Difference-in-Differences with Compositional Changes: Supplemental Appendix

Pedro H. C. Sant'Anna* Qi Xu†

April 26, 2023

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

Notation: 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 $f_X(x) = f_{X_c|X_d}(x_c|x_d) \cdot \mathbb{P}(X_d = x_d)$, $\mathbb{N}_n = \{1, 2, ..., n\}$, and $\iota(d,t) = \mathbb{1}\{d = 1, t = 0\} + 2 \cdot \mathbb{1}\{d = 0, t = 1\} + 3 \cdot \mathbb{1}\{d = 0, t = 0\}$. The notation $a_n \lesssim b_n$ implies that $a_n \leqslant cb_n$ for some positive constant c when n is sufficiently large. The symbol $a_n \sim b_n$ denotes that $a_n/b_n \to 1$ as $n \to \infty$. We define $f \in L_2(\mathcal{U})$ to indicate that $\int_{\mathcal{U}} f^2 d\mu$ is finite, and let the L_2- and sup-norm of f to denote $\|f\|_{L_2}$ and $\|f\|_\infty$, respectively. Denote the ATT by τ, i.e.,

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

A Proofs for main results in the text

Let

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

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

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

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

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

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

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

Proof of Lemma A.1:

*Emory University. E-mail: pedrosantanna@causal-solutions.com

†Vanderbilt University. E-mail: qi.xu.1@vanderbilt.edu.

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

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

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

Propensity score estimand: Let $p(1,1) = \mathbb{P}(D = 1, T = 1)$. 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,

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

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 1:

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

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

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

$$+ t_\theta(x),$$

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

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

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

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

$$p_{dt}(\cdot) \in L_2(\mathcal{X}), \text{ with } \sum_{(d,t)\in\mathcal{S}} \int p_{dt}(x)f(x)dx = 0, \tag{A.2}$$

and

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

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 A.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

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

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

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

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

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

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 $F_\tau(\cdot) \in \mathcal{T}$. By setting

$$\begin{aligned} s_{11}(y,x) &= \frac{y - m_{1,1}(x)}{p(1,1)}, \\ p_{11}(x) &= p(1,1)^{-1} \sum_{(d,t)\in\mathcal{S}} (-1)^{d+t}\left(m_{d,t}(x) - \int_{\mathcal{X}} m_{d,t}(x) f(x) dx\right), \\ s_{dt}(y,x) &= (-1)^{d+t} \frac{p(1,1,x)(y - m_{d,t}(x))}{p(d,t,x)p(1,1)}, \\ p_{dt}(x), s(x) &= 0, \end{aligned}$$

for $(d,t) \in \mathcal{S}_-$, it is straightforward to show that (A.1)-(A.3) hold, which leads to the desired result.

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 1: The proof follows directly from the LIE as displayed in the main text. ■

Proof of Proposition 2:
It follows by Theorem 1 that

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

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

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

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 $V_{1,sz} \equiv \mathbb{E}\left[D(\tau(X) - \tau)^2\right]/p^2$, and

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

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

$$V_{1,dr} = \frac{1}{\lambda p^2}\,\mathbb{E}[D(\tau(X)-\tau)^2], \tag{A.6}$$

and therefore,

$$V_{1,dr} - V_{1,sz} = \frac{1-\lambda}{p^2\lambda}\,\mathbb{E}[D(\tau(X)-\tau)^2]. \tag{A.7}$$

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

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

where the first equality follows because $p(d,t,x) = \mathbb{P}(D=d, X=x)\cdot\mathbb{P}(T=t)$ under Assumption 3. The desired result then follows from (A.7) and (A.8). ∎

Proof of Lemma 3.1:

Let $\psi_{d,t}(W;w,m) = \mathbb{1}\{dt=1\}w_{1,1}(D,T)Y + \mathbb{1}\{dt\neq 1\}\{w_{d,t}(D,T,X)(Y - m_{d,t}(X)) + w_{1,1}(D,T)m_{d,t}(X)\}$, and $\tilde{\tau}_{dr} = \sum_{(d,t)\in\mathcal{S}}(-1)^{d+t}\psi_{d,t}(W;w,m)$. Using $\tilde{\tau}_{dr}$, we decompose $\widehat{\tau}_{dr}$ as

$$\widehat{\tau}_{dr} - \tau = (\widehat{\tau}_{dr} - \tilde{\tau}_{dr}) + (\tilde{\tau}_{dr} - \tau)\,. \tag{A.9}$$

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

$$\begin{aligned}
\psi_{d,t}(W;\widehat{w},\widehat{m}) - \psi_{d,t}(W;w,m) =& (Y - m_{d,t}(X))\,(\widehat{w}_{d,t} - w_{d,t})\,(W) + m_{d,t}(X)\,(\widehat{w}_{1,1} - w_{1,1})\,(W)\\
&+ (w_{1,1} - w_{d,t})\,(W)\,(\widehat{m}_{d,t} - m_{d,t})\,(X)\\
&+ \{(\widehat{w}_{1,1} - w_{1,1})\,(W) - (\widehat{w}_{d,t} - w_{d,t})\,(W)\}\,(\widehat{m}_{d,t} - m_{d,t})\,(X)\\
\equiv& \Delta^{\psi,1}_{d,t}(W) + \Delta^{\psi,2}_{d,t}(W) + \Delta^{\psi,3}_{d,t}(W),
\end{aligned}$$

