EC 709: DiD–Parallel Trend Assumption

Liang Zhong[1]

Boston University

samzl@bu.edu

September 2023

[1]Thanks to Rubaiyat Alam for kindly sharing his lecture notes

Overview

Table of Contents

Causal Inference with Panel data

Simplest Case: Panel data on Y_{it} for $t = 1, 2$, $d = 0, 1$ and $i = 1, ..., N$

- Pre vs. Post comparisons:
  - Compares: same individuals/communities/groups of units before and after program
  - Implementation is very simple: just a comparison of mean across time periods
  - Drawback: Does not account for potential trends in outcomes
  - More reasonable if we study very short-run effects
- Treated vs. Untreated comparisons
  - Compares: participants to those who have not experienced treatment (at least not yet)
  - Usually implemented via lookalike/matching or regressions or machine learning methods
  - Drawback: Rule out selection on unobservables
  - Need to have data on everything that affects treatment timing and outcomes of interest (unconfoundedness assumption)

About Difference-in-Differences

- Difference-in-Differences (DID) combines previous approaches to avoid their pitfalls
  - Exploit variation in time (before vs. after) and across groups (treated vs. untreated) to recover the causal effects of interest
- Advantage: Allow for selection on time-invariant unobservables and for time-trends
- *DID*

$$
\begin{aligned}
&= (\hat{Y}_{d=1,t=2} - \hat{Y}_{d=1,t=1}) - (\hat{Y}_{d=0,t=2} - \hat{Y}_{d=0,t=1}) \\
&\overset{p}{=} (E[Y_{i,t=2}|D_i = 1] - E[Y_{i,t=1}|D_i = 1]) - (E[Y_{i,t=2}|D_i = 0] - E[Y_{i,t=1}|D_i = 0])
\end{aligned}
$$

  - $\hat{Y}_{d=s,t=j}$: the sample mean of the outcome Y for units in group s in time period j
  - $\overset{p}{=}$: the 2nd term is the sample analog of the 3rd term
- What are the assumptions to make DID estimate ATET?

Assumptions for DID to be ATET-SUTVA

Assumption 1 (SUTVA)

$Y_{i,t} = \Sigma_{d \in \{0,1\}} 1\{D_i = d\} Y_{i,t}(d)$

$\Rightarrow$ Implicitly implies that potential outcomes for unit i are not affected by the treatment of unit j

- Rules out interference and spillover effects across units
- May be problematic in some applications

$\Leftarrow$ See Butts(2023) for a DID framework without SUTVA

$\Rightarrow$ *DID*

$$= (E[Y_{i,t=2}|D_i = 1] - E[Y_{i,t=1}|D_i = 1]) - (E[Y_{i,t=2}|D_i = 0] - E[Y_{i,t=1}|D_i = 0])$$
$$= (E[Y_{i,t=2}(1)|D_i = 1] - E[Y_{i,t=1}(1)|D_i = 1]) - (E[Y_{i,t=2}(0)|D_i = 0] - E[Y_{i,t=1}(0)|D_i = 0])$$

Parallel Trend Assumption

Assumption 2 (Parallel Trends Assumption (PT))

$$E[Y_{i,t=2}(0)|D_i = 1] - E[Y_{i,t=1}(0)|D_i = 1] = E[Y_{i,t=2}(0)|D_i = 0] - E[Y_{i,t=1}(0)|D_i = 0]$$

$\Rightarrow$ In the absence of treatment, the evolution of the outcomes among the treated units is, on average, the same as the evolution of the outcomes among the untreated units

- Confounding canceled out w/a **particular functional form**
- **the selection bias** to be constant over time

$\Rightarrow$ *DID*

$$= (E[Y_{i,t=2}(1)|D_i = 1] - E[Y_{i,t=1}(1)|D_i = 1]) - (E[Y_{i,t=2}(0)|D_i = 0] - E[Y_{i,t=1}(0)|D_i = 0])$$

$$= (E[Y_{i,t=2}(1)|D_i = 1] - E[Y_{i,t=1}(1)|D_i = 1]) - (E[Y_{i,t=2}(0)|D_i = 1] - E[Y_{i,t=1}(0)|D_i = 1])$$

No-Anticipation Assumption

Assumption 3 (No-Anticipation)

For all units i, $Y_{i,t}(1) = Y_{i,t}(0)$ for all groups in their pre-treatment periods

$\Rightarrow$ Unit-specific treatment effects are zero in all pre-treatment periods

