Better Understanding Triple Differences Estimators

Marcelo Ortiz-Villavicencio
Emory University

Pedro H.C. Sant’Anna
Emory University

Nov 1, 2024
Midwest Econometrics Group Conference 2024

Motivation

DiD Boom

- In the last few years, we have seen a big boom in DiD methods:
  - Role of covariates in DiD;
  - Variation in treatment timing;
  - Diagnostics for TWFE regressions;
  - Sensitivity Analysis for Violations of Parallel Trends (PT);
  - Non-binary treatments: Continuous and multi-valued treatments;
  - Nonlinear DiD.
- All these involve comparing groups of units with and without treatment across time.
- All these involve relying on a PT assumption between treated and untreated groups.

What about other DiD designs like Triple Differences?

Triple Differences

- Triple Differences (DDD) extend to cases where the PT assumption in DiD *may not hold*.

Triple Differences

- Triple Differences (DDD) extend to cases where the PT assumption in DiD *may not hold*.
  - **Ex**: PT can be violated due to the presence of a *time-varying confounder* that *changes differently across states*.

Triple Differences

- Triple Differences (DDD) extend to cases where the PT assumption in DiD *may not hold*.
  - **Ex**: PT can be violated due to the presence of a *time-varying confounder* that *changes differently across states*.
- When the PT assumption is questionable, researchers often augment the design by adding another *placebo* comparison group to ”clean” the bias introduced by the confounder.

Triple Differences

- Triple Differences (DDD) extend to cases where the PT assumption in DiD *may not hold*.
  - **Ex**: PT can be violated due to the presence of a *time-varying confounder* that *changes differently across states*.
- When the PT assumption is questionable, researchers often augment the design by adding another *placebo* comparison group to ”clean” the bias introduced by the confounder.
- DDD designs address this issue by finding a *within-state comparison group* that **is not exposed** to the treatment but is **affected** by the time-varying confounder.

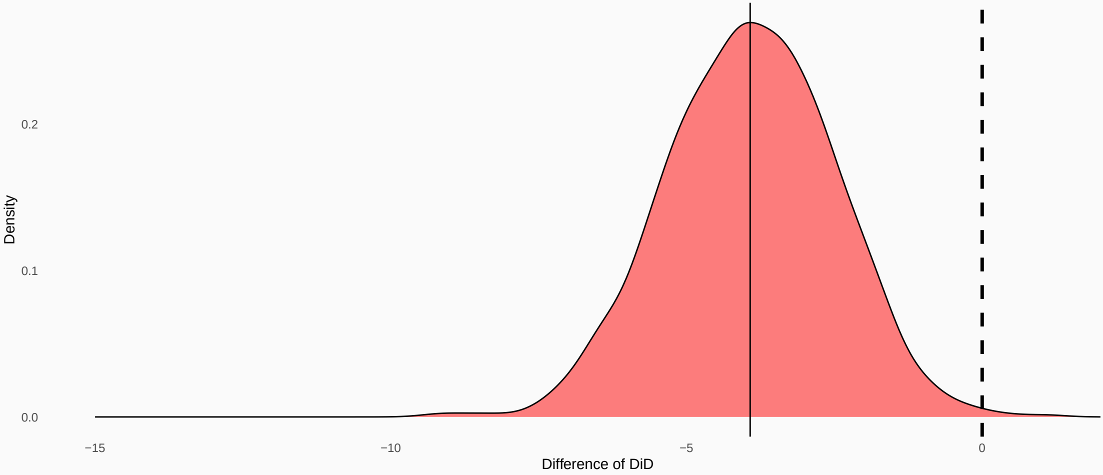
But… is not just a matter of adding another interaction term?

But… is not just a matter of adding another interaction term?

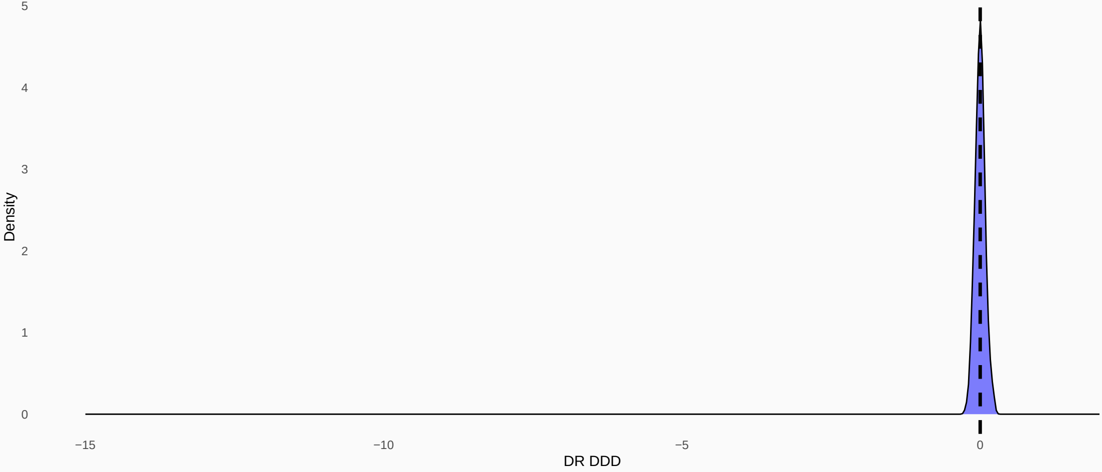
Is not DDD just a difference of 2 DiD?

