Federate Report

19331053 纪传宇

2023 年 8 月 4 日

目录

1 DID

- DiD methods exploit variation in time (before vs. after) and across groups (treated vs. untreated) to recover causal effects of interest. DiD combines previous approaches to avoid their pitfalls.

- Advantage: Allow for selection on time-invariant unobservables and for time-trends.

We need to assume that, absent the treatment and conditional on covariates (features), outcome of interest would grow similarly across groups/cohorts - parallel trends assumption.

What is Difference-in-differences?

- A non-random treatment is applied to one or more groups

- A group of units do not receive the units at the same time (either never, or not yet, for comparison purposes)

- Observations are taken before and after for each group

- Researcher differences before and after, then differences the difference - hence the name

Conditional Average Treatment Effects

Definition 5: Average Treatment Effect on the Treated (ATT) The average treatment effect on the treatment group is equal to the average treatment effect conditional on being a treatment group member:

$$\begin{aligned} E[\delta \mid D=1] &= E\left[Y^1-Y^0 \mid D=1\right] \\ &= E\left[Y^1 \mid D=1\right]-E\left[Y^0 \mid D=1\right] \end{aligned}$$

Definition 6: Average Treatment Effect on the Untreated (ATU) The average treatment effect on the untreated group is equal to the average treatment effect conditional on being untreated:

$$\begin{aligned} E[\delta \mid D=0] &= E\left[Y^1 - Y^0 \mid D=0\right] \\ &= E\left[Y^1 \mid D=0\right] - E\left[Y^0 \mid D=0\right] \end{aligned}$$

1.1 Fixed effect regression

- Our unobserved effects model is:

$$y_{it} = x_{it}\beta + c_i + \varepsilon_{it}; t = 1, 2, \ldots, T$$

If we have data on multiple time periods, we can think of c_i as fixed effects to be estimated - OLS estimation with fixed effects yields

$$\left(\widehat{\beta}, \widehat{c}_1, \ldots, \widehat{c}_N\right) = \underset{b, m_1, \ldots, m_N}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - x_{it}b - m_i\right)^2$$

this amounts to including N individual dummies in regression of y_{it} on x_{it}

Derivation: fixed effects regression Therefore, for $i = 1, \ldots, N$,

$$\widehat{c}_i = \frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - x_{it}\widehat{\beta}\right) = \bar{y}_i - \bar{x}_i\widehat{\beta},$$

where $\bar{x}_i \equiv \frac{1}{T} \sum_{t=1}^{T} x_{it}; \bar{y}_i \equiv \frac{1}{T} \sum_{t=1}^{T} y_{it}$ Plug this result into the first FOC to obtain:

$$\begin{gathered} \widehat{\beta} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{it} - \bar{x}_j\right)' \left(x_{it} - \bar{x}_i\right)\right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{it} - \bar{x}_i\right)' \left(y_{it} - \bar{y}\right)\right) \\ \widehat{\beta} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it}\ddot{x}_{it}\right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it}' \ddot{x}_{it}\right) \end{gathered}$$

with time-demeaned variables $\ddot{x}_{it} \equiv x_{it} - \bar{x}, \ddot{y}_{it} \equiv y_{it} - \bar{y}_i$

1.2 potential outcomes

Parallel trends bias

$$\begin{aligned} \widehat{\delta}_{kU}^{2\times 2} = & \underbrace{E\left[Y_k^1 \mid \text{ Post }\right] - E\left[Y_k^0 \mid \text{ Post }\right]}_{\text{ATT}} \\ & + \underbrace{\left[E\left[Y_k^0 \mid \text{ Post }\right] - E\left[Y_k^0 \mid \text{ Pre }\right]\right] - \left[E\left[Y_U^0 \mid \text{ Post }\right] - E\left[Y_U^0 \mid \text{ Pre }\right]\right]}_{\text{Non-parallel trends bias in } 2\times 2 \text{ case}} \end{aligned}$$