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 $\psi_{1,1}(W;\widehat{w},\widehat{m})$ around $\mathbb{E}[DT]$ yields that

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

$$= -\frac{\mathbb{E}[DTY]}{\mathbb{E}[DT]^2} \cdot (\mathbb{E}_n[DT] - \mathbb{E}[DT]) + o_p(n^{-1/2}). \tag{A.10}$$

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

$$\begin{aligned}
\mathbb{E}_n\left[\Delta_{d,t}^{\psi,1}(W)\right] &= \mathbb{E}_n\left[(Y - m_{d,t}(X))\left(\widehat{w}_{d,t} - w_{d,t}\right)(W) + m_{d,t}(X)\left(\widehat{w}_{1,1} - w_{1,1}\right)(W)\right] \\
&= \mathbb{E}_n\left[(Y - m_{d,t}(X))\left(\widehat{w}_{d,t} - w_{d,t}\right)(W)\right] \\
&\quad - \frac{\mathbb{E}_n\left[DTm_{d,t}(X)\right]}{\mathbb{E}[DT]^2}(\mathbb{E}_n[DT] - \mathbb{E}[DT]) + O_p(|\mathbb{E}_n[DT] - \mathbb{E}[DT]|^2) \\
&= -\frac{\mathbb{E}\left[DTm_{d,t}(X)\right]}{\mathbb{E}[DT]^2}(\mathbb{E}_n[DT] - \mathbb{E}[DT]) + o_p(n^{-1/2}),
\end{aligned} \tag{A.11}$$

where the last equation holds under Assumption 4.2(i).

Next, note that $\Delta_{1,1}^{\psi,2}(\cdot) = 0$, and when $(d,t) \in \mathcal{S}_-$, we deduce from Assumption 4.2(ii) that

$$\mathbb{E}_n\left[\Delta_{d,t}^{\psi,2}(W)\right] = \mathbb{E}_n\left[(w_{1,1} - w_{d,t})(W)\left(\widehat{m}_{d,t} - m_{d,t}\right)(X)\right] = o_p(n^{-1/2}). \tag{A.12}$$

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

$$\begin{aligned}
&\mathbb{E}_n\left[\Delta_{d,t}^{\psi,3}(W)\right] \\
&= \mathbb{E}_n\left[\left(\left(\widehat{w}_{1,1} - w_{1,1}\right)(W) - \left(\widehat{w}_{d,t} - w_{d,t}\right)(W)\right)\left(\widehat{m}_{d,t} - m_{d,t}\right)(X)\right] \\
&= \mathbb{E}_n\left[\frac{DT}{\mathbb{E}[DT]^2}\left(\widehat{m}_{d,t} - m_{d,t}\right)(X)\right] \cdot (\mathbb{E}_n[DT] - \mathbb{E}[DT]) + O_p(|\mathbb{E}_n[DT] - \mathbb{E}[DT]|^2) \qquad \text{(A.13)} \\
&\quad - \mathbb{E}_n\left[\left(\widehat{w}_{d,t} - w_{d,t}\right)(W) \cdot \left(\widehat{m}_{d,t} - m_{d,t}\right)(X)\right], \qquad \text{(A.14)}
\end{aligned}$$

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 (A.13) 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

$$w_{d,t}^\dagger(W) = \frac{I_{d,t}\widehat{p}(1,1,X)}{p(1,1)\widehat{p}(d,t,X)}, \tag{A.15}$$

based on which, we have the following decomposition

$$\mathbb{E}_n\left[\left(w_{d,t}^\dagger - w_{d,t}\right)(W) \cdot \left(\widehat{m}_{d,t} - m_{d,t}\right)(X)\right] + \mathbb{E}_n\left[\left(\widehat{w}_{d,t} - w_{d,t}^\dagger\right)(W) \cdot \left(\widehat{m}_{d,t} - m_{d,t}\right)(X)\right] = \Delta_{w,m}^{1,n} + \Delta_{w,m}^{2,n}. \tag{A.16}$$

We consider the L_2-norm first. Under Assumption 4.2(iii),

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

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

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

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,

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

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 $\sup_{x\in\mathcal{X}}|\widehat{p}(d,t,x) - p(d,t,x)| \leqslant p^{min}_{d,t}/2$ with probability at least $1-\epsilon$, whenever $n \geqslant N_\epsilon$. Thus, when n is sufficiently large, we have

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

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 $\Delta^{1,n}_{w,m} = O_p\left(r_n s_n\right)$, without the necessity of imposing Assumption 4.2(iii).

