Doubly Robust Difference-in-Differences Estimators: Supplemental Appendix

Pedro H. C. Sant'Anna* Jun B. Zhao†

Vanderbilt University Vanderbilt University

November 29, 2018

This supplemental appendix is organized as follows. Appendix A collects all the (detailed) proofs of the main results of the paper. Appendix B contains the closed form expression of some (generic) functions appearing in the main text. Finally, we present a small scale Monte Carlo simulation in Appendix C.

*Department of Economics, Vanderbilt University. E-mail: pedro.h.santanna@vanderbilt.edu.

†Department of Economics, Vanderbilt University. E-mail: jun.zhao@vanderbilt.edu.

A Proofs of Main Results

Proof of Theorem 1: We prove the results for panel and repeated cross section data separately.

Case 1: Pane data is available and propensity score model is correctly specified

In this case, we have that $\pi(X) = p(X)$ $a.s.$. In order to show that $\tau_{att}^{dr,p} = ATT$, first note that, by the law of iterated expectations,

$$\begin{aligned}
\mathbb{E}\left[\frac{\pi(X)(1-D)}{1-\pi(X)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{p(X)(1-D)}{1-p(X)}\middle|X\right]\right] \\
&= \mathbb{E}\left[\frac{p(X)(1-\mathbb{E}[D|X])}{1-p(X)}\right] \\
&= \mathbb{E}\left[\frac{p(X)(1-p(X))}{1-p(X)}\right] \\
&= \mathbb{E}[p(X)] \\
&= \mathbb{E}[D],
\end{aligned}$$

which yields

$$w_{cont}^{p}(D,X;\pi) = \frac{\pi(X)(1-D)}{1-\pi(X)} \Big/ \mathbb{E}\left[\frac{\pi(X)(1-D)}{(1-\pi(X)}\right] = \frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D]}.$$

Therefore, by the law of total expectation and repetitions of the law of iterated expectations, we have that

$$\begin{aligned}
&\tau_{att}^{dr,p} \\
=&\mathbb{E}\left[(w_{tr}^{p}(D) - w_{cont}^{p}(D,X;\pi))(\Delta Y - \mu_{\Delta}^{p}(X))\right] \\
=&\mathbb{E}\left[\left(\frac{D}{\mathbb{E}[D]} - \frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D]}\right)(\Delta Y - \mu_{\Delta}^{p}(X))\right] \\
=&\mathbb{E}\left[\frac{D-p(X)}{\mathbb{E}[D](1-p(X))}(\Delta Y - \mu_{\Delta}^{p}(X))\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[\frac{D-p(X)}{\mathbb{E}[D](1-p(X))}(\Delta Y - \mu_{\Delta}^{p}(X))\middle|X\right]\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[\frac{1}{\mathbb{E}[D]}(\Delta Y - \mu_{\Delta}^{p}(X))\middle|D=1,X\right]p(X)\right.
\end{aligned}$$

$$
\begin{aligned}
&+\mathbb{E}\left[\frac{-p(X)}{\mathbb{E}[D](1-p(X))}(\Delta Y-\mu_\Delta^p(X))\middle|D=0,X\right](1-p(X))\Bigg]\\
=&\mathbb{E}\left[\frac{p(X)}{\mathbb{E}[D]}\cdot\left[\mathbb{E}[\Delta Y-\mu_\Delta^p(X)|D=1,X]-\mathbb{E}\left[\frac{1-p(X)}{1-p(X)}(\Delta Y-\mu_\Delta^p(X))\middle|D=0,X\right]\right]\right]\\
=&\mathbb{E}\left[\frac{p(X)}{\mathbb{E}[D]}\cdot\left[\mathbb{E}[\Delta Y-\mu_\Delta^p(X)|D=1,X]-\mathbb{E}\left[\Delta Y-\mu_\Delta^p(X)\middle|D=0,X\right]\right]\right]\\
=&\mathbb{E}\left[\frac{p(X)}{\mathbb{E}[D]}\cdot\left[\mathbb{E}[\Delta Y|D=1,X]-\mathbb{E}\left[\Delta Y|D=0,X\right]\right]\right]\\
=&\mathbb{E}\left[\frac{p(X)}{\mathbb{E}[D]}\cdot\mathbb{E}\left[\frac{D-p(X)}{p(X)(1-p(X))}\Delta Y\middle|X\right]\right]\\
=&\mathbb{E}\left[\frac{p(X)}{\mathbb{E}[D]}\cdot\frac{D-p(X)}{p(X)(1-p(X))}\Delta Y\right]\\
=&\mathbb{E}\left[\frac{D-p(X)}{\mathbb{E}[D]\,(1-p(X))}\Delta Y\right].
\end{aligned}
$$

By Lemma 3.1 and equation (10) in Abadie (2005), i.e. equation (7) in our main text, we obtain that $\tau_{att}^{dr,p}=ATT$.

Case 2: Pane data is available and outcome regression models are correctly specified

In this case, we have that $\mu_\Delta^p(X)=m_{0,1}^p(X)-m_{0,0}^p(X)$ *a.s.*, i.e. the outcome regression models are correctly specified. To show that $\tau_{att}^{dr,p}=ATT$, recall that $m_{0,t}(x)=\mathbb{E}[Y|D=0,T=t,X=x]$, for $t=0,1$. Therefore, again by the law of total expectation and repetitions of the law of iterated expectations, we have that

$$
\begin{aligned}
&\tau_{att}^{dr,p}\\
=&\mathbb{E}[(w_{tr}^p(D)-w_{cont}^p(D,X;\pi))(\Delta Y-\mu_\Delta^p(X))]\\
=&\mathbb{E}[(w_{tr}^p(D)-w_{cont}^p(D,X;\pi))(\Delta Y-m_{0,1}^p(X)+m_{0,0}^p(X))]\\
=&\mathbb{E}\Big[p(X)\cdot\mathbb{E}[(w_{tr}^p(1)-w_{cont}^p(1,X;\pi))(\Delta Y-m_{0,1}^p(X)+m_{0,0}^p(X))|D=1,X]\\
&+(1-p(X))\cdot\mathbb{E}[(w_{tr}^p(0)-w_{cont}^p(0,X;\pi))(\Delta Y-m_{0,1}^p(X)+m_{0,0}^p(X))|D=0,X]\Big]\\
=&\mathbb{E}\Big[p(X)\cdot\mathbb{E}[(w_{tr}^p(1)-w_{cont}^p(1,X;\pi))(\Delta Y-m_{0,1}^p(X)+m_{0,0}^p(X))|D=1,X]\\
&+(1-p(X))(w_{tr}^p(0)-w_{cont}^p(0,X;\pi))\cdot\left[\mathbb{E}[\Delta Y|D=0,X]-\left(m_{0,1}^p(X)-m_{0,0}^p(X)\right)\right]\Big]
\end{aligned}
$$

$$=\mathbb{E}\left[\frac{p(X)}{\mathbb{E}[D]}\cdot\mathbb{E}[\Delta Y-\left(m^p_{0,1}(X)-m^p_{0,0}(X)\right)|D=1,X]\right]$$

$$=\mathbb{E}\left[\frac{p(X)}{\mathbb{E}[D]}\cdot\mathbb{E}\left[\Delta Y-\mathbb{E}[\Delta Y|D=0,X]|D=1,X\right]\right]$$

$$=\mathbb{E}\left[\frac{p(X)}{\mathbb{E}[D]}\cdot\left(\mathbb{E}[\Delta Y|D=1,X]-\mathbb{E}[\Delta Y|D=0,X]\right)\right]$$

$$=\mathbb{E}\left[\mathbb{E}[\Delta Y|D=1,X]-\mathbb{E}[\Delta Y|D=0,X]|D=1\right],$$

where the fifth equality holds because $w^p_{cont}(1,X;\pi)=0$ *a.s.*, and $\left|w^p_{tr}(0)-w^p_{cont}(0,X;\pi)\right|=\pi(X)/(1-\pi(X))\big/\mathbb{E}\left[\pi(X)/(1-\pi(X))\right]<\infty$ almost surely. The result that $\tau^{dr,p}_{att}=ATT$ now follows from the conditional parallel trends assumption, i.e. Assumption 2 in the main text.

Case 3: Repeated cross section data and propensity score model is correctly specified

In this case, we have that $\pi(X)=p(X)$ *a.s.*. First of all, recall that $\lambda=\mathbb{P}(T=1)=\mathbb{E}[T]$, and notice that, by the law of iterated expectations and stationarity of (D,X),

$$\mathbb{E}\left[\frac{\pi(X)(1-D)T}{1-\pi(X)}\right]=\mathbb{E}\left[\frac{p(X)(1-D)}{1-p(X)}\middle|T=1\right]\lambda$$

$$=\mathbb{E}[p(X)]\lambda$$

$$=\mathbb{E}[D]\lambda.$$

Similarly, we have that

$$\mathbb{E}\left[\frac{\pi(X)(1-D)(1-T)}{1-\pi(X)}\right]=\mathbb{E}[D](1-\lambda),$$

together with $\mathbb{E}[DT]=\mathbb{E}[D]\lambda$ and $\mathbb{E}[D(1-T)]=\mathbb{E}[D](1-\lambda)$. Thus, when the propensity score is correctly specified, we can write the four weights in $\tau^{dr,rc}_{att}$ as

$$w^{rc}_{tr,1}(D,T)=\frac{DT}{\mathbb{E}[D]\lambda},$$

$$w^{rc}_{tr,0}(D,T)=\frac{D(1-T)}{\mathbb{E}[D](1-\lambda)},$$

$$w^{rc}_{cont,1}(D,T,X;\pi)=\frac{\pi(X)(1-D)T}{(1-\pi(X))\mathbb{E}[D]\lambda},$$

$$w_{cont,0}^{rc}(D,T,X;\pi)=\frac{\pi(X)(1-D)(1-T)}{(1-\pi(X))\mathbb{E}[D](1-\lambda)}.$$

Next, by the law of total expectation and repetitions of the law of iterated expectations, we have that

$$\begin{aligned}
&\tau_{att}^{dr,rc}\\
=&\mathbb{E}[(w_{tr,1}^{rc}(D,T)-w_{tr,0}^{rc}(D,T)-w_{cont,1}^{rc}(D,T,X;\pi)+w_{cont,0}^{rc}(D,T,X;\pi))(Y-\mu_Y^{rc}(X,T))]\\
=&\mathbb{E}\left[\left(\frac{DT}{\mathbb{E}[D]\lambda}-\frac{D(1-T)}{\mathbb{E}[D](1-\lambda)}-\frac{\pi(X)(1-D)T}{(1-\pi(X))\mathbb{E}[D]\lambda}+\frac{\pi(X)(1-D)(1-T)}{(1-\pi(X))\mathbb{E}[D](1-\lambda)}\right)(Y-\mu_Y^{rc}(X,T))\right]\\
=&\mathbb{E}\left[\left(\frac{D}{\mathbb{E}[D]\lambda}-\frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D]\lambda}\right)(Y-\mu_Y^{rc}(X,1)\middle|T=1\right]\lambda\\
&+\mathbb{E}\left[\left(-\frac{D}{\mathbb{E}[D](1-\lambda)}+\frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D](1-\lambda)}\right)(Y-\mu_Y^{rc}(X,0))\middle|T=0\right](1-\lambda)\\
=&\mathbb{E}\left[\frac{T}{\lambda}\left(\frac{D}{\mathbb{E}[D]}-\frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D]}\right)(Y-\mu_Y^{rc}(X,T))\right]\\
&+\mathbb{E}\left[\frac{1-T}{1-\lambda}\left(-\frac{D}{\mathbb{E}[D]}+\frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D]}\right)(Y-\mu_Y^{rc}(X,T))\right]\\
=&\mathbb{E}\left[\frac{T-\lambda}{\lambda(1-\lambda)}\left(\frac{D}{\mathbb{E}[D]}-\frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D]}\right)(Y-\mu_Y^{rc}(X,T))\right],\\
=&\mathbb{E}\left[\frac{T-\lambda}{\lambda(1-\lambda)}\left(\frac{D}{\mathbb{E}[D]}-\frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D]}\right)Y\right]\\
&-\mathbb{E}\left[\frac{T-\lambda}{\lambda(1-\lambda)}\left(\frac{D}{\mathbb{E}[D]}-\frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D]}\right)\mu_Y^{rc}(X,T)\right]\\
=&\mathbb{E}\left[\frac{T-\lambda}{\lambda(1-\lambda)}\left(\frac{D}{\mathbb{E}[D]}-\frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D]}\right)Y\right]\\
&-\mathbb{E}\left[\frac{T-\lambda}{\lambda(1-\lambda)}\mu_Y^{rc}(X,T)\mathbb{E}\left[\frac{D}{\mathbb{E}[D]}-\frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D]}\middle|X,T\right]\right]\\
=&\mathbb{E}\left[\frac{T-\lambda}{\lambda(1-\lambda)}\left(\frac{D}{\mathbb{E}[D]}-\frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D]}\right)Y\right]\\
&-\mathbb{E}\left[\frac{T-\lambda}{\lambda(1-\lambda)}\mu_Y^{rc}(X,T)\left(\frac{p(X)}{\mathbb{E}[D]}-\frac{p(X)}{\mathbb{E}[D]}\right)\right]\\
=&\mathbb{E}\left[\frac{T-\lambda}{\lambda(1-\lambda)}\left(\frac{D}{\mathbb{E}[D]}-\frac{p(X)(1-D)}{(1-p(X))\mathbb{E}[D]}\right)Y\right]
\end{aligned}$$

Then, by Lemma 3.2 and equation (12) in Abadie (2005), i.e. equation (8) in our main text, we obtain that $\tau_{att}^{dr,p} = ATT$.

Case 4: Repeated cross section data and outcome regression models are correctly specified

We now consider the case where the outcome models are correctly specified, i.e., $\mu_Y^{rc}(X,T) = T \cdot m_{0,1}^{rc}(X) + (1-T) \cdot m_{0,0}^{rc}(X)$ $a.s.$, where $m_{d,t}^{rc}(x) = \mathbb{E}[Y|D=d,T=t,X=x]$ for $t=0,1$.

