Identification and Estimation

Linear Panel Event Studies

Liyang (Sophie) Sun (CEMFI)
Jesse M. Shapiro (Harvard and NBER)

Outline

- Lay out key identifying assumptions for the simplest difference-in-differences estimator
  - “no anticipation” assumption and its economic content
  - “parallel trends” assumption and its economic content
- Generalize assumptions for popular extensions to the estimator when
  - treatment lasts several periods
  - treatment is introduced to different units at different times

Basics

Classical example: Card and Krueger (1994)

- Measured employment before and after minimum wage increase for a sample of fast-food restaurants
- Motivated difference-in-differences (DID) estimator by the following

> Moreover, since seasonal patterns of employment are similar in New Jersey and eastern Pennsylvania, as well as across high- and low-wage stores within New Jersey, our comparative methodology effectively "differences out" any seasonal employment effects.

A simple DID estimator

- Table 3 of Card and Krueger (1994)

Variable	PA (i)	NJ (ii)	Difference, NJ − PA (iii)
	Stores by state		
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	−2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	−0.14 (1.07)
3. Change in mean FTE employment	−2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

- Binary treatment D_i
  - For PA, $D_i = 0$; for NJ, $D_i = 1$
- Two periods: $t \in \{-1, 0\}$ and treatment is implemented at $t = 0$
- Four sample averages of the outcome $\bar{y}_{t,D}$:
  - before vs after and PA vs NJ

A simple DID estimator

- Row 3 Column (iii) is their DID estimate

Variable	Stores by state PA (i)	NJ (ii)	Difference, NJ − PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	−2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	−0.14 (1.07)
3. Change in mean FTE employment	−2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

- We can write the estimator as

$$\hat{\beta}^{\text{DID}} = \left(\bar{y}_{t=0,D=1} - \bar{y}_{t=-1,D=1}\right) - \left(\bar{y}_{t=0,D=0} - \bar{y}_{t=-1,D=0}\right)$$

What is the DID estimator estimating?

- The DID estimator is

$$\hat{\beta}^{\text{DID}} = (\bar{y}_{t=0,D=1} - \bar{y}_{t=-1,D=1}) - (\bar{y}_{t=0,D=0} - \bar{y}_{t=-1,D=0})$$

- Potential outcomes $y_{i,t}(d)$ for $d \in \{0, 1\}$
  - The employment that would have been if minimum wage increased ($d = 1$) and did not increase ($d = 0$)
  - For PA, observe $y_{i,t}(0)$; for NJ, observe $y_{i,t}(1)$
- Interested in the average impact for NJ after the minimum wage increased, formally, the average treatment effect on the treated

$$\text{ATT: } E[\overbrace{\underbrace{y_{i,0}(1)}_{\text{observed}} - \underbrace{y_{i,0}(0)}_{\text{counterfactual}}}^{\text{treatment effect}} \mid D_i = 1]$$

- Since counterfactual outcomes are never observed, we need to impose some assumptions to estimate the ATT

Sufficient assumptions (1): No anticipation

- "no anticipation" assumption:
  - the outcome is not affected by the treatment prior to its implementation: $y_{i,-1}(0) = y_{i,-1}(1)$ for all i with $D_i = 1$
- Assuming "no anticipation," outcomes we observe $y_{i,t}$ can be written as

	PA $D_i = 0$	NJ $D_i = 1$
before $t = -1$	$y_{i,-1}(0)$	$y_{i,-1}(0)$
after $t = 0$	$y_{i,0}(0)$	$y_{i,0}(1)$

- Example violation: fast food restaurants laying off minimum wage workers in advance of increase in wage
  - Other examples: consumption smoothing for anticipated job loss (Hendren 2017)

Sufficient assumptions (2): Parallel trends

- "parallel trends" assumption:

$$
\begin{aligned}
&E[y_{i,0}(0) - y_{i,-1}(0) \mid D_i = 1] && \text{(NJ counterfactual trend)} \\
=&E[y_{i,0}(0) - y_{i,-1}(0) \mid D_i = 0] && \text{(PA trend)}
\end{aligned}
$$

