Causal Inference using Difference-in-Differences

Lecture 4: Parallel Trends and Functional Form

Pedro H. C. Sant'Anna
Emory University

January 2025

Introduction

Introduction

- **Difference-in-differences** (DiD) is one of the most popular strategies for estimating causal effects in non-experimental contexts.
- The reliability of DiD methods depends on the **parallel trends** assumption.
- Random assignment of treatment (unconfoundedness) is not necessary for parallel trends to hold.

What does parallel trends impose if treatment is not randomly assigned?

Introduction

- There are potentially many ways of tackling this question.
- A natural one focuses on the extent to which the validity of DiD depends on **functional form** restrictions.
- Following Athey and Imbens (2006), we will say parallel trends is *insensitive* to functional form if when it holds for potential outcomes $Y(\infty)$, it also holds for potential outcomes $s(Y(\infty))$ for any strictly monotonic s.
- Intuitively, this says that parallel trends holds regardless of the units in which one measures the outcome.

Why study sensitivity to functional form

- Studying sensitivity to functional form helps clarify the different ways that a researcher can justify the validity of a DiD design:
  - Can verify conditions that ensure PT holds for all functional forms.
  - If sensitive to functional form, can justify the particular choice.

WWhy study sensitivity to functional form

- It often **may not be clear from subject-specific knowledge** what is the "right" transformation for PT to hold.
- Example: different labor market studies have measured **earnings in levels, logs, or percentiles** relative to national wage distribution.
- The choice of transformation may be motivated by which ATT is "most relevant", but not always obvious that policy variation will generate PT for the same transformation
- Moreover, we might want to use the same policy variation to study the ATT for multiple transformations of the same outcome.
- We will use Meyer, Viscusi and Durbin (1995) as a running example in the next slides: interested in studying whether changes in weekly benefit amounts affected the duration of time out of work in Michigan and Kentucky.

Parallel Trends in levels

- **Parallel trends assumption (in levels):**

$$\mathbb{E}\left[Y_{i,t=2}(\infty)|G_i=2\right]-\mathbb{E}\left[Y_{i,t=1}(\infty)|G_i=2\right]=\mathbb{E}\left[Y_{i,t=2}(\infty)|G_i=\infty\right]-\mathbb{E}\left[Y_{i,t=1}(\infty)|G_i=\infty\right]$$

- If Y is the duration of claims measured in weeks, and treatment is an increase of cap (PT in levels)
  - PT would suggest that the average untreated claims' duration among workers who are affected by the increased cap would evolve the same as the average untreated claims' duration among workers who are not affected by the change in the cap.
  - If the average change in untreated claims' duration among workers who are not affected by the change in the cap is 0.05 weeks, these would serve as counterfactual changes for the average untreated claims' duration among workers who are affected by the change in cap
  - ATT would provide the average treatment effect (in weeks) among workers who are affected by the change in cap.

Parallel Trends in logs

- **Parallel trends assumption (in logs):**

$$\mathbb{E}\left[\ln Y_{i,t=2}(\infty) \mid G_i = 2\right] - \mathbb{E}\left[\ln Y_{i,t=1}(\infty) \mid G_i = 2\right] = \mathbb{E}\left[\ln Y_{i,t=2}(\infty) \mid G_i = \infty\right] - \mathbb{E}\left[\ln Y_{i,t=1}(\infty) \mid G_i = \infty\right]$$

$$\mathbb{E}\left[\ln Y_{i,t=2}(\infty) - \ln Y_{i,t=1}(\infty) \mid G_i = 2\right] = \mathbb{E}\left[\ln Y_{i,t=2}(\infty) - \ln Y_{i,t=1}(\infty) \mid G_i = \infty\right]$$

$$\mathbb{E}\left[\ln \frac{Y_{i,t=2}(\infty)}{Y_{i,t=1}(\infty)} \,\middle|\, G_i = 2\right] = \mathbb{E}\left[\ln \frac{Y_{i,t=2}(\infty)}{Y_{i,t=1}(\infty)} \,\middle|\, G_i = \infty\right]$$

- Under parallel trends (in logs), the ATT would take the format:

$$ATT = \mathbb{E}\left[\ln Y_{i,t=2}(2) - \ln Y_{i,t=2}(\infty) \mid G = 2\right] = \mathbb{E}\left[\ln \frac{Y_{i,t=2}(2)}{Y_{i,t=2}(\infty)} \,\middle|\, G = 2\right].$$

- ATT is measured in relative terms when you have PT in logs.

