Recent Advances in Difference-in-Differences

Wenzhi Wang [*]

August 20, 2024

1. Introduction and Learning Resources

Roth et al. (2023)

de Chaisemartin and D'Haultfœuille (2023)

2. The Canonical DiD Setup

2.1. Treatment Assignment and Timing

- There are two time periods, $t = 1, 2$.
- Units indexed by i are drawn from one of two groups.
  - Units from the treated population receive a treatment of interest in period $t = 2$, denoted by $G_i = 2$.
  - Units from the untreated (a.k.a. comparison or control) population remain untreated in both time periods, denoted by $G_i = \infty$.
  - In this notation, G refers to the treatment time.
  - The econometrician observes an outcome $Y_{i,t}$ and treatment status G_i for a panel of units, $i = 1, \ldots, N$ and $t = 1, 2$. [1]

[*]This note is written in my pre-doc period at the University of Chicago Booth School of Business.

[1]Although DiD methods also accommodate the case where only repeated cross-sectional data is available, or where the panel is unbalanced, we focus on the simpler setup with balanced panel data for ease of exposition.

2.2. Potential Outcomes and Target Parameter

Let $Y_{i,t}(g)$ denote unit i's potential outcome in period t if unit i was exposed to treatment for the first time in period g. [2]

- $Y_{i,1}(2)$ denotes unit i's (who is first treated in period $t = 2$) potential outcome in period $t = 1$.
- $Y_{i,2}(\infty)$ denotes unit i's (who is never treated in both periods) potential outcome in period $t = 2$.
- This notation implicitly encodes the stable unit treatment value assumption (SUTVA) that unit i's outcomes do not depend on the treatment status of unit $j \neq i$, which rules out spillover and general equilibrium effects.

The observed outcome is given by

$$Y_{i,t} = \sum_{g \in \mathcal{G}} \mathbb{I}(G_i = g) Y_{i,t}(g),$$

where $\mathcal{G}$ is the set containing all treatment periods, and in the canonical 2×2 setting, $\mathcal{G} = \{2, \infty\}$.

The causal estimand of primary interest in the canonical DiD setup is the average treatment effect on the treated (ATT) in period $t = 2$,

$$\text{ATT} = \underbrace{\mathbb{E}\left[Y_{i,2}(2) \mid G_i = 2\right]}_{\text{estimable from the data}} - \underbrace{\mathbb{E}\left[Y_{i,2}(\infty) \mid G_i = 2\right]}_{\textit{counterfactual component}}. \tag{2.2.1}$$

It simply measures the average causal effect on treated units in the period that they are treated ($t = 2$).

2.3. Identification Under Two Assumptions

Assumption 1 (Parallel Trends)**.**

$$\mathbb{E}\left[Y_{i,2}(\infty) - Y_{i,1}(\infty) \mid G_i = 2\right] = \mathbb{E}\left[Y_{i,2}(\infty) - Y_{i,1}(\infty) \mid G_i = \infty\right]. \tag{2.3.2}$$

Assumption 1 states that the average outcome for the treated and untreated populations would have evolved in parallel if treatment has not occurred. It can be rationalized by imposing a particular generative model for the untreated potential outcomes:

$$Y_{i,t}(\infty) = \alpha_i + \phi_t + \varepsilon_{i,t},$$

where $\varepsilon_{i,t}$ is mean-independent of G_i. [3]

[2] Compared with the standard textbook notation, this seemingly weird notation actually has more advantages. This point will be clear in the next Section when the staggered adoption designs are discussed.

[3] Note that this parametric model about untreated counterfactual outcomes allows treatment to be assigned non-randomly based on characteristics that affect the level of the outcome (α_i), but requires the treatment assignment to be mean-independent of variables that affect the *trend* in the outcome ($\varepsilon_{i,t}$). In other words, parallel trends allows for the presence of selection bias, but the bias from selecting into treatment must be the same in period $t = 1$ as it is in period $t = 2$.

Another important assumption required for identification of ATT is the no-anticipation assumption, which states that the treatment has no causal effect prior to its implementation.

Assumption 2 (No Anticipation Effects)**.**

$$Y_{i,1}(2) = Y_{i,1}(\infty) \quad \text{for all } i \text{ with } G_i = 2. \tag{2.3.3}$$

- Assumption 2 says that for these treated units, in the pre-treatment units, the potential outcomes given that they are treated or untreated are the same.
- That is, in pre-treatment units, units cannot act differently based on whether they are going to be treated in the future.
- Assumption 2 is important for identification of ATT, since otherwise the changes in the outcome for the treatment between period 1 and 2 could reflect not just the causal effect in period $t = 2$ but also the anticipatory effect in period $t = 1$.