By the law of total expectation and repetitions of the law of iterated expectations, we have that

$$
\begin{aligned}
&\tau_{att}^{dr,rc} \\
=&\mathbb{E}[\left(w_{tr,1}^{rc}(D,T) - w_{tr,0}^{rc}(D,T) - w_{cont,1}^{rc}(D,T,X;\pi) + w_{cont,0}^{rc}(D,T,X;\pi)\right)(Y - \mu_Y^{rc}(X,T))] \\
=&\mathbb{E}[(w_{tr,1}^{rc}(D,1) - w_{cont,1}^{rc}(D,1,X;\pi))(Y - \mu_Y^{rc}(X,1))|T=1]\lambda \\
&-\mathbb{E}[(w_{tr,0}^{rc}(D,0) - w_{cont,0}^{rc}(D,0,X;\pi))(Y - \mu_Y^{rc}(X,0))|T=0](1-\lambda). \qquad \text{(A.1)}
\end{aligned}
$$

Next, note that

$$
\begin{aligned}
&\mathbb{E}\left[\left(w_{tr,1}^{rc}(D,1) - w_{cont,1}^{rc}(D,1,X;\pi)\right)\left(Y - m_{0,1}^{rc}(X)\right)\middle|T=1\right]\lambda \\
&= \mathbb{E}\left[p(X) \cdot \mathbb{E}\left[\left(w_{tr,1}^{rc}(D,1) - w_{cont,1}^{rc}(D,1,X;\pi)\right)\left(Y - m_{0,1}^{rc}(X)\right)\middle|D=1,T=1,X\right]\middle|T=1\right]\lambda \\
&+\mathbb{E}\left[(1-p(X)) \cdot \mathbb{E}\left[\left(w_{tr,1}^{rc}(D,1) - w_{cont,1}^{rc}(D,1,X;\pi)\right)\left(Y - m_{0,1}^{rc}(X)\right)\middle|D=0,T=1,X\right]\middle|T=1\right]\lambda \\
&= \mathbb{E}\left[\frac{p(X)}{\mathbb{E}[D]} \cdot \left(\mathbb{E}[Y|D=1,T=1,X] - m_{0,1}^{rc}(X)\right)\middle|T=1\right] \\
&- \mathbb{E}\left[(1-p(X)) \cdot \frac{\dfrac{\pi(X)}{1-\pi(X)}}{\mathbb{E}\left[\dfrac{\pi(X)(1-D)}{1-\pi(X)}\right]}\left(\mathbb{E}[Y|D=0,T=1,X] - m_{0,1}^{rc}(X)\right)\middle|T=1\right] \\
&= \mathbb{E}\left[\frac{T}{\lambda}\frac{p(X)}{\mathbb{E}[D]} \cdot \left(\mathbb{E}[Y|D=1,T=1,X] - m_{0,1}^{rc}(X)\right)\right] \\
&- \mathbb{E}\left[\frac{T}{\lambda}(1-p(X)) \cdot \frac{\dfrac{\pi(X)}{1-\pi(X)}}{\mathbb{E}\left[\dfrac{\pi(X)(1-D)}{1-\pi(X)}\right]}\left(\mathbb{E}[Y|D=0,T=1,X] - m_{0,1}^{rc}(X)\right)\right] \\
&= \mathbb{E}\left[\frac{T}{\lambda}\frac{p(X)}{\mathbb{E}[D]} \cdot \left(m_{1,1}^{rc}(X) - m_{0,1}^{rc}(X)\right)\right]
\end{aligned}
$$

$$
-\mathbb{E}\left[\frac{T}{\lambda}(1-p(X))\cdot\frac{\dfrac{\pi(X)}{1-\pi(X)}}{\mathbb{E}\left[\dfrac{\pi(X)(1-D)}{1-\pi(X)}\right]}\left(m^{rc}_{0,1}(X)-m^{rc}_{0,1}(X)\right)\right]
$$

$$
=\mathbb{E}\left[\frac{T}{\lambda}\frac{p(X)}{\mathbb{E}[D]}\cdot\left(m^{rc}_{1,1}(X)-m^{rc}_{0,1}(X)\right)\right] \tag{A.2}
$$

Analogously,

$$
\begin{aligned}
&\mathbb{E}[\left(w^{rc}_{tr,0}(D,0)-w^{rc}_{cont,0}(D,0,X;\pi)\right)(Y-\mu^{rc}_{Y}(X,0))\big|T=0](1-\lambda)\\
&=\mathbb{E}\left[\frac{1-T}{1-\lambda}\cdot\frac{p(X)}{\mathbb{E}[D]}\cdot\left(m^{rc}_{1,0}(X)-m^{rc}_{0,0}(X)\right)\right]\\
&-\mathbb{E}\left[\frac{1-T}{1-\lambda}\cdot(1-p(X))\cdot\frac{\dfrac{\pi(X)}{1-\pi(X)}}{\mathbb{E}\left[\dfrac{\pi(X)(1-D)}{1-\pi(X)}\right]}\left(m^{rc}_{0,0}(X)-m^{rc}_{0,0}(X)\right)\right]\\
&=\mathbb{E}\left[\frac{1-T}{1-\lambda}\cdot\frac{p(X)}{\mathbb{E}[D]}\cdot\left(m^{rc}_{1,0}(X)-m^{rc}_{0,0}(X)\right)\right].
\end{aligned} \tag{A.3}
$$

Plugging (A.2) and (A.3) into (A.1), we have

$$
\begin{aligned}
\tau^{dr,rc}_{att}&=\mathbb{E}\left[\frac{T}{\lambda}\frac{p(X)}{\mathbb{E}[D]}\cdot\left(m^{rc}_{1,1}(X)-m^{rc}_{0,1}(X)\right)\right]\\
&-\mathbb{E}\left[\frac{1-T}{1-\lambda}\cdot\frac{p(X)}{\mathbb{E}[D]}\cdot\left(m^{rc}_{1,0}(X)-m^{rc}_{0,0}(X)\right)\right]\\
&=\mathbb{E}\left[m^{rc}_{1,1}(X)-m^{rc}_{0,1}(X)\big|T=1,D=1\right]\\
&-\mathbb{E}\left[m^{rc}_{1,0}(X)-m^{rc}_{0,0}(X)\big|T=0,D=1\right]\\
&=\mathbb{E}\left[\left(m^{rc}_{1,1}(X)-m^{rc}_{1,0}(X)\right)-\left(m^{rc}_{0,1}(X)-m^{rc}_{0,0}(X)\right)\big|D=1\right].
\end{aligned}
$$

where the last equality follows from the stationarity of (D,X). The result now follows from the conditional parallel trends assumption, i.e. Assumption 2 in the main text. ■

Proof of Theorem 2: Under Assumptions 1-5, we prove the large sample properties of our proposed DR estimator, $\widehat{\tau}^{dr,p}_{att}$, when panel data are available.

First of all, recall that the estimator takes the following form:

$$
\widehat{\tau}^{dr,p}_{att}=\mathbb{E}_n\left[\left(\widehat{w}^{p}_{tr}(D)-\widehat{w}^{p}_{cont}(D,X;\widehat{\gamma})\right)\left(\Delta Y-\mu^{p}_{\Delta}(X;\widehat{\beta}^{p}_{0},\widehat{\beta}^{p}_{1})\right)\right],
$$

where

$$\widehat{w}^p_{tr}(D) = \frac{D}{\mathbb{E}_n[D]},$$

$$\widehat{w}^p_{cont}(D,X;\gamma) = \frac{\pi(X;\gamma)(1-D)}{1-\pi(X;\gamma)} \Bigg/ \mathbb{E}_n\left[\frac{\pi(X;\gamma)(1-D)}{1-\pi(X;\gamma)}\right];$$

and where $\widehat{\gamma}$, $\widehat{\beta}^p_0$, and $\widehat{\beta}^p_1$ are estimators for pseudo-true γ_0, $\beta^p_{0,0}$, and $\beta^p_{1,0}$, and for generic β^p_0 and β^p_1, $\mu^p_\Delta(\cdot;\beta^p_0,\beta^p_1) = \mu^p_1(\cdot;\beta^p_1) - \mu^p_0(\cdot;\beta^p_0)$.

By weak law of large numbers and continuous mapping theorem, as $n \to \infty$,

$$\mathbb{E}_n[D] \xrightarrow{p} \mathbb{E}[D],$$

$$\mathbb{E}_n\left[\frac{\pi(X;\hat{\gamma})(1-D)}{1-\pi(X;\hat{\gamma})}\right] \xrightarrow{p} \mathbb{E}\left[\frac{\pi(X;\gamma_0)(1-D)}{1-\pi(X;\gamma_0)}\right],$$

and

$$\hat{\tau}^{dr,p}_{att} \xrightarrow{p} \mathbb{E}_n\left[\left(w^p_{tr}(D) - w^p_{cont}(D,X;\gamma_0)\right)\left(\Delta Y - \mu^p_\Delta(X;\beta^p_{0,0},\beta^p_{1,0})\right)\right],$$

where

$$w^p_{tr}(D) = \frac{D}{\mathbb{E}[D]},$$

$$w^p_{cont}(D,X;\gamma_0) = \frac{\pi(X;\gamma_0)(1-D)}{1-\pi(X;\gamma_0)} \Bigg/ \mathbb{E}\left[\frac{\pi(X;\gamma_0)(1-D)}{1-\pi(X;\gamma_0)}\right].$$

Therefore, if either $\pi(X;\gamma_0) = p(X)$ a.s. or $\mu^p_\Delta(X;\beta^p_{0,0},\beta^p_{1,0}) = m^p_{0,1}(X) - m^p_{0,0}(X)$ *a.s.*, from Theorem 1, we have that

$$\mathbb{E}\left[\left(w^p_{tr}(D) - w^p_{cont}(D,X;\gamma_0)\right)\left(\Delta Y - \mu^p_\Delta(X;\beta^p_{0,0},\beta^p_{1,0})\right)\right] \equiv \tau^{dr,p}_{att} = ATT,$$

which completes the consistency proof.

Next, we derive the asymptotically linear representation of $\widehat{\tau}^{dr,p}_{att}$, which leads to the asymptotic distribution result. Towards this end, first notice that

$$\begin{aligned}
&\widehat{\tau}^{dr,p}_{att} - \tau^{dr,p}_{att} \\
&= \left(\mathbb{E}_n[\widehat{w}^p_{tr}(D)\Delta Y] - \mathbb{E}[w^p_{tr}(D)\Delta Y]\right) \\
&\quad - \left(\mathbb{E}_n[\widehat{w}^p_{cont}(D,X;\widehat{\gamma})\Delta Y] - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right)
\end{aligned}$$

$$-\left(\mathbb{E}_n[\widehat{w}^p_{tr}(D)\mu^p_\Delta(X;\widehat{\beta}^p_0,\widehat{\beta}^p_1)] - \mathbb{E}[w^p_{tr}(D)\mu^p_\Delta(X;\beta^p_{0,0},\beta^p_{1,0})]\right)$$

$$+\left(\mathbb{E}_n[\widehat{w}^p_{cont}(D,X;\widehat{\gamma})\mu^p_\Delta(X;\widehat{\beta}^p_0,\widehat{\beta}^p_1)] - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\mu^p_\Delta(X;\beta^p_{0,0},\beta^p_{1,0}]\right)$$

$$\equiv\left(\widehat{ATT}_1 - ATT_1\right) - \left(\widehat{ATT}_2 - ATT_2\right) - \left(\widehat{ATT}_3 - ATT_3\right) + \left(\widehat{ATT}_4 - ATT_4\right).$$

Therefore,

$$\begin{aligned}&\sqrt{n}(\hat{\tau}^{dr,p}_{att} - \tau^{dr,p}_{att})\\ =&\sqrt{n}\left(\widehat{ATT}_1 - ATT_1\right) - \sqrt{n}\left(\widehat{ATT}_2 - ATT_2\right) - \sqrt{n}\left(\widehat{ATT}_3 - ATT_3\right) + \sqrt{n}\left(\widehat{ATT}_4 - ATT_4\right),\end{aligned} \tag{A.4}$$

and we next obtain the asymptotically linear representation for each component in the above decomposition.

We first analyze $\sqrt{n}\left(\widehat{ATT}_1 - ATT_1\right)$. Note that by classical central limit theorem, it follows that

$$\begin{aligned}\mathbb{E}_n[D] - \mathbb{E}[D] &= O_p(n^{-1/2}),\\ \mathbb{E}_n\left[\frac{D}{E[D]^2}\Delta Y\right] - \mathbb{E}\left[\frac{D}{E[D]^2}\Delta Y\right] &= O_p(n^{-1/2}) = o_p(1),\end{aligned} \tag{A.5}$$

which yields that

$$(\mathbb{E}_n[D] - \mathbb{E}[D])^2 = O_p(n^{-1}) = o_p(n^{-1/2}).$$

Next, by a second-order Taylor expansion of $\widehat{ATT}_1$ around $\mathbb{E}[D]$, we have that

$$\begin{aligned}\widehat{ATT}_1 &= \mathbb{E}_n\left[\frac{D}{\mathbb{E}_n[D]}\Delta Y\right]\\ &= \mathbb{E}_n\left[\frac{D}{\mathbb{E}[D]}\Delta Y\right] - (\mathbb{E}_n[D] - \mathbb{E}[D])\mathbb{E}_n\left[\frac{D}{\mathbb{E}[D]^2}\Delta Y\right] + O_p((\mathbb{E}_n[D] - \mathbb{E}[D])^2)\\ &= \mathbb{E}_n\left[\frac{D}{\mathbb{E}[D]}\Delta Y\right] - (\mathbb{E}_n[D] - \mathbb{E}[D])\mathbb{E}_n\left[\frac{D}{\mathbb{E}[D]^2}\Delta Y\right] + o_p(n^{-1/2}),\end{aligned}$$

which leads to

$$\sqrt{n}\left(\widehat{ATT}_1 - ATT_1\right)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{D_i}{\mathbb{E}[D]}\Delta Y_i - (D_i - \mathbb{E}[D])\mathbb{E}_n\left[\frac{D}{\mathbb{E}[D]^2}\Delta Y\right] - \mathbb{E}\left[\frac{D}{\mathbb{E}[D]}\Delta Y\right]\right) + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{D_i}{\mathbb{E}[D]}\Delta Y_i - (D_i - \mathbb{E}[D])\mathbb{E}\left[\frac{D}{\mathbb{E}[D]^2}\Delta Y\right] - \mathbb{E}\left[\frac{D}{\mathbb{E}[D]}\Delta Y\right]\right) + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{D_i}{\mathbb{E}[D]}\Delta Y_i - \frac{D_i}{\mathbb{E}[D]}\mathbb{E}\left[\frac{D}{\mathbb{E}[D]}\Delta Y\right]\right) + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(w_{tr}^{p}(D_i)\Delta Y_i - w_{tr}^{p}(D_i)\mathbb{E}[w_{tr}^{p}(D)\Delta Y]\right) + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{tr}^{p}(D_i)\left(\Delta Y_i - \mathbb{E}[w_{tr}^{p}(D)\Delta Y]\right) + o_p(1), \tag{A.6}$$

where the second equality follows from (A.5).