Parallel Trends in logs

- **Parallel trends assumption (in logs):**

$$\mathbb{E}\left[\ln \frac{Y_{i,t=2}(\infty)}{Y_{i,t=1}(\infty)} \middle| G_i = 2\right] = \mathbb{E}\left[\ln \frac{Y_{i,t=2}(\infty)}{Y_{i,t=1}(\infty)} \middle| G_i = \infty\right]$$

- If Y is the duration of claims measured in weeks, and treatment is an increase of cap (PT in levels)
  - PT would suggest that the average log relative growth of untreated claims' duration among workers who are treated would be the same as the average log relative growth of untreated claims' duration among workers who are not treated
  - If the average log relative growth of untreated claims' duration among workers who are not treated is 0.008, these would serve as counterfactual changes for the average log relative growth of untreated claims' duration for workers that are treated.
  - ATT would provide the average treatment effect (in relative terms) among workers who are treated.

Which PT should we pick?

What if we take other transformations?

The rest of the lecture will build on

Roth and Sant'Anna (2023), but with

different notation.

Setup

Model setup

- We consider the 2x2 DiD setup:
  - 2 time periods: $t = 1$ (before treatment) and $t = 2$ (after treatment);
  - 2 groups: $G = 2$ (treated at period 2) and $G = \infty$ (untreated by period 2);
- Potential outcomes: $Y_{i,t}(2), Y_{i,t}(\infty)$. Observe $Y_{i,t} = 1_{\{G_i=1\}} Y_{i,t}(2) + 1_{\{G_i=\infty\}} Y_{i,t}(\infty)$.
- Let's assume No-anticipation: $Y_{i,t=1}(2) = Y_{i,t=1}(\infty)$.
- Target parameter is the ATT in period $t = 2$,

$$ATT = \mathbb{E}\left[Y_{i,t=2}(2) - Y_{i,t=2}(\infty) \mid G = 2\right].$$

More general models

- We consider a 2-period, 2-group model for expositional simplicity.
- More recent papers have considered settings with multiple periods and staggered adoption.
  - Typically impose a version of the 2-group, 2-period parallel trends assumption for many periods/groups
    (de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Borusyak, Jaravel and Spiess, 2024; Wooldridge, 2021).
  - Thus, 2x2 results have immediate implications for the generalized PT assumption in the staggered case.
- The following results remain valid if all probability statements are implicitly conditional on X, as when one assumes conditional parallel trends
  (Heckman, Ichimura and Todd, 1997; Abadie, 2005; Sant'Anna and Zhao, 2020; Callaway and Sant'Anna, 2021).

Parallel Trends for all transformations of $Y(\infty)$

Parallel Trends and Insensitivity to Functional Form

- Following the definition in Athey and Imbens (2006), we say parallel trends is **insensitive to functional form** (a.k.a. invariant to transformations) if

$$\mathbb{E}\left[s(Y_{i,t=2}(\infty)) \mid G_i = 2\right] - \mathbb{E}\left[s(Y_{i,t=1}(\infty)) \mid G_i = 2\right]$$
$$=$$
$$\mathbb{E}\left[s(Y_{i,t=2}(\infty)) \mid G_i = \infty\right] - \mathbb{E}\left[s(Y_{i,t=1}(\infty)) \mid G_i = \infty\right]$$

  for all strictly monotonic s such that the expectations exist and are finite.

  - s could be levels, logs, percentiles of a reference distribution, etc.

Insensitivity of Parallel Trends

Roth and Sant'Anna (2023) established the following characterization relating PT and functional form.

