An Introduction to Difference-in-Differences and Event Studies

Francesco Ruggieri

April 4, 2022

Definitions

- **Time periods** indexed by $t \in \{1, \ldots, \bar{t}\}$
- $D_t \in \{0, 1\}$ indicates **treatment assignment** at the *beginning* of period t
- The treatment is **absorbing**, i.e., $D_t = 1 \implies D_\tau = 1$ for all $\tau \in \{t + 1, \ldots, \bar{t}\}$
- $Y_t \in \mathbb{R}$ denotes an **outcome** observed at the *end* of period t

Definitions

1. **Difference-in-Differences (DiD)**
   - There exists **one and only one** time period t^* at which one can receive the treatment
   - If a unit is untreated at $t = t^*$, it will never be treated
   - Example: policies that are implemented all at once
2. **Event Study (ES)**
   - **Staggered** assignment of the treatment
   - **Cohorts** are implied by the **timing** of treatment assignment (including never- and always-treated)
   - Example: policies that are implemented at different times for different groups

The Canonical Two-Period Difference-in-Differences Design

- Two **time periods** indexed by $t \in \{1, 2\}$
- The treatment is assigned in $t = 2$, i.e., $\mathbb{P}\left(D_1 = 0\right) = 1$ and $0 < \mathbb{P}\left(D_2 = 1\right) < 1$

The Canonical Two-Period Difference-in-Differences Design

- **Potential treatments** $D_1 = 0$ (degenerate) and $D_2(0)$ (nondegenerate)
- It is common to define a **control group** $(G = 0)$ and a **treatment group** $(G = 1)$

$$G = 0 \iff (D_1, D_2(0)) = (0, 0)$$
$$G = 1 \iff (D_1, D_2(0)) = (0, 1)$$

- Thus, the treatment can be defined as $D_t \equiv G \times \mathbb{I}[t = 2]$
- **Potential outcomes** $Y_t(0, D_2(0))$ for $t \in \{1, 2\}$
  - $Y_1(0, 0)$, $Y_1(0, 1)$, $Y_2(0, 0)$, $Y_2(0, 1)$ depend on the **full path** of treatment states

The Canonical Two-Period Difference-in-Differences Design

- **Target parameter**: $\mathrm{ATT}_2 \equiv \mathbb{E}\left[Y_2\left(0,1\right) - Y_2\left(0,0\right) | D_1 = 0, D_2(0) = 1\right]$

- The first conditional mean is observed:

$$\mathbb{E}\left[Y_2\left(0,1\right) | D_1 = 0, D_2(0) = 1\right] = \mathbb{E}\left[Y_2 | D_1 = 0, D_2 = 1\right]$$

- To identify the second conditional mean, assume **common trends**:

$$\mathbb{E}\left[Y_2\left(0,0\right) - Y_1\left(0,0\right) | D_1 = 0, D_2(0) = 0\right] = \mathbb{E}\left[Y_2\left(0,0\right) - Y_1\left(0,0\right) | D_1 = 0, D_2(0) = 1\right]$$

- Equivalently,

$$\mathbb{E}\left[Y_2\left(0,0\right) - Y_1\left(0,0\right) | G = 0\right] = \mathbb{E}\left[Y_2\left(0,0\right) - Y_1\left(0,0\right) | G = 1\right]$$

The Canonical Two-Period Difference-in-Differences Design

- The left-hand side is observed, so

$$\mathbb{E}\left[Y_2(0,0)\,|D_1=0, D_2(0)=1\right] = \mathbb{E}\left[Y_2 - Y_1|D_1=0, D_2=0\right] + \mathbb{E}\left[Y_1(0,0)\,|D_1=0, D_2(0)=1\right]$$

