Doubly Robust Difference-in-Differences Estimators

Pedro H. C. Sant'Anna
Vanderbilt University

Jun Zhao
Vanderbilt University

Journal of Econometrics, 2020

Tools can be implemented via our companion R package, `DRDID`.

Difference-in-Differences: framework

- Framework:
  - 2 groups: $D = 0$ and $D = 1$
  - 2 times periods: $T = 0$ and $T = 1$
- Potential outcomes framework
  - $Y_{it}(1)$ is a treated potential outcome at time t
  - $Y_{it}(0)$ is an untreated potential outcome at time t
- (Y_{i1}, Y_{i0}) are a vector of outcomes at times $t = 1$ and $t = 0$ that are generated by

$$
\begin{aligned}
Y_{i1} &= D_i Y_{i1}(1) + (1 - D_i) Y_{i1}(0) \\
Y_{i0} &= Y_{i0}(0)
\end{aligned}
$$

- X_i is a vector of time invariant covariates
- D_i is a binary “treatment indicator”

Difference-in-Differences: framework

- Assume that either:
  - the data $\{Y_{i0}, Y_{i1}, D_i, X_i\}_{i=1}^{n}$ are independent and identically distributed (iid) **(panel data)**
  - conditional on $T = t$, $t = 0, 1$, the data are *iid* from the distribution of (Y_t, D, X), and $(D, X) \,|\, T = 0 \overset{d}{\sim} (D, X) \,|\, T = 1$ **(repeated cross-section stationary data)**
- Assume that, for some $\varepsilon > 0$, $\mathbb{P}(D = 1) > \varepsilon$ and $\mathbb{P}(D = 1|X) \equiv p(X) \leq 1 - \varepsilon$ a.s.
- Assume SUTVA holds

Difference-in-Differences: parameter of interest

- Parameter of interest:

$$\tau = ATT \equiv \underbrace{\mathbb{E}\left[Y_1(1) \mid D=1\right]}_{\text{estimable from the data}} - \underbrace{\mathbb{E}\left[Y_1(0) \mid D=1\right]}_{\text{counterfactual component}}$$

- Identification of the ATT achieved via the **Conditional Parallel Trends Assumption:**

$$\begin{aligned}
\mathbb{E}\left[Y_1(0) - Y_0(0) \mid D=1, X\right] &= \mathbb{E}\left[Y_1(0) - Y_0(0) \mid D=0, X\right] \\
&\Updownarrow \\
\mathbb{E}\left[Y_1(0) \mid D=1, X\right] &= \mathbb{E}\left[Y_0(0) \mid D=1, X\right] + \mathbb{E}\left[Y_1(0) - Y_0(0) \mid D=0, X\right]
\end{aligned}$$

Difference-in-Differences in practice

- Canonical DID without covariates:

$$\widehat{ATT_n} = \mathbb{E}_n\left[Y_1 - Y_0 | D = 1\right] - \mathbb{E}_n\left[Y_1 - Y_0 | D = 0\right]$$

- In the absence of covariates, i.e., rulling out covariate-specific trends,

$$Y_{i,t} = \alpha + \gamma D_i + \lambda 1\left\{t = 1\right\} + \underbrace{\beta}_{\equiv ATT}\left(D_i \cdot 1\left\{t = 1\right\}\right) + \varepsilon_{i,t}$$

But many times the PTA is only satisfied after conditioning on X!

DID and the Two-Way Fixed-Effects Linear Regression Model

- It may be tempting to "extrapolate" and use the TWFE linear regression model:

$$Y_{i,t} = \alpha + \gamma D_i + \lambda 1\{t=1\} + \underbrace{\beta}_{????} \cdot (D_i \cdot 1\{t=1\}) + \theta \cdot X_i + \varepsilon_{i,t}$$

- This TWFE specification severely restricts the DGP well beyond the conditional PTA:
  - $\mathbb{E}[Y_1 - Y_0 | D=1, X] = \mathbb{E}[Y_1 - Y_0 | D=1]$ (no X-specific trends for the treated group)
  - $\mathbb{E}[Y_1 - Y_0 | D=0, X] = \mathbb{E}[Y_1 - Y_0 | D=0]$ (no X-specific trends for the control group)
  - $\mathbb{E}[Y_1(1) - Y_1(0) | D=1, X] = \mathbb{E}[Y_1(1) - Y_1(0) | D=1]$ *a.s.* (homogeneous treatment effects)

TWFE estimator with Treatment Effect Heterogeneity

TWFE estimator with treatment effect heterogeneity

Semiparametric Difference-in-Differences

Clear separation between identification and estimation is key for causal inference

- Semiparametric DID estimators for the ATT can therefore have "many faces"

- For an overview, see e.g., Imbens and Wooldridge (2009) and Lechner (2010).

Semiparametric Difference-in-Differences

- Outcome regression estimator as in Heckman, Ichimura and Todd (1997):

$$\widehat{\tau}^{reg} = \bar{Y}_{1,1} - \left[\bar{Y}_{1,0} + \frac{1}{n_{treat}} \sum_{i|D_i=1} \left(\widehat{\mu}_{0,1}\left(X_i\right) - \widehat{\mu}_{0,0}\left(X_i\right)\right) \right],$$

$$\widehat{\tau}^{reg,2} = \frac{1}{n_{treat}} \sum_{i|D_i=1} \left[\left(\widehat{\mu}_{1,1}\left(X_i\right) - \widehat{\mu}_{1,0}\left(X_i\right)\right) - \left(\widehat{\mu}_{0,1}\left(X_i\right) - \widehat{\mu}_{0,0}\left(X_i\right)\right) \right].$$

- IPW estimator as in Abadie (2005):

$$\widehat{\tau}^{ipw,p} = \widehat{p}^{-1} \cdot \mathbb{E}_n \left[\frac{D - \widehat{\pi}\left(X\right)}{1 - \widehat{\pi}(X)} \left(Y_1 - Y_0\right) \right],$$

$$\widehat{\tau}^{ipw,rc} = \widehat{p}^{-1} \cdot \mathbb{E}_n \left[\frac{D - \widehat{\pi}\left(X\right)}{1 - \widehat{\pi}(X)} \frac{T - \widehat{\lambda}}{\widehat{\lambda}\left(1 - \widehat{\lambda}\right)} Y \right],$$

$\widehat{p} = \mathbb{E}_n\left[D\right],\ \widehat{\lambda} = \mathbb{E}_n\left[T\right].$


OUTCOME REGRESSION
PROPENSITY SCORE
JAKE-CLARK.TUMBLR
imgflip.com

First contribution of the paper: DR DID estimators

- **Combine both outcome regression and IPW** approaches to form more robust estimators

- Estimators are **Doubly Robust consistent**: they are consistent for the ATT if either (but not necessarily both)
  - Regression working model for outcome dynamics is correctly specified
  - Propensity score working model is correctly specified
- Available literature on DR methods mostly focuses on cross-section designs.
  - For an overview, see e.g., Słoczyński and Wooldridge (2018) and Seaman and Vansteelandt (2018)

How do the DR DID estimators look like?