- Plausible in many setups, especially if treatment is not announced in advance

$\Rightarrow$ *DID*

$$= (E[Y_{i,t=2}(1)|D_i = 1] - E[Y_{i,t=1}(1)|D_i = 1]) - (E[Y_{i,t=2}(0)|D_i = 1] - E[Y_{i,t=1}(0)|D_i = 1])$$
$$= (E[Y_{i,t=2}(1)|D_i = 1] - E[Y_{i,t=1}(0)|D_i = 1]) - (E[Y_{i,t=2}(0)|D_i = 1] - E[Y_{i,t=1}(0)|D_i = 1])$$
$$= E[Y_{i,t=2}(1)|D_i = 1] - E[Y_{i,t=2}(0)|D_i = 1] = ATET$$

Table of Contents

Functional form and PT

1. Confounding canceled out w/a **particular functional form**?

- Consider an example:
    - In period 0, all control units have outcome 10; all treated units have outcome 5.
    - In period 1, all control units have outcome 15.
    - If treatment hadn't occurred, would the treated units' outcome have increased by 5 also (PT in levels)?
    - Or would they have increased by 50% ($\sim$ PT in logs)?

- The situation becomes more complicated when dealing with a binary outcome variable:
    1. Both groups move by the same absolute probability (LPM)
    2. Both groups move by the same standard deviations (of the standard normal error) (Probit)
    3. Both groups move by the same logit expression (Logit)

Functional form for binary outcome variables

- Suppose $P(Y = 1|D = 0)$ increases from 0.8 (pre-treatment period) to 0.82 (post-treatment period)
- Suppose $P(Y = 1|D = 1) = 0.5$ in the pre-treatment period.

$\Rightarrow$ What should be the counterfactual in the post-treatment period if no treatment was assigned?

- By the same absolute probability: increase to 0.52 in counterfactual
- By the same standard deviations: increase to 0.53 in counterfactual
    - $\Phi^{-1}(0.8) = 0.842$ and $\Phi^{-1}(0.82) = 0.915$. This is a movement of 0.073 standard deviations
    - $\Phi^{-1}(0.5) = 0$, thus treatment moves from 0 to 0.073
    - The probability in counterfactual is $\Phi(0.073) = 0.528 \approx 0.53$

Choosing Functional form

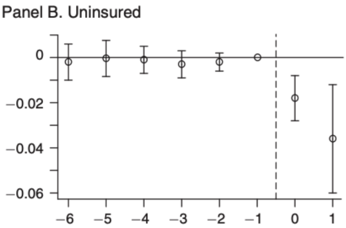
- In general, the counterfactual is sensitive to the functional form
- Roth and Sant'Anna (2023) show that PT will depend on the functional form unless:
  - **Randomization:** treated and control groups have the same distribution of $Y(0)$ in each period
  - **No time effects:** distribution of $Y(0)$ doesn't change over time for either group
  - **A hybrid:** θ fraction of the population is as good as randomized; the other $1-\theta$ fraction has no time effects
- Absent these conditions, PT will be violated for at least some functional form
  - chose the right one by "theory" or by pre-trend

Pre-trend and Parallel Trend Assumption

- Luckily, in most DiD applications we have several periods before anyone is treated
- We can test whether the groups were moving in parallel prior to the treatment
  - If so, then the assumption that confounding factors are stable seems more plausible
  - If not, then it's relatively implausible that would have magically started moving in parallel after the treatment date
- Testing for pre-trends provides a natural plausibility check on the parallel trends assumption
- Most widely used method: **Event-Study plot**
  - Treat year prior to treatment as the base year
  - Estimate difference between the control and treatment groups in each previous year relative to the base year
  - Essentially a set of placebo DIDs done together

Event-Study plot in Carey, Miller, and Wherry (2020)

Comparing states that expanded Medicaid in 2014 to states that didn't



$$Y_{its} = \phi_t + \lambda_s + \sum_{r \neq -1} D_i \times 1[t = 2014 + r]\beta_r + \epsilon_{it}$$

- Y_{its}: insurance for person i in year t in state s, and $D_i = 1$ if in an expansion state
- Insignificant effect before 2014 $\Rightarrow$ pre-periods have the same effect as base year $\Rightarrow$ No pre-trend, PT might hold
- At least choose the functional form that no pre-trend in the event-study plot

Selection bias and PT

2. **the selection bias** to be constant over time?

- There might be different confounding factors in period 1 as in period 0

  E.g.. states that passing a minimum wage increase might also change unemployment insurance (UI) at the same time
  - Then UI is a confound in period 1 but not in period 0
- The same confounding factors may have different effects on the outcome in different time periods
  - Suppose people who enroll in a job training program are more motivated to find a job
  - Motivation might matter more in a bad economy than in a good economy
- Researchers often test it by looking at the event-study plot as well
- However, DID wouldn't be valid regardless of the results of testing Pre-trend

Limitations on Testing Pre-trend

1. Parallel pre-trends don't necessarily imply parallel (counterfactual) post-treatment trends
    - If other policies change at the same time as the one of interest — e.g. min wage and UI reform together — can produce parallel pre-trends but non-parallel post-trends
    - Likewise, could be that treated/control groups are differentially exposed to recessions, but there is only a recession in the post-treatment period
    - So for the selection bias cases above, typically useless to test pre-trend

- Kahn-Lane and Lang (2019): Need logical reasoning about why parallel trends should apply

⇒ "authors should perform a thorough comparison of the differences between the treatment and control groups including demographic composition, other factors that could have differentially affected each group, and comparison of trends as far back as possible".