Challenging folks wisdom: 3WFE?

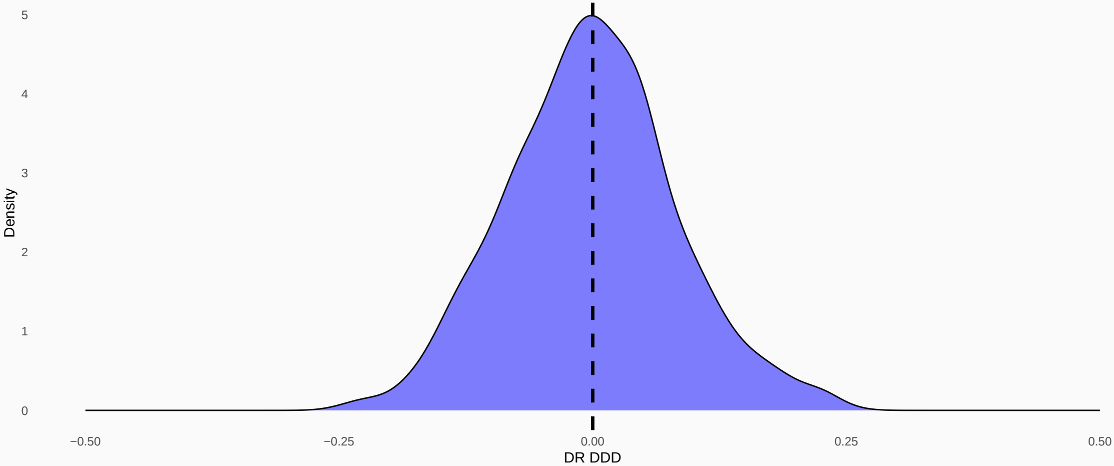
Challenging folks wisdom: Difference of 2 DiD?



Challenging folks wisdom: DR DDD!



Challenging folks wisdom: DR DDD!



Contribution

- Although it is also widely used in empirical work, DDD hasn't received as much attention as DiD.
- The key question in this paper is: How can we leverage our DiD knowledge to approach DDD?
  - We study identification, estimation, and inference procedures for DDD designs.
  - We derive the semiparametric efficiency bound for DDD designs and demonstrate that DDD estimators using a doubly robust representation reach this bound.
  - We extend our framework to staggered DDD designs.

Running Example: Muralidharan and Prakash (2017)

- **Program**: Impact of giving bicycles on girls' *secondary school enrollment* in Bihar, India.
- **Pre-program gap**:
  - Enrollment declines as the distance to school increases.
  - Higher attrition rates for women compared to men before the program.
- **DDD to the rescue**:
  - PT Assumption challenged by concurrent economic growth and *increased education spending* in Bihar, unlike the control state, Jharkhand.
  - Boys in Bihar *are not* exposed to the policy but are *affected* by the expansion in education spending.

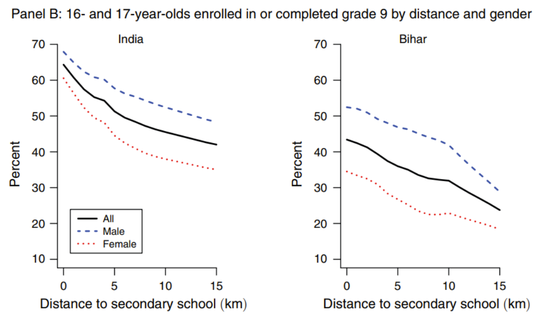


Fig. 1 Distance on Female Enrollment

What's the appeal of DDD compared to DiD?

Putting everything together, in DDD we allow to use all the information to control for location-specific trends and partition-specific trends, which otherwise would arise questionable results using DiD.