How do the DR DID estimators look like?

- $\pi(X)$: an arbitrary model for the propensity score.

- **Panel data**:
  - $\Delta Y = Y_1 - Y_0$
  - $\mu^p_{d,\Delta}(X) \equiv \mu^p_{d,1}(X) - \mu^p_{d,0}(X)$, where $\mu^p_{d,t}(x)$ is a model for
$$m^p_{d,t}(x) \equiv \mathbb{E}[Y_t|D = d, X = x], t = 0, 1, d = 0, 1.$$

- **Repeated cross-section data**:
  - $\mu^{rc}_{d,\Delta}(X) \equiv \mu^{rc}_{d,1}(X) - \mu^{rc}_{d,0}(X)$, where $\mu^{rc}_{d,t}(x)$ is a model for
$$m^{rc}_{d,t}(x) \equiv \mathbb{E}[Y|D = d, T = t, X = x], t = 0, 1, d = 0, 1.$$
  - For $d = 0, 1$, let $\mu^{rc}_{d,Y}(T, X) \equiv T \cdot \mu^{rc}_{d,1}(X) + (1 - T) \cdot \mu^{rc}_{d,0}(X)$

DR DID estimand with panel data

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

where

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

DR DID estimand with repeated cross-section data

- We propose two different estimators with repeated cross-section data
- First estimator "mimics" the panel data one:

$$\tau_1^{dr,rc} = \mathbb{E}\left[\left(w_1^{rc}\left(D,T\right) - w_0^{rc}\left(D,T,X;\pi\right)\right)\left(Y - \mu_{0,Y}^{rc}(T,X)\right)\right],$$

where

$$\begin{aligned} w_1^{rc}\left(D,T\right) &= w_{1,1}^{rc}\left(D,T\right) - w_{1,0}^{rc}\left(D,T\right), \\ w_0^{rc}\left(D,T,X;g\right) &= w_{0,1}^{rc}\left(D,T,X;g\right) - w_{0,0}^{rc}\left(D,T,X;g\right), \end{aligned}$$

and, for $t = 0, 1$,

$$\begin{aligned} w_{1,t}^{rc}\left(D,T\right) &= \frac{D \cdot 1\left\{T = t\right\}}{\mathbb{E}\left[D \cdot 1\left\{T = t\right\}\right]}, \\ w_{0,t}^{rc}\left(D,T,X;g\right) &= \frac{g(X)\left(1-D\right) \cdot 1\left\{T = t\right\}}{1 - g(X)} \Bigg/ \mathbb{E}\left[\frac{g(X)\left(1-D\right) \cdot 1\left\{T = t\right\}}{1 - g(X)}\right]. \end{aligned}$$

DR DID estimand with repeated cross-section data

Second estimator relies on outcome regression models for the treated unit:

$$\tau_2^{dr,rc} = \tau_1^{dr,rc}$$

$$+ \left(\mathbb{E}\left[\mu_{1,1}^{rc}(X) - \mu_{0,1}^{rc}(X) \middle| D = 1\right] - \mathbb{E}\left[\mu_{1,1}^{rc}(X) - \mu_{0,1}^{rc}(X) \middle| D = 1, T = 1\right]\right)$$

$$- \left(\mathbb{E}\left[\mu_{1,0}^{rc}(X) - \mu_{0,0}^{rc}(X) \middle| D = 1\right] - \mathbb{E}\left[\mu_{1,0}^{rc}(X) - \mu_{0,0}^{rc}(X) \middle| D = 1, T = 0\right]\right),$$

DR DID estimands

First main result: Doubly Robust DID Estimands for the ATT

Let the assumptions we mentioned before hold. Then:

1. When panel data are available, $\tau^{dr,p} = \tau$ if either (but not necessarily both) $\pi(X) = p(X)$ *a.s.* or $\mu_{\Delta}^{p}(X) = m_{0,1}^{p}(X) - m_{0,0}^{p}(X)$ *a.s.*;

2. When repeated cross-section data are available, $\tau_1^{dr,rc} = \tau_2^{dr,rc} = \tau$ if either (but not necessarily both) $\pi(X) = p(X)$ *a.s.* or $\mu_{0,\Delta}^{rc}(X) = m_{0,1}^{rc}(X) - m_{0,0}^{rc}(X)$ *a.s.*.

Can we say something about efficiency?

Can we say something about efficiency?

- Development of DR estimators was initially motivated by the goal of improving efficiency
    - See e.g., Robins, Rotnitzky and Zhao (1994) and Ai and Chen (2003, 2007, 2012)
- Provide a benchmark that researchers can use to assess whether any given (regular) semiparametric DID estimator for the ATT is fully exploiting the empirical content of Assumptions we are making
- **Derive the semiparametric efficiency bound for the ATT in DID setups**
    - Panel vs. Repeated cross-section data
- We build on Newey (1990), Hahn (1998), and Chen, Hong and Tarozzi (2008).

Second main result: Semiparametric efficiency bound for the ATT in DID

Let the assumptions we mentioned before hold.

1. When **panel data** are available, the efficient influence function for the ATT is

$$\eta^{e,p}\left(Y_1, Y_0, D, X\right) = w_1^p\left(D\right)\left(m_{1,\Delta}^p\left(X\right) - m_{0,\Delta}^p\left(X\right) - \tau\right)$$

$$+ w_1^p\left(D\right)\left(\Delta Y - m_{1,\Delta}^p(X)\right) - w_0^p\left(D, X; p\right)\left(\Delta Y - m_{0,\Delta}^p(X)\right),$$

Second main result: Semiparametric efficiency bound for the ATT in DID (cont.)

Let the assumptions we mentioned before hold.

1. When **repeated cross-section data** are available, the efficient influence function for the ATT is

$$\eta^{e,rc}(Y,D,T,X) = \frac{D}{\mathbb{E}[D]}\left(m^{rc}_{1,\Delta}(X) - m^{rc}_{0,\Delta}(X) - \tau\right)$$

$$+ \left(w^{rc}_{1,1}(D,T)\left(Y - m^{rc}_{1,1}(X)\right) - w^{rc}_{1,0}(D,T)\left(Y - m^{rc}_{1,0}(X)\right)\right)$$

$$- \left(w^{rc}_{0,1}(D,T,X;p)\left(Y - m^{rc}_{0,1}(X)\right) - w^{rc}_{0,0}(D,T,X;p)\left(Y - m^{rc}_{0,0}(X)\right)\right),$$