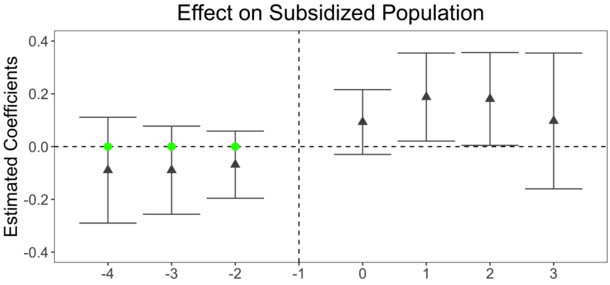
Limitations on Testing Pre-trend (cont.)

2. Low power: even if pre-trends are non-zero, we may fail to detect it statistically
3. Pre-testing issues: if we only analyze cases without statistically significant pre-trends, this introduces a form of selection bias (which can make things worse)

- If we fail the pre-test, what next? May still want to write a paper
- I will talk about them one by one in the rest of the talk

  Reference: DID Resources by Jonathan Roth

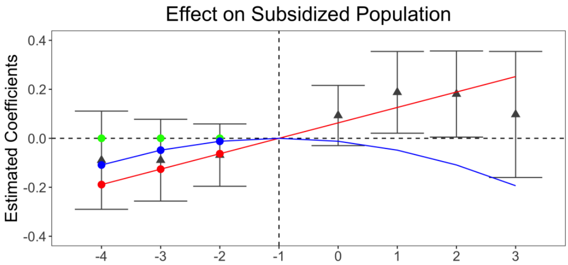
Low power issue



- He & Wang (2017) study impacts of placing college grads as village officials in China
- P-value for H_0 : β_{pre} = green dots (no pre-trend): 0.81

$\Rightarrow$ can't reject zero pre-trend

Low power issue (cont.)

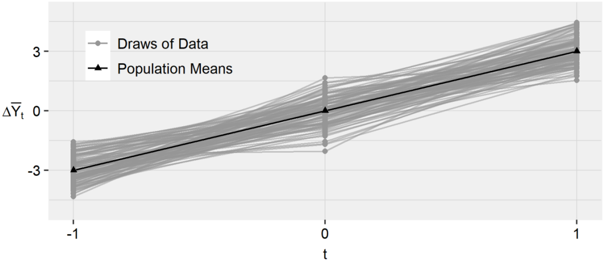


- P-value for H0 : β_{pre} = red dots (has linear pre-trend): 0.81
- P-value for H0 : β_{pre} = blue dots (has quadratic pre-trend): 0.81

$\Rightarrow$ May not find significant pre-trend even if PT is violated

- Under smooth extrapolations to the post-treatment period would produce substantial bias

Distortions from Pre-testing

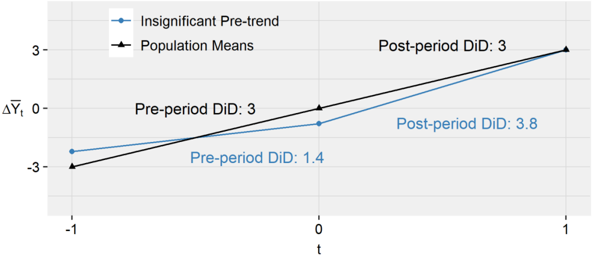


- Example: In population, there is a linear difference in trend with slope 3

$\Rightarrow$ PT is violated, treatment occurs at $t = 1$ but no causal effect

- Due to sampling variation, there will be noise around this line (grey lines)
- In some cases, the difference between period -1 and 0 will be insignificant

Distortions from Pre-testing (cont.)