	No Eligible	Eligible
Treated	(Bihar, boys)	(Bihar, girls)
Control	(Jharkhand, boys)	(Jharkhand, girls)

Framework

Some notation

We have access to a sample of n units available, $i = 1, 2, \cdots, n$

- T time periods: $t = 1, 2, \ldots, T$.
- Different groups adopt a policy in different time periods g. Let $G \in \mathcal{G} \subset \{2, ..., T\} \cup \{\infty\}$ denote the time when group g is first-adopt the policy, with the notion that if a group is *"never-treated"*, $G = \infty$.
- Within each set of groups, we have two partitions (defined by some well-known criterion), $\ell \in P \equiv \{0, 1\}$. This determines *eligibility status*.
- Let $D_i \in \mathcal{D} \subset \{2, ..., T\} \cup \{\infty\}$ denote the time unit i is first-treated, with the notion that if a unit is *"never-treated"*, $D_i = \infty$.
- Note that $\mathcal{D} = \mathcal{G}$ such that:

$$D = \begin{cases} d & \text{if } G = g \equiv d \wedge P = 1, \\ \infty & \text{if } (G = g \wedge P = 0) \text{ or } (G = g' \wedge P \in \{0, 1\}, \text{ with } g' > g) \end{cases}$$

Ex: Units in $G = 2$ with $P = 1$ are treated at time $D = 2$, otherwise the unit remains untreated ($D = \infty$).

Building block of the analysis

- Let $Y_{i,t}(d)$ be the potential outcome for unit i, at time t, if this unit is first treated at time period d.
- A parameter that is interesting and has clear economic interpretation is the $ATT(g,t)$ (Callaway and Sant'Anna, 2021).

Definition (Parameter of interest: ATT(g,t))

Average Treatment Effect at time t of starting treatment at time g, among the units that indeed started treatment at time g.

$$ATT(g,t) := \mathbb{E}\left[Y_t(d) - Y_t(\infty) | G = g, P = 1\right], \textit{ for } t \geq g.$$

Building block of the analysis

- Let $Y_{i,t}(d)$ be the potential outcome for unit i, at time t, if this unit is first treated at time period d.
- A parameter that is interesting and has clear economic interpretation is the $ATT(g,t)$ (Callaway and Sant'Anna, 2021).

Definition (Parameter of interest: ATT(g,t))

Average Treatment Effect at time t of starting treatment at time g, among the units that indeed started treatment at time g.

$$ATT(g,t) := \mathbb{E}\left[Y_t(d) - Y_t(\infty) | G = g, P = 1\right], \textit{ for } t \geq g.$$

- Then, our identification problem comes from the fact that we never observe $\mathbb{E}\left[Y_t\left(\infty\right)|G = g, P = 1\right]$ in $t \geq g$.

Recovering the ATT using 3WFE Regression

- For simplicity, let us focus on the canonical 2x2x2 DDD. Thus, $G = \{2, \infty\}$ and $P = \{0, 1\}$.
- Let θ be the ATT.
- When there are only 2 time periods and *no covariates*, the following three-way fixed-effects (3WFE) regression specification can be used to recover the ATT:

$$
\begin{aligned}
Y_{i,t} = \; & \alpha_0 + \gamma_{0,1} 1_{\{G_i=2\}} + \gamma_{0,2} 1_{\{P_i=1\}} + \gamma_{0,3} 1_{\{T_i=2\}} \\
& + \gamma_{0,4} 1_{\{G_i=2\}} 1_{\{P_i=1\}} + \gamma_{0,5} 1_{\{G_i=2\}} 1_{\{T_i=2\}} + \gamma_{0,6} 1_{\{P_i=1\}} 1_{\{T_i=2\}} \\
& + \beta_0^{3wfe} 1_{\{G_i=2\}} 1_{\{P_i=1\}} 1_{\{T_i=2\}} + \varepsilon_{i,t},
\end{aligned}
$$

- One can show that $\beta_0^{3wfe} = \theta$ (Olden and Møen, 2022).

Recovering the ATT using 3WFE Regression

- For simplicity, let us focus on the canonical 2*x*2*x*2 DDD. Thus, $G = \{2, \infty\}$ and $P = \{0, 1\}$.
- Let θ be the ATT.
- When there are only 2 time periods and *no covariates*, the following three-way fixed-effects (3WFE) regression specification can be used to recover the ATT:

$$
\begin{aligned}
Y_{i,t} \quad = \quad & \alpha_0 + \gamma_{0,1}1_{\{G_i=2\}} + \gamma_{0,2}1_{\{P_i=1\}} + \gamma_{0,3}1_{\{T_i=2\}} \\
& +\gamma_{0,4}1_{\{G_i=2\}}1_{\{P_i=1\}} + \gamma_{0,5}1_{\{G_i=2\}}1_{\{T_i=2\}} + \gamma_{0,6}1_{\{P_i=1\}}1_{\{T_i=2\}} \\
& +\beta_0^{3wfe}1_{\{G_i=2\}}1_{\{P_i=1\}}1_{\{T_i=2\}} + \varepsilon_{i,t},
\end{aligned}
$$

- One can show that $\beta_0^{3wfe} = \theta$ (Olden and Møen, 2022).

In general, one can recover θ in the canonical DDD either (i) by a saturated 3WFE regression or (2) by a difference of 2 DiDs.

Difference of 2 DiDs

What happens when covariates play an important role?

Assumptions

- Adding covariates in the above 3WFE specification would imply additional restrictions to the DGP:
  - Homogeneous treatment effects in covariates.
  - Rule out covariate-specific trends in both the treated and comparison groups.