Semiparametric efficiency: Panel vs. Repeated cross-section data

Assume that T is independent of (Y_1, Y_0, D, X), and some other regularity conditions hold. Then,

$$
\begin{aligned}
&\mathbb{E}\left[\eta^{e,rc}(Y, D, T, X)^2\right] - \mathbb{E}\left[\eta^{e,p}(Y_1, Y_0, D, X)^2\right] \\
&\quad = \frac{1}{\mathbb{E}[D]^2}\mathbb{E}\left[D\left(\sqrt{\frac{1-\lambda}{\lambda}}\left(Y_1 - m_{1,1}(X)\right) + \sqrt{\frac{\lambda}{1-\lambda}}\left(Y_0 - m_{1,0}(X)\right)\right)^2\right. \\
&+ \left.\frac{(1-D)\,p(X)^2}{(1-p(X))^2}\left(\sqrt{\frac{1-\lambda}{\lambda}}\left(Y_1 - m_{0,1}(X)\right) + \sqrt{\frac{\lambda}{1-\lambda}}\left(Y_0 - m_{0,0}(X)\right)\right)^2\right] \geq 0.
\end{aligned} \tag{1}
$$

Estimation and Inference

Estimation and inference when panel data are available

1. Estimate the propensity score $p(x)$ by $\pi(x;\widehat{\gamma})$, and the outcome regression models by $\mu^{p}_{0,t}\left(x;\widehat{\beta}^{p}_{0,t}\right)$, $t=0,1$.
   - We focus on the case where one uses parametric models (**curse of dimensionality**)

2. With the nuisance parameters on hands, estimate the *ATT* using the plug-in principle:

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

where

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

Inference when panel data are available

Under relatively weak regularity conditions (smoothness and existence of moments), provided that *at least one of the two models is correctly specified*,

$$\sqrt{n}\left(\widehat{\tau}^{dr,p} - ATT\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \eta^{p}(\mathcal{W}_i; \gamma^*, \beta_0^*) + o_p(1)$$

where, for a generic γ and $\beta_0 = \left(\beta_{0,1}^{'}, \beta_{0,0}^{'}\right)^{'}$, let

$$\eta^{p}\left(W; \gamma, \beta_0\right) = \eta_1^{p}\left(W; \beta_0\right) - \eta_0^{p}\left(W; \gamma, \beta_0\right) - \eta_{est}^{p}\left(W; \gamma, \beta_0\right), \tag{2}$$

where the estimation effect from the nuisance parameters is given by

$$\begin{aligned}\eta_{est}^{p}\left(W; \gamma, \beta_0\right) &= l_{reg}\left(\beta_0\right)' \cdot \mathbb{E}\left[\left(w_1^{p} - w_0^{p}\left(\gamma\right)\right) \cdot \dot{\mu}_{0,\Delta}^{p}\left(\beta_0\right)\right] \\ &+ l_{ps}\left(\gamma\right)' \cdot \mathbb{E}\left[\alpha_{ps}^{p}\left(\gamma\right)\left(\left(\Delta Y - \mu_{0,\Delta}^{p}\left(\beta_0\right)\right) - \mathbb{E}\left[w_0^{p}\left(\gamma\right) \cdot \left(\Delta Y - \mu_{0,\Delta}^{p}\left(\beta_0\right)\right)\right]\right) \cdot \dot{\pi}(\gamma)\right],\end{aligned} \tag{3}$$

Inference when panel data are available

- From the above asymptotic linear representation and the application of the CLT, we have

$$\sqrt{n}\left(\widehat{\tau}^{dr,p} - ATT\right) \xrightarrow{d} N(0, V^p)$$

where

$$\Sigma_p = \mathbb{E}[\eta^p(\mathcal{W})\eta^p(\mathcal{W})'].$$

- When both models are correctly specified, $\eta^p = \eta^{e,p}_{np}$
    - **DR DID estimator for the ATT is locally semiparametric efficient, when panel data are available**

Inference when repeated cross-section data are available

- When only repeated cross-section data are available, we have similar results

$$\sqrt{n}\left(\widehat{\tau}_1^{dr,rc} - ATT\right) \xrightarrow{d} N(0, V_1^{rc})$$

  and

$$\sqrt{n}\left(\widehat{\tau}_2^{dr,rc} - ATT\right) \xrightarrow{d} N(0, V_2^{rc})$$

- However, our proposed estimator $\widehat{\tau}_1^{dr,rc}$, in general, does not achieve the semiparametric efficiency bound when all nuisance models are correctly specified

- On the other hand, $\widehat{\tau}_2^{dr,rc}$ is locally semiparametrically efficient

“Price” of mimicking the panel data setup

When all working models are correctly specified and some weak regularity conditions are satisfied,

$$
\begin{aligned}
& V_1^{rc} - V_2^{rc} \\
&= p^{-1} \cdot Var\left[\sqrt{\frac{1-\lambda}{\lambda}} \left(m_{1,1}^{rc}(X) - m_{0,1}^{rc}(X) \right) + \sqrt{\frac{\lambda}{1-\lambda}} \left(m_{1,0}^{rc}(X) - m_{0,0}^{rc}(X) \right) \middle| D = 1 \right] \\
& \geq 0.
\end{aligned}
$$

Can we do even better?

Generality vs Specificity

- The aforementioned procedures are generic and can accomodate a rich family of first-step estimators
- Generality is always good, but sometimes it can be hard to choose

- **In some situations, we can do better by tailoring the first step estimation procedure**

Improved DR DID estimators

- **We propose "improved" DR DID estimators:**
  - Build on Graham, Pinto and Egel (2012), and Vermeulen and Vansteelandt (2015)
  - Assume that PS is logistic, OR is linear, and the same vector of covariates are used in both models
  - Goal is to "kill" the estimation effect, regardless of which model are correctly specified
  - Resulting estimator is also **DR for inference** (no need to adjust for estimation effect)

Improved DR DID estimators when panel data are available

- The first two steps are given by

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

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

  where

$$\pi\left(X,\gamma\right) = \Lambda\left(X'\gamma\right) = \frac{\exp\left(X'\gamma\right)}{1+\exp\left(X'\gamma\right)}.$$

  and

$$\mu_{\Delta}^{p}\left(X;\beta_0^p,\beta_1^p\right) = X'\beta^p$$

- Proposed improved DR DID estimator for the ATT is

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

Improved DR DID estimators when panel data are available

- The first order conditions for the IPT and WLS estimators are given by

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{D}{\mathbb{E}[D]} - \frac{\exp(X'\gamma^*)(1-D)}{\mathbb{E}[\exp(X'\gamma^*)(1-D)]}\right)X\right] &= 0, \\
\mathbb{E}\left[\exp(X'\gamma^*)\left(\Delta Y - \mu_{0,\Delta}^{lin,p}(X;\beta_{0,\Delta}^*)\right)X \,\middle|\, D=0\right] &= 0.
\end{aligned}
$$

- These two moment conditons imply that $\eta_{est}^{p}\left(W;\gamma^{*,ipt},\beta_{0,\delta}^{*,wls,p}\right) = 0$ a.s., regardless of which model is correctly specified.

