Lecture 14: Causal inference

Economics 326 — Introduction to Econometrics II

Vadim Marmer, UBC

Motivation: causal questions

- Many important questions in economics are **causal**:
  - Does job training increase earnings?
  - Does a new drug improve health outcomes?
  - Does building an incinerator reduce nearby house prices?
- The **fundamental challenge**: we can never observe the same individual both with and without treatment at the same time.

Potential outcomes

- For each individual i, define two **potential outcomes**:
  - $Y_i(1)$: outcome if individual i receives treatment ($D_i = 1$),
  - $Y_i(0)$: outcome if individual i does not receive treatment ($D_i = 0$).
- The **individual treatment effect** for person i is:

$$Y_i(1) - Y_i(0).$$

- Example: if Y_i is earnings and D_i indicates job training, then $Y_i(1) - Y_i(0)$ is the causal effect of training on earnings for person i.

The fundamental problem

- Let $D_i \in \{0, 1\}$ denote treatment status. The **observed outcome** is:

$$Y_i = D_i \, Y_i(1) + (1 - D_i) \, Y_i(0).$$

- If $D_i = 1$, we observe $Y_i(1)$ but not $Y_i(0)$.
- If $D_i = 0$, we observe $Y_i(0)$ but not $Y_i(1)$.
- We can never observe both potential outcomes for the same individual. This is the **fundamental problem of causal inference**.

Average treatment effects

- Since individual treatment effects $Y_i(1) - Y_i(0)$ are unobservable, we focus on averages.
- **Average Treatment Effect (ATE):**

$$\text{ATE} = \mathrm{E}\left[Y_i(1) - Y_i(0)\right].$$

- **Average Treatment Effect on the Treated (ATT):**

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

- ATE averages over the entire population; ATT averages only over those who actually receive treatment.

Selection bias

- A naive comparison of treated and untreated outcomes yields:

$$\begin{aligned}
&\text{E}\left[Y_i \mid D_i = 1\right] - \text{E}\left[Y_i \mid D_i = 0\right] \\
&= \text{E}\left[Y_i(1) \mid D_i = 1\right] - \text{E}\left[Y_i(0) \mid D_i = 0\right].
\end{aligned}$$

- Add and subtract $\text{E}\left[Y_i(0) \mid D_i = 1\right]$:

$$\begin{aligned}
&= \underbrace{\text{E}\left[Y_i(1) - Y_i(0) \mid D_i = 1\right]}_{\text{ATT}} \\
&\quad + \underbrace{\text{E}\left[Y_i(0) \mid D_i = 1\right] - \text{E}\left[Y_i(0) \mid D_i = 0\right]}_{\text{Selection bias}}.
\end{aligned}$$

- **Selection bias** arises when the treatment and control groups differ in their baseline outcomes $Y_i(0)$. For example, people who choose to enroll in job training may differ systematically from those who do not.

Random assignment

- Under **random assignment**, treatment D_i is independent of potential outcomes:

$$\text{E}\left[Y_i(0) \mid D_i = 1\right] = \text{E}\left[Y_i(0) \mid D_i = 0\right] = \text{E}\left[Y_i(0)\right].$$

- The selection bias vanishes, and the simple difference in means equals the ATE:

$$\text{E}\left[Y_i \mid D_i = 1\right] - \text{E}\left[Y_i \mid D_i = 0\right] = \text{ATE} = \text{ATT}.$$

- Randomized experiments (like the National Supported Work program) achieve this by randomly assigning individuals to treatment and control groups.

Regression with a treatment dummy

- With random assignment, the ATE can be estimated by regressing Y_i on a treatment dummy:

$$Y_i = \alpha + \tau D_i + U_i,$$

where $\text{E}\left[U_i \mid D_i\right] = 0$ under randomization.

- The OLS estimate of τ equals the difference in sample means:

$$\hat{\tau} = \bar{Y}_1 - \bar{Y}_0,$$

where $\bar{Y}_1$ and $\bar{Y}_0$ are the sample averages for the treated and control groups.

Example: Lalonde data

- The **National Supported Work (NSW)** demonstration recruited disadvantaged workers (long-term unemployed, high-school dropouts, former drug users, ex-offenders).
- Among eligible applicants, some were **randomly assigned** to receive job training (treated group), and the rest formed the **experimental control group**. Both groups come from the same disadvantaged population.
- We use the `jtrain2` dataset from the `wooldridge` package, based on the Lalonde (1986) study:

```
library(wooldridge)
data(jtrain2)
head(jtrain2[, c("train", "re78", "educ", "age", "black", "married")], n = 10)
```

```
   train     re78 educ age black married
1      1  9.93005   11  37     1       1
2      1  3.59589    9  22     0       0
3      1 24.90950   12  30     1       0
4      1  7.50615   11  27     1       0
5      1  0.28979    8  33     1       0
6      1  4.05649    9  22     1       0
7      1  0.00000   12  23     1       0
8      1  8.47216   11  32     1       0
9      1  2.16402   16  22     1       0
10     1 12.41810   12  33     0       1
```

- `train`: 1 if randomly assigned to job training, 0 if assigned to control.
- `re78`: real earnings in 1978 (thousands of dollars).

