Lecture 16: Difference-in-differences

Economics 326 — Introduction to Econometrics II

Vadim Marmer, UBC

April 5, 2026

Motivation

- Previous lecture: treatment effects from **cross-sectional** data, assuming selection on observables (treatment is as good as random after controlling for covariates).
- Often hard to justify. Alternative: exploit data over time:
  - **Panel data**: repeated observations on the exact same units over time.
  - **Repeated cross-sections**: observations on different units from the same populations at different points in time.
- **Difference-in-differences (DID)**: compare changes over time between a treatment group and a control group.

DID basic setup

- Two time periods: $t \in \{0, 1\}$ (before and after treatment).
- Two groups: $D_i \in \{0, 1\}$ (control and treatment).
- Treatment occurs between periods 0 and 1; only the treatment group ($D_i = 1$) is affected.
- Y_{it}: outcome for individual i at time t.

DID regression model

- DID regression:

$$Y_{it} = \alpha + \delta \cdot t + \gamma D_i + \beta (t \cdot D_i) + U_{it},$$

  where $\mathrm{E}[U_{it} \mid D_i] = 0$.
- Regressors:
  - t: time (0 = before, 1 = after).
  - D_i: group (0 = control, 1 = treatment).
  - $t \cdot D_i$: interaction, equals 1 only for the treatment group after treatment.

Interpreting the coefficients

- $$Y_{it} = \alpha + \delta \cdot t + \gamma D_i + \beta (t \cdot D_i) + U_{it},$$
- Evaluate $\mathrm{E}[Y_{it} \mid D_i]$ for each combination of t and D_i:

$$
\begin{aligned}
t=0,\ D_i=0: &\quad \mathrm{E}[Y_{i0} \mid D_i=0]=\alpha, \\
t=0,\ D_i=1: &\quad \mathrm{E}[Y_{i0} \mid D_i=1]=\alpha+\gamma, \\
t=1,\ D_i=0: &\quad \mathrm{E}[Y_{i1} \mid D_i=0]=\alpha+\delta, \\
t=1,\ D_i=1: &\quad \mathrm{E}[Y_{i1} \mid D_i=1]=\alpha+\delta+\gamma+\beta.
\end{aligned}
$$

- Summarized as a 2×2 table:

	$D_i = 0$ (Control)	$D_i = 1$ (Treatment)
$t = 0$	α	$\alpha + \gamma$
$t = 1$	$\alpha + \delta$	$\alpha + \delta + \gamma + \beta$

- α: baseline (control group, $t = 0$).
- γ: pre-existing **group difference** at $t = 0$.
- δ: **time effect** — change in the control group from $t = 0$ to $t = 1$ (common trend).
- Change over time for each group:
  - **Treatment:** $\mathrm{E}[Y_{i1} \mid D_i = 1] - \mathrm{E}[Y_{i0} \mid D_i = 1] = \delta + \beta$.
  - **Control:** $\mathrm{E}[Y_{i1} \mid D_i = 0] - \mathrm{E}[Y_{i0} \mid D_i = 0] = \delta$.
- Subtract control's change from treatment's change:

$$
(\delta + \beta) - \delta = \beta.
$$

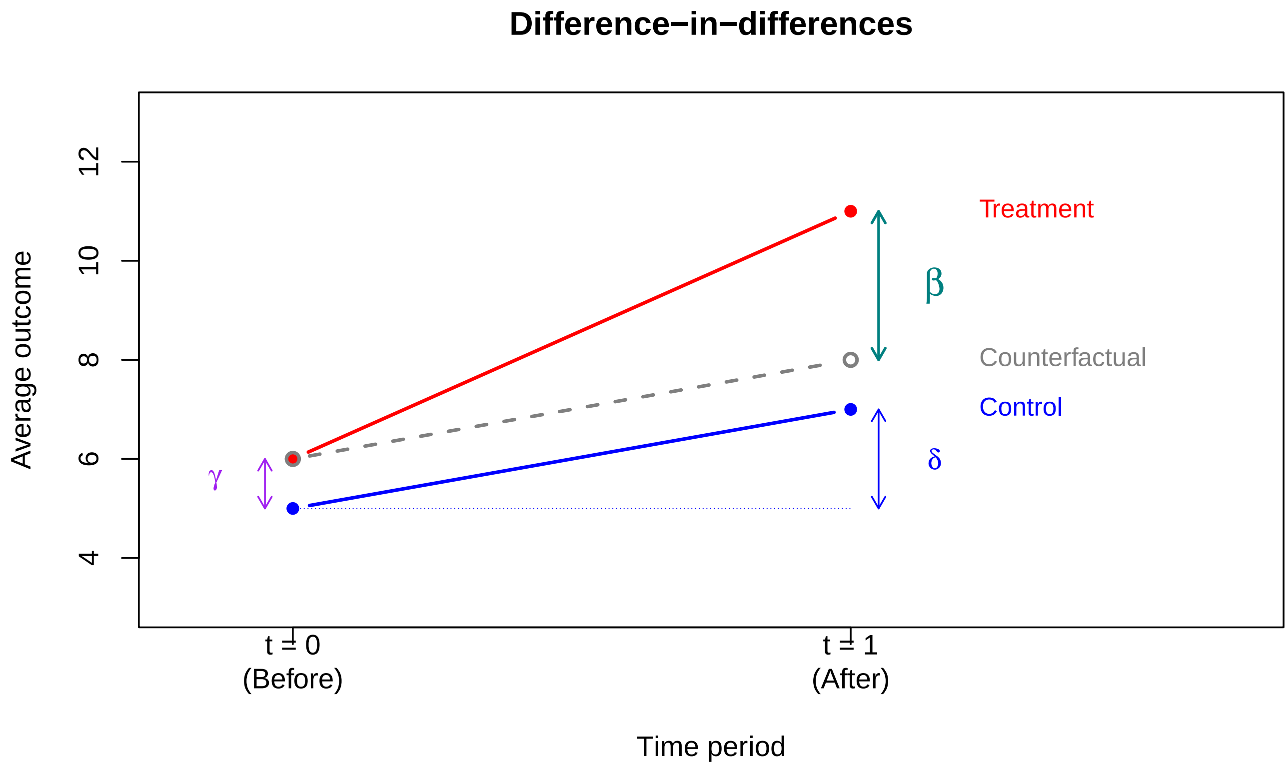
- DID estimand as a double difference:

$$
\beta = \mathrm{E}[Y_{i1} - Y_{i0} \mid D_i = 1] - \mathrm{E}[Y_{i1} - Y_{i0} \mid D_i = 0].
$$

- The common trend δ cancels, isolating β.

DID diagram

- DID regression: $Y_{it} = \alpha + \delta \cdot t + \gamma D_i + \beta(t \cdot D_i) + U_{it}$.
- DID diagram — control, treatment, and counterfactual:



- Counterfactual (dashed gray): treatment group's baseline $\alpha + \gamma$ plus the control group's change δ.

DID and potential outcomes

- **Panel potential outcomes**: $Y_{it}(d)$ — outcome for individual i at time t if assigned to group $d \in \{0, 1\}$.
- Four potential outcomes: $Y_{i0}(0)$, $Y_{i1}(0)$, $Y_{i0}(1)$, $Y_{i1}(1)$.
- The observed outcome is:

$$Y_{it} = D_i\, Y_{it}(1) + (1 - D_i)\, Y_{it}(0).$$

- What we observe for each group:

	Control ($D_i = 0$)	Treatment ($D_i = 1$)
$t = 0$	$Y_{i0}(0)$	$Y_{i0}(1)$
$t = 1$	$Y_{i1}(0)$	$Y_{i1}(1)$

- Treatment effect on the treated at $t = 1$:

$$\text{ATT} = \text{E}\left[Y_{i1}(1) - Y_{i1}(0) \mid D_i = 1\right].$$

  The counterfactual $Y_{i1}(0)$ is unobserved for the treated group.