- **Third main result of the paper: improved DR DID estimators are also DR for inference!**

Monte Carlo Simulations

Simulations

- Data generating processes are similar to those of Kang and Schafer (2007)
- We compare our proposed DR DID estimators with IPW (standardized and non-standardized), outcome regression, and TWFE estimators
- Samples sizes $n = 1,000$
- For each design, we consider $10,000$ Monte Carlo experiments
- Available data are $\{Y_1, Y_0, D, Z\}_{i=1}^{n}$, where $Y_1 = D \cdot Y_1(1) + (1 - D)\, Y_1(0)$ and $Y_0 = Y_0(0)$
- We estimate the pscore using a logit specification, and the OR assuming a linear specification

DGPs

- Let $Z_j = \left(\tilde{Z} - \mathbb{E}\left[\tilde{Z}\right]\right) / \sqrt{Var\left(\tilde{Z}\right)}$, $j = 1, 2, 3, 4$, where

$$
\begin{aligned}
\tilde{Z}_1 &= \exp\left(\frac{X_1}{2}\right) \\
\tilde{Z}_2 &= \frac{X_2}{1 + \exp\left(X_1\right)} + 10 \\
\tilde{Z}_3 &= \left(\frac{X_1 X_3}{25} + 0.6\right)^3 \\
\tilde{Z}_4 &= \left(X_2 + X_4 + 20\right)^2
\end{aligned}
$$

and $X_j \sim N\left(0, 1\right)$, $j = 1, 2, 3, 4$.

DGPs

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

$$f_{reg}(W) = 210 + 27.4 \cdot W_1 + 13.7 \cdot (W_2 + W_3 + W_4)$$

$$f_{ps}(W) = 0.75 \cdot (-W_1 + 0.5 \cdot W_2 - 0.25 \cdot W_3 - 0.1 \cdot W_4)$$

- Also

$$v(Z, D) \overset{d}{\sim} N(D \cdot f_{reg}(Z), 1)$$
$$v(X, D) \overset{d}{\sim} N(D \cdot f_{reg}(X), 1)$$
$$\varepsilon_0 \overset{d}{\sim} N(0, 1)$$
$$\varepsilon_1(1) \overset{d}{\sim} N(0, 1)$$
$$\varepsilon_1(0) \overset{d}{\sim} N(0, 1)$$
$$U \overset{d}{\sim} U(0, 1)$$

DGPs

- We consider 4 DGPs
- $DGP1$:

$$
\begin{aligned}
Y_{0,i}(0) &= f_{reg}(Z_i) + v_i(Z_i, D_i) + \varepsilon_{0,i} \\
Y_{1,i}(0) &= 2 \cdot f_{reg}(Z_i) + v_i(Z_i, D_i) + \varepsilon_{1,i}(0) \\
Y_{1,i}(1) &= 2 \cdot f_{reg}(Z_i) + v_i(Z_i, D_i) + \varepsilon_{1,i}(1) \\
p(Z) &= \frac{\exp(f_{ps}(Z_i))}{1 + \exp(f_{ps}(Z_i))} \\
D &= 1\{p(Z) \geq U\}
\end{aligned}
$$

DGPs

- *DGP2*:

$$
\begin{aligned}
Y_{0,i}(0) &= f_{reg}(Z_i) + v_i(Z_i, D_i) + \varepsilon_{0,i} \\
Y_{1,i}(0) &= 2 \cdot f_{reg}(Z_i) + v_i(Z_i, D_i) + \varepsilon_{1,i}(0) \\
Y_{1,i}(1) &= 2 \cdot f_{reg}(Z_i) + v_i(Z_i, D_i) + \varepsilon_{1,i}(1) \\
p(X_i) &= \frac{\exp(f_{ps}(X_i))}{1 + \exp(f_{ps}(X_i))} \\
D_i &= 1\{p(X_i) \geq U_i\}
\end{aligned}
$$

- Only OR model is correctly specified

DGPs

- *DGP3*:

$$
\begin{aligned}
Y_{0,i}(0) &= f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{0,i} \\
Y_{1,i}(0) &= 2 \cdot f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{1,i}(0) \\
Y_{1,i}(1) &= 2 \cdot f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{1,i}(1) \\
p(Z_i) &= \frac{\exp(f_{ps}(Z_i))}{1 + \exp(f_{ps}(Z_i))} \\
D_i &= 1\{p(Z_i) \geq U_i\}
\end{aligned}
$$

- Only pscore model is correctly specified

DGPs

- *DGP*4:

$$
\begin{aligned}
Y_{0,i}(0) &= f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{0,i} \\
Y_{1,i}(0) &= 2 \cdot f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{1,i}(0) \\
Y_{1,i}(1) &= 2 \cdot f_{reg}(X_i) + v_i(X_i, D_i) + \varepsilon_{1,i}(1) \\
p(X_i) &= \frac{\exp(f_{ps}(X_i))}{1 + \exp(f_{ps}(X_i))} \\
D_i &= 1\{p(X_i) \geq U_i\}
\end{aligned}
$$

- Both pscore and OR models are misspecified

Table 1: Monte Carlo Simulations, DGP1: Both pscore and OR are correctly specified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-20.9518	21.1227	2.5271	0.0000	9.9061
$\widehat{\tau}^{reg}$	-0.0012	0.1005	0.1010	0.9500	0.3960
$\widehat{\tau}^{ipw,p}$	0.0257	2.7743	2.6636	0.9518	10.4412
$\widehat{\tau}^{ipw,p}_{std}$	0.0075	1.1320	1.0992	0.9476	4.3090
$\widehat{\tau}^{dr,p}$	-0.0014	0.1059	0.1052	0.9473	0.4124
$\widehat{\tau}^{dr,p}_{imp}$	-0.0013	0.1057	0.1043	0.9451	0.4088

Figure 1: Monte Carlo for DID estimators, DGP1: Both pscore and OR are correctly specified

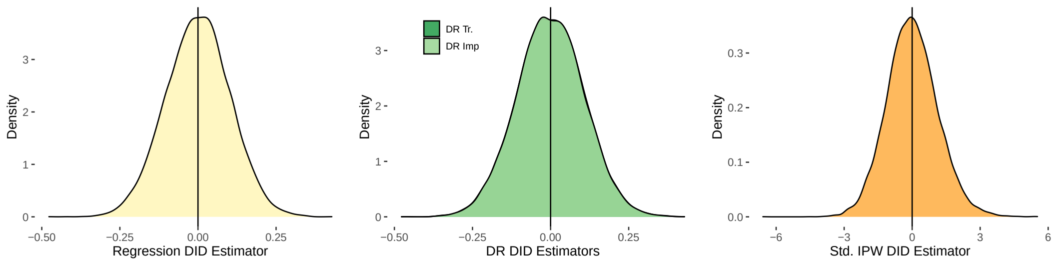


Table 2: Monte Carlo Simulations, DGP2: Only OR is correctly specified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-19.2859	19.4683	2.5754	0.0000	10.0955
$\widehat{\tau}^{reg}$	-0.0008	0.0997	0.1004	0.9492	0.3937
$\widehat{\tau}^{ipw,p}$	2.0100	3.2982	2.5049	0.8376	9.8193
$\widehat{\tau}^{ipw,p}_{std}$	-0.7942	1.2253	0.9241	0.8564	3.6226
$\widehat{\tau}^{dr,p}$	-0.0008	0.1036	0.1031	0.9469	0.4043
$\widehat{\tau}^{dr,p}_{imp}$	-0.0007	0.1042	0.1030	0.9445	0.4039

