Introduction to Doubly Robust DiD

Sant'Anna and Zhao(2020)

M1 Inoue Jin

Hitotsubashi University

May 17, 2022

Doubly Robust DiD

- Doubly Robust DiD was proposed by Sant'Anna and Zhao(2020)
- This paper suggests doubly robust estimators for the ATT in DiD research designs.
- Authors also state that one can construct doubly robust DiD estimators for the ATT that are also doubly robust for inference.

Notation

- Y_{it} : the outcome of interest for unit i at time t
- t : researchers have access to outcome data in a pre-treatment period $t = 0$ and in a post-treatment period $t = 1$.
- $D_{it} = 1$ if i is treated before time t and $D_{it} = 0$ otherwise.
  - $D_{i0} = 0$ for every i at time t and then, $D_i = D_{i1}$
- $Y_{it} = D_i Y_{it}(1) + (1 - D_i) Y_{it}(0)$
  - $Y_{i0} = Y_{i0}(0)$ for all i
  - $Y_{i1} = D_i Y_{i1}(1) + (1 - D_i) Y_{i1}(0)$
- X_i :a vector of pre-treatment covariates.

The parameter of interest

- Difference in Differences identifies the average treatment effect for the treated (ATT).

The parameter of interest:ATT

$$\tau = \mathbb{E}[Y_{i1}(1) - Y_{i1}(0)|D_i = 1] = \mathbb{E}[Y_1|D = 1] - \mathbb{E}[Y_1(0)|D = 1]$$

Assumption 1.

Assumption 1. (Random Sampling)

the data $\{Y_{i0}, Y_{i1}, D_i, X_i\}_{i=1}^{n}$ are independent and identically distributed (iid)

Assumption 2.

Assumption 2. (Conditional PTA)

$\mathbb{E}[Y_1(0) - Y_0(0)|D = 1, X] = \mathbb{E}[Y_1(0) - Y_0(0)|D = 0, X]$ almost surely (a.s.).

- This assumption is called "Conditional Parallel Trend Assumption"
- It is crucial for most of DiD literature.

Assumption 3.

Assumption 3. (Overlap Condition)

For some $\varepsilon > 0$, $\mathbb{P}(D = 1) > \varepsilon$ and $\mathbb{P}(D = 1|X) \leq 1 - \varepsilon$ a.s.

- This assumption states that at least a small fraction of the population is treated and that for every value of the covariates X, there is at least a small probability that the unit is not treated.
- These assumptions are standard in conditional DID methods.

Doubly Robust DiD estimand:Notation

- $\pi(X)$: an arbitrary model for the true, unknown propensity score
- $\Delta Y = Y_1 - Y_0$:the difference of observed outcomes
- $\mu^p_{d,\Delta} \equiv \mu^p_{d,1}(X) - \mu^p_{d,0}(X)$
  - $\mu^p_{d,t}$: a model for the true, unknown outcome regression $m^p_{d,t} \equiv \mathbb{E}[Y_t | D = d, X = x]], d, t = 0, 1$

Doubly Robust DiD estimand

DR DID estimand when panel data are available

$$\tau^{dr,p} = \mathbb{E}[(w_1^p(D) - w_0^p(D,X;\pi))(\Delta Y - \mu_{0,\Delta}^p(X)))]$$

where, for a generic g,

$$w_1^p(D) = \frac{D}{\mathbb{E}[D]}, w_0^p(D,X;g) = \frac{\frac{g(X)(1-D)}{1-g(X)}}{\mathbb{E}[\frac{g(X)(1-D)}{1-g(X)}]}$$

Doubly Robust DiD estimand

Theorem 1.