Proposition (PT and functional form)

Parallel trends is insensitive to functional form if and only if parallel trends of CDFs is satisfied, i.e.

$$\underbrace{F_{Y_{i,t=2}(\infty)|G_i=2}(y) - F_{Y_{i,t=1}(\infty)|G_i=2}(y)}_{\text{Change in CDF for treated group}} = \underbrace{F_{Y_{i,t=2}(\infty)|G_i=\infty}(y) - F_{Y_{i,t=1}(\infty)|G_i=\infty}(y)}_{\text{Change in CDF for comparison group}}, \text{ for all } y \in \mathbb{R} \quad (1)$$

where $F_{Y_{i,t}(\infty)|G_i=g}$ is the cumulative distribution function of $Y_{i,t}(\infty) \mid G_i = g$.

Note that if $Y(\infty)$ is continuous (discrete), this is equivalent to parallel trends of PDFs (PMFs).

What Generates PT of CDFs?

What Generates PT of CDFs?

- Under minor regularity conditions, Roth and Sant'Anna (2023) shows that parallel trends of CDFs holds if and only if

$$F_{Y_{i,t}(\infty)|G_i=g}(y) = \theta J_t(y) + (1-\theta)H_g(y) \text{ for all } y \in \mathbb{R} \text{ and } g \times t \in \{2, \infty\} \times \{1, 2\}. \quad (2)$$

  for some $\theta \in [0, 1]$ and CDFs $J_t(y)$ and $H_g(y)$ depending only on time and group, respectively.

- This says that the distribution of $Y(\infty)$ for group g in period t is a mixture of a time-dependent distribution (not depending on g) and a group-dependent distribution (not depending on t).

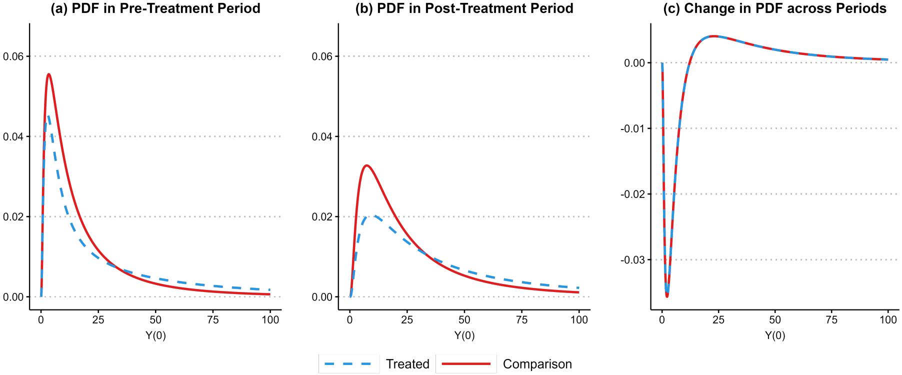
Cases

This implies that PT is insensitive to funct form iff we are in the following three cases:

- **Case 1: (As-If) Randomized Treatment ($\theta = 1$).** The distribution of $Y_{i,t}(\infty)|G = g$ is the same for both groups ($g = 2, \infty$)
- **Case 2: Stationary $Y(\infty)$ ($\theta = 0$).** For each group, the distribution of $Y_{i,t}(\infty)|G = g$ doesn't depend on t.
- **Case 3: A hybrid. ($\theta \in (0, 1)$).**
  θ fraction of the population is as-if randomized btwn treatment and control
  $1 - \theta$ fraction of the population is non-randomized in treatment and control but have stationary $Y(\infty)$ distributions (conditional on group)
  - Perhaps plausible if there is effectively an experiment among a sub-population with time trends (e.g. younger workers), and endogenous selection into treatment among sub-populations with stable earnings over time (e.g. older workers).

Numerical Illustration of Case 3

- $\theta = \frac{1}{2}$ (e.g. share of younger workers)
  $J_t \sim lognormal(1 + t, 1)$ (e.g. wages of younger workers in period t)
  $H_g \sim lognormal(3 + 1_{\{g=2\}}, 1)$ (e.g. wages of older workers in state g)
  $Y_{i,t}(\infty)|G_i = g \sim \theta J_t + (1 - \theta)H_g$



Can we test PT in CDFs?

Testable Implications

- The parallel trends of CDFs condition implies that

$$\underbrace{F_{Y_{i,t=2}(\infty)|G_i=2}(y)}_{\text{Counterfactual}} = \underbrace{F_{Y_{i,t=1}(\infty)|G_i=2}(y) + F_{Y_{i,t=2}(\infty)|G_i=\infty}(y) - F_{Y_{i,t=1}(\infty)|G_i=\infty}(y)}_{\text{Identified}} \text{ for all } y \in \mathbb{R} \tag{3}$$

- A (sharp) testable implication of PT of CDFs is that the RHS is monotonically increasing.

