)dx}{p(1,1)}.\end{aligned}$$

For any $w = (y,d,t,x) \in \mathcal{W}$, define

$$\begin{aligned}F_\tau(w) = &\frac{dt(y-m_{1,1}(x))}{p(1,1)} + \frac{p(1,1,x)}{p(1,1)}\left\{-\frac{d(1-t)(y-m_{1,0}(x))}{p(1,0,x)}\right. \\ &\left.-\frac{(1-d)t(y-m_{0,1}(x))}{p(0,1,x)} + \frac{(1-d)(1-t)(y-m_{0,0}(x))}{p(0,0,x)}\right\} \\ &+ \frac{dt}{p(1,1)}\sum_{(d,t)\in\mathcal{S}}(-1)^{d+t}\left(m_{d,t}(x) - \int_{\mathcal{X}} m_{d,t}(x)f(x|1,1)dx\right).\end{aligned}$$

It can be readily verified that $\left.\frac{d\tau(\theta)}{d\theta}\right|_{\theta=0} = \mathbb{E}[F_\tau(W)s_0(Y,D,T,X)]$, thereby showing $\tau(\theta)$ is path-wise differentiable.

In Step 3, we show that $F_\tau(W)$ is the efficient influence function for τ, which we will accomplish by invoking Theorem 3.1 in Newey (1990). To apply this theorem, we need to verify that $F_\tau(\cdot) \in \mathcal{T}$. Let $c_m(x) = p(1,1)^{-1}\cdot\sum_{(d,t)\in\mathcal{S}}(-1)^{d+t}\left(m_{d,t}(x) - \int_{\mathcal{X}} m_{d,t}(x)f(x|1,1)dx\right)$. By setting

$$\begin{aligned}&s_{11}(y,x) = \frac{y-m_{1,1}(x)}{p(1,1)}, \\ &p_{11}(x) = (1-p(1,1,x))\cdot c_m(x), \\ &s_{dt}(y,x) = (-1)^{d+t}\cdot\frac{p(1,1,x)(y-m_{d,t}(x))}{p(d,t,x)p(1,1)}, \text{ for } (d,t)\in\mathcal{S}_-, \\ &p_{dt}(x) = -p(1,1,x)\cdot c_m(x), \text{ for } (d,t)\in\mathcal{S}_-, \\ &l(x) = p(1,1,x)\cdot c_m(x),\end{aligned}$$

it is straightforward to show that (B.1)–(B.3) hold. Here, we show (B.3) explicitly,

$$\begin{aligned}\int_{\mathcal{X}} l(x)f(x)dx &= \int_{\mathcal{X}} \frac{p(1,1,x)}{p(1,1)} \left\{ \sum_{(d,t)\in\mathcal{S}} (-1)^{d+t} \left(m_{d,t}(x) - \int_{\mathcal{X}} m_{d,t}(x) f(x|1,1)dx \right) \right\} f(x)dx \\ &= \int_{\mathcal{X}} f(x|1,1) \left\{ \sum_{(d,t)\in\mathcal{S}} (-1)^{d+t} m_{d,t}(x) \right\} dx - \tau \\ &= \tau - \tau = 0.\end{aligned}$$

Thus, $F_\tau(\cdot) \in \mathcal{T}$.

Finally, since $p(1,1) = \mathbb{E}\left[I_{d,t}p(1,1,X)p(d,t,X)^{-1}\right]$, for $(d,t) \in \mathcal{S}$, direct manipulation yields that $F_\tau(W) = \eta_{\text{eff}}(W)$. Now, we take the expectation of $\eta^2_{\text{eff}}(W)$ and the semi-parametric efficiency bound follows by standard manipulation. This completes the proof. ■

Proof of Proposition 2.1: The proof follows directly from the LIE as displayed in the main text. ■

Proof of Proposition 2.2:
It follows by Theorem 2.1 that

$$\begin{aligned}\mathbb{E}[\eta_{\text{eff}}(W)^2] &= \frac{1}{\mathbb{E}[DT]^2}\,\mathbb{E}\left[DT(\tau_{dr}(Y,X)-\tau)^2 + \sum_{(d,t)\in\mathcal{S}_-} \frac{I_{d,t}p(1,1,X)^2}{p(d,t,X)^2}(Y-m_{d,t}(X))^2 \right] \\ &= \frac{1}{\mathbb{E}[DT]^2}\,\mathbb{E}[DT(\tau(X)-\tau)^2] \\ &\quad + \mathbb{E}\left[w_{1,1}(D,T)^2(Y-m_{1,1}(X))^2 + \sum_{(d,t)\in\mathcal{S}_-} w_{d,t}(D,T,X,p)^2(Y-m_{d,t}(X))^2 \right] \\ &\equiv V_{1,dr} + V_{2,dr},\end{aligned}$$

where the second equality follows from direct manipulations and the fact that

$$\begin{aligned}&\mathbb{E}[DT\cdot(Y-m_{1,1}(X))\cdot(m_{d,t}(X)-\mathbb{E}[m_{d,t}(X)|D=1,T=1])] \\ &\quad = \mathbb{E}[\mathbb{E}[p(1,1,X)\cdot(m_{1,1}(X)-m_{1,1}(X))\cdot(m_{d,t}(X)-\mathbb{E}[m_{d,t}(X)|D=1,T=1])|X]] = 0,\end{aligned}$$

for $(d,t) \in \mathcal{S}$.

Meanwhile, from Part (b) of Proposition 1 in Sant'Anna and Zhao (2020), we have the following decomposition,

$$\mathbb{E}[\eta_{sz}(W)^2] = V_{1,sz} + V_{2,sz},$$

where $V_{1,sz} \equiv \mathbb{E}[D]^{-2}\cdot\mathbb{E}\left[D(\tau(X)-\tau)^2\right]$, $\tau(x) = (m_{1,1}(x)-m_{1,0}(x)) - (m_{0,1}(x)-m_{0,0}(x))$, and

$$\begin{aligned}V_{2,sz} \equiv& \,\mathbb{E}[D]^{-2}\cdot\mathbb{E}\left[\frac{DT}{\mathbb{E}[T]^2}(Y-m_{1,1}(X))^2 + \frac{D(1-T)}{(1-\mathbb{E}[T])^2}(Y-m_{1,0}(X))^2 \right. \\ &\left. + \frac{(1-D)T\tilde{p}(X)^2}{(1-\tilde{p}(X))^2\,\mathbb{E}[T]^2}(Y-m_{0,1}(X))^2 + \frac{(1-D)(1-T)\tilde{p}(X)^2}{(1-\tilde{p}(X))^2(1-\mathbb{E}[T])^2}(Y-m_{0,0}(X))^2\right].\end{aligned} \tag{B.5}$$

Under Assumption 3, we have that $\mathbb{E}[\mathbb{1}\{T=t\}g(X)] = \mathbb{P}(T=t)\,\mathbb{E}[g(X)]$, $\mathbb{E}[I_{d,t}Yg(X)] = \mathbb{P}(T=t)\,\mathbb{E}[\mathbb{1}\{D=d\}Y_t g(X)]$, and $p(d,t,x) = (\mathbb{1}\{t=1\}\,\mathbb{E}[T] + \mathbb{1}\{t=0\}(1-\mathbb{E}[T]))\cdot$

$(\mathbb{1}\{d=1\}\tilde{p}(x)+\mathbb{1}\{d=0\}(1-\tilde{p}(x)))$. It then follows that

$$V_{1,dr}=\frac{1}{\mathbb{E}[D]^2\,\mathbb{E}[T]}\,\mathbb{E}\left[D(\tau(X)-\tau)^2\right], \tag{B.6}$$

and therefore,

$$V_{1,dr}-V_{1,sz}=\frac{1-\mathbb{E}[T]}{\mathbb{E}[D]^2\,\mathbb{E}[T]}\,\mathbb{E}\left[D(\tau(X)-\tau)^2\right]. \tag{B.7}$$

We now focus on $V_{2,dr}$. Observe that

$$\begin{aligned}V_{2,dr}=&\frac{1}{\mathbb{E}[T]^2\,\mathbb{E}[D]^2}\left\{\mathbb{E}[DT(Y_1-m_{1,1}(X))^2]+\mathbb{E}\left[\frac{D(1-T)\,\mathbb{E}[T]^2\tilde{p}(X)^2}{(1-\mathbb{E}[T])^2\tilde{p}(X)^2}(Y_0-m_{1,0}(X))^2\right]\right.\\&\left.+\,\mathbb{E}\left[\frac{(1-D)T\,\mathbb{E}[T]^2\tilde{p}(X)^2}{\mathbb{E}[T]^2(1-\tilde{p}(X))^2}(Y_1-m_{0,1}(X))^2\right]+\mathbb{E}\left[\frac{(1-D)(1-T)\,\mathbb{E}[T]^2\tilde{p}(X)^2}{(1-\mathbb{E}[T])^2(1-\tilde{p}(X))^2}(Y_0-m_{0,0}(X))^2\right]\right\}\\=&\frac{1}{\mathbb{E}[D]^2}\,\mathbb{E}\left[\frac{DT}{\mathbb{E}[T]^2}(Y-m_{1,1}(X))^2+\frac{D(1-T)}{(1-\mathbb{E}[T])^2}(Y-m_{1,0}(X))^2\right.\\&\left.+\,\frac{(1-D)T\tilde{p}(X)^2}{(1-\tilde{p}(X))^2\,\mathbb{E}[T]^2}(Y-m_{0,1}(X))^2+\frac{(1-D)(1-T)\tilde{p}(X)^2}{(1-\tilde{p}(X))^2(1-\mathbb{E}[T])^2}(Y-m_{0,0}(X))^2\right]=V_{2,sz},\end{aligned} \tag{B.8}$$

where the first equality follows again by Assumption 3. The desired result then follows from (B.7) and (B.8). ■

Lemma B.2 (Generic Convergence Rates under Stochastic Equicontinuity)
Suppose the convergence rates r_n and s_n from (3.4)–(3.5) hold for $e=\infty$, and that Assumptions 1, 2, 4.1, and 4.3 are satisfied. In addition, assume the following conditions hold for all $(d,t)\in\mathcal{S}_-$:

$$\begin{aligned}&(i)\qquad \mathbb{G}_n\left\{(Y-m_{d,t}(X))\cdot(\widehat{w}_{d,t}-w_{d,t})(W)\right\}=o_p(1).\\&(ii)\qquad \mathbb{G}_n\left\{(w_{1,1}-w_{d,t})(W)\cdot(\widehat{m}_{d,t}-m_{d,t})(X)\right\}=o_p(1).\end{aligned}$$

Then, the expansion in (3.6) holds. Furthermore, if the following conditions also hold for all $(d,t)\in\mathcal{S}_-$:

$$\begin{aligned}&(iii)\qquad \mathbb{G}_n\left\{I_{d,t}\cdot\left(\frac{\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)}-\frac{p(1,1,X)}{p(d,t,X)}\right)\cdot(\widehat{m}_{d,t}-m_{d,t})(X)\right\}=o_p(1).\\&(iv)\qquad \mathbb{G}_n\left\{w_{d,t}(W)\cdot(\widehat{m}_{d,t}-m_{d,t})(X)\right\}=o_p(1).\\&(v)\qquad \mathbb{G}_n\left\{I_{d,t}\cdot\left(\frac{\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)}-\frac{p(1,1,X)}{p(d,t,X)}\right)\right\}=o_p(1).\end{aligned}$$

Then expansion in (3.6) remains valid when (3.4)–(3.5) hold for $e=L_2$.

Proof of Lemma B.2:
Let

$$\begin{aligned}\psi_{d,t}(W;w,m)=\;&\mathbb{1}\{dt=1\}w_{1,1}(D,T)Y+\mathbb{1}\{dt\neq 1\}\\&\cdot\left\{w_{d,t}(D,T,X)(Y-m_{d,t}(X))+w_{1,1}(D,T)m_{d,t}(X)\right\},\end{aligned} \tag{B.9}$$

and $\tilde{\tau}_{dr} = \sum_{(d,t)\in\mathcal{S}}(-1)^{d+t}\,\mathbb{E}_n\left[\psi_{d,t}(W;w,m)\right]$. Using $\tilde{\tau}_{dr}$, we decompose $\hat{\tau}_{dr}$ as

$$\hat{\tau}_{dr} - \tau = (\hat{\tau}_{dr} - \tilde{\tau}_{dr}) + (\tilde{\tau}_{dr} - \tau)\,. \tag{B.10}$$

Note first that the second term, $\tilde{\tau}_{dr} - \tau$, has *i.i.d.* centered summands with bounded variance under Assumptions 2(iii) and 4.3; thus, it is $O_p(n^{-1/2})$. Now we investigate the behavior of $\hat{\tau}_{dr} - \tilde{\tau}_{dr}$, for which we make the following decomposition

$$\begin{aligned}\psi_{d,t}(W;\hat{w},\hat{m}) - \psi_{d,t}(W;w,m) =& (Y - m_{d,t}(X))\left(\hat{w}_{d,t} - w_{d,t}\right)(W) + m_{d,t}(X)\left(\hat{w}_{1,1} - w_{1,1}\right)(W) & \text{(B.11)}\\ &+ \left(w_{1,1} - w_{d,t}\right)(W)\left(\hat{m}_{d,t} - m_{d,t}\right)(X) & \text{(B.12)}\\ &+ \left\{\left(\hat{w}_{1,1} - w_{1,1}\right)(W) - \left(\hat{w}_{d,t} - w_{d,t}\right)(W)\right\}\left(\hat{m}_{d,t} - m_{d,t}\right)(X) & \text{(B.13)}\\ \equiv& \Delta^{\psi,1}_{d,t}(W) + \Delta^{\psi,2}_{d,t}(W) + \Delta^{\psi,3}_{d,t}(W),\end{aligned}$$

for $(d,t)\in\mathcal{S}$. Here, we use the unifying notation $w_{d,t}(W)$ to denote $w_{d,t}(D,T,X)$ when $(d,t)\in\mathcal{S}_-$ and $w_{1,1}(D,T)$ otherwise. We proceed by establishing convergence rates for each component in the above decomposition.

We first analyze $\Delta^{\psi,1}_{d,t}$. A second-order Taylor expansion of $\psi_{1,1}(W;\hat{w},\hat{m})$ around $\mathbb{E}[DT]$ yields that

$$\begin{aligned}\mathbb{E}_n\left[\Delta^{\psi,1}_{1,1}(W)\right] &= \mathbb{E}_n\left[Y\left(\frac{DT}{\mathbb{E}_n[DT]} - \frac{DT}{\mathbb{E}[DT]}\right)\right]\\ &= -\frac{\mathbb{E}_n\left[DTY\right]}{\mathbb{E}[DT]^2}\cdot\left(\mathbb{E}_n[DT] - \mathbb{E}[DT]\right) + O_p(|\mathbb{E}_n[DT] - \mathbb{E}[DT]|^2)\\ &= -\frac{\mathbb{E}\left[DTY\right]}{\mathbb{E}[DT]^2}\cdot\left(\mathbb{E}_n[DT] - \mathbb{E}[DT]\right) + o_p(n^{-1/2}).\end{aligned} \tag{B.14}$$

When $(d,t)\in\mathcal{S}_-$, similar analysis reveals that

$$\begin{aligned}\mathbb{E}_n\left[\Delta^{\psi,1}_{d,t}(W)\right] =& \mathbb{E}_n\left[(Y - m_{d,t}(X))\left(\hat{w}_{d,t} - w_{d,t}\right)(W) + m_{d,t}(X)\left(\hat{w}_{1,1} - w_{1,1}\right)(W)\right]\\ =& \mathbb{E}_n\left[(Y - m_{d,t}(X))\left(\hat{w}_{d,t} - w_{d,t}\right)(W)\right]\\ &- \frac{\mathbb{E}_n\left[DTm_{d,t}(X)\right]}{\mathbb{E}[DT]^2}\left(\mathbb{E}_n[DT] - \mathbb{E}[DT]\right) + O_p(|\mathbb{E}_n[DT] - \mathbb{E}[DT]|^2)\\ =& -\frac{\mathbb{E}\left[DTm_{d,t}(X)\right]}{\mathbb{E}[DT]^2}\left(\mathbb{E}_n[DT] - \mathbb{E}[DT]\right) + o_p(n^{-1/2}),\end{aligned} \tag{B.15}$$

where the last equation holds by Condition (i) of this lemma and by Lemma D.1(i).

Next, note that $\Delta^{\psi,2}_{1,1}(\cdot) = 0$, and when $(d,t)\in\mathcal{S}_-$, we deduce from Condition (ii) and Lemma D.1(ii) that

$$\mathbb{E}_n\left[\Delta^{\psi,2}_{d,t}(W)\right] = \mathbb{E}_n\left[(w_{1,1} - w_{d,t})(W)\left(\hat{m}_{d,t} - m_{d,t}\right)(X)\right] = o_p(n^{-1/2}). \tag{B.16}$$

Analogously, $\Delta^{\psi,3}_{1,1}(\cdot)$ is identically zero, and therefore, we only need to focus the other three cases, for which we have

$$\begin{aligned}&\mathbb{E}_n\left[\Delta^{\psi,3}_{d,t}(W)\right]\\ &= \mathbb{E}_n\left[\left(\left(\hat{w}_{1,1} - w_{1,1}\right)(W) - \left(\hat{w}_{d,t} - w_{d,t}\right)(W)\right)\left(\hat{m}_{d,t} - m_{d,t}\right)(X)\right]\end{aligned}$$

$$= \mathbb{E}_n\left[\frac{DT}{\mathbb{E}[DT]^2}\left(\widehat{m}_{d,t}-m_{d,t}\right)(X)\right]\cdot\left(\mathbb{E}_n[DT]-\mathbb{E}[DT]\right)+O_p(|\mathbb{E}_n[DT]-\mathbb{E}[DT]|^2) \tag{B.17}$$

$$- \mathbb{E}_n\left[\left(\widehat{w}_{d,t}-w_{d,t}\right)(W)\cdot\left(\widehat{m}_{d,t}-m_{d,t}\right)(X)\right], \tag{B.18}$$

where the second equality follows from a second-order Taylor expansion of $\mathbb{E}_n[DT]$ around $\mathbb{E}[DT]$.

Taking the fact that $\mathbb{E}[DT]>0$ under Assumption 2(iii) and that $\widehat{m}_{d,t}$ is uniformly convergent to $m_{d,t}$, we obtain

$$\left|\mathbb{E}_n\left[\frac{DT}{\mathbb{E}[DT]^2}\left(\widehat{m}_{d,t}-m_{d,t}\right)(X)\right]\right|\leqslant\mathbb{E}_n\left[\left|\frac{DT}{\mathbb{E}[DT]^2}\right|\cdot\left|\left(\widehat{m}_{d,t}-m_{d,t}\right)(X)\right|\right]\lesssim\|\widehat{m}_{d,t}-m_{d,t}\|_\infty=o_p(1).$$

Combining this result with $\mathbb{E}_n[DT]-\mathbb{E}[DT]=O_p\left(n^{-1/2}\right)$, we conclude that (B.17) is $o_p\left(n^{-1/2}\right)$.

Next, we study $\mathbb{E}_n\left[\left(\widehat{w}_{d,t}-w_{d,t}\right)(W)\cdot\left(\widehat{m}_{d,t}-m_{d,t}\right)(X)\right]$. Let

$$w^\dagger_{d,t}(W)=\frac{I_{d,t}\widehat{p}(1,1,X)}{p(1,1)\widehat{p}(d,t,X)}, \tag{B.19}$$

based on which, we have the following decomposition

$$\mathbb{E}_n\left[\left(w^\dagger_{d,t}-w_{d,t}\right)(W)\cdot\left(\widehat{m}_{d,t}-m_{d,t}\right)(X)\right]+\mathbb{E}_n\left[\left(\widehat{w}_{d,t}-w^\dagger_{d,t}\right)(W)\cdot\left(\widehat{m}_{d,t}-m_{d,t}\right)(X)\right]=\Delta^{1,n}_{w,m}+\Delta^{2,n}_{w,m}. \tag{B.20}$$

We consider the L_2-norm first. Under Condition (iii),

$$\Delta^{1,n}_{w,m}=\underbrace{\mathbb{E}\left[\left(w^\dagger_{d,t}-w_{d,t}\right)(W)\cdot\left(\widehat{m}_{d,t}-m_{d,t}\right)(X)\right]}_{\equiv\Delta^1_{w,m}}+o_p\left(n^{-1/2}\right).$$

Since $\hat{a}/\hat{b}-a/b=(\hat{a}-a)/b-a(\hat{b}-b)/b^2-(\hat{a}-a)(\hat{b}-b)/(\hat{b}b)+a(\hat{b}-b)^2/(\hat{b}b^2)$, we have

$$\begin{aligned}
\Delta^1_{w,m}=&\mathbb{E}\left[\frac{\delta_{d,t}(W)}{p(d,t,X)}\left(\widehat{p}(1,1,X)-p(1,1,X)\right)\right]\\
&-\mathbb{E}\left[\frac{\delta_{d,t}(W)p(1,1,X)}{p^2(d,t,X)}\left(\widehat{p}(d,t,X)-p(d,t,X)\right)\right]\\
&-\mathbb{E}\left[\frac{\delta_{d,t}(W)}{\widehat{p}(d,t,X)p(d,t,X)}\left(\widehat{p}(1,1,X)-p(1,1,X)\right)\left(\widehat{p}(d,t,X)-p(d,t,X)\right)\right]\\
&+\mathbb{E}\left[\frac{\delta_{d,t}(W)p(1,1,X)}{\widehat{p}(d,t,X)p(d,t,X)^2}\left(\widehat{p}(d,t,X)-p(d,t,X)\right)^2\right]\\
\equiv&\Delta^{1,1}_{w,m}+\Delta^{1,2}_{w,m}+\Delta^{1,3}_{w,m}+\Delta^{1,4}_{w,m},
\end{aligned}$$

where $\delta_{d,t}(W)=p(1,1)^{-1}I_{d,t}\left(\widehat{m}_{d,t}-m_{d,t}\right)(X)$.

For $\Delta^{1,1}_{w,m}$,

$$\begin{aligned}
\left|\Delta^{1,1}_{w,m}\right|\leqslant&p(1,1)^{-1}\left(p^{min}_{d,t}\right)^{-1}\mathbb{E}\left[\left|\left(\widehat{p}(1,1,X)-p(1,1,X)\right)\left(\widehat{m}_{d,t}-m_{d,t})(X)\right|\right]\\
\leqslant&O(1)\cdot\|\widehat{p}(1,1,\cdot)-p(1,1,\cdot)\|_{L_2}\cdot\|\widehat{m}_{d,t}-m_{d,t}\|_{L_2}\\
=&O_p\left(r_ns_n\right),
\end{aligned}$$

where $p^{min}_{d,t}=\inf_{x\in\mathcal{X}}|p(d,t,x)|$. The first inequality holds under Assumption 2(iii), and the second one is due to the Cauchy-Schwarz inequality.

Likewise,

$$\begin{aligned}\left|\Delta_{w,m}^{1,2}\right| \leqslant & p(1,1)^{-1} \sup_{x\in\mathcal{X}} |p(1,1,x)| \left\{\inf_{x\in\mathcal{X}} |p(d,t,x)|\right\}^{-2} \mathbb{E}\left[\left|\left(\widehat{p}(d,t,X)-p(d,t,X)\right)(\widehat{m}_{d,t}-m_{d,t})(X)\right|\right] \\ \leqslant & O(1)\cdot \|\widehat{p}(d,t,\cdot)-p(d,t,\cdot)\|_{L_2}\cdot \|\widehat{m}_{d,t}-m_{d,t}\|_{L_2} \\ = & O_p\left(r_n s_n\right).\end{aligned}$$

To analyze the convergence of the remaining two terms, we can use a similar approach to the one used for the previous two terms. However, to complete the analysis, we need to show that $\widehat{p}(d,t,x)$ is uniformly bounded away from 0 across $\mathcal{X}$, with high probability. Due to the uniform convergence, for any given $\epsilon \in (0,1/2)$, there is $N_\epsilon > 0$ such that $\sup_{x\in\mathcal{X}} |\widehat{p}(d,t,x)-p(d,t,x)| \leqslant p_{d,t}^{min}/2$ with probability at least $1-\epsilon$, whenever $n \geqslant N_\epsilon$. Thus, when n is sufficiently large, we have

$$\inf_{x\in\mathcal{X}} |\widehat{p}(d,t,x)| \geqslant \inf_{x\in\mathcal{X}} |p(d,t,x)| - \sup_{x\in\mathcal{X}} |\widehat{p}(d,t,x)-p(d,t,x)| \geqslant p_{d,t}^{min}/2 > 0,$$

with probability $1-\epsilon$, leading to our desired claim.

The sup-norm case can be handled analogously. Different from the L_2-norm, it is now possible to work directly with the empirical measure, leading to the conclusion that $\Delta_{w,m}^{1,n} = O_p\left(r_n s_n\right)$, without invoking Conditions (iii)-(v) of this lemma.

Next,we examine the estimation effect of the normalizing weight as given in $\Delta_{w,m}^{2,n}$. Let $\widehat{p}(1,1) = \mathbb{E}_n\left[I_{d,t}\frac{\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)}\right]$. Again, we focus on L_2-norm first. By definition,

$$\Delta_{w,m}^{2,n} = -\widehat{p}(1,1)^{-1}\cdot \underbrace{\mathbb{E}_n\left[w_{d,t}^{\dagger}(W)\cdot\left(\widehat{m}_{d,t}-m_{d,t}\right)(X)\right]}_{\Delta_{w,m}^{2,1,n}}\cdot\left(\widehat{p}(1,1)-p(1,1)\right).$$

We can further decompose $\Delta_{w,m}^{2,1,n}$ into

$$\begin{aligned}\Delta_{w,m}^{2,1,n} =& \Delta_{w,m}^{1,n} && \text{(B.21)}\\ &+ \left(\mathbb{E}_n-\mathbb{E}\right)\left[w_{d,t}(W)\cdot\left(\widehat{m}_{d,t}-m_{d,t}\right)(X)\right] && \text{(B.22)}\\ &+ \mathbb{E}\left[w_{d,t}(W)\cdot\left(\widehat{m}_{d,t}-m_{d,t}\right)(X)\right]. && \text{(B.23)}\\ =& O_p\left(s_n\right)+O_p\left(r_n s_n\right)+o_p\left(n^{-1/2}\right)\end{aligned}$$

By Conditions (iii) and (iv), (B.21) and (B.22) are $O_p\left(r_n s_n\right)$ and $o_p\left(n^{-1/2}\right)$, respectively. Since $p_{d,t}(\cdot)$ is uniformly bounded over $\mathcal{X}$, (B.23) is $O_p\left(s_n\right)$ by the Cauchy-Schwartz inequality.