Assumptions

- Adding covariates in the above 3WFE specification would imply additional restrictions to the DGP:
  - Homogeneous treatment effects in covariates.
  - Rule out covariate-specific trends in both the treated and comparison groups.

- Our goal is to introduce an estimator for DDD under the condition that the PT assumption is valid after controlling for covariates, i.e., $X \in \mathcal{X} \subseteq \mathrm{R}^d$.

Assumptions

- Adding covariates in the above 3WFE specification would imply additional restrictions to the DGP:
  - Homogeneous treatment effects in covariates.
  - Rule out covariate-specific trends in both the treated and comparison groups.

- Our goal is to introduce an estimator for DDD under the condition that the PT assumption is valid after controlling for covariates, i.e., $X \in \mathcal{X} \subseteq \mathrm{R}^d$.

Assumption (Conditional Parallel Trends Assumption for DDD)

$$\mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | G = 2, P = 1, X\right] - \mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | G = 2, P = 0, X\right]$$
$$=$$
$$\mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | G = \infty, P = 1, X\right] - \mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | G = \infty, P = 0, X\right] \ a.s.$$

Assumptions

- Adding covariates in the above 3WFE specification would imply additional restrictions to the DGP:
  - Homogeneous treatment effects in covariates.
  - Rule out covariate-specific trends in both the treated and comparison groups.

- Our goal is to introduce an estimator for DDD under the condition that the PT assumption is valid after controlling for covariates, i.e., $X \in \mathcal{X} \subseteq \mathrm{R}^d$.

Assumption (Conditional Parallel Trends Assumption for DDD)

$$\begin{aligned}
\mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | G=2, P=1, X\right] \quad &- \quad \mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | G=2, P=0, X\right] \\
&= \\
\mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | G=\infty, P=1, X\right] \quad &- \quad \mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | G=\infty, P=0, X\right] \; a.s.
\end{aligned}$$

Additional Assumptions: Strong Overlap & No Anticipation.

Doubly Robust and Semiparametric Efficiency

- Under the previous assumptions, the ATT can be identified via Regression Adjustments or IPW or any convex combination between them. RA & IPW

- However, this depends on the researcher's ability to accurately model outcome regression or propensity scores.
  - How would you choose this combination if the goal was to achieve **Doubly Robustness** (DR)?
  - How would you choose this combination if the goal was **efficiency**?

- We tackle these questions by deriving the *semiparametric efficiency bound* for the ATT in DDD setups.

- That usually leads to DR estimands, too.

Semiparametric Efficiency Bound

Proposition (Semiparametric Efficiency Bound for DDD)

Suppose that conditional PT, no-anticipation, and strong overlap assumptions are satisfied, and balanced panel data is available. Let $\theta(\Delta Y, X) := \Delta Y - m_{\Delta}^{G=2,P=0}(X) - m_{\Delta}^{G=\infty,P=1}(X) + m_{\Delta}^{G=\infty,P=0}(X)$*,* $S := (\Delta Y, G, P, X)$*. Then, the efficient influence function for the ATT is given by*

$$
\begin{aligned}
\eta_{eff}(S) = \omega_1^{G=2,P=1} \cdot \Big(\theta(\Delta Y, X) - \theta \Big) \\
- \omega_0^{G=2,P=0}(X) \cdot \Big(\Delta Y - m_{\Delta}^{G=2,P=0}(X) \Big) \\
- \omega_0^{G=\infty,P=1}(X) \cdot \Big(\Delta Y - m_{\Delta}^{G=\infty,P=1}(X) \Big) \\
+ \omega_0^{G=\infty,P=0}(X) \cdot \Big(\Delta Y - m_{\Delta}^{G=\infty,P=0}(X) \Big).
\end{aligned}
$$

Furthermore, the semiparametric efficiency bound for the set of all regular, and asymptotic linear estimators of the ATT is $\mathbb{E}[\eta_{eff}(S)^2]$.

Weights Notation

DR DDD as a function of 3 DiDs

- We can take the expected value of $\eta_{\text{eff}}(W)$ and isolate θ given that any influence function has mean zero.
- Let's conveniently rewrite the propensity scores $\forall(g', \ell) \in \{(\infty, 0), (\infty, 1), (2, 0)\}$ as

$$p_{g',\ell}(X) = \mathbb{P}[G = 2, P = 1 | X, (G = 2, P = 1) \cup (G = g', P = \ell)],$$