- In the selected sample, we tend to underestimate the difference between treatment and control at $t = 0$

$\Rightarrow$ DiD between period 0 and 1 tends to be particularly large when we get an insignificant pre-trend

$\Rightarrow$ Selection bias from only analyzing cases with insignificant pre-trend

Roth's comments on Pre-testing

- Proposed a formal method with better power to test pre-trends
    - *Pretrends* package and Shiny App
    - Need to specify the hypothesized trend. Will sometimes be difficult to summarize over many of these
- Does not avoid the issues of statistical distortions from pretesting

  $\Rightarrow$ Consider alternative approaches that attempt to avoid the pretesting
    - These are also the approaches we can try when we fail the pre-test

1. Freyaidenhoven, Hansen, and Shapiro (2019):
    - Unobserved Confounders $\Rightarrow$ Endogenity on treatment $\Rightarrow$ nonzero pre-trends
    - Find a covariate x_{it} (e.g., adult employment) that is affected by the unobserved confounder (e.g., labor demand) η_{it} but not by the treatment z_{it}
    - Assume the dynamic relationship of x_{it} to z_{it} mirrors the dynamic relationship of η_{it} to z_{it}
    - Use covariate to adjust for the counterfactual difference in trends

A sensitivity analysis Approach

2. Rambachan and Roth (2022):
    - The intuition motivating pre-trends testing is that the pre-trends are informative about counterfactual post-treatment trends
    - Formalize this by imposing the restriction that the counterfactual difference in trends can't be "too different" than the pre-trend

- Denote: δ_1: magnitude of violation of PT (unobservable)

$\Rightarrow$ δ_{-1}: pre-treatment analog (observable)

- Two Types of Relaxation by Roth:
    - Bounds on relative magnitudes: Require that $|\delta_1| \leq \hat{M}|\delta_{-1}|$
    - Smoothness restriction: Bound how far δ_1 can deviate from a **linear extrapolation** of the pre-trend:$\delta_1 \in [-\delta_{-1} - M, -\delta_{-1} + M]$

Rewrite Parallel Trend Assumption

Assumption 4 (Rewrite Parallel Trends Assumption)

$$\delta_1 = E[Y_{i,t=2}(0) - Y_{i,t=1}(0)|D_i = 1] - E[Y_{i,t=2}(0) - Y_{i,t=1}(0)|D_i = 0] = 0$$

- Pre-treatment analog
  $\delta_{-1} = E[Y_{i,t=1}(0) - Y_{i,t=0}(0)|D_i = 1] - E[Y_{i,t=1}(0) - Y_{i,t=0}(0)|D_i = 0]$

$\Rightarrow$ With No anticipation effect , DID

$$\begin{aligned}
&= (E[Y_{i,t=2}(1)|D_i = 1] - E[Y_{i,t=1}(1)|D_i = 1]) - (E[Y_{i,t=2}(0)|D_i = 0] - E[Y_{i,t=1}(0)|D_i = 0]) \\
&= (E[Y_{i,t=2}(1)|D_i = 1] - E[Y_{i,t=1}(0)|D_i = 1]) - (E[Y_{i,t=2}(0)|D_i = 1] - E[Y_{i,t=1}(0)|D_i = 1] - \delta_1) \\
&= ATET + \delta_1
\end{aligned}$$

- Previously, we assumed $\delta_1 = 0$, so $DID = ATET$

$\Rightarrow$ If $\delta_1 \in [-\delta_{-1} - M, -\delta_{-1} + M]$, $DID \in [ATET - \delta_{-1} - M, ATET - \delta_{-1} + M]$

$\Rightarrow$ Bound on ATET: $[DID + \delta_{-1} - M, DID + \delta_{-1} + M]$

Some comments on this approach

Similar idea to all other sensitivity analysis papers (e.g., Oster (2019))

1. Introducing an alternative assumption and the sensitivity parameter M
2. Bound the treatment effect and obtain uniformly valid confidence sets
    - Confidence sets directly account for the uncertainty over the magnitude of the pretreatment trend and thus avoid the need to test whether the pre-trends are zero
    - The robust CIs tend to be wider the larger are the confidence intervals on the pre-trends — intuitive, since if we know less about the pre-trends, we should have more uncertainty

    $\Rightarrow$ Contrasts with pre-trends tests, where you're less likely to reject the null that $\beta_{pre} = 0$ when the SEs are larger!
3. Robustness measure of the results: How different would the counterfactual trend have to be from the pre-trends to negate a conclusion (e.g. a positive effect)?

- R Package Available: *HonestDiD* package

Maybe a weaker assumption?

- All the previous discussions center around the parallel trend assumption

$\Leftarrow$ It is quite restrictive, especially there is no covariates X in it

Assumption 5 (Conditional Parallel Trends Assumption)

$$E[Y_{i,t=2}(0)|D_i = 1, X] - E[Y_{i,t=1}(0)|D_i = 1, X] = E[Y_{i,t=2}(0)|D_i = 0, X] - E[Y_{i,t=1}(0)|D_i = 0, X]$$

$\Rightarrow$ Intuitively, a weaker assumption than PT

  - In the absence of treatment, conditional on X, the evolution of the outcome among the treated units is, on average, the same as the evolution of the outcome among the untreated units
  - Allows for covariate-specific trends

- However, estimation is very tricky

$\Rightarrow$ Simply add covariates X in the TWFE specification can introduce huge bias

- Will be discussed next time

Thank You!