Difference-in-Differences IV: Staggered Treatment Adoption and Remedies

Lingguo Cheng

Nanjing University

December 22, 2025

Introduction

- Staggered Treatment Adoption: Treated groups adopt treatment at different times; $N(N > 2)$ groups, $T(T > 2)$ period; for example, suppose sulfa drugs were introduced in different states at different times.
- TWFE: Two-way fixed effect model, with group and period fixed effects.
- Goodman-Bacon (2018, 2021) shows the combination of staggered adoption and time-varying treatment effects can introduce confounded comparison into the TWFE regression estimator.
- It is helpful to conceive of a staggered adoption design as a collection of simpler DID comparisons and to take control over the subexperiments that contribute to their analysis.

A stylized Example

- A hypothetical example: Sulfa drags were introduced in all US states in 1937. Instead, we suppose the US states gained access to sulfa drugs in 1930, 1940, 1945, and 1950, respectively, along with a group of states that never gained access.
- (Treatment Effect Heterogeneity) Treatment effect could vary with time since treatment, geography, calendar time of treatment, etc.
  - Varies over time in that treatment gradually reduce (time-varying treatment effects)
  - Varies by the year in which treatment was introduced (i.e., by timing group) in that treatment that began earlier is more effective that treatment that began later.
  - Different timing groups vary in their mortality rate, introducing a source of geographic heterogeneity.

A Stylized Example

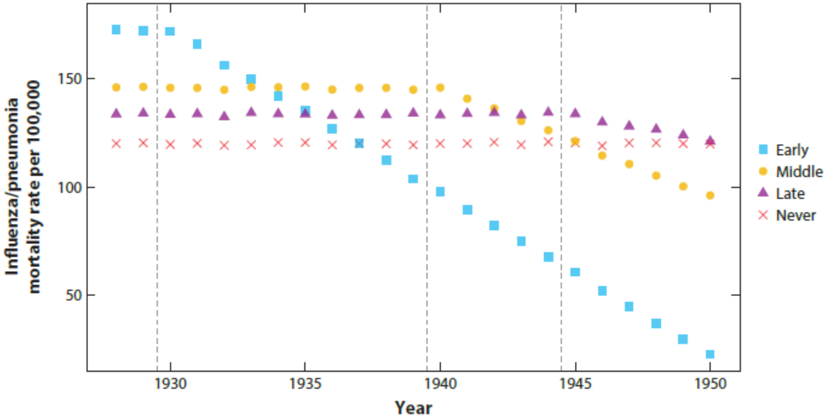

Figure 1: Pneumonia/influenza mortality and sulfa drugs

Source: Wing et al.(2024)

- It helps to view the staggered adoption design as a collection of simpler 2×2 DID designs, which we refer to as subexperiments.
- Researchers should develop principles for actively deciding which 2×2 DID designs will contribute to the analysis and be on guard against DID comparisons that may be confounded.
  - Use $i = 1, \ldots, N$ to index individual observations; $s = 1, \ldots, S$, the collection of groups, and $t = T_1, \ldots, T_T$, calendar time periods.
  - A_s, the calendar period when units in group s are first exposed to treatment, and set $A_s = \infty$ for groups that never adopt treatment during the study period
  - $D_{st} = \mathbf{1}\{t \geq A_s\}$, a binary treatment variable indicating whether treatment is active in group s in period t.

- $Y_{ist}(0)$ represents the outcome person i from group s would experience in calendar period t under a hypothetical scenario in which the person's group is never exposed to treatment.
- $Y_{ist}(a)$ represents the outcome that the same person would experience at t if they were first exposed to treatment in calendar period a.
- The causal effect of adopting treatment in period a compared to never adopting treatment is

$$\beta_{ist}(a) = Y_{ist}(a) - Y_{ist}(0)$$

  Notice the subscripts in $\beta_{ist}(a)$, which emphasize that the treatment effect can differ across units, groups, and time periods.
- The "average treatment effect on the treated" (ATT) evaluated at a particular calendar date is

$$ATT(a,t) = E[\beta_{ist}(a)|A_s = a].$$

- The realized outcome depends on the adoption date, so that we have

$$Y_{ist} = Y_{ist}(0) + \sum_{a=1}^{T} \mathbf{1}\{A_s = a\} \times \mathbf{1}\{t \geq a\}\beta_{ist}(a).$$

- This is the untreated outcome plus the treatment effect if the treatment is actually in place.