- Finally, we get the following *DR-DDD* estimand for the ATT,

$$\begin{aligned}
\theta^{DR} &= \mathbb{E}\left[\left(\omega_1^{G=2,P=1} - \omega_0^{G=2,P=0}\left(p_{2,0}\left(X\right)\right)\right)\left(\Delta Y - m_{\Delta}^{G=2,P=0}\left(X\right)\right)\right] \Leftarrow DRDiD_{(2,1)}^{(2,0)} \\
&+ \mathbb{E}\left[\left(\omega_1^{G=2,P=1} - \omega_0^{G=\infty,P=1}\left(p_{\infty,1}(X)\right)\right)\left(\Delta Y - m_{\Delta}^{G=\infty,P=1}\left(X\right)\right)\right] \Leftarrow DRDiD_{(2,1)}^{(\infty,1)} \\
&- \mathbb{E}\left[\left(\omega_1^{G=2,P=1} - \omega_0^{G=\infty,P=0}\left(p_{\infty,0}(X)\right)\right)\left(\Delta Y - m_{\Delta}^{G=\infty,P=0}\left(X\right)\right)\right] \Leftarrow DRDiD_{(2,1)}^{(\infty,0)}
\end{aligned}$$

- From here, we need to estimate the nuisance functions to get an estimator for θ.

What happens when we have variation in Treatment Timing?

Staggered DDD with never-treated groups

- Building on the previous intuition for $T = 2$, let's go back to the case with multiple time periods.
- We then explore two alternative assumptions that place constraints on how untreated potential outcomes develop over time:

Assumption (Conditional Parallel Trends for DDD based on "never-treated" groups)

For each $t \in \{2, \dots, T\}$, $g \in \mathcal{G}_{treated}$ such that $t \geq g$, with probability one,

$$\begin{aligned} \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty)|G = g, P = 1, X\right] &- \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty)|G = g, P = 0, X\right] \\ &= \\ \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty)|G = \infty, P = 1, X\right] &- \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty)|G = \infty, P = 0, X\right]. \end{aligned}$$

Staggered DDD with not-yet-treated groups

Assumption (Conditional Parallel Trends for DDD based on "not-yet-treated" groups)

For each $(g', t) \in \mathcal{G} \times \{2, \ldots, \}$, $g \in \mathcal{G}_{treated}$ *such that* $g' > t \geq g$, *with probability one,*

$$\begin{aligned} \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | G = g, P = 1, X\right] \quad &- \quad \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | G = g, P = 0, X\right] \\ &= \\ \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | G = g', P = 1, X\right] \quad &- \quad \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | G = g', P = 0, X\right]. \end{aligned}$$

DR DDD for ATT(g,t) with never-treated as comparison group

- Let

$$m_{g,t}^{G=g',P=\ell}(X) := \mathbb{E}\left[Y_t - Y_{g-1} | G = g', P = \ell, X\right];$$
$$p_{g,g',\ell} := \mathbb{P}\big[G = g, P = 1 | X, (G = g, P = 1) \cup (G = g', P = \ell)\big];$$
$$\Delta Y_{g,t} := Y_t - Y_{g-1}$$

- With this notation, we can introduce a DR DDD estimand leveraging *never-treated* states as a comparison group.
- For any $g \in \mathcal{G}_{\text{treated}}$ and $g' \in \{g, \infty\}$ with eligibility status $\ell \in P$,

$$\begin{aligned} ATT^{nt}(g,t) = {} & \mathbb{E}\left[\left(\omega_1^{G=g,P=1} - \omega_0^{G=g,P=0}\left(p_{g,g,0}(X)\right)\right)\left(\Delta Y_{g,t} - m_{g,t}^{G=g,P=0}(X)\right)\right] \Leftarrow CSDiD_{(g,1)}^{(g,0)} \\ & + \mathbb{E}\left[\left(\omega_1^{G=g,P=1} - \omega_0^{G=\infty,P=1}\left(p_{g,\infty,1}(X)\right)\right)\left(\Delta Y_{g,t} - m_{g,t}^{G=\infty,P=1}(X)\right)\right] \Leftarrow CSDiD_{(g,1)}^{(\infty,1)} \\ & - \mathbb{E}\left[\left(\omega_1^{G=g,P=1} - \omega_0^{G=\infty,P=0}\left(p_{g,\infty,0}(X)\right)\right)\left(\Delta Y_{g,t} - m_{g,t}^{G=\infty,P=0}(X)\right)\right] \Leftarrow CSDiD_{(g,1)}^{(\infty,0)} \end{aligned}$$