Let Assumptions 1-3 hold. When panel data are available, $\tau^{dr,p} = \tau$ if either (but not necessarily both) $\pi(X) = p(X)$ a.s. or $\mu^p_\Delta(X) = m^p_{0,1}(X) - m^p_{0,0}(X)$

- This theorem states that at least one of the outcome regression model or propensity score model is correctly specified, we can recover the average treatment effect for the treated(ATT).
  - Case1:the propensity score model is correctly specified($\pi(X) = p(X) = P(D = 1|X)$), but the outcome regression model is miss-specified.
  - Case2:the outcome regression model is correcly specified($\mu^p_{0,\Delta}(X) = m^p_{0,\Delta}(X) = \mathbb{E}[Y_1|D = 0, X] - \mathbb{E}[Y_0|D = 0, X]$), but the propensity score model is miss-specified.

Proof of Theorem 1.(Case 1)

In Case 1, the propensity score model $\pi(X)$ is equivalent to the true propensity score $p(X) = P(D = 1|X)$.
To begin with, calculate the difference of two weights $w_1^p(D) - w_0^p(D, X)$:

$$
\begin{aligned}
w_1^p(D) - w_0^p(D, X) &= \frac{D}{\mathbb{E}[D]} - \frac{\frac{\pi(X)(1-D)}{1-\pi(X)}}{\mathbb{E}[\frac{\pi(X)(1-D)}{1-\pi(X)}]} \\
&= \frac{D}{\mathbb{E}[D]} - \frac{\frac{p(X)(1-D)}{1-p(X)}}{\mathbb{E}[D]} \\
&= \frac{1}{\mathbb{E}[D]}\left(\frac{(1-p(X))D}{1-p(X)} - \frac{(1-D)p(X)}{1-p(X)}\right) \\
&= \frac{D - p(X)}{\mathbb{E}[D](1-p(X))}
\end{aligned}
$$

Proof of Theorem 1.(Case 1)

Therefore,

$$\begin{aligned}
\tau^{dr,p} &= \mathbb{E}\left[(w_1^p(D) - w_0^p(D,X))(\Delta Y - \mu_{0,\Delta}^p(X))\right] \\
&= \mathbb{E}\left[\left(\frac{D - p(X)}{\mathbb{E}[D](1 - p(X))}\right)\left(\Delta Y - \mu_{0,\Delta}^p(X)\right)\right] \\
&= \mathbb{E}\left[\left(\frac{D - p(X)}{\mathbb{E}[D](1 - p(X))}\right)\Delta Y\right] - \mathbb{E}\left[\left(\frac{D - p(X)}{\mathbb{E}[D](1 - p(X))}\right)\mu_{0,\Delta}^p(X)\right]
\end{aligned}$$

The first term is equivalent to the Abadie(2005)'s IPW DID estimator. The second term seems to be "bias term" and we want this term to be zero.

Proof of Theorem 1.(Case 1)

Then, the bias term is:

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

The first line is obtained by the law of iterated expectation, and the third line reduces to 0 because $p(X)=P(D=1|X)=\mathbb{E}[D|X]$.

Proof of Theorem 1.(Case 1)

Next, check whether the IPW estimator is equivalent to ATT:

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

Note that $\Delta Y = Y_1 - Y_0 = DY_1(1) + (1-D)Y_1(0) - Y_0(0)$ (applying the equation of potential outcomes.)

Proof of Theorem 1.(Case 1)

We can calculate the second expectation by definition:

$$\mathbb{E}_D\left[\mathbb{E}\left[\frac{D-p(X)}{\mathbb{E}[D](1-p(X))}\left(DY_1(1)+(1-D)Y_1(0)-Y_0(0)\right)\middle|D,X\right]\middle|X\right]$$
$$=\frac{p(X)}{P(D=1)}\mathbb{E}\left[Y_1(1)-Y_0(0)|D=1,X\right]-\frac{p(X)}{P(D=1)}\mathbb{E}\left[Y_1(0)-Y_0(0)|D=0,X\right]$$
$$=\frac{p(X)}{P(D=1)}\{\mathbb{E}[Y_1(1)-Y_0(0)|D=1,X]-\mathbb{E}[Y_1(0)-Y_0(0)|D=1,X]\}$$
$$=\frac{p(X)}{P(D=1)}\mathbb{E}[Y_1(1)-Y_1(0)|D=1,X]$$

The first equality follows from the direct calculation of the outer conditional expectation, and we can obtain the second equality by applying the conditional PTA(Assumption 2.).