We next analyze $\sqrt{n}\left(\widehat{ATT}_2 - ATT_2\right)$. Following the same arguments as above, we have that

$$\sqrt{n}\left(\widehat{ATT}_2 - ATT_2\right)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\widehat{w}_{cont}^{p}(D_i,X_i;\widehat{\gamma})\Delta Y_i - \mathbb{E}[w_{cont}^{p}(D,X;\gamma_0)\Delta Y]\right)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\widetilde{w}_{cont}^{p}(D_i,X_i;\widehat{\gamma})\Delta Y_i - \mathbb{E}[w_{cont}^{p}(D,X;\gamma_0)\Delta Y]\right)$$

$$- \sqrt{n}\left(\mathbb{E}_n\left[\frac{\pi(X;\widehat{\gamma})(1-D)}{1-\pi(X;\widehat{\gamma})}\right] - \mathbb{E}\left[\frac{\pi(X;\gamma_0)(1-D)}{1-\pi(X;\gamma_0)}\right]\right)\cdot\frac{\mathbb{E}\left[\frac{\pi(X;\gamma_0)(1-D)}{1-\pi(X;\gamma_0)}\Delta Y\right]}{\mathbb{E}\left[\frac{\pi(X;\gamma_0)(1-D)}{1-\pi(X;\gamma_0)}\right]^2} + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\widetilde{w}_{cont}^{p}(D_i,X_i;\widehat{\gamma})\Delta Y_i - \mathbb{E}[w_{cont}^{p}(D,X;\gamma_0)\Delta Y]\right)$$

$$- \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\left(\widetilde{w}_{cont}^{p}(D_i,X_i;\widehat{\gamma}) - 1\right)\mathbb{E}[w_{cont}^{p}(D,X;\gamma_0)\Delta Y]\right) + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{w}_{cont}^{p}(D_i,X_i;\widehat{\gamma})\left(\Delta Y_i - \mathbb{E}[w_{cont}^{p}(D,X;\gamma_0)\Delta Y]\right) + o_p(1), \tag{A.7}$$

where

$$\widetilde{w}_{cont}^{p}(D,X;\widehat{\gamma}) = \frac{\pi(X;\widehat{\gamma})(1-D)}{1-\pi(X;\widehat{\gamma})}\Bigg/\mathbb{E}\left[\frac{\pi(X;\gamma_0)(1-D)}{1-\pi(X;\gamma_0)}\right].$$

Then, by doing a second-order Taylor expansion of the above expression around pseudo-true γ_0, we have that

$$\sqrt{n}\left(\widehat{ATT}_2 - ATT_2\right)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w^p_{cont}(D_i,X_i;\gamma_0)\left(\Delta Y_i - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right)$$
$$+(\widehat{\gamma}-\gamma_0)'\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \dot{w}^p_{cont}(D_i,X_i;\gamma_0)\left(\Delta Y_i - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right)+o_p(1)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w^p_{cont}(D_i,X_i;\gamma_0)\left(\Delta Y_i - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right)$$
$$+\sqrt{n}(\widehat{\gamma}-\gamma_0)'\cdot\frac{1}{n}\sum_{i=1}^{n} \dot{w}^p_{cont}(D_i,X_i;\gamma_0)\left(\Delta Y_i - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right)+o_p(1)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w^p_{cont}(D_i,X_i;\gamma_0)\left(\Delta Y_i - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_{ps}(W_i;\gamma_0)'\cdot\mathbb{E}\left[\dot{w}^p_{cont}(D,X;\gamma_0)\left(\Delta Y - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right)\right]+o_p(1)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w^p_{cont}(D_i,X_i;\gamma_0)\left(\Delta Y_i - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_{ps}(W_i;\gamma_0)'\cdot\mathbb{E}\left[\alpha^p_{ps}(D,X;\gamma_0)\left(\Delta Y - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right)\dot{\pi}(X;\gamma_0)\right]+o_p(1)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(w^p_{cont}(D_i,X_i;\gamma_0)\left(\Delta Y_i - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right)$$
$$+l_{ps}(W_i;\gamma_0)'\cdot\mathbb{E}\left[\alpha^p_{ps}(D,X;\gamma_0)\left(\Delta Y - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right)\dot{\pi}(X;\gamma_0)\right]\Big)+o_p(1), \tag{A.8}$$

where the third equality follows from Assumption 4, and the fourth equality holds because the derivative of $w^p_{cont}(D,X;\gamma)$ with respect to γ, denoted by $\dot{w}^p_{cont}(D,X;\gamma)$, can be written as

$$\dot{w}^p_{cont}(D,X;\gamma) = \alpha^r_{ps}(D,X;\gamma)\dot{\pi}(X;\gamma),$$

where

$$\alpha^p_{ps}(D,X;\gamma) = \frac{(1-D)}{(1-\pi(X;\gamma))^2}\Bigg/\,\mathbb{E}\left[\frac{\pi(X;\gamma)(1-D)}{1-\pi(X;\gamma)}\right].$$

For $\sqrt{n}\left(\widehat{ATT}_3 - ATT_3\right)$, we have that, by using similar arguments as in (A.6), it follows that

$$\sqrt{n}\left(\widehat{ATT}_3 - ATT_3\right) = \frac{1}{n}\sum_{i=1}^{n} w_{tr}^{p}(D_i)\left(\mu_{\Delta}^{p}(X_i;\widehat{\beta}_0^{p},\widehat{\beta}_1^{p}) - \mathbb{E}[w_{tr}^{p}(D)\mu_{\Delta}^{p}(X;\beta_{0,0}^{p},\beta_{1,0}^{p})]\right) + o_p(1).$$

Denote $\widehat{\beta} \equiv (\widehat{\beta}_1^{p},\widehat{\beta}_0^{p})$. Then, from a second-order Taylor expansion of the above expression around $\beta_0 \equiv (\beta_{1,0}^{p},\beta_{0,0}^{p})$, we have that

$$\begin{aligned}
\sqrt{n}\left(\widehat{ATT}_3 - ATT_3\right) =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{tr}^{p}(D_i)\left(\mu_{\Delta}^{p}(X_i;\beta_0) - \mathbb{E}[w_{tr}^{p}(D)\mu_{\Delta}^{p}(X;\beta_0)]\right) \\
&+ \sqrt{n}\left(\widehat{\beta}-\beta_0\right)' \cdot \mathbb{E}_n[w_{tr}^{p}(D)\dot{\mu}_{\Delta}^{p}(X;\beta_0)] + o_p(1) \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{tr}^{p}(D_i)\left(\mu_{\Delta}^{p}(X_i;\beta_0) - \mathbb{E}[w_{tr}^{p}(D)\mu_{\Delta}^{p}(X;\beta_0)]\right) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_{reg}(W_i;\beta_0)' \cdot \mathbb{E}[w_{tr}^{p}(D)\dot{\mu}_{\Delta}^{p}(X;\beta_0)] + o_p(1) \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left(w_{tr}^{p}(D_i)\left(\mu_{\Delta}^{p}(X_i;\beta_0) - \mathbb{E}[w_{tr}^{p}(D)\mu_{\Delta}^{p}(X;\beta_0)]\right)\right. \\
&\left. + l_{reg}(W_i;\beta_0)' \cdot \mathbb{E}[w_{tr}^{p}(D)\dot{\mu}_{\Delta}^{p}(X;\beta_0)]\right) \} + o_p(1), \qquad \text{(A.9)}
\end{aligned}$$

where the second equality holds because of Assumption 4.

Finally, for $\sqrt{n}\left(\widehat{ATT}_4 - ATT_4\right)$, using the same arguments as in (A.7)-(A.9), we have that

$$\begin{aligned}
&\sqrt{n}\left(\widehat{ATT}_4 - ATT_4\right) \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \widetilde{w}_{cont}^{p}(D_i,X_i;\widehat{\gamma})\left(\mu_{\Delta}^{p}(X_i;\widehat{\beta}) - \mathbb{E}[w_{cont}^{p}(D,X;\gamma_0)\mu_{\Delta}^{p}(X;\beta_0)]\right) + o_p(1) \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{cont}^{p}(D_i,X_i;\gamma_0)\left(\mu_{\Delta}^{p}(X_i;\beta_0) - \mathbb{E}[w_{cont}^{p}(D,X;\gamma_0)\mu_{\Delta}^{p}(X;\beta_0)]\right) \\
&+ \sqrt{n}(\widehat{\gamma}-\gamma_0)' \cdot \mathbb{E}\left[\alpha_{ps}^{p}(D,X;\gamma_0)\left(\mu_{\Delta}^{p}(X;\beta_0) - \mathbb{E}[w_{cont}^{p}(D,X;\gamma_0)\mu_{\Delta}^{p}(X;\beta_0)]\right)\dot{\pi}(X;\gamma_0)\right] \\
&+ \sqrt{n}\left(\widehat{\beta}-\beta_0\right)' \cdot \mathbb{E}[w_{cont}^{p}(D,X;\gamma_0)\dot{\mu}_{\Delta}^{p}(X;\beta_0)] + o_p(1) \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left(w_{cont}^{p}(D_i,X_i;\gamma_0)\left(\mu_{\Delta}^{p}(X_i;\beta_0) - \mathbb{E}[w_{cont}^{p}(D,X;\gamma_0)\mu_{\Delta}^{p}(X;\beta_0)]\right)\right. \\
&+ l_{ps}(W_i;\gamma_0)' \cdot \mathbb{E}\left[\alpha_{ps}^{p}(D,X;\gamma_0)\left(\mu_{\Delta}^{p}(X;\beta_0) - \mathbb{E}[w_{cont}^{p}(D,X;\gamma_0)\mu_{\Delta}^{p}(X;\beta_0)]\right)\dot{\pi}(X;\gamma_0)\right]
\end{aligned}$$

$$+l_{reg}(W_i;\beta_0)' \cdot \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\dot{\mu}^p_\Delta(X;\beta_0)]\big) + o_p(1). \tag{A.10}$$

Taking (A.6)-(A.10) together, we conclude that

$$\sqrt{n}(\widehat{\tau}^{dr,p}_{att} - \tau^{dr,p}_{att})$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(w^p_{tr}(D_i) \left(\Delta Y_i - \mathbb{E}[w^p_{tr}(D)\Delta Y]\right)$$
$$- w^p_{cont}(D_i,X_i;\gamma_0) \left(\Delta Y_i - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right)$$
$$- l_{ps}(W_i;\gamma_0)' \cdot \mathbb{E}\left[\alpha^p_{ps}(D,X;\gamma_0) \left(\Delta Y - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\Delta Y]\right) \dot{\pi}(X;\gamma_0)\right]$$
$$- w^p_{tr}(D_i) \left(\mu^p_\Delta(X_i;\beta_0) - \mathbb{E}[w^p_{tr}(D)\mu^p_\Delta(X;\beta^p_{0,0},\beta^p_{1,0})]\right)$$
$$- l_{reg}(W_i;\beta_0)' \cdot \mathbb{E}[w^p_{tr}(D)\dot{\mu}^p_\Delta(X;\beta_0)]$$
$$+ w^p_{cont}(D_i,X_i;\gamma_0) \left(\mu^p_\Delta(X_i;\beta_0) - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\mu^p_\Delta(X;\beta_0)]\right)$$
$$+ l_{ps}(W_i;\gamma_0)' \cdot \mathbb{E}\left[\alpha^p_{ps}(D,X;\gamma_0) \left(\mu^p_\Delta(X;\beta_0) - \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\mu^p_\Delta(X;\beta_0)]\right) \dot{\pi}(X;\gamma_0)\right]$$
$$+ l_{reg}(W_i;\beta_0)' \cdot \mathbb{E}[w^p_{cont}(D,X;\gamma_0)\dot{\mu}^p_\Delta(X;\beta_0)]\big) + o_p(1).$$

After rearrangement and dropping the dependence of the functionals on W within $\mathbb{E}[\cdot]$, we have that

$$\sqrt{n}(\widehat{\tau}^{dr,p}_{att} - \tau^{dr,p}_{att})$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(w^p_{tr}(D_i) \left(\Delta Y_i - \mu^p_\Delta(X_i;\beta_0) - \mathbb{E}[w^p_{tr} \left(\Delta Y - \mu^p_\Delta(\beta_0)\right)]\right)$$
$$- w^p_{cont}(D_i,X_i;\gamma_0) \left(\Delta Y_i - \mu^p_\Delta(X_i;\beta_0) - \mathbb{E}[w^p_{cont}(\gamma_0) \left(\Delta Y - \mu^p_\Delta(\beta_0)\right)]\right)$$
$$- l_{reg}(W_i;\beta_0)' \cdot \mathbb{E}[\left(w^p_{tr} - w^p_{cont}(\gamma_0)\right) \dot{\mu}^p_\Delta(\beta_0)]$$
$$- l_{ps}(W_i;\gamma_0)' \cdot \mathbb{E}\left[\alpha^p_{ps}(\gamma_0) \left(\Delta Y - \mu^p_\Delta(\beta_0) - \mathbb{E}[w^p_{cont}(\gamma_0) \left(\Delta Y - \mu^p_\Delta(\beta_0)\right)]\right) \dot{\pi}(\gamma_0)\right] + o_p(1)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta^p(W_i;\gamma_0,\beta_0) + o_p(1), \tag{A.11}$$

since

$$\eta^p(W_i;\gamma_0,\beta_0) \equiv \left(w^p_{tr}(D_i)\left(\Delta Y_i - \mu^p_\Delta(X_i;\beta_0) - \mathbb{E}[w^p_{tr}\left(\Delta Y - \mu^p_\Delta(\beta_0)\right)]\right)\right.$$

$$- w^p_{cont}(D_i,X_i;\gamma_0)\left(\Delta Y_i - \mu^p_\Delta(X_i;\beta_0) - \mathbb{E}[w^p_{cont}(\gamma_0)\left(\Delta Y - \mu^p_\Delta(\beta_0)\right)]\right)$$

$$- l_{reg}(W_i;\beta_0)' \cdot \mathbb{E}[\left(w^p_{tr} - w^p_{cont}(\gamma_0)\right)\dot{\mu}^p_\Delta(\beta_0)]$$

$$- l_{ps}(W_i;\gamma_0)' \cdot \mathbb{E}\left[\alpha^p_{ps}(\gamma_0)\left(\Delta Y - \mu^p_\Delta(\beta_0) - \mathbb{E}[w^p_{cont}(\gamma_0)\left(\Delta Y - \mu^p_\Delta(\beta_0)\right)]\right)\dot{\pi}(\gamma_0)\right]$$

$$= \eta^p_{tr}(W_i;\beta_0) - \eta^p_{cont}(W_i;\gamma_0,\beta_0) - \eta^p_{est}(W_i;\gamma_0,\beta_0),$$

where

$$\eta^p_{tr}(W_i;\beta_0) \equiv w^p_{tr}(D_i)\left(\Delta Y_i - \mu^p_\Delta(X_i;\beta_0) - \mathbb{E}[w^p_{tr}\left(\Delta Y - \mu^p_\Delta(\beta_0)\right)]\right),$$

$$\eta^p_{cont}(W_i;\gamma_0,\beta_0) \equiv w^p_{cont}(D_i,X_i;\gamma_0)\left(\Delta Y_i - \mu^p_\Delta(X_i;\beta_0) - \mathbb{E}\left[w^p_{cont}(\gamma_0)\left(\Delta Y - \mu^p_\Delta(\beta_0)\right)\right]\right),$$

as well as

$$\eta^p_{est}(W_i;\gamma_0,\beta_0) \equiv l_{reg}(W_i;\beta_0)' \cdot \mathbb{E}[\left(w^p_{tr} - w^p_{cont}(\gamma_0)\right)\dot{\mu}^p_\Delta(\beta_0)]$$

$$+ l_{ps}(W_i;\gamma_0)' \cdot \mathbb{E}\left[\alpha^p_{ps}(\gamma_0)\left(\Delta Y - \mu^p_\Delta(\beta_0) - \mathbb{E}[w^p_{cont}(\gamma_0)\left(\Delta Y - \mu^p_\Delta(\beta_0)\right)\right)\right)\dot{\pi}(\gamma_0)\right].$$