Analogously, we have

$$\begin{aligned}\widehat{p}(1,1)-p(1,1) =& \left(\mathbb{E}_n-\mathbb{E}\right)\left[I_{d,t}\left(\frac{\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)}-\frac{p(1,1,X)}{p(d,t,X)}\right)\right] && \text{(B.24)}\\ &+ \left(\mathbb{E}_n-\mathbb{E}\right)\left[I_{d,t}\frac{p(1,1,X)}{p(d,t,X)}\right] && \text{(B.25)}\\ &+ \mathbb{E}\left[I_{d,t}\left(\frac{\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)}-\frac{p(1,1,X)}{p(d,t,X)}\right)\right] && \text{(B.26)}\\ =& O_p\left(r_n\right)+O_p\left(n^{-1/2}\right)+o_p\left(n^{-1/2}\right).\end{aligned}$$

Under Condition (v), (B.24) is $o_p\left(n^{-1/2}\right)$. Since (B.25) is a centered *i.i.d.* summand, it is $O_p\left(n^{-1/2}\right)$.

Arguing along the same line as for $\Delta^1_{w,m}$, we get (B.26) is $O_p(r_n)$. Collecting these results, we conclude that both $\Delta^{1,n}_{w,m}$ and $\Delta^{2,n}_{w,m}$ are $O_p(r_n s_n)$.

Once again, analysis under the sup-norm rely directly on empirical measure, thus eliminating the need for conditions on the empirical process. Further details are not provided here for brevity.

To finish the proof of this lemma, we gather the results in (B.10), (B.14), (B.15), (B.16), (B.18), and (B.20), which leads to

$$\begin{aligned}\widehat{\tau}_{dr} - \tau &= \mathbb{E}_n\left[\sum_{(d,t)\in\mathcal{S}}(-1)^{d+t}\psi_{d,t}(W;w,m) - \tau\right] + \tau\left(1 - \frac{\mathbb{E}_n[DT]}{\mathbb{E}[DT]}\right) + O_p(r_n s_n) + o_p\left(n^{-1/2}\right)\\ &= \mathbb{E}_n[\eta_{\text{eff}}(W)] + O_p(r_n s_n) + o_p\left(n^{-1/2}\right).\end{aligned}$$

■

Lemma B.3 Under Assumptions 2 and 4, the stochastic equicontinuity conditions (i)–(v) stated in Lemma B.2 hold for all $(d,t) \in \mathcal{S}_-$.

Proof of Lemma B.3:
The proof proceeds by verifying that the assumptions of Lemma 19.24 in van der Vaart (1998) are implied by our assumptions for each of the required conditions. For brevity, we explicitly verify Conditions (i) and (iii); the remaining conditions follow by analogous arguments.
<u>Proof of Condition (i)</u>: Let $\mathcal{E}$ denote the event that $(\widehat{p}, \{\widehat{m}_{d,t}\}_{(d,t)\in\mathcal{S}_-}) \in \mathcal{F}$. Given the definition of $w^\dagger$ in (B.19), we decompose the difference between $\widehat{w}_{d,t}$ and $w_{d,t}$ in Condition (i) as

$$\begin{aligned}&\mathbb{G}_n\left\{(Y - m_{d,t}(X))\left(\widehat{w}_{d,t} - w_{d,t}\right)(W)\right\}\\ &\quad= \mathbb{G}_n\left\{(Y - m_{d,t}(X))\left(w^\dagger_{d,t} - w_{d,t}\right)(W)\right\} + \mathbb{G}_n\left\{(Y - m_{d,t}(X))\left(\widehat{w}_{d,t} - w^\dagger_{d,t}\right)(W)\right\}\\ &\quad\equiv \Delta^1_{g_1} + \Delta^2_{g_1},\end{aligned}$$

We analyze the two terms separately.

Note that

$$\Delta^1_{g_1} = \mathbb{G}_n\left\{\underbrace{\frac{I_{d,t}(Y - m_{d,t}(X))}{p(1,1)}\frac{\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)}}_{\equiv \hat{f}^{g_1}_{d,t}(W)} - \underbrace{\frac{I_{d,t}(Y - m_{d,t}(X))}{p(1,1)}\frac{p(1,1,X)}{p(d,t,X)}}_{\equiv f^{g_1}_{d,t}(W)}\right\}.$$

Define the function class

$$\mathcal{F}^{g_1}_{d,t} = \left\{(\tilde{d},\tilde{t},\tilde{x},\tilde{y}) \in \{0,1\}^2 \times \mathcal{X}\mathcal{Y} \mapsto \left(\frac{\mathbb{1}\{\tilde{d} = d, \tilde{t} = t\}(\tilde{y} - m_{d,t}(\tilde{x}))}{p(1,1)}\right)\cdot\frac{\tilde{p}(1,1,\tilde{x})}{\tilde{p}(d,t,\tilde{x})} : \tilde{p} \in \mathcal{F}^p\left(\{0,1\}^2 \times \mathcal{X}\right)\right\},$$

for $(d,t) \in \mathcal{S}_-$. By construction, both $\hat{f}^{g_1}_{d,t}$ and $f^{g_1}_{d,t}$ belong to $\mathcal{F}^{g_1}_{d,t}$ on $\mathcal{E}$.

By the construction of $\mathcal{F}^p$ and $\mathcal{F}^m$ in Assumptions 4.2, both sets are uniformly bounded with finite uniform entropy integral, so both are P-Donsker by Corollary 2.3.12 in van der Vaart and Wellner (1996).

Moreover: (a) by Assumption 4.2(i), $\mathcal{F}^p$ is uniformly bounded away from zero, implying $\{1/f : f \in \mathcal{F}^p\}$ is also uniformly bounded; (b)

$$\sup_{f\in\mathcal{F}_{d,t}^{g_1}} \|f\|_{L_2}^2 \leqslant C \cdot \mathbb{E}[\mathbb{V}\mathrm{ar}\,[Y|D=d,T=t,X]] \cdot \sup_{\tilde{p}\in\mathcal{F}^p}\left\{\sup_{x\in\mathcal{X}} \tilde{p}(1,1,x)\right\} \cdot \sup_{\tilde{p}\in\mathcal{F}^p}\left\{\left(\inf_{x\in\mathcal{X}} \tilde{p}(d,t,x)\right)^{-1}\right\} < \infty,$$

where the first inequality follows from Assumptions 2(iii), and the second follows from Assumptions 4.2–4.3; (c) $\mathcal{F}_{d,t}^{g_1}$ consists of Lipschitz transformations of functions in $\mathcal{F}^p$, $\mathcal{F}^m$, and singleton function sets. By Theorem 2.10.6 in van der Vaart and Wellner (1996), $\mathcal{F}_{d,t}^{g_1}$ is therefore uniformly-bounded and P-Donsker.

Next, we have

$$\begin{aligned}\left\|\hat{f}_{d,t}^{g_1} - f_{d,t}^{g_1}\right\|_{L_2} \leqslant& C \cdot \mathbb{E}[\mathbb{V}\mathrm{ar}\,[Y|D=d,T=t,X]]^{1/2} \cdot \\ &\{C_1 \|\widehat{p}(1,1,\cdot) - p(1,1,\cdot)\|_\infty + C_2 \|\widehat{p}(d,t,\cdot) - p(d,t,\cdot)\|_\infty\} \\ =& o_p(1),\end{aligned}$$

where the inequality holds under Assumptions 2(iii) and 4.2(i). The convergence follows from Assumptions 4.1 and 4.3. Therefore, $\left\|\hat{f}_{d,t}^{g_1} - f_{d,t}^{g_1}\right\|_{L_2} = o_p(1)$ on $\mathcal{E}$.

Applying Lemma 19.24 of van der Vaart (1998) with $\hat{f}_n = \hat{f}_{d,t}^{g_1}$ and $f_0 \equiv f_{d,t}^{g_1}$, we conclude that $\Delta_{g_1}^1 = o_p(1)$ on the event $\mathcal{E}$. Since $\mathbb{P}\left((\widehat{p}, \{\widehat{m}_{d,t}\}_{(d,t)\in\mathcal{S}_-}) \notin \mathcal{F}\right) \leqslant \varepsilon_n^* \to 0$, it follows that $\Delta_{g_1}^1 = o_p(1)$ unconditionally.

For $\Delta_{g_1}^2$, direct manipulation yields

$$\begin{aligned}\Delta_{g_1}^2 =& \left(\frac{1}{\mathbb{E}_n\left[\frac{I_{d,t}\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)}\right]} - \frac{1}{p(1,1)}\right) \cdot \mathbb{G}_n\left\{I_{d,t}(Y - m_{d,t}(X)) \cdot \frac{\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)}\right\} \\ =& -\Bigg(\underbrace{\mathbb{E}_n\left[\frac{I_{d,t}\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)}\right] \cdot \mathbb{E}\left[\frac{I_{d,t}p(1,1,X)}{p(d,t,X)}\right]}_{\equiv\Delta_{g_1}^{21}}\Bigg)^{-1} \cdot \Bigg(\underbrace{\mathbb{E}_n\left[\frac{I_{d,t}\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)}\right] - \mathbb{E}\left[\frac{I_{d,t}p(1,1,X)}{p(d,t,X)}\right]}_{\equiv\Delta_{g_1}^{22}}\Bigg) \cdot \\ &\Bigg(\underbrace{\mathbb{G}_n\left\{I_{d,t}(Y - m_{d,t}(X)) \cdot \left(\frac{\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)} - \frac{p(1,1,X)}{p(d,t,X)}\right)\right\}}_{\equiv\Delta_{g_1}^{23}} + \underbrace{\mathbb{G}_n\left\{I_{d,t}(Y - m_{d,t}(X)) \cdot \frac{p(1,1,X)}{p(d,t,X)}\right\}}_{\equiv\Delta_{g_1}^{24}}\Bigg).\end{aligned}$$

By Assumptions 2(iii) and 4.2(i), $\Delta_{g_1}^{21}$ is uniformly bounded and bounded away from zero, so $\left(\Delta_{g_1}^{21}\right)^{-1} = O_p(1)$ on $\mathcal{E}$. Analogous arguments to those used for $\Delta_{g_1}^1$ and Lemma 19.24 in van der Vaart (1998) imply $\Delta_{g_1}^{23} = o_p(1)$ on $\mathcal{E}$. Furthermore, we have $\Delta_{g_1}^{24} = O_p(1)$ by the CLT. It remains to bound $\Delta_{g_1}^{22}$, for which we have

$$\begin{aligned}\Delta_{g_1}^{22} =& (\mathbb{E}_n - \mathbb{E})\left[\frac{I_{d,t}\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)}\right] + \mathbb{E}\left[I_{d,t} \cdot \left(\frac{\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)} - \frac{p(1,1,X)}{p(d,t,X)}\right)\right] \\ \leqslant& o_p(1) + C_1 \cdot \|\widehat{p}(1,1,\cdot) - p(1,1,\cdot)\|_\infty + C_2 \cdot \|\widehat{p}(d,t,\cdot) - p(d,t,\cdot)\|_\infty \\ =& o_p(1)\end{aligned} \tag{B.27}$$

on $\mathcal{E}$. The first term on the right hand side of (B.27) is $o_p(1)$ because both $\widehat{p}$ and $1/\widehat{p}$ lie in uniformly bounded P-Donsker classes with probability approaching one under Assumption 4. Hence, by Theorem 2.10.6 of van der Vaart and Wellner (1996) (see also Examples 2.10.8–2.10.9 therein), it is $o_p\left(n^{-1/2}\right)$. The second term on the right hand side is bounded by direct application of the triangle inequality and Assumption 4.2(i). The last equality holds by Assumption 4.1.

Collecting results on $\Delta^{21}_{g_1} - \Delta^{24}_{g_1}$, we obtain $\Delta^2_{g_1} = O_p(1) \cdot o_p(1) \cdot (o_p(1) + O_p(1)) = o_p(1)$ on $\mathcal{E}$. Again, since $\mathbb{P}(\mathcal{E}) \to 1$, the conclusion $\Delta^2_{g_1} = o_p(1)$ holds unconditionally.

Since both $\Delta^1_{g_1}$ and $\Delta^2_{g_1}$ are $o_p(1)$, Condition (i) holds.

Proof of Condition (iii): We first define the function class,

$$\mathcal{F}^{g_3}_{d,t} = \Bigg\{(\tilde{d}, \tilde{t}, \tilde{x}, \tilde{y}) \in \{0,1\}^2 \times \mathcal{XY} \mapsto \mathbb{1}\{\tilde{d} = d, \tilde{t} = t\} \cdot \left(\frac{\tilde{p}(1,1,\tilde{x})}{\tilde{p}(d,t,\tilde{x})} - \frac{p(1,1,\tilde{x})}{p(d,t,\tilde{x})}\right) \cdot (\tilde{m}_{d,t}(\tilde{x}) - m_{d,t}(\tilde{x})) : \\ \tilde{p} \in \mathcal{F}^p\left(\{0,1\}^2 \times \mathcal{X}\right), \tilde{m} \in \mathcal{F}^m(\mathcal{X})\Bigg\},$$

for $(d,t) \in \mathcal{S}_-$. Define $\hat{f}^{g_3}_{d,t}(W) \equiv I_{d,t}\left(\frac{\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)} - \frac{p(1,1,X)}{p(d,t,X)}\right) \cdot (\widehat{m}_{d,t}(X) - m_{d,t}(X))$, for $(d,t) \in \mathcal{S}_-$. Following arguments analogous to those in the proof of Condition (i), we deduce that $\mathcal{F}^{g_3}_{d,t}$ is a uniformly-bounded and P-Donsker. Moreover,

$$\begin{aligned}\left\|\hat{f}^{g_3}_{d,t}\right\|_{L_2} &\leqslant (C_1 \cdot \|\widehat{p}(1,1,\cdot) - p(1,1,\cdot)\|_\infty + C_2 \cdot \|\widehat{p}(d,t,\cdot) - p(d,t,\cdot)\|_\infty) \cdot \|\widehat{m}_{d,t} - m_{d,t}\|_{L_2} \\ &= o_p(1)\end{aligned}$$

on $\mathcal{E}$. The first inequality follows under Assumptions 2(iii) and 4.2(i)-(ii), and the convergence follows by Assumption 4.1.

Applying Lemma 19.24 of van der Vaart (1998) with $\hat{f}_n = \hat{f}^{g_3}_{d,t}$ and $f_0 \equiv 0$ and using $\mathbb{P}(\mathcal{E}) \to 1$, the desired result follows immediately. ■

Proof of Lemma 3.1: The result follows directly from Lemmas B.2 and B.3. ■

Proof of Theorem 3.1:
Lemma B.2 establishes that the expansion in (3.6) holds under the sup-norm when Assumptions 4.1, 4.3, and Conditions (i)–(ii) of that lemma are satisfied, without requiring direct verification of the uniform entropy integral conditions in Assumption 4.2.

Assumption 4.1 follows from Lemma C.2 and Assumption 5.5, while Assumption 4.3 is implied by Assumptions 5.1-3. Stochastic equicontinuity conditions (i)–(ii) in Lemma B.2 are verified in Lemmas D.2 and D.3. With the bandwidth rate conditions in Assumption 5.5 guaranteeing that the leading remainder term is $O_p(r_n s_n) = o_p\left(n^{-1/2}\right)$, the asymptotic normality follows directly from the CLT. ■

Proof of Theorem 4.1:
Proof of Part (a): In Theorem 3.1, we have already shown that under Assumption 5, for every $P \in \mathbf{P}_0$, the following holds: $\widehat{\tau}_{dr} - \tau = \mathbb{E}_n[\eta_{\text{eff}}(W)] + o_p\left(n^{-1/2}\right)$. By applying a similar line of reasoning and using the additional conditions on Assumption 5 specified in this theorem, one can similarly establish that, for any $P \in \mathbf{P}_0$: $\widehat{\tau}_{sz} - \tau = \mathbb{E}_n[\eta_{sz}(W)] + o_p\left(n^{-1/2}\right)$. Now, by the CLT, we have

$$\sqrt{n}\,(\widehat{\tau}_{dr} - \widehat{\tau}_{sz}) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[(\eta_{\text{eff}}(W) - \eta_{sz}(W))^2\right]\right).$$

It remains to show that

$$\widehat{V}_n \xrightarrow{p} V, \tag{B.28}$$

and

$$V = \rho_{sz} > 0. \tag{B.29}$$

First, it is implied from the proof of Theorem 3.1 that $\widehat{\eta}_{eff}(w) \xrightarrow{p} \eta_{\text{eff}}(w)$, uniformly in $w \in \mathcal{W}$. In a similar vein, $\widehat{\eta}_{sz}(w) \xrightarrow{p} \eta_{sz}(w)$ uniformly over $\mathcal{W}$, under $H_{0,P}$. Combining these two results, (B.28) then follows by the CMT and the weak LLN.

From Proposition 1 in Sant'Anna and Zhao (2020), we know that $\eta_{sz}(\cdot)$ is the efficient influence function for all regular estimators of the ATT, which makes $\widehat{\tau}_{sz}$ an efficient estimator of τ under $H_{0,P}$. Moreover, since both $\widehat{\tau}_{dr}$ and $\widehat{\tau}_{sz}$ are consistent under $H_{0,P}$, it follows from Lemma 2.1 in Hausman (1978) that $\mathbb{E}[\eta_{\text{eff}}(W)\eta_{sz}(W)] = \mathbb{E}[\eta_{sz}(W)^2]$. Hence, $\mathbb{E}\left[(\eta_{\text{eff}}(W) - \eta_{sz}(W))^2\right] = \mathbb{E}\left[\eta_{\text{eff}}(W)^2\right] - \mathbb{E}\left[\eta_{sz}(W)^2\right]$. Given this result, (B.29) now follows by Proposition 2.2 and the condition that $\mathbb{V}\text{ar}\left[\tau(X)|D=1\right] > 0$.

<u>Proof of Part (b)</u>: We proceed by establishing: (i) $\widehat{\tau}_{sz} - \widehat{\tau}_{dr} \xrightarrow{p} \tau_{sz} - \tau_{dr} \neq 0$; (ii) $\widehat{V}_n \xrightarrow{p} V < \infty$, under $H_{1,P}$.

Under the assumptions of the theorem, for any $P \in \mathbf{P}_1$, $\widehat{p}(d,t,x) \xrightarrow{p} p(d,t,x)$ and $\widehat{m}_{d,t}(x) \xrightarrow{p} m_{d,t}(x)$, uniformly in $x \in \mathcal{X}$, for $(d,t) \in \mathcal{S}$. Now, applying the LLN, we get $\widehat{\tau}_{dr} \xrightarrow{p} \tau_{dr}$ and $\widehat{\tau}_{sz} \xrightarrow{p} \tau_{sz}$. Result (i) then follows from the CMT. Next, we deduce from the uniform consistency of $\widehat{p}$ and $\widehat{m}$, the CMT, and LLN, that (B.28) holds under $H_{1,P}$. Furthermore, Assumptions 2(iii) and 5.3 ensure that both $\widehat{\eta}_{sz}$ and $\widehat{\eta}_{dr}$ are uniformly bounded, which leads to $V < \infty$. This concludes the proof of part (b). ■

Proof of Lemma 7.1:
The proof follows a similar line of reasoning as in Lemma B.2. With the generic first-step estimator now replaced by cross-fitted estimators, we can leverage the empirical process theory developed in Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey and Robins (2017) to simplify the analysis.

Since J is a fixed integer, it suffices to examine the behavior of $\mathbb{E}_{n,j}[\psi_{d,t}(W;\widehat{w}_j,\widehat{m}_j)]$, where $\psi_{d,t}$ is defined in (B.9), for any given $j \in \{1,...,J\}$. Define the weight factor for a given $\tilde{p}$ as $\tilde{w}_{d,t}(w) = \frac{I_{d,t}\tilde{p}(1,1,x)}{\mathbb{E}[DT]\tilde{p}(d,t,x)}$ and for the cross-fitted estimators: $w^\dagger_{d,t,j}(w) = \frac{I_{d,t}\widehat{p}_j(1,1,x)}{\mathbb{E}[DT]\widehat{p}_j(d,t,x)}$. Note that $\tilde{w}_{1,1}(W) = w^\dagger_{1,1,j}(W) = w_{1,1}(W) = \frac{DT}{\mathbb{E}[DT]}$.

Let $\mathcal{E}_n$ denote the event that $(\widehat{p}_j,\widehat{m}_j) \in \mathcal{J}_n$ for all $j \in \{1,...,J\}$. Under Assumption 6.1, we have $\mathbb{P}(\mathcal{E}_n) \geqslant 1 - J\varepsilon'_n \to 1$.

Define

$$\begin{aligned}
\bar{\psi}^1_{d,t}(W;w^\dagger_j,\widehat{m}_j) &= \psi_{d,t}(W;w^\dagger_{d,t,j},\widehat{m}_{d,t,j}) - \psi_{d,t}(W;w_{d,t},m_{d,t}),\\
\bar{\psi}^2_{d,t}(W;\widehat{w}_j,w^\dagger_j,\widehat{m}_j) &= \psi_{d,t}(W;\widehat{w}_{d,t,j},\widehat{m}_{d,t,j}) - \psi_{d,t}(W;w^\dagger_{d,t,j},\widehat{m}_{d,t,j}).
\end{aligned}$$

By the triangular inequality, we have

$$\left|\mathbb{E}_{n,j}\left[\bar{\psi}^1_{d,t}(W;w^\dagger_j,\widehat{m}_j)\right]\right| \leqslant n_J^{-1/2}\left(G^1_{d,t,j} + G^2_{d,t,j}\right), \tag{B.30}$$

where

$$\begin{aligned}
G^1_{d,t,j} &= \left|\mathbb{G}_{n,j}[\bar{\psi}^1_{d,t}(W;w^\dagger_j,\widehat{m}_j)]\right|,\\
G^2_{d,t,j} &= n_J^{1/2}\left|\mathbb{E}\left[\psi_{d,t}(W;w^\dagger_{d,t,j},\widehat{m}_{d,t,j})|(W_i)_{i\in\mathfrak{I}_{-j}}\right] - \mathbb{E}\left[\psi_{d,t}(W;w_{d,t},m_{d,t})\right]\right|,
\end{aligned}$$

$$\mathbb{G}_{n,j}[f(W)] = n_J^{1/2}\left(\mathbb{E}_{n,j}[f(W)] - \mathbb{E}[f(W)|(W_i)_{i\in\mathfrak{I}_{-j}}]\right).$$

Note that $\bar{\psi}^1_{1,1} = 0$, so it suffices to consider the remaining three cases. On the event $\mathcal{E}_n$, which holds with probability approaching 1, we have

$$\begin{aligned}
\mathbb{E}\left[\left(G^1_{d,t,j}\right)^2 |(W_i)_{i\in\mathfrak{I}_{-j}}\right] &\leqslant \mathbb{E}\left[\left\|\bar{\psi}^1_{d,t}(W; w^\dagger_j, \widehat{m}_j)\right\|^2 |(W_i)_{i\in\mathfrak{I}_{-j}}\right] \\
&\leqslant \sup_{(\tilde{p},\tilde{m})\in\mathcal{J}_n} \mathbb{E}\left[\left\|\bar{\psi}^1_{d,t}(W;\tilde{w},\tilde{m})\right\|^2 |(W_i)_{i\in\mathfrak{I}_{-j}}\right] \\
&\leqslant \sup_{(\tilde{p},\tilde{m})\in\mathcal{J}_n} \mathbb{E}\left[\left\|\bar{\psi}^1_{d,t}(W;\tilde{w},\tilde{m})\right\|^2\right]
\end{aligned}$$

Using the decomposition provided in (B.11)–(B.13), we write

$$\bar{\psi}^1_{d,t}(W;\tilde{w},\tilde{m}) = \Delta^{\psi,1}_{d,t}(W;\tilde{w},\tilde{m}) + \Delta^{\psi,2}_{d,t}(W;\tilde{w},\tilde{m}) + \Delta^{\psi,3}_{d,t}(W;\tilde{w},\tilde{m}).$$

On the event $\mathcal{E}_n$, we have

$$\begin{aligned}
&\left\|\Delta^{\psi,1}_{d,t}(W;\tilde{w},\tilde{m})\right\|_{L_2} \\
&\leqslant \|I_{d,t}(Y - m_{d,t}(X))\|_{L_2} \cdot \|\tilde{w}_{d,t} - w_{d,t}\|_{L_2} \\
&\leqslant C \cdot \|I_{d,t}(Y - m_{d,t}(X))\|_{L_2} \cdot \Big(\|\tilde{p}(1,1,\cdot) - p(1,1,\cdot)\|_{L_2} \cdot \|p(d,t,\cdot)\|^{-1}_{L_2} \\
&\quad + \|\tilde{p}(d,t,\cdot) - p(d,t,\cdot)\|_{L_2} \cdot \|\tilde{p}(d,t,\cdot)\|^{-2}_{L_2}\Big) \\
&\leqslant C_1 \sup_{\tilde{p}\in\mathcal{J}^p_n} \|\tilde{p}(1,1,\cdot) - p(1,1,\cdot)\|_{L_2} + C_2 \sup_{\tilde{p}\in\mathcal{J}^p_n} \|\tilde{p}(d,t,\cdot) - p(d,t,\cdot)\|_{L_2} \\
&= O(r_{n,1,1} + r_{n,d,t}),
\end{aligned}$$

where the third inequality follows because

$$\begin{aligned}
\|I_{d,t}(Y - m_{d,t}(X))\|_{L_2} &= \left(\mathbb{E}[p(d,t,X)\mathbb{V}\mathrm{ar}\left[Y|D=d,T=t,X\right]]\right)^{1/2} \\
&\leqslant \left(p(d,t)\cdot\|\mathbb{V}\mathrm{ar}\left[Y|D=d,T=t,X=\cdot\right]\|_\infty\right)^{1/2} < C,
\end{aligned}$$

under Assumption 6.2. The last equality is due to Assumption 6.1(i).