Next,we examine the estimation effect of the normalizing weight as given in $\Delta^{2,n}_{w,m}$. 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,

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

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

$$\begin{aligned}\Delta^{2,1,n}_{w,m} =& \Delta^{1,n}_{w,m} && \text{(A.17)}\\ &+ (\mathbb{E}_n - \mathbb{E})\left[w_{d,t}(W)\cdot(\widehat{m}_{d,t} - m_{d,t})(X)\right] && \text{(A.18)}\\ &+ \mathbb{E}\left[w_{d,t}(W)\cdot(\widehat{m}_{d,t} - m_{d,t})(X)\right]. && \text{(A.19)}\\ =& O_p(r_n) + O_p(r_n s_n) + o_p\left(n^{-1/2}\right)\end{aligned}$$

Under Assumptions 4.2(iii, iv), (A.17) and (A.18) are $O_p(r_n s_n)$ and $o_p\left(n^{-1/2}\right)$, respectively. Since $p_{d,t}(\cdot)$ is uniformly bounded over $\mathcal{X}$, (A.19) is $O_p(r_n)$ by the Cauchy-Schwartz inequality.

Analogously, we have

$$\begin{aligned}\widehat{p}(1,1) - p(1,1) =& (\mathbb{E}_n - \mathbb{E})\left[I_{d,t}\left(\frac{\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)} - \frac{p(1,1,X)}{p(d,t,X)}\right)\right] && \text{(A.20)}\\ &+ (\mathbb{E}_n - \mathbb{E})\left[I_{d,t}\frac{p(1,1,X)}{p(d,t,X)}\right] && \text{(A.21)}\\ &+ \mathbb{E}\left[I_{d,t}\left(\frac{\widehat{p}(1,1,X)}{\widehat{p}(d,t,X)} - \frac{p(1,1,X)}{p(d,t,X)}\right)\right] && \text{(A.22)}\\ =& O_p(s_n) + O_p\left(n^{-1/2}\right) + o_p\left(n^{-1/2}\right).\end{aligned}$$

Under Assumption 4.2(v), (A.20) is $o_p\left(n^{-1/2}\right)$. Since (A.21) 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 (A.22) is $O_p(s_n)$. 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 (A.9), (A.10), (A.11), (A.12), (A.14), and (A.16), which leads to

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

■

Proof of Theorem 2:

We proceed by applying Lemma 3.1. As we are working with the sup-norm, we need to verify the first two conditions in Assumption 4.2. Lemmas C.2 and C.3 provide the required verification for these conditions. With the bandwidth rate conditions in Assumption 5.5 guaranteeing that the leading remainder term is $O_p(r_n s_n) = o_p\left(n^{-1/2}\right)$, we can then derive the asymptotic normality directly from the CLT. ■

Proof of Theorem 3:

Proof of Part (a): We have already shown in Theorem 2 that $\widehat{\tau}_{dr} - \tau = \mathbb{E}_n[\eta_{\text{eff}}(W)] + o_p\left(n^{-1/2}\right)$. Following a similar line of reasoning, one can easily demonstrate that $\widehat{\tau}_{sz} - \tau = \mathbb{E}_n[\eta_{sz}(W)] + o_p\left(n^{-1/2}\right)$, under Assumptions 1, 2, 5, Condition (i), and the null hypothesis, $\mathbf{H}_0$. Now, by the CLT, we have

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

It remains to show that

$$\widehat{V}_n \xrightarrow{p} V, \tag{A.23}$$

and

$$V = \rho_{sz} > 0. \tag{A.24}$$

First, it is implied from the proof of Theorem 2 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 $\mathbf{H}_0$. Combining these two results, (A.23) 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 τ_{sz}, which is equal to τ under $\mathbf{H}_0$. Moreover, since both $\widehat{\tau}_{dr}$ and $\widehat{\tau}_{sz}$ are consistent for τ_{sz} under $\mathbf{H}_0$, 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, (A.24) now follows by Proposition 2 and the condition that $\mathbb{V}\text{ar}\left[\tau(X)|D=1\right] > 0$.

Proof of Part (b): 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 $\mathbf{H}_1$.

Under Assumption 5, and Condition (i) of the theorem, $\widehat{p}(d,t,x) \xrightarrow{p} p(d,t,x)$ and $\widehat{m}_{d,t}(x) \xrightarrow{p} m_{d,t}(x)$, uniformly in x, 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 (A.23) holds under $\mathbf{H}_1$. 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). ■

B 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 B.2.

B.1 Rates of convergence: U-Statistic

Let $\{X_i\}_{i=1}^n$ be a random sample from an unknown distribution. Given a real-valued function $h(x_1, ..., x_r)$ that possibly depends on the sample size, define

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

as a r-th order U-statistic with kernel h, where the summation is over $S(n,r)$, the set of permutation $(s_1, ..., s_r)$ of size r of the set $\{1, ..., n\}$. Since a given function h can always be replaced by a symmetric one, we restrict attention to symmetric kernels in what follows. That is, U_n can be equivalently represented as

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

where $\mathcal{C}(n,r)$ is the set of combinations $(s_1, ..., s_r)$ of size r of the set $\{1, ..., n\}$.

For $1 \leqslant s \leqslant r$, define the quantities h_s and σ_s by

$$h_s(x_1, ..., x_s) = \mathrm{E}[h(x_1, ..., x_s, X_{s+1}, ..., X_r)] \quad \text{and} \quad \sigma_s = \mathbb{V}\mathrm{ar}\left[h_s(X_1, ..., X_s)\right]^{1/2}.$$

We call U_n with kernel h is s^*'th order degenerate if $\sigma_s = 0$ for all $s \leqslant s^*$.