Thus, from (A.11) we conclude the proof of the asymptotic linear representation of $\sqrt{n}(\widehat{\tau}^{dr,p}_{att} - \tau^{dr,p}_{att})$. The asymptotic normality now follows directly from an application of the classical central limit theorem. ■

Proof of Theorem 3: By the conditional multiplier central limit theorem, see Lemma 2.9.5 in van der Vaart and Wellner (1996), as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i \cdot \eta^p(W_i;\gamma_0,\beta^p_{0,0},\beta^p_{1,0}) \xrightarrow{d} N(0,V^p).$$

Thus in order to show that

$$\sqrt{n}(\widehat{\tau}^{dr,p,*}_{att} - \widehat{\tau}^{dr,p}_{att}) \underset{*}{\xrightarrow{d}} N(0,V^p),$$

it is sufficient to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i \cdot [\widehat{\eta}^p(W_i;\widehat{\gamma},\widehat{\beta}^p_0,\widehat{\beta}^p_1) - \eta^p(W_i;\gamma_0,\beta^p_{0,0},\beta^p_{1,0})] = o_{p^*}(1).$$

Towards this goal, note that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot[\widehat{\eta}^p(W_i;\widehat{\gamma},\widehat{\beta}_0^p,\widehat{\beta}_1^p)-\eta^p(W_i;\gamma_0,\beta_{0,0}^p,\beta_{1,0}^p)]\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot[\widehat{\eta}_{tr}^p(W_i;\widehat{\beta}_0^p,\widehat{\beta}_1^p)-\eta_{tr}^p(W_i;\beta_{0,0}^p,\beta_{1,0}^p)] \qquad \text{(A.12)}\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot[\widehat{\eta}_{cont}^p(W_i;\widehat{\gamma},\widehat{\beta}_0^p,\widehat{\beta}_1^p)-\eta_{cont}^p(W_i;\gamma_0,\beta_{0,0}^p,\beta_{1,0}^p)] \qquad \text{(A.13)}\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot[\widehat{\eta}_{est}^p(W_i;\widehat{\gamma},\widehat{\beta}_0^p,\widehat{\beta}_1^p)-\eta_{est}^p(W_i;\gamma_0,\beta_{0,0}^p,\beta_{1,0}^p)], \qquad \text{(A.14)}
\end{aligned}$$

where $\widehat{\eta}_{tr}^p(W_i;\widehat{\beta}_0^p,\widehat{\beta}_1^p)$, $\widehat{\eta}_{cont}^p(W_i;\widehat{\gamma},\widehat{\beta}_0^p,\widehat{\beta}_1^p)$ and $\widehat{\eta}_{est}^p(W_i;\widehat{\gamma},\widehat{\beta}_0^p,\widehat{\beta}_1^p)$ are the sample analogue of the corresponding terms $\eta_{tr}^p(W_i;\beta_{0,0}^p,\beta_{1,0}^p)$, $\eta_{cont}^p(W_i;\gamma_0,\beta_{0,0}^p,\beta_{1,0}^p)$ and $\eta_{est}^p(W_i;\gamma_0,\beta_{0,0}^p,\beta_{1,0}^p)$, where population expectations are replaced by their empirical analogue, the pseudo-true parameters γ_0, $\beta_{0,0}^p$ and $\beta_{1,0}^p$ are replaced by their estimators $\widehat{\gamma}$, $\widehat{\beta}_0^p$ and $\widehat{\beta}_1^p$, respectively, and we also denote estimators for the derivatives $\dot{\pi}$ and $\dot{\mu}_t^p$, $t=0,1$, by $\widehat{\dot{\pi}}$ and $\widehat{\dot{\mu}}_t^p$, respectively. We next show that, under Assumptions 1-5, all three terms of (A.12)-(A.14) are $o_{p^*}(1)$.

We first show that (A.12) is $o_{p^*}(1)$. In fact, note that (A.12) can be decomposed into

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot[\widehat{\eta}_{tr}^p(W_i;\widehat{\beta}_0^p,\widehat{\beta}_1^p)-\eta_{tr}^p(W_i;\beta_{0,0}^p,\beta_{1,0}^p)]\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left(\frac{D_i}{\mathbb{E}_n[D]}\Delta Y_i-\frac{D_i}{\mathbb{E}[D]}\Delta Y_i\right)\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left(\frac{D_i}{\mathbb{E}_n[D]}\mu_\Delta^p(X_i;\widehat{\beta}_0^p,\widehat{\beta}_1^p)-\frac{D_i}{\mathbb{E}[D]}\mu_\Delta^p(X_i;\beta_{0,0}^p,\beta_{1,0}^p)\right)\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left(\frac{D_i}{\mathbb{E}_n[D]^2}\mathbb{E}_n[D(\Delta Y-\mu_\Delta^p(X_i;\widehat{\beta}_0^p,\widehat{\beta}_1^p))]-\frac{D_i}{\mathbb{E}[D]^2}\mathbb{E}[D(\Delta Y-\mu_\Delta^p(X_i;\beta_{0,0}^p,\beta_{1,0}^p))]\right)\\
\equiv&A_{1n}-A_{2n}-A_{3n},
\end{aligned}$$

and it suffices to show that A_{1n}, A_{2n} and A_{3n} are $o_{p^*}(1)$.

It is easy to get the following decompositions for A_{1n}, A_{2n} and A_{3n}:

$$A_{1n}=\left(\frac{1}{\mathbb{E}_n[D]}-\frac{1}{\mathbb{E}[D]}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot D_i\Delta Y_i,$$

and

$$A_{2n}=\left(\frac{1}{\mathbb{E}_n[D]}-\frac{1}{\mathbb{E}[D]}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\mu_{\Delta}^{p}(X_i;\beta_{0,0}^{p},\beta_{1,0}^{p})$$

$$+\frac{1}{\mathbb{E}_n[D]}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot D_i[\mu_{\Delta}^{p}(X_i;\widehat{\beta}_0^{p},\widehat{\beta}_1^{p})-\mu_{\Delta}^{p}(X_i;\beta_{0,0}^{p},\beta_{1,0}^{p})],$$

and

$$A_{3n}=\left(\frac{\mathbb{E}_n[D(\Delta Y-\mu_{\Delta}^{p}(X_i;\widehat{\beta}_0^{p},\widehat{\beta}_1^{p}))]}{\mathbb{E}_n[D]^2}-\frac{\mathbb{E}[D(\Delta Y-\mu_{\Delta}^{p}(X_i;\beta_{0,0}^{p},\beta_{1,0}^{p}))]}{\mathbb{E}[D]^2}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot D_i.$$

By weak law of large numbers, continuous mapping theorem, and Theorem 2.9.5 in van der Vaart and Wellner (1996), each term in the above expressions is $o_{p^*}(1)$, implying that

$$A_{1n}-A_{2n}-A_{3n}=o_{p^*}(1).$$

Next we show that (A.13) is $o_{p^*}(1)$. Again, notice that (A.13) can be decomposed into

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot[\widehat{\eta}_{cont}^{p}(W_i;\widehat{\gamma},\widehat{\beta}_0^{p},\widehat{\beta}_1^{p})-\eta_{cont}^{p}(W_i;\gamma_0,\beta_{0,0}^{p},\beta_{1,0}^{p})]\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left[\left(\widehat{w}_{cont}^{p}(D_i,X_i;\widehat{\gamma})-w_{cont}^{p}(D_i,X_i;\gamma_0)\right)\Delta Y_i\right]\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left[\left(\widehat{w}_{cont}^{p}(D_i,X_i;\widehat{\gamma})\,\mu_{\Delta}^{p}(X_i;\widehat{\beta}_0^{p},\widehat{\beta}_1^{p})-w_{cont}^{p}(D_i,X_i;\gamma_0)\mu_{\Delta}^{p}(X_i;\beta_{0,0}^{p},\beta_{1,0}^{p})\right)\right.\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left[\left(\widehat{w}_{cont}^{p}(D_i,X_i;\widehat{\gamma})\right)\mathbb{E}_n\left[\widehat{w}_{cont}^{p}(D,X;\widehat{\gamma})\left(\Delta Y-\mu_{\Delta}^{p}(X;\widehat{\beta}_0^{p},\widehat{\beta}_1^{p})\right)\right]\right.\\
&\qquad\left.-w_{cont}^{p}(D_i,X_i;\gamma_0)\mathbb{E}\left[w_{cont}^{p}(D,X;\gamma_0)\left(\Delta Y-\mu_{\Delta}^{p}(X;\beta_{0,0}^{p},\beta_{1,0}^{p})\right)\right]\right]\\
\equiv&B_{1n}-B_{2n}-B_{3n},
\end{aligned}$$

and it suffices to show that B_{1n}, B_{2n} and B_{3n} are $o_{p^*}(1)$.

It is easy to get the following decompositions for B_{1n}, B_{2n} and B_{3n}:

$$B_{1n}=\left(\mathbb{E}_n\left[\frac{\pi(X;\widehat{\gamma})(1-D)}{1-\pi(X;\widehat{\gamma})}\right]^{-1}-\mathbb{E}\left[\frac{\pi(X;\gamma_0)(1-D)}{1-\pi(X;\gamma_0)}\right]^{-1}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\frac{\pi(X_i;\gamma_0)(1-D_i)}{1-\pi(X_i;\gamma_0)}\Delta Y_i$$

$$+\mathbb{E}_n\left[\frac{\pi(X;\widehat{\gamma})(1-D)}{1-\pi(X;\widehat{\gamma})}\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left(\frac{\pi(X_i;\widehat{\gamma})(1-D_i)}{1-\pi(X_i;\widehat{\gamma})}-\frac{\pi(X_i;\gamma_0)(1-D_i)}{1-\pi(X_i;\gamma_0)}\right)\Delta Y_i,$$

and

$$
\begin{aligned}
&B_{2n}\\
=&\left(\mathbb{E}_n\left[\frac{\pi(X;\widehat{\gamma})(1-D)}{1-\pi(X;\widehat{\gamma})}\right]^{-1}-\mathbb{E}\left[\frac{\pi(X;\gamma_0)(1-D)}{1-\pi(X;\gamma_0)}\right]^{-1}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\frac{\pi(X_i;\gamma_0)(1-D_i)}{1-\pi(X_i;\gamma_0)}\mu_{\Delta}^{p}(X_i;\beta_0)\\
&+\mathbb{E}_n\left[\frac{\pi(X;\widehat{\gamma})(1-D)}{1-\pi(X;\widehat{\gamma})}\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left(\frac{\pi(X_i;\widehat{\gamma})(1-D_i)}{1-\pi(X_i;\widehat{\gamma})}\mu_{\Delta}^{p}(X_i;\widehat{\beta})-\frac{\pi(X_i;\gamma_0)(1-D_i)}{1-\pi(X_i;\gamma_0)}\mu_{\Delta}^{p}(X_i;\beta_0)\right),
\end{aligned}
$$

also

$$
\begin{aligned}
B_{3n}=&\tilde{B}_{3n,1}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\frac{\pi(X_i;\gamma_0)(1-D_i)}{1-\pi(X_i;\gamma_0)}\\
&+\frac{\mathbb{E}_n\left[\frac{\pi(X;\widehat{\gamma})(1-D)}{1-\pi(X;\widehat{\gamma})}\left(\Delta Y-\mu_{\Delta}^{p}(X;\widehat{\beta})\right)\right]}{\mathbb{E}_n\left[\frac{\pi(X;\widehat{\gamma})(1-D)}{1-\pi(X;\widehat{\gamma})}\right]^2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left(\frac{\pi(X_i;\widehat{\gamma})(1-D_i)}{1-\pi(X_i;\widehat{\gamma})}-\frac{\pi(X_i;\gamma_0)(1-D_i)}{1-\pi(X_i;\gamma_0)}\right),
\end{aligned}
$$

where $\widehat{\beta}=(\widehat{\beta}_1^p,\widehat{\beta}_0^p)$, $\beta_0=(\beta_{1,0}^p,\beta_{0,0}^p)$, and

$$
\tilde{B}_{3n,1}=\frac{\mathbb{E}_n\left[\widehat{w}_{cont}^{p}(D_i,X_i;\widehat{\gamma})\left(\Delta Y-\mu_{\Delta}^{p}(X;\widehat{\beta})\right)\right]}{\mathbb{E}_n\left[\frac{\pi(X;\widehat{\gamma})(1-D)}{1-\pi(X;\widehat{\gamma})}\right]}-\frac{\mathbb{E}\left[w_{cont}^{p}(D_i,X_i;\gamma_0)\left(\Delta Y-\mu_{\Delta}^{p}(X;\beta_0)\right)\right]}{\mathbb{E}\left[\frac{\pi(X;\gamma_0)(1-D)}{1-\pi(X;\gamma_0)}\right]}.
$$

By weak law of larger numbers, continuous mapping theorem, and Theorem 2.9.5 in van der Vaart and Wellner (1996), we have that each term in the above expressions is $o_{p^*}(1)$, implying that

$$
B_{1n}-B_{2n}-B_{3n}=o_{p^*}(1).
$$

Finally we need to show that (A.14) is $o_{p^*}(1)$. As before, (A.14) can be decomposed into