  - if minimum wage never increased for NJ, average trends would coincide between NJ and PA
- Example violation: NJ labor market was improving compared to PA
  - Other examples: downward trend in wage income leading to participation in job training programs (Ashenfelter's dip)

Sufficient assumptions (2): Parallel trends

- Very different from the unconfoundedness assumption that is common in RCTs:
  - Random assignment $\{y_i(1), y_i(0)\} \perp D_i$
- Parallel trends assumption allows for potentially non-zero selection bias:

$$\underbrace{E[y_{i,-1}(0) \mid D_i = 1] - E[y_{i,-1}(0) \mid D_i = 0]}_{\text{selection bias at } t=-1}$$
$$= \underbrace{E[y_{i,0}(0) \mid D_i = 1] - E[y_{i,0}(0) \mid D_i = 0]}_{\text{selection bias at } t=0}$$

- Sensitive to the scale: if parallel trends holds for level of employment, it might fail for log of employment, and vice versa (Roth and Sant'Anna 2023)

DID is unbiased for ATT

- The DID estimator

$$\hat{\beta}^{\text{DID}} = \left(\bar{y}_{t=0,D=1} - \bar{y}_{t=-1,D=1}\right) - \left(\bar{y}_{t=0,D=0} - \bar{y}_{t=-1,D=0}\right)$$

- is therefore unbiased for

$$\begin{aligned}
& E[y_{i,0} - y_{i,-1} \mid D_i = 1] - E[y_{i,0} - y_{i,-1} \mid D_i = 0] \\
=& E[y_{i,0}(1) - \underbrace{y_{i,-1}(0)}_{\text{"no anticipation"}} \mid D_i = 1] - E[y_{i,0}(0) - y_{i,-1}(0) \mid D_i = 0] \\
=& \underbrace{E[y_{i,0}(1) - y_{i,0}(0) \mid D_i = 1]}_{\text{ATT}} + \\
& \underbrace{E[y_{i,0}(0) - y_{i,-1}(0) \mid D_i = 1] - E[y_{i,0}(0) - y_{i,-1}(0) \mid D_i = 0]}_{=0 \text{ under "parallel trends"}}
\end{aligned}$$

Estimation with two periods

Regression representation

- Recall the DID estimator:

$$\hat{\beta}^{\text{DID}} = \left(\bar{y}_{t=0,D=1} - \bar{y}_{t=-1,D=1}\right) - \left(\bar{y}_{t=0,D=0} - \bar{y}_{t=-1,D=0}\right)$$

- Can implement via regression as follows
- Define z_{it} as
  - 1 if i is treated ($D_i = 1$) and t is after treatment ($t = 0$)
  - 0 otherwise
- Estimate $y_{it} = \alpha_d + \gamma_t + \beta z_{it} + \varepsilon_{it}$
  - Group fixed effect α_d for $d \in \{0, 1\}$
  - Time fixed effect γ_t
- The OLS estimate $\hat{\beta}$ is numerically equivalent to $\hat{\beta}^{\text{DID}}$

Grouped data and repeated cross sections

- This regression representation is also useful for non-panel datasets
- For repeated cross sections, $\hat{\beta}^{\text{DID}}$ still unbiased estimate of ATT and so is the regression representation
- We can also collapse to group-level and obtain group-level panel data
  - WLS coincides exactly with $\hat{\beta}^{\text{DID}}$

Two-way Fixed Effects (TWFE)

- Common to implement DID via Two-way Fixed Effects (TWFE) regression
- Estimate $y_{it} = \alpha_i + \gamma_t + \beta z_{it} + \varepsilon_{it}$
  - Unit fixed effect α_i
  - Time fixed effect γ_t
- Large subsequent literature on minimum wage (for example, Neumark and Wascher 2007) estimates this model allowing for continuous treatment, covariates, multiple time periods, etc.
- Will return to some of these topics later in lecture
- For now focus on **multiple time periods**