Figure 2: Monte Carlo for DID estimators, DGP2: Only OR is correctly specified

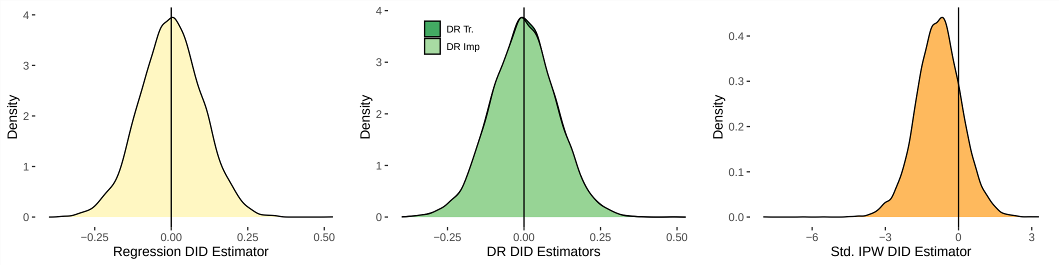


Table 3: Monte Carlo Simulations, DGP3: Only PS is correctly specified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-13.1703	13.3638	3.5611	0.0035	13.9596
$\widehat{\tau}^{reg}$	-1.3843	1.8684	1.2286	0.8001	4.8159
$\widehat{\tau}^{ipw,p}$	0.0114	3.1982	3.0043	0.9468	11.7769
$\widehat{\tau}^{ipw,p}_{std}$	-0.0299	1.4270	1.3990	0.9447	5.4840
$\widehat{\tau}^{dr,p}$	-0.0513	1.2142	1.1768	0.9416	4.6132
$\widehat{\tau}^{dr,p}_{imp}$	-0.0709	1.0151	0.9842	0.9423	3.8581

Figure 3: Monte Carlo for DID estimators, DGP3: Only PS is correctly specified

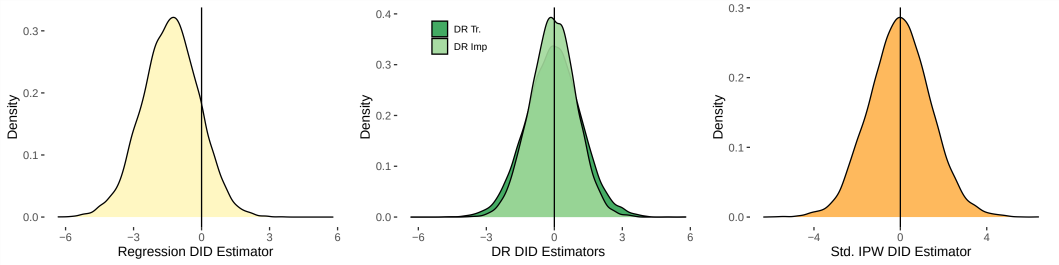
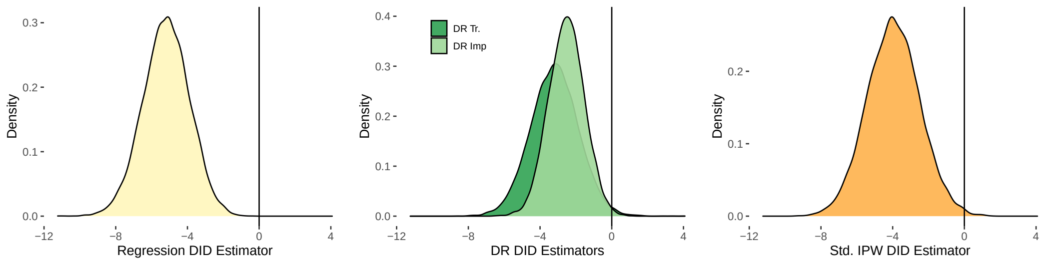


Table 4: Monte Carlo Simulations, DGP4: Both OR and PS are misspecified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-16.3846	16.5383	3.6268	0.0000	14.2169
$\widehat{\tau}^{reg}$	-5.2045	5.3641	1.2890	0.0145	5.0531
$\widehat{\tau}^{ipw,p}$	-1.0846	2.6557	2.3746	0.9487	9.3084
$\widehat{\tau}^{ipw,p}_{std}$	-3.9538	4.2154	1.4585	0.2282	5.7172
$\widehat{\tau}^{dr,p}$	-3.1878	3.4544	1.2946	0.3076	5.0749
$\widehat{\tau}^{dr,p}_{imp}$	-2.5291	2.7202	0.9837	0.2737	3.8561

Figure 4: Monte Carlo for DID estimators, DGP4: Both OR and PS are misspecified



Monte Carlo simulations for repeated cross-section data

- Same DGPs as before, but now, we observe a sample either from $T = 1$ or from $T = 0$ with probability 0.5.

Table 5: Monte Carlo Simulations, DGP1: Both pscore and OR are correctly specified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-20.7916	21.0985	3.5705	0.0002	13.9962
$\widehat{\tau}^{reg}$	0.0263	7.5878	7.5702	0.9510	29.6751
$\widehat{\tau}^{ipw,rc}$	-0.6619	55.9708	55.5516	0.9493	217.7621
$\widehat{\tau}^{ipw,rc}_{std}$	-0.0502	9.6477	9.5815	0.9487	37.5596
$\widehat{\tau}^{dr,rc}_{1}$	0.0129	3.0414	3.0340	0.9504	11.8934
$\widehat{\tau}^{dr,rc}_{2}$	0.0041	0.2159	0.2102	0.9441	0.8239
$\widehat{\tau}^{dr,rc}_{1,imp}$	0.0136	3.0413	3.0337	0.9507	11.8921
$\widehat{\tau}^{dr,rc}_{2,imp}$	0.0047	0.2163	0.2049	0.9371	0.8032

Figure 5: Monte Carlo for DID estimators, DGP1: Both pscore and OR are correctly specified

