Difference-in-Differences Designs

Extra Lecture 4: Triple-Differences

Pedro H. C. Sant'Anna

August 2022

Introduction

DiD plethora

- We have already seen different DiD facets
  - With and without covariates;
  - With and without several time periods;
  - With and without variation in treatment timing.
- 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 (or not-yet-treated) groups.

What about other DiD designs like Triple-Differences?

How can we leverage our DiD knowledge to tackle DDD?

Setup

Setup

Framework

What is Triple-Differences?

- To make things simple, let's consider a two-periods setup, see, e.g., Olden and Moen (2022) for a review of this simple DDD setup.
  - There are n units available, $i = 1, 2, \ldots, n$ (think of different people) .
  - There are two time periods available, $t = 1$ and $t = 2$.
  - Units belong to two different sets of "states", $s = \infty$, and $s = 2$ (states is just a generic name here). Here, s indexes the date in which treatment start at that set/cohort of states.
  - Within each set of states, we have two partitions (defined by some well-known criteria), $\ell = A$ and $\ell = B$ (think of age profiles or gender).
  - Nobody is treated in period $t = 1$.
  - In period $t = 2$, units in states $s = 2$ with $\ell = B$ are exposed to treatment. Everybody else remains untreated.

CAUSAL SOLUTIONS

Example: Effect of policies providing maternity benefits on hourly wages

- Triple-Differences were introduced by Gruber (1994).
- Main idea of Gruber (1994):
  - Study several state mandates which stipulated that childbirth be covered comprehensively in health insurance plans, raising the relative cost of insuring women of childbearing age. Want to learn the effect on hourly wages.
  - Data on individuals (used CPS data).
  - Exploit data before the mandates ($t = 1$), and after the mandates ($t = 2$).
  - Exploit data on states who have passed mandates ($s = 2$) (treated states), and those who have not passed mandates ($s = \infty$) (untreated states).
  - Exploit data on married women, 20-40yo ($\ell = B$), and all individuals over 40yo, and single males 20-40yo ($\ell = A$).

Setup

Potential outcomes and parameters of interest

Potential outcomes

- As usual, we will talk about potential outcomes.
- Potential outcomes will reflect time units are first-treated (as we have done in almost the entire course).
- Let $Y_{i,t}(g)$ be the potential outcome for unit i, at time t, if this unit is first treated at time period g.
- Units in states $s = 2$ that have $\ell = B$ are treated at time $g = 2$ (treated units), and all other units remain untreated ($g = \infty$).
    - So observations need to satisfy two criteria to be treated: $s = 2$ **and** $\ell = B$.
- For units that are treated in time period $g = 2$, we observe $Y_{i,t=1}(2)$ and $Y_{i,t=2}(2)$.
- For the "never treated" units $g = \infty$, we observe $Y_{i,t=1}(\infty)$ and $Y_{i,t=2}(\infty)$.

CAUSAL SOLUTIONS

Parameter of interest

- In this context, we still care about the ATT:

$$ATT = \mathbb{E}[Y_{t=2}(2) - Y_{t=2}(\infty)|G = 2].$$

- S is a "state" indicator and L is a "partition" indicator.
- Interestingly, note that by conditioning on $G = 2$, we are restricting our attention to those units with $S = 2$, and $\ell = B$, since $G = 2 \iff S = 2$ and $L = B$.
- Thus, by we have that

$$\begin{aligned} ATT &= \mathbb{E}[Y_{t=2}(2) - Y_{t=2}(\infty)|G = 2] \\ &= \mathbb{E}[Y_{t=2}(2) - Y_{t=2}(\infty)|S = 2, L = B], \end{aligned}$$

<u>Average Treatment effect in time period $t = 2$, among units who are treated in period $g = 2$, i.e, those units within the set of states with $s = 2$ that belong to partition $\ell = B$.</u>

Parameter of interest in the context of Gruber (1994)

- In the context of Gruber (1994):

$$\begin{aligned} ATT &= \mathbb{E}[Y_{t=2}(2) - Y_{t=2}(\infty)|G = 2] \\ &= \mathbb{E}[Y_{t=2}(2) - Y_{t=2}(\infty)|S = 2, L = B], \end{aligned}$$

- Average Treatment effect of state maternity mandate on hourly wages in post-treatment period, among married women, with 20-40 years old, who live in states who has a maternity mandate by period $s = 2$.

Can we map this to standard DiDs?

Let's map triple-differences back to a standard DiD

- The first natural thing to do is to just drop data from states with $s = \infty$.
  - There are two time periods available, $t = 1$ and $t = 2$.
  - Units can belong to two different sets of "states", $s = \infty$ or $s = 2$. **We only keep data from units with $s = 2$.**
  - Within each set of state, we have two partitions (defined by some well-known criteria), $\ell = A$ and $\ell = B$ (think of different age profiles or gender) .
  - Nobody is treated in period $t = 1$.
  - In period $t = 2$, units in the set of states $s = 2$ with $\ell = B$ are exposed to treatment. Everybody else remains untreated.
  - We are now back to a 2x2 DiD setup, as, among those in state $s = 2$, $G = 2 \iff L = B$.

CAUSAL SOLUTIONS

Implied assumptions here

- If we were to follow this path, we would be (implicitly) making the following assumptions:

Assumption (No-Anticipation)

For all units i, $Y_{i,t}(g) = Y_{i,t}(\infty)$ *for all groups in their pre-treatment periods, i.e., for all* $t < g$.

Assumption (Parallel Trends Assumption within " treated states")

Within the set of treated states, the average evolution of untreated potential outcomes is the same between units with $\ell = A$ *and* $\ell = B$*, that is,*

$$\mathbb{E}\left[Y_{t=2}(\infty)|S=2, L=B\right] - \mathbb{E}\left[Y_{t=1}(\infty)|S=2, L=B\right] = \mathbb{E}\left[Y_{t=2}(\infty)|S=2, L=A\right] - \mathbb{E}\left[Y_{t=1}(\infty)|S=2, L=A\right].$$

Example: Effect of policies providing maternity benefits on hourly wages

- Within the context of Gruber (1994), these assumptions imply that
  - Women and employers in the treated states do not change their behavior in periods before the state maternity mandate has passed (i.e., they do not act on the potential knowledge of this policy intervention before it actually happens).
  - In the absence of the state maternity mandate, the evolution of average hourly wage among married women, with 20-40 years old, who live in a treated state is the same as the same as the average hourly wage among all individuals over 40yo, and single males 20-40yo, who live in a treated state.
- This PT assumption rule-out gender-and-age-specific trends!

Can we do something else that is still a simple DiD?