Next, for $\Delta^{\psi,2}_{d,t}$,

$$\begin{aligned}
\left\|\Delta^{\psi,2}_{d,t}(W;\tilde{w},\tilde{m})\right\|_{L_2} &\leqslant \|w_{1,1} - w_{d,t}\|_{L_2} \cdot \|\tilde{m}_{d,t} - m_{d,t}\|_{L_2} \\
&\leqslant C \sup_{\tilde{m}\in\mathcal{J}^m_n} \|\tilde{m}_{d,t} - m_{d,t}\|_{L_2} \\
&= O(s_{n,d,t}),
\end{aligned}$$

on the event $\mathcal{E}_n$. Here, the second inequality follows by Assumption 2(iii) and the last equality is due to Assumption 6.1(ii).

For $\Delta^{\psi,3}_{d,t}$, we have

$$\begin{aligned}
\left\|\Delta^{\psi,3}_{d,t}(W;\tilde{w},\tilde{m})\right\|_{L_2} &\leqslant \|\tilde{w}_{d,t} - w_{d,t}\|_{L_2} \cdot \|\tilde{m}_{d,t} - m_{d,t}\|_{L_2} \\
&\leqslant \left(C_1\|\tilde{p}(1,1,\cdot) - p(1,1,\cdot)\|_{L_2} + C_2\|\tilde{p}(d,t,\cdot) - p(d,t,\cdot)\|_{L_2}\right) \cdot \|\tilde{m}_{d,t} - m_{d,t}\|_{L_2}
\end{aligned}$$

$$
\begin{aligned}
&\leqslant C_1 \sup_{(\tilde{p},\tilde{m})\in\mathcal{J}_n} \|\tilde{p}(1,1,\cdot)-p(1,1,\cdot)\|_{L_2}\cdot\|\tilde{m}_{d,t}-m_{d,t}\|_{L_2}\\
&\quad + C_2 \sup_{(\tilde{p},\tilde{m})\in\mathcal{J}_n} \|\tilde{p}(d,t,\cdot)-p(d,t,\cdot)\|_{L_2}\cdot\|\tilde{m}_{d,t}-m_{d,t}\|_{L_2}\\
&=O(r_{n,1,1}\cdot s_{n,d,t}+r_{n,d,t}\cdot s_{n,d,t}),
\end{aligned}
$$

where the last equality holds under Assumption 6.1(iii).

Collecting these results, we deduce that $\sup_{(\tilde{p},\tilde{m})\in\mathcal{J}_n}\mathbb{E}\left[\left\|\bar{\psi}^1_{d,t}(W;\tilde{w},\tilde{m})\right\|^2\right] = O((r_{n,1,1}+r_{n,d,t}+s_{n,d,t})^2)$ on the event $\mathcal{E}_n$. By Markov's inequality and Lemma 6.1 in Chernozhukov et al. (2017), it follows that $G^1_{d,t,j}=O_p(r_{n,1,1}+r_{n,d,t}+s_{n,d,t})=o_p(1)$.

To bound $G^2_{d,t,j}$, we first define

$$
g_{\psi,j}(a)\equiv\mathbb{E}\left[\psi_{d,t}(W;aw^\dagger_{d,t,j}+(1-a)w_{d,t},a\widehat{m}_{d,t,j}+(1-a)m_{d,t})|(W_i)_{i\in\mathfrak{I}_{-j}}\right]-\mathbb{E}\left[\psi_{d,t}(W;w_{d,t},m_{d,t})\right],
$$

for $a\in[0,1]$. By a second-order Taylor expansion, we have

$$
g_{\psi,j}(1)=g_{\psi,j}(0)+g'_{\psi,j}(0)+g''_{\psi,j}(\tilde{a})/2,
$$

for some $\tilde{a}\in(0,1)$. Noting that $g_j(0)=0$, and further

$$
\begin{aligned}
g'_{\psi,j}(a)=&\,\mathbb{E}\Big[(Y-m_{d,t}(X)-a(\widehat{m}_{d,t,j}-m_{d,t})(X))\cdot(w^\dagger_{d,t,j}-w_{d,t})(W)\\
&\qquad+(w_{1,1}-w_{d,t}(W)-a(w^\dagger_{d,t,j}-w_{d,t})(W))\cdot(\widehat{m}_{d,t,j}-m_{d,t})(X)|(W_i)_{i\in\mathfrak{I}_{-j}}\Big]\\
g''_{\psi,j}(a)=&-2\,\mathbb{E}[(w^\dagger_{d,t,j}-w_{d,t})(W)\cdot(\widehat{m}_{d,t,j}-m_{d,t})(X)|(W_i)_{i\in\mathfrak{I}_{-j}}],
\end{aligned}
$$

From this, we deduce by the LIE that

$$
\begin{aligned}
g'_{\psi,j}(0)&=\mathbb{E}\left[(Y-m_{d,t}(X))\cdot(w^\dagger_{d,t,j}-w_{d,t})(W)+(w_{1,1}-w_{d,t})(W)\cdot(\widehat{m}_{d,t,j}-m_{d,t})(X)|(W_i)_{i\in\mathfrak{I}_{-j}}\right]\\
&=0.
\end{aligned}
$$

On the event $\mathcal{E}_n$, we have

$$
\begin{aligned}
\left\|G^2_{d,t,j}\right\|_{L_2}=&\,n_J^{1/2}\cdot\|g_{\psi,j}(1)\|_{L_2}=n_J^{1/2}\cdot\left\|g''_{\psi,j}(\tilde{a})/2\right\|_{L_2}\\
=&\,n_J^{1/2}\cdot\left\|\mathbb{E}[(w^\dagger_{d,t,j}-w_{d,t})(W)\cdot(\widehat{m}_{d,t,j}-m_{d,t})(X)|(W_i)_{i\in\mathfrak{I}_{-j}}]\right\|_{L_2}\\
\leqslant&\,n_J^{1/2}\cdot\sup_{(\tilde{p},\tilde{m})\in\mathcal{J}_n}\|\tilde{w}_{d,t}-w_{d,t}\|_{L_2}\cdot\|\tilde{m}_{d,t}-m_{d,t}\|_{L_2}\\
\leqslant&\,n_J^{1/2}\cdot\left(\sup_{(\tilde{p},\tilde{m})\in\mathcal{J}_n}C_1\|\tilde{p}_{1,1}-p_{1,1}\|_{L_2}\cdot\|\tilde{m}_{d,t}-m_{d,t}\|_{L_2}\right.\\
&\left.\quad+\sup_{(\tilde{p},\tilde{m})\in\mathcal{J}_n}C_2\|\tilde{p}_{d,t}-p_{d,t}\|_{L_2}\cdot\|\tilde{m}_{d,t}-m_{d,t}\|_{L_2}\right)\\
=&\,O(n^{1/2}\cdot s_{n,d,t}\cdot(r_{n,1,1}+r_{n,d,t}))=o(1),
\end{aligned}
$$

where the last equality holds under Assumption 6.1(iii). This implies, by Markov's inequality, that $G^2_{d,t,j}=o_p(1)$. As a result,

$$
\mathbb{E}_{n,j}\left[\bar{\psi}^1_{d,t}(W;w^\dagger_j,\widehat{m}_j)\right]=o_p\left(n^{-1/2}\right). \tag{B.31}
$$

Next, we analyze $\bar{\psi}^2_{d,t}$. Using a second-order Taylor expansion, we have

$$J^{-1}\sum_{j=1}^{J}\mathbb{E}_{n,j}[\bar{\psi}^2_{1,1}(W;\widehat{w}_j,w_j^{\dagger},\widehat{m}_j)] = -\frac{\mathbb{E}[DTY]}{\mathbb{E}[DT]^2}\cdot(\mathbb{E}_n[DT]-\mathbb{E}[DT]) + O_p(|\mathbb{E}_n[DT]-\mathbb{E}[DT]|^2) = O_p\left(n^{-1/2}\right) \tag{B.32}$$

When $(d,t)\in\mathcal{S}_-$,

$$\begin{aligned}\bar{\psi}^2_{d,t}(W;\widehat{w}_j,w_j^{\dagger},\widehat{m}_j) &= (Y-\widehat{m}_{d,t,j}(X))(\widehat{w}_{d,t,j}-w^{\dagger}_{d,t,j})(W)+\widehat{m}_{d,t,j}(X)(\widehat{w}_{1,1,j}-w^{\dagger}_{1,1,j})(W)\\ &\equiv \bar{\psi}^{2,1}_{d,t}(W;\widehat{w}_j,w_j^{\dagger},\widehat{m}_j)+\bar{\psi}^{2,2}_{d,t}(W;\widehat{w}_j,w_j^{\dagger},\widehat{m}_j).\end{aligned}$$

We first bound $\bar{\psi}^{2,2}_{d,t}$. Applying a first-order Taylor expansion gives

$$\begin{aligned}&J^{-1}\sum_{j=1}^{J}\mathbb{E}_{n,j}[\bar{\psi}^{2,2}_{d,t}(W;\widehat{w}_j,w_j^{\dagger},\widehat{m}_j)]\\ &= -\frac{(\mathbb{E}_n[DT]-\mathbb{E}[DT])}{\mathbb{E}_n[DT]\,\mathbb{E}[DT]}\cdot J^{-1}\sum_{j=1}^{J}\mathbb{E}_{n,j}\left[\widehat{m}_{d,t}(X)DT\right]\\ &= -\frac{\mathbb{E}[DTm_{d,t}(X)]}{\mathbb{E}[DT]^2}\cdot(\mathbb{E}_n[DT]-\mathbb{E}[DT])\\ &\quad -\frac{(\mathbb{E}_n[DT]-\mathbb{E}[DT])}{\mathbb{E}[DT]^2}\cdot J^{-1}\sum_{j=1}^{J}\mathbb{E}_{n,j}\left[\underbrace{DT(\widehat{m}_{d,t,j}-m_{d,t})(X)}_{\Delta^{m,1}_{d,t,j}(W)}\right]+O_p(|\mathbb{E}_n[DT]-\mathbb{E}[DT]|^2)\\ &= -\frac{\mathbb{E}[DTm_{d,t}(X)]}{\mathbb{E}[DT]^2}\cdot(\mathbb{E}_n[DT]-\mathbb{E}[DT])+O_p(n^{-1/2}s_{n,d,t}),\end{aligned} \tag{B.33}$$

on $\mathcal{E}_n$. The last equality follows from bounding the cross-fitted average of $\Delta^{m,1}_{d,t,j}$ as

$$\begin{aligned}\mathbb{E}\left[\left(\mathbb{G}_{n,j}[\Delta^{m,1}_{d,t,j}(W)]\right)^2|(W_i)_{i\in\mathfrak{I}_{-j}}\right] &\leqslant \mathbb{E}\left[\left\|\Delta^{m,1}_{d,t,j}(W)\right\|^2|(W_i)_{i\in\mathfrak{I}_{-j}}\right]\\ &\leqslant \sup_{(\tilde{p},\tilde{m})\in\mathcal{J}_n}\|\tilde{m}_{d,t}-m_{d,t}\|^2_{L_2}=O((s_{n,d,t})^2),\\ \text{and}\qquad \left|\mathbb{E}\left[\Delta^{m,1}_{d,t,j}(W)|(W_i)_{i\in\mathfrak{I}_{-j}}\right]\right| &\leqslant C\sup_{(\tilde{p},\tilde{m})\in\mathcal{J}_n}\|\tilde{m}_{d,t}-m_{d,t}\|_{L_2}=O(s_{n,d,t})=o(1),\end{aligned}$$

which implies $J^{-1}\sum_{j=1}^{J}\mathbb{E}_{n,j}\left[\Delta^{m,1}_{d,t,j}(W)\right]=O_p(s_{n,d,t}+n^{-1/2}s_{n,d,t})$.

Next, we bound $\bar{\psi}^{2,1}_{d,t}$. Let $\widehat{p}_J(1,1)=J^{-1}\sum_{j=1}^{J}\mathbb{E}_{n,j}\left[\frac{I_{d,t}\widehat{p}_j(1,1,X)}{\widehat{p}_j(d,t,X)}\right]$. Then,

$$\begin{aligned}&J^{-1}\sum_{j=1}^{J}\mathbb{E}_{n,j}[\bar{\psi}^{2,1}_{d,t}(W;\widehat{w}_j,w_j^{\dagger},\widehat{m}_j)]\\ &= -\frac{(\widehat{p}_J(1,1)-p(1,1))}{\widehat{p}_J(1,1)}\cdot\left(J^{-1}\sum_{j=1}^{J}\mathbb{E}_{n,j}\left[w^{\dagger}_{d,t,j}(W)(Y-\widehat{m}_{d,t,j}(X))\right]\right)\end{aligned} \tag{B.34}$$

$$
\begin{aligned}
&= -\,\widehat{p}_J(1,1)^{-1} \cdot \left(J^{-1} \sum_{j=1}^{J} \mathbb{E}_{n,j} \left[\underbrace{\frac{I_{d,t}\widehat{p}_j(1,1,X)}{\widehat{p}_j(d,t,X)} - \frac{I_{d,t}p(1,1,X)}{p(d,t,X)}}_{\Delta^{m,2}_{d,t,j}(X)} \right] + (\mathbb{E}_n - \mathbb{E}) \left[\frac{I_{d,t}p(1,1,X)}{p(d,t,X)} \right] \right) \\
&\quad \cdot \left(J^{-1} \sum_{j=1}^{J} \mathbb{E}_{n,j} \left[\underbrace{(w^{\dagger}_{d,t,j} - w_{d,t})(W)(Y - m_{d,t}(X))}_{\Delta^{m,3}_{d,t,j}(X)} - \underbrace{(w^{\dagger}_{d,t,j} - w_{d,t})(W)(\widehat{m}_{d,t,j} - m_{d,t})(X)}_{\Delta^{m,4}_{d,t,j}(X)} \right.\right. \\
&\quad \left.\left. - \underbrace{w_{d,t}(W)(\widehat{m}_{d,t,j} - m_{d,t})(X)}_{\Delta^{m,5}_{d,t,j}(X)} \right] + \mathbb{E}_n \left[w_{d,t}(W)(Y - m_{d,t}(X)) \right] \right) \\
&= O_p(1) \cdot \Big(O_p(r_{n,1,1} + r_{n,d,t}) + O_p(n^{-1/2}) \Big) \cdot \Big(O_p(s_{n,d,t}) + O_p(n^{-1/2}) \Big) \\
&= O_p(r_{n,1,1}s_{n,d,t} + r_{n,d,t}s_{n,d,t}) + o_p\left(n^{-1/2}\right) = o_p\left(n^{-1/2}\right), \qquad \text{(B.35)}
\end{aligned}
$$

where the last equality follows from the bounds on the terms involving $\Delta^{m,2}_{d,t}$ through $\Delta^{m,5}_{d,t}$.

We now proceed to bound these terms. Starting with the asymptotic variance part,

$$
\begin{aligned}
&\mathbb{E}\left[\Big(\mathbb{G}_{n,j}[\Delta^{m,2}_{d,t,j}(W)] \Big)^2 \right] \leqslant \mathbb{E}\left[\left\| \Delta^{m,2}_{d,t,j}(W) \right\|^2 |(W_i)_{i\in\mathfrak{I}_{-j}} \right] \\
&\leqslant \left(C_1 \sup_{\tilde{p}\in\mathcal{J}^p_n} \|\tilde{p}(1,1,\cdot) - p(1,1,\cdot)\|^2_{L_2} + C_2 \sup_{\tilde{p}\in\mathcal{J}^p_n} \|\tilde{p}(d,t,\cdot) - p(d,t,\cdot)\|^2_{L_2} \right) \\
&= O((r_{n,1,1})^2 + (r_{n,d,t})^2), \\
&\mathbb{E}\left[\Big(\mathbb{G}_{n,j}[\Delta^{m,3}_{d,t,j}(W)] \Big)^2 |(W_i)_{i\in\mathfrak{I}_{-j}} \right] \leqslant \mathbb{E}\left[\left\| \Delta^{m,3}_{d,t,j}(W) \right\|^2 |(W_i)_{i\in\mathfrak{I}_{-j}} \right] \\
&\leqslant \left(C_1 \sup_{\tilde{p}\in\mathcal{J}^p_n} \|\tilde{p}(1,1,\cdot) - p(1,1,\cdot)\|^2_{L_2} + C_2 \sup_{\tilde{p}\in\mathcal{J}^p_n} \|\tilde{p}(d,t,\cdot) - p(d,t,\cdot)\|^2_{L_2} \right) \\
&= O((r_{n,1,1})^2 + (r_{n,d,t})^2), \\
&\mathbb{E}\left[\Big(\mathbb{G}_{n,j}[\Delta^{m,4}_{d,t,j}(W)] \Big)^2 |(W_i)_{i\in\mathfrak{I}_{-j}} \right] \leqslant \mathbb{E}\left[\left\| \Delta^{m,4}_{d,t,j}(W) \right\|^2 |(W_i)_{i\in\mathfrak{I}_{-j}} \right] \\
&\leqslant \left(C_1 \sup_{\tilde{p}\in\mathcal{J}^p_n} \|\tilde{p}(1,1,\cdot) - p(1,1,\cdot)\|^2_{L_2} + C_2 \sup_{\tilde{p}\in\mathcal{J}^p_n} \|\tilde{p}(d,t,\cdot) - p(d,t,\cdot)\|^2_{L_2} \right) \cdot \sup_{\tilde{m}\in\mathcal{J}^m_n} \|\tilde{m}_{d,t} - m_{d,t}\|^2_{L_2} \\
&= O((r_{n,1,1})^2(s_{n,d,t})^2 + (r_{n,d,t})^2(s_{n,d,t})^2), \\
&\mathbb{E}\left[\Big(\mathbb{G}_{n,j}[\Delta^{m,5}_{d,t,j}(W)] \Big)^2 |(W_i)_{i\in\mathfrak{I}_{-j}} \right] \leqslant \mathbb{E}\left[\left\| \Delta^{m,5}_{d,t,j}(W) \right\|^2 |(W_i)_{i\in\mathfrak{I}_{-j}} \right] \leqslant C \sup_{\tilde{m}\in\mathcal{J}^m_n} \|\tilde{m}_{d,t} - m_{d,t}\|^2_{L_2} \\
&= O((s_{n,d,t})^2),
\end{aligned}
$$

Their respective bias parts are given by

$$
\begin{aligned}
\left| \mathbb{E}\left[\Delta^{m,2}_{d,t,j}(W) |(W_i)_{i\in\mathfrak{I}_{-j}} \right] \right| &\leqslant C_1 \sup_{\tilde{p}\in\mathcal{J}^p_n} \|\tilde{p}(1,1,\cdot) - p(1,1,\cdot)\|_{L_2} + C_2 \sup_{\tilde{p}\in\mathcal{J}^p_n} \|\tilde{p}(d,t,\cdot) - p(d,t,\cdot)\|_{L_2} \\
&= O(r_{n,1,1} + r_{n,d,t}), \\
\left| \mathbb{E}\left[\Delta^{m,3}_{d,t,j}(W) |(W_i)_{i\in\mathfrak{I}_{-j}} \right] \right| &= 0,
\end{aligned}
$$

$$\left|\mathbb{E}\left[\Delta^{m,4}_{d,t,j}(W)|(W_i)_{i\in\mathfrak{I}_{-j}}\right]\right| \leqslant \left(C_1 \sup_{\tilde{p}\in\mathcal{J}^p_n}\|\tilde{p}(1,1,\cdot)-p(1,1,\cdot)\|_{L_2} + C_2 \sup_{\tilde{p}\in\mathcal{J}^p_n}\|\tilde{p}(d,t,\cdot)-p(d,t,\cdot)\|_{L_2}\right)$$

$$\cdot \sup_{\tilde{m}\in\mathcal{J}^m_n}\|\tilde{m}_{d,t}-m_{d,t}\|_{L_2} = O(r_{n,1,1}s_{n,d,t}+r_{n,d,t}s_{n,d,t}),$$

$$\left|\mathbb{E}\left[\Delta^{m,5}_{d,t,j}(W)|(W_i)_{i\in\mathfrak{I}_{-j}}\right]\right| \leqslant C_1 \sup_{\tilde{m}\in\mathcal{J}^m_n}\|\tilde{m}_{d,t}-m_{d,t}\|_{L_2} = O(s_{n,d,t}).$$

Consequently, each term exhibits the following convergence rates on $\mathcal{E}_n$:

$$J^{-1}\sum_{j=1}^{J}\mathbb{E}_{n,j}\left[\Delta^{m,2}_{d,t,j}(W)\right] = O_p(r_{n,1,1}+r_{n,d,t}),$$

$$J^{-1}\sum_{j=1}^{J}\mathbb{E}_{n,j}\left[\Delta^{m,3}_{d,t,j}(W)\right] = O_p(n^{-1/2}(r_{n,1,1}+r_{n,d,t})),$$

$$J^{-1}\sum_{j=1}^{J}\mathbb{E}_{n,j}\left[\Delta^{m,4}_{d,t,j}(W)\right] = O_p(r_{n,1,1}s_{n,d,t}+r_{n,d,t}s_{n,d,t}),$$

$$J^{-1}\sum_{j=1}^{J}\mathbb{E}_{n,j}\left[\Delta^{m,5}_{d,t,j}(W)\right] = O_p(s_{n,d,t}).$$

Thus, the leading rate of (B.34) is $O_p(r_{n,1,1}s_{n,d,t}+r_{n,d,t}s_{n,d,t})$, which is $o_p\left(n^{-1/2}\right)$ under Assumption 6.1(iii).

To complete the proof, we collect the results from (B.31), (B.32), (B.33), and (B.35). From these, we deduce

$$\begin{aligned}\hat{\tau}^{cf}_{dr,J}-\tau =& n^{-1}\sum_{i=1}^{n}\psi_{d,t}(W;w,m)-\tau+\tau\cdot\left(1-\frac{\mathbb{E}_n[DT]}{\mathbb{E}[DT]}\right)+o_p\left(n^{-1/2}\right)\\ =& n^{-1}\sum_{i=1}^{n}\eta_{\text{eff}}(W_i)+o_p\left(n^{-1/2}\right).\end{aligned}$$

Asymptotic normality then follows directly by applying the CLT to the linear expansion. ■

Proof of Theorem 7.1:

Proof of Part (a): The derivation of the EIF proceeds by the same steps used in Theorem 2.1.

Step 1: The density of the observed data $W^{oc}=(RY_0,RY_1,(1-R)Y,(1-R)T,D,X,R)$ is given by

$$f(y,y_1,y_0,t,d,r,x) = (f^p(y_1,y_0,d,x)\cdot\bar{r})^r\cdot(f^{rc}(y,d,t,x)\cdot(1-\bar{r}))^{1-r},$$

where $\bar{r}=\mathbb{P}(R=1)$ and

$$\begin{aligned}f^p(y_1,y_0,d,x) =& f(y_1,y_0|1,1,x)^d\cdot p_1(x)^d\cdot f(y_1,y_0|0,1,x)^{1-d}\cdot(1-p_1(x))^{1-d}f(x|1)\\ f^{rc}(y,d,t,x) =& f(y|1,1,0,x)^{dt}\cdot f(y|1,0,0,x)^{d(1-t)}\cdot f(y|0,1,0,x)^{(1-d)t}\cdot f(y|0,0,0,x)^{(1-d)(1-t)}\\ &\cdot p_0(1,1,x)^{dt}\cdot p_0(1,0,x)^{d(1-t)}\cdot p_0(0,1,x)^{(1-d)t}\cdot p_0(0,0,x)^{(1-d)(1-t)}\cdot f(x|0),\end{aligned}$$

where $f(y_1,y_0|d,r,x)$, $f(y|d,t,r,x)$, $f(x|r)$ are shorthand for $f(y_1,y_0|D=d,R=r,X=x)$, $f(y|D=d,T=t,R=r,X=x)$, and $f(x|R=r)$, respectively.