Lemma B.1 Let $h : \mathcal{X}^r \to \mathbb{R}$ be a permutation-symmetric, measurable function of r arguments such that $\mathrm{E}[h(X_1, ..., X_r)] = 0$, and $\sigma_r < \infty$, then $U_n = O_p\left(\sum_{s=1}^{r} \frac{\sigma_s}{n^{s/2}}\right)$.

Note that if the U-statistic is s^*-th order degenerate, its convergence rate is $\sum_{s=s^*+1}^{r} \frac{\sigma_s}{n^{s/2}}$. The lemma follows directly from Markov's inequality, and therefore, we omit the proof.

B.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

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

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 $G^{(ps)}_{1,1}(x) = -\sum_{(d,t)\in\mathcal{S}_-} G^{(ps)}_{d,t}(x)$. Additionally, we define, for a given observation X_j

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

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

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

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

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

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

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

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

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

$$\sup_{j\in\mathbb{N}_n}\left|B^{(or)}_{n,d,t}(X_j)\right| = O_p(b^{q+1}_{d,t} + \vartheta_{d,t,o} + \vartheta_{d,t,u}),$$

$$\sup_{j\in\mathbb{N}_n}\left|S^{(or)}_{n,d,t}(X_j)\right| = O_p\left(\sqrt{\log n/(nb_{d,t}{}^{\upsilon_c})}\right),$$

$$\sup_{j\in\mathbb{N}_n}\left|R^{(or)}_{n,d,t}(X_j)\right| = O_p\left(\left(b^{p+1}_{d,t} + \vartheta_{d,t,o} + \vartheta_{d,t,u} + \sqrt{\log n/(nb_{d,t}{}^{\upsilon_c})}\right)^2\right).$$

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

$$\mathbf{Q}_\ell(x_c) = \begin{pmatrix} \mathbf{Q}^{(0,0)}(\mathcal{S}_{x_c,h}) & \cdots & \mathbf{Q}^{(0,\ell)}(\mathcal{S}_{x_c,h}) \\ \vdots & \ddots & \vdots \\ \mathbf{Q}^{(\ell,0)}(\mathcal{S}_{x_c,h}) & \cdots & \mathbf{Q}^{(\ell,\ell)}(\mathcal{S}_{x_c,h}) \end{pmatrix}, \tag{B.6}$$

$$\mathbf{T}_\ell(x_c) = \begin{pmatrix} \mathbf{T}^{(0,0)}(\mathcal{S}_{x_c,h}) & \dots & \mathbf{T}^{(0,\ell)}(\mathcal{S}_{x_c,h}) \\ \vdots & \ddots & \vdots \\ \mathbf{T}^{(\ell,0)}(\mathcal{S}_{x_c,h}) & \dots & \mathbf{T}^{(\ell,\ell)}(\mathcal{S}_{x_c,h}) \end{pmatrix},$$

$$\mathbf{M}_{\ell,k}(x_c) = \begin{pmatrix} \mathbf{Q}^{(0,k)}(\mathcal{S}_{x_c,h}) \\ \dots \\ \mathbf{Q}^{(\ell,k)}(\mathcal{S}_{x_c,h}) \end{pmatrix},$$

where $\mathbf{Q}_\ell^{(i,j)}(\mathcal{S})$ and $\mathbf{T}_\ell^{(i,j)}(\mathcal{S})$ 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 B.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 (2019) 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 (B.3), we have

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

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

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

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

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

and thus, $\sup_{x\in\mathcal{X}} \left\| \Sigma^{ps}(x)^{-1} \right\| = O(1)$. Now, from Lemma B.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 B.3 and the CMT to deduce (B.4).

To establish (B.5), the proof proceed through three steps. First, we demonstrate the existence of a global maximizer for the local log-likelihood function defined in (3.6). 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

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

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

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

where

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

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 B.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

$$\inf_{x,y\in\mathcal{X}} \mathcal{I}(x,y,\gamma(y)) > \inf_{x,y\in\mathcal{X}} \left\{ \prod_{(d,t)\in\mathcal{S}_-} \Psi_{d,t}(\{\underline{\mathbf{x}}(y)'\gamma_{d,t}(y)\}_{(d,t)\in\mathcal{S}_-}) \right\} \cdot I_3, \tag{B.10}$$

in the sense that their difference 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 right-hand side of (B.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 $\mathbb{P}(E_{1n}(c_1)) > 1 - \epsilon/2$, $\mathbb{P}(E_{2n}(c_2)) > 1 - \epsilon/2$, and (B.10) is satisfied. Now, set $\delta_0 > 2c_1 c_2^{-1} c_3^{-1}$. 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 $\sup_{x\in\mathcal{X}} \{\mathcal{L}_n^{ps}(\bar{\gamma}(x); x) - \mathcal{L}_n^{ps}(\bar{\gamma}^*(x); x)\} < 0$, with a probability of at least $1 - \epsilon$. This is because

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

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, due to the concavity of $\mathcal{L}_n^{ps}(\cdot; x)$, $\hat{\bar{\gamma}}(x)$ 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 $\|\hat{\bar{\gamma}}(\cdot) - \bar{\gamma}^*(\cdot)\|_\infty = O_p(\kappa_n)$.