Estimation with multiple periods

Multiple periods but one treatment group

- For example, Seattle minimum wage increase (Jardim et al. 2022)
- Suppose for those $D_i = 1$, treatment starts at $t = t^*$
- Define z_{it} as before
  - 1 if i is treated ($D_i = 1$) and t is after treatment ($t \geq t^*$)
  - 0 otherwise
- Then relative time indicator $\Delta z_{i,t-k} = 1$ if treatment happens in period $t - k$
  - $k = 0$: contemporaneous
  - $k > 0$: indicator for start of treatment k periods ago
  - $k < 0$: indicator for start of treatment $|k|$ periods in future

“Dynamic” specification

- Estimate a “dynamic” specification

$$y_{it} = \alpha_i + \gamma_t + \sum_{-\infty}^{\infty} \delta_k \Delta z_{i,t-k} + \varepsilon_{it}$$

  - Unit (or group) fixed effect α_i
  - Time fixed effect γ_t
  - Normalize $\delta_{-1} = 0$ so δ_k is in normalized differences
- Each regression coefficient estimator can still be thought of as a DID estimator:

$$\hat{\delta}_k = \left(\bar{y}_{t=t^*+k,D=1} - \bar{y}_{t=t^*-1,D=1}\right) - \left(\bar{y}_{t=t^*+k,D=0} - \bar{y}_{t=t^*-1,D=0}\right)$$

Generalized “no anticipation” and “parallel trends”

- “No anticipation” assumption:
  - Treatment has no causal effect prior to its implementation: $y_{it}(0) = y_{it}(1)$ for all i with $D_i = 1$ for all $t < t^*$
- “Parallel trends” assumption:

$$E[y_{it'}(0) - y_{it}(0) \mid D_i = 1] = E[y_{it'}(0) - y_{it}(0) \mid D_i = 0]$$

  for all $t \neq t'$

- Under “no anticipation” and “parallel trends”, can interpret δ_k for $k \geq 0$ as cumulative ATT:

$$E[y_{i,t^*+k}(1) - y_{i,t^*+k}(0) \mid D_i = 1]$$

- But also implies δ_k for $k < -1$ should be zero, which is the basis for pre-trends testing that we discuss later

Multiple periods and multiple treatment groups

- For example, minimum wage increase was introduced gradually across states
- Suppose we want to estimate the impact of having experienced any increase in minimum wage ("staggered adoption")
  - Recall that $z_{it} \in \{0, 1\}$ indicates whether unit i is treated in period t
  - "Staggered adoption" implies that z_{it} is non-decreasing in t
- Can categorize units uniquely into treatment timing groups by $g(i) = \min\{t : z_{it} = 1\}$, the earliest period at which unit i has received treatment
- Takes on values $g(i) \in \{0, 1, \ldots, \infty\}$ where
  - ∞ is never-treated

Generalized "no anticipation" and "parallel trends"

- "no anticipation" assumption:
  - Treatment has no causal effect prior to its implementation: $y_{it}(\infty) = y_{it}(g)$ for all i for all $t < g(i)$
- "parallel trends" assumption (strong version):

$$E[y_{it'}(\infty) - y_{it}(\infty) \mid g(i) = g] = E[y_{it'}(\infty) - y_{it}(\infty) \mid g(i) = \infty]$$

for all $t \neq t'$ and for all adoption groups $g < \infty$

  - The never-treated counterfactual would evolve in parallel for all adoption groups, as well as the never-treated group

Generalized "no anticipation" and "parallel trends"

- "no anticipation" assumption:
  - Treatment has no causal effect prior to its implementation: $y_{it}(\infty) = y_{it}(g)$ for all i for all $t < g(i)$
- "parallel trends" assumption (weak version):

$$E[y_{it'}(\infty) - y_{it}(\infty) \mid g(i) = g] = E[y_{it'}(\infty) - y_{it}(\infty) \mid g(i) = g']$$

for all $t \neq t'$ and for all adoption groups $g, g' < \infty$

  - The never-treated counterfactual would evolve in parallel for all adoption groups
  - Might not be parallel with the never-treated group

Estimate the dynamic effect: staggered adoption

- Under "staggered adoption", suppose we estimate a "dynamic" specification

$$y_{it} = \alpha_i + \gamma_t + \sum_{-\infty}^{\infty} \delta_k \Delta z_{i,t-k} + \varepsilon_{it}$$