- If the RHS is non-monotonic, then there is no possible counterfactual distribution $Y_{i,t=2}(\infty)|G_i = 2$ such that parallel trends is insensitive to functional form!

- Roth and Sant'Anna (2023) show that we can use this to test for cases where it is clear from data we need to justify the particular choice of functional form

Testing in Practice

- Consider the case where $Y(\infty)$ has finite support.
- Then, testing that the implied CDF is increasing is equivalent to testing that the implied mass is non-negative at all support points, i.e.

$$f_{Y_{i,t=1}|G_i=2}(y) + f_{Y_{i,t=2}|G_i=\infty}(y) - f_{Y_{i,t=2}|G_i=\infty}(y) \geq 0 \text{ for all } y,$$

  where $f_{Y_{i,t}|G_i=g}(y)$ is the probability mass function of $Y_{i,t}|G_i = g$.
- To test, we can merely replace the mass functions with sample analogs and apply tools from the **moment inequality** literature to test that

$$\mathbb{E}[f_{Y_{i,t=1}|G_i=2}(y) + f_{Y_{i,t=2}|G_i=\infty}(y) - f_{Y_{i,t=2}|G_i=\infty}(y)] \geq 0 \text{ for all } y.$$

- With continuous support, can likewise use methods for testing a continuum of inequalities (e.g. Andrews and Shi (2013)).

Caveats

- These tests may be useful for detecting when parallel trends is sensitive to functional form.

- **But failure to reject does not mean that we don't need to worry about functional form!**

- PT of CDFs is *falsifiable but not verifiable*:
  - Null is that there is some possible distribution for $Y_{i,t=2}(\infty)|G_i = 2$ such that it holds.

- Like tests of pre-trends, such pre-tests may be *underpowered*, and relying on them can introduce distortions from *pre-testing* (Roth, 2022).

EMORY UNIVERSITY

Empirical Illustration

Empirical Illustration

- Stylized analysis of the impact of state-level minimum wage changes on wage distribution

- Testing PT of CDFs is interesting both because it determines whether PT is sensitive to functional form and because DiD has been used to estimate distributional impacts in this context.

- Set-up:
  - The pre-period is either 2007 or 2010. Post-period is 2015
  - Treatment is whether the state raised MW between Pre and Post.

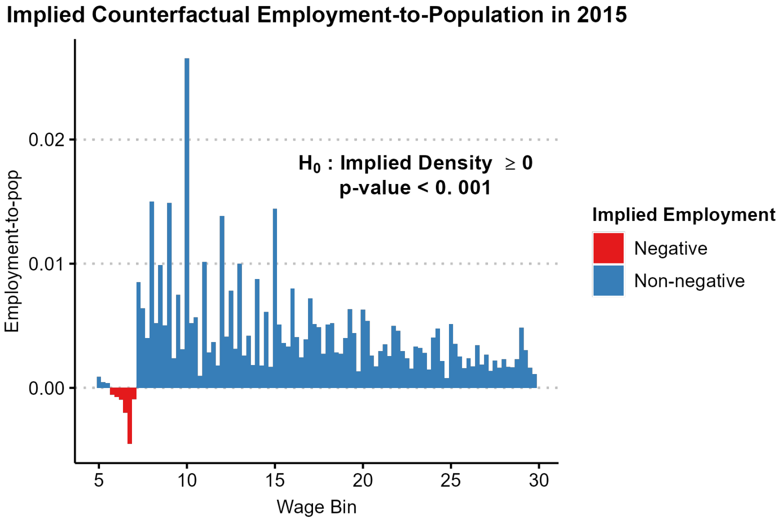
Empirical Illustration

- Panel data from Cengiz, Dube, Lindner and Zipperer (2019) with state-level MW changes and employment-to-population ratios for 25c wage-bins (in 2016 dollars) at state-level
- If W_i is person i's wage if employed and 0 otherwise, then employment-to-pop ratio at wage w is density of W_i at w.
- Estimate counterfactual employment-to-population ratio in bin w under PT of CDFs as:

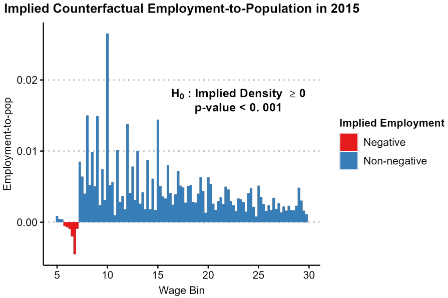
$$\hat{f}_{post,D=1}(w) = \hat{f}_{pre,D=1}(w) + \hat{f}_{post,D=0}(w) - \hat{f}_{pre,D=0}(w),$$

  weighting states by population size
- Conduct moment inequality tests by comparing the minimum studentized moment to "least-favorable" critical values (assuming all moments have mean zero)

Results: Pre = 2007, Post = 2015

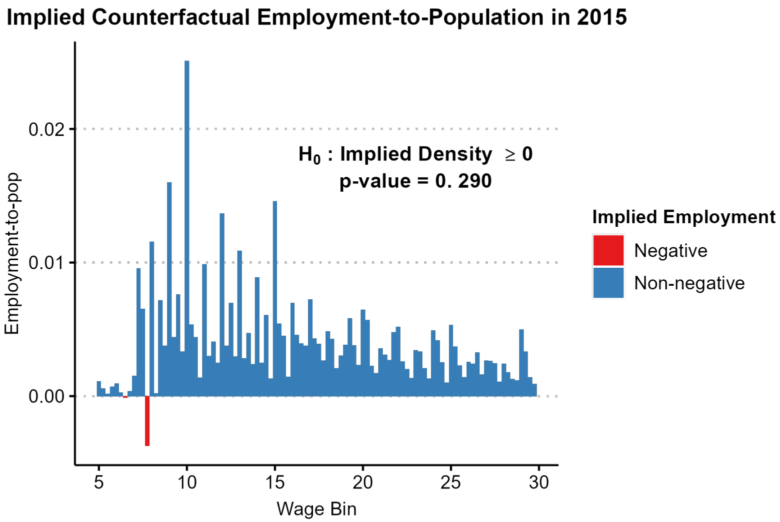


Results: Pre $= 2007$, Post $= 2015$



- Implied density is negative for wages \$5-7.
- Intuitively, employment declines in control states are larger than initial levels in treatment states (likely b/c of differential effects of change in federal MW)

Results: Pre = 2010, Post = 2015



R package

- Jon Roth and I have prepared the R package didFF to help you use these tests.
- The package covers a variety of setups:
  - Multiple time periods;
  - Staggered treatment adoption;
  - PT plausible only after conditioning on covariates.
- Please check it out at `https://github.com/pedrohcgs/didFF`

References

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

Andrews, Donald W. K. and Xiaoxia Shi, “Inference Based on Conditional Moment Inequalities,” *Econometrica*, 2013, *81* (2), 609–666.

Athey, Susan and Guido Imbens, “Identification and Inference in Nonlinear Difference-in-Differences Models,” *Econometrica*, 2006, *74* (2), 431–497.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, “Revisiting Event Study Designs: Robust and Efficient Estimation,” *Review of Economic Studies*, 2024, *Forthcoming*.

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

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, “The Effect of Minimum Wages on Low-Wage Jobs,” *The Quarterly Journal of Economics*, August 2019, *134* (3), 1405–1454.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd, "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, October 1997, *64* (4), 605–654.

Meyer, Bruce D., W. Kip Viscusi, and David L. Durbin, "Workers' Compensation and Injury Duration: Evidence from a Natural Experiment," *The American Economic Review*, 1995, *85* (3), 322–340.

Roth, Jonathan, "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends," *American Economic Review: Insights*, 2022, *Forthcoming.*

— **and Pedro H. C. Sant'Anna**, "When Is Parallel Trends Sensitive to Functional Form?," *Econometrica*, 2023, *91* (2), 737–747.

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

Sun, Liyan and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2).

Wooldridge, Jeffrey M, "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators," *Working Paper*, 2021, pp. 1–89.