DID as a treatment effect

- $\beta = \text{E}\left[Y_{i1} - Y_{i0} \mid D_i = 1\right] - \text{E}\left[Y_{i1} - Y_{i0} \mid D_i = 0\right].$

- Substituting observed outcomes with potential outcomes:

$$\beta = \mathrm{E}\left[Y_{i1}(1) - Y_{i0}(1) \mid D_i = 1\right] \\ - \mathrm{E}\left[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0\right].$$

- To relate β to the ATT, add and subtract $Y_{i1}(0)$ and $Y_{i0}(0)$ inside the first expectation:

$$\begin{aligned}\beta &= \mathrm{E}\left[Y_{i1}(1) - Y_{i0}(1) \mid D_i = 1\right] - \mathrm{E}\left[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0\right] \\ &= \mathrm{E}\left[Y_{i1}(1) \underbrace{- Y_{i1}(0) + Y_{i1}(0)}_{=0} \underbrace{- Y_{i0}(0) + Y_{i0}(0)}_{=0} - Y_{i0}(1) \mid D_i = 1\right] \\ &\quad - \mathrm{E}\left[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0\right].\end{aligned}$$

- Rearranging and splitting the expectation:

$$\begin{aligned}\beta = &\underbrace{\mathrm{E}\left[Y_{i1}(1) - Y_{i1}(0) \mid D_i = 1\right]}_{\text{ATT}} \\ &+ \underbrace{\mathrm{E}\left[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 1\right] - \mathrm{E}\left[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0\right]}_{\text{difference in trends}} \\ &+ \underbrace{\mathrm{E}\left[Y_{i0}(0) - Y_{i0}(1) \mid D_i = 1\right]}_{\text{anticipation effect}}.\end{aligned}$$

- For β to equal the ATT, the difference in trends and anticipation effect must be zero. This requires two assumptions.

Assumption 1: Parallel trends

- Difference in trends is zero if both groups would have experienced the same change over time absent treatment:

$$\mathrm{E}\left[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 1\right] = \mathrm{E}\left[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0\right].$$

- Under parallel trends, the decomposition reduces to:

$$\beta = \text{ATT} + \underbrace{\mathrm{E}\left[Y_{i0}(0) - Y_{i0}(1) \mid D_i = 1\right]}_{\text{anticipation effect}}.$$

- Cannot be directly tested: $Y_{i1}(0)$ is unobserved for the treated group.
- With multiple pre-treatment periods, can check whether trends were parallel before treatment.

Assumption 2: No anticipation

- Anticipation effect is zero if pre-treatment outcomes are not affected by future treatment assignment:

$$\mathrm{E}\left[Y_{i0}(1) \mid D_i = 1\right] = \mathrm{E}\left[Y_{i0}(0) \mid D_i = 1\right].$$

- Treatment assignment does not change pre-treatment outcomes in expectation.
- Under both parallel trends and no anticipation:

$$\beta = \text{ATT}.$$

Example: incinerator and house prices

- Kiel and McClain (1995): did a garbage incinerator in North Andover, MA reduce nearby house prices?
- This is a **repeated cross-section**: the houses sold in 1981 (after) are different from the houses sold in 1978 (before).
- Data: `kielmc` from the `wooldridge` package:

```
library(wooldridge)
data(kielmc)
# Show 2 observations from each of the 4 groups
rows <- c(
  head(which(kielmc$y81 == 0 & kielmc$nearinc == 0), 2),
  head(which(kielmc$y81 == 0 & kielmc$nearinc == 1), 2),
  head(which(kielmc$y81 == 1 & kielmc$nearinc == 0), 2),
  head(which(kielmc$y81 == 1 & kielmc$nearinc == 1), 2)
)
kielmc[rows, c("rprice", "y81", "nearinc", "age")]
```

```
      rprice y81 nearinc age
14  52000.00   0       0  32
15  49000.00   0       0  18
1   60000.00   0       1  48
2   40000.00   0       1  83
187 90245.77   1       0   1
188 46082.95   1       0  41
180 37634.41   1       1  81
181 39938.55   1       1  71
```

- `rprice`: house price in 1978 dollars (Y_{it}).
- `y81`: 1 if year is 1981 (after incinerator announced), 0 if 1978 ($t = 1$ if 1981, 0 if 1978).
- `nearinc`: 1 if house is near the incinerator site (D_i).