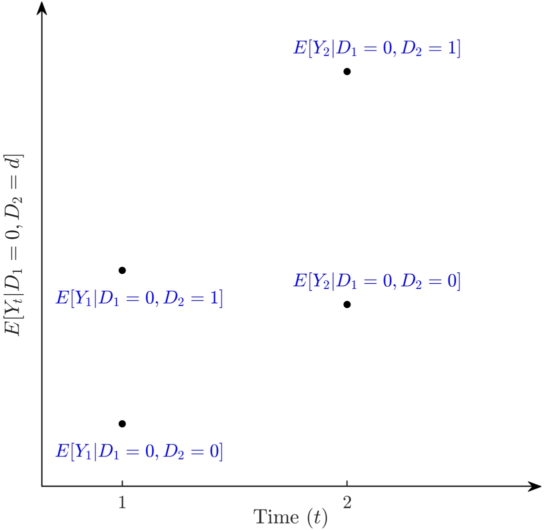
- To identify $\mathbb{E}\left[Y_1(0,0)\,|D_1=0, D_2(0)=1\right]$, assume **no anticipation**:

$$\mathbb{E}\left[Y_1(0,0)|D_1=0, D_2(0)=1\right] = \mathbb{E}\left[Y_1(0,1)|D_1=0, D_2(0)=1\right]$$

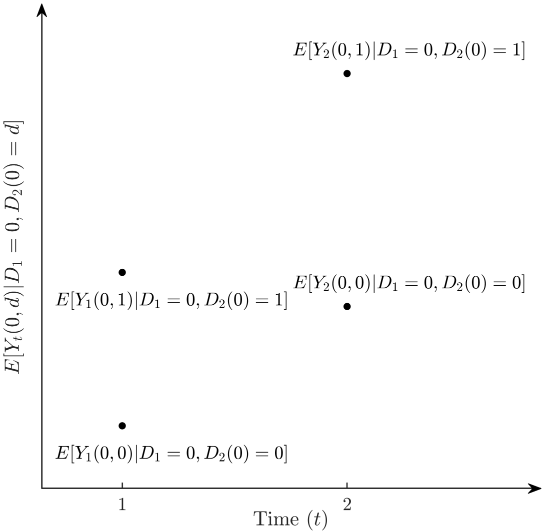
- The right-hand side is observed, so the target parameter is identified by the **DiD estimand**

$$\mathrm{ATT}_2 = \mathbb{E}\left[Y_2 - Y_1|D_1=0, D_2=1\right] - \mathbb{E}\left[Y_2 - Y_1|D_1=0, D_2=0\right]$$

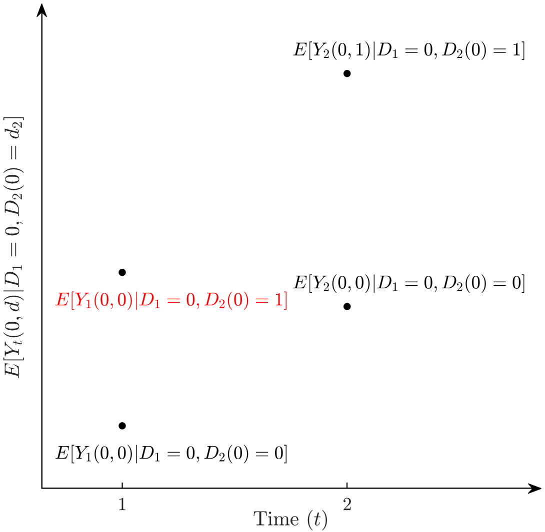
Observed Conditional Means



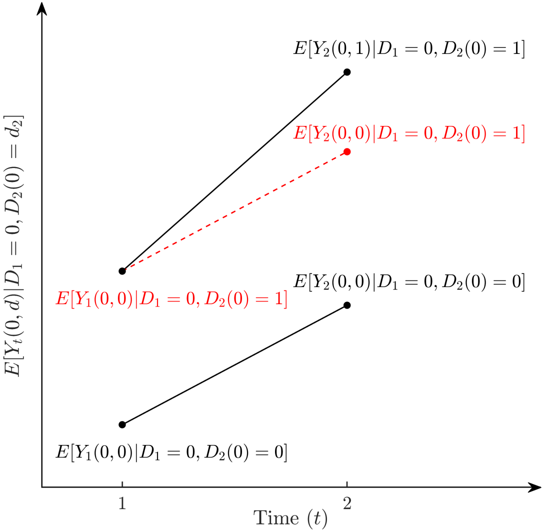
Identified Conditional Means



Imposing No Anticipation



Imposing Common Trends



Implementation with Linear Regression

- The difference-in-differences estimand

$$\text{DiD} = \mathbb{E}\left[Y_2 - Y_1 | D_1 = 0, D_2 = 1\right] - \mathbb{E}\left[Y_2 - Y_1 | D_1 = 0, D_2 = 0\right]$$