Consider the regular sub-model parameterized by $\theta \geqslant 0$, with the true model indexed by $\theta_0 = 0$,

$$f_\theta(w^{oc}) = f_\theta^p(y_1, y_0, d, x)^r \cdot f_\theta^{rc}(y, d, t, x)^{1-r} \cdot \bar{r}_\theta^r \cdot (1 - \bar{r}_\theta)^{1-r}.$$

The corresponding score function is given by

$$s_\theta(w^{oc}) = r \cdot s_\theta^p(y_1, y_0, d, x) + (1 - r) \cdot s_\theta^{rc}(y, d, t, x) + s_\theta^r(r),$$

where

$$\begin{aligned}
s_\theta^p(y_1, y_0, d, x) =& d \cdot s_\theta(y_1, y_0|1, 1, x) + (1 - d) \cdot s_\theta(y_1, y_0|0, 1, x) + \frac{d - p_{1,\theta}(x)}{p_{1,\theta}(x)(1 - p_{1,\theta}(x))}\dot{p}_{1,\theta}(x) + t_{1,\theta}(x), \\
s_\theta^{rc}(y, d, t, x) =& dt \cdot s_\theta(y|1, 1, 0, x) + d(1 - t) \cdot s_\theta(y|1, 0, 0, x) \\
& + (1 - d)t \cdot s_\theta(y|0, 1, 0, x) + (1 - d)(1 - t) \cdot s_\theta(y|0, 0, 0, x) \\
& + dt \cdot \frac{\dot{p}_{0,\theta}(1, 1, x)}{p_{0,\theta}(1, 1, x)} + (1 - d)t \cdot \frac{\dot{p}_{0,\theta}(0, 1, x)}{p_{0,\theta}(0, 1, x)} \\
& + d(1 - t) \cdot \frac{\dot{p}_{0,\theta}(1, 0, x)}{p_{0,\theta}(1, 0, x)} + (1 - d)(1 - t) \cdot \frac{\dot{p}_{0,\theta}(0, 0, x)}{p_{0,\theta}(0, 0, x)} \\
& + t_{0,\theta}(x), \\
s_\theta^r(r) =& \frac{r - \bar{r}_\theta}{\bar{r}_\theta \cdot (1 - \bar{r}_\theta)} \cdot \dot{\bar{r}}_\theta,
\end{aligned}$$

$s_\theta(y_1, y_0|d, 1, x) = \partial \log f_\theta(y_1, y_0|d, 1, x)/\partial\theta$, $s_\theta(y|d, t, 0, x) = \partial \log f_\theta(y|d, t, 0, x)/\partial\theta$, $\dot{p}_{1,\theta}(x) = \partial p_{1,\theta}(x)/\partial\theta$, $\dot{p}_{0,\theta}(d, t, x) = \partial p_{0,\theta}(d, t, x)/\partial\theta$, $t_{r,\theta}(x) = \partial \log f_\theta(x|r)/\partial\theta$, and $\dot{\bar{r}}_\theta = d\bar{r}_\theta/d\theta$. Given the score functions, the tangent space of this parametric model can be characterized by

$$\begin{aligned}
\mathcal{T}^{oc} = \{r \cdot S^p(y_1, y_0, d, x) + (1 - r) \cdot S^{rc}(y, d, t, x) + c_1 \cdot (r - c_2) : S^p \in \mathcal{T}^p, S^{rc} \in \mathcal{T}^{rc}, \\
c_1 \in \mathbb{R}, c_2 \in (0, 1)\},
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{T}^p =& \{ds_{1,1}(y_1, y_0, x) + (1 - d)s_{1,0}(y_1, y_0, x) + a_1(x)(d - a_2(x)) + l_1(x)\}, \\
\mathcal{T}^{rc} =& \{dts_{0,11}(y, x) + d(1 - t)s_{0,10}(y, x) + (1 - d)ts_{0,01}(y, x) + (1 - d)(1 - t)s_{0,00}(y, x) \\
& + dtp_{0,11}(x) + d(1 - t)p_{0,10}(x) + (1 - d)tp_{0,01}(x) + (1 - d)(1 - t)p_{0,00}(x) + l_0(x)\},
\end{aligned}$$

for any functions$\{s_{1,d}(\cdot, \cdot, \cdot), s_{0,dt}(\cdot, \cdot), p_{0,dt}(\cdot), a_1(\cdot), a_2(\cdot), l_1(\cdot), l_0(\cdot)\}_{(d,t)\in\mathcal{S}}$ that satisfy the following constraints:

$$s_{1,d}(\cdot, \cdot, \cdot) \in L_2(\mathcal{Y} \times \mathcal{Y} \times \mathcal{X}), \text{ with } \int\int s_{1,d}(y_1, y_0, x) f(y_1, y_0|d, 1, x) dy_1 dy_0 = 0, \ \forall x \in \mathcal{X}, \tag{B.36}$$

$$s_{0,dt}(\cdot, \cdot) \in L_2(\mathcal{Y} \times \mathcal{X}), \text{ with } \int s_{0,dt}(y, x) f(y|d, t, 0, x) dy = 0, \ \forall x \in \mathcal{X}, \tag{B.37}$$

$$p_{0,dt}(\cdot) \in L_2(\mathcal{X}), \text{ with } \sum_{(d,t)\in\mathcal{S}} p_{0,dt}(x) p_0(d, t, x) = 0, \ \forall x \in \mathcal{X}, \tag{B.38}$$

$$l_r(\cdot) \in L_2(\mathcal{X}), \text{ with } \int l_r(x) f(x|r) dx = 0, r \in \{0, 1\}, \tag{B.39}$$

$0 < a_2(\cdot) < 1$, and $a_1(\cdot) \in L_2(\mathcal{X})$.

In Step 2, we show that the target parameter associated with the parametric sub-model is *path-wise* *differentiable*. Following standard arguments such as those used in Lemma B.1, we can show that the ATT can be identified by

$$\begin{aligned}\tau^{oc} =& \mathbb{P}(R=1)\cdot\mathbb{E}\left[\mathbb{E}\left[\Delta Y|D=1,R=1,X\right]-\mathbb{E}\left[\Delta Y|D=0,R=1,X\right]|R=1\right]\\&+\mathbb{P}(R=0)\cdot\mathbb{E}\left[\sum_{(d,t)\in\mathcal{S}}(-1)^{d+t}\,\mathbb{E}\left[Y|D=d,T=t,R=0,X\right]|R=0\right]\end{aligned}$$

under Assumption 7. In the parametric sub-model, we can represent the target parameter as follows:

$$\begin{aligned}\tau^{oc}(\theta) =& \bar{r}_\theta\cdot\underbrace{\left(\sum_{d\in\{0,1\}}(-1)^{d+1}\frac{\int\int\int(y_1-y_0)f_\theta(y_1,y_0|d,1,x)p_{1,\theta}(x)f_\theta(x|1)dy_1dy_0dx}{\int p_{1,\theta}(x)f_\theta(x|1)dx}\right)}_{\equiv\tau_p^{oc}(\theta)}\\&+(1-\bar{r}_\theta)\cdot\underbrace{\left(\sum_{(d,t)\in\mathcal{S}}(-1)^{d+t}\frac{\int\int yf_\theta(y|d,t,0,x)p_{0,\theta}(1,1,x)f_\theta(x|0)dydx}{\int p_{0,\theta}(1,1,x)f_\theta(x|0)dx}\right)}_{\equiv\tau_{rc}^{oc}(\theta)}.\end{aligned}$$

The derivative of $\tau^{oc}(\theta)$ with respect to θ, evaluated at $\theta_0=0$, is then given by

$$\left.\frac{d\tau^{oc}(\theta)}{d\theta}\right|_{\theta=0}=\bar{r}_0\cdot\left.\frac{d\tau_p^{oc}(\theta)}{d\theta}\right|_{\theta=0}+(1-\bar{r}_0)\cdot\left.\frac{d\tau_{rc}^{oc}(\theta)}{d\theta}\right|_{\theta=0}+\dot{\bar{r}}_0\cdot(\tau_p^{oc}(0)-\tau_{rc}^{oc}(0)),$$

where

$$\begin{aligned}\left.\frac{d\tau_p^{oc}(\theta)}{d\theta}\right|_{\theta=0}=&\sum_{d\in\{0,1\}}(-1)^{d+1}\frac{\int\int\int(y_1-y_0)s(y_1,y_0|d,1,x)f(y_1,y_0|d,1,x)p_1(x)f(x|1)dy_1dy_0dx}{p_1}\\&+\frac{\int(m_{1,\Delta}^p(x)-m_{0,\Delta}^p(x)-\tau_p^{oc})\cdot\dot{p}_1(x)f(x|1)dx}{p_1}\\&+\frac{\int(m_{1,\Delta}^p(x)-m_{0,\Delta}^p(x)-\tau_p^{oc})\cdot p_1(x)t_1(x)f(x|1)dx}{p_1},\\\left.\frac{d\tau_{rc}^{oc}(\theta)}{d\theta}\right|_{\theta=0}=&\sum_{(d,t)\in\mathcal{S}}(-1)^{d+t}\frac{\int\int ys(y|d,t,0,x)f(y|d,t,0,x)p_0(1,1,x)f(x|R=0)dydx}{p_0(1,1)}\\&+\frac{\int(\sum_{(d,t)\in\mathcal{S}}m_{d,t}^{rc}(x)-\tau_{rc}^{oc})\cdot\dot{p}_0(1,1,x)f(x|0)dx}{p_0(1,1)}\\&+\frac{\int(\sum_{(d,t)\in\mathcal{S}}m_{d,t}^{rc}(x)-\tau_{rc}^{oc})\cdot p_0(1,1,x)t_0(x)f(x|0)dx}{p_0(1,1)},\end{aligned}$$

and $p_1=\mathbb{P}(D=1|R=1)$. We now define the following candidate EIFs:

$$\begin{aligned}&F^{oc}(w^{oc})=r\cdot F^p(y_1,y_0,d,x)+(1-r)\cdot F^{rc}(y,d,t,x)+F^r(r),\\&F^p(y_1,y_0,d,x)=\Big\{w_1^p(d)\left(m_{1,\Delta}^p(x)-m_{0,\Delta}^p(x)-\tau_p^{oc}\right)\\&\qquad\qquad+w_1^p(d)(\Delta y-m_{1,\Delta}^p(x))-w_0^p(d,x)(\Delta y-m_{0,\Delta}^p(x))\Big\},\\&F^{rc}(y,d,t,x)=w_{1,1}^{rc}(d,t)(\tau^{rc}(y,x)-\tau_{rc}^{oc})+\sum_{(d,t)\in\mathcal{S}_-}(-1)^{(d+t)}w_{d,t}^{rc}(d,t,x)(y-m_{d,t}^{rc}(x)),\\&F^r(r)=r\cdot\tau_p^{oc}+(1-r)\cdot\tau_{rc}^{oc}-\tau^{oc}=(r-\bar{r})\cdot(\tau_p^{oc}-\tau_{rc}^{oc}).\end{aligned}$$

From direct calculation, we obtain,

$$\begin{aligned}\mathbb{E}[F^{oc}(W^{oc})\cdot s_0(W^{oc})] =& \mathbb{E}[(R\cdot F^p(Y_1,Y_0,D,X)+(1-R)\cdot F^{rc}(Y,D,T,X)+F^r(R))\\ &\cdot(R\cdot s_0^p(Y_1,Y_0,D,X)+(1-R)\cdot s_0^{rc}(Y,D,T,X)+s_0^r(R))]\\ =&\mathbb{E}[(R\cdot F^p(Y_1,Y_0,D,X)\cdot s_0^p(Y_1,Y_0,D,X)\\ &+(1-R)\cdot F^{rc}(Y,D,T,X)\cdot s_0^{rc}(Y,D,T,X)\\ &+F^r(R)\cdot s_0(W^{oc})\\ &+(R\cdot F^p(Y_1,Y_0,D,X)+(1-R)\cdot F^{rc}(Y,D,T,X))\cdot s_0^r(R)].\end{aligned}$$

Arguing along the same lines as in the proof of Theorem 2.1 and that of Proposition 1(a) in Sant'Anna and Zhao (2020), we deduce that

$$\begin{aligned}\mathbb{E}[R\cdot F^p(Y_1,Y_0,D,X)\cdot s_0^p(Y_1,Y_0,D,X)] &= \bar{r}\cdot\mathbb{E}[F^p(Y_1,Y_0,D,X)\cdot s_0^p(Y_1,Y_0,D,X)|R=1]\\ &=\bar{r}\cdot\left.\frac{d\tau_p^{oc}(\theta)}{d\theta}\right|_{\theta=0}, \qquad\qquad (B.40)\\ \mathbb{E}[(1-R)\cdot F^{rc}(Y,D,T,X)\cdot s_0^{rc}(Y,D,T,X)] &= (1-\bar{r})\cdot\mathbb{E}[F^{rc}(Y,D,T,X)\cdot s_0^{rc}(Y,D,T,X)|R=0]\\ &=(1-\bar{r})\cdot\left.\frac{d\tau_{rc}^{oc}(\theta)}{d\theta}\right|_{\theta=0}. \qquad\qquad (B.41)\end{aligned}$$

Furthermore,

$$\begin{aligned}\mathbb{E}[F^r(R)\cdot s_0(W^{oc})] &= \mathbb{E}[F^r(R)\cdot(R\cdot s_0^p(Y_1,Y_0,D,X)+(1-R)\cdot s_0^{rc}(Y,D,T,X)+s_0^r(R))]\\ &=\mathbb{E}[F^r(R)\cdot s_0^r(R)]\\ &=\mathbb{E}\left[\frac{(R-\bar{r})^2}{\bar{r}(1-\bar{r})}\cdot(\tau_p^{oc}-\tau_{rc}^{oc})\cdot\dot{\bar{r}}\right]\\ &=\dot{\bar{r}}\cdot(\tau_p^{oc}-\tau_{rc}^{oc}), \qquad\qquad (B.42)\end{aligned}$$

where the second equality is due to

$$\begin{aligned}&\mathbb{E}[F^r(R)\cdot R\cdot s_0^p(Y_1,Y_0,D,X)] = \mathbb{E}[R\cdot F^R(1)]\cdot\mathbb{E}[s_0^p(Y_1,Y_0,D,X)|R=1]=0,\\ &\mathbb{E}[F^r(R)\cdot(1-R)\cdot s_0^{rc}(Y,D,T,X)] = \mathbb{E}[(1-R)\cdot F^R(0)]\cdot\mathbb{E}[s_0^{rc}(Y,D,T,X)|R=0]=0.\end{aligned}$$

These two lines follows directly from the fact that the true score functions s_0^p and s_0^{rc} satisfy $\mathbb{E}[s_0^p(Y_1,Y_0,D,X)|R=1]=0$ and $\mathbb{E}[s_0^{rc}(Y,D,T,X)|R=0]=0$, respectively.

Next,

$$\begin{aligned}\mathbb{E}[R\cdot F^p(Y_1,Y_0,D,X)\cdot s_0^r(R)] &= \mathbb{E}\left[\frac{R}{\bar{r}}\cdot\dot{\bar{r}}\cdot\mathbb{E}[F^p(Y_1,Y_0,D,X)|R=1]\right]\\ &=\dot{\bar{r}}\cdot\mathbb{E}[F^p(Y_1,Y_0,D,X)|R=1]\\ &=0, \qquad\qquad (B.43)\end{aligned}$$

where the last equation follows due to $\mathbb{E}[F^p(Y_1,Y_0,D,X)|R=1]=0$. Analogously, since $\mathbb{E}[F^{rc}(Y,D,T,X)|R=0]=0$, we have

$$\mathbb{E}[(1-R)\cdot F^{rc}(Y,D,T,X)\cdot s_0^r(R)] = \mathbb{E}\left[-\frac{1-R}{1-\bar{r}}\cdot\dot{\bar{r}}\cdot\mathbb{E}[F^{rc}(Y,D,T,X)|R=0]\right]$$

$$= -\dot{\bar{r}} \cdot \mathbb{E}[F^{rc}(Y, D, T, X)|R = 0]$$
$$= 0. \tag{B.44}$$

Collecting results in (B.40) through (B.44), we deduce that

$$\mathbb{E}[F^{oc}(W^{oc}) \cdot s_0(W^{oc})] = \left.\frac{d\tau^{oc}(\theta)}{d\theta}\right|_{\theta=0}.$$

In Step 3, we show $F^{oc}(W^{oc})$ is indeed the efficient influence function for τ^{oc}, by verifying that $F^{oc} \in \mathcal{T}^{oc}$. To that end, we let $c_m^p(x) = p_1{}^{-1} \cdot \left(m_{1,\Delta}^p(x) - m_{0,\Delta}^p(x) - \tau_p^{oc}\right)$ and $c_m^{rc}(x) = p_0(1,1)^{-1} \cdot \left(\sum_{(d,t)\in\mathcal{S}}(-1)^{d+t} m_{d,t}^{rc}(x) - \tau_{rc}^{oc}\right)$. Moreover, leveraging the following quantities

$$\begin{aligned}
&s_{1,1}(y_1, y_0, x) = p_1^{-1} \cdot (y_1 - y_0 - m_{1,\Delta}^p(x)), \quad s_{1,0}(y_1, y_0, x) = \frac{p_1(x)(y_1 - y_0 - m_{0,\Delta}^p(x))}{p_1 \cdot (1 - p_1(x))}, \\
&a_1(x) = c_m^p(x), \quad a_2(x) = p_1(x), \quad l_1(x) = p_1(x) \cdot c_m^p(x), \\
&s_{0,11}(y, x) = \frac{y - m_{1,1}^{rc}(x)}{p_0(1,1)}, \quad s_{0,dt}(y, x) = (-1)^{d+t} \cdot \frac{p_0(1,1,x) \cdot (y - m_{d,t}^{rc}(x))}{p_0(d,t,x) \cdot p_0(1,1)}, \text{ for } (d,t) \in \mathcal{S}_-, \\
&p_{0,11}(x) = (1 - p_0(1,1,x)) \cdot c_m^{rc}(x), \quad p_{0,dt}(x) = -p_0(1,1,x) \cdot c_m^{rc}(x), \text{ for } (d,t) \in \mathcal{S}_-, \\
&l_0(x) = p_0(1,1,x) \cdot c_m^{rc}(x), \quad c_1 = \left(\tau_p^{oc} - \tau_{rc}^{oc}\right), \text{ and } c_2 = \bar{r},
\end{aligned}$$

one can readily verify that constraints imposed on the tangent space $\mathcal{T}^{oc}$ (i.e., (B.36)–(B.39)) are satisfied under these function choices.

To prove part (b), we note that the semiparametric efficiency bound can be computed as follows

$$\begin{aligned}
\mathbb{E}[\eta^{oc}(W^{oc})^2] &= \mathbb{E}[R \cdot \eta_p(Y_1, Y_0, D, X)^2] + \mathbb{E}[(1 - R) \cdot \eta_{rc}(Y, D, T, X)^2] + \mathbb{E}[\eta_r(R)^2] \\
&\quad + \mathbb{E}[(R \cdot \eta_p(Y_1, Y_0, D, X) + (1 - R) \cdot \eta_{rc}(Y, D, T, X)) \cdot \eta_r(R)] \\
&= \mathbb{E}[R] \cdot \mathbb{E}[\eta_p(Y_1, Y_0, D, X)^2 | R = 1] + \mathbb{E}[1 - R] \cdot \mathbb{E}[\eta_{rc}(Y, D, T, X)^2 | R = 0] + V_r^{oc} \\
&= \mathbb{E}[R] \cdot V_p^{oc} + \mathbb{E}[1 - R] \cdot V_{rc}^{oc} + V_r^{oc},
\end{aligned}$$

where the second last equality holds because, by the LIE, the cross-product term in the second line has mean zero. ■

C Results on asymptotic linear expansion of local polynomial estimators

In the next subsection, we provide some well-known results about the U-statistics, based on which, we derive uniform stochastic expansions of local polynomial estimators in Section C.2.

C.1 Rates of convergence: U-Statistic

Let $\{X_i\}_{i=1}^n$ be a random sample from an unknown distribution. Let

$$h : \mathcal{X}^r \mapsto \mathbb{R}, \qquad (x_1, \ldots, x_r) \mapsto h(x_1, \ldots, x_r),$$

be a (possibly sample-size dependent) real-valued measurable function. Define the rth-order U-statistic with kernel h as

$$U_n = \frac{(n-r)!}{n!} \sum_{(s_1,\ldots,s_r)\in S(n,r)} h(X_{s_1},\ldots,X_{s_r}),$$

where $S(n,r)$ is the set of all $n!/(n-r)!$ ordered permutations $(s_1,\ldots,s_r)$ of size r from the index set $\{1,\ldots,n\}$.

Since any kernel h can be symmetrized, we can replace it by its symmetrized version

$$h^{(s)}(x_1,\ldots,x_r) = \frac{1}{r!} \sum_{\pi\in S(r,r)} h(x_{\pi(1)},\ldots,x_{\pi(r)}),$$

where the summation ranges over all permutations of $\{1,\ldots,r\}$. Restricting attention to symmetric kernels, U_n can be equivalently written as

$$U_n = \binom{n}{r}^{-1} \sum_{(s_1,\ldots,s_r)\in C(n,r)} h^{(s)}(X_{s_1},\ldots,X_{s_r}),$$

where $C(n,r)$ is the set of all $\binom{n}{r}$ unordered combinations $(s_1,\ldots,s_r)$ of size r from $\{1,\ldots,n\}$.

For $1 \leqslant s \leqslant r$, define

$$h_s(x_1,\ldots,x_s) = \mathbb{E}\big[h(x_1,\ldots,x_s,X_{s+1},\ldots,X_r)\big], \qquad \sigma_s = \sqrt{\mathbb{V}\mathrm{ar}\,[h_s(X_1,\ldots,X_s)]}.$$

We say that a U-statistic with kernel h is s^**-th order degenerate* if $\sigma_s = 0$ for all $s \leqslant s^*$.

Lemma C.1 Let $h : \mathcal{X}^r \to \mathbb{R}$ be a measurable, permutation-symmetric function of r arguments such that $\mathbb{E}[h(X_1,\ldots,X_r)] = 0$ and $\sigma_r < \infty$. Then $U_n = O_p\big(\sum_{s=1}^r n^{-s/2}\sigma_s\big)$. In particular, if the U-statistic is s^*th-order degenerate, its convergence rate simplifies to $U_n = O_p\big(\sum_{s=s^*+1}^r n^{-s/2}\sigma_s\big)$.

This rate is a standard result in U-statistic theory and follows directly from a variance calculation combined with Markov's inequality. For detailed derivations and formal results, see e.g., Appendix B.1 of Rothe and Firpo (2019a), Section 12.1 of van der Vaart (1998), and Sherman (1994).

C.2 Asymptotic linear expansion of local polynomial estimators

In this section, we provide some results on the asymptotic expansion of the local polynomial estimators.

For $(d,t) \in \mathcal{S}_-$, we define the summand of the (local) score function as

$$\widetilde{A}_{d,t}(W,x,\gamma) = \left(I_{d,t} - \frac{\exp(\underline{\mathbf{X}}(x_c)'\gamma_{d,t})}{1+\sum_{(d',t')\in\mathcal{S}_-}\exp(\underline{\mathbf{X}}(x_c)'\gamma_{d',t'})} \right) H(h)\underline{\mathbf{X}}(x_c)\widetilde{K}_{ps}(X;x,h,\lambda),$$

where $H(h)$ is a diagonal matrix with the main diagonal entries being $h^{-|\mathbf{k}|}$, for lexicographic-ordered $\mathbf{k}$, with $0 \leqslant |\mathbf{k}| \leqslant p$. Here, we have dropped the subscript of $\underline{\mathbf{X}}$ to ease notational burden. We let $\boldsymbol{\iota}_-(\{S_{d,t}\}_{(d,t)\in\mathcal{S}_-}) = (S'_{1,0}, S'_{0,1}, S'_{0,0})'$. The local Fisher information matrix evaluated at x can be approximated as

$$\mathcal{I}(x) = diag(\mathbf{p}_-(x)) - \mathbf{p}_-(x)\mathbf{p}'_-(x), \tag{C.1}$$

where $\mathbf{p}_-(x) = (p(1,0,x), p(0,1,x), p(0,0,x))$. In addition, we define the local hessian as

$$\Sigma^{ps}(x) = \mathbb{E}[\mathcal{I}(X) \otimes H(h)\underline{\mathbf{X}}(x_c)\underline{\mathbf{X}}(x_c)'H(h)\widetilde{K}_{ps}(X;x,h,\lambda)].$$

With these notations in hand, we can introduce several quantities associated with the linear expansion of the PS estimator. For each $(d,t) \in \mathcal{S}_{-}$,

$$\begin{aligned}
A_{d,t}(W,x) &= (e_{3,\iota(d,t)} \otimes e_{N_p,1})' \Sigma^{ps}(x)^{-1} \widetilde{\boldsymbol{A}}_{-}(W,x,\gamma^*(x)), \\
G^{(ps)}_{d,t}(W,x) &= e'_{3,\iota(d,t)} \mathcal{I}(x) \mathbf{A}_{-}(W,x),
\end{aligned}$$

where $\widetilde{\boldsymbol{A}}_{-}(W,x,\gamma) = \boldsymbol{\iota}_{-}(\{\tilde{A}_{d,t}(W,x,\gamma)\}_{(d,t)\in\mathcal{S}_{-}})$, and $\mathbf{A}_{-}(W,x) = \boldsymbol{\iota}_{-}(\{A_{d,t}(W,x)\}_{(d,t)\in\mathcal{S}_{-}})$. For the treated group in $t=1$, let $G^{(ps)}_{1,1}(W,x) = -\sum_{(d,t)\in\mathcal{S}_{-}} G^{(ps)}_{d,t}(W,x)$. Additionally, we define, for a given observation X_j

$$\begin{aligned}
B^{(ps)}_{n,d,t}(X_j) &= \mathbb{E}[G^{(ps)}_{d,t}(W_i,X_j)|X_j], \qquad \text{(C.2)}\\
S^{(ps)}_{n,d,t}(X_j) &= \frac{1}{n-1}\sum_{i\neq j} G^{(ps)}_{d,t}(W_i,X_j) - \mathbb{E}[G^{(ps)}_{d,t}(W_i,X_j)|X_j], \\
R^{(ps)}_{n,d,t}(X_j) &= \hat{p}(d,t,X_j) - p(d,t,X_j) - B^{(ps)}_{n,d,t}(X_j) - S^{(ps)}_{n,d,t}(X_j).
\end{aligned}$$

The three quantities represent the bias, the first-order stochastic part, and the remaining terms derived from the decomposition of the PS estimator, respectively.

Focusing on the OR model, for $(d,t) \in \mathcal{S}$, the leave-one-out local polynomial estimator has a closed-form solution given by

$$\widehat{m}_{d,t}(X_j) = \frac{1}{n-1}\sum_{i\neq j} e'_{N_q,1} \hat{\Sigma}^{or\;-1}_{d,t}(X_j) H(b_{d,t}) \underline{\mathbf{X}}_i(X_j) I_{d,t,i} Y_i \widetilde{K}_{or}(X_i;X_j,b_{d,t},\vartheta_{d,t}),$$

where $\widehat{\Sigma}^{or}_{d,t}(X_j) = \frac{1}{n-1}\sum_{i\neq j} I_{d,t,i} H(b_{d,t}) \underline{\mathbf{X}}_i(x_c) \underline{\mathbf{X}}_i(x_c)' H(b_{d,t}) \widetilde{K}_{or}(X_i;X_j,b_{d,t},\vartheta_{d,t})$.