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

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

Furthermore,

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

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

By the triangular inequality and the Cauchy-Schwarz inequality,

$$\begin{aligned}
\sup_{x\in\mathcal{X}}\|\mathbf{C}_n(x)\| \leqslant &\frac{1}{2n}\sum_{i=1}^{n}\sum_{(d,t)\in\mathcal{S}_-}\sum_{(d',t')\in\mathcal{S}_-}\left\|\dot{\mathcal{I}}_{\iota(d,t),\iota(d',t')}(X_{c,i},x_c,\tilde{\gamma}(x))\right\| \\
&\cdot\left\|\hat{\bar{\gamma}}_{d,t}(x)-\bar{\gamma}^*_{d,t}(x)\right\|\cdot\left\|\hat{\bar{\gamma}}_{d',t'}(x)-\bar{\gamma}^*_{d',t'}(x)\right\|\cdot\|H(h)\underline{\mathbf{X}}_i(x_c)\|^3\cdot\left|\widetilde{K}_{ps}(X_i;x,h,\lambda)\right| \\
\lesssim &\max_{(d,t),(d',t')\in\{0,1\}}\sup_{x,z\in\mathcal{X}}\left\{\left\|\dot{\mathcal{I}}_{\iota(d,t),\iota(d',t')}(z_c,x_c,\tilde{\gamma}(x))\right\|\cdot\left\|\hat{\bar{\gamma}}_{d,t}(x)-\bar{\gamma}^*_{d,t}(x)\right\|\cdot\left\|\hat{\bar{\gamma}}_{d',t'}(x)-\bar{\gamma}^*_{d',t'}(x)\right\|\right\}
\end{aligned} \tag{B.11}$$

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

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 (B.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 (B.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 (B.5). This completes the proof of the lemma. ■

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

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

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

Proof of Lemma B.3:
The proof of (B.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 $\kappa_n = (\log n / (nh^{v_c}))^{1/2}$. 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)

$$\max_{i \in \mathbb{N}_n} \left| \widetilde{A}_{d,t}^{\dagger,(\mathbf{k})}(W_i, x_c, \gamma^*(x)) - \widetilde{A}_{d,t}^{\dagger,(\mathbf{k})}(W_i, x_c', \gamma^*(x')) \right| \leqslant h^{v_c} \kappa_n \epsilon / 3, \tag{B.15}$$
$$\left| \mathbb{E}\left[\widetilde{A}_{d,t}^{\dagger,(\mathbf{k})}(W, x_c, \gamma^*(x)) \right] - \mathbb{E}\left[\widetilde{A}_{d,t}^{\dagger,(\mathbf{k})}(W, x_c', \gamma^*(x')) \right] \right| \leqslant h^{v_c} \kappa_n \epsilon / 3, \tag{B.16}$$

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 $J_n = O(n^{\kappa_b})$, $\kappa_b > 0$, 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 $x \in \mathcal{B}(x_j, \delta_n) \cap \mathcal{X}$, and for all $x' \in \mathcal{B}(x_j, \delta_n)$, $x_d' = x_{d,j}$. As a result, $\mathcal{X} = \bigcup_{j=1}^{J_n} (\mathcal{B}(x_j, \delta_n) \cap \mathcal{X})$.

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

$$\sup_{x \in \mathcal{X}} \left| \frac{1}{n} \sum_{i=1}^{n} \widetilde{A}_{d,t}^{(\mathbf{k})}(W_i, x, \gamma^*(x)) - \mathbb{E}[\widetilde{A}_{d,t}^{(\mathbf{k})}(W, x, \gamma^*(x))] \right|$$
$$\leqslant \max_{j \in \mathbb{N}_{J_n}} \left| \frac{1}{n} \sum_{i=1}^{n} \widetilde{A}_{d,t}^{(\mathbf{k})}(W_i, x_j, \gamma^*(x_j)) - \mathbb{E}[\widetilde{A}_{d,t}^{(\mathbf{k})}(W, x_j, \gamma^*(x_j))] \right| \tag{B.17}$$
$$+ \max_{j \in \mathbb{N}_{J_n}} \sup_{x \in \mathcal{B}(x_j, \delta_n) \cap \mathcal{X}} \left| \frac{1}{n} \sum_{i=1}^{n} \left(\widetilde{A}_{d,t}^{(\mathbf{k})}(W_i, x, \gamma^*(x)) - \widetilde{A}_{d,t}^{(\mathbf{k})}(W_i, x_j, \gamma^*(x_j)) \right) \right| \tag{B.18}$$
$$+ \max_{j \in \mathbb{N}_{J_n}} \sup_{x \in \mathcal{B}(x_j, \delta_n) \cap \mathcal{X}} \left| \mathbb{E}[\widetilde{A}_{d,t}^{(\mathbf{k})}(W, x, \gamma^*(x))] - \mathbb{E}[\widetilde{A}_{d,t}^{(\mathbf{k})}(W, x_j, \gamma^*(x_j))] \right|. \tag{B.19}$$

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

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

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

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

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

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

---

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 $t \geqslant 0$,

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

With these two results in hand, we have

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

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, (B.17) $\leqslant\kappa_n\epsilon/3$.