- **Assumption 1 (No anticipation or strict exogeneity):** The average causal effect of adopting treatment in period a is equal to zero for all calendar periods prior to period a. For periods $t < a$, we have

$$E[Y_{ist}(a) - Y_{ist}(0) \mid A_s = a] = 0$$

- **Comment:**
  - The strict exogeneity assumption in the context of panel data models
  - Often referred to as no pretrends assumption
  - Could fail if treatment exposure occurs in response to volatility in the outcome variable or if the behavior changes due to expectations of future treatment.

- **Assumption 2 (Common trends):** In the absence of treatment exposure, the average change across posttreatment time periods would be the same in the treatment group ($A_s = a$) and the comparison group ($A_s > a$). For periods $t > a$, we have

$$
\begin{aligned}
E[Y_{ist}(0) - Y_{ist-1}(0) \mid A_s = a]& \\
= E[Y_{ist}(0) - Y_{ist-1}(0) \mid A_s > a]&
\end{aligned}
$$

- Comments:
  - The treatment group makes up of all units that adopt treatment in period a; the control group, have not yet adopted treatment, including never-treated and/or not-yet-treated.
  - $E[Y_{ist}(0) - Y_{ist-1}(0) \mid A_s = a](t > a)$ represents the time trend that the treated group would have experienced in the absence of treatment exposure. This is a counterfactual that we cannot observe directly. The common trends assumption implies that the counterfactual trend is equal to the observed trend in the control group.

2×2 DID

- The 2×2 DID design has two periods ($t = [1, 2]$) and two groups ($s = [1, 2]$). The first group is never treated, so we have $A_1 = \infty$. The second group has $A_2 = 2$, meaning it is first exposed in period 2. Periods 1 and 2 are the pre- and post-period, respectively.
- The DID estimator is the difference between the expected pre–post change in realized outcomes in the treatment group and control group:

$$\begin{aligned}
\Delta_{DID} &= E[Y_{i22} - Y_{i21} | A_s = 2] - E[Y_{i12} - Y_{i11} | A_s = \infty] \\
&= E[Y_{i22}(2) - Y_{i21}(2) | A_s = 2] - E[Y_{i12}(0) - Y_{i11}(0) \mid A_s = \infty] \\
&= E[Y_{i22}(2) - Y_{i21}(0) | A_s = 2] - E[Y_{i12}(0) - Y_{i11}(0) | A_s = \infty] \\
&= E[\beta_{i22}(2) | A_s = 2] \\
&+ \{E[Y_{i22}(0) - Y_{i21}(0) | A_s = 2] - E[Y_{i12}(0) - Y_{i11}(0) | A_s = \infty]\} \\
&= E[\beta_{i22}(2) | A_s = 2] \\
&= ATT(2, 2)
\end{aligned}$$

- Adding structure makes the DID more intuitive. Write the untreated outcome as $Y_{ist}(0) = c_s + b_t + \epsilon_{ist}$. Then the treated outcome is $Y_{ist}(a) = Y_{ist}(0) + \beta_{ist}(a)$. Clearly, the time trend is $E[Y_{is2}(0) - Y_{is1}(0)] = b_2 - b_1$ in both groups: That is what the common trends assumption looks like in this case. The group difference is $E[Y_{i2t}(0) - Y_{i1t}(0)] = c_2 - c_1$ in both periods, showing that the DID assumptions do not require the groups to be comparable. What matters is that disparities do not change over time.
- The 2×2 DID can also be formed using regressions. Define $\text{Treat}_s = \mathbf{1}(A_s = 2)$ and $\text{Post}_t = \mathbf{1}(t = 2)$ and estimate an ordinary least squares (OLS) regression

$$Y_{ist} = \beta_0 + \beta_1 \text{Treat}_s + \beta_2 \text{Post}_t + \beta_3 (\text{Treat}_s \times \text{Post}_t) + \epsilon_{ist}.$$

In the 2×2 DID case, we have $\beta_3 = \Delta_{DID}$.