Analogous to the PS case, we use $B^{(or)}_{n,d,t}$, $S^{(or)}_{n,d,t}$, and $R^{(or)}_{n,d,t}$ to represent the bias, the first-order stochastic and the remainder terms, respectively. For a given observation X_j, these quantities are specified as

$$\begin{aligned}
B^{(or)}_{n,d,t}(X_j) &= \mathbb{E}[G^{(or)}_{d,t}(W_i,X_j)|X_j], \\
S^{(or)}_{n,d,t}(X_j) &= \frac{1}{n-1}\sum_{i\neq j} G^{(or)}_{n,d,t}(W_i,X_j) - \mathbb{E}[G^{(or)}_{d,t}(W_i,X_j)|X_j], \\
R^{(or)}_{n,d,t}(X_j) &= \hat{m}_{d,t}(X_j) - m_{d,t}(X_j) - B^{(or)}_{n,d,t}(X_j) - S^{(or)}_{n,d,t}(X_j),
\end{aligned}$$

where

$$\begin{aligned}
G^{(or)}_{d,t}(W_i,X_j) &= e'_{N_q,1}\Sigma^{or}_{d,t}(X_j)^{-1} H(b_{d,t}) \underline{\mathbf{X}}_i(X_j) I_{d,t,i} \xi^{or}_{d,t,i}(X_j) \widetilde{K}_{or}(X_i;X_j,b_{d,t},\vartheta_{d,t}), \\
\Sigma^{or}_{d,t}(x) &= \mathbb{E}[I_{d,t,i} H(b_{d,t}) \underline{\mathbf{X}}_i(x_c) \underline{\mathbf{X}}_i(x_c)' H(b_{d,t}) \widetilde{K}_{or}(X;x,b_{d,t},\vartheta_{d,t})], \\
\xi^{or}_{d,t}(x) &= (Y - \underline{\mathbf{X}}(x)'\beta^*_{d,t}).
\end{aligned}$$

Lemma C.2 Suppose Assumptions 1, 2, and 5 are satisfied. In addition, Assumptions 5.2(ii) and 5.5(iv)-(vii) hold for $(d,t) = (1,1)$. Then, for $(d,t) \in \mathcal{S}$,

$$\sup_{j\in\mathbb{N}_n} \left| B^{(ps)}_{n,d,t}(X_j) \right| = O_p(h^{p+1} + \lambda_o + \lambda_u), \qquad \text{(C.3)}$$

$$\sup_{j\in\mathbb{N}_n} \left| S^{(ps)}_{n,d,t}(X_j) \right| = O_p\left(\sqrt{\log n/\left(nh^{v_c}\right)}\right), \qquad \text{(C.4)}$$

$$\sup_{j\in\mathbb{N}_n}\left|R^{(ps)}_{n,d,t}(X_j)\right| = O_p\left(\left(h^{p+1}+\lambda_o+\lambda_u+\sqrt{\log n/\left(nh^{\upsilon_c}\right)}\right)^2\right), \tag{C.5}$$
$$\sup_{j\in\mathbb{N}_n}\left|B^{(or)}_{n,d,t}(X_j)\right| = O_p(b_{d,t}^{q+1}+\vartheta_{d,t,o}+\vartheta_{d,t,u}),$$
$$\sup_{j\in\mathbb{N}_n}\left|S^{(or)}_{n,d,t}(X_j)\right| = O_p\left(\sqrt{\log n/\left(nb_{d,t}{}^{\upsilon_c}\right)}\right),$$
$$\sup_{j\in\mathbb{N}_n}\left|R^{(or)}_{n,d,t}(X_j)\right| = O_p\left(\left(b_{d,t}^{q+1}+\vartheta_{d,t,o}+\vartheta_{d,t,u}+\sqrt{\log n/\left(nb_{d,t}{}^{\upsilon_c}\right)}\right)^2\right).$$

Before stating the proof, we need to introduce some additional notations. Since kernel functions K and L are supported on $[-1,1]^{\upsilon_c}$, the effective support of $K((\cdot-x_c)/h)$ is $\mathcal{S}_{x_c,h}=\{z: x_c+hz\in\mathcal{X}\}\cap[-1,1]^{\upsilon_c}$. When $\mathcal{S}_{x_c,h}=[-1,1]^{\upsilon_c}$, x is an interior point, otherwise x lies close to the boundary. For any measurable set $\mathcal{S}\subset[-1,1]^{\upsilon_c}$, let $\nu_{\mathbf{k}}(\mathcal{S})=\int_{\mathcal{S}}\mathbf{u}^{\mathbf{k}}K(\mathbf{u})d\mathbf{u}$ and $\varkappa_{\mathbf{k}}(\mathcal{S})=\int_{\mathcal{S}}\mathbf{u}^{\mathbf{k}}K^2(\mathbf{u})d\mathbf{u}$. Now we let the $N_\ell\times N_\ell$ matrices $\mathbf{Q}_\ell(x_c)$ and $\mathbf{T}_\ell(x_c)$, and the $N_\ell\times n_k$ matrix $\mathbf{M}_{\ell,k}(x_c)$ be defined as

$$\mathbf{Q}_\ell(x_c)=\begin{pmatrix}\mathbf{Q}^{(0,0)}(\mathcal{S}_{x_c,h}) & \dots & \mathbf{Q}^{(0,\ell)}(\mathcal{S}_{x_c,h})\\ \vdots & \ddots & \vdots\\ \mathbf{Q}^{(\ell,0)}(\mathcal{S}_{x_c,h}) & \dots & \mathbf{Q}^{(\ell,\ell)}(\mathcal{S}_{x_c,h})\end{pmatrix}, \tag{C.6}$$
$$\mathbf{T}_\ell(x_c)=\begin{pmatrix}\mathbf{T}^{(0,0)}(\mathcal{S}_{x_c,h}) & \dots & \mathbf{T}^{(0,\ell)}(\mathcal{S}_{x_c,h})\\ \vdots & \ddots & \vdots\\ \mathbf{T}^{(\ell,0)}(\mathcal{S}_{x_c,h}) & \dots & \mathbf{T}^{(\ell,\ell)}(\mathcal{S}_{x_c,h})\end{pmatrix},$$
$$\mathbf{M}_{\ell,k}(x_c)=\begin{pmatrix}\mathbf{Q}^{(0,k)}(\mathcal{S}_{x_c,h})\\ \dots\\ \mathbf{Q}^{(\ell,k)}(\mathcal{S}_{x_c,h})\end{pmatrix},$$

where $\mathbf{Q}^{(i,j)}_\ell(\mathcal{S})$ and $\mathbf{T}^{(i,j)}_\ell(\mathcal{S})$ are $n_i\times n_j$ matrices with their respective (l,m)-th element given by $\nu_{\pi_i(l)+\pi_j(m)}(\mathcal{S})$ and $\varkappa_{\pi_i(l)+\pi_j(m)}(\mathcal{S})$. When x is a boundary point, these quantities are not invariant to x, and thus, capture the boundary effects.

Proof of Lemma C.2:
Given that our data is a random sample, it is straightforward to show the "leave-one-out" estimators considered in the lemma is asymptotically equivalent to the usual "leave-in" estimators. See Rothe and Firpo (2019b) for a detailed exposition. We therefore focus on the "leave-in" versions in what follows.

We prove the results for PS only. The case for OR follows by generalizing Proposition 7 of Fan and Guerre (2016) to the case where discrete covariates are accommodated. This generalization can be achieved by employing the techniques similar to those presented here.

For (C.3), we have

$$\sup_{x\in\mathcal{X}}\left\|B^{(ps)}_{n,d,t}(x)\right\| = \sup_{x\in\mathcal{X}}\left\|e'_{3,\iota(d,t)}\mathcal{I}(x)(I_3\otimes e'_{N_p,1})\Sigma^{ps}(x)^{-1}\,\mathbb{E}[\widetilde{\boldsymbol{A}}_-(W,x,\gamma^*(x))]\right\|$$
$$\leqslant \sup_{x\in\mathcal{X}}\left\|e'_{3,\iota(d,t)}\mathcal{I}(x)\right\|\cdot\sup_{x\in\mathcal{X}}\left\|(I_3\otimes e'_{N_p,1})\Sigma^{ps}(x)^{-1}\right\|\cdot\sup_{x\in\mathcal{X}}\left\|\mathbb{E}[\widetilde{\boldsymbol{A}}_-(W,x,\gamma^*(x))]\right\|.$$

By definition, $\sup_{x\in\mathcal{X}}\|\mathcal{I}(x)\|=O(1)$. Standard change of variable gives

$$\Sigma^{ps}(x)=\mathcal{I}(x)\otimes\mathbf{Q}_p(x_c)f_X(x)+O\left(h+\lambda_o+\lambda_u\right). \tag{C.7}$$

Since $\inf_{x\in\mathcal{X}} \lambda_{min}(\mathcal{I}(x)\otimes\mathbf{Q}_p(x_c)) = \inf_{x\in\mathcal{X}} \lambda_{min}(\mathcal{I}(x))\cdot\inf_{x_c\in\mathcal{X}_c} \lambda_{min}(\mathbf{Q}_p(x_c)) > 0$ and $\inf_{x\in\mathcal{X}} f_X(x) > 0$ under Assumptions 2(iii), 5.6, and 5.1, we get

$$\sup_{x\in\mathcal{X}} \left\|\mathcal{I}(x)^{-1}\otimes\mathbf{Q}_p(x_c)^{-1}\cdot f_X(x)^{-1}\right\| = O(1), \tag{C.8}$$

and thus, $\sup_{x\in\mathcal{X}}\left\|\Sigma^{ps}(x)^{-1}\right\| = O(1)$. Now, from Lemma C.3, we conclude that $\sup_{x\in\mathcal{X}}\left\|B^{(ps)}_{n,d,t}(x)\right\| = O\left(h^{p+1}+\lambda_o+\lambda_u\right)$.

Having just demonstrated that $\Sigma^{ps}(x)^{-1}$ is uniformly bounded over $\mathcal{X}$, we can now apply Lemma C.3 and the CMT to deduce (C.4).

To establish (C.5), the proof proceed through three steps. First, we demonstrate the existence of a global maximizer for the local log-likelihood function defined in (3.8). Subsequently, we obtain the uniform asymptotic linear expansion for the local maximum likelihood estimator. Finally, we apply the delta method to verify that the remainder term exhibits the required rate.

Step 1: Define $\bar{\gamma} = (I_3\otimes H(h)^{-1})\gamma$ and $\bar{\gamma}^*(\cdot) = (I_3\otimes H(h)^{-1})\gamma^*(\cdot)$. Using the scaled parameters, we rewrite the likelihood as

$$\begin{aligned}\mathcal{L}^{ps}_n(\bar{\gamma};x) =& \frac{1}{n}\sum_{i=1}^n \sum_{(d',t')\in\mathcal{S}_-} I_{d,t}H(h)\underline{\mathbf{X}}(x_c)'\bar{\gamma}_{d,t}\\ &-\log\left(1+\sum_{(d',t')\in\mathcal{S}_-}\exp\left(H(h)\underline{\mathbf{X}}(x_c)'\bar{\gamma}_{d',t'}\right)\right)\widetilde{K}_{ps}(X_i;x,h,\lambda).\end{aligned} \tag{C.9}$$

The gradient and hessian of $\mathcal{L}^{ps}_n(\bar{\gamma};x)$ with respect to $\bar{\gamma}$ are given by

$$\nabla_{\bar{\gamma}}\mathcal{L}^{ps}_n(\bar{\gamma};x) = \frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{A}}_-(W_i,x,\gamma), \qquad \nabla^2_{\bar{\gamma}\bar{\gamma}'}\mathcal{L}^{ps}_n(\bar{\gamma};x) = -\frac{1}{n}\sum_{i=1}^n \mathbf{H}(W_i,x,\gamma),$$

where

$$\begin{aligned}\mathbf{H}(X,x,\gamma) &= \mathcal{I}(X_c,x_c,\gamma)\otimes\widetilde{H}(X,x,h,\lambda),\\ \widetilde{H}(X,x,h,\lambda) &= H(h)\underline{\mathbf{X}}(x_c)\underline{\mathbf{X}}(x_c)'H(h)\widetilde{K}_{ps}(X;x,h,\lambda),\\ \mathcal{I}(X_c,x_c,\gamma) &= diag(\boldsymbol{\Psi}_-(X_c,x_c,\gamma)) - \boldsymbol{\Psi}_-(X_c,x_c,\gamma)\boldsymbol{\Psi}_-(X_c,x_c,\gamma)',\\ \boldsymbol{\Psi}_-(X,x,\gamma) &= \boldsymbol{\iota}_-(\{\Psi_{d,t}(\underline{\mathbf{X}}(x),\gamma)\}_{(d,t)\in\mathcal{S}_-}),\\ \Psi_{d,t}(x,\gamma) &= \frac{\mathbb{1}\{dt\neq 1\}\exp\left(x'\gamma_{d,t}\right)+\mathbb{1}\{dt=1\}}{1+\sum_{(d',t')\in\mathcal{S}_-}\exp\left(x'\gamma_{d',t'}\right)}.\end{aligned}$$

Next, we define the following two events

$$\begin{aligned}E_{1n}(c) &= \left\{\sup_{x\in\mathcal{X}}\left\|\frac{1}{n}\sum_{i=1}^n\widetilde{\boldsymbol{A}}_-(W_i,x,\gamma^*(x))\right\| < c\kappa_n\right\},\\ E_{2n}(c) &= \left\{\inf_{x\in\mathcal{X}}\lambda_{min}\left(\frac{1}{n}\sum_{i=1}^n\widetilde{H}(X;x,h,\lambda)\right) > c\right\},\end{aligned}$$

for $c>0$ and $\kappa_n = \sqrt{\log n/\left(nh^{v_c}\right)}+h^{p+1}+\lambda_u+\lambda_o$.

By Lemma C.3, we deduce that $\mathbb{P}\left(E_{1n}(c_1)\right)\to 1$, for any fixed $c_1>0$.

Now, standard change-of-variable analysis gives

$$\mathbb{E}[\widetilde{H}(X;x,h,\lambda)] = \mathbf{Q}_p(x_c)f_X(x) + O\left(h+\lambda_o+\lambda_u\right).$$

Under Assumptions 5.1 and 5.6, $\inf_{x\in\mathcal{X}} f_X(x) > 0$ and $\inf_{x_c\in\mathcal{X}_c} \lambda_{min}\left(\mathbf{Q}_p(x_c)\right) > 0$. As a result, there exists $c_2 > 0$ such that $\inf_{x\in\mathcal{X}} \lambda_{min}(\mathbb{E}[\widetilde{H}(X;x,h,\lambda)]) \geqslant c_2$, when n is sufficiently large. Coupled with the fact that

$$\sup_{x\in\mathcal{X}}\left\|\frac{1}{n}\sum_{i=1}^{n}\widetilde{H}(X_i;x,h,\lambda) - \mathbb{E}[\widetilde{H}(X;x,h,\lambda)]\right\| = O_p\left(\sqrt{\log n/\left(nh^{v_c}\right)}\right).$$

which is a consequence of Lemma 5 from Fan and Guerre (2016), we deduce that $\mathbb{P}\left(E_{2n}(c)\right) \to 1$, for $c \leqslant c_2$.

Next, we define a neighborhood of $\bar{\gamma}^*(\cdot)$,

$$\Gamma(\delta) = \left\{\gamma(\cdot) : \|\bar{\gamma}(\cdot) - \bar{\gamma}^*(\cdot)\|_\infty \leqslant \delta\kappa_n\right\}.$$

Theorem 1 in Tanabe and Sagae (1992) implies that all the eigenvalues of $\mathcal{I}(x,y,\gamma(y))$ lies in $(0,1)$, $\det\left(\mathcal{I}(x,y,\gamma(y))\right) = \prod_{(d,t)\in\mathcal{S}} \Psi_{d,t}(\underline{\mathbf{x}}(y),\gamma(y))$, and therefore

$$\inf_{x,y\in\mathcal{X}} \mathcal{I}(x,y,\gamma(y)) > \inf_{x,y\in\mathcal{X}} \det\left(\mathcal{I}(x,y,\gamma(y))\right) = \inf_{x,y\in\mathcal{X}}\left\{\prod_{(d,t)\in\mathcal{S}} \Psi_{d,t}(\underline{\mathbf{x}}(y),\gamma(y))\right\}\cdot I_3. \tag{C.10}$$

The inequality holds in the sense that the difference between the two sides is positive definite. For any $\delta > 0$, if $\gamma \in \Gamma(\delta)$, Assumption 5.5(ii) implies that $\|\gamma(\cdot) - \gamma^*(\cdot)\|_\infty = o(1)$. This further suggests that, when n is sufficiently large, the far-right-hand side of (C.10) is bounded from below by $c_3 I_3$, for some positive constant c_3.

The analysis leading up to this point demonstrates that for for a given $c_1 > 0$, it is possible to select n large enough such that $\mathbb{P}\left(E_{1n}(c_1)\right) > 1-\epsilon/2$, $\mathbb{P}\left(E_{2n}(c_2)\right) > 1-\epsilon/2$, and (C.10) is satisfied. Now, set $\delta_0 > 2c_1c_2^{-1}c_3^{-1}$. Then, for any $\gamma(\cdot) \in \partial\Gamma(\delta_0)$, i.e., $\|\bar{\gamma}(x) - \bar{\gamma}^*(x)\| = \delta_0\kappa_n$, for all $x \in \mathcal{X}$, we have $\sup_{x\in\mathcal{X}}\left\{\mathcal{L}_n^{ps}(\bar{\gamma}(x);x) - \mathcal{L}_n^{ps}(\bar{\gamma}^*(x);x)\right\} < 0$, with a probability of at least $1-\epsilon$. This is because

$$\begin{aligned}
&\sup_{x\in\mathcal{X}}\left\{\mathcal{L}_n^{ps}(\bar{\gamma}(x);x) - \mathcal{L}_n^{ps}(\bar{\gamma}^*(x);x)\right\}\\
=&\sup_{x\in\mathcal{X}}\left\{\nabla_{\bar{\gamma}}\mathcal{L}_n^{ps}(\bar{\gamma}^*(x);x)(\bar{\gamma}-\bar{\gamma}^*(x)) - (\bar{\gamma}(x)-\bar{\gamma}^*(x))'\left(-\nabla^2_{\bar{\gamma}\bar{\gamma}'}\mathcal{L}_n^{ps}(\bar{\gamma}^\dagger;x)\right)(\bar{\gamma}(x)-\bar{\gamma}^*(x))/2\right\}\\
\leqslant&\left(\sup_{x\in\mathcal{X}}\left\|\frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{A}}_-(W_i,x,\gamma^*(x))\right\| - c_1\kappa_n\right)\cdot\delta_0\kappa_n\\
<&0,
\end{aligned}$$

where $\bar{\gamma}^\dagger$, dependent on x, lies between $\bar{\gamma}(x)$ and $\bar{\gamma}^*(x)$. Since $\mathcal{L}_n^{ps}(\bar{\gamma};x)$ is continuous, a local maximum, denoted by $\hat{\bar{\gamma}}(x)$, exists within the compact set $\{\bar{\gamma} : \|\bar{\gamma} - \bar{\gamma}^*(x)\| \leqslant \delta_0\kappa_n\}$, for any $x \in \mathcal{X}$. Furthermore, since $H(h)\underline{\mathbf{X}}(x_c)\underline{\mathbf{X}}(x_c)'H(h)$ is always positive semi-definite, (C.10) then implies that $\nabla^2_{\bar{\gamma}\bar{\gamma}'}\mathcal{L}_n^{ps}(\bar{\gamma};x)$ is negative semi-definite and thus $\mathcal{L}_n^{ps}(\cdot;x)$ is globally concave. This implies that $\hat{\bar{\gamma}}(x)$ maximizes $\mathcal{L}_n^{ps}(\cdot;x)$ over $\mathbb{R}^{3N_p}$ for any $x \in \mathcal{X}$. Hence, $\hat{\bar{\gamma}}(\cdot)$ is the global maximizer of $\mathcal{L}_n^{ps}(\bar{\gamma}(\cdot);\cdot)$ with a probability exceeding $1-\epsilon$. As ϵ is arbitrary and δ_0 is independent of x, it can be inferred that $\left\|\hat{\bar{\gamma}}(\cdot) - \bar{\gamma}^*(\cdot)\right\|_\infty = O_p(\kappa_n)$.

Step 2: We proceed to derive the uniform asymptotic linear expansion of $\hat{\bar{\gamma}}(\cdot) - \bar{\gamma}^*(\cdot)$. Expanding

$\mathcal{L}_n^{ps}(\bar{\gamma};x)$ using a third-order Taylor series and rearranging the terms lead to

$$\hat{\bar{\gamma}}(x) - \bar{\gamma}^*(x) = \frac{1}{n}\sum_{i=1}^{n}\Sigma^{ps}(x)^{-1}\widetilde{\boldsymbol{A}}_-(W_i,x,\gamma^*(x)) + R^{\gamma}(X_j),$$

where

$$\begin{aligned}
R^{\gamma}(x) =& -\left(\Sigma_n^{ps}(x)^{-1} - \Sigma^{ps}(x)^{-1}\right)\cdot\frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{A}}_-(W_i,x,\gamma^*(x)) - \Sigma_n^{ps}(x)^{-1}\mathbf{C}_n(x),\\
\mathbf{C}_n(x) =& \frac{1}{2n}\sum_{i=1}^{n}\sum_{(d,t)\in\mathcal{S}_-}\sum_{(d',t')\in\mathcal{S}_-}(\hat{\bar{\gamma}}_{d,t}(x)-\bar{\gamma}^*_{d,t}(x))'H(h)\underline{\mathbf{X}}_i(x_c)\underline{\mathbf{X}}_i(x_c)'H(h)(\hat{\bar{\gamma}}_{d',t'}(x)-\bar{\gamma}^*_{d',t'}(x))\\
&\cdot\dot{\mathcal{I}}_{\iota(d,t),\iota(d',t')}(X_{c,i},x_c,\tilde{\gamma})\otimes\underline{\mathbf{X}}_i(x_c)H(h)\widetilde{K}_{ps}(X_i;x,h,\lambda),
\end{aligned}$$

for an intermediate point $\tilde{\gamma}$ lying between $\hat{\gamma}(x)$ and $\gamma^*(x)$, $\Sigma_n^{ps}(\cdot) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{H}(W_i,\cdot,\gamma^*(\cdot))$, and

$$\begin{aligned}
&\dot{\mathcal{I}}_{\iota(d_1,t_1),\iota(d_2,t_2)} = \boldsymbol{\iota}_-\left(\left\{\dot{\mathcal{I}}^{(d_3,t_3)}_{\iota(d_1,t_1),\iota(d_2,t_2)}\right\}_{(d_3,t_3)\in\mathcal{S}_-}\right),\\
&\dot{\mathcal{I}}^{(d_3,t_3)}_{\iota(d_1,t_1),\iota(d_2,t_2)}(X_c,x_c,\gamma) = \mathbb{1}\{(d_1,t_1)=(d_2,t_2)\}\Psi_{d_1,t_1}(\underline{\mathbf{X}}(x_c),\gamma)(\mathbb{1}\{(d_1,t_1)=(d_3,t_3)\}-\Psi_{d_3,t_3}(\underline{\mathbf{X}}(x_c),\gamma))\\
&+\sum_{\ell_1,\ell_2\in\{1,2\},\ell_1\neq\ell_2}\Psi_{d_{\ell_1},t_{\ell_1}}(\underline{\mathbf{X}}(x_c),\gamma)\Psi_{d_{\ell_2},t_{\ell_2}}(\underline{\mathbf{X}}(x_c),\gamma)(\mathbb{1}\{(d_{\ell_2},t_{\ell_2})=(d_3,t_3)\}-\Psi_{d_3,t_3}(\underline{\mathbf{X}}(x_c),\gamma)).
\end{aligned}$$

In view of (C.7) and (C.8), $\left\|\Sigma^{ps}(\cdot)^{-1}\right\| = O(1)$. Taking this into account, along with Lemma C.3, we obtain

$$\begin{aligned}
&\sup_{x\in\mathcal{X}}\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{A}_-(W_i,x,\gamma^*) - \mathbb{E}[\mathbf{A}_-(W,x,\gamma^*)]\right\| = O_p\left(\sqrt{\log n/(nh^{\upsilon_c})}\right),\\
&\sup_{x\in\mathcal{X}}\|\mathbb{E}[\mathbf{A}_-(W,x,\gamma^*)]\| = O_p\left(h^{p+1}+\lambda_o+\lambda_u\right).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\sup_{x\in\mathcal{X}}\left\|\Sigma_n^{ps}(x)^{-1} - \Sigma^{ps}(x)^{-1}\right\|\\
&\quad\leqslant \sup_{x\in\mathcal{X}}\|\Sigma_n^{ps}(x)\|^{-1}\cdot\sup_{x\in\mathcal{X}}\|\Sigma_n^{ps}(x)-\Sigma^{ps}(x)\|\cdot\sup_{x\in\mathcal{X}}\|\Sigma^{ps}(x)\|^{-1}\\
&\quad= O_p(1)\cdot O_p\left(\sqrt{\log n/(nh^{\upsilon_c})}\right)\cdot O(1)\\
&\quad= O_p\left(\sqrt{\log n/(nh^{\upsilon_c})}\right).
\end{aligned}$$

where the first inequality is a result of the relationship $A^{-1} - B^{-1} = -A^{-1}(A-B)B^{-1}$ and the submultiplicativity of matrix norm. The next line is derived from (C.7) and (C.8), and arguments similar to those employed in the proof of Lemma 5 in Fan and Guerre (2016).