Let's map triple-differences back to another standard DiD

- We can drop data from units with $\ell = A$ instead of states with $s = \infty$.
  - There are two time periods available, $t = 1$ and $t = 2$.
  - Units can belong to two different sets of "states", $s = \infty$, $s = 2$.
  - Within each set of state, we have two partitions (defined by some well-known criteria), $\ell = A$ and $\ell = B$. **We only keep data from units with $\ell = B$.**
  - Nobody is treated in period $t = 1$.
  - In period $t = 2$, units (with $\ell = B$) in the set of states $s = 2$ are exposed to treatment. Everybody else remains untreated.
  - We are now back to a 2x2 DiD setup, as, among those with $\ell = B$, $G = 2 \iff S = 2$.

Implied assumptions here

- If we were to follow this path, we would be (implicitly) making the following assumptions:

Assumption (No-Anticipation)

For all units i, $Y_{i,t}(g) = Y_{i,t}(\infty)$ *for all groups in their pre-treatment periods, i.e., for all* $t < g$.

Assumption (Parallel Trends Assumption within "partitions")

Within the set of units with $\ell = B$*, the average evolution of untreated potential outcomes is the same between units in "treated" and "untreated" states, that is,*

$$\mathbb{E}\left[Y_{t=2}(\infty)|S=2, L=B\right] - \mathbb{E}\left[Y_{t=1}(\infty)|S=2, L=B\right] = \mathbb{E}\left[Y_{t=2}(\infty)|S=\infty, L=B\right] - \mathbb{E}\left[Y_{t=1}(\infty)|S=\infty, L=B\right].$$

CAUSAL SOLUTIONS

Example: Effect of policies providing maternity benefits on hourly wages

- Within the context of Gruber (1994), these assumptions imply that
  - Women and employers in the treated states do not change their behavior in periods before the state maternity mandate has passed (i.e., they do not act on the potential knowledge of this policy intervention before it actually happens).
  - In the absence of the state maternity mandate, the evolution of average hourly wage among married women, with 20-40 years old, who live in the set of treated states is the same as the same as the average hourly wage among married women, with 20-40 years old, who live in the set of untreated states.
- This PT assumption rule-out state-specific trends (or state-specific shocks).

Triple-differences for the rescue

Main idea of triple-differences

- Main appeal of a triple-differences procedure is to fully exploit all available data to allow for:
  - "location"-specific trends (think about state-specific trends) ;
  - "partition"-specific trends (think about gender-and-age-specific trends) ;
- Of course, we will need to restrict how these two sets of trends interact.
- The key condition is to assume that the **difference** of the average untreated outcome among units with $\ell = A$ and $\ell = B$ in the treated states $s = 2$ evolves in the same way as the **difference** of the average untreated outcome among units with $\ell = A$ and $\ell = B$ in the untreated states $s = \infty$.

CAUSAL SOLUTIONS

Let's be explicit about our assumptions

Assumption (No-Anticipation)

For all units i, $Y_{i,t}(g) = Y_{i,t}(\infty)$ for all groups in their pre-treatment periods, i.e., for all $t < g$.

Assumption (Parallel Trends Assumption for Triple-Differences)

The difference of the evolution of the average untreated outcome among units with $\ell = A$ and $\ell = B$ in the treated states ($s = 2$) is the same as the difference of the evolution of the average untreated outcome among units with $\ell = A$ and $\ell = B$ in the untreated states ($s = \infty$). That is,

$$\begin{aligned}
\mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | S = 2, L = B\right] \quad &- \quad \mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | S = 2, L = A\right] \\
&= \\
\mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | S = \infty, L = B\right] \quad &- \quad \mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | S = \infty, L = A\right].
\end{aligned}$$

CAUSAL SOLUTIONS

Example: Effect of policies providing maternity benefits on hourly wages

- Within the context of Gruber (1994), these DDD assumptions imply that
  - Women and employers in the treated states do not change their behavior in periods before the state maternity mandate has passed (i.e., they do not act on the potential knowledge of this policy intervention before it actually happens).
  - In the absence of the state maternity mandate, the difference of the evolution of average hourly wage between married women with 20-40 years old and and all individuals over 40 years old, and single males with 20-40 years old is the same in treated and untreated states.
- This PT assumption imposes that differential average trends of untreated outcomes between treated and untreated "target populations" is the same across states who implemented and did not implement the policy. This seems more plausible.

But how can this parallel trends assumption help us?

Parallel Trends and the ATT

- We will follow similar steps as in Lecture 2 (slide 28).
- From the definition of the ATT and SUTVA, and the fact that $G = 2 \iff S = 2$ and $L = B$, we have that

$$\begin{aligned} ATT &\equiv \mathbb{E}\left[Y_{i,t=2}\left(2\right)|S=2, L=B\right] - \mathbb{E}\left[Y_{i,t=2}\left(\infty\right)|S=2, L=B\right] \\ &= \underbrace{\mathbb{E}\left[Y_{i,t=2}|S=2, L=B\right]}_{\text{by SUTVA}} - \mathbb{E}\left[Y_{i,t=2}\left(\infty\right)|S=2, L=B\right] \end{aligned}$$

- Green object is estimable from data (under SUTVA).
- Red object still depends on potential outcomes, and we will show how our assumptions allow us to identify it.

CAUSAL SOLUTIONS

Parallel Trends and the ATT

1) First, let's write our PT exploring the linearity of expectations:

$$\mathbb{E}\left[Y_{t=2}(\infty)|S=2,L=B\right]-\mathbb{E}\left[Y_{t=1}(\infty)|S=2,L=B\right] - \left(\mathbb{E}\left[Y_{t=2}(\infty)|S=2,L=A\right]-\mathbb{E}\left[Y_{t=1}(\infty)|S=2,L=A\right]\right)$$
$$=$$
$$\mathbb{E}\left[Y_{t=2}(\infty)|S=\infty,L=B\right]-\mathbb{E}\left[Y_{t=1}(\infty)|S=\infty,L=B\right] - \left(\mathbb{E}\left[Y_{t=2}(\infty)|S=\infty,L=A\right]-\mathbb{E}\left[Y_{t=1}(\infty)|S=\infty,L=A\right]\right).$$

2) By simple manipulation, we can write it as

$$\begin{aligned}\mathbb{E}\left[Y_{t=2}(\infty)|S=2,L=B\right] \;=\; & \mathbb{E}\left[Y_{t=1}(\infty)|S=2,L=B\right]\\ & +\left(\mathbb{E}\left[Y_{t=2}(\infty)|S=2,L=A\right]-\mathbb{E}\left[Y_{t=1}(\infty)|S=2,L=A\right]\right)\\ & +\left(\mathbb{E}\left[Y_{t=2}(\infty)|S=\infty,L=B\right]-\mathbb{E}\left[Y_{t=1}(\infty)|S=\infty,L=B\right]\right)\\ & -\left(\mathbb{E}\left[Y_{t=2}(\infty)|S=\infty,L=A\right]-\mathbb{E}\left[Y_{t=1}(\infty)|S=\infty,L=A\right]\right).\end{aligned}$$