- Alternatively, we could represent TWFE regressions as

$$Y_{ist} = \beta_{FE} D_{st} + c_s + b_t + \epsilon_{ist},$$

where c_s and b_t represent unobserved fixed effects and $\beta_{FE} = \Delta_{DID}$.

Two-group Event Studies

- A basic event study has two groups $(s = [1, 2])$ but multiple time periods $t = T_1, \ldots, T_T$. In our $2 \times T$ example, group 1 (never treated) has $A_1 = \infty$. Group 2 is treated at $A_2 = a$, with $a > T_1$. The pre-period runs from T_1 to $a - 1$, and the post-period runs from $t = a$ to T_T.
- By adding periods to the 2×2 design, the event study makes it possible to learn more about time-varying treatment effects and also enables some partial tests of the core DID assumptions.
- We use $ATT(a, t^*) = E[Y_{ist^*}(a) - Y_{ist^*}(0) \mid As = a]$ to represent the average effect of adopting treatment in period a on outcomes experienced in calendar period t^* among units in ***timing group*** $A_s = a$. With this notation, $ATT(a, a)$ represents the immediate effect, and $ATT(a, a + k)$ represents the effect k periods after initial adoption.

- Under Assumptions 1 (no anticipation) and 2 (common trends), $ATT(a, a+k)$ is identified for each $k = 0, \ldots, T_{T-a}$.
- The DID estimator of $ATT(a, a+k)$ is

$$
\begin{aligned}
\delta_{ES}^{a+k} &= E[Y_{i2,a+k} - Y_{i2,a-1} \mid A_s = a] - E[Y_{i1,a+k} - Y_{i1,a-1} \mid A_s = \infty] \\
&= E[Y_{i2,a+k}(a) - Y_{i2,a-1}(a) \mid A_s = a] - E[Y_{i1,a+k}(0) - Y_{i1,a-1}(0) \mid A_s = \infty] \\
&= E[Y_{i2,a+k}(a) - Y_{i2,a-1}(0) \mid A_s = a] - E[Y_{i1,a+k}(0) - Y_{i1,a-1}(0) \mid A_s = \infty] \\
&= E[Y_{i2,a+k}(a) - Y_{i2,a+k}(0) + Y_{i2,a+k}(0) - Y_{i2,a-1}(0) \mid A_s = a] \\
&\quad - E[Y_{i1,a+k}(0) - Y_{i1,a-1}(0) \mid A_s = \infty] \\
&= E[\beta_{i2,a+k}(a) \mid A_s = a] \\
&\quad + E[Y_{i2,a+k}(0) - Y_{i2,a-1}(0) \mid A_s = a] - E[Y_{i1,a+k}(0) - Y_{i1,a-1}(0) \mid A_s = \infty] \\
&= E[\beta_{i2,a+k}(a) \mid A_s = a]
\end{aligned}
$$

- Under Assumptions 1 (no anticipation) and 2 (common trends), $ATT(a, a+k)$ is identified for each $k = 0, \ldots, T_{T-a}$.
- The DID estimator of $ATT(a, a+k)$ is

$$
\begin{aligned}
\delta_{ES}^{a+k} &= E[Y_{i2,a+k} - Y_{i2,a-1} \mid A_s = a] - E[Y_{i1,a+k} - Y_{i1,a-1} \mid A_s = \infty] \\
&= E[Y_{i2,a+k}(a) - Y_{i2,a-1}(a) \mid A_s = a] \\
&- E[Y_{i1,a+k}(0) - Y_{i1,a-1}(0) \mid A_s = \infty] \\
&= E[Y_{i2,a+k}(a) - Y_{i2,a-1}(0) \mid A_s = a] \\
&= E[Y_{i2,a+k}(a) - Y_{i2,a+k}(0) + Y_{i2,a+k}(0) - Y_{i2,a-1}(0) \mid A_s = a] \\
&- E[Y_{i1,a+k}(0) - Y_{i1,a-1}(0) \mid A_s = \infty] \\
&= E[\beta_{i2,a+k}(a) \mid A_s = a] \\
&+ E[Y_{i2,a+k}(0) - Y_{i2,a-1}(0) \mid A_s = a] - E[Y_{i1,a+k}(0) - Y_{i1,a-1}(0) \mid A_s = \infty] \\
&= E[\beta_{i2,a+k}(a) \mid A_s = a]
\end{aligned}
$$