Observe that

$$\begin{aligned}
\sup_{x\in\mathcal{X}}\|\mathbf{C}_n(x)\| \leqslant& \frac{1}{2n}\sum_{i=1}^{n}\sum_{(d,t)\in\mathcal{S}_-}\sum_{(d',t')\in\mathcal{S}_-}\left\|\dot{\mathcal{I}}_{\iota(d,t),\iota(d',t')}(X_{c,i},x_c,\tilde{\gamma}(x))\right\|\\
&\cdot\left\|\hat{\bar{\gamma}}_{d,t}(x)-\bar{\gamma}^*_{d,t}(x)\right\|\cdot\left\|\hat{\bar{\gamma}}_{d',t'}(x)-\bar{\gamma}^*_{d',t'}(x)\right\|\cdot\|H(h)\underline{\mathbf{X}}_i(x_c)\|^3\cdot\left|\widetilde{K}_{ps}(X_i;x,h,\lambda)\right|
\end{aligned}$$

$$\lesssim \max_{(d,t),(d',t')\in\{0,1\}} \sup_{x,z\in\mathcal{X}} \left\{ \left\| \dot{\mathcal{I}}_{\iota(d,t),\iota(d',t')}(z_c, x_c, \tilde{\gamma}(x)) \right\| \cdot \left\| \hat{\bar{\gamma}}_{d,t}(x) - \bar{\gamma}^*_{d,t}(x) \right\| \cdot \left\| \hat{\bar{\gamma}}_{d',t'}(x) - \bar{\gamma}^*_{d',t'}(x) \right\| \right\} \tag{C.11}$$

$$\cdot \frac{1}{n}\sum_{i=1}^{n} \sup_{x\in\mathcal{X}} \left\{ \left| K_h^{ps}(\underline{\mathbf{X}}_i^{(1)}(x_c)) \right| \cdot \| H(h)\underline{\mathbf{X}}_i(x_c) \|^3 \right\}. \tag{C.12}$$

When $\tilde{\gamma}$ converges uniformly to γ^*, as established in the first step, $\left\| \dot{\mathcal{I}}_{\iota(d,t),\iota(d',t')}(z_c, x_c, \tilde{\gamma}(x)) \right\|$ in (C.11) is asymptotically bounded, uniformly in $x, z \in \mathcal{X}$, and for each $(d,t),(d',t') \in \mathcal{S}_-$. In addition, we can deduce from a standard change of variable argument that (C.12) is $O_p(1)$. Hence, it can be concluded that $\sup_{x\in\mathcal{X}} \|\mathbf{C}_n(x)\| = O_p\left(\kappa_n^2\right)$. As a result, we obtain $\sup_{x\in\mathcal{X}} \|R^\gamma(x)\| = O_p\left(\kappa_n^2\right)$.

Step 3: We note that $\widehat{p}(d,t,x) - p(d,t,x) = \Psi_{d,t}(e_{N_p,1}, \widehat{\gamma}(x)) - \Psi_{d,t}(e_{N_p,1}, \gamma^*(x))$ and $\nabla_{\gamma_{d,t}} \Psi_{d,t}(e_{N_p,1}, \gamma^*(x)) = e'_{3,\iota(d,t)}\mathcal{I}(x)$. Utilizing the delta method in conjunction with the uniform expansion obtained in Step 2 then establishes (C.5). This completes the proof of the lemma. ■

Lemma C.3 Suppose that the conditions of Lemma C.2 hold. Then

$$\sup_{x\in\mathcal{X}} \left\| \frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{A}}_-(W_i, x, \gamma^*(x)) - \mathbb{E}[\widetilde{\boldsymbol{A}}_-(W, x, \gamma^*(x))] \right\| = O_p\left((\log n / (nh^{\upsilon_c}))^{1/2} \right), \tag{C.13}$$

$$\sup_{x\in\mathcal{X}} \left\| \mathbb{E}[\widetilde{\boldsymbol{A}}_-(W, x, \gamma^*(x))] \right\| = O\left(h^{p+1} + \lambda_o + \lambda_u \right). \tag{C.14}$$

Proof of Lemma C.3:
The proof of (C.13) proceeds along similar lines as in Lemma 5 of Fan and Guerre (2016). For any given vector $\mathbf{k}$ with $0 \leqslant |\mathbf{k}| \leqslant p$, define

$$\widetilde{A}_{d,t}^{(\mathbf{k})}(W, x, \gamma) = (I_{d,t} - \Psi_{d,t}(\underline{\mathbf{X}}(x_c), \gamma))\, h^{-|\mathbf{k}|}(X_c - x_c)^{\mathbf{k}} \widetilde{K}\,(X; x, h, \lambda)\,,$$

$$\widetilde{A}_{d,t}^{\dagger,(\mathbf{k})}(W, x_c, \gamma) = (I_{d,t} - \Psi_{d,t}(\underline{\mathbf{X}}(x_c), \gamma))\, h^{-|\mathbf{k}|}(X_c - x_c)^{\mathbf{k}} K\left(\frac{X_c - x_c}{h} \right),$$

for $(d,t) \in \mathcal{S}_-$, and let $\kappa_n = (\log n / (nh^{\upsilon_c}))^{1/2}$. Assumption 5.5 implies that $\kappa_n \to 0$. Moreover, under Assumptions 5.1, 5.2, and 5.4, we have that, for any $\epsilon > 0$, there exists $\delta_n = n^{-\kappa_a}$ such that (i)

$$\max_{i\in\mathbb{N}_n} \left| \widetilde{A}_{d,t}^{\dagger,(\mathbf{k})}(W_i, x_c, \gamma^*(x)) - \widetilde{A}_{d,t}^{\dagger,(\mathbf{k})}(W_i, x'_c, \gamma^*(x')) \right| \leqslant h^{\upsilon_c}\kappa_n\epsilon/3, \tag{C.15}$$

$$\left| \mathbb{E}\left[\widetilde{A}_{d,t}^{\dagger,(\mathbf{k})}(W, x_c, \gamma^*(x)) \right] - \mathbb{E}\left[\widetilde{A}_{d,t}^{\dagger,(\mathbf{k})}(W, x'_c, \gamma^*(x')) \right] \right| \leqslant h^{\upsilon_c}\kappa_n\epsilon/3, \tag{C.16}$$

for $(d,t) \in \mathcal{S}_-$ and for all $x, x' \in \mathcal{X}$ such that $x_d = x'_d$ and $\|x_c - x'_c\| \leqslant \delta_n$; (ii) there is a positive integer $J_n = O\left(n^{\kappa_b}\right)$, $\kappa_b > 0$, and a set $\{x_j\}_{j=1}^{J_n} \subset \mathcal{X}$, such that for all $x \in \mathcal{X}$, there exists a j satisfying $x \in \mathcal{B}\left(x_j, \delta_n\right) \cap \mathcal{X}$, and for all $x' \in \mathcal{B}\left(x_j, \delta_n\right)$, $x'_d = x_{d,j}$. As a result, $\mathcal{X} = \bigcup_{j=1}^{J_n} \left(\mathcal{B}\left(x_j, \delta_n\right) \cap \mathcal{X}\right)$.

Now, observe that, for $(d,t) \in \mathcal{S}_-$

$$\sup_{x\in\mathcal{X}} \left| \frac{1}{n}\sum_{i=1}^{n} \widetilde{A}_{d,t}^{(\mathbf{k})}(W_i, x, \gamma^*(x)) - \mathbb{E}[\widetilde{A}_{d,t}^{(\mathbf{k})}(W, x, \gamma^*(x))] \right|$$

$$\leqslant \max_{j\in\mathbb{N}_{J_n}} \left| \frac{1}{n}\sum_{i=1}^{n} \widetilde{A}_{d,t}^{(\mathbf{k})}(W_i, x_j, \gamma^*(x_j)) - \mathbb{E}[\widetilde{A}_{d,t}^{(\mathbf{k})}(W, x_j, \gamma^*(x_j))] \right| \tag{C.17}$$

$$+ \max_{j\in\mathbb{N}_{J_n}} \sup_{x\in\mathcal{B}(x_j,\delta_n)\cap\mathcal{X}} \left| \frac{1}{n}\sum_{i=1}^{n} \left(\widetilde{A}_{d,t}^{(\mathbf{k})}(W_i, x, \gamma^*(x)) - \widetilde{A}_{d,t}^{(\mathbf{k})}(W_i, x_j, \gamma^*(x_j)) \right) \right| \tag{C.18}$$

$$+\max_{j\in\mathbb{N}_{J_n}}\sup_{x\in\mathcal{B}(x_j,\delta_n)\cap\mathcal{X}}\left|\mathbb{E}[\widetilde{A}^{(\mathbf{k})}_{d,t}(W,x,\gamma^*(x))]-\mathbb{E}[\widetilde{A}^{(\mathbf{k})}_{d,t}(W,x_j,\gamma^*(x_j))]\right|. \tag{C.19}$$

In view of (C.15), (C.18) is bounded from above by

$$\max_{i\in\mathbb{N}_n,j\in\mathbb{N}_{J_n}}\sup_{x\in\mathcal{B}(x_j,\delta_n)\cap\mathcal{X}}h^{-\upsilon_c}\left|\widetilde{A}^{\dagger,(\mathbf{k})}_{d,t}(W_i,x_c,\gamma^*(x))-\widetilde{A}^{\dagger,(\mathbf{k})}_{d,t}(W_i,x_{c,j},\gamma^*(x_j))\right|\leqslant\kappa_n\epsilon/3.$$

Meanwhile, since $x_d=x_{d,j}$, whenever $x\in\mathcal{B}(x_j,\delta_n)$, (C.16) then implies that (C.19) $\leqslant\kappa_n\epsilon/3$.

To bound (C.17), we apply Bernstein's inequality.[1] Since the support of K is bounded, we have that $\left|\widetilde{A}^{(\mathbf{k})}_{d,t}(W,x,\gamma^*(x))\right|\leqslant C\|K\|_\infty$, for a sufficiently large positive constant C. Additionally, standard calculation gives

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\widetilde{A}_{d,t}(W,x,\gamma^*(x))\right]=&\mathbb{E}[(I_{d,t}-p(d,t,(X_c,x_d)))^2H(h)\underline{\mathbf{X}}(x_c)\underline{\mathbf{X}}(x_c)'H(h)K_h(\underline{\mathbf{X}}_i(x_c))^2\mathbb{1}\{X_d=x_d\}]\\&+o\left(h^{-\upsilon_c}\right)\\=&h^{-\upsilon_c}\mathcal{I}(x)_{\iota(d,t),\iota(d,t)}\mathbf{T}_p(x_c)f_X(x)+o\left(h^{-\upsilon_c}\right).\end{aligned}$$

Hence, $\mathbb{V}\mathrm{ar}\left[\widetilde{A}^{(\mathbf{k})}_{d,t}(W,x,\gamma^*(x))\right]\leqslant Ch^{-\upsilon_c}$ under Assumption 5.4.

With these two results in hand, we have

$$\begin{aligned}&\mathbb{P}\left(\max_{j\in\mathbb{N}_{J_n}}\left|\frac{1}{n}\sum_{i=1}^n\widetilde{A}^{(\mathbf{k})}_{d,t}(W_i,x_j,\gamma^*(x_j))-\mathbb{E}[\widetilde{A}^{(\mathbf{k})}_{d,t}(W,x_j,\gamma^*(x_j))]\right|\geqslant\kappa_n\epsilon/3\right)\\\leqslant&\sum_{j=1}^{J_n}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n\widetilde{A}^{(\mathbf{k})}_{d,t}(W_i,x_j,\gamma^*(x_j))-\mathbb{E}[\widetilde{A}^{(\mathbf{k})}_{d,t}(W,x_j,\gamma^*(x_j))]\right|\geqslant\kappa_n\epsilon/3\right)\\\leqslant&2J_n\exp\left(-\frac{\epsilon^2\log n}{C+C\epsilon(\log n\cdot n^{-1}h^{-\upsilon_c})^{1/2}}\right)\leqslant2\exp\left(-\frac{(\epsilon^2-\kappa_b)\log n}{C}\right),\end{aligned}$$

where the first inequality is due to the Bonferoni inequality and the second is by Bernstein's inequality. The far right side goes to 0 when $\epsilon^2>\kappa_b$. Hence, (C.17) $\leqslant\kappa_n\epsilon/3$.

Combining (C.17)-(C.19) gives

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}}\left|\frac{1}{n}\sum_{i=1}^n\widetilde{A}^{(\mathbf{k})}_{d,t}(W_i,x,\gamma^*(x))-\mathbb{E}[\widetilde{A}^{(\mathbf{k})}_{d,t}(W,x,\gamma^*(x))]\right|\geqslant\kappa_n\epsilon\right)\to0. \tag{C.20}$$

This complete the proof for (C.13).

Next, we establish (C.14). Define $I_o(x_d,z_d)=\sum_{s=1}^{\upsilon_o}\mathbb{1}\{|x_{o,s}-z_{o,s}|=1\}\prod_{l\neq s}\mathbb{1}\{x_{o,l}=z_{o,l}\}$, and $I_u(x_d,z_d)=\sum_{s=1}^{\upsilon_u}\mathbb{1}\{x_{u,s}\neq z_{u,s}\}\prod_{l\neq s}\mathbb{1}\{x_{u,l}=z_{u,l}\}$. From a Taylor expansion of order $p+1$, we deduce that, uniformly in $x\in\mathcal{X}$,

$$\begin{aligned}&\mathbb{E}[\widetilde{A}_{d,t}(W,x,\gamma^*(x))]\\=&\frac{1}{(p+1)!}\sum_{(d',t')\in\mathcal{S}_-}\mathbb{E}\left[\mathcal{I}(X_c,x_d)_{\iota(d,t),\iota(d',t')}\mathbf{g}^{(p+1)}_{d',t'}(X_c,x_d)'\underline{\mathbf{X}}^{(p+1)}(x_c)H(h)\underline{\mathbf{X}}(x_c)K_h(\underline{\mathbf{X}}^{(1)}(x_c))\mathbb{1}\{X_d=x_d\}\right]\end{aligned}$$

---

1 Let $\{X_i\}_{i=1}^n$ be independent zero-mean random variables. Suppose $|X_i|\leqslant M$ almost surely, for $i\in\mathbb{N}_n$. Then, Bernstein's inequality states that for all $t\geqslant0$,

$$\mathbb{P}\left(\sum_{i=1}^nX_i\geqslant t\right)\leqslant\exp\left(-\frac{t^2/2}{\sum_{i=1}^n\mathbb{E}[X_i^2]+Mt/3}\right).$$

$$
\begin{aligned}
&+\sum_{z_d\in\mathcal{X}_d\backslash x_d}\sum_{j=o,u}\lambda_j I_j(x_d,z_d)\left(p(d,t,x)-p(d,t,(x_c,z_d))\right)\mathbb{E}\left[H(h)\underline{\mathbf{X}}(x_c)K_h(\underline{\mathbf{X}}^{(1)}(x_c))\mathbb{1}\{X_d=z_d\}\right]\\
&+(s.o.)\\
=&\frac{h^{p+1}}{(p+1)!}\sum_{(d',t')\in\mathcal{S}_-}\mathcal{I}(x)_{\iota(d,t),\iota(d',t')}\mathbf{M}_{p,p+1}(x_c)\mathbf{g}^{(p+1)}_{d',t'}(x)f_X(x)\\
&+\sum_{z_d\in\mathcal{X}_d\backslash x_d}\sum_{j=o,u}\lambda_j I_j(x_d,z_d)\left(p(d,t,x)-p(d,t,(x_c,z_d))\right)\mathbf{M}_{p,0}(x_c)f_X(x_c,z_d)\\
&+o(h^{p+1}+\lambda_o+\lambda_u)\\
=&O(h^{p+1}+\lambda_o+\lambda_u),
\end{aligned}
$$

where the last equality is due to Assumptions 5.2 and 5.4. ■

D Auxiliary lemmas and results

D.1 Auxiliary lemmas

Lemma D.1 Under Assumptions 1 and 2, for $d,t\in\{0,1\}$ and any measurable function $h:\mathcal{X}\to\mathbb{R}$,

$$(i)\ \ \mathbb{E}\left[I_{d,t}(Y-m_{d,t}(X))h(X)\right]=0, \tag{D.1}$$

$$(ii)\ \ \mathbb{E}\left[(w_{1,1}-w_{d,t})(W)h(X)\right]=0. \tag{D.2}$$

Proof of Lemma D.1: This lemma follows immediately from the LIE. ■

Lemma D.2 Suppose the conditions of Theorem 3.1 hold. Then, for $\widehat{w}$ defined in (3.1) with $\widehat{p}$ given by (3.9), we have

$$\mathbb{E}_n[(Y-m_{d,t}(X))\left(\widehat{w}_{d,t}-w_{d,t}\right)(W)]=o_p(n^{-1/2}),$$

for $(d,t)\in\mathcal{S}_-$.

Proof of Lemma D.2:

Recall the definition of $w^\dagger$ as given in (B.19), and decompose the difference between $\widehat{w}_{d,t}$ and $w_{d,t}$ as

$$
\begin{aligned}
&\mathbb{E}_n[(Y-m_{d,t}(X))\left(\widehat{w}_{d,t}-w_{d,t}\right)(W)]\\
&\quad=\mathbb{E}_n\left[(Y-m_{d,t}(X))\left(w^\dagger_{d,t}-w_{d,t}\right)(W)\right]+\mathbb{E}_n\left[(Y-m_{d,t}(X))\left(\widehat{w}_{d,t}-w^\dagger_{d,t}\right)(W)\right]\\
&\quad\equiv\Delta^1_w+\Delta^2_w.
\end{aligned}
$$

We bound the two terms in turn. By a third-order Taylor expansion of Δ^1_w around $p(d,t,x)$, we get

$$
\begin{aligned}
\Delta^1_w=&\mathbb{E}_n\left[\frac{I_{d,t}(Y-m_{d,t}(X))}{p(d,t,X)p(1,1)}\left(\widehat{p}(1,1,X)-p(1,1,X)\right)\right]\\
&-\mathbb{E}_n\left[\frac{I_{d,t}p(1,1,X)(Y-m_{d,t}(X))}{p^2(d,t,X)p(1,1)}\left(\widehat{p}(d,t,X)-p(d,t,X)\right)\right]+R_{n,d,t}\\
\equiv&\Delta^{11}_w+\Delta^{12}_w+R_{n,d,t},
\end{aligned}
$$

where the remainder term, $R_{n,d,t}$, collects the second-order terms. Specifically,

$$
\begin{aligned}
R_{n,d,t} =& \mathbb{E}_n\left[(Y - m_{d,t}(X))\frac{I_{d,t}}{p(1,1)}\left(-\frac{(\widehat{p}(1,1,X)-p(1,1,X))(\widehat{p}(d,t,X)-p(d,t,X))}{p^2(d,t,X)}\right)\right] \\
&+ \mathbb{E}_n\left[(Y - m_{d,t}(X))\frac{I_{d,t}}{p(1,1)}\left(\frac{p(1,1,X)(\widehat{p}(d,t,X)-p(d,t,X))^2}{\tilde{p}^3(d,t,X)}\right)\right],
\end{aligned}
$$

where the intermediate point $\tilde{p}(d,t,x)$ lying between $\widehat{p}(d,t,x)$ and $p(d,t,x)$. Under Assumptions 2(iii) and 5.1, both $\widehat{p}(d,t,x)$ and $p(d,t,x)$ are (asymptotically) bounded away from zero, uniformly over $\mathcal{X}$ and for $(d,t) \in \mathcal{S}$. Moreover, $\mathbb{E}[|Y - m_{d,t}(X)|] = O(1)$ under Assumption 5.3. We deduce that $R_{n,d,t} = O_p\left(\|\widehat{p}(1,1,\cdot) - p(1,1,\cdot)\|_\infty^2\right) + O_p\left(\|\widehat{p}(d,t,\cdot) - p(d,t,\cdot)\|_\infty^2\right)$, which is $o_p\left(n^{-1/2}\right)$ by Lemma C.2 and Assumption 5.5.

The first two terms in the decomposition of Δ_w^1 share a similar structure. We only derive the stochastic limit for Δ_w^{11}.

Using the asymptotic expansion of local polynomial estimators in Lemma C.2, we obtain

$$
\Delta_w^{11} = \frac{1}{n}\sum_{i=1}^n \left\{\frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}\left(B_{n,1,1}^{(ps)}(X_i) + S_{n,1,1}^{(ps)}(X_i) + R_{n,1,1}^{(ps)}(X_i)\right)\right\}.
$$

We proceed by establishing bounds for the convergence rate of the terms involving the bias, the first-order stochastic and the remainder, respectively.

To analyze the bias, we first apply Chebyshev's inequality and obtain

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}B_{n,1,1}^{(ps)}(X_i) =& \mathbb{E}\left[\frac{I_{d,t}(Y - m_{d,t}(X))}{p(d,t,X)p(1,1)}B_{n,1,1}^{(ps)}(X)\right] \\
&+ O_p\left(n^{-1/2}(h^{p+1} + \lambda_o + \lambda_u)\right),
\end{aligned}
$$

where, due to Lemma D.1(i), the mean on the right-hand side is zero. The rate of the remainder term follows from a standard variance calculation:

$$
\begin{aligned}
&\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n \frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}B_{n,1,1}^{(ps)}(X_i)\right] \\
=& \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n \frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}B_{n,1,1}^{(ps)}(X_i)\right)^2\right] \\
=& n^{-1}\,\mathbb{E}\left[\left(\frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}B_{n,1,1}^{(ps)}(X_i)\right)^2\right] \\
=& n^{-1}\,\mathbb{E}\left[\left(\frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}\right)^2\left(-\sum_{k=1}^3\left(\mathcal{B}_k^{ps}(X_i,h,\lambda) + R_{\mathcal{B},k}^{ps}(X_i,h,\lambda)\right)\right)^2\right] \\
=& n^{-1}\left\{\mathbb{E}\left[\left(\frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}\right)^2\left(\sum_{k=1}^3 \mathcal{B}_k^{ps}(X_i,h,\lambda)\right)^2\right] + (s.o.)\right\} \\
=& O\left(n^{-1}(h^{p+1} + \lambda_o + \lambda_u)^2\right) + o\left(n^{-1}(h^{p+1} + \lambda_o + \lambda_u)^2\right),
\end{aligned}
$$

where the first equality follows by Lemma D.1(i); The second equality is due to *i.i.d.* sampling; The third follows from the bias expansion in (D.15), with $\mathcal{B}^{ps}$ (as defined in (D.10)) representing the leading bias terms and $R_{\mathcal{B}}^{ps}$ collecting the higher-order terms, and uses the fact that $B_{n,1,1}^{(ps)} = -\sum_{(d,t)\in\mathcal{S}_-} B_{n,d,t}^{(ps)}$;

The last line is obtained by directly evaluating the leading term using the explicit expression of $\mathcal{B}^{ps}$.

Consequently,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}B_{n,1,1}^{(ps)}(X_i) = O_p\left(n^{-1/2}(h^{p+1}+\lambda_o+\lambda_u)\right). \tag{D.3}$$

Under the bandwidth restrictions in Assumption 5.5, this term is $o_p\left(n^{-1/2}\right)$.

We now introduce the term $\psi_{w1,d,t}(W_i,W_j)$, which represents the summand of the first-order stochastic term as follows

$$\psi_{w1,d,t}(W_i,W_j) = \frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}\left(G_{1,1}^{(ps)}(W_j,X_i) - \mathbb{E}[G_{1,1}^{(ps)}(W_j,X_i)|X_i]\right). \tag{D.4}$$

By its definition, we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}S_{n,1,1}^{(ps)}(X_i) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\psi_{w1,d,t}(W_i,W_j). \tag{D.5}$$

By construction, $\mathbb{E}\left[\psi_{w,d,t}(W_i,W_j)|W_i\right] = 0$. Moreover, it follows by Lemma D.1(i) that $\mathbb{E}\left[\psi_{w,d,t}(W_i,W_j)|W_j\right] = 0$. Hence, (D.5) represents a second-order U-statistic with first-order degenerate kernel. To obtain its convergence rate, we apply Lemma C.1, which requires evaluating the variance rate:

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}\left[\psi_{w1,d,t}(W_i,W_j)\right]\\
=&\,\mathbb{E}\left[\mathbb{E}\left[(\psi_{w1,d,t}(W_i,W_j))^2\,|W_i\right]\right]\\
=&\,\mathbb{E}\left[\left(\frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}\right)^2\cdot\left\{(-\mathbf{1}_3)'\left(\mathcal{V}^{ps}(X_i,h,\lambda)+R_{\mathcal{V}}^{ps}(X_i,h,\lambda)\right)(-\mathbf{1}_3)\right\}\right]\\
=&\,\mathbb{E}\left[\left(\frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}\right)^2\cdot\left(\sum_{k=1}^{3}\sum_{l=1}^{3}\mathcal{V}_{k,l}^{ps}(X_i,h,\lambda)\right)\right]+(s.o.)\\
=&\,O\left(h^{-\upsilon_c}\right)+o\left(h^{-\upsilon_c}\right),
\end{aligned}$$

where the first equality follows by the LIE and Lemma D.1(i); The second equality follows from the variance expansion in (D.16) with $\mathcal{V}^{ps}$ (as defined in (D.11)) representing the leading variance terms and $R_{\mathcal{V}}^{ps}$ collecting the higher-order terms, and $S_{n,1,1}^{(ps)} = -\sum_{(d,t)\in\mathcal{S}_-}S_{n,d,t}^{(ps)}$; The last equality is obtained by directly evaluating the leading term using the explicit form of $\mathcal{V}^{ps}$.

Applying Lemma C.1 with $s^* = 1$ then yields,

$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\psi_{w1,d,t}(W_i,W_j) = O_p\left(n^{-1}h^{-\upsilon_c/2}\right), \tag{D.6}$$

Under our bandwidth assumptions, this term is $o_p\left(n^{-1/2}\right)$.

Under Assumption 2(iii), $p(d,t,x)$ is uniformly bounded away from zero for all $x\in\mathcal{X}$ and for all $(d,t)\in\mathcal{S}_-$. Also, under Assumption 5.3, we have $\mathbb{E}[|Y - m_{d,t}(X)|] = O(1)$. Consequently, we can deduce that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}R_{n,1,1}^{(ps)}(X_i) = O_p\left(\sup_{i\in\mathbb{N}_n}\left|R_{n,1,1}^{(ps)}(X_i)\right|\right)$$

$$= O_p\left(\left(h^{p+1} + \lambda_o + \lambda_u + \sqrt{\log n/\left(nh^{\upsilon_c}\right)}\right)^2\right) \tag{D.7}$$

which is $o_p\left(n^{-1/2}\right)$ under Assumption 5.5.

Combining (D.3), (D.6), and (D.7), we can conclude that $\Delta_w^{11} = o_p\left(n^{-1/2}\right)$.