Proof of Theorem 1.(Case 1.)

Finally, ATT is recovered by direct calculation.

$$
\begin{aligned}
&\mathbb{E}_X\left[\frac{p(X)}{P(D=1)}\mathbb{E}[Y_1(1)-Y_1(0)|D=1,X]\right] \\
&= \int_x \frac{P(D=1|X)}{P(D=1)} \int_y (y_1(1)-y_1(0))\, f(y|d=1,x) f(x) dydx \\
&= \int_y \int_x \frac{f(d=1|x)}{f(d=1)} \frac{f(y,d=1,x)}{f(d=1,x)} f(x)\,(y_1(1)-y_1(0))\, dxdy \\
&= \int_y \int_x \frac{f(d=1|x)}{f(d=1)} \frac{f(y,d=1,x)f(x)}{f(d=1|x)f(x)}\,(y_1(1)-y_1(0))\, dxdy \\
&= \int_y \frac{y_1(1)-y_1(0)}{f(d=1)} \left\{ \int_x f(y,d=1,x)dx \right\} dy \\
&= \int_y (y_1(1)-y_1(0)) f(y|d=1) dy \\
&= \mathbb{E}[Y_1(1)-Y_1(0)|D=1]
\end{aligned}
$$

Estimation and Inference

DR DiD estimator

$$\widehat{\tau}^{dr,p} = \mathbb{E}_n \left[(\widehat{w}_1^p(D) - \widehat{w}_0^p(D, X; \widehat{\gamma}))(\Delta Y - \mu_{0,\Delta}^p(X; \widehat{\beta}_{0,0}^p, \widehat{\beta}_{0,1}^p)) \right],$$

where
$\widehat{w}_1^p(D) = \frac{D}{\mathbb{E}_n[D]}$, $\widehat{w}_0^p(D, X; \gamma) = \frac{\pi(X;\gamma)(1-D)}{1-\pi(X;\gamma)} \Big/ \mathbb{E}_n \left[\frac{\pi(X;\gamma)(1-D)}{1-\pi(X;\gamma)} \right]$

- $\widehat{\tau}^{dr,p}$ is doubly robust, and also locally semiparametrically efficient when the working models for the nuisance functions are correctly specified.(Theorem A.1 in Appendix A)

Semiparametric efficiency bound

- This result provides the semiparametric analogue of the Cramér-Rao lower bound commonly used in fully parametric procedures.
- Thus, this provides a benchmark that researchers can use to assess whether any given (regular) semiparametric DiD estimator works well.

Proposition 1

Let Assumptions 1-3 hold. Then, when panel data are available, the efficient influence function for the ATT is

$$\begin{aligned}\eta^{e,p}(Y_1, Y_0, D, X) = {} & w_1^p(D)(m_{1,\Delta}^p(X) - m_{0.\Delta}^p(X) - \tau) \\ & + w_1^p(D)(\Delta Y - m_{1,\Delta}^p(X)) - w_0^p(D, X; p)(\Delta Y - m_{0,\Delta}^p(X)),\end{aligned}$$

and the semiparametric efficiency bound for all regular estimators for the ATT is

$$\begin{aligned}\mathbb{E}[\eta^{e,p}(Y_1, Y_0, D, X)^2] = {} & \frac{1}{\mathbb{E}[D]^2}\mathbb{E}\Big[D(m_{1,\Delta}^p(X) - m_{0,\Delta}^p(X) - \tau)^2 \\ & + D(\Delta Y - m_{1,\Delta}^p(X))^2 + \frac{(1-D)p(X)^2}{(1-p(X))^2}(\Delta Y - m_{0.\Delta}^p(X))^2\Big]\end{aligned}$$

Note that $m_{0.\Delta}^p(X) \equiv m_{0,1}^p(X) - m_{0,0}^p(X)$

Estimation and Inference

- Practitioners need to choose a particular estimation procedure to be implemented.
- In this paper, authors propose the improved DR DiD estimator that can be easily implemented and is not only doubly robust for consistency but also doubly robust for inference.
- Doubly Robustness for inference practically means that the asymptotic variance of the DR DiD estimator for the ATT is invariant to which working models for the nuisance functions are correctly specified.
- In the following derivation of the improved DR DiD estimators, we consider the case where we use linear model for the outcome regression, logistic model for the propensity score, and use covariates X for all the nuisance models in a symmetric manner.