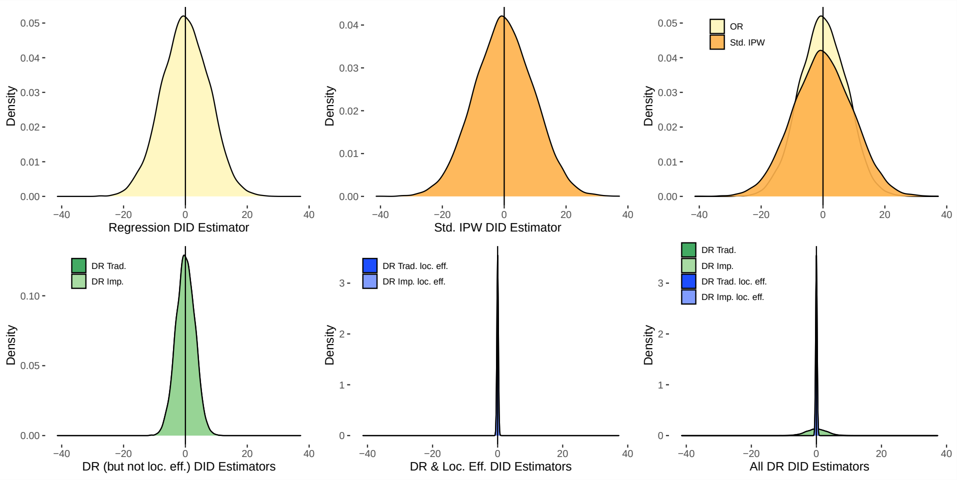


Figure 6: Monte Carlo for DID estimators, DGP1: Both pscore and OR are correctly specified

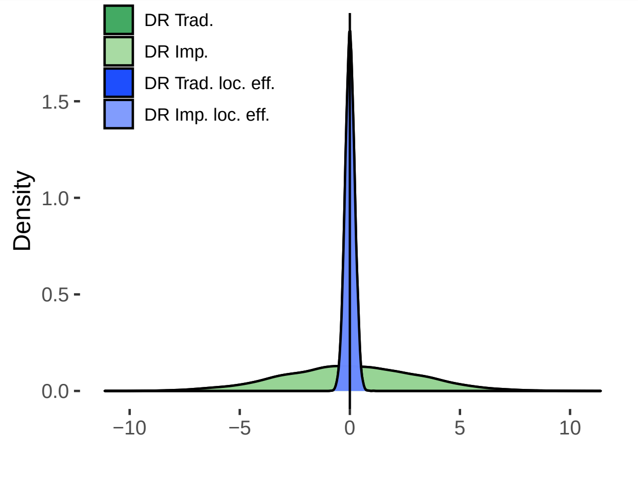


Figure 7: Monte Carlo for DID estimators, DGP1: Both pscore and OR are correctly specified

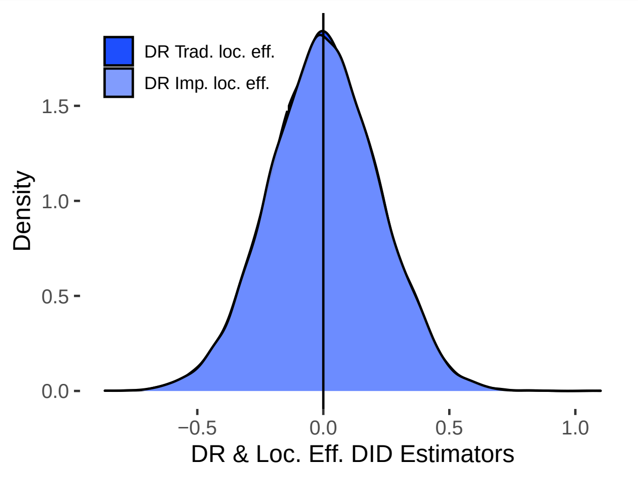


Table 6: Monte Carlo Simulations, DGP2: Only OR is correctly specified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-19.1783	19.5289	3.6345	0.0005	14.2472
$\widehat{\tau}^{reg}$	-0.0244	8.1906	8.1493	0.9481	31.9454
$\widehat{\tau}^{ipw,rc}$	1.8203	55.0496	54.9614	0.9491	215.4486
$\widehat{\tau}^{ipw,rc}_{std}$	-0.8119	9.8141	9.7018	0.9459	38.0310
$\widehat{\tau}^{dr,rc}_{1}$	-0.0102	3.2814	3.2651	0.9486	12.7991
$\widehat{\tau}^{dr,rc}_{2}$	-0.0002	0.2108	0.2054	0.9454	0.8051
$\widehat{\tau}^{dr,rc}_{1,imp}$	-0.0095	3.2818	3.2650	0.9488	12.7989
$\widehat{\tau}^{dr,rc}_{2,imp}$	0.0002	0.2127	0.2030	0.9403	0.7958