To show why ATT is identified, first re-arrange terms in equation (2.3.2),

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

Second, by the no anticipatory assumption, $\mathbb{E}\left[Y_{i,1}(\infty) \mid G_i = 2\right] = \mathbb{E}\left[Y_{i,1}(2) \mid G_i = 2\right]$. It follows that

$$\begin{aligned}
\mathbb{E}\left[Y_{i,2}(0) \mid G_i = 2\right] &= \mathbb{E}\left[Y_{i,1}(2) \mid G_i = 2\right] + \mathbb{E}\left[Y_{i,2}(\infty) - Y_{i,1}(\infty) \mid G_i = \infty\right] \\
&= \mathbb{E}\left[Y_{i,1} \mid G_i = 2\right] + \mathbb{E}\left[Y_{i,2} - Y_{i,1} \mid G_i = \infty\right],
\end{aligned}$$

where the second equality uses the fact that we observe $Y_{i,1}(2)$ for treated units and $Y_{i,t}(\infty), t = 1, 2$ for untreated units.

Therefore, ATT can be identified as

$$\text{ATT} = \underbrace{\mathbb{E}\left[Y_{i,2} - Y_{i,1} \mid G_i = 2\right]}_{\text{Changes in the treated group}} - \underbrace{\mathbb{E}\left[Y_{i,2} - Y_{i,1} \mid G_i = \infty\right]}_{\text{Changes for in the control group}}, \tag{2.3.4}$$

i.e., the “difference-in-differences” of population means.

2.4. Estimation and Inference

Equation (2.3.4) gives an expression for ATT in terms of a “difference-in-differences” of population expectations. Therefore, a natural way to estimate ATT is to replace expectations with their sample analogs,

$$\widehat{\text{ATT}} = \left(\overline{Y}_{t=2,G=2} - \overline{Y}_{t=1,G=2}\right) - \left(\overline{Y}_{t=2,G=\infty} - \overline{Y}_{t=1,G=\infty}\right), \tag{2.4.5}$$

where $\overline{Y}_{t=t',G=g}$ is the sample mean of Y for group g in period t'.

A popular way of computing $\widehat{\mathrm{ATT}}$, which facilitates the computation of standard errors, is to use the two-way fixed effects (TWFE) regression specification

$$Y_{i,t} = \alpha_i + \phi_t + [\mathbb{I}(t=2) \cdot \mathbb{I}(G_i = 2)]\, \beta + \varepsilon_{i,t}, \tag{2.4.6}$$

In this canonical DiD setup, it is straightforward to show that the ordinary least squares (OLS) coefficient $\widehat{\beta}$ is equivalent to $\widehat{\mathrm{ATT}}$.

With a balanced panel, the OLS coefficient on β is also numerically identical to the following regression:

$$Y_{i,t} = \alpha + \mathbb{I}(G_i = 2)\theta + \mathbb{I}(t=2)\xi + [\mathbb{I}(t=2) \cdot \mathbb{I}(G_i = 2)]\, \beta + \varepsilon_{i,t}. \tag{2.4.7}$$

This regression can be generalized to repeated cross-sectional data.

Assumption 3 (Independent Sampling)**.** Let $W_i = (Y_{i,2}, Y_{i,1}, G_i)'$ denote the vector of outcomes and treatment status for unit i We observe a sample of N i.i.d. draws $W_i \sim F$ for some distribution F satisfying parallel trends.

Under Assumptions 1 - 3 and mild regularity conditions,

$$\sqrt{n}\left(\widehat{\beta} - \mathrm{ATT}\right) \to_d \mathcal{N}\left(0, \sigma^2\right)$$

in the asymptotic as $N \to \infty$ and T is fixed. [4]

3. Staggered Adoption Designs

3.1. Notations and Traditional Approaches

I will talk about *staggered adoption designs* here – cases in which being treated is an absorbing state: once a unit adopts the binary treatment, it remains exposed to the treatment for all periods afterwards.

Notations:

- There are in total T observable periods: $t \in \mathcal{T} \coloneqq \{1, 2, \ldots, T\}$.
- For each unit i, his first treatment date is denoted by $G_i \in \mathcal{G} \coloneqq \mathcal{T} \bigcup \{\infty\}$, where $G_i = \infty$ refers to a "never-treated" unit.
- $K_{it} = t - G_i$ is the relative time index to the treatment date.

[4] The variance σ^2 is consistently estimable using standard clustering methods that allow for arbitrary correlation at the unit level. For example, we can easily extend the independent sampling case (Assumption 3) to cases where the observations are individual units who are members of independently-sampled clusters (e.g. states), and the standard errors are clustered at the appropriate level, provided that the number of treated and untreated clusters both grow large.

- $D_{it} = \mathbb{I}(K_{it} \geq 0)$ is a dummy indexing if the unit i is exposed to treatment in period t.
- For $g \in \mathcal{G}$, all units with $G_i = g$ are referred to as a *cohort*, or a *timing group*, denoted by g, with a little abuse of notations.