Estimation and Inference

As discussed above, we consider the following working models for the nuisance functions:

$$\pi(X;\gamma) = \Lambda(X'\gamma) \equiv \frac{exp(X'\gamma)}{1 + exp(X'\gamma)}, \text{ and } \mu_{0,\Delta}^{p} = \mu_{0,\Delta}^{lin,p} \equiv X'\beta_{0,\Delta}^{p}.$$

Improved DR DiD estimator

$$\widehat{\tau}_{imp}^{dr,p} = \mathbb{E}_n[(\widehat{w}_1^p(D) - \widehat{w}_0^p(D, X; \widehat{\gamma}^{ipt}))(\Delta Y - \mu_{0,\Delta}^{lin,p}(X; \widehat{\beta}_{0,\Delta}^{wls,p}))]$$

Estimation and Inference

the improved DR DiD estimator consists of three estimation procedures. The first two-steps consist of computing the optimization problems stated below.

$$\widehat{\gamma}^{ipt} = \arg\max_{\gamma \in \Gamma} \mathbb{E}_n[DX'\gamma - (1-D)exp(X'\gamma)],$$

$$\widehat{\beta}^{wls,p}_{0,\Delta} = \arg\min_{b \in \Theta} \mathbb{E}_n \left[\frac{\Lambda(X'\widehat{\gamma}^{ipt})}{1-\Lambda(X'\widehat{\gamma}^{ipt})} (\Delta Y - X'b)^2 \middle| D = 0 \right],$$

Then, in the third step, one plugs the fitted values of the logit model and linear model into the sample analogue of $\tau^{dr,p}$

Here, note that $\widehat{\gamma}^{ipt}$ is the inverse probability tilting estimator proposed by Graham et al.(2012) in a different context.

Estimation and Inference

Theorem 2.

Suppose Assumptions 1-3 and Assumption A.1-A.2 stated in Appendix A hold, and that the working models are logit model and linear regression model. Then,
(a) If either $\Lambda(X'\gamma^{*,ipt}) = p(X)$ a.s. or $X'\beta_{0,\Delta}^{*,wls,p} = m_{0,\Delta}^{p}(X)$ a.s., then, as $n \to \infty$,

$$\widehat{\tau}_{imp}^{dr,p} \xrightarrow{p} \tau$$

and

$$\sqrt{n}(\widehat{\tau}_{imp}^{dr,p} - \tau_{imp}^{dr,p}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_{imp}^{dr,p}(W; \gamma^{*,ipt}, \beta_{0,\Delta}^{*,wls,p}, \tau_{imp}^{dr,p}) + o_p(1)$$
$$\xrightarrow{d} N(0, V_{imp}^{p}),$$

where $V_{imp}^{p} = \mathbb{E}\left[\eta_{imp}^{dr,p}(W; \gamma^{*,ipt}, \beta_{0,\Delta}^{*,wls,p}, \tau_{imp}^{dr,p})^2\right]$.
(b) If both $\Lambda(X'\gamma^{*,ipt}) = p(X) a.s.$ and $X'\beta_{0,\Delta}^{*,wls,p} = m_{0,\Delta}^{p}(X) a.s.$, then $\eta_{imp}^{dr,p}(W; \gamma^{*,ipt}, \beta^{*,wls,p}, \tau_{imp}^{dr,p}) = \eta^{e,p}(Y_1, Y_0, D, X) a.s.$ and V_{imp}^{p} is equal to the semiparametric efficient bound.