Figure 8: Monte Carlo for DID estimators, DGP2: Only OR is correctly specified

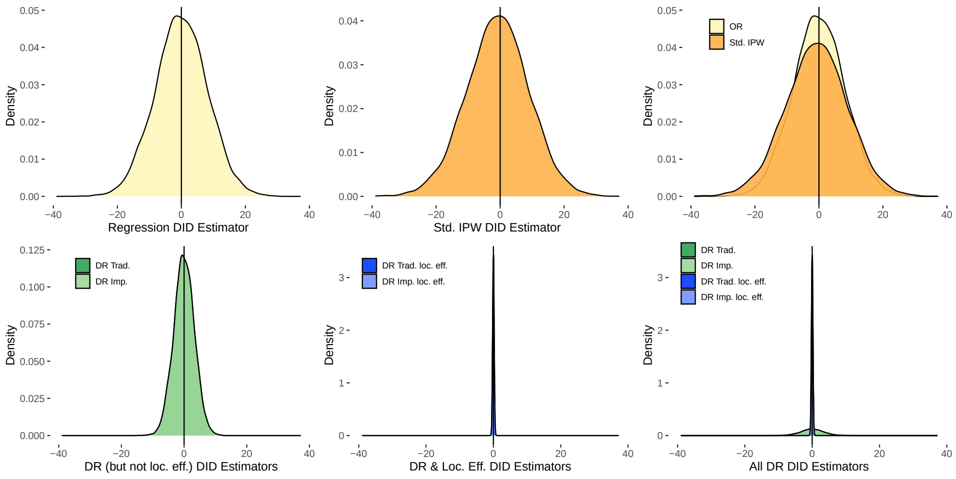


Figure 9: Monte Carlo for DID estimators, DGP2: Only OR is correctly specified

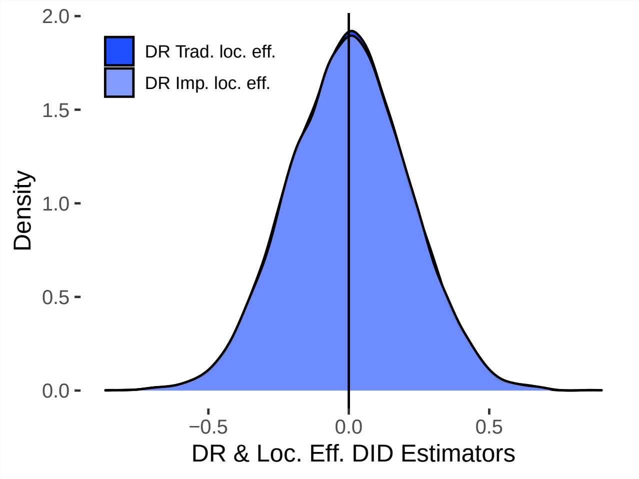


Table 7: Monte Carlo Simulations, DGP3: Only PS is correctly specified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-13.1310	14.0577	5.0424	0.2598	19.7664
$\widehat{\tau}^{reg}$	-1.3763	8.1367	8.0046	0.9421	31.3782
$\widehat{\tau}^{ipw,rc}$	-0.9734	57.2618	56.9005	0.9465	223.0501
$\widehat{\tau}^{ipw,rc}_{std}$	0.0508	9.4283	9.3068	0.9431	36.4826
$\widehat{\tau}^{dr,rc}_{1}$	-0.0855	5.6917	5.6276	0.9453	22.0602
$\widehat{\tau}^{dr,rc}_{2}$	-0.0289	4.7419	4.6585	0.9416	18.2613
$\widehat{\tau}^{dr,rc}_{1,imp}$	-0.1191	4.8371	4.7970	0.9450	18.8042
$\widehat{\tau}^{dr,rc}_{2,imp}$	-0.0762	4.0623	3.9669	0.9436	15.5503

Figure 10: Monte Carlo for DID estimators, DGP3: Only PS is correctly specified

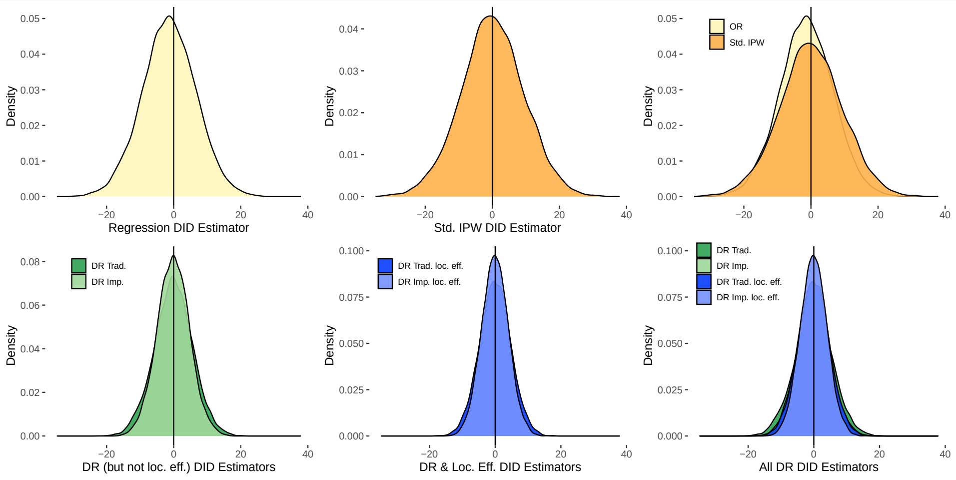


Figure 11: Monte Carlo for DID estimators, DGP3: Only PS is correctly specified

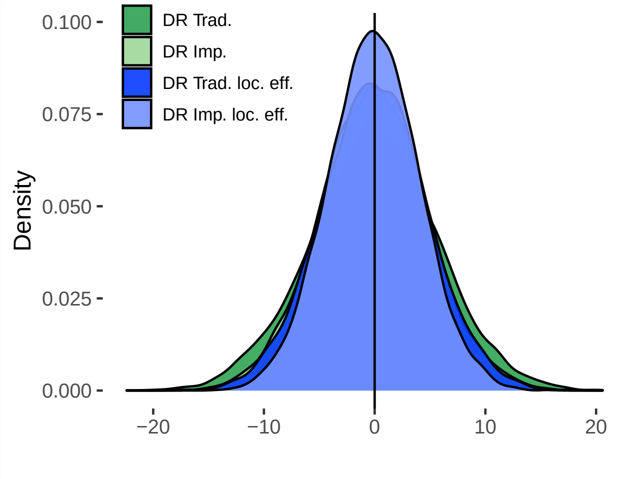


Figure 12: Monte Carlo for DID estimators, DGP3: Only PS is correctly specified

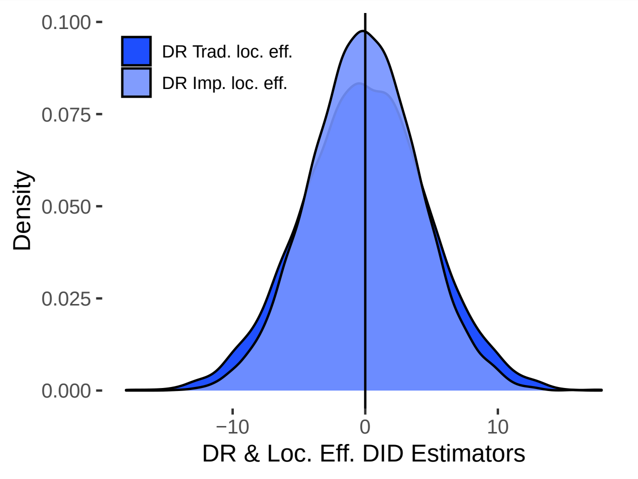


Table 8: Monte Carlo Simulations, DGP4: Both OR and PS are misspecified

	Bias	RMSE	Std. error	Coverage	CI length
$\widehat{\tau}^{fe}$	-16.3305	17.1263	5.1307	0.1138	20.1123
$\widehat{\tau}^{reg}$	-5.3378	9.9773	8.5196	0.9075	33.3969
$\widehat{\tau}^{ipw,rc}$	-1.3912	55.1777	55.6717	0.9518	218.2330
$\widehat{\tau}^{ipw,rc}_{std}$	-4.1487	10.5195	9.6864	0.9304	37.9707
$\widehat{\tau}^{dr,rc}_{1}$	-3.3422	7.0709	6.1963	0.9157	24.2897
$\widehat{\tau}^{dr,rc}_{2}$	-3.2751	6.0158	4.8876	0.8863	19.1593
$\widehat{\tau}^{dr,rc}_{1,imp}$	-2.6888	5.5642	4.8416	0.9134	18.9790
$\widehat{\tau}^{dr,rc}_{2,imp}$	-2.6138	4.8453	3.9673	0.8923	15.5519