- Similarly, we can estimate the pre-period ATTs using

$$\delta_{ES}^{a-h} = E[Y_{i2,a-h} - Y_{i2,a-1} \mid A_s = a] - E[Y_{i1,a-h} - Y_{i1,a-1} \mid A_s = \infty]$$

- Rejecting the null hypothesis that these pre-period DIDs are equal to zero implies that the common trends and no anticipation assumptions are not met.
- In practice, it is convenient to estimate these event study DIDs using a single linear regression:

$$\begin{aligned} Y_{ist} = &\sum_{h=1}^{a-1} \alpha_h \mathbf{1}[A_s = a] \times \mathbf{1}[t = b] + \sum_{k=a}^{T_T} \beta_k \mathbf{1}[A_s = a] \times \mathbf{1}[t = k] \\ &+ c_s + b_t + \epsilon_{ist} \end{aligned}$$

- In this specification, each $\beta_k = \delta_{ES}^{a+k} = ATT(a, a+k)$ and each $\alpha_h = \delta_{ES}^{a-h} = ATT(a, a-h)$.

Staggered Adoption Designs

- A law that has been adopted in a set of states at different times
- The workhorse specification is the two-way fixed effects model:

$$Y_{ist} = \beta_{FE} D_{st} + c_s + b_t + \epsilon_{ist}.$$

- Comments:
  - Recent work by Goodman-Bacon (2021) and de Chaisemartin & D'Haultfoeuille (2021) provides a clearer account of how the FE parameter represents a variance of treatment–weighted combination of underlying heterogeneous effects.
  - At a broad level, the recent literature shifts the focus away from matters of statistical modeling and toward the research design itself.

The key building block for all staggered DD estimation

- The group $\times$ time treatment effect, $ATT(a,t)$, is a key building block for interpreting the staggered adoption design. The group-time ATT has the same meaning in the staggered adoption case as it did in the 2×2 and event study cases. The difference is that the staggered design distinguishes between multiple $ATT(a,t)$ parameters because there are more adoption groups and periods.
- In principle, the staggered adoption design makes it possible to identify a collection of different $ATT(a,t)$ parameters using the same no anticipation and common trends assumptions used in the simpler designs. The trick is to apply the standard DID estimator to the correct combination of periods and groups.

- For a generic $ATT(a,t)$ effect, the DID comparison is

$$
\begin{aligned}
&\delta_{SA}^{a+k} \\
&\quad = E[Y_{is,a+k} - Y_{is,a-1} | A_s = a] - E[Y_{is,a+k} - Y_{is,a-1} | A_s > a+k] \\
&\quad = E[Y_{is,a+k}(a) - Y_{is,a-1}(a) | A_s = a] - E[Y_{is,a+k}(0) - Y_{is,a-1}(0) | A_s > a+k] \\
&\quad = E[Y_{is,a+k}(a) - Y_{is,a-1}(0) | A_s = a] - E[Y_{is,a+k}(0) - Y_{is,a-1}(0) | A_s > a+k] \\
&\quad = E[Y_{is,a+k}(a) - Y_{is,a+k}(0) + Y_{is,a+k}(0) - Y_{is,a-1}(0) | A_s = a] \\
&\qquad - E[Y_{is,a+k}(0) - Y_{is,a-1}(0) | A_s > a+k] \\
&\quad = E[\beta_{is,a+k}(a) | A_s = a] + E[Y_{is,a+k}(0) - Y_{is,a-1}(0) | A_s = a] \\
&\qquad - E[Y_{is,a+k}(0) - Y_{is,a-1}(0) | A_s > a+k] \\
&\quad = E[\beta_{is,a+k}(a) | A_s = a] \\
&\quad = ATT(a, a+k)
\end{aligned}
$$

Understanding of Threats to Validity

- Goodman-Bacon (2021) showed the TWFE estimator is a weighted average of these underlying DIDs, working out the weights assigned to each underlying DID comparison, which is helpful in understanding which policy changes drive the overall estimate.
- Goodman-Bacon (2021) categorizes the 2×2 DID comparisons that contribute to the TWFE estimator into three types: (a) treated vs. never-treated DIDs $(A_s = a\, vs.\, A_s = \infty)$, (b) early vs. late DIDs $(A_s = a\, vs.\, A_s = c > a)$, and (c) late vs. early DIDs $(A_s = a\, vs.\, A_s = b < a)$. These three types of comparisons are weighted together to form a single summary coefficient, β_{FE}.