By the same reasoning, we can demonstrate that Δ_w^{12} is dominated by the first-order stochastic term. Define

$$\psi_{w2,d,t}(W_i, W_j) = -\frac{I_{d,t}p(1,1,X_i)(Y_i - m_{d,t}(X_i))}{p^2(d,t,X_i)p(1,1)}\left(G_{d,t}^{(ps)}(W_j, X_i) - \mathbb{E}[G_{d,t}^{(ps)}(W_j, X_i)|X_i]\right), \tag{D.8}$$

As a result, the leading term is given by $n^{-1}(n-1)^{-1}\sum_{i=1}^n\sum_{j\neq i}^n \psi_{w2,d,t}(W_i, W_j)$, which has an order of $O_p\left(n^{-1}h^{-\upsilon_c/2}\right) = o_p\left(n^{-1/2}\right)$. The detailed proof is omitted for brevity.

Now, let's consider Δ_w^2. Define $\widehat{p}(1,1) = \mathbb{E}_n\left[\dfrac{I_{d,t}\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)}\right]$.

$$\begin{aligned}
\Delta_w^2 &= \mathbb{E}_n\left[\frac{I_{d,t}\widehat{p}(1,1,X)(Y - m_{d,t}(X))}{\widehat{p}(d,t,X)}\left(\frac{1}{\widehat{p}(1,1)} - \frac{1}{p(1,1)}\right)\right] \\
&= \mathbb{E}_n\left[\frac{I_{d,t}\widehat{p}(1,1,X)(Y - m_{d,t}(X))}{\widehat{p}(d,t,X)}\right]\cdot O_p\left(|\widehat{p}(1,1) - p(1,1)|\right),
\end{aligned}$$

where the second line follows by a first-order Taylor expansion of the right-hand side of the first equality in $\widehat{p}(1,1)$ around $p(1,1)$. In the proof of Lemma 3.1, it is established that when $\widehat{p}$ is uniformly convergent to p, $|\widehat{p}(1,1) - p(1,1)| = o_p(1)$. The uniform convergence follows by Lemma C.2 under the rate conditions specified in Assumption 5.5.

To study the first term, we can use an approach similar to the proof of Δ_w^1, and show that

$$\mathbb{E}_n\left[\frac{I_{d,t}\widehat{p}(1,1,X)(Y - m_{d,t}(X))}{\widehat{p}(d,t,X)}\right] = \mathbb{E}_n\left[\frac{I_{d,t}p(1,1,X)(Y - m_{d,t}(X))}{p(d,t,X)}\right] + o_p\left(n^{-1/2}\right).$$

Due to Lemma D.1(i), the first term on the right-hand side of the preceding equation has a mean of zero. Consequently, this term is of order $O_p\left(n^{-1/2}\right)$. This completes our proof. ■

Lemma D.3 Suppose the conditions of Theorem 3.1 hold, then with $\widehat{m}$ given by (3.11),

$$\mathbb{E}_n[(w_{1,1} - w_{d,t})(W)\cdot(\widehat{m}_{d,t} - m_{d,t})] = o_p(n^{-1/2}),$$

for $(d,t) \in \mathcal{S}_-$.

Proof of Lemma D.3:

The proof closely resembles the first part of Lemma D.2. We first decompose the estimation error for the OR functions as

$$\mathbb{E}_n[(w_{1,1} - w_{d,t})(W)\left(\widehat{m}_{d,t} - m_{d,t}\right)(X)] = \frac{1}{n}\sum_{i=1}^n\left\{(w_{1,1} - w_{d,t})(W_i)\left(B_{n,d,t}^{(or)}(X_i) + S_{n,d,t}^{(or)}(X_i) + R_{n,d,t}^{(or)}(X_i)\right)\right\}.$$

We address the three terms individually. For the bias term

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n\left\{(w_{1,1} - w_{d,t})(W_i)B_{n,d,t}^{(or)}(X_i)\right\} &= \mathbb{E}\left[(w_{1,1} - w_{d,t})(W)B_{n,d,t}^{(or)}(X)\right] + O_p\left(n^{-1/2}(b_{d,t}^{q+1} + \vartheta_{o,d,t} + \vartheta_{u,d,t})\right) \\
&= O_p\left(n^{-1/2}(b_{d,t}^{q+1} + \vartheta_{o,d,t} + \vartheta_{u,d,t})\right) = o_p\left(n^{-1/2}\right),
\end{aligned}$$

where the first equality follows from Chebyshev's inequality, and the second is derived from Lemma D.1(ii) and the following variance calculation:

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{i=1}^{n}\left\{(w_{1,1}-w_{d,t})(W_i)B^{(or)}_{n,d,t}(X_i)\right\}\right]\\
=&n^{-1}\,\mathbb{E}\left[\left((w_{1,1}-w_{d,t})(W_i)B^{(or)}_{n,d,t}(X_i)\right)^2\right]\\
=&n^{-1}\,\mathbb{E}\left[((w_{1,1}-w_{d,t})(W_i))^2\cdot\left(\mathcal{B}^{or}_{d,t}(X_i,b_{d,t},\vartheta_{d,t})+R^{or}_{\mathcal{B},d,t}(X_i,b_{d,t},\vartheta_{d,t})\right)^2\right]\\
=&n^{-1}\left\{\mathbb{E}\left[((w_{1,1}-w_{d,t})(W_i))^2\,\mathcal{B}^{or}_{d,t}(X_i,b_{d,t},\vartheta_{d,t})^2\right]+(s.o.)\right\}\\
=&\,O\left(n^{-1}(b^{q+1}_{d,t}+\vartheta_{o,d,t}+\vartheta_{u,d,t})^2\right)+o\left(n^{-1}(b^{q+1}_{d,t}+\vartheta_{o,d,t}+\vartheta_{u,d,t})^2\right),
\end{aligned}
$$

where the first equality follows by Lemma D.1(ii) and $i.i.d.$ sampling; The second equality follows from the bias expansion in (D.17) with $\mathcal{B}^{or}_{d,t}$ (as defined in (D.12)) representing the leading bias terms and $R^{or}_{\mathcal{B},d,t}$ collecting the higher-order terms; The last line is obtained by directly evaluating the leading term using the explicit expression of $\mathcal{B}^{or}_{d,t}$.

Next, for the first-order stochastic term, we define

$$
\psi_{m,d,t}(W_i,W_j)=(w_{1,1}-w_{d,t})(W_i)\left(G^{(or)}_{d,t}(W_j,X_i)-\mathbb{E}[G^{(or)}_{d,t}(W_j,X_i)|X_i]\right), \tag{D.9}
$$

By definition,

$$
\frac{1}{n}\sum_{i=1}^{n}\left\{(w_{1,1}-w_{d,t})(W_i)S^{(or)}_{n,d,t}(X_i)\right\}=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\psi_{m,d,t}(W_i,W_j).
$$

By construction, $\mathbb{E}\left[\psi_{m,d,t}(W_i,W_j)|W_i\right]=0$. Moreover, Lemma D.1(ii) implies $\mathbb{E}\left[\psi_{m,d,t}(W_i,W_j)|W_j\right]=0$, so the right-hand side of the above equation is a second-order U-statistic with a degenerate first-order kernel. To obtain its convergence rate, we apply Lemma C.1, which requires evaluating the variance rate:

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}\left[\psi_{m,d,t}(W_i,W_j)\right]\\
=&\,\mathbb{E}\left[\mathbb{E}\left[(\psi_{m,d,t}(W_i,W_j))^2\,|W_i\right]\right]\\
=&\,\mathbb{E}\left[((w_{1,1}-w_{d,t})(W_i))^2\cdot\left\{\mathcal{V}^{or}_{d,t}(X_i,b_{d,t},\vartheta_{d,t})+R^{or}_{\mathcal{V},d,t}(X_i,b_{d,t},\vartheta_{d,t})\right\}\right]\\
=&\,\mathbb{E}\left[((w_{1,1}-w_{d,t})(W_i))^2\cdot\mathcal{V}^{or}_{d,t}(X_i,b_{d,t},\vartheta_{d,t})\right]+(s.o.)\\
=&\,O\left(b^{-\upsilon_c}_{d,t}\right)+o\left(b^{-\upsilon_c}_{d,t}\right),
\end{aligned}
$$

where the first equality follows by the LIE and Lemma D.1(ii); The second equality follows from the variance expansion in (D.18) with $\mathcal{V}^{or}_{d,t}$ (as defined in (D.13)) representing the leading variance terms and $R^{or}_{\mathcal{V},d,t}$ collecting the higher-order terms; The last equality is obtained by directly evaluating the leading term using the explicit form of $\mathcal{V}^{or}_{d,t}$.

Applying Lemma C.1 with $s^*=1$ then leads to,

$$
\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\psi_{m,d,t}(W_i,W_j)=O_p\left(n^{-1}b^{-\upsilon_c/2}_{d,t}\right),
$$

which is $o_p\left(n^{-1/2}\right)$ due to our bandwidth restrictions.

Finally, as $p(d,t,x)$ is uniformly bounded away from zero under Assumption 2(iii), we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left\{(w_{1,1}-w_{d,t})(W_i)R_{n,d,t}^{(or)}(X_i)\right\} &= O_p\left(\sup_{i\in\mathbb{N}_n}\left|R_{n,d,t}^{(or)}(X_i)\right|\right)\\ &= O_p\left(\left((b_{d,t}^{q+1}+\vartheta_{o,d,t}+\vartheta_{u,d,t})+\sqrt{\log n/\left(nb_{d,t}^{\upsilon_c}\right)}\right)^2\right),\end{aligned}$$

which is $o_p\left(n^{-1/2}\right)$ under Assumption 5.5. This completes our proof. ■

D.2 Mean integrated squared error

Cross-validated bandwidth asymptotically minimizes the mean integrated squared errors (MISE). Given user-specified weight functions $\omega^{ps}(\cdot),\omega_{d,t}^{or}(\cdot):\mathcal{X}\to\mathbb{R}_+$, MISE is defined as

$$\begin{aligned}\chi(h,\lambda,\{b_{d,t},\vartheta_{d,t}\}_{(d,t)\in\mathcal{S}_-}) &= \int_{\mathcal{X}}\mathbb{E}\left[\|\hat{\mathbf{p}}_-(x)-\mathbf{p}_-(x)\|^2\right]\omega^{ps}(x)dx\\ &\quad+\sum_{(d,t)\in\mathcal{S}_-}\int_{\mathcal{X}}\mathbb{E}\left[|\hat{m}_{d,t}(x)-m_{d,t}(x)|^2\right]\omega_{d,t}^{or}(x)dx.\end{aligned}$$

Let $(h^*,\lambda^*,\{b_{d,t}^*,\vartheta_{d,t}^*\}_{(d,t)\in\mathcal{S}_-})$ denote the minimizer of the MISE. In the subsequent analysis, we investigate the properties of these optimal smoothing parameters.

For $(d,t)\in\mathcal{S}_-$, we represent the $n_k\times 1$ vector of k-th derivatives $p(d,t,x)$ as $\mathbf{p}_{d,t}^{(k)}(x)$, ordered lexicographically according to the method discussed earlier in the paper. Define $\mathbf{g}_-^{(k)}(x)=\left(\mathbf{g}_{1,0}^{(k)}(x),\mathbf{g}_{0,1}^{(k)}(x),\mathbf{g}_{0,0}^{(k)}(x)\right)$. For $j=p,q$, let $\varrho_{j,1}^b(x_c)=e_{N_j,1}'\mathbf{Q}_j(x_c)^{-1}\mathbf{M}_{j,j+1}(x_c)$, $\varrho_{j,2}^b(x_c)=e_{N_j,1}'\mathbf{Q}_j(x_c)^{-1}\mathbf{M}_{j,0}(x_c)$, and $\varrho_j^v(x_c)=e_{N_j,1}'\mathbf{Q}_j(x_c)^{-1}\mathbf{T}_j(x_c)\mathbf{Q}_j(x_c)^{-1}e_{N_j,1}$. Additionally, we define terms associated with the asymptotic bias and variance of $\hat{\mathbf{p}}_-(x)-\mathbf{p}_-(x)$ as follows

$$\begin{aligned}\mathcal{B}^{ps}(x,h,\lambda) =&\frac{h^{p+1}}{(p+1)!}\varrho_{p,1}^b(x_c)\mathbf{g}_-^{(p+1)}(x)\mathcal{I}(x)\\ &+\sum_{z_d\in\mathcal{X}_d\backslash x_d}\sum_{j=o,u}\frac{f_X(x_c,z_d)}{f_X(x)}\lambda_j I_j(x_d,z_d)\varrho_{p,2}^b(x_c)\left(\mathbf{p}_-(x)-\mathbf{p}_-(x_c,z_d)\right),\end{aligned}\tag{D.10}$$

$$\mathcal{V}^{ps}(x,h,\lambda) =\frac{\mathcal{I}(x)\varrho_p^v(x_c)}{h^{\upsilon_c}f_X(x)}.\tag{D.11}$$

For the OR functions, we define

$$\begin{aligned}\mathcal{B}_{d,t}^{or}(x,b,\vartheta) =&\frac{b^{q+1}}{(q+1)!}\left(\varrho_{q,1}^b(x_c)\mathbf{m}_{d,t}^{(q+1)}(x)\right)\\ &+\sum_{z_d\in\mathcal{X}_d\backslash x_d}\sum_{j=o,u}\frac{f_X(x_c,z_d)}{f_X(x)}\vartheta_j I_j(x_d,z_d)\varrho_{q,2}^b(x_c)\left(m_{d,t}(x)-m_{d,t}(x_c,z_d)\right),\end{aligned}\tag{D.12}$$

$$\mathcal{V}_{d,t}^{or}(x,b,\vartheta) =\frac{\sigma_{d,t}^2(x)\varrho_q^v(x_c)}{b^{\upsilon_c}f_X(x)},\tag{D.13}$$

where $\sigma_{d,t}^2(x)=\mathbb{E}[I_{d,t}(Y-m_{d,t}(X))^2|X=x]$.

Finally, we define a first-order approximation of the MISE as

$$\chi^*(h,\lambda,\{b_{d,t},\vartheta_{d,t}\}_{(d,t)\in\mathcal{S}_-}) = \int_{\mathcal{X}} \left\{\|\mathcal{B}^{ps}(x,h,\lambda)\|^2 + \mathrm{tr}(\mathcal{V}^{ps}(x,h,\lambda))\right\}\omega^{ps}(x)dx + \sum_{(d,t)\in\mathcal{S}_-}\int_{\mathcal{X}}\left\{\mathcal{B}^{or}_{d,t}(x,b_{d,t},\vartheta_{d,t})^2 + \mathcal{V}^{or}_{d,t}(x,b_{d,t},\vartheta_{d,t})\right\}\omega^{or}_{d,t}(x)dx. \quad \text{(D.14)}$$

We denote the constrained minimizer of χ^* as $(h^o,\lambda^o,b^o_{d,t},\vartheta^o_{d,t}{}_{(d,t)\in\mathcal{S}_-})$, where each argument of the function is constrained to be non-negative.

Assumption D.1 1. The constrained minimizer of χ^*, denoted as $(h^o,\lambda^o,\{b^o_{d,t},\vartheta^o_{d,t}\}_{(d,t)\in\mathcal{S}_-})$, is uniquely determined and finite.

2. The constrained minimizer resides in $[0,\delta_n]^{12}$, where $n^\epsilon\delta_n\to\infty$ for any $\epsilon>0$.

Theorem D.1 (a) Assuming that Assumptions 1, 5.1-5.4, 5.6, and D.1 hold and both p and q are odd, the optimal bandwidths $(h^*,\lambda^*,\{b^*_{d,t},\vartheta^*_{d,t}\}_{(d,t)\in\mathcal{S}_-})$ satisfy

$$\begin{aligned} h^* &\sim h^o n^{-1/(2p+\upsilon_c+2)}, & \lambda^* &\sim \lambda^o n^{-(p+1)/(2p+\upsilon_c+2)}, \\ b^*_{d,t} &\sim b^o_{d,t} n^{-1/(2q+\upsilon_c+2)}, & \vartheta^*_{d,t} &\sim \vartheta^o_{d,t} n^{-(q+1)/(2q+\upsilon_c+2)}, \quad \text{for } (d,t)\in\mathcal{S}_-. \end{aligned}$$

(b) Furthermore, assume that $0<\upsilon_c<2+2\min\{p,q\}$ and $\zeta>2+\upsilon_c/(q+1)$. Under these optimal bandwidth choices, we have:

1. $(h^*)^{p+1} = o\left(n^{-1/4}\right)$, $\lambda^* = o\left(n^{-1/4}\right)$, $\lambda^*/(h^*)^p = o(1)$, $\log n/\left(n(h^*)^{\upsilon_c+2p}\right) = o(1)$ and $\log n/\left(n(h^*)^{\upsilon_c}\right) = o(n^{-1/2})$.
2. $(b^*_{d,t})^{q+1} = o\left(n^{-1/4}\right)$, $\vartheta^*_{d,t} = o\left(n^{-1/4}\right)$, $\log n/\left(n^{1-2/\zeta}(b^*_{d,t})^{\upsilon_c}\right) = o(1)$, and $\log n/\left(n(b^*_{d,t})^{\upsilon_c}\right) = o\left(n^{-1/2}\right)$, for all $(d,t)\in\mathcal{S}_-$.

Thus, the rate conditions in Assumption 5.5 are satisfied by the optimal bandwidths.

Proof of Theorem D.1:
<u>Proof of Part (a)</u>: From the uniform linear expansions of Lemma C.2, we know that

$$\mathbb{E}\left[\|\hat{\mathbf{p}}_-(x)-\mathbf{p}_-(x)\|^2\right] = \|\mathbb{E}[\mathcal{I}(x)\mathbf{A}_-(W,x)]\|^2 + n^{-1}\,\mathrm{tr}\left(\mathbb{V}\mathrm{ar}\left[\mathcal{I}(x)\mathbf{A}_-(W,x)\right]\right) + (s.o.),$$

where

$$\begin{aligned} &\mathbb{E}[\mathcal{I}(x)\mathbf{A}_-(W,x)] = \mathcal{I}(x)(I_3\otimes e_{N_p,1})'\Sigma^{ps}(x)^{-1}\,\mathbb{E}[\widetilde{\mathbf{A}}_-(W,x,\gamma^*(x))] \\ &= \frac{h^{p+1}}{(p+1)!}\mathcal{I}(x)(I_3\otimes e_{N_p,1})'(\mathcal{I}(x)\otimes\mathbf{Q}_p(x_c)f_X(x))^{-1}\Big\{(\mathcal{I}(x)\otimes\mathbf{M}_{p,p+1}(x_c))\,\mathrm{vec}\left(\mathbf{g}_-^{(p+1)}(x)\right)f_X(x) \\ &\quad + \sum_{z_d\in\mathcal{X}_d\backslash x_d}\sum_{j=o,u}\lambda_j I_j(x_d,z_d)\left(\mathbf{p}_-(x)-\mathbf{p}_-(x_c,z_d)\right)\otimes\mathbf{M}_{p,0}(x_c)f_X(x_c,z_d)\Big\} + o\left(h^{p+1}+\lambda_o+\lambda_u\right) \\ &= \frac{h^{p+1}}{(p+1)!}e'_{N_p,1}\mathbf{Q}_p(x_c)^{-1}\mathbf{M}_{p,p+1}(x_c)\mathbf{g}_-^{(p+1)}(x)\mathcal{I}(x) \\ &\quad + \sum_{z_d\in\mathcal{X}_d\backslash x_d}\sum_{j=o,u}\frac{f_X(x_c,z_d)}{f_X(x)}\lambda_j I_j(x_d,z_d)e'_{N_p,1}\mathbf{Q}_p(x_c)^{-1}\mathbf{M}_{p,0}(x_c)\left(\mathbf{p}_-(x)-\mathbf{p}_-(x_c,z_d)\right) \\ &\quad + o\left(h^{p+1}+\lambda_o+\lambda_u\right) \end{aligned}$$

$$=\mathcal{B}^{ps}(x,h,\lambda)+o\left(h^{p+1}+\lambda_o+\lambda_u\right), \tag{D.15}$$

and

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\mathcal{I}(x)\mathbf{A}_-(W,x)\right]=&h^{-\upsilon_c}\mathcal{I}(x)(I_3\otimes e_{N_p,1})'\Sigma^{ps}(x)^{-1}\left(\mathcal{I}(x)\otimes\mathbf{T}_p(x_c)f_X(x)\right)\Sigma^{ps}(x)^{-1}(I_3\otimes e_{N_p,1})\mathcal{I}(x)\\ =&h^{-\upsilon_c}\mathcal{I}(x)(I_3\otimes e_{N_p,1})'(\mathcal{I}(x)\otimes\mathbf{Q}_p(x_c)f_X(x))^{-1}\left(\mathcal{I}(x)\otimes\mathbf{T}_p(x_c)f_X(x)\right)\\ &\cdot\left(\mathcal{I}(x)\otimes\mathbf{Q}_p(x_c)f_X(x)\right)^{-1}(I_3\otimes e_{N_p,1})\mathcal{I}(x)+o\left(h^{-\upsilon_c}\right)\\ =&h^{-\upsilon_c}f_X(x)^{-1}\mathcal{I}(x)e_{N_p,1}'\mathbf{Q}_p(x_c)^{-1}\mathbf{T}_p(x_c)\mathbf{Q}_p(x_c)^{-1}e_{N_p,1}+o\left(h^{-\upsilon_c}\right)\\ =&\mathcal{V}^{ps}(x,h,\lambda)+o\left(h^{-\upsilon_c}\right).\end{aligned} \tag{D.16}$$

Analogously, for $(d,t)\in\mathcal{S}_-$

$$\mathbb{E}\left[|\widehat{m}_{d,t}(x)-m_{d,t}(x)|^2\right]=\left|\mathbb{E}[G_{d,t}^{(or)}(W,x)]\right|^2+n^{-1}\mathbb{V}\mathrm{ar}\left[G_{d,t}^{(or)}(W,x)\right]+(s.o.),$$

where

$$\begin{aligned}&\mathbb{E}[G_{d,t}^{(or)}(W,x)]=e_{N_q,1}'\Sigma_{d,t}^{or}(x)^{-1}\,\mathbb{E}[H(b_{d,t})\underline{\mathbf{X}}(X_j)I_{d,t}\xi_{d,t}^{or}(x)\widetilde{K}_{or}(X;x,b_{d,t},\vartheta_{d,t})]\\ =&\frac{b_{d,t}^{q+1}}{(q+1)!}e_{N_q,1}'(\mathbf{Q}_q(x_c)f_X(x))^{-1}\left\{\mathbf{M}_{q,q+1}(x_c)\mathbf{m}_{d,t}^{(q+1)}(x)f_X(x)\right.\\ &\left.+\sum_{z_d\in\mathcal{X}_d\backslash x_d}\sum_{j=o,u}\vartheta_{d,t,j}I_j(x_d,z_d)\left(m_{d,t}(x)-m_{d,t}(x_c,z_d)\right)\mathbf{M}_{q,0}(x_c)f_X(x_c,z_d)\right\}\\ &+o\left(b_{d,t}^{q+1}+\vartheta_{d,t,o}+\vartheta_{d,t,u}\right)\\ =&\frac{b_{d,t}^{q+1}}{(q+1)!}\left(e_{N_q,1}'\mathbf{Q}_q(x_c)^{-1}\mathbf{M}_{q,q+1}(x_c)\mathbf{m}_{d,t}^{(q+1)}(x)\right)\\ &+\sum_{z_d\in\mathcal{X}_d\backslash x_d}\sum_{j=o,u}\frac{f_X(x_c,z_d)}{f_X(x)}\vartheta_jI_j(x_d,z_d)e_{N_q,1}'\mathbf{Q}_q(x_c)^{-1}\mathbf{M}_{q,0}(x_c)\left(m_{d,t}(x)-m_{d,t}(x_c,z_d)\right)\\ &+o\left(b_{d,t}^{q+1}+\vartheta_{d,t,o}+\vartheta_{d,t,u}\right),\\ =&\mathcal{B}_{d,t}^{or}(x,b_{d,t},\vartheta_{d,t})+o\left(b_{d,t}^{q+1}+\vartheta_{d,t,o}+\vartheta_{d,t,u}\right),\end{aligned} \tag{D.17}$$

and

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[G_{d,t}^{(or)}(W,x)\right]=&b_{d,t}^{-\upsilon_c}e_{N_q,1}'\Sigma_{d,t}^{or}(x)^{-1}\,\mathbb{E}[H(b_{d,t})\underline{\mathbf{X}}(X_j)I_{d,t}(Y-m_{d,t}(X))^2\\ &+H(b_{d,t})\underline{\mathbf{X}}(X_j)'\widetilde{K}_{or}(X;x,b_{d,t},\vartheta_{d,t})^2]\Sigma_{d,t}^{or}(x)^{-1}e_{N_q,1}+o\left(b^{-\upsilon_c}\right)\\ =&b_{d,t}^{-\upsilon_c}e_{N_q,1}'(\mathbf{Q}_q(x_c)f_X(x))^{-1}\left(\sigma_{d,t}^2(x)\mathbf{T}_q(x_c)f_X(x)\right)(\mathbf{Q}_q(x_c)f_X(x))^{-1}+o\left(b^{-\upsilon_c}\right)\\ =&b_{d,t}^{-\upsilon_c}f_X(x)^{-1}\sigma_{d,t}^2(x)e_{N_q,1}'\mathbf{Q}_q(x_c)^{-1}\mathbf{T}_q(x_c)\mathbf{Q}_q(x_c)^{-1}e_{N_q,1}+o\left(b^{-\upsilon_c}\right)\\ =&\mathcal{V}_{d,t}^{or}(x,b_{d,t},\vartheta_{d,t})+o\left(b^{-\upsilon_c}\right).\end{aligned} \tag{D.18}$$

Now, we define

$$(h^\dagger,\lambda^\dagger,\{b_{d,t}^\dagger,\vartheta_{d,t}^\dagger\}_{(d,t)\in\mathcal{S}_-})=\left(n^{\frac{1}{2p+\upsilon_c+2}}h,n^{\frac{p+1}{2p+\upsilon_c+2}}\lambda,\left\{n^{\frac{1}{2q+\upsilon_c+2}}b_{d,t},n^{\frac{q+1}{2q+\upsilon_c+2}}\vartheta_{d,t}\right\}_{(d,t)\in\mathcal{S}_-}\right).$$

It follows from (D.15)-(D.18) and standard analysis that

$$
\begin{aligned}
&\chi(h, \lambda, \{b_{d,t}, \vartheta_{d,t}\}_{(d,t)\in\mathcal{S}_-}) \\
&\quad = n^{-\frac{2(p+1)}{2p+\upsilon_c+2}} \int_{\mathcal{X}} \left\{ \left\| \mathcal{B}^{ps}(x, h^\dagger, \lambda^\dagger) \right\|^2 + \operatorname{tr}(\mathcal{V}^{ps}(x, h^\dagger, \lambda^\dagger)) \right\} \omega^{ps}(x) dx \\
&\quad + o\left(n^{-\frac{2(p+1)}{2p+\upsilon_c+2}} \left(h^\dagger + \lambda_o^\dagger + \lambda_u^\dagger \right) \right) \\
&\quad + n^{-\frac{2(q+1)}{2q+\upsilon_c+2}} \sum_{(d,t)\in\mathcal{S}_-} \int_{\mathcal{X}} \left\{ \mathcal{B}^{or}_{d,t}(x, b^\dagger_{d,t}, \vartheta^\dagger_{d,t})^2 + \mathcal{V}^{or}_{d,t}(x, b^\dagger_{d,t}, \vartheta^\dagger_{d,t}) \right\} \omega^{or}_{d,t}(x) dx \\
&\quad + \sum_{(d,t)\in\mathcal{S}_-} o\left(n^{-\frac{2(q+1)}{2q+\upsilon_c+2}} \left(b^\dagger_{d,t} + \vartheta^\dagger_{d,t,o} + \vartheta^\dagger_{d,t,u} \right) \right),
\end{aligned}
$$

uniformly over $[0, \delta_n]^{12}$. Since χ^* is separable in (h, λ) and $(\{b_{d,t}, \vartheta_{d,t}\}_{(d,t)\in\mathcal{S}_-})$, and its constrained minimizer is well-defined, unique, and finite under Assumption D.1, the proof is completed by minimizing χ with respect to $(h^\dagger, \lambda^\dagger, \{b^\dagger_{d,t}, \vartheta^\dagger_{d,t}\}_{(d,t)\in\mathcal{S}_-})$ and recalling the definition of $(h^o, \lambda^o, \{b^o_{d,t}, \vartheta^o_{d,t}\}_{(d,t)\in\mathcal{S}_-})$.