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

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}}\left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{A}^{(\mathbf{k})}_{d,t}(W_i,x,\gamma^*(x))-\mathbb{E}[\widetilde{A}^{(\mathbf{k})}_{d,t}(W,x,\gamma^*(x))]\right|\geqslant\kappa_n\epsilon\right)\to 0. \tag{B.20}$$

This complete the proof for (B.13).

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

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

where $(s.o.)$ stands for smaller order terms. The last equality is due to Assumptions 5.2 and 5.4. ■

C Auxiliary lemmas and results

C.1 Auxiliary lemmas

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

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

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

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

Lemma C.2 Suppose the conditions of Theorem 2 hold. Then, for $\widehat{w}$ defined in (3.1) with $\widehat{p}$ given by (3.7), we have

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

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

Proof of Lemma C.2:

Recall the definition of $w^\dagger$ as given in (A.15), and decompose the difference between $\widehat{w}_{d,t}$ and $w_{d,t}$ as

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

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

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

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 $R_{n,d,t} = O_p\left(\|\widehat{p}(1,1,\cdot) - p(1,1,\cdot)\|^2_\infty\right) + O_p\left(\|\widehat{p}(d,t,\cdot) - p(d,t,\cdot)\|^2_\infty\right)$, which is $o_p\left(n^{-1/2}\right)$ by Lemma B.2 and Assumption 5.5.

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

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

$$\Delta^{11}_w = \frac{1}{n}\sum_{i=1}^n \left\{\frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}\left(B^{(ps)}_{n,1,1}(X_i) + S^{(ps)}_{n,1,1}(X_i) + R^{(ps)}_{n,1,1}(X_i)\right)\right\}.$$

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

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

where the rate of the remainder comes from standard variance calculation. Owning to Lemma C.1(i), the mean on the right-hand side is zero, which leads to

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

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

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

By its definition, we have

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

Given the construction, we have $\mathbb{E}\left[\psi_{w,d,t}(W_i, W_j)|W_i\right] = 0$. Moreover, by Lemma C.1(i), we also have that $\mathbb{E}\left[\psi_{w,d,t}(W_i, W_j)|W_j\right] = 0$. Hence, (C.5) represents a second-order U-statistic with first-order degenerate kernel. Lemma B.1 and standard variance calculation then gives that

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

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

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\frac{I_{d,t,i}(Y_i - m_{d,t}(X_i))}{p(d,t,X_i)p(1,1)}R^{(ps)}_{n,1,1}(X_i) &= O_p\left(\sup_{i\in\mathbb{N}_n}\left|R^{(ps)}_{n,1,1}(X_i)\right|\right)\\ &= O_p\left(\left(h^{p+1} + \lambda_o + \lambda_u + \sqrt{\log n/\left(nh^{v_c}\right)}\right)^2\right)\end{aligned} \tag{C.7}$$

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

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

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

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

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 $O_p\left(n^{-1}h^{-v_c/2}\right) = o_p\left(n^{-1/2}\right)$. The detailed proof is omitted for brevity.

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

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

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 B.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 C.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 C.3 Suppose the conditions of Theorem 2 hold, then with $\widehat{m}$ given by (3.9),

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

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

Proof of Lemma C.3:
The proof closely resembles the first part of Lemma C.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 C.1(ii).

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

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

By definition,

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

In view of Lemma C.1(ii), the right-hand side of the above equation is a second-order U-statistic with a degenerate first-order kernel. A standard variance calculation shows that it is of the order $O_p\left(n^{-1}b_{d,t}^{\upsilon_c/2}\right)$, 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. ■

C.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

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

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

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

For the OR functions, we define

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

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

$$\chi^*(h,\lambda,\{b_{d,t},\vartheta_{d,t}\}_{(d,t)\in\mathcal{S}_-}) = \int_{\mathcal{X}} \left\{\|\mathcal{B}^{ps}(x,h,\lambda)\|^2 + \mathrm{tr}(\mathcal{V}^{ps}(x,h,\lambda))\right\} \omega^{ps}(x)dx + \sum_{(d,t)\in\mathcal{S}_-} \int_{\mathcal{X}} \left\{\mathcal{B}^{or}_{d,t}(x,b_{d,t},\vartheta_{d,t})^2 + \mathcal{V}^{or}_{d,t}(x,b_{d,t},\vartheta_{d,t})\right\} \omega^{or}_{d,t}(x)dx. \tag{C.10}$$

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 C.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 C.1 Assuming that Assumptions 1, 5, and C.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