- The type of late adopter vs. earlier adopter DID can create problems if treatment effects vary with time since the event. For example, consider a DID in which timing group $A_s = a$ is compared to an already treated control group with $A_s = b < a$:

$$
\begin{aligned}
\Delta^{a,a+k}_{Bad} &= E[Y_{is,a+k} - Y_{is,a-1}|A_s = a] - E[Y_{is,a+k} - Y_{is,a-1}|A_s = b] \\
&= E[\beta_{is,a+k}(a) + Y_{is,a+k}(0) - Y_{is,a-1}(0)|A_s = a] \\
&\quad - E[(Y_{is,a+k}(0) + \beta_{is,a+k}(b)) - (Y_{is,a-1}(0) + \beta_{is,a-1}(b))|A_s = b] \\
&= E[\beta_{is,a+k}(a)|A_s = a] + E[\beta_{is,a+k}(b) - \beta_{is,a-1}(b)|A_s = b] \\
&\quad + (E[Y_{is,a+k}(0) - Y_{is,a-1}(0)|A_s = a] - E[Y_{is,a+k}(0) - Y_{is,a-1}(0)|A_s = b]) \\
&= \mathsf{ATT}(a, a+k) + E[\beta_{is,a+k}(b) - \beta_{is,a-1}(b)|A_s = b]
\end{aligned}
$$

- Comments:
  - The final equality shows that $\Delta_{Bad}^{a,a+k}$ is equal to $ATT(a, a+k)$ plus a bias term that is driven by time-varying treatment effects in the early treatment comparison group.
  - Depending on the sign and magnitude of the bias term, $\Delta_{Bad}^{a,a+k}$ can be biased up or down compared to $ATT(a, a+k)$. Sign flips are possible, for example. The bias occurs despite the fact that both the common trends assumption and the no anticipation assumption are valid.
  - A key point here is that in the staggered adoption design some of the within variation does not identify a causal effect.

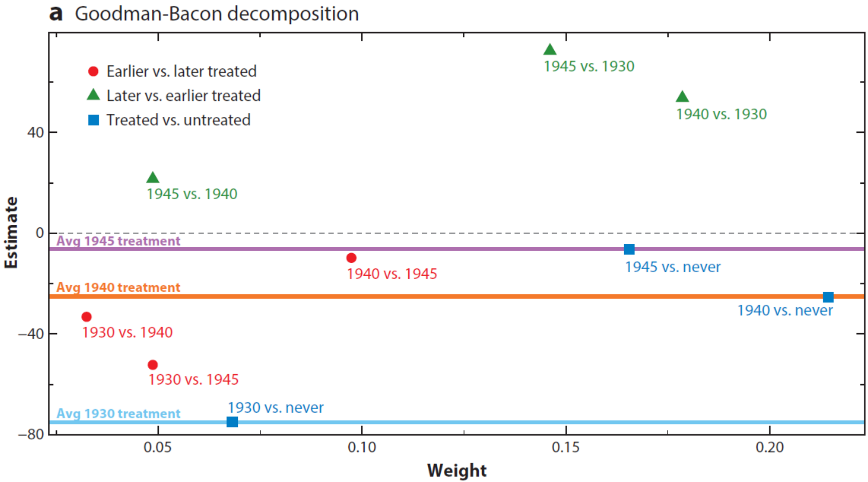

Figure 2.a: Goodman-Bacon decomposition of the TWFE estimate.
Source: Wing et al.(2024)