- Linear combination of **four conditional means**
- Could be estimated nonparametrically (binning), but linear regression is more convenient
- A **saturated** regression that exactly replicates realizations of $\mathbb{E}\left[Y_t | D_1 = 0, D_2\right]$ for $t \in \{1, 2\}$
  - Four bins and four regressors $\rightarrow$ no need to approximate the conditional means of Y

Implementation with Linear Regression

- One possible specification is

$$\begin{aligned}\mathbb{E}\left[Y|D_1, D_2, T\right] = {} & \alpha_1 \times \mathbb{I}\left[D_1 = 0, D_2 = 0, T = 1\right] + \alpha_2 \times \mathbb{I}\left[D_1 = 0, D_2 = 0, T = 2\right] \\ & + \alpha_3 \times \mathbb{I}\left[D_1 = 0, D_2 = 1, T = 1\right] + \alpha_4 \times \mathbb{I}\left[D_1 = 0, D_2 = 1, T = 2\right]\end{aligned}$$

The target parameter (ATT_2) is identified by $\boldsymbol{(\alpha_4 - \alpha_3) - (\alpha_2 - \alpha_1)}$

- For a more convenient interpretation,

$$\begin{aligned}\mathbb{E}\left[Y|D_1, D_2, T\right] = {} & \beta_1 \times 1 + \beta_2 \times \underbrace{\mathbb{I}\left[D_1 = 0, D_2 = 1\right]}_{\text{treated group}} \\ & + \beta_3 \times \underbrace{\mathbb{I}\left[T = 2\right]}_{\text{post period}} + \beta_4 \times \underbrace{\mathbb{I}\left[D_1 = 0, D_2 = 1, T = 2\right]}_{\text{treated group \& post period}}\end{aligned}$$

The target parameter (ATT_2) is identified by $\boldsymbol{\beta_4} = (\alpha_4 - \alpha_3) - (\alpha_2 - \alpha_1)$

Extension to Covariates

- The **common trends** assumption may be more plausible within bins implied by **covariates**
  - Ideally **predetermined** because time-varying covariates may be caused by the treatment

- Let $X \in \mathbb{R}^{d_x}$ be a vector of predetermined (time-invariant) covariates

- **Conditional common trends**. With probability one,

$$\mathbb{E}\left[Y_2(0,0) - Y_1(0,0) \,|\, D_1 = 0, D_2(0) = 0, X\right] = \mathbb{E}\left[Y_2(0,0) - Y_1(0,0) \,|\, D_1 = 0, D_2(0) = 1, X\right]$$

- Assume an **overlap condition**

$$0 < \mathbb{P}\left(D_1 = 0, D_2(0) = 1 | X = x\right) < 1 \quad \text{for all } x \in \operatorname{supp}(X)$$

Intuition: for each possible realization of X, both control and treatment groups are "populated"

Extension to Covariates

- The **conditional** target parameter $\mathrm{ATT}_2(x)$ is identified by

$$\mathrm{ATT}_2(x) = \mathbb{E}[Y_2 - Y_1 | D_1 = 0, D_2 = 1, X = x] - \mathbb{E}[Y_2 - Y_1 | D_1 = 0, D_2 = 0, X = x]$$

- By the Law of Iterated Expectations, the **unconditional** target parameter is

$$\mathrm{ATT}_2 = \mathbb{E}[\mathrm{ATT}_2(X) | D_1 = 0, D_2 = 1]$$

$$= \underbrace{\mathbb{E}[Y_2 - Y_1 | D_1 = 0, D_2 = 1]}_{\text{easy}} - \underbrace{\mathbb{E}[\mathbb{E}[Y_2 - Y_1 | D_1 = 0, D_2 = 0, X] | D_1 = 0, D_2 = 1]}_{\text{not so easy}}$$

- In **finite samples**, it may not be easy to compute the second term if X has **large support**
  - Curse of dimensionality, the estimator will likely have **high variance**