Monte Carlo simulation study

- Monte Carlo experiments in order to study the finite sample property of DiD estimators including DR DiD estimator.
- $\widehat{\tau}^{dr,p}$, $\widehat{\tau}^{dr,p}_{imp}$:DR DiD estimators.
- $\widehat{\tau}^{reg} = \bar{Y}_{1,1} - \left[\bar{Y}_{1,0} + \frac{1}{n_{treat}} \sum_{i|D_i=1} (\widehat{\mu}_{0,1}(X_i) - \widehat{\mu}_{0,0}(X_i)) \right]$, where $\bar{Y}_{d,t} = \sum_{i|D_i=d,T_i=t} Y_{it}/n_{d,t}$ is the sample average outcome among units in treatment group d and time t, see e.g. Heckman et al.(1997).
- $\widehat{\tau}^{ipw,p} = \frac{1}{\mathbb{E}_n[D]} \mathbb{E}_n \left[\frac{D-p(X)}{1-p(X)} (Y_1 - Y_0) \right]$ see e.g. Abadie(2005).
- $\widehat{\tau}^{ipw,p}_{std} = \mathbb{E}_n[(\widehat{w}_1^p(D) - \widehat{w}_0^p(D, X; \widehat{\gamma}))(Y_1 - Y_0)]$
- $\widehat{\tau}^{fe}$:Two way fixed effect model.

Monte Carlo simulation study

- All observed covariates enter the working models linearly.
- n = 1000 for each design and conduct 10000 Monte Carlo simulations.

Simulation design when panel data are available

For a generic $W = (W_1, W_2, W_3, W_4)'$, let

$$f_{reg}(W) = 210 + 27.4 \cdot W_1 + 13.7 \cdot (W_2 + W_3 + W_4),$$
$$f_{ps}(W) = 0.75 \cdot (-W_1 + 0.5 \cdot W_2 - 0.25 \cdot W_3 - 0.1 \cdot W_4).$$

Let $\boldsymbol{X} = (X_1, X_2, X_3, X_4)'$ be distributed as $N(\mathbf{0}, I_4)$, and I_4 be the 4×4 identity matrix.

For $j = 1, 2, 3, 4$, let $Z_j = (\tilde{Z}_j - \mathbb{E}[\tilde{Z}_j]) \Big/ \sqrt{Var(\tilde{Z})}$, where

$\tilde{Z}_1 = exp(0.5X_1), \tilde{Z}_2 = 10 + X_2/(1 + exp(X_1)), \tilde{Z}_3 = (0.6 + X_1X_2/25)^3$ and $\tilde{Z}_4 = (20 + X_2 + X_4)^2$.

Monte Carlo simulation study

Data generating process(DGP)

$$Y_0(0) = f_{reg}(Z) + v(Z,D) + \varepsilon_0, \quad Y_1(d) = 2 \cdot f_{reg}(Z) + v(Z,D) + \varepsilon_1(d), d = 0,1,$$
$$p(Z) = \frac{exp(f_{ps}(Z))}{1 + exp(f_{ps}(Z))}, \quad D = 1\{p(Z) \geq U\} : DGP1$$

$$Y_0(0) = f_{reg}(Z) + v(Z,D) + \varepsilon_0, \quad Y_1(d) = 2 \cdot f_{reg}(Z) + v(Z,D) + \varepsilon_d, d = 0,1,$$
$$p(X) = \frac{exp(f_{ps}(X))}{1 + exp(f_{ps}(X))}, \quad D = 1\{p(X) \geq U\} : DGP2$$

$$Y_0(0) = f_{reg}(X) + v(X,D) + \varepsilon_0, \quad Y_1(d) = 2 \cdot f_{reg}(X) + v(X,D) + \varepsilon_1(d), d = 0,1,$$
$$p(Z) = \frac{exp(f_{ps}(Z))}{1 + exp(f_{ps}(Z))}, \quad D = 1\{p(Z) \geq U\} : DGP3$$

$$Y_0(0) = f_{reg}(X) + v(X,D) + \varepsilon_0, \quad Y_1(d) = 2 \cdot f_{reg}(X) + v(X,D) + \varepsilon_1(d), d = 0,1,$$
$$p(X) = \frac{exp(f_{ps}(X))}{1 + exp(f_{ps}(X))}, \quad D = 1\{p(X) \geq U\} : DGP4$$

Monte Carlo simulation study

Table 1
Monte Carlo results under designs $DGP1 - DGP4$ with panel data. Sample size $n = 1000$.