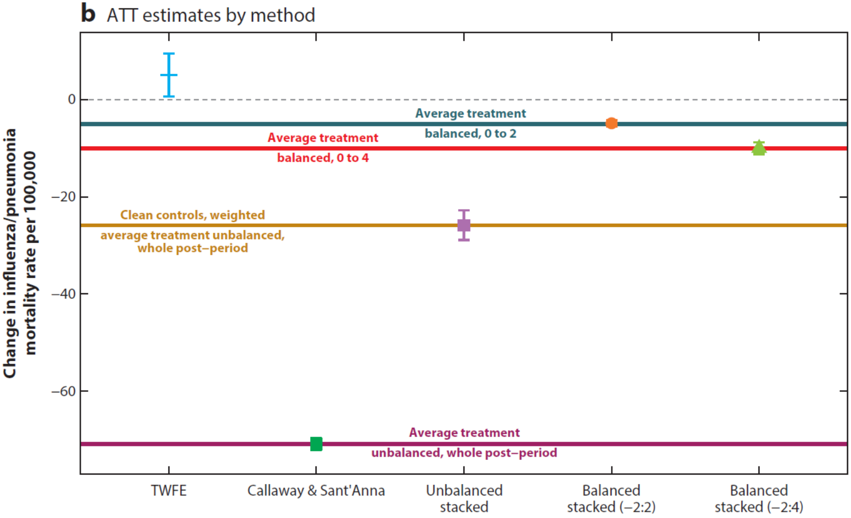

Figure 2.b: Comparison of ATT across methods.
Source: Wing et al.(2024)

Key insight of Goodman-Bacon's (2018, 2021) work

- We can avoid this bias by simply removing these potentially biased comparisons from our estimation strategy.
- This insight underlies one of the key design principles that have emerged from the new DID literature: the importance of using so-called clean controls.
- The idea is that, in a staggered adoption setting, causal inferences should be based on DIDs that compare treated timing groups to never-treated comparison units, to future-treated comparison units, or to both.

Stacked DID

Basic setting

切成小块，扔掉坏的

- In the stacked DID framework, each policy adoption is viewed as a separate subexperiment, and each subexperiment is designed to be free from confounded DID comparisons.
- For group s, the maximum length of time before treatment adoption is $A_s - T_1$, and the the maximum length of time before treatment adoption is $T_T - A_s$; the length of the feasible pre- and post- periods varies across adoption groups.

Impose a fixed event time

- Impose a fixed event time window that will be used across all subexperiments. Let k_a be the length of the pretreatment period and k_b be the length of the posttreatment period. A shorter k allows more policy events to be studied; A longer window allows researcher to study treatment effects that vary with time while for a smaller subset of adoption events.
- Let $\Omega_A = \{A_s \mid T_1 + k_a \leqslant A_s \leqslant T - k_b\}$ represent the set of policy changes that are feasible to study given a choice of k_a and k_b. Use $d \in \omega_A$ to index the admissible subexperiments.

Build a separate data set for each subexperiment $d \in \Omega_A$

- Inclusion criterion (IC) for observations included in a given subexperiment d
  - IC 1 (homogeneous treatment timing): Treatment adoption dates are homogeneous and non-staggered.
  - IC 2 (clean control): The control group consists of units that are not exposed to treatment during the event study period running from $d - k_a$ to $d + k_b$.
  - IC 3 (admissible calendar periods): All observations on treated and control units come from calendar time periods that fall inside the event window so that $d - k_a \leqslant t \leqslant d + k_b$.

- Define IC variables
  - Under IC1, define $T_{sd} = 1(A_S = d)$ indicating that an observation from group s is member of the treatment group in subexperiment d.
  - Under IC2, define $C_{sd} = 1(A_S > d + k_b)$ to indicate group s is a valid clean control for subexperiment d.
  - Under IC3, define $M_{td} = 1(d - k_a \leqslant t \leqslant d + k_b)$ to indicate that calender period t falls inside the event window for subexperiment d.
  - Put pieces together, define $I_{istd} = M_{td}(T_{sd} + C_{sd})$ as a binary inclusion variable indicating whether observation i from group s in calender period t belongs in subexperiment d.

- Applying the inclusion rule using the IC variables to the raw data repeatedly for each subexperiment yields a collection of subexperiment data sets, each centered around a specific policy change and including data only on clean controls and treated units for the approriate calendar time periods.
- The subexperimental data sets are then vertically concatenated into a single stacked analytic data set. (Note: Some units will appear as control observations in multiple subexperimental data sets.)

Stacked DID: Estimation

- Run the following regression model with the stacked data:

$$Y_{ised} = \sum_{\substack{b=-\kappa_{\text{pre}}\ldots\kappa_{\text{post}} \\ b\neq -1}} \left[\beta_e^{\text{stacked}}(D_{sd} \times 1[e=b])\right] + a_{sd} + b_{de} + \epsilon_{ised}$$

  - Y_{ised} denotes the observed outcome for unit i from state s in event time period $e = t - d$ in subexperiment d.
  - a_{sd}, a set of group $\times$ subexperiment fixed effects
  - b_{de}, a set of event time $\times$ subexperiment fixed effects.