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left[\widehat{\eta}_{est}^{p}(W_i;\widehat{\gamma},\widehat{\beta}_0^p,\widehat{\beta}_1^p)-\eta_{est}^{p}(W_i;\gamma_0,\beta_{0,0}^p,\beta_{1,0}^p)\right]\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left[l_{reg}(W_i;\widehat{\beta})'\mathbb{E}_n\left[\left(\widehat{w}_{tr}^{p}-\widehat{w}_{cont}^{p}(\widehat{\gamma})\right)\widehat{\dot{\mu}}_{\Delta}^{p}\left(\widehat{\beta}\right)\right]-l_{reg}(W_i;\beta_0)'\mathbb{E}\left[\left(w_{tr}^{p}-w_{cont}^{p}(\gamma_0)\right)\dot{\mu}_{\Delta}^{p}(\beta_0)\right]\right]\\
+&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left[l_{ps}(W_i;\widehat{\gamma})'\mathbb{E}_n\left[\widehat{\alpha}^{p}(\widehat{\gamma})\left(\Delta Y-\mu_{\Delta}^{p}(\widehat{\beta})-\mathbb{E}_n[\widehat{w}_{cont}^{p}(\widehat{\gamma})(\Delta Y-\mu_{\Delta}^{p}(\widehat{\beta}))]\right)\widehat{\dot{\pi}}(\widehat{\gamma})\right]\right.\\
&\left.-l_{ps}(W_i;\gamma_0)'\mathbb{E}\left[\alpha^{p}(\gamma_0)\left(\Delta Y-\mu_{\Delta}^{p}(\beta_0)-\mathbb{E}\left[w_{cont}^{p}(\gamma_0)\left(\Delta Y-\mu_{\Delta}^{p}(\beta_0)\right)\right]\right)\dot{\pi}(\gamma_0)\right]\right]
\end{aligned}
$$

$$\equiv C_{1n} + C_{2n},$$

where the dependence of the functionals on W within $\mathbb{E}_n[\cdot]$ and $\mathbb{E}[\cdot]$ is dropped to ease the notation. To conclude the proof, it suffices to show that C_{1n} and C_{2n} are $o_{p^*}(1)$. Given that,

$$C_{1n} = \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i \cdot l_{reg}(W_i;\beta_0)'\right] \cdot \left[\mathbb{E}_n\left[\left(\widehat{w}_{tr}^p - \widehat{w}_{cont}^p(\widehat{\gamma})\right)\widehat{\dot{\mu}}_{\Delta}^p\left(\widehat{\beta}\right)\right] - \mathbb{E}\left[(w_{tr}^p - w_{cont}^p(\gamma_0))\dot{\mu}_{\Delta}^p(\beta_0)\right]\right]$$

$$+ \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i \cdot [l_{reg}(W_i;\widehat{\beta}) - l_{reg}(W_i;\beta_0)]'\right] \cdot \mathbb{E}_n\left[\left(\widehat{w}_{tr}^p - \widehat{w}_{cont}^p(\widehat{\gamma})\right)\widehat{\dot{\mu}}_{\Delta}^p\left(\widehat{\beta}\right)\right],$$

and

$$C_{2n}$$

$$= \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i \cdot l_{ps}(W_i;\gamma_0)'\right] \cdot \left[\mathbb{E}_n\left[\widehat{\alpha}^p(\widehat{\gamma})\left(\Delta Y - \mu_{\Delta}^p(\widehat{\beta}) - \mathbb{E}_n[\widehat{w}_{cont}^p(\widehat{\gamma})(\Delta Y - \mu_{\Delta}^p(\widehat{\beta}))]\right)\dot{\pi}(\widehat{\gamma})\right]\right.$$

$$\left. - \mathbb{E}\left[\alpha^p(\gamma_0)(\Delta Y - \mu_{\Delta}^p(\beta_0) - \mathbb{E}[w_{cont}^p(\gamma_0)(\Delta Y - \mu_{\Delta}^p(\beta_0))])\dot{\pi}(\gamma_0)\right]\right]$$

$$+ \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i \cdot [l_{ps}(W_i;\widehat{\gamma}) - l_{ps}(W_i;\gamma_0)]'\right] \cdot \mathbb{E}_n\left[\widehat{\alpha}^p(\widehat{\gamma})(\Delta Y - \mu_{\Delta}^p(\widehat{\beta}) - \mathbb{E}_n[\widehat{w}_{cont}^p(\widehat{\gamma})(\Delta Y - \mu_{\Delta}^p(\widehat{\beta}))])\dot{\pi}(\widehat{\gamma})\right],$$

by the weak law of large numbers, continuous mapping theorem, and Theorem 2.9.5 in van der Vaart and Wellner (1996), each term in the above expressions is $o_{p^*}(1)$, implying that

$$C_{1n} + C_{2n} = o_{p^*}(1),$$

which concludes the proof. ■

Proof of Theorem 4: Under Assumptions 1-5 and the additional assumption that $n_1/n \to \lambda \in (0,1)$, we establish the large sample properties of the estimator $\widehat{\tau}_{att}^{dr,rc}$ when repeated cross section data are available.

First of all, recall that the estimator takes the following form:

$$\widehat{\tau}_{att}^{dr,rc} = \mathbb{E}_n\left[(\widehat{w}_{tr}^{rc}(D,T) - \widehat{w}_{cont}^{rc}(D,T,X;\widehat{\gamma}))\left(Y - \mu_Y^{rc}(X,T;\widehat{\beta}_0^{rc},\widehat{\beta}_1^{rc})\right)\right],$$

where

$$\widehat{w}_{tr}^{rc}(D,T) = \widehat{w}_{tr,1}^{rc}(D,T) - \widehat{w}_{tr,0}^{rc}(D,T),$$

$$\widehat{w}^{rc}_{cont}(D,T,X;\gamma) = \widehat{w}^{rc}_{cont,1}(D,T,X;\gamma) - \widehat{w}^{rc}_{cont,0}(D,T,X;\gamma);$$

with

$$\begin{aligned}
\widehat{w}^{rc}_{tr,1}(D,T) &= \frac{DT}{\mathbb{E}_n[DT]}, \\
\widehat{w}^{rc}_{tr,0}(D,T) &= \frac{D(1-T)}{\mathbb{E}_n[D(1-T)]}, \\
\widehat{w}^{rc}_{cont,1}(D,T,X;\gamma) &= \frac{\pi(X;\gamma)(1-D)T}{1-\pi(X;\gamma)} \Bigg/ \mathbb{E}_n\left[\frac{\pi(X;\gamma)(1-D)T}{1-\pi(X;\gamma)}\right], \\
\widehat{w}^{rc}_{cont,1}(D,T,X;\gamma) &= \frac{\pi(X;\gamma)(1-D)(1-T)}{1-\pi(X;\gamma)} \Bigg/ \mathbb{E}_n\left[\frac{\pi(X;\gamma)(1-D)(1-T)}{1-\pi(X;\gamma)}\right];
\end{aligned}$$

and $\widehat{\gamma}$, $\widehat{\beta}^{rc}_0$ and $\widehat{\beta}^{rc}_1$ are estimators for the pseudo-true parameters γ_0, $\beta^{rc}_{0,0}$ and $\beta^{p}_{1,0}$; and, for generic β^{rc}_0 and β^{rc}_1, $\mu^{rc}_Y(\cdot,T;\beta^{rc}_0,\beta^{rc}_1) = T\cdot\mu^{rc}_1(\cdot;\beta^{rc}_1) + (1-T)\cdot\mu^{rc}_0(\cdot;\beta^{rc}_0)$.

By the weak law of large numbers and continuous mapping theorem, we have that, as $n\to\infty$,

$$\widehat{\tau}^{dr,rc}_{att} \xrightarrow{p} \mathbb{E}\left[\left(w^{rc}_{tr}(D,T) - w^{rc}_{cont}(D,T,X;\gamma_0)\right)\left(Y - \mu^{rc}_Y(T,X;\beta^{rc}_{0,0},\beta^{rc}_{1,0})\right)\right],$$

where

$$w^{rc}_{tr}(D,T) = w^{rc}_{tr,1}(D,T) - w^{rc}_{tr,0}(D,T),$$

$$w^{rc}_{cont}(D,T,X;\gamma) = w^{rc}_{cont,1}(D,T,X;\gamma) - w^{rc}_{cont,0}(D,T,X;\gamma),$$

and

$$\begin{aligned}
w^{rc}_{tr,1}(D,T) &= \frac{DT}{\mathbb{E}[DT]}, \\
w^{rc}_{tr,0}(D,T) &= \frac{D(1-T)}{\mathbb{E}[D(1-T)]}, \\
w^{rc}_{cont,1}(D,T,X;\gamma) &= \frac{\pi(X;\gamma)(1-D)T}{1-\pi(X;\gamma)} \Bigg/ \mathbb{E}[\frac{\pi(X;\gamma)(1-D)T}{1-\pi(X;\gamma)}], \\
w^{rc}_{cont,0}(D,T,X;\gamma) &= \frac{\pi(X;\gamma)(1-D)(1-T)}{1-\pi(X;\gamma)} \Bigg/ \mathbb{E}[\frac{\pi(X;\gamma)(1-D)(1-T)}{1-\pi(X;\gamma)}].
\end{aligned}$$

Thus, if either $\pi(X;\gamma_0) = p(X)$ a.s. or $\mu^{rc}_Y(T,X;\beta^{rc}_{0,0},\beta^{rc}_{1,0}) = T\cdot m^{rc}_{0,1}(X) + (1-T)\cdot m^{rc}_{0,0}(X)$

a.s., it follows from Theorem 1 that

$$\mathbb{E}\left[\left(w_{tr}^{rc}(D,T)-w_{cont}^{rc}(D,T,X;\gamma_0)\right)\left(Y-\mu_Y^{rc}(T,X;\beta_{0,0}^{rc},\beta_{1,0}^{rc})\right)\right]\equiv\tau_{att}^{dr,rc}=ATT,$$

which completes the convergence in probability result.

Next, we establish the asymptotically linear representation of $\widehat{\tau}_{att}^{dr,rc}$. Following the same procedure of the proof of Theorem 2, we first obtain the decomposition

$$\begin{aligned}
&\widehat{\tau}_{att}^{dr,rc}-\tau_{att}^{dr,rc}\\
=&\left(\mathbb{E}_n\left[\widehat{w}_{tr,1}^{rc}(D,T)Y\right]-\mathbb{E}[w_{tr,1}^{rc}(D,T)Y]\right)\\
&-\left(\mathbb{E}_n\left[\widehat{w}_{tr,0}^{rc}(D,T)Y\right]-\mathbb{E}[w_{tr,0}^{rc}(D,T)Y]\right)\\
&-\left(\mathbb{E}_n\left[\widehat{w}_{cont,1}^{rc}(D,T,X;\widehat{\gamma})Y\right]-\mathbb{E}\left[w_{cont,1}^{rc}(D,T,X;\gamma_0)Y\right]\right)\\
&+\left(\mathbb{E}_n\left[\widehat{w}_{cont,0}^{rc}(D,T,X;\widehat{\gamma})Y\right]-\mathbb{E}[w_{cont,0}^{rc}(D,T,X;\gamma_0)Y]\right)\\
&-\left(\mathbb{E}_n\left[\widehat{w}_{tr,1}^{rc}(D,T)\mu_Y^{rc}\left(X,T;\widehat{\beta}_0^{rc},\widehat{\beta}_1^{rc}\right)\right]-\mathbb{E}[w_{tr,1}^{rc}(D,T)\mu_Y^{rc}(X,T;\beta_{0,0}^{rc},\beta_{1,0}^{rc})]\right)\\
&+\left(\mathbb{E}_n\left[\widehat{w}_{tr,0}^{rc}(D,T)\mu_Y^{rc}\left(X,T;\widehat{\beta}_0^{rc},\widehat{\beta}_1^{rc}\right)\right]-\mathbb{E}[w_{tr,0}^{rc}(D,T)\mu_Y^{rc}(X,T;\beta_{0,0}^{rc},\beta_{1,0}^{rc})]\right)\\
&+\left(\mathbb{E}_n\left[\widehat{w}_{cont,1}^{rc}(D,T,X;\widehat{\gamma})\mu_Y^{rc}\left(X,T;\widehat{\beta}_0^{rc},\widehat{\beta}_1^{rc}\right)\right]-\mathbb{E}[w_{cont,1}^{rc}(D,T,X;\gamma_0)\mu_Y^{rc}(X,T;\beta_{0,0}^{rc},\beta_{1,0}^{rc})]\right)\\
&-\left(\mathbb{E}_n\left[\widehat{w}_{cont,0}^{rc}(D,T,X;\widehat{\gamma})\mu_Y^{rc}\left(X,T;\widehat{\beta}_0^{rc},\widehat{\beta}_1^{rc}\right)\right]-\mathbb{E}[w_{cont,0}^{rc}(D,T,X;\gamma_0)\mu_Y^{rc}(X,T;\beta_{0,0}^{rc},\beta_{1,0}^{rc})]\right)\\
\equiv&\left(\widehat{ATT}_1-ATT_1\right)-\left(\widehat{ATT}_2-ATT_2\right)-\left(\widehat{ATT}_3-ATT_3\right)+\left(\widehat{ATT}_4-ATT_4\right)\\
&-\left(\widehat{ATT}_5-ATT_5\right)+\left(\widehat{ATT}_6-ATT_6\right)+\left(\widehat{ATT}_7-ATT_7\right)-\left(\widehat{ATT}_8-ATT_8\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sqrt{n}(\widehat{\tau}_{att}^{dr,rc}-\tau_{att}^{dr,rc})\\
=&\sqrt{n}\left(\widehat{ATT}_1-ATT_1\right)-\sqrt{n}\left(\widehat{ATT}_2-ATT_2\right)-\sqrt{n}\left(\widehat{ATT}_3-ATT_3\right)+\sqrt{n}\left(\widehat{ATT}_4-ATT_4\right)\\
&-\sqrt{n}\left(\widehat{ATT}_5-ATT_5\right)+\sqrt{n}\left(\widehat{ATT}_6-ATT_6\right)+\sqrt{n}\left(\widehat{ATT}_7-ATT_7\right)-\sqrt{n}\left(\widehat{ATT}_8-ATT_8\right),
\end{aligned}$$

and we next obtain the asymptotically linear representation for each component in the above decomposition.