DR DDD for ATT(g,t) with not-yet-treated as comparison group

- Similarly, we can introduce a DR DDD estimand leveraging *not-yet-treated* states as a comparison group.
- For any $g \in \mathcal{G}_{\text{treated}}$ and $g' > t \geq g$ with eligibility status $\ell \in P$,

$$
\begin{aligned}
ATT^{nyt}(g,t) = \mathbb{E}&\left[\left(\omega_1^{G=g,P=1} - \omega_0^{G=g,P=0}\left(p_{g,g,0}(X)\right)\right)\left(\Delta Y_{g,t} - m_{g,t}^{G=g,P=0}(X)\right)\right] \Leftarrow CSDiD_{(g,1)}^{(g,0)} \\
+\, \mathbb{E}&\left[\left(\omega_1^{G=g,P=1} - \omega_0^{G=g',P=1}\left(p_{g,g',1}(X)\right)\right)\left(\Delta Y_{g,t} - m_{g,t}^{G=g',P=1}(X)\right)\right] \Leftarrow CSDiD_{(g,1)}^{(g',1)} \\
-\, \mathbb{E}&\left[\left(\omega_1^{G=g,P=1} - \omega_0^{G=g',P=0}\left(p_{g,g',0}(X)\right)\right)\left(\Delta Y_{g,t} - m_{g,t}^{G=g',P=0}(X)\right)\right] \Leftarrow CSDiD_{(g,1)}^{(g',0)}
\end{aligned}
$$

Conclusion

Highlights

- DDD is widely used in empirical research, but its properties have receive little attention.
- In its basic format, it is equivalent to running two separate DiD estimators and subtracting one from another.
  - This equivalence breaks down when covariates play an important role in the analysis.
- We derived semiparametric efficiency bound for DDD and proposed DR DDD estimands.
- We can leverage these results to tackle staggered treatment setups, too.
- A very fast package implementation in R is under construction.

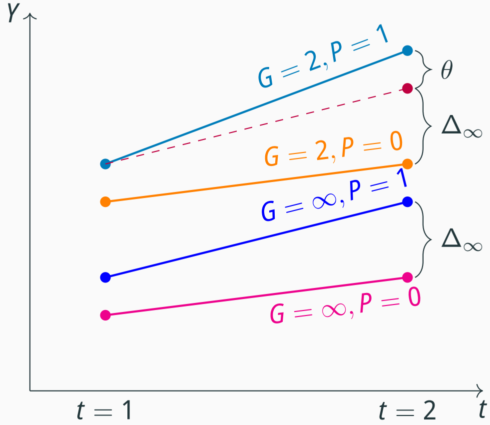
Thanks!

marcelo.ortiz@emory.edu
marcelortiz.com
@marcelortizv

Appendix

2x2x2 DDD without covariates

Let, $g \in G = \{2, \infty\}$ and $\ell \in P = \{0, 1\}$.



$$
\begin{aligned}
\theta = & \Big[\Big(\mathbb{E}[Y_{t=2} | G = 2, P = 1] - \mathbb{E}[Y_{t=1} | G = 2, P = 1] \Big) \\
& - \Big(\mathbb{E}[Y_{t=2} | G = 2, P = 0] - \mathbb{E}[Y_{t=1} | G = 2, P = 0] \Big) \Big] \\
& - \Big[\Big(\mathbb{E}[Y_{t=2} | G = \infty, P = 1] - \mathbb{E}[Y_{t=1} | G = \infty, P = 1] \Big) \\
& - \Big(\mathbb{E}[Y_{t=2} | G = \infty, P = 0] - \mathbb{E}[Y_{t=1} | G = \infty, P = 0] \Big) \Big]
\end{aligned}
$$

- Note that this is the difference of two DiD's: one among $G = 2$ across P groups, and one among $G = \infty$ across P groups.

Go Back

Some additional notation

- For $(g, \ell) \in \{2, \infty\} \times \{0, 1\}$, let $\Delta Y = Y_{t=2} - Y_{t=1}$, and

$$
\begin{aligned}
m_{\Delta}^{G=g,P=\ell}(X) &:= \mathbb{E}\left[\Delta Y | G = g, P = \ell, X\right], \quad \text{(outcome regression)}. \\
p^{G=g,P=\ell}(X) &:= \mathbb{P}[G = g, P = \ell | X] \quad \text{(multi-valued propensity score)}.
\end{aligned}
$$

- For $(g, \ell) \in \mathcal{S}_c \equiv \{(\infty, 0), (\infty, 1), (2, 0)\}$, let

$$
\omega_1^{G=2,P=1} := \frac{1_{\{G=2,P=1\}}}{\mathbb{E}[1_{\{G=2,P=1\}}]},
$$

$$
\omega_0^{G=g,P=\ell}(X) := \frac{\dfrac{1_{\{G=g,P=\ell\}} \cdot p^{G=2,P=1}(X)}{p^{G=g,P=\ell}(X)}}{\mathbb{E}\left[\dfrac{1_{\{G=g,P=\ell\}} \cdot p^{G=2,P=1}(X)}{p^{G=g,P=\ell}(X)}\right]}
$$

- Let $\theta(\Delta Y, X) := \Delta Y - m_{\Delta}^{G=2,P=0}(X) - m_{\Delta}^{G=\infty,P=1}(X) + m_{\Delta}^{G=\infty,P=0}(X)$, $S := (\Delta Y, G, P, X)$.