Stacked DID: Estimation

- Comments:
  - This method uses what looks like a typical TWFE regression estimate, but because of the structure of the data, it only incorporates clean controls.
  - One way to think about this regression is as a way of estimating all of the $ATT(a, t)$ parameters and then immediately aggregating them into a single set of event time parameters, $\beta_e^{stacked}$.
  - Although free of confounding from late vs. early adoption comparisons, the aggregations produced by these models are still based on implicit variance of treatment weights, which may not be an intuitive or appealing way to summarize the underlying group-time ATTs. (– Wing, et al. 2024, NBER Working Paper)

Callaway & Sant'Anna's Aggregation

- Although there are many ways that $ATT(a,t)$ parameters could be aggregated, applied researchers are often be interested in examining dynamic treatment effects using what Callaway & Sant'Anna (2021) call a balanced event study aggregation.
- Using κ_a and κ_b to represent an event study window of interest, the balanced event study aggregate ATT at a specific event time that is q periods away from treatment adoption is the following:

$$\delta_q^{\kappa_a,\kappa_b} = \sum_a 1[T_1 + \kappa_a \leq a \leq T_T - \kappa_b] \times ATT(a, a+q) \\ \times Pr(A_s = a | T_1 + \kappa_a \leq a \leq T_T - \kappa_b)$$

- $\delta_q^{\kappa_a,\kappa_b}$ is an average of $ATT(a, a+q)$ parameters across groups with different values of $a \in A_s$.

- In practice, the idea is to estimate a family of $\delta_q^{\kappa_a,\kappa_b}$ parameters for values of $q = -\kappa_a \ldots 0 \ldots \kappa_b$ and then plot these parameters in an event study graph. The summation cycles over each of the treatment adoption dates, $a \in \Omega_A$.
- The first term trims out any adoption dates a that occurs outside.
- The third term is the weight assigned to the surviving $ATT(a, a+q)$, and the weight is the sample size of the adoption group relative to the total sample size of all admissible adoption groups.

- Stacked DID, Cengiz et al.(2019), Deshpande & Li (2019)
- Callaway & Sant'Anna (2021)
- Gardner (2022)
- Sun & Abraham (2021)
- Wooldridge (2021)
- de Chaisemartin & D'Haultfoeuille (2020)

Do you need a software package to implement?

- Stacked DID: **STACKEDEV** in Stata or reg directly
- Callaway & Sant'Anna (2021): **Did** in R and **cysdid** in Stata
- Gardner (2022): **Did2s** in R or reg directly
- Sun & Abraham (2021):**eventstudyinteract** in R or reg directly
- Wooldridge (2021): reg directly
- de Chaisemartin & D'Haultfoeuille (2020): **did_multiplegt** in Stata

What are the primary outputs?

- Stacked DID: Single aggregated treatment effect or event study estimates.
- Callaway & Sant'Anna (2021): Individual treatment effects by group and period (group-time ATTs).
- Gardner (2022): Single aggregated treatment effect or event study estimates.
- Sun & Abraham (2021):Event study coefficients (which can be interpreted as group-time ATTs in the post-period).
- Wooldridge (2021): Individual treatment effects by group and period (group-time ATTs).
- de Chaisemartin & D'Haultfoeuille (2020): Single aggregated treatment effect

What are the primary outputs?

- Stacked DID: Single aggregated treatment effect or event study estimates.
- Callaway & Sant'Anna (2021): Individual treatment effects by group and period (group-time ATTs).
- Gardner (2022): Single aggregated treatment effect or event study estimates.
- Sun & Abraham (2021):Event study coefficients (which can be interpreted as group-time ATTs in the post-period).
- Wooldridge (2021): Individual treatment effects by group and period (group-time ATTs).
- de Chaisemartin & D'Haultfoeuille (2020): Single aggregated treatment effect

How do the outputs relate to a well-known estimand, such as the ATT calculated in a 2×2 DID?