For $\sqrt{n}\left(\widehat{ATT}_1 - ATT_1\right)$, following the analogous steps to derive (A.6) in the panel data case, we have that

$$\sqrt{n}\left(\widehat{ATT}_1 - ATT_1\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{tr,1}^{rc}(D_i,T_i)\left(Y_i - \mathbb{E}[w_{tr,1}^{rc}(D,T)Y]\right) + o_p(1). \tag{A.15}$$

Analogously, for $\sqrt{n}\left(\widehat{ATT}_2 - ATT_2\right)$ we have that

$$\sqrt{n}\left(\widehat{ATT}_2 - ATT_2\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{tr,0}^{rc}(D_i,T_i)\left(Y_i - \mathbb{E}[w_{tr,0}^{rc}(D,T)Y]\right) + o_p(1). \tag{A.16}$$

The case of $\sqrt{n}\left(\widehat{ATT}_3 - ATT_3\right)$ is similar to (A.7)-(A.8) in the panel data case. That is, by following similar steps as above and a second-order Taylor expansion argument, we have that

$$\begin{aligned}
&\sqrt{n}\left(\widehat{ATT}_3 - ATT_3\right)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \widetilde{w}_{cont,1}^{rc}(D_i,T_i,X_i;\widehat{\gamma})\left(Y_i - \mathbb{E}[w_{cont,1}^{rc}(D,T,X;\gamma_0)Y]\right) + o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left(w_{cont,1}^{rc}(D_i,T_i,X_i;\gamma_0)\left(Y_i - \mathbb{E}[w_{cont,1}^{rc}(D,T,X;\gamma_0)Y]\right)\right.\\
&+ l_{ps}(W_i;\gamma_0)' \cdot \mathbb{E}\left[\alpha_{ps,1}^{rc}(D,T,X;\gamma_0)\left(Y - \mathbb{E}\left[w_{cont,1}^{rc}(D,T,X;\gamma_0)Y\right]\right)\dot{\pi}(X;\gamma_0)\right]\Big) + o_p(1)
\end{aligned} \tag{A.17}$$

where

$$\begin{aligned}
\widetilde{w}_{cont,1}^{rc}(D,T,X;\widehat{\gamma}) &= \frac{\pi(X;\widehat{\gamma})(1-D)T}{1-\pi(X;\widehat{\gamma})} \Big/ \mathbb{E}\left[\frac{\pi(X;\gamma_0)(1-D)T}{1-\pi(X;\gamma_0)}\right],\\
\alpha_{ps,1}^{rc}(D,T,X;\gamma) &= \frac{(1-D)T}{(1-\pi(X;\gamma))^2} \Big/ \mathbb{E}\left[\frac{\pi(X;\gamma)(1-D)T}{1-\pi(X;\gamma)}\right].
\end{aligned}$$

Using analogous arguments as in (A.17), we get

$$\begin{aligned}
&\sqrt{n}\left(\widehat{ATT}_4 - ATT_4\right)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left(w_{cont,0}^{rc}(D_i,T_i,X_i;\gamma_0)\left(Y_i - \mathbb{E}[w_{cont,0}^{rc}(D,T,X;\gamma_0)Y]\right)\right.\\
&+ l_{ps}(W_i;\gamma_0)' \cdot \mathbb{E}\left[\alpha_{ps,0}^{rc}(D,T,X;\gamma_0)\left(Y - \mathbb{E}[w_{cont,0}^{rc}(D,T,X;\gamma_0)Y]\right)\dot{\pi}(X;\gamma_0)\right]\Big) + o_p(1),
\end{aligned} \tag{A.18}$$

where

$$\alpha^{rc}_{ps,0}(D,T,X;\gamma) = \frac{(1-D)(1-T)}{(1-\pi(X;\gamma))^2} \Big/ \mathbb{E}\left[\frac{\pi(X;\gamma)(1-D)(1-T)}{1-\pi(X;\gamma)}\right].$$

We next derive the asymptotic linear representation of $\sqrt{n}\left(\widehat{ATT}_5 - ATT_5\right)$ by following similar steps as in (A.9). First, it is easy to show that

$$\begin{aligned}&\sqrt{n}\left(\widehat{ATT}_5 - ATT_5\right)\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w^{rc}_{tr,1}(D_i,T_i)\left(\mu^{rc}_Y(X_i,T_i;\widehat{\beta}^{rc}_0,\widehat{\beta}^{rc}_1) - \mathbb{E}[w^{rc}_{tr,1}(D,T)\mu^{rc}_Y(X,T;\beta^{rc}_{0,0},\beta^{rc}_{1,0})]\right) + o_p(1).\end{aligned}$$

Let $\widehat{\beta} = (\widehat{\beta}^{rc}_1, \widehat{\beta}^{rc}_0)$, and $\beta_0 = (\beta^{rc}_{1,0}, \beta^{rc}_{0,0})$. Next, from a second-order Taylor expansion argument, we obtain that

$$\begin{aligned}&\sqrt{n}\left(\widehat{ATT}_5 - ATT_5\right)\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(w^{rc}_{tr,1}(D_i,T_i)\left(\mu^{rc}_Y(X_i,T_i;\beta_0) - \mathbb{E}[w^{rc}_{tr,1}(D,T)\mu^{rc}_Y(X,T;\beta_0)]\right)\\ &+l_{reg}(W_i;\beta_0)'\cdot\mathbb{E}\left[w^{rc}_{tr,1}(D,T)\dot{\mu}^{rc}_Y(X,T;\beta_0)\right]\Big) + o_p(1).\end{aligned} \tag{A.19}$$

Using the same arguments as in (A.19), we have

$$\begin{aligned}&\sqrt{n}\left(\widehat{ATT}_6 - ATT_6\right)\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(w^{rc}_{tr,0}(D_i,T_i)\left(\mu^{rc}_Y(X_i,T_i;\beta_0) - \mathbb{E}[w^{rc}_{tr,0}(D,T)\mu^{rc}_Y(X,T;\beta_0)]\right)\\ &+l_{reg}(W_i;\beta_0)'\cdot\mathbb{E}\left[w^{rc}_{tr,0}(D,T)\dot{\mu}^{rc}_Y(X,T;\beta_0)\right]\Big) + o_p(1).\end{aligned} \tag{A.20}$$

The asymptotically linear representation of $\sqrt{n}\left(\widehat{ATT}_7 - ATT_7\right)$ can be derived following similar steps as in (A.10). More precisely, one can easily show that

$$\begin{aligned}&\sqrt{n}\left(\widehat{ATT}_7 - ATT_7\right)\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{w}^{rc}_{cont,1}(D_i,T_i,X_i;\widehat{\gamma})\left(\mu^{rc}_Y(X_i,T_i;\widehat{\beta}) - \mathbb{E}[w^{rc}_{cont,1}(D,T,X;\gamma_0)\mu^{rc}_Y(X,T;\beta_0)]\right) + o_p(1).\end{aligned}$$

Then, by doing a second-order Taylor expansion of the above expression around pseudo-true γ_0 and β_0, and plugging the asymptotically linear representations of $\sqrt{n}(\widehat{\gamma}-\gamma_0)$ and $\sqrt{n}\left(\widehat{\beta}-\beta_0\right)$, we have

$$
\begin{aligned}
&\sqrt{n}\left(\widehat{ATT}_7 - ATT_7\right)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(w^{rc}_{cont,1}(D_i,T_i,X_i;\gamma_0)\left(\mu^{rc}_Y(X_i,T_i;\beta_0) - \mathbb{E}[w^{rc}_{cont,1}(\gamma_0)\mu^{rc}_Y(\beta_0)]\right)\right.\\
&+l_{ps}(W_i;\gamma_0)'\cdot\mathbb{E}\left[\alpha^{rc}_{ps,1}(\gamma_0)\left(\mu^{rc}_Y(\beta_0) - \mathbb{E}[w^{rc}_{cont,1}(\gamma_0)\mu^{rc}_Y(\beta_0)]\right)\dot{\pi}(\gamma_0)\right]\\
&\left.+l_{reg}(W_i;\beta_0)'\cdot\mathbb{E}\left[w^{rc}_{cont,1}(\gamma_0)\dot{\mu}^{rc}_Y(\beta_0)\right]\right) + o_p(1), \qquad \text{(A.21)}
\end{aligned}
$$

where the dependence of the functionals on W within $\mathbb{E}[\cdot]$ is dropped to ease the notation.

Analogously, we have that

$$
\begin{aligned}
&\sqrt{n}\left(\widehat{ATT}_8 - ATT_8\right)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(w^{rc}_{cont,0}(D_i,T_i,X_i;\gamma_0)\left(\mu^{rc}_Y(X_i,T_i;\beta_0) - \mathbb{E}[w^{rc}_{cont,0}(\gamma_0)\mu^{rc}_Y(\beta_0)]\right)\right.\\
&+l_{ps}(W_i;\gamma_0)'\cdot\mathbb{E}\left[\alpha^{rc}_{ps,0}(\gamma_0)\left(\mu^{rc}_Y(\beta_0) - \mathbb{E}[w^{rc}_{cont,0}(\gamma_0)\mu^{rc}_Y(\beta_0)]\right)\dot{\pi}(\gamma_0)\right]\\
&\left.+l_{reg}(W_i;\beta_0)'\cdot\mathbb{E}\left[w^{rc}_{cont,0}(\gamma_0)\dot{\mu}^{rc}_Y(\beta_0)\right]\right) + o_p(1), \qquad \text{(A.22)}
\end{aligned}
$$

Finally, note that by combining and rearranging (A.15)-(A.22), we have that

$$
\begin{aligned}
&\sqrt{n}(\widehat{\tau}^{dr,rc}_{att} - \tau^{dr,rc}_{att})\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(w^{rc}_{tr,1}(D_i,T_i)\left(Y_i - \mu^{rc}_Y(X_i,T_i;\beta_0) - \mathbb{E}[w^{rc}_{tr,1}[Y - \mu^{rc}_Y(\beta_0)]]\right)\right.\\
&- w^{rc}_{tr,0}(D_i,T_i)\left(Y_i - \mu^{rc}_Y(X_i,T_i;\beta_0) - \mathbb{E}[w^{rc}_{tr,0}(Y - \mu^{rc}_Y(\beta_0))]\right)\\
&- w^{rc}_{cont,1}(D_i,T_i,X_i;\gamma_0)\left(Y_i - \mu^{rc}_Y(X_i,T_i;\beta_0) - \mathbb{E}[w^{rc}_{cont,1}(\gamma_0)(Y - \mu^{rc}_Y(\beta_0))]\right)\\
&+ w^{rc}_{cont,0}(D_i,T_i,X_i;\gamma_0)\left(Y_i - \mu^{rc}_Y(X_i,T_i;\beta_0) - \mathbb{E}[w^{rc}_{cont,0}(\gamma_0)(Y - \mu^{rc}_Y(\beta_0))]\right)\\
&- l_{reg}(W_i;\beta_0)'\cdot\mathbb{E}[(w^{rc}_{tr,1} - w^{rc}_{tr,0}) - (w^{rc}_{cont,1}(\gamma_0) - w^{rc}_{cont,0}(\gamma_0))\dot{\mu}^{rc}_Y(\beta_0)]\\
&- l_{ps}(W_i;\gamma_0)'\cdot\mathbb{E}\left[\alpha^{rc}_{ps,1}(\gamma_0)\left(Y - \mu^{rc}_Y(\beta_0) - \mathbb{E}[w^{rc}_{cont,1}(\gamma_0)(Y - \mu^{rc}_Y(\beta_0))]\right)\dot{\pi}(\gamma_0)\right]\\
&\left.+l_{ps}(W_i;\gamma_0)'\cdot\mathbb{E}\left[\alpha^{rc}_{ps,0}(\gamma_0)\left(Y - \mu^{rc}_Y(\beta_0) - \mathbb{E}[w^{rc}_{cont,0}(\gamma_0)(Y - \mu^{rc}_Y(\beta_0))]\right)\dot{\pi}(\gamma_0)\right]\right)
\end{aligned}
$$

$$
\begin{aligned}
&+o_p(1) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta^{rc}(W_i;\gamma_0,\beta_0)+o_p(1),
\end{aligned} \tag{A.23}
$$

since

$$
\begin{aligned}
&\eta^{rc}(W_i;\gamma_0,\beta_0) \\
\equiv&w_{tr,1}^{rc}(D_i,T_i)\left(Y_i-\mu_Y^{rc}(X_i,T_i;\beta_0)-\mathbb{E}[w_{tr,1}^{rc}\left(Y-\mu_Y^{rc}(\beta_0)\right)]\right) \\
&-w_{tr,0}^{rc}(D_i,T_i)\left(Y_i-\mu_Y^{rc}(X_i,T_i;\beta_0)-\mathbb{E}[w_{tr,0}^{rc}\left(Y-\mu_Y^{rc}(\beta_0)\right)]\right) \\
&-w_{cont,1}^{rc}(D_i,T_i,X_i;\gamma_0)\left(Y_i-\mu_Y^{rc}(X_i,T_i;\beta_0)-\mathbb{E}[w_{cont,1}^{rc}(\gamma_0)\left(Y-\mu_Y^{rc}(\beta_0)\right)]\right) \\
&+w_{cont,0}^{rc}(D_i,T_i,X_i;\gamma_0)\left(Y_i-\mu_Y^{rc}(X_i,T_i;\beta_0)-\mathbb{E}[w_{cont,0}^{rc}(\gamma_0)\left(Y-\mu_Y^{rc}(\beta_0)\right)]\right) \\
&-l_{reg}(W_i;\beta_0)'\cdot\mathbb{E}[(w_{tr,1}^{rc}-w_{tr,0}^{rc})-(w_{cont,1}^{rc}(\gamma_0)-w_{cont,0}^{rc}(\gamma_0))\dot{\mu}_Y^{rc}(\beta_0)] \\
&-l_{ps}(W_i;\gamma_0)'\cdot\mathbb{E}\left[\alpha_{ps,1}^{rc}(\gamma_0)\left(Y-\mu_Y^{rc}(\beta_0)-\mathbb{E}[w_{cont,1}^{rc}(\gamma_0)\left(Y-\mu_Y^{rc}(\beta_{0,0}^{rc},\beta_{1,0}^{rc})\right)]\right)\dot{\pi}(\gamma_0)\right] \\
&+l_{ps}(W_i;\gamma_0)'\cdot\mathbb{E}\left[\alpha_{ps,0}^{rc}(\gamma_0)\left(Y-\mu_Y^{rc}(\beta_0)-\mathbb{E}[w_{cont,0}^{rc}(\gamma_0)\left(Y-\mu_Y^{rc}(\beta_0)\right)]\right)\dot{\pi}(\gamma_0)\right] \\
=&\eta_{tr}^{rc}(W_i;\beta_0)-\eta_{cont}^{rc}(W_i;\gamma_0,\beta_0)-\eta_{est}^{rc}(W_i;\gamma_0,\beta_0),
\end{aligned}
$$

