Introduction to Doubly Robust DiD

Sant'Anna and Zhao(2020)

M1 Inoue Jin

Hitotsubashi University

May 16, 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-treated 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 (Random Sampling)

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

Assumption 2.

Assumption (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 (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:

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

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:

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

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

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

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

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

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

Estimation and Inference

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

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

where $V_{imp}^{p} = \mathbb{E}\left[\eta_{imp}^{dr,p}(W; \gamma^{*,ipt}, \beta_{0,\Delta}^{*,wls,p}, \tau_{imp}^{dr,p})^2\right]$.

Monte Carlo simulation study

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.