3) Now, exploiting No-Anticipation:

$$
\begin{aligned}
\mathbb{E}\left[Y_{t=2}(\infty)|S=2,L=B\right] \quad = \quad & \mathbb{E}\left[Y_{t=1}(2)|\underbrace{S=2,L=B}_{iff\ G=2}\right] \\
& + \left(\mathbb{E}\left[Y_{t=2}(\infty)|\underbrace{S=2,L=A}_{implies\ G=\infty}\right] - \mathbb{E}\left[Y_{t=1}(\infty)|\underbrace{S=2,L=A}_{implies\ G=\infty}\right]\right) \\
& + \left(\mathbb{E}\left[Y_{t=2}(\infty)|\underbrace{S=\infty,L=B}_{implies\ G=\infty}\right] - \mathbb{E}\left[Y_{t=1}(\infty)|\underbrace{S=\infty,L=B}_{implies\ G=\infty}\right]\right) \\
& - \left(\mathbb{E}\left[Y_{t=2}(\infty)|\underbrace{S=\infty,L=A}_{implies\ G=\infty}\right] - \mathbb{E}\left[Y_{t=1}(\infty)|\underbrace{S=\infty,L=A}_{implies\ G=\infty}\right]\right).
\end{aligned}
$$

4) Now, exploiting SUTVA (which we are always keeping under the hood):

$$\begin{aligned}
\mathbb{E}\left[Y_{t=2}(\infty)|S=2, L=B\right] \quad = \quad & \mathbb{E}\left[Y_{t=1}|S=2, L=B\right] \\
& +(\mathbb{E}\left[Y_{t=2}|S=2, L=A\right]-\mathbb{E}\left[Y_{t=1}|S=2, L=A\right]) \\
& +(\mathbb{E}\left[Y_{t=2}|S=\infty, L=B\right]-\mathbb{E}\left[Y_{t=1}|S=\infty, L=B\right]) \\
& -(\mathbb{E}\left[Y_{t=2}|S=\infty, L=A\right]-\mathbb{E}\left[Y_{t=1}|S=\infty, L=A\right]).
\end{aligned}$$

Parallel Trends and the ATT

Plugging the red term of slide 21 onto the ATT as defined in slide 18, we have that, under SUTVA + No-Anticipation + PT assumptions, it follows that

$$
\begin{aligned}
\textbf{ATT} = \; & \mathbb{E}\left[Y_{t=2} | S=2, L=B\right] - \Big(\mathbb{E}\left[Y_{t=1} | S=2, L=B\right] \\
& + \mathbb{E}\left[Y_{t=2} | S=2, L=A\right] - \mathbb{E}\left[Y_{t=1} | S=2, L=A\right] \\
& + \mathbb{E}\left[Y_{t=2} | S=\infty, L=B\right] - \mathbb{E}\left[Y_{t=1} | S=\infty, L=B\right] \\
& - \mathbb{E}\left[Y_{t=2} | S=\infty, L=A\right] - \mathbb{E}\left[Y_{t=1} | S=\infty, L=A\right] \Big).
\end{aligned}
$$

Parallel Trends and the ATT

- Now, just rearranging the terms:

$$
\begin{aligned}
\text{ATT} \quad = \quad & \Bigg[\Big(\mathbb{E}\left[Y_{t=2}|S=2, L=B\right] - \mathbb{E}\left[Y_{t=1}|S=2, L=B\right]\Big) \\
& -\Big(\mathbb{E}\left[Y_{t=2}|S=2, L=A\right] - \mathbb{E}\left[Y_{t=1}|S=2, L=A\right]\Big)\Bigg] \\
& -\Bigg[\Big(\mathbb{E}\left[Y_{t=2}|S=\infty, L=B\right] - \mathbb{E}\left[Y_{t=1}|S=\infty, L=B\right]\Big) \\
& -\Big(\mathbb{E}\left[Y_{t=2}|S=\infty, L=A\right] - \mathbb{E}\left[Y_{t=1}|S=\infty, L=A\right]\Big)\Bigg]
\end{aligned}
$$

- This is the Triple-Differences estimand!

CAUSAL SOLUTIONS

Parallel Trends and the ATT

$$\begin{aligned}
\text{ATT} \quad = \quad & \Big[\Big(\mathbb{E}\left[Y_{t=2}|S=2, L=B\right] - \mathbb{E}\left[Y_{t=1}|S=2, L=B\right]\Big) \\
& -\Big(\mathbb{E}\left[Y_{t=2}|S=2, L=A\right] - \mathbb{E}\left[Y_{t=1}|S=2, L=A\right]\Big)\Big] \\
& -\Big[\Big(\mathbb{E}\left[Y_{t=2}|S=\infty, L=B\right] - \mathbb{E}\left[Y_{t=1}|S=\infty, L=B\right]\Big) \\
& -\Big(\mathbb{E}\left[Y_{t=2}|S=\infty, L=A\right] - \mathbb{E}\left[Y_{t=1}|S=\infty, L=A\right]\Big)\Big]
\end{aligned}$$

- Note that this is the difference of two DiD's: one among $S = 2$ and one among $S = \infty$. However, we do not need two different PTs.

CAUSAL SOLUTIONS

Estimation and Inference

"Brute-force" Triple-Differences estimator

- Let $\overline{Y}_{s=a,\ell=b,t=j}$ be the sample mean of the outcome Y for units in the set of states s, with $\ell = b$, in time period j,

$$\overline{Y}_{s=a,\ell=b,t=j} = \frac{1}{N_{s=a,\ell=b,t=j}} \sum_{i=1}^{N \cdot T} Y_i 1\{S_i = a\} 1\{L_i = b\} 1\{T_i = j\},$$

with

$$N_{s=a,\ell=b,t=j} = \sum_{i=1}^{N \cdot T} 1\{S_i = a\} 1\{L_i = b\} 1\{T_i = j\},$$

S_i, L_i and T_i being "state", "partition", and time dummies, respectively, and Y_i is the "poolled" outcome data.