Go Back

Regression Adjustment and IPW identification

- We can show that if conditional PT, no-anticipation, and strong overlap assumptions are satisfied and balanced panel data is available, the ATT is identified via regression adjustments or IPW:

$$\theta = ATT^{RA} = ATT^{IPW},$$

where

$$ATT^{RA} := \mathbb{E}[\Delta Y | G = 2, P = 1] - \mathbb{E}\left[m_{\Delta}^{G=2,P=0}(X) + \left(m_{\Delta}^{G=\infty,P=1}(X) - m_{\Delta}^{G=\infty,P=0}(X) \right) \Big| G = 2, P = 1 \right],$$

$$ATT^{IPW} := \mathbb{E}\left[\left(\left(w^{G=2,P=1}(G,P) - w^{G=2,P=0}(G,P,X) \right) - \left(w^{G=\infty,P=1}(G,P,X) - w^{G=\infty,P=0}(G,P,X) \right) \right) \Delta Y \right].$$

Go Back

DR DDD as a function of 3 DiDs

- To get a *DR-DDD* estimand for the ATT, isolate θ given that any influence function has mean zero.

Go Back

- We can conveniently rewrite the propensity scores $\forall(g', \ell) \in \mathcal{S}_c$ as

$$p_{g',\ell}(X) = \mathbb{P}[G=2, P=1 | X, (G=2, P=1) \cup (G=g', P=\ell)],$$

$$\begin{aligned}
\implies \theta^{DR} \quad &= \mathbb{E}\left[\left(\frac{1_{\{G=2,P=1\}}}{\mathbb{E}\left[1_{\{G=2,P=1\}}\right]} - \frac{\frac{p_{2,0}(X)\cdot 1_{\{G=2,P=0\}}}{1-p_{2,0}(X)}}{\mathbb{E}\left[\frac{p_{2,0}(X)\cdot 1_{\{G=2,P=0\}}}{1-p_{2,0}(X)}\right]}\right)\left(\Delta Y - m_{\Delta}^{G=2,P=0}(X)\right)\right] \\
&+ \mathbb{E}\left[\left(\frac{1_{\{G=2,P=1\}}}{\mathbb{E}\left[1_{\{G=2,P=1\}}\right]} - \frac{\frac{p_{\infty,1}(X)\cdot 1_{\{G=\infty,P=1\}}}{1-p_{\infty,1}(X)}}{\mathbb{E}\left[\frac{p_{\infty,1}(X)\cdot 1_{\{G=\infty,P=1\}}}{1-p_{\infty,1}(X)}\right]}\right)\left(\Delta Y - m_{\Delta}^{G=\infty,P=1}(X)\right)\right] \\
&- \mathbb{E}\left[\left(\frac{1_{\{G=2,P=1\}}}{\mathbb{E}\left[1_{\{G=2,P=1\}}\right]} - \frac{\frac{p_{\infty,0}(X)\cdot 1_{\{G=\infty,P=0\}}}{1-p_{\infty,0}(X)}}{\mathbb{E}\left[\frac{p_{\infty,0}(X)\cdot 1_{\{G=\infty,P=0\}}}{1-p_{\infty,0}(X)}\right]}\right)\left(\Delta Y - m_{\Delta}^{G=\infty,P=0}(X)\right)\right]
\end{aligned}$$

Monte Carlo Simulations

Simulations for $T = 2$ with covariates

- For simplicity, we consider the scenario for panel data with $T = 2$ and we have access to generic data $W = (W_1, W_2, W_3, W_4)'$.
- WLOG, consider that the *eligibity of treatment* is given by binary well-know criterion $P = \{0, 1\}$ and let $(g, \ell) \in \mathcal{S}_c := \{(\infty, 0), (\infty, 1), (2, 0)\}$ and $\mathcal{S} := \mathcal{S}_c \cup \{(2, 1)\}$.
- Since we have 4 partitions in the data, we can model the selection into treatment as multinomial logistic link function.
- Outcome process can be modeled as linear regression onto W.
- We consider 4 DGPs:
  - **both** models are correctly specified;
  - Only **propensity score** is correctly specified;
  - Only **outcome model** is correctly specified;
  - **both** models are wrong.
- We compare our DR DDD estimator with:
  - 3WFE specification.
  - Difference between 2 Doubly Robust DiD (Sant'Anna & Zhao, 2020).