Extension to Covariates

Some possible solutions:

1. The good old **linear regression**. A commonly adopted specification is

$$\mathbb{E}\left[Y|D_1, D_2, T, X\right] \approx \gamma_1 \times 1 + \gamma_2 \times \mathbb{I}\left[D_1 = 0, D_2 = 1\right] + \gamma_3 \times \mathbb{I}\left[T = 2\right] + \gamma_4 \times \mathbb{I}\left[D_1 = 0, D_2 = 1, T = 2\right] + X'\delta$$

This linear regression is **no longer saturated** (optimal MSE is positive, not zero)!

- If treatment effect $Y_2(0,1) - Y_2(0,0)$ is a **deterministic constant**, no problem
- However, $Y_2(0,1) - Y_2(0,0)$ is very likely to be a **nondegenerate random variable**
- Coefficients in unsaturated regressions are often **hard to interpret** in this case...
- ...even in extremely simple specifications. Consider, for instance, Słoczyński (2020)

Extension to Covariates

2. **Matching** on X (if discrete and with small support)

3. **Matching** on the **propensity score** $p(X) \equiv \mathbb{P}(D_1 = 0, D_2 = 1 | X)$
   - The propensity score is often estimated with a logistic regression

4. **Propensity score weighting**. Given $G = 1 \iff D_1 = 0, D_2 = 1$, the target parameter is

$$\mathrm{ATT}_2 = \frac{1}{\mathbb{P}(G = 1)} \mathbb{E}\left[\left(G - \frac{(1 - G)\, p(X)}{1 - p(X)}\right)(Y_2 - Y_1)\right]$$

In practice, replace population moments with their sample counterparts (plug-in method)

Extension to Multiple Time Periods

- The (absorbing) treatment is assigned in period $t^* \in \{1, \ldots, \bar{t}\}$
- Thus, the treatment can be defined as $D_t \equiv G \times \mathbb{I}\left[t \geq t^*\right]$
- Given multiple time periods, easier to index potential treatments and outcomes by $G \in \{0, 1\}$
- For any t, **potential treatments** $D_t\left(G\right)$ and **potential outcomes** $Y_t\left(G\right)$
- Because this a sharp design, $D_t\left(G\right)$ is a deterministic function of G
  - E.g. for any $t \geq t^*$, $D_t\left(0\right) = 0$ and $D_t\left(1\right) = 1$ with probability one

Extension to Multiple Time Periods

- **Multiple target parameters**: for any $t \geq t^*$,

$$\mathrm{ATT}_t \equiv \mathbb{E}\left[Y_t(1) - Y_t(0) \,|\, G = 1\right]$$

where $Y_t(g)$ indicates the period-t potential outcome in group $G = g$

- A **generalized common trends** assumption. For any $s < t^*$ and $t \geq t^*$,

$$\mathbb{E}\left[Y_t(0) - Y_s(0) \,|\, G = 0\right] = \mathbb{E}\left[Y_t(0) - Y_s(0) \,|\, G = 1\right]$$

- A **generalized no anticipation** assumption
  - In words: on average, today's potential outcome is not affected by future treatment states

Extension to Multiple Time Periods

- The identification argument is analogous to the two-period case. For any $s < t^*$ and $t \geq t^*$,

$$\mathrm{ATT}_t = \mathbb{E}\left[Y_t - Y_s | G = 1\right] - \mathbb{E}\left[Y_t - Y_s | G = 0\right]$$

- Implementation with **linear regression** is more convenient in this case

- By the common trends assumption, for $g \in \{0, 1\}$ and any $t \in \{1, \ldots, \bar{t}\}$,

$$\mathbb{E}\left[Y(0) | G = g, T = t\right] = \mathbb{E}\left[Y(0) | G = g, T = 1\right] + \underbrace{\mathbb{E}\left[Y(0) | T = t\right] - \mathbb{E}\left[Y(0) | T = 1\right]}_{\text{pure time indicators}}$$