“Brute-force” Triple-Differences estimator

- By using the analogy (or plug-in) principle, we have that our triple-difference **estimator** for the ATT is given by

$$
\begin{aligned}
\widehat{\theta}_n^{DDD} &= \left[\left(\overline{Y}_{S=2,\ell=B,t=2} - \overline{Y}_{S=2,\ell=B,t=1}\right) - \left(\overline{Y}_{S=2,\ell=A,t=2} - \overline{Y}_{S=2,\ell=A,t=1}\right) \right] \\
&\quad - \left[\left(\overline{Y}_{S=\infty,\ell=B,t=2} - \overline{Y}_{S=\infty,\ell=B,t=1}\right) - \left(\overline{Y}_{S=\infty,\ell=A,t=2} - \overline{Y}_{S=\infty,\ell=A,t=1}\right) \right].
\end{aligned}
$$

- This is very intuitive and literally consist of using the difference of two DiD estimators!

TWFE DDD Regression Estimator

- When there is only 2 time periods and no covariates, the following two-way fixed-effects (TWFE) regression specification can be used to recover the ATT:

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

  where $\mathbb{E}[\varepsilon_{i,t} | S_i, L_i, T_i] = 0$ *almost surely*.

- We can show that, under our assumptions, β_0^{twfe} is equal to the DDD estimand in the canonical triple-differences setup.
- Intuition: this regression is just-identified - 8 parameters for 8 unknowns.
- Implication: Inference is now "standard" (at least when there are many clusters).

CAUSAL SOLUTIONS

What if we want to allow for covariate-specific trends?

Allowing for covariate-specific trends

- Just like in the 2x2 DiD, perhaps the valid of the PT for DDD is only plausible after you condition on a vector of (pre-treatment) covariates X.
- This way, you allow for covariate-specific trends, too.
- Interestingly, I have not seen much work on this direction being done.
- Well, let's close this gap (paper is coming soon)!

Allowing for covariate-specific trends

Assumption (Conditional Parallel Trends Assumption for Triple-Differences)

Conditional on X, the difference of the evolution of the average untreated outcome among units with $\ell = A$ and $\ell = B$ in the treated states ($s = 2$) is the same as the difference of the evolution of the average untreated outcome among units with $\ell = A$ and $\ell = B$ in the untreated states ($s = \infty$). That is, with probability one,

$$\begin{aligned} \mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | S = 2, L = B, X\right] &- \mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | S = 2, L = A, X\right] \\ &= \\ \mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | S = \infty, L = B, X\right] &- \mathbb{E}\left[Y_{t=2}(\infty) - Y_{t=1}(\infty) | S = \infty, L = A, X\right]. \end{aligned}$$

Strong Overlap

- Just like in Lecture 4, we will introduce an additional assumption stating that every unit has a strictly positive probability of being in the untreated group.

Assumption (Strong Overlap Assumption)

For $(s, \ell) \in \{(\infty, A), (\infty, B), (2, A)\}$ *and for some* $\epsilon > 0$, $\mathbb{P}[S = s, L = \ell | X] > \epsilon$ *a.s.*

- The covariates X here are the same as those used to justify the conditional PT assumption!

- For identification purpose, we can take $\epsilon = 0$. For (standard) inference, though, we would have problems, unless we rely on "extrapolation"; see, e.g., Khan and Tamer (2010).

Let's now talk about regression adjustments, IPW and DR procedures.

Let's assume we have balanced panel data available

What if we want to allow for covariate-specific trends?

Triple-Differences Regression adjustment

DDD Regression adjustment procedure

- By building on the conditional PT assumption, no-anticipation and strong overlap, we can show that, when balanced panel data is available,

$$
\begin{aligned}
ATT &= \mathbb{E}\left[Y_{t=2} - Y_{t=1} \middle| S=2, L=B\right] \\
&\quad - \mathbb{E}\left[m_{\Delta}^{S=2,L=A}(X) - \left(m_{\Delta}^{S=\infty,L=B}(X) - m_{\Delta}^{S=\infty,L=A}(X) \right) \middle| S=2, L=B \right],
\end{aligned}
$$

where

$$
m_{\Delta}^{S=s,L=\ell}(X) \equiv \mathbb{E}\left[Y_{t=2} - Y_{t=1} \middle| S=s, L=\ell, X\right]
$$

- Now, it is a matter of estimating the unknown **regression functions** $m_{\Delta}^{S=s,L=\ell}(X)$ with you favorite estimation method - it can be parametric, semiparametric, or nonparametric!

Regression-adjusted DDD estimators

rely on researchers ability to model

the outcome evolution for 3 groups!

CAUSAL SOLUTIONS

What if we want to allow for covariate-specific trends?

Inverse Probability Weighting procedure

Inverse probability weighting procedures

- Similarly to Abadie (2005) and Sant'Anna and Zhao (2020), we can get DDD IPW estimators.
- Let's consider the following 4 partitions of the data:
  - $PA_i = 1$ if $S_i = \infty$ and $L_i = A$;
  - $PA_i = 2$ if $S_i = \infty$ and $L_i = B$;
  - $PA_i = 3$ if $S_i = 2$ and $L_i = A$;
  - $PA_i = 4$ if $S_i = 2$ and $L_i = B$.
- Consider the following propensity scores $p_a(X) \equiv \mathbb{P}(PA_i = 4 | X, (PA_i \in \{a, 4\})$.
- $p_a(X)$ measures the probability of belonging to partition 4, given X and given that a unit belongs to partition a or partition 4.

Inverse probability weighting procedures

$$
\begin{aligned}
ATT &= \mathbb{E}\left[\left(\frac{1\{PA=4\}}{\mathbb{E}\left[1\{PA=4\}\right]} - \frac{\frac{p_3(X)\left(1\{PA=3\}\right)}{1-p_3(X)}}{\mathbb{E}\left[\frac{p_3(X)1\{PA=3\}}{1-p_3(X)}\right]}\right)\left(Y_{t=2}-Y_{t=1}\right)\right] \\
&\quad -\mathbb{E}\left[\left(\frac{\frac{p_2(X)\left(1\{PA=2\}\right)}{1-p_2(X)}}{\mathbb{E}\left[\frac{p_2(X)1\{PA=2\}}{1-p_2(X)}\right]} - \frac{\frac{p_1(X)\left(1\{PA=1\}\right)}{1-p_1(X)}}{\mathbb{E}\left[\frac{p_1(X)1\{PA=1\}}{1-p_1(X)}\right]}\right)\left(Y_{t=2}-Y_{t=1}\right)\right],
\end{aligned}
$$

Inverse probability weighting procedures

- These formulas suggest a simple two-step estimation procedure, too!

  1. Choose your favorite method to estimate the unknown propensity scores $p_1(X), p_2(X), p_3(X)$.

  2. Plug-in the estimated fitted values of the propensity scores into the *ATT* equation, and replace the population expectations by their sample analogue.

IPW-adjusted DDD estimators

rely on researchers ability to model

the three propensity score models.