	DGP1: OR correct, PS correct Semiparametric efficiency bound: 11.1						DGP2: OR correct, PS incorrect Semiparametric efficiency bound: 11.6					
	Av. Bias	Med. Bias	RMSE	Asy. V	Cover	CIL	Av. Bias	Med. Bias	RMSE	Asy. V	Cover	CIL
$\widehat{\tau}^{fe}$	−20.952	−20.965	21.123	6392.2	0.000	9.906	−19.286	−19.287	19.468	6640.3	0.000	10.095
$\widehat{\tau}^{reg}$	−0.001	−0.001	0.100	10.2	0.950	0.396	−0.001	−0.001	0.100	10.1	0.949	0.394
$\widehat{\tau}^{ipw,p}$	0.026	0.195	2.774	8078.0	0.952	10.441	2.010	2.054	3.298	7048.3	0.838	9.819
$\widehat{\tau}^{ipw,p}_{std}$	0.008	−0.013	1.132	1286.4	0.948	4.309	−0.794	−0.798	1.225	891.7	0.856	3.623
$\widehat{\tau}^{dr,p}$	−0.001	0.000	0.106	11.1	0.947	0.412	−0.001	−0.002	0.104	10.7	0.947	0.404
$\widehat{\tau}^{dr,p}_{imp}$	−0.001	0.000	0.106	10.9	0.945	0.409	−0.001	−0.001	0.104	10.6	0.945	0.404
	DGP3: OR incorrect, PS correct Semiparametric efficiency bound: 11.1						DGP4: OR incorrect, PS incorrect Semiparametric efficiency bound: 11.6					
	Av. Bias	Med. Bias	RMSE	Asy. V	Cover	CIL	Av. Bias	Med. Bias	RMSE	Asy. V	Cover	CIL
$\widehat{\tau}^{fe}$	−13.170	−13.194	13.364	12 687.9	0.004	13.960	−16.385	−16.393	16.538	13 160.7	0.000	14.217
$\widehat{\tau}^{reg}$	−1.384	−1.365	1.868	1514.4	0.800	4.816	−5.204	−5.171	5.364	1666.6	0.015	5.053
$\widehat{\tau}^{ipw,p}$	0.011	0.158	3.198	10 062.5	0.947	11.777	−1.085	−1.017	2.656	6151.4	0.949	9.308
$\widehat{\tau}^{ipw,p}_{std}$	−0.030	−0.032	1.427	1988.0	0.945	5.484	−3.954	−3.949	4.215	2156.5	0.228	5.717
$\widehat{\tau}^{dr,p}$	−0.051	−0.046	1.214	1400.9	0.942	4.613	−3.188	−3.183	3.454	1704.9	0.308	5.075
$\widehat{\tau}^{dr,p}_{imp}$	−0.071	−0.064	1.015	971.2	0.942	3.858	−2.529	−2.514	2.720	970.1	0.274	3.856

Notes: Simulations based on 10,000 Monte Carlo experiments. $\widehat{\tau}^{fe}$ is the TWFE outcome regression estimator of τ^{fe} in (2.5), $\widehat{\tau}^{reg}$ is the OR-DID estimator (2.2), $\widehat{\tau}^{dr,p}$ is the IPW DID estimator (2.4), $\widehat{\tau}^{ipw,p}_{std}$ is the standardized IPW DID estimator (4.1), $\widehat{\tau}^{dr,p}$ is our proposed DR DID estimator (3.1), and $\widehat{\tau}^{dr,p}_{imp}$ is our proposed DR DID estimator (3.7). We use a linear OR working model and a logistic PS working model, where the unknown parameters are estimated via OLS and maximum likelihood, respectively, except for $\widehat{\tau}^{dr,p}_{imp}$, where we use the estimation methods described in Section 3.1. Finally, "Av. Bias", "Med. Bias", "RMSE", "Asy. V", "Cover" and "CIL', stand for the average simulated bias, median simulated bias, simulated root mean-squared errors, average of the plug-in estimator for the asymptotic variance, 95% coverage probability, and 95% confidence interval length, respectively. See the main text for further details.