- Stacked DID: The estimate is interpreted as an average of group-time ATTs, where each is weighted by variance.
- Callaway & Sant'Anna (2021): The group-time ATTs can be aggregated to a single ATT that is analogous to the canonical DID.
- Gardner (2022): The estimate is interpreted as an average of group-time ATTs, where each is given equal weight.
- Sun & Abraham (2021):The event study coefficients in the post-period can be aggregated to a single ATT, weighted by the share of each group by period.
- Wooldridge (2021):The group-time ATTs can be aggregated to a single ATT, where each is given equal weight.
- de Chaisemartin & D'Haultfoeuille (2020): The estimate is interpreted as the average treatment effect of all switching cells.

What observations are included in the control group?

- Stacked DID: Researcher can specify their desired inclusion criteria for control observations.
- Callaway & Sant'Anna (2021): Researcher can specify never-treated or not-yet-treated units as the control observations.
- Gardner (2022): Both never-treated and not-yet-treated units are included in the control observations.
- Sun & Abraham (2021):Researcher can specify never-treated or last-to-be-treated units as the control observations.
- Wooldridge (2021):Both never-treated and not-yet-treated units are included in the control observations. Last-to-be-treated units can be used if all groups are eventually treated.
- de Chaisemartin & D'Haultfoeuille (2020): Units whose treatment status does not change between two periods are used as the control observations.

What is the role (if any) for covariates?

- Stacked DID: Covariates are not an explicit part of the method.
- Callaway & Sant'Anna (2021): Time-constant covariates can be included to allow for conditional parallel trends.
- Gardner (2022): Time-constant covariates can be included to allow for conditional parallel trends.
- Sun & Abraham (2021): Covariates are not an explicit part of the method.
- Wooldridge (2021): Time-constant covariates can be included to allow for conditional parallel trends.
- de Chaisemartin & D'Haultfoeuille (2020): Covariates require an assumption that the treatment effect is homogeneous.

Practical Advice

- Explore the Raw Data
  - Most researchers first begin to probe the key assumptions of no anticipation and common trends has been to plot the raw data over calendar time among treatment and control units.
  - Under new understandings of DID, this should still be the first step, but with the added advice to separate the data into timing groups to make a figure akin to Figure 1.

- Assess If and Why Bias Is Likely to Be an Issue in Your Setting
  - First, researchers can visually compare trends between timing groups (and an untreated group if applicable) in pretreatment years to assess no anticipation and common trends assumptions.
  - Researchers also get a sense of whether or not any potential treatment effects change over time by visualizing whether treated groups experience a different trend after treatment. (plotting the raw data in calendar time, not in an event study.)
  - It is also worth noting that even if adoption is staggered and treatment effects are time varying, the severity of confounding also depends on how much the adoptions spread over time. An emerging best practice is to also plot the adoption timeline or to include a table with adopting groups and period of adoption [as in Cook et al. (2020), table 1].

- Use an Empirical Diagnostic Method to Explore the Possible Bias
  - Goodman-Bacon decomposition
  - Sun & Abraham (2021) show how event study coefficients can be similarly decomposed into their underlying group-time ATTs and associated weights,
  - de Chaisemartin & D'Haultfoeuille (2020) propose a ratio statistic that assesses how robust the two-way fixed effects estimate is to treatment effect heterogeneity.

- Plan to Estimate Your Treatment Effect with at Least Two Techniques
    - The simple two-way fixed effects regression and one of the new estimators
    - If visualizing the raw data and other diagnostics suggest that confounding due to time-varying treatment effects is likely to be large, we recommend that an estimator robust to this confounding be considered the main estimate and that TWFE estimates be potentially included with explicit acknowledgment that they are likely confounded.
    - By contrast, if potential confounding seems likely to be small and authors feel the TWFE specification will be more intuitive to readers, then the latter could be included as the main estimate with at least one alternative specification from the more robust methods.

- Consider New Robustness Checks to Test for Pretrends
  - First, researchers can use the methods used by Roth (2022) to determine the smallest trend that can be detected in the pre-period, with a given power, for their sample.
  - Second, scholars can estimate bounds on their treatment effect in the post-period given a range of possible trends in the pre-period (Roth, 2023).

Thanks for Your Attentions!

Any questions or comments please write to:

chenglingguo@nju.edu.cn