Doubly Robust DDD Procedures

- **Combine both outcome regression and IPW** approaches to form more robust estimators.
- Build on the DR DiD estimator of Sant'Anna and Zhao (2020)
- Estimators are **Doubly Robust consistent**: they are consistent for the ATT if either (but not necessarily both)
  - The regression working models for outcome dynamics are correctly specified.
  - The propensity score working models are correctly specified.
- This is definitely more demanding than standard DiD because we now have 3 OR and PS models instead of 1 OR and 1 PS.

Doubly Robust DDD procedure with Panel data

$$
\begin{aligned}
ATT \quad = \quad & \mathbb{E}\left[\left(\frac{1\{PA=4\}}{\mathbb{E}\left[1\{PA=4\}\right]} - \frac{\frac{p_3(X)\left(1\{PA=3\}\right)}{1-p_3(X)}}{\mathbb{E}\left[\frac{p_3(X)1\{PA=3\}}{1-p_3(X)}\right]}\right)\left(Y_{t=2}-Y_{t=1}-m_{\Delta}^{S=2,L=A}(X)\right)\right] \\
& -\mathbb{E}\left[\left(\frac{1\{PA=4\}}{\mathbb{E}\left[1\{PA=4\}\right]} - \frac{\frac{p_2(X)\left(1\{PA=2\}\right)}{1-p_2(X)}}{\mathbb{E}\left[\frac{p_2(X)1\{PA=2\}}{1-p_2(X)}\right]}\right)\left(Y_{t=2}-Y_{t=1}-m_{\Delta}^{S=\infty,L=B}(X)\right)\right] \\
& +\mathbb{E}\left[\left(\frac{1\{PA=4\}}{\mathbb{E}\left[1\{PA=4\}\right]} - \frac{\frac{p_1(X)\left(1\{PA=1\}\right)}{1-p_1(X)}}{\mathbb{E}\left[\frac{p_1(X)1\{PA=1\}}{1-p_1(X)}\right]}\right)\left(Y_{t=2}-Y_{t=1}-m_{\Delta}^{S=\infty,L=A}(X)\right)\right]
\end{aligned}
$$

Open question:

What is the semiparametric efficiency bound for DDD?

CAUSAL SOLUTIONS

What if we want to allow for covariate-specific trends?

Main take-away message

DDD Procedures with covariates

- Without any covariates, we have seen that we can split the data into two, those with $s = 2$ and those with $s = \infty$, run a separate DiD procedure for each, and then take the difference of the 2 DiD procedures.
- This would be numerically the same as the triple-difference estimator in slide 26, and also the one associated with the TWFE specification in slide 27.
- However, when our PT assumption hold only after conditioning on covariates, this equivalence of doing two separate DiD breaks!
- Intuition: we need to integrate the covariates over the distribution of those with $S = 2, L = B$, for every single piece.
- Consequence: we need to run **three** different DiD's, especially for DR procedures.

What if we have multiple time periods?

DDD setup with multiple time periods

- So far we have considered the simplest triple-differences setup.
- Now, let's consider the case where we have more time periods, but treatment can happen at a fixed point in time (so we still have two set of states):
  - T time periods: $t = 1, 2, \ldots, T$.
  - Treatment may happen at a given point in time, say g.
  - Pre-treatment periods: $t = 1, 2, \ldots, g - 1$.
    Post-treatment periods: $t = g, g + 1, \ldots, T$.
  - 2 set of states: $s = g$ (treated at period g) and $s = \infty$ (untreated by period T)
  - Within each set of states, we have two partitions (defined by some well-known criteria), $\ell = A$ and $\ell = B$.
  - Units in partition $\ell = B$, in states with $s = g$ are exposed to treatment $(G = g)$.
    Everybody else remains untreated $(G = \infty)$.

Parameters of interest

- We know how to handle this case by simply leveraging the insights in Lecture 5!
- Now we have multiple post-treatment periods, so we will talk about time (and group) specific ATT's:

$$ATT(g,t) \equiv \mathbb{E}\left[Y_t(g)|G=g\right] - \mathbb{E}\left[Y_t(\infty)|G=g\right]$$

Average Treatment Effect among units treated at time g, at time t.

- In our DDD context, since $G = g \iff S = g$ and $\ell = B$, we have

$$ATT(g,t) \equiv \mathbb{E}\left[Y_t(g)|S=g, L=B\right] - \mathbb{E}\left[Y_t(\infty)|S=g, L=B\right]$$

- Since we only have one treatment time, we could omit the index g; I prefer to keep it though.

More aggregated parameters of interest

- We can also further aggregate these ATT's across time-periods:

$$\theta_s^{agg} = \frac{\sum_{e=0}^{s} ATT(g, g+e)}{s+1}, \quad s \geq 0.$$

- If you want to "discount" more distant periods, that is easy, as long as these $w_e's$ are estimable or known:

$$\theta_{s,w}^{agg} = \frac{\sum_{e=0}^{s} w_e \cdot ATT(g, g+e)}{\sum_{e=0}^{s} w_e}, \quad s \geq 0.$$

- Advantage of these aggregations: one-dimensional summary parameters.

Let's be explicit about our assumptions with multiple-time periods

Assumption (No-Anticipation)

For all units i, $Y_{i,t}(g) = Y_{i,t}(\infty)$ *for all groups in their pre-treatment periods, i.e., for all* $t < g$.

Assumption (Parallel Trends Assumption for Triple-Differences)

For all $t \geq g$,

$$\begin{aligned} \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = g, L = B\right] &- \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = g, L = A\right] \\ &= \\ \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = \infty, L = B\right] &- \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = \infty, L = A\right]. \end{aligned}$$

Parallel Trends and the ATT(g,t)

- By building on Lecture 5 and on slides 18-24, we can show that, for every $t \geq g$,

$$
\begin{aligned}
\text{ATT}(g,t) = \; & \Bigg[\Big(\mathbb{E}\left[Y_t | S = g, L = B\right] - \mathbb{E}\left[Y_{g-1} | S = g, L = B\right]\Big) \\
& - \Big(\mathbb{E}\left[Y_t | S = g, L = A\right] - \mathbb{E}\left[Y_{g-1} | S = g, L = A\right]\Big)\Bigg] \\
& - \Bigg[\Big(\mathbb{E}\left[Y_t | S = \infty, L = B\right] - \mathbb{E}\left[Y_{g-1} | S = \infty, L = B\right]\Big) \\
& - \Big(\mathbb{E}\left[Y_t | S = \infty, L = A\right] - \mathbb{E}\left[Y_{g-1} | S = \infty, L = A\right]\Big)\Bigg]
\end{aligned}
$$

- It is just a matter of using long-differences!
- Now we can do event-study plots with triple-differences!

CAUSAL SOLUTIONS

Implementation details

- Implementation here is straightforward, too.