Appendix A

Assumption A.1.

(i) $g(x) = g(x;\theta)$ is a parametric model, where $\theta \in \Theta \subset \mathbb{R}^k$, Θ being compact; (ii) $g(x;\theta)$ is *a.s.* continuous at each $\theta \in \Theta$; (iii) there exists a unique pseudo-true parameter $\theta^* \in int(\Theta)$; (iv) $g(x;\theta)$ is *a.s.* twice continuously differentiable in a neighborhood of $\theta^*, \Theta^* \subset \Theta$; ($v$) the estimator $\widehat{\theta}$ is strongly consistent for the θ^* and satisfies the following linear expansion:

$$\sqrt{n}(\widehat{\theta} - \theta^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l_g(W_i;\theta^*) + o_p(1),$$

where $l_g(\cdot;\theta)$ is such that $\mathbb{E}[l_g(W;\theta^*)] = 0$, $\mathbb{E}[l_g(W;\theta^*) l_g(W;\theta^*)']$ exists and is positive definite and $lim_{\delta \to \infty} \mathbb{E}\left[\sup_{\theta \in \Theta^* : \|\theta - \theta^*\| \leq \delta} \|l_g(W;\theta) - l_g(W;\theta^*)\|^2\right] = 0$. In addition, ($vi$) for some $\varepsilon > 0$, $0 < \pi(X;\gamma) \leq 1 - \varepsilon$ *a.s.*, for all $\gamma \in int(\Theta^{ps})$,where Θ^{ps} denotes the parameter space of γ.

Appendix A

Assumption A.2.

When panel data are available, assume that $\mathbb{E}[\|h^p(W;\kappa^{*,p})\|^2] < \infty$ and $\mathbb{E}[sup_{\kappa \in \Gamma^{*,p}}|\dot{h}^p(W;\kappa)|] < \infty$,where $\Gamma^{*,p}$ is a small neighborhood of $\kappa^{*,p}$.

Appendix A

Theorem A.1.

Suppose Assumptions 1-3 and Assumptions A.1-A.2 stated in Appendix A hold. Consider the following claims:

$\exists \gamma^* \in \Theta^{ps} : \mathbb{P}(\pi(X;\gamma^*) = p(X)) = 1,$
$\exists (\beta_{0,1}^{*,p}, \beta_{0,0}^{*,p}) \in \Theta_{0,1}^{reg} \times \Theta_{0,0}^{reg} : \mathbb{P}(\mu_{0,1}^{p}(X;\beta_{0,1}^{*,p}) - \mu_{0,0}^{p}(X;\beta_{0,0}^{*,p}) = m_{0,1}^{p}(X) - m_{0,0}^{p}(X)) = 1.$

(a) Provided that either one of claims stated above is true, as $n \to \infty$,

$$\widehat{\tau}^{dr,p} \xrightarrow{p} \tau.$$

Furthermore,

$$\begin{aligned} \sqrt{n}(\widehat{\tau}^{dr,p} - \tau^{dr,p}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta^p(W_i;\gamma^*, \beta_0^{*,p}) + o_p(1) \\ &\xrightarrow{d} N(0, V^p), \end{aligned}$$

where $V^p = \mathbb{E}[\eta^p(W;\gamma^*, \beta_0^{*,p})^2]$.
(b) When both of the claims above are true, $\eta^p(W;\gamma^*, \beta_0^{*,p}) = \eta^{e,p}(Y_1, Y_0, D, X) a.s.$ and V^p is equal to the semiparametric efficient bound.