where

$$
\begin{aligned}
\eta_{tr}^{rc}(W_i;\beta_0)&=\eta_{tr,1}^{rc}(W_i;\beta_0)-\eta_{tr,0}^{rc}(W_i;\beta_0), \\
\eta_{cont}^{rc}(W_i;\gamma_0,\beta_0)&=\eta_{cont,1}^{rc}(W_i;\gamma_0,\beta_0)-\eta_{cont,0}^{rc}(W_i;\gamma_0,\beta_0), \\
\eta_{est}^{rc}(W_i;\gamma_0,\beta_0)&=\eta_{est,reg}^{rc}(W_i;\gamma_0,\beta_0)+\eta_{est,ps}^{rc}(W_i;\gamma_0,\beta_0),
\end{aligned}
$$

and, for $t=0,1$,

$$
\begin{aligned}
\eta_{tr,t}^{rc}(W_i;\gamma_0,\beta_0)&=w_{tr,t}^{rc}(D_i,T_i)\left(Y_i-\mu_Y^{rc}(X_i,T_i;\beta_0)-\mathbb{E}[w_{tr,t}^{rc}\left(Y-\mu_Y^{rc}(\beta_0)\right)]\right), \\
\eta_{cont,t}^{rc}(W_i;\gamma_0,\beta_0)&=w_{cont,t}^{rc}(D_i,T_i,X_i;\gamma_0)\left(Y_i-\mu_Y^{rc}(X_i,T_i;\beta_0)-\mathbb{E}[w_{cont,t}^{rc}(\gamma_0)\left(Y-\mu_Y^{rc}(\beta_0)\right)]\right),
\end{aligned}
$$

and

$$\eta^{rc}_{est,reg}(W_i;\gamma_0,\beta_0) = l_{reg}(W_i;\beta_0)' \cdot \mathbb{E}[(w^{rc}_{tr,1} - w^{rc}_{tr,0}) - (w^{rc}_{cont,1}(\gamma_0) - w^{rc}_{cont,0}(\gamma_0))\dot{\mu}^{rc}_Y(\beta_0)],$$

$$\begin{aligned}
&\eta^{rc}_{est,ps}(W_i;\gamma_0,\beta_0) \\
=&l_{ps}(W_i;\gamma_0)' \cdot \mathbb{E}\left[\alpha^{rc}_{ps,1}(\gamma_0)\left(Y - \mu^{rc}_Y(\beta_0) - \mathbb{E}[w^{rc}_{cont,1}(\gamma_0)(Y - \mu^{rc}_Y(\beta_0))]\right)\dot{\pi}(\gamma_0)\right] \\
&- l_{ps}(W_i;\gamma_0)' \cdot \mathbb{E}\left[\alpha^{rc}_{ps,0}(\gamma_0)\left(Y - \mu^{rc}_Y(\beta_0) - \mathbb{E}[w^{rc}_{cont,0}(\gamma_0)(Y - \mu^{rc}_Y(\beta_0))]\right)\dot{\pi}(\gamma_0)\right].
\end{aligned}$$

The asymptotic normality result now follows immediately from (A.23) and the classical central limit theorem. ■

Proof of Theorem 5: The proof goes along the exact same lines as that of Theorem 3 and is therefore omitted. ■

B Closed-form Expressions

In this section, we first focus on the asymptotic results in Section 3, and give explicit expressions of functions $h^p(W;\kappa^p)$ and $h^{rc}(W;\kappa^{rc})$ before Assumption 5. Recall that

$$h^p(W;\kappa^p) = \left(w^p_{treat}(D) - w^p_{control}(D,X;\gamma)\right)\left(\Delta Y - \mu^p_\Delta(X;\beta^p_0,\beta^p_1)\right),$$

$$h^{rc}(W;\kappa^{rc}) = (w^{rc}_{treat}(D,T) - w^{rc}_{control}(D,T,X;\gamma))(Y - \mu^{rc}_Y(X,T;\beta^{rc}_0,\beta^{rc}_1),$$

where $\kappa^p = (\gamma',\beta^{p'}_0,\beta^{p'}_1)'$, and $\kappa^{rc} = (\gamma',\beta^{rc'}_0,\beta^{rc'}_1)'$; additionally, let $\kappa^p_0 = (\gamma'_0,\beta^{p'}_{0,0},\beta^{p'}_{1,0})'$ and $\kappa^{rc}_0 = (\gamma'_0,\beta^{rc'}_{0,0},\beta^{rc'}_{1,0})'$ denote the corresponding pseudo-true parameters. We next explicitly define $\dot{h}^p(W;\kappa^p) = \partial h^p(W;\kappa^p)/\partial\kappa^p$ and $\dot{h}^{rc}(W;\kappa^{rc}) = \partial h^{rc}(W;\kappa^{rc})/\partial\kappa^{rc}$.

Define

$$\alpha^p_{ps}(D,X;\gamma) = \frac{(1-D)}{(1-\pi(X;\gamma))^2}\bigg/\mathbb{E}\left[\frac{\pi(X;\gamma)(1-D)}{1-\pi(X;\gamma)}\right],$$

$$\alpha^{rc}_{ps,1}(D,T,X;\gamma) = \frac{(1-D)T}{(1-\pi(X;\gamma))^2}\bigg/\mathbb{E}\left[\frac{\pi(X;\gamma)(1-D)T}{1-\pi(X;\gamma)}\right],$$

$$\alpha^{rc}_{ps,0}(D,T,X;\gamma) = \frac{(1-D)(1-T)}{(1-\pi(X;\gamma))^2}\bigg/\mathbb{E}\left[\frac{\pi(X;\gamma)(1-D)(1-T)}{1-\pi(X;\gamma)}\right],$$

and let $\dot{\pi}(X;\gamma) = \partial\pi(X;\gamma)/\partial\gamma$, and $\dot{\mu}_t^p(X;\beta_t^p)$ and $\dot{\mu}_t^{rc}(X;\beta_t^{rc})$, $t = 0,1$, are defined analogously. Then, we have that

$$\dot{h}^p(W;\kappa^p) = \big(-\alpha_{ps}^p(D,X;\gamma)\left(\Delta Y - \mu_\Delta^p(X;\beta_0^p,\beta_1^p)\right)\dot{\pi}(X;\gamma)',$$

$$\left(w_{treat}^p(D) - w_{control}^p(D,X;\gamma)\right)\dot{\mu}_0^p(X;\beta_0^p)',$$

$$-\left(w_{treat}^p(D) - w_{control}^p(D,X;\gamma)\right)\dot{\mu}_1^p(X;\beta_0^p)'\big)',$$

and

$$\dot{h}^{rc}(W;\kappa^{rc}) = \big(-\left(\alpha_{ps,1}^{rc}(D,T,X;\gamma) - \alpha_{ps,0}^{rc}(D,T,X;\gamma)\right)(Y - \mu_Y^{rc}(X,T;\beta_0^{rc},\beta_1^{rc}))\,\dot{\pi}(X;\gamma)',$$

$$-\left(w_{treat}^{rc}(D,T) - w_{control}^{rc}(D,T,X;\gamma)\right)\dot{\mu}_0^{rc}(X;\beta_0^{rc})',$$

$$-\left(w_{treat}^{rc}(D,T) - w_{control}^{rc}(D,T,X;\gamma)\right)\dot{\mu}_1^{rc}(X;\beta_0^{rc})'\big)'.$$

Next, we derive the closed-form expression for the estimator $\widehat{\eta}^p(W;\widehat{\gamma},\widehat{\beta}_0^p,\widehat{\beta}_1^p)$ when panel data are available; in the repeated cross section data case, $\hat{\eta}^{rc}(W;\hat{\gamma},\widehat{\beta}_0^{rc},\widehat{\beta}_1^{rc})$ is defined analogously to $\widehat{\eta}^p(W;\widehat{\gamma},\widehat{\beta}_0^p,\widehat{\beta}_1^p)$, so we omit the details for brevity. We consider two scenarios: the first is when one assumes a linear outcome regression model and estimates $\beta_{t,0}^p$ $t = 0,1$, via ordinary least squares, and assumes a linear index model such as logit/probit and estimates the propensity score parameters γ_0 via maximum likelihood; the second case is when one assumes linear index models for all nuisance functions and estimates all parameters via maximum likelihood - this second case is particularly appealing when the outcome of interest is binary.

Firstly, let us consider the case where the propensity score model is of the form $\pi(X;\gamma) = G(X\gamma_0)$, where $G(\cdot)$ is a known link function - this setup allows one to use logit or probit models, for example. Given the binary nature of D, γ_0 can be estimated via maximum likelihood, that is,

$$\widehat{\gamma} = \arg\max_{\gamma}\sum_{i=1}^{n} D_i ln(G(X_i\gamma)) + (1-D_i)ln(1-G(X_i\gamma)).$$

When this model is correctly specified, under Assumption 4, the asymptotically linear representation is given by

$$\sqrt{n}(\widehat{\gamma}-\gamma_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}l_{ps}(W_i;\gamma_0)+o_p(1),$$

where

$$l_{ps}(W;\gamma)=\mathbb{E}\left[\frac{\dot{G}(X\gamma)^2}{G(X\gamma)(1-G(X\gamma))}X'X\right]^{-1}\frac{(D-G(X\gamma))\dot{G}(X\gamma)}{G(X\gamma)(1-G(X\gamma))}X',$$

and $\dot{G}(u)=\partial G(u)/\partial u$.

Then, we turn our attention to the estimation of the outcome regression functions $\mu_t^p(X;\beta_{t,0}^p)$, $t=0,1$, and their corresponding asymptotically linear representations $l_{reg,t}(W;\beta_{t,0}^p)$. We consider two cases: the first one is a linear regression model where unknown parameters are estimated using ordinary least squares; the second case is when Y is binary and one uses a linear index model such as logit or probit, and estimates the unknown parameters using maximum likelihood.

Let $n_0=\sum_{i=1}^{n}D_i=0$ be the sample size of the control group. When $\mu_t^p(X;\beta_{t,0}^p)=X\beta_{t,0}^p$ a.s., and we obtain the estimators $\widehat{\beta}_t^p$, $t=0,1$, via ordinary least squares,

$$\widehat{\beta}_t^p=\arg\min_{\beta_t^p}\sum_{i|D_i=0}(Y_{it}-X_i\beta_t^p)^2,$$

the asymptotically linear representation is given by

$$\sqrt{n}(\widehat{\beta}_t^p-\beta_{t,0}^p)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}l_{reg,t}(W_i;\beta_{t,0}^p)+o_p(1),$$

where, for $t=0,1$,

$$l_{reg,t}(W;\beta_t^p)=\mathbb{E}\left[(1-D)X'X\right]^{-1}X'(1-D)(Y_t-X\beta_{t,0}^p).$$

On the other hand, when Y is binary, $\mu_t^p(X;\beta_{t,0}^p)=\Lambda_t(X\beta_{t,0}^p)$ a.s. and one obtains the estimators $\widehat{\beta}_t^p$ for $t=0,1$ via maximum likelihood,

$$\widehat{\beta}_t^p=\arg\max_{\beta_t^p}\sum_{i|D_i=0}Y_{it}ln(\Lambda_t(X_i\beta_t^p))+(1-Y_{it})ln(1-\Lambda_t(X_i\beta_t^p)),$$

the asymptotically linear representation is given by

$$\sqrt{n}(\widehat{\beta}_t^p-\beta_{t,0}^p)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}l_{reg,t}(W_i;\beta_{t,0}^p)+o_p(1),$$

where, for $t = 0, 1$,

$$l_{reg,t}(W;\beta_t^p) = \mathbb{E}\left[\frac{(1-D)\dot{\Lambda}_t(X\beta_t^p)^2}{\Lambda_t(X\beta_t^p)(1-\Lambda_t(X\beta_t^p))}X'X\right]^{-1}\frac{(1-D)(Y_t-\Lambda_t(X\beta_t^p))\dot{\Lambda}_t(X\beta_t^p)}{\Lambda_t(X\beta_t^p)(1-\Lambda_t(X\beta_t^p))}X',$$

and $\dot{\Lambda}_t(u) = \partial\Lambda_t(u)/\partial u$.

Finally, we derive the closed form expression for the estimator $\widehat{\eta}^p(W;\widehat{\gamma},\widehat{\beta}_0^p,\widehat{\beta}_1^p)$ when panel data are available. We drop the dependence of all following functionals on W, and let

$$\begin{aligned}
\widehat{w}_{tr}^p &= \frac{D}{\mathbb{E}_n[D]},\\
\widehat{w}_{cont}^p(\widehat{\gamma}) &= \frac{G(X\widehat{\gamma})(1-D)}{1-G(X\widehat{\gamma})}\Bigg/\mathbb{E}_n\left[\frac{G(X\widehat{\gamma})(1-D)}{1-G(X\widehat{\gamma})}\right],\\
\widehat{\alpha}_{ps}^p(\hat{\gamma}) &= \frac{(1-D)}{(1-G(X\widehat{\gamma}))^2}\Bigg/\mathbb{E}_n\left[\frac{G(X\widehat{\gamma})(1-D)}{1-G(X\widehat{\gamma})}\right],
\end{aligned}$$

denote the estimated weights and the element in the estimation effect of $\widehat{\gamma}$. Let $\widehat{\mu}_\Delta^p(\widehat{\beta}) = X(\widehat{\beta}_1^p - \widehat{\beta}_0^p)$ when one uses OLS estimators for the outcomes, or $\widehat{\mu}_\Delta^p(\widehat{\beta}) = \Lambda_1(X\widehat{\beta}_1^p) - \Lambda_0(X\widehat{\beta}_0^p)$ when one uses binary maximum likelihood estimators for the outcomes.

One can easily obtain estimators for the asymptotically linear representation of the first-step estimators as $l_{ps}(W;\widehat{\gamma})$ and $l_{reg,t}(W;\widehat{\beta}_t^p)$, $t = 0, 1$, by replacing population expectations $\mathbb{E}[\cdot]$ with their sample counterparts $\mathbb{E}_n[\cdot]$, and by replacing unknown pseudo-true parameters γ_0 and $\beta_{t,0}^p$ with their respect estimators, $\widehat{\gamma}$ and $\widehat{\beta}_t^p$, $t = 0, 1$. Using the same principle, we can construct estimators for the derivatives as $\widehat{\dot{\pi}}(\widehat{\gamma}) = \dot{G}(X\widehat{\gamma})X'$, and for $t = 0, 1$, either $\widehat{\dot{\mu}}_t^p(\widehat{\beta}_t^p) = X'$ or $\widehat{\dot{\mu}}_t^p(\widehat{\beta}_t^p) = \dot{\Lambda}(X\widehat{\beta}_t^p)X'$, depending on which models for the outcomes are assumed.