1. Split the data into those with $L = B$ and those with $L = A$.
2. Among those with $L = A$, do a DiD for the time periods you care using $S = g$ as "treatment status". .
3. Among those with $L = B$, do a DiD for the time periods you care using $S = g$ as "treatment status".
4. Subtract the DiD event-study estimators for those with $L = A$ from those with $L = B$.
5. This final difference is your DDD estimator for the $ATT(g, t)$'s.
   Importantly, the partition-specific DiDs do not have a causal interpretation under our assumptions.
6. To do inference, leverage their influence functions and use the multiplier bootstrap.
   See R file posted on member's area for a template using the **did** package.

CAUSAL SOLUTIONS

What if we have multiple time periods?

Incorporating covariates

Conditional PT and strong overlap

Assumption (Conditional Parallel Trends Assumption for Triple-Differences)

For all $t \geq g$, with probability one,

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

Assumption (Strong Overlap Assumption)

For $(s, \ell) \in \{(\infty, A), (\infty, B), (g, A)\}$ and for some $\epsilon > 0$, $\mathbb{P}[S = s, L = \ell|X] > \epsilon$ a.s.

CAUSAL SOLUTIONS

Parallel Trends and the ATT(g,t)

- By building on Lecture 5 and on slides 31-39, we can get DDD outcome regressions, IPW and DR procedures.
- It is all about replacing “short-differences” with “long-differences” in slides 31-39.
- Let’s illustrate this procedure with the DR DDD procedure.
- Let $m_t^{S=s,L=\ell}(X) \equiv \mathbb{E}\left[Y_t - Y_{g-1} | S=s, L=\ell, X\right]$
- As before, let $p_a(X) \equiv \mathbb{P}(PA_i = 4 | X, (PA_i \in \{a,4\})$, where
  - $PA_i = 1$ if $S_i = \infty$ and $L_i = A$;
  - $PA_i = 2$ if $S_i = \infty$ and $L_i = B$;
  - $PA_i = 3$ if $S_i = g$ and $L_i = A$;
  - $PA_i = 4$ if $S_i = g$ and $L_i = B$.

Doubly Robust DDD procedure with Panel data and multiple time periods

For every $t \geq g$, under conditional PT, strong overlap, no-anticipation and SUTVA,

$$
\begin{aligned}
ATT(g,t) \quad = \quad & \mathbb{E}\left[\left(\frac{1\{PA=4\}}{\mathbb{E}\left[1\{PA=4\}\right]} - \frac{\frac{p_3(X)\left(1\{PA=3\}\right)}{1-p_3(X)}}{\mathbb{E}\left[\frac{p_3(X)1\{PA=3\}}{1-p_3(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_t^{S=g,L=A}(X)\right)\right] \\
& -\mathbb{E}\left[\left(\frac{1\{PA=4\}}{\mathbb{E}\left[1\{PA=4\}\right]} - \frac{\frac{p_2(X)\left(1\{PA=2\}\right)}{1-p_2(X)}}{\mathbb{E}\left[\frac{p_2(X)1\{PA=2\}}{1-p_2(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_t^{S=\infty,L=B}(X)\right)\right] \\
& +\mathbb{E}\left[\left(\frac{1\{PA=4\}}{\mathbb{E}\left[1\{PA=4\}\right]} - \frac{\frac{p_1(X)\left(1\{PA=1\}\right)}{1-p_1(X)}}{\mathbb{E}\left[\frac{p_1(X)1\{PA=1\}}{1-p_1(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_t^{S=\infty,L=A}(X)\right)\right].
\end{aligned}
$$

Estimation and Inference

- For estimation, all we need to do is to apply the plug-in principle:
  Replace pscores and outcome regressions with fitted values using your favorite method, and the replace population expectations with sample means.

- For inference, we need to account for the estimation uncertainty that arises from estimating the nuisance functions.

- We also need to account for “multiple-testing” as we are very likely to make inference about the path of ATT(g,t) over time.

- All this can be done using the multiplier bootstrap procedure that leverages influence functions, as discussed in detail in Lecture 5.

What if Treatment is staggered?

DDD setup with multiple time periods

- To make things fun, let's now consider the case where we have staggered treatment adoption like in Lecture 7 and 8 [treatment is irreversible (or people do not forget it)].
  - T time periods: $t = 1, 2, \ldots, T$.
  - Different "states" adopt a policy in different time periods g. Let $S \in \mathcal{S} \subset \{2, ..., T\} \cup \{\infty\}$ denote the time state s is first-adopt the policy, with the notion that if a state is "never-treated", $S = \infty$.
  - Within each set of states, we have two partitions (defined by some well-known criteria), $\ell = A$ and $\ell = B$.
  - Let $G_i \in \mathcal{G} \subset \{2, ..., T\} \cup \{\infty\}$ denote the time unit i is first-treated, with the notion that if a unit is "never-treated", $G_i = \infty$. Note that $\mathcal{G} = \mathcal{S}$.
  - Units in partition $\ell = B$, in states with $S = g$ are exposed to treatment at time g, $(G = g)$. Units in partition $\ell = A$ are "never-treated" $(G = \infty)$.

Building block of the analysis

- This part builds on Lecture 8 and on Callaway and Sant'Anna (2021)
- In staggered setups, a parameter that is interesting and has clear economic interpretation is the $ATT(g,t)$

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

- Average Treatment Effect at time t of starting treatment at time g, among the units that indeed started treatment at time g.
- In our DDD context, since $G=g \iff S=g$ and $\ell=B$, we have

$$ATT(g,t)\equiv\mathbb{E}\left[Y_t(g)|S=g,L=B\right]-\mathbb{E}\left[Y_t(\infty)|S=g,L=B\right]$$

CAUSAL SOLUTIONS

Identifying Assumptions: No-Anticipation

- Given that we never observe $Y(\infty)$ in post-treatment periods among units that have been treated, we need to make assumptions to identify $ATT(g,t)$'s
- **No-Anticipation Assumption**: For all i,t and $t < g, g'$, $Y_{i,t}(g) = Y_{i,t}(g')$.
- Unit treatment effects are zero before treatment takes place.
- Exactly the same content as before!

CAUSAL SOLUTIONS

Parallel Trends for DDD based on a "never treated" state

Assumption (Parallel Trends for Triple-Differences based on a "never-treated" state)

For each $t \in \{2, \ldots, T\}$, $g \in \mathcal{S}$ *such that* $t \geq g$,

$$\begin{aligned} \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = g, L = B\right] \quad &- \quad \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = g, L = A\right] \\ &= \\ \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = \infty, L = B\right] \quad &- \quad \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = \infty, L = A\right]. \end{aligned}$$

CAUSAL SOLUTIONS

Parallel Trends for DDD based on not-yet treated states

Let D_k be a dummy variable that takes value 1 if a state has adopted the treatment by time k, and takes value 0 otherwise. That is, $D_k = 1\{S \leq k\}$.