Proof of Part (b): Given the optimal bandwidths and the condition that $\upsilon_c < 2 + 2\min\{p, q\}$ and $\zeta > 2 + \upsilon_c/(q+1)$, we have

$$
\begin{aligned}
&(h^*)^{p+1} = O\left(n^{-(p+1)/(2p+\upsilon_c+2)}\right) = o\left(n^{-1/4}\right), \\
&\lambda^* = O\left(n^{-(p+1)/(2p+\upsilon_c+2)}\right) = o\left(n^{-1/4}\right), \\
&\lambda^*/(h^*)^p = O\left(n^{-1/(2p+\upsilon_c+2)}\right) = o(1), \\
&\log n / \left(n(h^*)^{\upsilon_c+2p}\right) = O\left(\log n \cdot \left(n^{-2/(2p+\upsilon_c+2)}\right)\right) = o(1), \\
&\log n / \left(n(h^*)^{\upsilon_c}\right) = O\left(\log n \cdot \left(n^{-2(p+1)/(2p+\upsilon_c+2)}\right)\right) = o\left(n^{-1/2}\right), \\
&(b^*_{d,t})^{q+1} = O\left(n^{-(q+1)/(2q+\upsilon_c+2)}\right) = o\left(n^{-1/4}\right), \\
&\vartheta^*_{d,t} = O\left(n^{-(q+1)/(2q+\upsilon_c+2)}\right) = o\left(n^{-1/4}\right), \\
&\log n / \left(n^{1-2/\zeta}(b^*_{d,t})^{\upsilon_c}\right) = O\left(\log n \cdot n^{2/\zeta-(2q+2)/(2q+\upsilon_c+2)}\right) = o(1), \\
&\log n / \left(n(b^*_{d,t})^{\upsilon_c}\right) = O\left(\log n \cdot \left(n^{-2(q+1)/(2q+\upsilon_c+2)}\right)\right) = o\left(n^{-1/2}\right).
\end{aligned}
$$

This concludes the proof. ■

D.3 DGPs for Monte Carlo simulation

In this section, we describe the DGP used by Simulation 1 in Section 5.1. Let $\mathbf{X} = (X_1, X_2, ..., X_6)$, where X_1 and X_2 are drawn from Uniform $[-1, 1]$, X_3 and X_4 are binary variables, following Bernoulli (0.5), and the remaining two, X_5 and X_6, are distributed as Binomial $(3, 0.5)$. The six variables are mutually independent.

Define

$$
f^{ps}_{1,0}(X) = 0.4 \sum_{s=1}^{2} (X_s - X_s^2) + 0.2 \sum_{k=3}^{6} X_k + 0.1 \left(\sum_{j\in\{3,5\}} (-1)^{j+1} X_j X_{j+1} \right.
$$

$$
\begin{aligned}
&+\sum_{l=1}^{2}\sum_{l'=3}^{6}(-1)^{l+1}X_lX_{l'}+\sum_{\ell=3}^{4}\sum_{\ell'=5}^{6}(-1)^{\ell+\ell'}X_\ell X_{\ell'}\Bigg),\\
f_{0,1}^{ps}(X)=&0.4(2X_1+X_2+X_1^2-X_2^2+X_1X_2)\\
&+0.2\sum_{k=3}^{6}(-1)^{k+1}X_k+0.1\left(\sum_{l=3}^{6}X_2X_l+\sum_{\ell=3}^{4}X_\ell X_6\right),\\
f_{0,0}^{ps}(X)=&0.4(X_1+2X_2-X_1^2+X_2^2-X_1X_2)\\
&+0.2\sum_{k=3}^{6}(-1)^{k}X_k+0.1\left(\sum_{l=3}^{6}X_1X_l+\sum_{\ell=3}^{4}X_\ell X_5\right),
\end{aligned}
$$

and for the OR models,

$$
\begin{aligned}
f_{base}^{or}(X)&=f_{het}^{or}(X)=27.4X_1+27.4X_2+13.7X_1^2+13.7X_2^2+13.7X_1X_2,\\
f_{att}^{or}(X)&=27.4X_1+13.7X_2+6.85\sum_{k=3}^{6}X_k-15.
\end{aligned}
$$

We consider the following data generating process

$$
p^{s1}(d,t,X)=\begin{cases}\dfrac{\exp(f_{d,t}^{ps}(X))}{1+\sum_{(d,t)\in\mathcal{S}_-}\exp(f_{d,t}^{ps}(X))}, & \text{if } (d,t)\in\mathcal{S}_-\\[2ex] \dfrac{1}{1+\sum_{(d,t)\in\mathcal{S}_-}\exp(f_{d,t}^{ps}(X))}, & \text{if } (d,t)=(1,1).\end{cases}
$$

Let $U\sim$ Uniform $[0,\,1]$. The treatment groups are assigned as follows

$$
(D,T)=\begin{cases}(1,0), & \text{if } U\leqslant p^{s1}(1,0,X),\\ (0,1), & \text{if } p^{s1}(1,0,X)<U\leqslant p^{s1}(1,0,X)+p^{s1}(0,1,X),\\ (0,0), & \text{if } p^{s1}(1,0,X)+p^{s1}(0,1,X)<U\leqslant 1-p^{s1}(1,1,X),\\ (1,1), & \text{if } 1-p^{s1}(1,1,X)<U.\end{cases}
$$

Next, building on Kang and Schafer (2007), we consider the following potential outcomes

$$Y_0(j)=210+f_{base}^{or}(X)+\epsilon_{het}+\epsilon_{j,0},\ \text{for } j=0,1, \tag{D.19}$$

$$Y_1(0)=210+2f_{base}^{or}(X)+\epsilon_{het}+\epsilon_{0,1}, \tag{D.20}$$

$$Y_1(1)=210+2f_{base}^{or}(X)+f_{att}^{or}(X)+\epsilon_{het}+\epsilon_{1,1}, \tag{D.21}$$

where $\epsilon_{het}\sim N(D\cdot f_{het}^{or},1)$ and $\epsilon_{d,t}$, $(d,t)\in\mathcal{S}$ are independent standard normal random variables.

D.4 Rescaled cross validation

To perform bandwidth selection using rescaled cross validation (RCV), the dataset is first split into a training set, $\mathfrak{I}_{tr}$ of size n_{tr}, and a validation set, $\mathfrak{I}_{val}$ of size $n_{val}=n-n_{tr}$. The rescaled loss functions are defined as follows:

$C_n^{rcv,ls}(h,\lambda,\{b_{d,t},\vartheta_{d,t}\}_{(d,t)\in\mathcal{S}_-})$

$$= \frac{1}{n_{val}} \sum_{i \in \mathfrak{I}_{val}} \left\{ \sum_{(d,t) \in \mathcal{S}} (I_{d,t,i} - \widehat{p}^{tr}(d,t,X_i))^2 + \sum_{(d,t) \in \mathcal{S}_-} I_{d,t,i}(Y_i - \widehat{m}^{tr}_{d,t}(X_i))^2 \right\}, \tag{D.22}$$

$$C_n^{rcv,ml}(h, \lambda, \{b_{d,t}, \vartheta_{d,t}\}_{(d,t) \in \mathcal{S}_-})$$

$$= \frac{1}{n_{val}} \sum_{i \in \mathfrak{I}_{val}} \left\{ - \sum_{(d,t) \in \mathcal{S}} I_{d,t,i} \log(\widehat{p}^{tr}(d,t,X_i)) + \sum_{(d,t) \in \mathcal{S}_-} I_{d,t,i}(Y_i - \widehat{m}^{tr}_{d,t}(X_i))^2 \right\}, \tag{D.23}$$

where $\widehat{p}^{tr}$ and $\widehat{m}^{tr}$ are calculated using (3.9) and (3.11), with the corresponding local polynomial coefficient estimates $\widehat{\gamma}^{tr}$ and $\widehat{\gamma}^{tr}$ obtained by

$$\widehat{\gamma}^{tr}(X_j) = \arg\max_{\gamma \in \mathbb{R}^{3N_p}} \frac{1}{n_{tr}} \sum_{i \in \mathfrak{I}_{tr}} \ell(W_i, X_j; \gamma) \widetilde{K}_{ps}(X_i; X_j, h, \lambda),$$

$$\widehat{\beta}^{tr}_{d,t}(X_j) = \arg\min_{\beta \in \mathbb{R}^{N_p}} \frac{1}{n_{tr}} \sum_{i \in \mathfrak{I}_{tr}} \left(Y_i - \underline{\mathbf{X}}_{q,i}(X_{c,j})'\beta\right)^2 I_{d,t,i} \widetilde{K}_{or}(X_i; X_j, b_{d,t}, \vartheta_{d,t}),$$

for each $j \in \mathfrak{I}_{val}$.

The rescaled cross-validated bandwidths, $(\widehat{h}^{rcv,j}, \widehat{\lambda}^{rcv,j}, \{\widehat{b}^{rcv,j}_{d,t}, \widehat{\vartheta}^{rcv,j}_{d,t}\}_{(d,t) \in \mathcal{S}})$, are defined by

$$\widehat{h}^{rcv,j} = \widetilde{h}^{rcv,j} \left(\frac{n_{tr}}{n}\right)^{\frac{1}{2p+\upsilon_c+2}}, \quad \widehat{\lambda}^{rcv,j} = \widetilde{\lambda}^{rcv,j} \left(\frac{n_{tr}}{n}\right)^{\frac{p+1}{2p+\upsilon_c+2}},$$

$$\widehat{b}^{rcv,j}_{d,t} = \widetilde{b}^{rcv,j}_{d,t} \left(\frac{n_{tr}}{n}\right)^{\frac{1}{2p+\upsilon_c+2}}, \quad \widehat{\vartheta}^{rcv,j}_{d,t} = \widetilde{\vartheta}^{rcv,j}_{d,t} \left(\frac{n_{tr}}{n}\right)^{\frac{p+1}{2p+\upsilon_c+2}},$$

where $(\widetilde{h}^{rcv,j}, \widetilde{\lambda}^{rcv,j}, \{\widetilde{b}^{rcv,j}_{d,t}, \widetilde{\vartheta}^{rcv,j}_{d,t}\}_{(d,t) \in \mathcal{S}})$ minimizes $C_n^{rcv,j}$ for $j = ls, ml$.

D.5 Plug-in estimators

When employing the frequency method (i.e., $\lambda = \vartheta_{d,t} = 0$), a straightforward plug-in rule can be used to determine the bandwidths $(h, \{b_{d,t}\}_{(d,t) \in \mathcal{S}_-})$. Notably, local polynomial estimators with an odd degree of fit are adaptive to boundaries, implying that the convergence rate of bias and variance remains constant regardless of the location of x. By solving Equation (D.14) and applying Theorem D.1, the following results are obtained

$$h^* = \left(\frac{\int \left\| \varrho^b_{p,1}(x_c) \mathbf{g}_-^{(p+1)}(x) \mathcal{I}(x) \right\|^2 \omega^{ps}(x) dx}{\int \mathrm{tr}\left(\mathcal{I}(x) \varrho^v_p(x_c)\right) / f_X(x) \cdot \omega^{ps}(x) dx} \frac{2(p+1)n}{\upsilon_c \{(p+1)!\}^2} \right)^{-1/(2p+\upsilon_c+2)},$$

$$b^*_{d,t} = \left(\frac{\int \left\| \varrho^b_{q,1}(x_c) \mathbf{m}^{(q+1)}_{d,t}(x) \right\|^2 \omega^{or}_{d,t}(x) dx}{\int \varrho^v_q(x_c) / f_X(x) \cdot \omega^{or}_{d,t}(x) dx} \frac{2(q+1)n}{\upsilon_c \{(q+1)!\}^2} \right)^{-1/(2q+\upsilon_c+2)}, \quad \text{for } (d,t) \in \mathcal{S}_-.$$

These bandwidths, however, are infeasible due to the presence of unknown quantities related to the derivatives of the nuisance functions and local Fisher information. To estimate the optimal bandwidths, preliminary approximations of these quantities are necessary. An additional challenge arises from the complicated dependence of the plug-in bandwidths on the location of x (through ϱ^b and ϱ^v). One possible solution is to substitute the values evaluated at a boundary point with those associated with interior points. This replacement has a negligible impact on the consistency of the optimal bandwidth in general. The bandwidth selection process can be outlined in the following algorithm:

Algorithm D.1 1. Let $\mathcal{X}_o$ collect all the unique values of $\{X_i\}_{i=1}^n$. Construct standard kernel estimates of covariate density with mixed data, $\widehat{f}_X(x)$, for $x \in \mathcal{X}_o$, following, e.g., Racine and Li (2004).

2. Use a polynomial multinomial logit regression of order $\ell = p + 2$ to get preliminary estimates $\breve{\mathcal{I}}(x), \breve{\mathbf{g}}_-^{(p+1)}(x), \breve{\mathbf{g}}_-^{(p+2)}(x)$, for $x \in \mathcal{X}_o$. Run polynomial regressions of order $\ell = q + 2$ to obtain $\breve{\mathbf{m}}_{d,t}^{(q+1)}(x)$ and $\breve{\mathbf{m}}_{d,t}^{(q+2)}(x)$, for $x \in \mathcal{X}_o$.

3. Compute preliminary bandwidths

$$\breve{h} = \left(\frac{\mathbb{E}_n\left[\left\|\varrho_{p,1}^b \breve{\mathbf{g}}_-^{(p+1)}(X)\breve{\mathcal{I}}(X)\right\|^2\right]}{\varrho_p^v\,\mathbb{E}_n\left[\widehat{f}_X^{-1}(X)\,\mathrm{tr}\left(\breve{\mathcal{I}}(X)\right)\right]} \frac{2(p+1)n}{\upsilon_c\{(p+1)!\}^2} \right)^{-1/(2p+\upsilon_c+2)},$$

$$\breve{b}_{d,t} = \left(\frac{\mathbb{E}_n\left[\left\|\varrho_{q,1}^b \breve{\mathbf{m}}_{d,t}^{(q+1)}(X)\right\|^2\right]}{\varrho_q^v\,\mathbb{E}_n\left[\widehat{f}_X^{-1}(X)\right]} \frac{2(q+1)n}{\upsilon_c\{(q+1)!\}^2} \right)^{-1/(2q+\upsilon_c+2)},$$

$$\tilde{h} = \left(\frac{\mathbb{E}_n\left[\left\|\boldsymbol{\varrho}_{p+1}^b \breve{\mathbf{g}}_-^{(p+2)}(X)\right\|^2\right]}{\mathbb{E}_n\left[\widehat{f}_X^{-1}(X)\,\mathrm{tr}\left(\breve{\mathcal{I}}(X)^{-1}\otimes \boldsymbol{\varrho}_{p+1}^v\right)\right]} \frac{2n}{\upsilon_c(2p+3)[(p+2)!]} \right)^{-1/(2p+\upsilon_c+4)},$$

$$\tilde{b}_{d,t} = \left(\frac{\mathbb{E}_n\left[\left\|\boldsymbol{\varrho}_{q+1}^b \breve{\mathbf{m}}_{d,t}^{(q+2)}(X)\right\|^2\right]}{\mathbb{E}_n\left[\left\|\widehat{f}_X^{-1}(X)\boldsymbol{\varrho}_{q+1}^v\right\|\right]} \frac{2n}{\upsilon_c(2q+3)[(q+2)!]} \right)^{-1/(2q+\upsilon_c+4)},$$

where we omitted the dependence of ϱ^b and ϱ^v on x_c to signify that the boundary effect is disregarded. Furthermore, in the preceding equations, $\boldsymbol{\varrho}_j^b = I_{N_j,\mathbf{j}}'\mathbf{Q}_j^{-1}\mathbf{M}_{j,j+1}$, $\boldsymbol{\varrho}_j^v = I_{N_j,\mathbf{j}}'\mathbf{Q}_j^{-1}\mathbf{T}_j\mathbf{Q}_j^{-1}I_{N_j,\mathbf{j}}$, and $I_{N_j,\mathbf{j}}$ is a $N_j \times n_j$ matrix consisting of the last n_j columns of the $N_j \times N_j$ identity matrix.

4. Run a local polynomial logistic regression of order $\ell = p + 1$, with bandwidth $\tilde{h}$, to obtain $\widehat{\mathbf{g}}_-^{(p+1)}(x)$. For each $(d,t) \in \mathcal{S}_-$, run a local polynomial regression of order $\ell = q + 1$, using bandwidth $\widehat{b}_{d,t}$, to get $\widehat{\mathbf{m}}_{d,t}^{(q+1)}(x)$, for $x \in \mathcal{X}_o$.

5. Run a local polynomial logistic regression of order $\ell = p$, with bandwidth $\breve{h}$, to obtain $\widehat{\mathcal{I}}(x)$, for $x \in \mathcal{X}_o$.

6. Compute the optimal bandwidth $\widehat{h}$ and $\widehat{b}_{d,t}$, following

$$\widehat{h} = \left(\frac{\mathbb{E}_n\left[\left\|\varrho_{p,1}^b \widehat{\mathbf{g}}_-^{(p+1)}(X)\widehat{\mathcal{I}}(X)\right\|^2\right]}{\varrho_p^v\,\mathbb{E}_n\left[\widehat{f}_X^{-1}(X)\,\mathrm{tr}\left(\widehat{\mathcal{I}}(X)\right)\right]} \frac{2(p+1)n}{\upsilon_c\{(p+1)!\}^2} \right)^{-1/(2p+\upsilon_c+2)},$$

$$\widehat{b}_{d,t} = \left(\frac{\mathbb{E}_n \left[\left\| \varrho_{q,1}^b \widehat{\mathbf{m}}_{d,t}^{(q+1)}(X) \right\|^2 \right]}{\varrho_q^v \, \mathbb{E}_n \left[\widehat{f}_X^{-1}(X) \right]} \frac{2(q+1)n}{v_c \left\{ (q+1)! \right\}^2} \right)^{-1/(2q+v_c+2)}.$$

D.6 Cluster-robust inference: bootstrap procedures

In this section, we introduce two bootstrap procedures that are suitable for cluster-robust inference. The first algorithm uses a multiplier-bootstrap method to compute studentized and cluster-robust standard errors. This method has been previously described in Kline and Santos (2012) and Callaway, Li and Oka (2018). The second procedure is a bootstrap Hausman-type test, which provides bootstrapped p-values.

Let $V_{i=1}^n$ be a sequence of $i.i.d.$ random variables with zero mean and unit variance, which is independent of the original sample. One example is $i.i.d.$ Bernoulli random variables with $P(V = v_0) = 1 - v_0/\sqrt{5}$ and $P(V = 1 - v_0) = v_0/\sqrt{5}$, where $v_0 = (\sqrt{5}+1)/2$, as suggested by Mammen (1993). Now, given a generic ATT estimator, $\widehat{\tau}$, and an estimator of its influence function, $\widehat{\eta}(\cdot)$, we compute the clustered standard errors as follows:

Algorithm D.2 1. In iteration b, draw a realization of V_b for each cluster. All observations within the same cluster share the same value of V_b.

2. Calculate a bootstrap estimate for ATT as

$$\widehat{\tau}_b^* = \widehat{\tau} + \mathbb{E}_n[V_b \cdot \widehat{\eta}(W)].$$

Form a bootstrap draw of the limiting distribution as

$$\widehat{R}_b^* = \sqrt{n}\left(\widehat{\tau}_b^* - \widehat{\tau}\right).$$

3. Repeat Steps 1-2 B times.

4. Calculate the bootstrapped standard error, $\hat{\sigma}^*$, as the bootstrap interquartile range normalized by the interquartile range of the standard normal distribution: $\hat{\sigma}^* = (q_{0.75}(\widehat{R}^*) - q_{0.25}(\widehat{R}^*))/(z_{0.75} - z_{0.25})$, where $q_p(\widehat{R}^*)$ is the p-th sample quantile of the bootstrap draws $\{\widehat{R}_b^*\}_{b=1}^B$, and z_p is the p-th quantile of the standard normal distribution.

Given the two DR DID estimators, $\widehat{\tau}_{dr}$ based on (3.1), $\widehat{\tau}_{sz}$ based on (4.1), and their respective linear expansions, $\widehat{\eta}_{dr}(\cdot)$ given in (3.13) and $\widehat{\eta}_{sz}(\cdot)$ given in (4.3), we conduct a cluster-robust Hausman-type test as follows

Algorithm D.3 1. Calculate the Hausman test statistic, $\mathcal{T}_n$, following (4.2).

2. In iteration b, generate a realization of V_b for each cluster. Observations within the same cluster share the same value of V_b.

3. Calculate bootstrap estimates of the ATT as

$$\widehat{\tau}_{j,b}^* = \widehat{\tau}_j + \mathbb{E}_n[V_b \cdot \widehat{\eta}_j(W)],$$

$$\hat{V}_b^* = \mathbb{E}_n[((V_b \cdot (\widehat{\eta}_{eff}(W) - \widehat{\eta}_{sz}(W)))^2].$$

Form a bootstrap test statistic, $\mathcal{T}_b^*$, as

$$\mathcal{T}_b^* = n\left(\widehat{\tau}_{dr,b}^* - \widehat{\tau}_{sz,b}^*\right)^2 / \hat{V}_b^*.$$

4. Repeat Steps 2-3 B times.

5. Calculate the bootstrapped p-value, p^*, as the proportion of the bootstrap test statistics, $\{\mathcal{T}_b^*\}_{b=1}^B$, that are greater than or equal to $\mathcal{T}_n$.

References

Andrews, Donald WK, "Asymptotics for semiparametric econometric models via stochastic equicontinuity," *Econometrica: Journal of the Econometric Society*, 1994, pp. 43–72.

Callaway, Brantly, Tong Li, and Tatsushi Oka, "Quantile treatment effects in difference in differences models under dependence restrictions and with only two time periods," *Journal of Econometrics*, oct 2018, *206* (2), 395–413.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins, "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, jun 2017, pp. 1–71.

Fan, Yangin and Emmanuel Guerre, "Multivariate local polynomial estimators: Uniform boundary properties and asymptotic linear representation," in "Essays in Honor of Aman Ullah," Emerald Group Publishing Limited, 2016.

Hahn, Jinyong, "On the role of the propensity score in efficient semiparametric estimation of average treatment effects," *Econometrica*, 1998, pp. 315–331.

Hausman, Jerry A., "Specification tests in econometrics," *Econometrica*, 1978, pp. 1251–1271.

Kang, Joseph D. Y. and Joseph L. Schafer, "Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data.," *Statistical Science*, 2007, *22* (4), 569–573.

Kline, Patrick and Andres Santos, "A Score Based Approach to Wild Bootstrap Inference," *Journal of Econometric Methods*, 2012, *1* (1), 1–40.

Newey, Whitney K., "Semiparametric efficiency bounds," *Journal of Applied Econometrics*, apr 1990, *5* (2), 99–135.

Racine, Jeff and Qi Li, "Nonparametric estimation of regression functions with both categorical and continuous data," *Journal of Econometrics*, 2004, *119* (1), 99–130.

Rothe, Christoph and Sergio Firpo, "Properties of doubly robust estimators when nuisance functions are estimated nonparametrically," *Econometric Theory*, 2019, *35* (5), 1048–1087.

_ **and** _ , "Properties of doubly robust estimators when nuisance functions are estimated nonparametrically," *Econometric Theory*, 2019, *35* (5), 1048–1087.

Sant'Anna, Pedro H. C. and Jun Zhao, "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, 2020, *219* (1), 101–122.

Sherman, Robert P, "Maximal inequalities for degenerate U-processes with applications to optimization estimators," *The Annals of Statistics*, 1994, *22* (1), 439–459.

Tanabe, Kunio and Masahiko Sagae, "An exact Cholesky decomposition and the generalized inverse of the variance–covariance matrix of the multinomial distribution, with applications," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1992, *54* (1), 211–219.

van der Vaart, Aad W., *Asymptotic Statistics*, Cambridge: Cambridge University Press, 1998.

— **and Jon A. Wellner**, *Weak Convergence and Empirical Processes*, New York: Springer, 1996.