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

Proof of Theorem C.1:
From the uniform linear expansions of Lemma B.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

$$\begin{aligned} &\mathbb{E}[\mathcal{I}(x)\mathbf{A}_-(W,x)] = \mathcal{I}(x)(I_3\otimes e_{N_p,1})'\Sigma^{ps}(x)^{-1}\,\mathbb{E}[\widetilde{\mathbf{A}}_-(W,x,\gamma^*(x))] \\ =&\frac{h^{p+1}}{(p+1)!}\mathcal{I}(x)(I_3\otimes e_{N_p,1})'(\mathcal{I}(x)\otimes\mathbf{Q}_p(x_c)f_X(x))^{-1}\Big\{(\mathcal{I}(x)\otimes\mathbf{M}_{p,p+1}(x_c))\,\mathrm{vec}\left(\mathbf{g}^{(p+1)}_-(x)\right)f_X(x) \\ &+\sum_{z_d\in\mathcal{X}_d\backslash x_d}\sum_{j=o,u}\lambda_j I_j(x_d,z_d)\left(\mathbf{p}_-(x)-\mathbf{p}_-(x_c,z_d)\right)\otimes\mathbf{M}_{p,0}(x_c)f_X(x_c,z_d)\Big\} + o\left(h^{p+1}+\lambda_o+\lambda_u\right) \\ =&\frac{h^{p+1}}{(p+1)!}e'_{N_p,1}\mathbf{Q}_p(x_c)^{-1}\mathbf{M}_{p,p+1}(x_c)\mathbf{g}^{(p+1)}_-(x)\mathcal{I}(x) \\ &+\sum_{z_d\in\mathcal{X}_d\backslash x_d}\sum_{j=o,u}\frac{f_X(x_c,z_d)}{f_X(x)}\lambda_j I_j(x_d,z_d)e'_{N_p,1}\mathbf{Q}_p(x_c)^{-1}\mathbf{M}_{p,0}(x_c)\left(\mathbf{p}_-(x)-\mathbf{p}_-(x_c,z_d)\right) \\ &+o\left(h^{p+1}+\lambda_o+\lambda_u\right) \\ =&\mathcal{B}^{ps}(x,h,\lambda)+o\left(h^{p+1}+\lambda_o+\lambda_u\right), \end{aligned} \tag{C.11}$$

and

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left[\mathcal{I}(x)\mathbf{A}_-(W,x)\right] =&h^{-v_c}\mathcal{I}(x)(I_3\otimes e_{N_p,1})'\Sigma^{ps}(x)^{-1}\left(\mathcal{I}(x)\otimes\mathbf{T}_p(x_c)f_X(x)\right)\Sigma^{ps}(x)^{-1}(I_3\otimes e_{N_p,1})\mathcal{I}(x) \\ =&h^{-v_c}\mathcal{I}(x)(I_3\otimes e_{N_p,1})'(\mathcal{I}(x)\otimes\mathbf{Q}_p(x_c)f_X(x))^{-1}\left(\mathcal{I}(x)\otimes\mathbf{T}_p(x_c)f_X(x)\right) \\ &\cdot(\mathcal{I}(x)\otimes\mathbf{Q}_p(x_c)f_X(x))^{-1}(I_3\otimes e_{N_p,1})\mathcal{I}(x)+o\left(h^{-v_c}\right) \\ =&h^{-v_c}f_X(x)^{-1}\mathcal{I}(x)e'_{N_p,1}\mathbf{Q}_p(x_c)^{-1}\mathbf{T}_p(x_c)\mathbf{Q}_p(x_c)^{-1}e_{N_p,1}+o\left(h^{-v_c}\right) \\ =&\mathcal{V}^{ps}(x,h,\lambda)+o\left(h^{-v_c}\right). \end{aligned} \tag{C.12}$$