In practice, two regressions are particularly popular, which are often referred to as *Two-Way Fixed Effects (TWFE)* regressions:

$$Y_{it} = \widetilde{\alpha}_i + \widetilde{\lambda}_t + \sum_{\substack{h=-a \\ h\neq -1}}^{b-1} \tau_h \, \mathbb{I}(K_{it} = h) + \tau_{b+} \, \mathbb{I}(K_{it} \geq b) + \tau_{a-} \, \mathbb{I}(K_{it} < a) + \varepsilon_{it} \tag{3.1.1}$$

I will further refer to this regression model as a "dynamic" TWFE regression equation. [5]

With $a = b = 0$, there is a "*static*" TWFE regression:

$$Y_{it} = \widetilde{\alpha}_i + \widetilde{\lambda}_t + \beta^{DD} D_{it} + \varepsilon_{it}. \tag{3.1.2}$$

Next, I will summarize how recent econometrics papers link the TWFE estimators to the causal estimand in the canonical 2×2 case. I will start with Goodman-Bacon (2021), which provide a set of algebraic decomposition results under minimal assumptions in Section 3.2. In Section 3.3, I will summarize results in Borusyak et al. (2024); Athey and Imbens (2022); de Chaisemartin and D'Haultfœuille (2020), which link TWFE estimators β^{DD} and $\{\tau_h\}$ to some reasonable causal estimands, under different assumptions.

Finally, I will talk about alternative estimators with better causal interpretation than the TWFE specifications, proposed by Borusyak et al. (2024); Callaway and Sant'Anna (2021); Sun and Abraham (2021).

3.2. Interpreting $\widehat{\beta}^{DD}$ I: Genearl Decomposition from Goodman-Bacon (2021)

TO BE SUMMARIZED FROM THE PAPER NOTES

3.3. Interpreting $\widehat{\beta}^{DD}$ II: Other Decomposition Results

TO BE SUMMARIZED FROM THE PAPER NOTES

[5] Here, $\widetilde{\alpha}_i, \widetilde{\lambda}_t$ are the unit and period fixed effects, $a \geq 0$ and $b \geq 0$ are the numbers of included "*leads*" and "*lags*" of the event indicator, respectively, and ε_{it} is the error term. The first lead, $\mathbb{I}(K_{it} = -1)$, is often excluded as a normalization, while the coefficients on the other leads (if present) are interpreted as measures of "pre-trends", and the hypothesis that $\tau_{a-} = \tau_{-1} = \ldots \tau_{-2} = 0$ is tested visually or statistically. Conditionally on this test passing, the coefficients on the lags are interpreted as a dynamic path of causal effects: at $h = 0, \ldots, b-1$ periods after treatment and, in the case of τ_{b+}, at longer horizons binned together.

3.4. Alternative Estimators I: Borusyak et al. (2024)

TO BE SUMMARIZED FROM THE PAPER NOTES

3.5. Alternative Estimators II: Callaway and Sant'Anna (2021)

TO BE SUMMARIZED FROM THE PAPER NOTES

3.6. Alternative Estimators III: Sun and Abraham (2021)

TO BE SUMMARIZED

References

Athey, Susan and Guido W. Imbens (2022) “Design-Based Analysis in Difference-in-Differences Settings with Staggered Adoption,” *Journal of Econometrics*, 226 (1), 62–79, 10.1016/j.jeconom.2020.10.012.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2024) “Revisiting Event-Study Designs: Robust and Efficient Estimation,” *Review of Economic Studies*, rdae007, 10.1093/restud/rdae007.

Callaway, Brantly and Pedro H.C. Sant’Anna (2021) “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, 225 (2), 200–230, 10.1016/j.jeconom.2020.12.001.

de Chaisemartin, Clément and Xavier D’Haultfœuille (2020) “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,” *American Economic Review*, 110 (9), 2964–2996, 10.1257/aer.20181169.

de Chaisemartin, Clément and Xavier D’Haultfœuille (2023) “Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey,” *The Econometrics Journal*, 26 (3), C1–C30, 10.1093/ectj/utac017.

Goodman-Bacon, Andrew (2021) “Difference-in-Differences with Variation in Treatment Timing,” *Journal of Econometrics*, 225 (2), 254–277, 10.1016/j.jeconom.2021.03.014.

Roth, Jonathan, Pedro H.C. Sant’Anna, Alyssa Bilinski, and John Poe (2023) “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature,” *Journal of Econometrics*, 235 (2), 2218–2244, 10.1016/j.jeconom.2023.03.008.

Sun, Liyang and Sarah Abraham (2021) “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects,” *Journal of Econometrics*, 225 (2), 175–199, 10.1016/j.jeconom.2020.09.006.