Results

	DGP 1: $\mathbb{E}\left[\eta_{\text{eff}}(W)^2\right] = 32.82$			DGP 2: $\mathbb{E}\left[\eta_{\text{eff}}(W)^2\right] = 32.52$			DGP 3: $\mathbb{E}\left[\eta_{\text{eff}}(W)^2\right] = 32.82$			DGP 4: $\mathbb{E}\left[\eta_{\text{eff}}(W)^2\right] = 32.52$		
	$\hat{\theta}_{ddd}$	$\hat{\theta}_{3wfe}$	$\hat{\theta}_{dr}$	$\hat{\theta}_{ddd}$	$\hat{\theta}_{3wfe}$	$\hat{\theta}_{dr}$	$\hat{\theta}_{ddd}$	$\hat{\theta}_{3wfe}$	$\hat{\theta}_{dr}$	$\hat{\theta}_{ddd}$	$\hat{\theta}_{3wfe}$	$\hat{\theta}_{dr}$
							$n = 1000$					
Bias	**0.0007**	-7.3298	-4.0199	**-0.0029**	-6.3178	-3.4484	**0.0255**	-3.8366	-1.9826	**0.0421**	-5.6247	-3.4447
RMSE	**0.1823**	8.4185	5.0331	**0.1799**	7.4545	4.5122	**1.4384**	5.1171	3.3235	**1.4498**	6.5893	4.3858
$\mathbb{E}[Var]$	43.5075	47341.2395	.	43.1213	47906.9263	.	2121.5766	45057.4548	.	2163.0339	45102.7697	.
Cov. 95	0.9650	0.9240	.	0.9730	0.9550	.	0.9600	0.9970	.	0.9530	0.9860	.
avg. length	**0.8110**	26.9471	.	**0.8071**	27.1033	.	**5.7012**	26.2999	.	**5.7553**	26.3134	.
							$n = 50000$					
Bias	**0.0008**	-7.3101	-4.0285	0.0007	-6.2891	-3.4389	-0.0039	-3.9647	-2.1148	0.1039	-5.4834	-3.3654
RMSE	**0.0257**	7.3348	4.0522	0.0257	6.3154	3.4636	0.2011	3.9929	2.1469	0.2345	5.5038	3.3857
$\mathbb{E}[Var]$	34.3777	47502.9417	.	33.9857	48240.7801	.	2059.7784	45339.0209	.	2113.8460	45453.0303	.
Cov. 95	0.9550	0.0000	.	0.9470	0.0000	.	0.9540	0.0000	.	0.9130	0.0000	.
avg. length	**0.1028**	3.8208	.	**0.1022**	3.8503	.	**0.7956**	3.7328	.	**0.8060**	3.7375	.

Monte Carlo Design

- Since we have 4 partitions in the data, we consider the following PS using a multinomial logistic link function:

$$p^{G=g,P=\ell}(W)=\begin{cases}\dfrac{\exp\left(f_{ps}^{g,\ell}(W)\right)}{1+\sum_{(g,\ell)\in\mathcal{S}_c}\exp\left(f_{ps}^{g,\ell}(W)\right)}, & \text{if } (g,\ell)\in\mathcal{S}_c\\ \dfrac{1}{1+\sum_{(g,\ell)\in\mathcal{S}_c}\exp\left(f_{ps}^{g,\ell}(W)\right)}, & \text{if } (g,\ell)=(2,1).\end{cases}$$

where, $f_{ps}^{g,\ell}(W)=\alpha_1^{g,\ell}W1+\alpha_2^{g,\ell}W_2+\alpha_3^{g,\ell}W_3+\alpha_4^{g,\ell}W_4$

Monte Carlo Design

- Let $U \sim$ Uniform $[0, 1]$. The partition groups are assigned as follows

$$(g, \ell) := \begin{cases} (\infty, 0), & \text{if } U \leqslant p^{G=\infty, P=0}(W), \\ (\infty, 1), & \text{if } p^{G=\infty, P=0}(W) < U \leqslant p^{G=\infty, P=0}(W) + p^{G=\infty, P=1)}(W), \\ (2, 0), & \text{if } p^{G=\infty, P=0}(W) + p^{G=\infty, P=1}(W) < U \leqslant 1 - p^{G=2, P=1}(W), \\ (2, 1), & \text{if } 1 - p^{G=2, P=1}(W) < U. \end{cases}$$

- For the Outcome Regression process, define

$$f^{g,\ell}_{reg, G=2}(W) = \beta^{g,\ell}_{11} W1 + \beta^{g,\ell}_{21} W_2 + \beta^{g,\ell}_{31} W_3 + \beta^{g,\ell}_{41} W_4, \forall (g, \ell) \in \{(2, \ell)\}$$
$$f^{g,\ell}_{reg, G=\infty}(W) = \beta^{g,\ell}_{10} W1 + \beta^{g,\ell}_{20} W_2 + \beta^{g,\ell}_{30} W_3 + \beta^{g,\ell}_{40} W_4, \forall (g, \ell) \in \{(\infty, \ell)\}$$

- Let *time-invariant unobserved heterogeneity* be defined as $\nu(W, G, P) \sim N\left(f^{g,\ell}_{het}(W), 1\right), \forall (g, \ell) \in \mathcal{S}$ where,

$$f^{g,\ell}_{het}(W) = 1_{\{G=2\}} \cdot 1_{\{P=1\}} \cdot f^{g,\ell}_{reg, G=2}(W) + (1 - 1_{\{G=2\}}) \cdot 1_{\{P=1\}} \cdot f^{s,\ell}_{reg, G=\infty}(W)$$

Go Back