Assumption (Parallel Trends for Triple-Differences based on "Not-Yet-Treated" states)

For each $(k, t) \in \{2, \ldots, T\} \times \{2, \ldots, T\}$, $g \in \mathcal{S}$ *such that* $t \geq g$, $k \geq t$

$$\begin{aligned} \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = g, L = B\right] \quad &- \quad \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = g, L = A\right] \\ &= \\ \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | D_k = 0, S \neq g, L = B\right] \quad &- \quad \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | D_k = 0, S \neq g, L = A\right]. \end{aligned}$$

ATT(g,t) Estimand: “never-treated” as comparison group

- Under no-anticipation and PT based on “never-treated”, we have

$$
\begin{aligned}
\text{ATT}(g,t) \quad = \quad & \Big[\Big(\mathbb{E}\left[Y_t | S = g, L = B\right] - \mathbb{E}\left[Y_{g-1} | S = g, L = B\right]\Big) \\
& - \Big(\mathbb{E}\left[Y_t | S = g, L = A\right] - \mathbb{E}\left[Y_{g-1} | S = g, L = A\right]\Big)\Big] \\
& - \Big[\Big(\mathbb{E}\left[Y_t | S = \infty, L = B\right] - \mathbb{E}\left[Y_{g-1} | S = \infty, L = B\right]\Big) \\
& - \Big(\mathbb{E}\left[Y_t | S = \infty, L = A\right] - \mathbb{E}\left[Y_{g-1} | S = \infty, L = A\right]\Big)\Big]
\end{aligned}
$$

ATT(g,t) Estimand: not-yet treated as comparison group

- If one wants to use an the units that have not-yet been exposed to treatment by time t, we have a different estimand.
- For all $t \geq g$, provided that we have a subset of units with $D_t = 0$,

$$
\begin{aligned}
\text{ATT}(g,t) = & \Bigg[\Big(\mathbb{E}\left[Y_t | S = g, L = B\right] - \mathbb{E}\left[Y_{g-1} | S = g, L = B\right]\Big) \\
& - \Big(\mathbb{E}\left[Y_t | S = g, L = A\right] - \mathbb{E}\left[Y_{g-1} | S = g, L = A\right]\Big)\Bigg] \\
& - \Bigg[\Big(\mathbb{E}\left[Y_t | D_t = 0, S \neq g, L = B\right] - \mathbb{E}\left[Y_{g-1} | S = \infty, L = B\right]\Big) \\
& - \Big(\mathbb{E}\left[Y_t | D_t = 0, S \neq g, L = A\right] - \mathbb{E}\left[Y_{g-1} | D_t = 0, S \neq g, L = A\right]\Big)\Bigg].
\end{aligned}
$$

Implementation details

- Implementation here is straightforward, too.

  1. Split the data into those with $L = B$ and those with $L = A$.
  2. Among those with $L = A$, estimate the "pseudo-ATT(g,t)'s" you care using $S = g$ as "treatment status". .
  3. Among those with $L = B$, estimate the "pseudo-ATT(g,t)'s" care using $S = g$ as "treatment status".
  4. Subtract the "pseudo-ATT(g,t)'s" for the units with $L = A$ from those with $L = B$.
  5. This final difference is your DDD estimator for the $ATT(g,t)$'s.
     Importantly, the partition-specific DiDs do not have a causal interpretation under our assumptions.
  6. To do inference, leverage their influence functions and use the multiplier bootstrap.
     See R file posted on member's area for a template using the **did** package.

What if Treatment is staggered?

Incorporating covariates

Parallel trends for DDD based on a "never treated" states

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

For each $t \in \{2, \ldots, T\}$, $g \in \mathcal{S}$ such that $t \geq g$, with probability one,

$$\mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = g, L = B, X\right] - \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = g, L = A, X\right]$$
$$=$$
$$\mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = \infty, L = B, X\right] - \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = \infty, L = A, X\right].$$

Assumption (Strong Overlap Assumption)

For $(s, \ell) \in \mathcal{S} \times \{A, B\}$ and for some $\epsilon > 0$, $\mathbb{P}[S = s, L = \ell | X] > \epsilon$ a.s.

It is possible to (slightly) relax this overlap condition, but we omit the details.

Parallel Trends for DDD based on not-yet treated states

Assumption (Conditional Parallel Trends for DDD based on "Not-Yet-Treated" states)

For each $(k, t) \in \{2, \ldots, T\} \times \{2, \ldots, T\}$, $g \in \mathcal{S}$ *such that* $t \geq g$, $k \geq t$, *with probability one,*

$$\mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = g, L = B, X\right] - \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | S = g, L = A, X\right]$$
$$=$$
$$\mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | D_k = 0, S \neq g, L = B, X\right] - \mathbb{E}\left[Y_t(\infty) - Y_{t-1}(\infty) | D_k = 0, S \neq g, L = A, X\right].$$

Assumption (Strong Overlap Assumption)

For $(s, \ell) \in \mathcal{S} \times \{A, B\}$ *and for some* $\epsilon > 0$, $\mathbb{P}[S = s, L = \ell | X] > \epsilon$ *a.s.*

It is possible to (slightly) relax this overlap condition, but we omit the details.

Identification results - never treated as comparison group

- To introduce a DR DDD estimand leveraging the never-treated states, we will need additional notation.
- Let $m_{g,t}^{S=s,L=\ell}(X) \equiv \mathbb{E}\left[Y_t - Y_{g-1} | S = s, L = \ell, X\right]$.
- Consider the following four group-specific partitions:
  - $PA_i^{(g,\infty)} = 1$ if $S_i = \infty$ and $L_i = A$ and $(S_i = \infty$ or $S_i = g)$;
  - $PA_i^{(g,\infty)} = 2$ if $S_i = \infty$ and $L_i = B$ and $(S_i = \infty$ or $S_i = g)$;
  - $PA_i^{(g,\infty)} = 3$ if $S_i = g$ and $L_i = A$ and $(S_i = \infty$ or $S_i = g)$;
  - $PA_i^{(g,\infty)} = 4$ if $S_i = g$ and $L_i = B$ and $(S_i = \infty$ or $S_i = g)$.
- Let $p_a^{(g,\infty)}(X) \equiv \mathbb{P}(PA_i^{(g,\infty)} = 4 | X, (PA_i^{(g,\infty)} \in \{a, 4\})$.