1.3 TWFE

The most conceptually simple estimator replaces population means with sample analogs:

$$\widehat{\tau}_{\text{DiD}} = \left(\bar{Y}_{12} - \bar{Y}_{11}\right) - \left(\bar{Y}_{02} - \bar{Y}_{01}\right)$$

where $\bar{Y}_{dt}$ is sample mean for group d in period t Conveniently, $\widehat{\tau}_{DID}$ is algebraically equal to OLS coefficient $\hat{\beta}$ from the TWFE regression

$$Y_{i,t} = \alpha_i + \phi_t + D_{i,t}\beta + \epsilon_{i,t},$$

where $D_{i,t} = D_i * 1[t = 2]$.

Bias in our go-to estimators OLS will identify the ATT with only two groups and no covariates

But the more common DD situation is one in which a treatment is adopted by different groups at different times OLS with panel and time fixed effects (”twoway fixed effects” or TWFE) is biased we now know I’ll discuss the bias of TWFE, discuss a new solution, and fingers crossed a simulation if we have time

OLS with twoway fixed effects Under parallel trends, OLS estimates the ATT for the two group case. Calculating standard errors is easy, multiple time periods is easy. But including covariates and time varying treatment (”differential timing”) will introduce problems.

1.4 Prepensity Score

Overview - Abadie (2005) proposed an alternative estimator to OLS to estimate the ATT - The method is a DD type estimator, but isn’t using TWFE - You need treatment and comparison group, before and after treatment - But you also need conditional parallel trends (based on X)

Why do this? - No randomization. Remember, DD doesn’t require randomization - it requires a version of parallel trends - Treatment is selecting on observable covariates

Why do this anyway? In a DD, we may need to control for X because treatment is only conditional on X But in TWFE, when you controlling for baseline X , it gets absorbed by the unit fixed effects

One way around this is to use time-varying controls, but this places restrictions on the DGP as we will see in Sant’Anna and Zhou (2020) Abadie (2005) proposes weighting the difference in means using the propensity score estimated with series logit or linear probability models

Terms

t is year of treatment which doesn’t vary across units (so no differential timing)

Y^1 and Y^0 are potential outcomes (counterfactual versus actual)

D is 1 or 0 based on group and time

b is the ”baseline” which is similar to CS using g as the one year pre-treatment

X are ”baseline” covariates only - they do not vary over time, which means propensity scores are estimated off the b period only

1.5 Covariate

Assumption CT (Common or Parallel Trends): With D the treatment indicator,

$$E[\Delta Y(0) \mid D = 1] = E[\Delta Y(0) \mid D = 0].$$

- No selection bias in the trend in the control state.
- Importantly, D can be correlated with $Y_1(0)$.
- D can be correlated with $\Delta Y(1) = Y_2(1) - Y_1(1)$.
- Conclusion: Under Assumptions NA and CT,

$$\operatorname{plim}\left(\hat{\tau}_{2,DD}\right) = \operatorname{plim}\left(\overline{\Delta Y}_{\text{treal }} - \overline{\Delta Y}_{\text{control }}\right) = \tau_{2,\text{ att}}$$

Two-Way Fixed Effects

- Recall the TWFE dummy variable regression where $ch_i = 1[h = i]$:

$$Y_{it} \text{ on } W_{it}, c1_i, c2_i, \ldots, cN_i, f2_t, t = 1, 2; i = 1, \ldots, N$$

- Equivalence of POLS and TWFE: It is enough to control for D_i rather than N unit dummies.

Adding Covariates

- Common to include controls (covariates) in DiD settings.
- Can help with failure of the CT Assumption.
- Current thinking: Only use pre-treatment controls.
- Can ruin the CT assumption by conditioning on controls affected by intervention.
- Let $\mathbf{X}_i$ be a $1 \times K$ vector of pre-intervention covariates.