The 2×2 table of means

- Compute the four group means:

```
means <- tapply(kielmc$rprice, list(kielmc$y81, kielmc$nearinc), mean)
colnames(means) <- c("Far (nearinc=0)", "Near (nearinc=1)")
rownames(means) <- c("1978 (y81=0)", "1981 (y81=1)")
round(means, 2)
```

```
             Far (nearinc=0) Near (nearinc=1)
1978 (y81=0)        82517.23         63692.86
1981 (y81=1)       101307.51         70619.24
```

- Computing the DID by hand:

```
diff_near <- means[2, 2] - means[1, 2]
diff_far  <- means[2, 1] - means[1, 1]
DID <- diff_near - diff_far
cat("Change (near):", round(diff_near, 2), "\n")
```

```
Change (near): 6926.38
```

```
cat("Change (far): ", round(diff_far, 2), "\n")
```

```
Change (far):  18790.29
```

```
cat("DID:          ", round(DID, 2), "\n")
```

```
DID:           -11863.9
```

DID regression

- The DID regression, where `y81nrinc` = y81 × nearinc is the interaction term:

```
options(scipen = 999)
reg_did <- lm(rprice ~ y81 + nearinc + y81nrinc, data = kielmc)
round(summary(reg_did)$coefficients, 4)
```

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  82517.23   2726.910 30.2603   0.0000
y81          18790.29   4050.065  4.6395   0.0000
nearinc     -18824.37   4875.322 -3.8612   0.0001
y81nrinc    -11863.90   7456.646 -1.5911   0.1126
```

- Coefficient on `y81nrinc` matches the DID from the 2×2 table.
- $\hat{\beta} = -\$11{,}864$ (SE $= 7{,}457$, $p = 0.113$): negative but not significant at 5%.

Assumptions in the incinerator example

- **Parallel trends:** without the incinerator, house prices near and far from the site would have followed the same trend over time.
- **No anticipation:** before the incinerator was announced, living near the future site did not affect house prices.

Why add covariates? Compositional changes

- Because this is a **repeated cross-section**, the houses sold in 1981 are different from those sold in 1978.
- What if the *mix* of houses sold changed over time? Look at the average age of houses sold:

```
means_age <- tapply(kielmc$age, list(kielmc$y81, kielmc$nearinc), mean)
colnames(means_age) <- c("Far (nearinc=0)", "Near (nearinc=1)")
rownames(means_age) <- c("1978 (y81=0)", "1981 (y81=1)")
round(means_age, 2)
```

```
             Far (nearinc=0) Near (nearinc=1)
1978 (y81=0)           12.75            39.79
1981 (y81=1)            8.50            27.95
```

- **Compositional change**: In the "Far" group, houses sold in 1981 were 4.25 years newer than in 1978. But in the "Near" group, they were **11.84 years newer**.
- The "Near" group got disproportionately newer. Since newer houses generally sell for more, this drastic change in composition artificially pushes the "Near" average prices up.
- This artificial price bump from the change in age composition partially **masked** the negative effect of the incinerator in our simple DID regression.

DID regression with covariates

- To remove this bias, we must control for age. This adjusts for the fact that the two groups experienced different compositional changes over time.

```
reg_did_cov <- lm(rprice ~ y81 + nearinc + y81nrinc + age + I(age^2),
                  data = kielmc)
round(summary(reg_did_cov)$coefficients, 4)
```

```
               Estimate Std. Error  t value Pr(>|t|)
(Intercept)  89116.5354  2406.0511  37.0385   0.0000
y81          21321.0418  3443.6311   6.1914   0.0000
```