Identification results - never treated as comparison group

$$
\begin{aligned}
ATT(g,t) \quad = \quad & \mathbb{E}\left[\left(\frac{1\{PA^{(g,\infty)}=4\}}{\mathbb{E}\left[1\{PA^{(g,\infty)}=4\}\right]} - \frac{\frac{p_3^{(g,\infty)}(X)\left(1\{PA^{(g,\infty)}=3\}\right)}{1-p_3^{(g,\infty)}(X)}}{\mathbb{E}\left[\frac{p_3^{(g,\infty)}(X)1\{PA^{(g,\infty)}=3\}}{1-p_3^{(g,\infty)}(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{S=g,L=A}(X)\right)\right] \\
& -\mathbb{E}\left[\left(\frac{1\{PA^{(g,\infty)}=4\}}{\mathbb{E}\left[1\{PA^{(g,\infty)}=4\}\right]} - \frac{\frac{p_2^{(g,\infty)}(X)\left(1\{PA^{(g,\infty)}=2\}\right)}{1-p_2(X)}}{\mathbb{E}\left[\frac{p_2^{(g,\infty)}(X)1\{PA^{(g,\infty)}=2\}}{1-p_2^{(g,\infty)}(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{S=\infty,L=B}(X)\right)\right] \\
& +\mathbb{E}\left[\left(\frac{1\{PA^{(g,\infty)}=4\}}{\mathbb{E}\left[1\{PA^{(g,\infty)}=4\}\right]} - \frac{\frac{p_1^{(g,\infty)}(X)\left(1\{PA^{(g,\infty)}=1\}\right)}{1-p_1^{(g,\infty)}(X)}}{\mathbb{E}\left[\frac{p_1^{(g,\infty)}(X)1\{PA^{(g,\infty)}=1\}}{1-p_1^{(g,\infty)}(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{S=\infty,L=A}(X)\right)\right].
\end{aligned}
$$

- Within states with $S = g$ or $S = \infty$, use the DR DDD estimand introduce in Slide 50.
  Generalize results of Callaway and Sant'Anna (2021) to the triple-differences setup!

Identification results - not-yet-treated as comparison group

- To introduce a DR DDD estimand leveraging the not-yet-treated states, we will need alternative (but similar) notation.
- Let $m_{g,t}^{S>s,L=\ell}(X) \equiv \mathbb{E}\left[Y_t - Y_{g-1} | S > s, L = \ell, X\right]$.
- Consider the following four group-specific partitions:
  - $PA_i^{(g,>t)} = 1$ if $S_i > t$ and $L_i = A$ and $(S_i > t$ or $S_i = g)$;
  - $PA_i^{(g,>t)} = 2$ if $S_i > t$ and $L_i = B$ and $(S_i > t$ or $S_i = g)$;
  - $PA_i^{(g,>t)} = 3$ if $S_i = g$ and $L_i = A$ and $(S_i > t$ or $S_i = g)$;
  - $PA_i^{(g,>t)} = 4$ if $S_i = g$ and $L_i = B$ and $(S_i > t$ or $S_i = g)$.
- Let $p_a^{(g,>t)}(X) \equiv \mathbb{P}(PA_i^{(g,>t)} = 4 | X, (PA_i^{(g,>t)} \in \{a, 4\})$.

Identification results - never treated as comparison group

$$
\begin{aligned}
ATT(g,t) \; = \; & \mathbb{E}\left[\left(\frac{1\{PA^{(g,>t)}=4\}}{\mathbb{E}\left[1\{PA^{(g,>t)}=4\}\right]} - \frac{\frac{p_3^{(g,>t)}(X)\left(1\{PA^{(g,>t)}=3\}\right)}{1-p_3^{(g,>t)}(X)}}{\mathbb{E}\left[\frac{p_3^{(g,>t)}(X)1\{PA^{(g,>t)}=3\}}{1-p_3^{(g,>t)}(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{S=g,L=A}(X)\right)\right] \\
& -\mathbb{E}\left[\left(\frac{1\{PA^{(g,>t)}=4\}}{\mathbb{E}\left[1\{PA^{(g,>t)}=4\}\right]} - \frac{\frac{p_2^{(g,>t)}(X)\left(1\{PA^{(g,>t)}=2\}\right)}{1-p_2(X)}}{\mathbb{E}\left[\frac{p_2^{(g,>t)}(X)1\{PA^{(g,>t)}=2\}}{1-p_2^{(g,>t)}(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{S>t,L=B}(X)\right)\right] \\
& +\mathbb{E}\left[\left(\frac{1\{PA^{(g,>t)}=4\}}{\mathbb{E}\left[1\{PA^{(g,>t)}=4\}\right]} - \frac{\frac{p_1^{(g,>t)}(X)\left(1\{PA^{(g,>t)}=1\}\right)}{1-p_1^{(g,>t)}(X)}}{\mathbb{E}\left[\frac{p_1^{(g,>t)}(X)1\{PA^{(g,>t)}=1\}}{1-p_1^{(g,>t)}(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{S>t,L=A}(X)\right)\right].
\end{aligned}
$$

- Within states with $S = g$ or $S > t$, use the DR DDD estimand introduce in Slide 50.
  Generalize results of Callaway and Sant'Anna (2021) to the triple-difference setup!

CAUSAL SOLUTIONS

What if Treatment is staggered?

Aggregation

Let’s aggregate these ATT(g,t)’s!

Summarizing ATT(g,t)

- $ATT(g,t)$ are very useful parameters that allow us to better understand treatment effect heterogeneity.
- We can also use these to summarize the treatment effects across groups, time since treatment, calendar time.
- In Lecture 8, we have discuss this in detail based on the procedure proposed by Callaway and Sant'Anna (2021)
- All the discussion there applies here directly, without any change!
- Same is true for estimation and inference steps!

CAUSAL SOLUTIONS

Take-way messages

DDD procedures

- Triple-Differences are commonly used in empirical research
- In its basic format, it is equivalent to running two separate DiD estimators, and subtracting one from another.
- However, it only needs one parallel trends to hold, not two - that is the whole point of triple-differences.
- Building on the first principles we have been developing throughout the course, we show can we can identify $ATT(g, t)$ in DDD setups with and without variation of treatment time.
- We show how we can reliably incorporate covariates to form DR DDD estimators, too!
- Now, we just need some volunteers to implement all these into statistical software!

References

Abadie, Alberto, "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies*, 2005, *72* (1), 1–19.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Gruber, Jonathan, "The Incidence of Mandated Maternity Benefits," *American Economic Review*, 1994, *84* (3), 622–641.

Khan, Shakeeb and Elie Tamer, "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 2010, *78* (6), 2021–2042.

Olden, Andreas and Jarle Moen, "The triple difference estimator," *Econometrics Journal*, 2022, *Forthcoming.*

Sant'Anna, Pedro H. C. and Jun Zhao, "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, November 2020, *219* (1), 101–122.