Let $\widehat{\beta} = (\widehat{\beta}_1^{p'}, \widehat{\beta}_0^{p'})'$, $l_{reg}(\widehat{\beta}) = (l_{reg,1}(\widehat{\beta}_1^p)', l_{reg,0}(\widehat{\beta}_0^p)')'$, and $\widehat{\dot{\mu}}_\Delta^p(\widehat{\beta}) = (\widehat{\dot{\mu}}_1^p(\widehat{\beta}_t^p)', -\widehat{\dot{\mu}}_0^p(\widehat{\beta}_t^p)')'$. Then, the estimator of the asymptotically linear representation is given by

$$\widehat{\eta}^p\left(W;\widehat{\gamma},\widehat{\beta}\right) = \widehat{\eta}_{tr}^p\left(\widehat{\beta}\right) - \widehat{\eta}_{cont}^p\left(\widehat{\gamma},\widehat{\beta}\right) - \widehat{\eta}_{est}^p\left(\widehat{\gamma},\widehat{\beta}\right),$$

where

$$\begin{aligned}
\widehat{\eta}_{tr}^p\left(\widehat{\beta}\right) &= \widehat{w}_{tr}^p\left(\left(\Delta Y - \widehat{\mu}_\Delta^p\left(\widehat{\beta}\right)\right) - \mathbb{E}_n\left[\widehat{w}_{tr}^p\cdot\left(\Delta Y - \widehat{\mu}_\Delta^p\left(\widehat{\beta}\right)\right)\right]\right),\\
\widehat{\eta}_{cont}^p\left(\widehat{\gamma},\widehat{\beta}\right) &= \widehat{w}_{cont}^p(\widehat{\gamma})\left(\left(\Delta Y - \widehat{\mu}_\Delta^p(\widehat{\beta})\right) - \mathbb{E}_n\left[\widehat{w}_{cont}^p(\widehat{\gamma})\cdot\left(\Delta Y - \widehat{\mu}_\Delta^p(\widehat{\beta})\right)\right]\right),
\end{aligned}$$

and

$$\widehat{\eta}^{p}_{est}\left(\widehat{\gamma},\widehat{\beta}\right) = l_{reg}\left(\widehat{\beta}\right)' \cdot \mathbb{E}_n\left[\left(\widehat{w}^{p}_{tr} - \widehat{w}^{p}_{cont}(\widehat{\gamma})\right) \cdot \dot{\widehat{\mu}}^{p}_{\Delta}\left(\widehat{\beta}\right)\right]$$
$$+ l_{ps}(W;\widehat{\gamma}) \cdot \mathbb{E}_n\left[\widehat{\alpha}^{p}_{ps}(\widehat{\gamma})\left(\Delta Y - \widehat{\mu}^{p}_{\Delta}\left(\widehat{\beta}\right) - \mathbb{E}_n\left[\widehat{w}^{p}_{cont}(\widehat{\gamma}) \cdot \left(\Delta Y - \widehat{\mu}^{p}_{\Delta}(\widehat{\beta})\right)\right]\right) \cdot \dot{\widehat{\pi}}(\widehat{\gamma})\right].$$

C Monte Carlo Simulation

In this section, we conduct a small scale Monte Carlo simulation to illustrate the finite sample properties of the doubly robust difference-in-differences estimator with panel data. In particular, we compare the performance of our doubly robust (DR) difference-in-differences (DID) estimator described in (14) with three other DID estimators for the ATT: the two-way fixed effect estimator (FE) based on (5), the DID regression estimator (REG) based on (6), and Abadie (2005) IPW estimator (IPW) described in (9). We consider the case where a researcher models the outcome dynamics using a linear model (where all available covariates enter linearly), and estimates the unknown parameters using OLS. For the propensity score model, we consider a logistic model where all available covariates enter linearly, and the unknown parameters are estimated via maximum likelihood.

We consider sample sizes n equal to 100, 400, and 1600. For each design, we conduct $1,000$ Monte Carlo simulations, and calculate the average ATT (which can be used to estimate the average bias), the Monte Carlo standard deviation (MCSD), root mean square error (RMSE), average analytical standard error (AIFSE), average standard error based on the multiplier bootstrap (AMBSE), 95% coverage probability based on analytical standard errors (Cov) and based on the multiplier bootstrap procedure (BCov).

Our simulation design is largely based on Kang and Schafer (2007), but adapted to the DID setup. Let $\mathbf{X} = (X_1, X_2,\ X_3, X_4)'$ be distributed as $N(0, I_4)$, and I_4 be the 4×4 identity matrix. The true propensity score model is given by

$$p(\mathbf{X}) = \frac{\exp(-X_1 + 0.5X_2 - 0.25X_3 - 0.1X_4)}{1 + \exp(-X_1 + 0.5X_2 - 0.25X_3 - 0.1X_4)},$$

and the treatment status D is generated as $D = 1\{p(\mathbf{X}) \geq U\}$, where U follows a uniform $(0,1)$

distribution. The potential outcomes $Y_0 = Y_0(0) = Y_0(1), Y_1(1)$ and $Y_1(0)$ are given by

$$Y_0 = (1 + X_1 + X_2 + X_3 + X_4) + v + \varepsilon_0$$

$$Y_1(0) = 2(1 + X_1 + X_2 + X_3 + X_4) + v + \varepsilon_1(0)$$

$$Y_1(1) = 3(1 + X_1 + X_2 + X_3 + X_4) + v + \varepsilon_1(1),$$

where ε_0, $\varepsilon_1(0)$ and $\varepsilon_1(1)$ are independent $N(0,1)$ random variables, and the time-invariant unobserved heterogeneity v follows a normal distribution with mean $D \times (1 + X_1 + X_2 + X_3 + X_4)$ and variance 0.5. The true ATT is given by

$$ATT = \frac{\mathbb{E}[p(X)\mathbb{E}[Y_1(1) - Y_1(0)|X]]}{\mathbb{E}[D]} = 0.658.$$

Finally, we assume that, instead of observing $\{(Y_{0i}, Y_{1i}, D_i, \mathbf{X}_i)\}_{i=1}^n$, one only observes $\{(Y_{0i}, Y_{1i}, D_i, \mathbf{Z}_i)\}_{i=1}^n$, where $\mathbf{Z} = (Z_1, Z_2, Z_3, Z_4)'$ with $Z_1 = \exp(X_1/2)$, $Z_2 = X_2/(1 + \exp(X_1))$, $Z_3 = (X_1 X_3/25 + 0.6)^3$, and $Z_4 = (X_1 + X_4 + 20)^2$. In other words, instead of observing the "true" covariates, one only observes their nonlinear transforms.

We consider four different scenarios to assess the sensibility of the DID estimators under misspecified models that are "nearly correct". First, we consider the case where both propensity score and outcome regressions models are correctly specified. Second, we consider the case where only the propensity score model is correctly specified. Third, we analyze the situation where only the outcome regression models are correctly specified. Finally, we consider the case where none of the nuisance models is correctly specified. The simulation results are shown in Table C.1.

The results in Table C.1 corroborate the theoretical results in main text. First, the FE estimator is not consistent for the ATT: it displays large bias, and its associated coverage probability is very close to zero. As mentioned in the main text, this finding is not surprising since the FE regression implicitly assumes that treatment effects are homogeneous across covariate values, and when such an assumption is false (as is the case in this simulation exercise), the FE approach does not recover an easy to interpret causal parameter.

Second, provided that either the propensity score, or the outcome regression models are

Table C.1: Monte Carlo simulation results.

	π correct, μ correct.							π correct, μ incorrect.						
n=100, unfeasible ATT=0.660:														
	π correct, μ correct.							π correct, μ incorrect.						
	ATT	MCSD	RMSE	AIFSE	AMBSE	Cov	BCov	ATT	MCSD	RMSE	AIFSE	AMBSE	Cov	BCov
FE	-0.006	0.690	0.962	0.674	0.681	0.810	0.814	-0.006	0.690	0.962	0.927	0.681	0.935	0.812
REG	0.671	0.471	0.471	0.447	0.450	0.935	0.934	0.316	0.598	0.692	0.571	0.571	0.881	0.895
IPW	0.659	0.724	0.724	0.574	0.587	0.935	0.939	0.659	0.724	0.724	0.574	0.587	0.935	0.939
DR	0.653	0.520	0.520	0.452	0.453	0.919	0.922	0.573	0.628	0.635	0.539	0.538	0.911	0.911
	π incorrect, μ correct.							π incorrect, μ incorrect.						
	ATT	MCSD	RMSE	AIFSE	AMBSE	Cov	BCov	ATT	MCSD	RMSE	AIFSE	AMBSE	Cov	BCov
FE	-0.006	0.690	0.962	0.674	0.681	0.810	0.814	-0.006	0.690	0.962	0.927	0.681	0.935	0.812
REG	0.671	0.471	0.471	0.447	0.450	0.935	0.934	0.316	0.598	0.692	0.571	0.571	0.881	0.895
IPW	0.417	0.650	0.696	0.607	0.602	0.903	0.903	0.417	0.650	0.696	0.607	0.602	0.903	0.903
DR	0.662	0.502	0.502	0.453	0.453	0.926	0.927	0.374	0.638	0.701	0.580	0.573	0.879	0.891
n=400, unfeasible ATT=0.658:														
	π correct, μ correct.							π correct, μ incorrect.						
	ATT	MCSD	RMSE	AIFSE	AMBSE	Cov	BCov	ATT	MCSD	RMSE	AIFSE	AMBSE	Cov	BCov
FE	-0.020	0.349	0.769	0.342	0.342	0.490	0.490	-0.020	0.349	0.769	0.473	0.341	0.765	0.484
REG	0.663	0.219	0.219	0.224	0.223	0.947	0.952	0.340	0.279	0.429	0.286	0.286	0.778	0.785
IPW	0.669	0.297	0.297	0.285	0.288	0.946	0.951	0.669	0.297	0.297	0.285	0.288	0.935	0.951
DR	0.663	0.239	0.239	0.235	0.234	0.935	0.948	0.648	0.301	0.301	0.285	0.285	0.942	0.941
	π incorrect, μ correct.							π incorrect, μ incorrect.						
	ATT	MCSD	RMSE	AIFSE	AMBSE	Cov	BCov	ATT	MCSD	RMSE	AIFSE	AMBSE	Cov	BCov
FE	-0.020	0.349	0.769	0.342	0.342	0.490	0.490	-0.020	0.349	0.769	0.473	0.341	0.765	0.484
REG	0.663	0.219	0.219	0.224	0.223	0.947	0.952	0.340	0.279	0.429	0.286	0.286	0.778	0.785
IPW	0.415	0.304	0.394	0.294	0.295	0.860	0.849	0.415	0.304	0.394	0.294	0.295	0.860	0.849
DR	0.662	0.233	0.233	0.230	0.232	0.940	0.946	0.422	0.308	0.393	0.295	0.296	0.847	0.855
n=1600, unfeasible ATT=0.658:														
	π correct, μ correct.							π correct, μ incorrect.						
	ATT	MCSD	RMSE	AIFSE	AMBSE	Cov	BCov	ATT	MCSD	RMSE	AIFSE	AMBSE	Cov	BCov
FE	-0.014	0.172	0.700	0.171	0.170	0.017	0.022	-0.014	0.172	0.700	0.237	0.171	0.115	0.024
REG	0.661	0.112	0.112	0.112	0.112	0.955	0.948	0.333	0.141	0.361	0.143	0.142	0.354	0.361
IPW	0.667	0.154	0.154	0.143	0.147	0.954	0.940	0.667	0.154	0.154	0.143	0.147	0.954	0.940
DR	0.662	0.123	0.123	0.119	0.120	0.958	0.947	0.661	0.159	0.159	0.148	0.152	0.949	0.940
	π incorrect, μ correct.							π incorrect, μ incorrect.						
	ATT	MCSD	RMSE	AIFSE	AMBSE	Cov	BCov	ATT	MCSD	RMSE	AIFSE	AMBSE	Cov	BCov
FE	-0.014	0.172	0.700	0.171	0.170	0.017	0.022	-0.014	0.172	0.700	0.237	0.171	0.115	0.024
REG	0.661	0.112	0.112	0.112	0.112	0.955	0.948	0.333	0.141	0.361	0.143	0.142	0.354	0.361
IPW	0.410	0.147	0.295	0.146	0.145	0.598	0.581	0.410	0.147	0.295	0.146	0.145	0.598	0.581
DR	0.661	0.118	0.118	0.116	0.116	0.955	0.949	0.419	0.147	0.286	0.149	0.149	0.617	0.606

Notes: 1. Results are based on 1,000 Monte Carlo simulations. 2. Unfeasible ATTs are calculated as $1/n\sum_{i=1}^{n}(Y_1(1)_i - Y_1(0)_i|D_i = 1)$ using Monte Carlo simulated draws and the DGP described before. 3. MCSD is Monte Carlo standard deviation. 4. RMSE is root mean square error. 5. AIFSE is the average of influence function standard errors. 6. AMBSE is the average of multiplier bootstrap standard errors. 7. Cov is the coverage probability of 95% confidence intervals. 8. BCov is the bootstrapped coverage probability of 95% confidence intervals. 9. FE method uses (5). 10. REG uses (6). 11. IPW uses estimator (9). 13. DR estimator uses (14). 14. Use 999 bootstrap draws.

correctly specified, our proposed DR estimator for the ATT performs relatively well: it displays low bias, relatively small standard errors, and coverage probability similar to the nominal level. Note that the outcome regression models are correctly specified, the DID regression and the DID DR estimators are close to each other in all criteria. However, when the outcome regression model is misspecified but the propensity score model is correct, the DR DID estimator strictly dominates the DID regression.

When compared to the IPW estimator, we note that when the propensity score is correctly specified, IPW and DR estimators yield consistent ATT estimates, but DR standard errors are smaller than those of IPW methods. Thus, this simulation illustrates that our DR DID estimator can indeed provide some important efficiency gains when compared to IPW estimators.

Finally, the results in Table C.1 show that when the sample size increases, all standard errors decline, and coverages rise, when the corresponding model is correct. Although not included in this table, when n grows further, larger than 1600, coverage probabilities do converge to the nominal level of 0.95, regardless if one uses analytical or bootstrap based inference procedures

Overall, this small scale simulation exercise highlights the fact that when either the propensity score or the outcome regression models are correctly specified, our proposed DR DID estimator remains attractive for policy evaluations. On the other hand, two-way fixed effects, IPW and regression based DID estimators can lead to misleading conclusions about the effectiveness of a given policy in the presence of model misspecification.

References

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

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

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.