- In addition, $\mathbb{E}\left[Y(0) | G, T = 1\right]$ has **two** possible realizations, so it can be expressed as

$$\mathbb{E}\left[Y(0) | G, T = 1\right] = \alpha + \beta G$$

Extension to Multiple Time Periods

- A hypothetical (intermediate) regression

$$\mathbb{E}\left[Y(0)|G,T\right] \approx \alpha + \beta G + \sum_{s=2}^{\bar{t}} \gamma_s \mathbb{I}\left[T = s\right]$$

$\mathbb{E}\left[Y(0)|G,T\right]$ has $2\bar{t}$ possible realizations, some regressors are missing to saturate it

- The treatment is a **deterministic function** of G and T. By the switching equation,

$$\begin{aligned}\mathbb{E}\left[Y|G,T\right] &= \mathbb{E}\left[Y(0) + D\left(Y(1) - Y(0)\right)|G,T\right] \\ &= \underbrace{\mathbb{E}\left[Y(0)|G,T\right]}_{\text{above}} + \underbrace{\mathbb{E}\left[D\left(Y(1) - Y(0)\right)|G,T\right]}_{\mathbb{P}(D=1|G=1,T=t)=1 \text{ for } t \geq t^*} \\ &= \mathbb{E}\left[Y(0)|G,T\right] + \sum_{s=t^*}^{\bar{t}} \underbrace{\mathbb{E}\left[Y(1) - Y(0)|G=1,T=s\right]}_{\equiv \mathrm{ATT}_s} \mathbb{I}\left[G=1,T=s\right]\end{aligned}$$

Extension to Multiple Time Periods

- A **Two-Way Fixed Effects (TWFE)** regression

$$\mathbb{E}[Y|G,T] \approx \alpha + \beta G + \sum_{s=2}^{\bar{t}} \gamma_s \mathbb{I}[T=s] + \sum_{s=t^*}^{\bar{t}} \delta_s \mathbb{I}[G=1, T=s]$$

  This specification is **not saturated** because **common trends** has been **assumed** to be **true**

- To determine if common trends is plausible, **saturate** it:

$$\mathbb{E}[Y|G,T] = \alpha + \beta G + \sum_{s=2}^{\bar{t}} \gamma_s \mathbb{I}[T=s] + \sum_{s=2}^{t^*-1} \eta_s \mathbb{I}[G=1, T=s] + \sum_{s=t^*}^{\bar{t}} \delta_s \mathbb{I}[G=1, T=s]$$

  Now $2\bar{t}$ realizations of $\mathbb{E}[Y|G,T]$ **and** $2\bar{t}$ regressors

- δ_s identifies ATT_s, while η_s will be equal to zero if common trends holds

Extension to Multiple Time Periods

- This is often referred to as **dynamic TWFE specification**
- An alternative is to consider the (unsaturated) **static TWFE specification**

$$\mathbb{E}\left[Y|G,T\right] = \alpha + \beta G + \sum_{s=2}^{\bar{t}} \gamma_s \mathbb{I}\left[T=s\right] + \sum_{s=2}^{t^*-1} \eta_s \mathbb{I}\left[G=1, T=s\right] + \delta \mathbb{I}\left[G=1, T \geq t^*\right]$$

- δ identifies $\frac{1}{\bar{t}-t^*}\sum_{s=t^*}^{\bar{t}} \mathrm{ATT}_s$, a **simple average** of ATTs in the post-period
  - Intuitively, fewer parameters to be estimated, so likely lower variance

Extension to Multiple Time Periods

- If the dynamic TWFE specification is **not** saturated, $\{\eta_s\}_{s=2}^{t^*-1}$ reflect **leads and lags**
  - Linear regression approximates – does not exactly replicate – the conditional outcome mean
- It may be inappropriate to test $\{\eta_s\}_{s=2}^{t^*-1}$ to assess the plausibility of common trends
- This is the case even if **common trends** is **in fact true**
- The issue disappears if the dynamic TWFE specification is **saturated**...
  - ...or if average effects are homogeneous over time ($\mathrm{ATT}_t = \mathrm{ATT}$ for all t)

Event Studies

- **Staggered** assignment of the treatment
- **Cohorts** are implied by the **timing** of treatment assignment (including never- and always-treated)
- **TWFE specifications** are **extremely problematic**...
- ...see you next Monday with Goodman-Bacon (2021)!