Figure 13: Monte Carlo for DID estimators, DGP4: Both OR and PS are misspecified

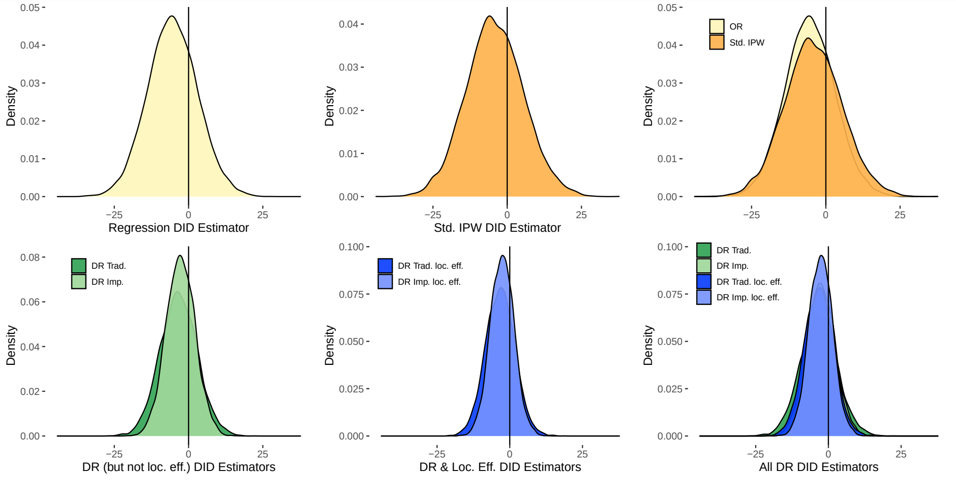


Figure 14: Monte Carlo for DID estimators, DGP4: Both OR and PS are misspecified

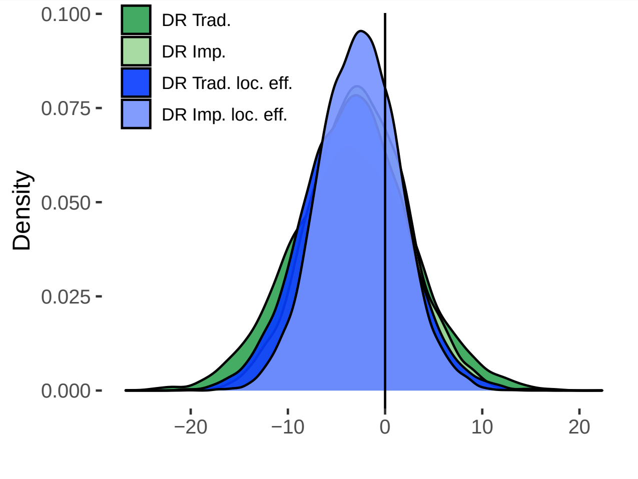
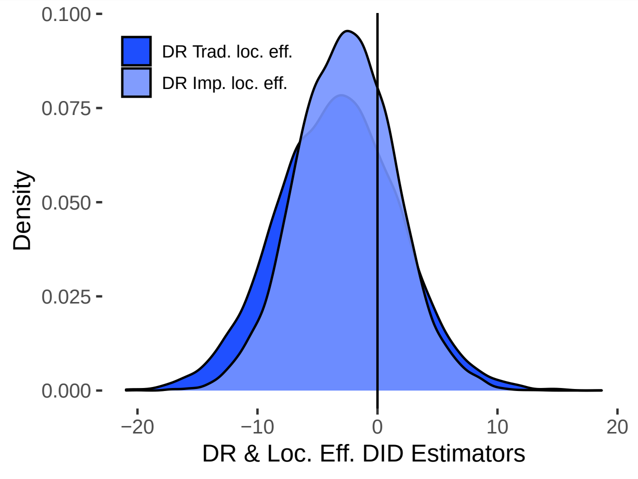


Figure 15: Monte Carlo for DID estimators, DGP4: Both OR and PS are misspecified



Empirical Illustration

The effect of job training on earnings

- Application inspired by LaLonde (1986), Dehejia and Wahba (1999) and Smith and Todd (2005)
- Use different DID estimators to assess which one does a "better" job reducing the "selection bias"
- Use randomized-out (control) individuals as pseudo-treated and CPS not-treated individuals as comparison units
- In the same spirit of Smith and Todd (2005), we consider different specifications, and different samples.

The effect of job training on earnings

Table 9: Results for Lalonde sample. NSW data with CPS comparison group.

	TWFE	REG	IPW	IPW.std	DR.tr	DR.imp
Linear Spec.	868	-1301	-1108	-1022	-871	-901
	(353)	(350)	(409)	(398)	(396)	(394)
	[98%]	[-147%]	[-125%]	[-115%]	[-98%]	[-102%]
DW Spec.	868	-830	-732	-564	-626	-591
	(359)	(360)	(534)	(487)	(496)	(467)
	[98%]	[-94%]	[-83%]	[-64%]	[-71%]	[-67%]
Aug. DW Spec.	868	-1041	-685	-558	-597	-599
	(352)	(358)	(523)	(485)	(491)	(470)
	[98%]	[-118%]	[-77%]	[-63%]	[-67%]	[-68%]

The effect of job training on earnings

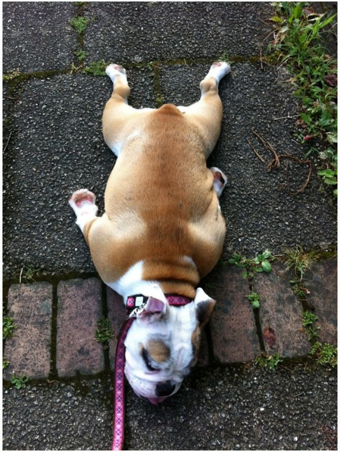
Table 10: Results for DW sample. NSW data with CPS comparison group.

	TWFE	REG	IPW	IPW.std	DR.tr	DR.imp
Linear Spec.	2092	-230	188	155	253	253
	(459)	(408)	(459)	(452)	(451)	(452)
	[117%]	[-13%]	[10%]	[9%]	[14%]	[14%]
DW Spec.	2092	402	-34	481	408	520
	(471)	(426)	(845)	(672)	(691)	(588)
	[117%]	[22%]	[-2%]	[27%]	[23%]	[29%]
Aug. DW Spec.	2092	27	97	502	514	524
	(458)	(428)	(793)	(653)	(663)	(582)
	[117%]	[2%]	[5%]	[28%]	[29%]	[29%]

The Effect of Job Training on Earnings

Table 11: Results for Early RA sample. NSW data with CPS comparison group.

	TWFE	REG	IPW	IPW.std	DR.tr	DR.imp
Linear Spec.	1136	-831	-516	-515	-434	-441
	(730)	(583)	(611)	(607)	(605)	(607)
	[41%]	[-30%]	[-19%]	[-19%]	[-16%]	[-16%]
DW Spec.	1136	-264	-495	-223	-246	-176
	(751)	(596)	(781)	(718)	(724)	(683)
	[41%]	[-10%]	[-18%]	[-8%]	[-9%]	[-6%]
Aug. DW Spec.	1136	-498	-337	-165	-148	-144
	(728)	(591)	(740)	(700)	(701)	(677)
	[41%]	[-18%]	[-12%]	[-6%]	[-5%]	[-5%]



Conclusion

Conclusion

- We proposed DR DID estimators for the ATT: remain consistent even if one (but not both) nuisance parameters are misspecified
- Proposed DR DID estimators achieve the semiparametric efficiency bound when both pscore and OR are correctly specified
- Proposed improved DR DID estimators applicable to some situations: Estimator is also DR for inference
- `R` **packgae** `DRDID` **available, so you can easily use our propoposed tools**