```
nearinc        9397.9359  4812.2218   1.9529   0.0517
y81nrinc     -21920.2700  6359.7454  -3.4467   0.0006
age           -1494.4240   131.8603 -11.3334   0.0000
I(age^2)          8.6913     0.8481  10.2476   0.0000
```

- With age controls: $\hat{\beta} = -\$21{,}920$ (SE = 6,360, $p = 0.0006$) — nearly twice as large, highly significant.
- Estimated bias from compositional change: $-\$11{,}864 - (-\$21{,}920) \approx +\$10{,}056$. The simple DID was severely biased upward because the "Near" houses sold in 1981 were unusually new.

Detecting anticipation and pre-trends

- With only two time periods, anticipation effects **cannot** be separately identified.
- With **multiple pre-treatment periods**, use an **event study** design. Data span $t = -T, \dots, -1, 0, 1, \dots, T'$, where $t = 0$ **is now the treatment date** (not "before" as in the two-period model).
- Replace the single interaction $\beta(t \cdot D_i)$ with a separate coefficient per period. The baseline is $t = -1$ (last pre-treatment period):

$$
\begin{aligned}
Y_{it} = \alpha + \sum_{s \neq -1} \delta_s \cdot \mathbb{1}[t = s] + \gamma D_i \\
+ \sum_{s \neq -1} \beta_s \cdot D_i \cdot \mathbb{1}[t = s] + U_{it}.
\end{aligned}
$$

- Expected outcomes for $t = -2$ and $t = -1$:

	$D_i = 0$ (Control)	$D_i = 1$ (Treatment)
$t = -1$	α	$\alpha + \gamma$
$t = -2$	$\alpha + \delta_{-2}$	$\alpha + \delta_{-2} + \gamma + \beta_{-2}$
Backward trend ($t{:}\,-1 \to -2$)	δ_{-2}	$\delta_{-2} + \beta_{-2}$

- β_{-2} is the difference in backward trends from $t = -1$ to $t = -2$ between treatment and control: $(\delta_{-2} + \beta_{-2}) - \delta_{-2} = \beta_{-2}$.
- More generally, β_s for $s \leq -2$ is the difference in backward trends from $t = -1$ to $t = s$ between treatment and control.
- No anticipation + parallel pre-trends $\implies$ $\beta_s = 0$ for all $s \leq -2$. However, we never observe $Y_{it}(0)$ for the treated group, so nonzero β_s tests the **joint** hypothesis: we cannot determine which assumption failed.
- The critical point: with multiple pre-treatment periods, we can **compare the two groups before treatment**. Since β_s for $s \leq -2$ are estimable from pre-treatment data, $\beta_s = 0$ is a **testable** implication of the joint hypothesis.
- β_s all equal and nonzero **suggests anticipation** at $t = -1$: backward trends are the same for all $s \leq -2$, but something shifts at the baseline period.
- β_s nonzero and unequal **suggests a parallel trends violation**: groups were already diverging before treatment.
- For $s \geq 0$: under parallel trends and no anticipation, β_s measures the **treatment effect** at period s relative to $t = -1$.

Time fixed effects

- In the two-period model, the term $\delta \cdot t$ creates two intercepts: α at $t = 0$ and $\alpha + \delta$ at $t = 1$.
- With multiple periods $t = -T, \dots, -1, 0, 1, \dots, T'$, we cannot use $\delta \cdot t$ because that forces a **linear** trend: the time effect at period s would be $\delta \cdot s$, with no flexibility.

- Instead, include a separate dummy for each period. The regression actually estimated is:

$$Y_{it} = \alpha + \cdots + \delta_{-2} \cdot \mathbb{1}[t = -2] + \delta_0 \cdot \mathbb{1}[t = 0] + \delta_1 \cdot \mathbb{1}[t = 1] + \cdots \\ + \gamma D_i + \cdots + U_{it}.$$

- The key property: $\mathbb{1}[t = s]$ equals 1 when $t = s$ and 0 otherwise. At $t = -1$ all dummies are zero (baseline). At any other t, exactly one dummy equals 1:

Period	Active dummy	Intercept
$t = -1$	none (baseline)	α
$t = 0$	$\mathbb{1}[t = 0] = 1$	$\alpha + \delta_0$
$t = 1$	$\mathbb{1}[t = 1] = 1$	$\alpha + \delta_1$