- In addition to "no anticipation" and "parallel trends for staggered setting", this specification also restricts homogeneity on treatment effects
    - The dynamic effect δ_k only depends on the relative time k, but not on the treatment timing g
    - Return later to cases where homogeneity is violated
- Can summarize "overall" ATT by taking averages of the estimated δ_k for $k \geq 0$

Estimate the “overall” effect: staggered adoption

- Another option is to estimate a **static** model

$$y_{it} = \alpha_i + \gamma_t + \beta_{post} z_{i,t} + \varepsilon_{it}$$

  - β_{post} is the correct summary for the “overall” effect if treatment effects are truly **static**, i.e., if $\delta_k = \mathbf{1}_{k \geq 0} \beta_{post}$
  - If treatment effects are not **static**, then misspecified
- In settings without a never-treated group, recent work found cases with severe misspecification:
  - Coefficient β_{post} may not correspond to *any* proper average of δ_k
  - For example, $\beta_{post} < 0$ even though $\delta_k > 0$, and vice versa
- See de Chaisemartin and D’Haultfœuille (2020) and Goodman-Bacon (2021) for diagnostics

Estimate the “overall” effect: staggered adoption

- For illustration, de Chaisemartin and D’Haultfœuille (2020) constructed an example of two adoption groups and three time periods
  - Suppose $\delta_0 = 1$ and $\delta_1 = 4$, can summarize the “overall” effect appropriately based on the **dynamic** model
  - Only relying on the **static** model can be misleading because β_{post} is equal to $-1/2$ instead
- If report estimates from both **static** and **dynamic** model, can combine these estimates into one while staying agnostic about the degree of misspecification by applying Armstrong, Kline and Sun (2023)

Alternative identifying assumptions

Alternative identifying assumptions

- If “parallel trends” assumption is not applicable, many proposals for alternative identifying assumptions:
    - Methods Lecture 2007 (Change-in-Changes, Semiparametric Difference-in-Differences,...)
    - Methods Lecture 2021 (Synthetic Controls)
- *Details are beyond the scope of this lecture*

Cohort comparison

- Instead of panel data or repeated cross sectional data,
  - Observe one cross section where units can be categorized by birth cohorts $g(i)$
- Sometimes leverage cross-cohort comparison, for example, Duflo (2001)
  - Late cohort in the treatment group is more exposed to the treatment than the early cohort
- de Chaisemartin and D'Haultfœuille (2017) propose alternative assumptions
  - Intuitively, cohort $g(i)$ plays the role of calendar time, but many differences

Randomized treatment / timing

- Treatment is randomized, and observe past outcomes
  - For example, baseline surveys
- Or the treatment timing is randomized, even though all units eventually receive treatment
  - For example, Parker, Souleles, Johnson, and McClelland (2013)
- Can now rely on random assignment $\{y_{i,t}(1), y_{i,t}(0)\} \perp g(i)$, which is invariant to scale
  - For example, McKenzie (2012) and Roth and Sant'Anna (Forthcoming)

Conclusion

Further Reading

Armstrong, Tim, Patrick Kline, and Liyang Sun. 2023. Adapting to Misspecification. In *arxiv [econ].*

de Chaisemartin, Clément and Xavier D'Haultfoeuille. 2017. Fuzzy difference-in-differences. In *The Review of Economic Studies.*

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. In *American Economic Review*.

Goodman-Bacon, Andrew. 2021. Difference-in-differences with variation in treatment timing. In *Journal of Econometrics*.

McKenzie, David. 2012. Beyond baseline and follow-up: The case for more T in experiments. In *Journal of Development Economics.*

Roth, Jonathan and Pedro H. C. Sant'Anna. Forthcoming. Efficient Estimation for Staggered Rollout Designs. In *Journal of Political Economy: Microeconomics*.

Roth, Jonathan and Pedro H. C. Sant'Anna. 2023. When Is Parallel Trends Sensitive to Functional Form? In *Econometrica*.

Today

- ~~Overview (Jesse)~~
- ~~Basics of identification and estimation (Liyang)~~
- Basics of plotting (Jesse)
- Pitfalls and some solutions
  - Confounds and pre-trend testing (Liyang)
  - Heterogeneous effects (Jesse)
- Conclusions (Liyang)