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

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

where

$$\begin{aligned}
&\mathbb{E}[G^{(or)}_{d,t}(W,x)] = e'_{N_q,1}\Sigma^{or}_{d,t}(x)^{-1}\,\mathbb{E}[H(b_{d,t})\underline{\mathbf{X}}(X_j)I_{d,t}\xi^{or}_{d,t}(x)\widetilde{K}_{or}(X;x,b_{d,t},\vartheta_{d,t})]\\
=&\frac{b^{q+1}_{d,t}}{(q+1)!}e'_{N_q,1}(\mathbf{Q}_q(x_c)f_X(x))^{-1}\Big\{\mathbf{M}_{q,q+1}(x_c)\mathbf{m}^{(q+1)}_{d,t}(x)f_X(x)\\
&+\sum_{z_d\in\mathcal{X}_d\backslash x_d}\sum_{j=o,u}\vartheta_{d,t,j}I_j(x_d,z_d)\left(m_{d,t}(x)-m_{d,t}(x_c,z_d)\right)\mathbf{M}_{q,0}(x_c)f_X(x_c,z_d)\Big\}\\
&+o\left(b^{q+1}_{d,t}+\vartheta_{d,t,o}+\vartheta_{d,t,u}\right)\\
=&\frac{b^{q+1}_{d,t}}{(q+1)!}\left(e'_{N_q,1}\mathbf{Q}_q(x_c)^{-1}\mathbf{M}_{q,q+1}(x_c)\mathbf{m}^{(q+1)}_{d,t}(x)\right)\\
&+\sum_{z_d\in\mathcal{X}_d\backslash x_d}\sum_{j=o,u}\frac{f_X(x_c,z_d)}{f_X(x)}\vartheta_jI_j(x_d,z_d)e'_{N_q,1}\mathbf{Q}_q(x_c)^{-1}\mathbf{M}_{q,0}(x_c)\left(m_{d,t}(x)-m_{d,t}(x_c,z_d)\right)\\
&+o\left(b^{q+1}_{d,t}+\vartheta_{d,t,o}+\vartheta_{d,t,u}\right),\\
=&\mathcal{B}^{or}_{d,t}(x,b_{d,t},\vartheta_{d,t})+o\left(b^{q+1}_{d,t}+\vartheta_{d,t,o}+\vartheta_{d,t,u}\right), && \text{(C.13)}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{V}\text{ar}\left[G^{(or)}_{d,t}(W,x)\right] =&b^{-v_c}_{d,t}e'_{N_q,1}\Sigma^{or}_{d,t}(x)^{-1}\,\mathbb{E}[H(b_{d,t})\underline{\mathbf{X}}(X_j)I_{d,t}(Y-m_{d,t}(X))^2\\
&+H(b_{d,t})\underline{\mathbf{X}}(X_j)'\widetilde{K}_{or}(X;x,b_{d,t},\vartheta_{d,t})^2]\Sigma^{or}_{d,t}(x)^{-1}e_{N_q,1}+o\left(b^{-v_c}\right)\\
=&b^{-v_c}_{d,t}e'_{N_q,1}(\mathbf{Q}_q(x_c)f_X(x))^{-1}\left(\sigma^2_{d,t}(x)\mathbf{T}_q(x_c)f_X(x)\right)(\mathbf{Q}_q(x_c)f_X(x))^{-1}+o\left(b^{-v_c}\right)\\
=&b^{-v_c}_{d,t}f_X(x)^{-1}\sigma^2_{d,t}(x)e'_{N_q,1}\mathbf{Q}_q(x_c)^{-1}\mathbf{T}_q(x_c)\mathbf{Q}_q(x_c)^{-1}e_{N_q,1}+o\left(b^{-v_c}\right)\\
=&\mathcal{V}^{or}_{d,t}(x,b_{d,t},\vartheta_{d,t})+o\left(b^{-v_c}\right). && \text{(C.14)}
\end{aligned}$$

Now, we define

$$(h^\dagger,\lambda^\dagger,\{b^\dagger_{d,t},\vartheta^\dagger_{d,t}\}_{(d,t)\in\mathcal{S}_-}) = (n^{1/(2p+v_c+2)}h, n^{2/(2p+v_c+2)}\lambda, \{n^{1/(2q+v_c+2)}b_{d,t}, n^{2/(2q+v_c+2)}\vartheta_{d,t}\}_{(d,t)\in\mathcal{S}_-}).$$

It follows from (C.11)-(C.14) and standard analysis that

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

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 C.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 $(h^o, \lambda^o, \{b^o_{d,t}, \vartheta^o_{d,t}\}_{(d,t)\in\mathcal{S}_-})$. ■

C.3 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 (C.10) and applying Theorem C.1, the following results are obtained

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

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

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 C.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

$$\breve{h} = \left(\frac{\mathbb{E}_n\left[\left\| \varrho^b_{p,1}\breve{\mathbf{g}}_-^{(p+1)}(X)\breve{\mathcal{I}}(X) \right\|^2 \right]}{\varrho^v_p\,\mathbb{E}_n\left[\widehat{f}_X^{-1}(X)\operatorname{tr}\left(\breve{\mathcal{I}}(X)\right) \right]} \frac{2(p+1)n}{\upsilon\{(p+1)!\}^2} \right)^{-1/(2p+\upsilon+2)},$$

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

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

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

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, $\boldsymbol{\varrho}^b_j = I'_{N_j,\mathbf{j}} \mathbf{Q}^{-1}_j \mathbf{M}_{j,j+1}$, $\boldsymbol{\varrho}^v_j = I'_{N_j,\mathbf{j}} \mathbf{Q}^{-1}_j \mathbf{T}_j \mathbf{Q}^{-1}_j I_{N_j,\mathbf{j}}$, 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 $\widehat{\mathbf{m}}^{(q+1)}_{d,t}(x)$, 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

$$\widehat{h} = \left(\frac{\mathbb{E}_n \left[\left\| \varrho^b_{p,1} \widehat{\mathbf{g}}^{(p+1)}_{-}(X) \widehat{\mathcal{I}}(X) \right\|^2 \right]}{\varrho^v_p \, \mathbb{E}_n \left[\widehat{f}^{-1}_X(X) \operatorname{tr}\left(\widehat{\mathcal{I}}(X) \right) \right]} \frac{2(p+1)n}{\upsilon\{(p+1)!\}^2} \right)^{-1/(2p+\upsilon+2)},$$

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

C.4 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 $V^n_{i=1}$ 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 C.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 $\widehat{R}_b$ in the B draws, 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 (3.11) and $\widehat{\eta}_{sz}(\cdot)$ given in (4.3), we conduct a cluster-robust Hausman-type test as follows

Algorithm C.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 1-2 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

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.

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.

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.

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

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.