- Each period gets its own intercept. The compact notation $\sum_{s \neq -1} \delta_s \cdot \mathbb{1}[t = s]$ writes the time dummies as a sum. The event study model with all the dummies written out:

$$Y_{it} = \alpha + \cdots + \delta_{-2}\mathbb{1}[t{=}-2] + \delta_0\mathbb{1}[t{=}0] + \delta_1\mathbb{1}[t{=}1] + \cdots \\ + \gamma D_i + \cdots + \beta_{-2} D_i \mathbb{1}[t{=}-2] + \beta_0 D_i \mathbb{1}[t{=}0] + \beta_1 D_i \mathbb{1}[t{=}1] + \cdots + U_{it}.$$

- At each t, exactly one time dummy survives (and at $t = -1$ none do). So $\alpha + \delta_t$ is the intercept at time t. Define $\lambda_t = \alpha + \delta_t$ (with $\delta_{-1} = 0$):

$$\ldots, \quad \lambda_{-1} = \alpha, \quad \lambda_0 = \alpha + \delta_0, \quad \lambda_1 = \alpha + \delta_1, \quad \ldots$$

The λ_t's are called **time fixed effects**. This is the regression we actually run in OLS, with a dummy for each period:

$$Y_{it} = \cdots + \lambda_{-1}\mathbb{1}[t{=}-1] + \lambda_0\mathbb{1}[t{=}0] + \lambda_1\mathbb{1}[t{=}1] + \cdots \\ + \gamma D_i + \cdots + \beta_{-2} D_i \mathbb{1}[t{=}-2] + \beta_0 D_i \mathbb{1}[t{=}0] + \beta_1 D_i \mathbb{1}[t{=}1] + \cdots + U_{it}.$$

- At each t, exactly one λ-dummy equals 1 and all others are zero:

$$\underbrace{\cdots + \lambda_t \cdot 1 + \cdots}_{\text{only } \lambda_t \text{ survives}} = \lambda_t.$$

The notation λ_t is shorthand for this — it represents whichever λ is active at time t:

$$Y_{it} = \lambda_t + \gamma D_i + \sum_{s \neq -1} \beta_s \cdot D_i \cdot \mathbb{1}[t = s] + U_{it}.$$

Individual fixed effects

- **Note**: The following requires **panel data** (observing the exact same individuals over time). It cannot be used with repeated cross-sections like the incinerator example.

- The same idea can be applied to individuals. In the event study model, the intercept for individual i is $\alpha + \gamma D_i$. This allows only **two** values:

Group	Intercept
Control ($D_i = 0$)	α
Treatment ($D_i = 1$)	$\alpha + \gamma$

All individuals within the same group share the same baseline. But individuals may differ even within a group (e.g., houses near the incinerator differ in size, age, neighborhood quality).

- **Individual fixed effects** give each individual its own dummy, just like time fixed effects give each period its own dummy. The regression is:

$$\begin{aligned} Y_{it} = {} & \alpha_1 \mathbb{1}[i=1] + \cdots + \alpha_n \mathbb{1}[i=n] \\ & + (\text{time and treatment terms}) + U_{it}. \end{aligned}$$

Only the dummy for individual i is active, so the intercept is α_i. Using time fixed effects λ_t from the previous slide:

$$\begin{aligned} Y_{it} = {} & \alpha_i + \lambda_t + \cancel{\gamma D_i} \\ & + \sum_{s \neq -1} \beta_s \cdot D_i \cdot \mathbb{1}[t = s] + U_{it}. \end{aligned}$$

- Since D_i does **not change over time** for any individual, it is already captured by α_i. Including both α_i and γD_i would be **perfect multicollinearity**, so we drop γD_i:

$$\begin{aligned} Y_{it} = {} & \alpha_i + \lambda_t \\ & + \sum_{s \neq -1} \beta_s \cdot D_i \cdot \mathbb{1}[t = s] + U_{it}. \end{aligned}$$

- The interaction terms $\beta_s \cdot D_i \cdot \mathbb{1}[t = s]$ survive because they vary across **both** individuals and time.

Two-way fixed effects

- Combining individual and time fixed effects gives the **two-way fixed effects** (TWFE) model:

$$\begin{aligned} Y_{it} = {} & \alpha_i + \lambda_t \\ & + \sum_{s \neq -1} \beta_s \cdot D_i \cdot \mathbb{1}[t = s] + U_{it}, \end{aligned}$$

where α_i is the individual fixed effect and λ_t is the time fixed effect.

- This is the standard **event study** specification used in the literature for DID with multiple periods of **panel data**.