Estimating the ATE

- Since treatment was randomly assigned, the coefficient on `train` estimates the ATE:

```
summary(lm(re78 ~ train, data = jtrain2))$coefficients
```

```
            Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 4.554802  0.4080460 11.162474 1.154113e-25
train       1.794343  0.6328536  2.835321 4.787524e-03
```

- The estimated ATE is approximately \$1,800 (recall that `re78` is in thousands): on average, participation in the job training program increased 1978 earnings by about \$1,800.

Observational studies

- In many settings, treatment is **not randomly assigned**. Individuals self-select into treatment.
- Example: workers who voluntarily enroll in job training may be more motivated or have lower baseline earnings than those who do not enroll.
- Self-selection generates **selection bias**, and the simple difference in means no longer estimates the ATE.
- To estimate treatment effects from observational data, we need to **control for covariates** that affect both treatment selection and outcomes.

Potential outcomes with a covariate

- Suppose the potential outcomes depend linearly on a covariate X_i:

$$
\begin{aligned}
Y_i(0) &= \alpha_0 + \beta_0 X_i + U_i(0),\\
Y_i(1) &= \alpha_1 + \beta_1 X_i + U_i(1),
\end{aligned}
$$

where $\mathrm{E}\left[U_i(0) \mid X_i, D_i\right] = 0$ and $\mathrm{E}\left[U_i(1) \mid X_i, D_i\right] = 0$.

- These assumptions mean that, after conditioning on X_i, treatment assignment D_i is as good as random. This is called **conditional mean independence** (or “selection on observables”).

The ATE with a covariate

- Taking expectations of the potential outcomes:

$$
\begin{aligned}
\mathrm{E}\left[Y_i(1)\right] &= \alpha_1 + \beta_1 \mathrm{E}\left[X_i\right], \\
\mathrm{E}\left[Y_i(0)\right] &= \alpha_0 + \beta_0 \mathrm{E}\left[X_i\right].
\end{aligned}
$$

- The ATE is:

$$
\begin{aligned}
\mathrm{ATE} &= \mathrm{E}\left[Y_i(1) - Y_i(0)\right] \\
&= (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)\mathrm{E}\left[X_i\right].
\end{aligned}
$$

- The ATE depends on the difference in intercepts **and** the difference in slopes, weighted by the population mean of X_i.

Two separate regressions

- One approach: estimate separate regressions for each group.
- **Control group** ($D_i = 0$): $Y_i = \alpha_0 + \beta_0 X_i + U_i(0)$.
- **Treatment group** ($D_i = 1$): $Y_i = \alpha_1 + \beta_1 X_i + U_i(1)$.
- The estimated ATE is:

$$
\begin{aligned}
\widehat{\mathrm{ATE}} = (\hat{\alpha}_1 - \hat{\alpha}_0) \\
+ (\hat{\beta}_1 - \hat{\beta}_0)\bar{X},
\end{aligned}
$$

where $\bar{X}$ is the overall sample mean of X_i.

Combined regression with interactions

- The observed outcome $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$ can be written as a single regression. Expanding:

$$
\begin{aligned}
Y_i &= \alpha_0(1 - D_i) + \alpha_1 D_i \\
&\quad + \beta_0 X_i (1 - D_i) + \beta_1 X_i D_i + \tilde{U}_i \\
&= \alpha_0 + (\alpha_1 - \alpha_0) D_i \\
&\quad + \beta_0 X_i + (\beta_1 - \beta_0) X_i D_i + \tilde{U}_i,
\end{aligned}
$$

where $\tilde{U}_i = (1 - D_i) U_i(0) + D_i U_i(1)$.

- This is a regression of Y_i on D_i, X_i, and the interaction $X_i D_i$.
- The coefficient on D_i is $\alpha_1 - \alpha_0$, which is **not** the ATE (unless $\beta_1 = \beta_0$).

The demeaning trick

- Write $X_i = \mathrm{E}[X_i] + (X_i - \mathrm{E}[X_i])$. Then the interaction term becomes:

$$\begin{aligned}(\beta_1 - \beta_0)X_i D_i = (\beta_1 - \beta_0)\mathrm{E}[X_i]\, D_i \\ + (\beta_1 - \beta_0)(X_i - \mathrm{E}[X_i])D_i.\end{aligned}$$

- Substituting into the regression:

$$\begin{aligned}Y_i = \alpha_0 + \Big[(\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)\mathrm{E}[X_i]\Big] D_i \\ + \beta_0 X_i + (\beta_1 - \beta_0)(X_i - \mathrm{E}[X_i])D_i + \tilde{U}_i.\end{aligned}$$

- Define $\tau = (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)\mathrm{E}[X_i]$ and $\delta = \beta_1 - \beta_0$:

$$\begin{aligned}Y_i = \alpha_0 + \tau D_i + \beta_0 X_i \\ + \delta(X_i - \mathrm{E}[X_i])D_i + \tilde{U}_i.\end{aligned}$$

- The coefficient τ on D_i is exactly the **ATE**.

Estimating the ATE with covariates

- In practice, replace $\mathrm{E}[X_i]$ with the sample mean $\bar{X}$ and run the regression:

$$\begin{aligned}Y_i = \hat{\alpha}_0 + \hat{\tau} D_i + \hat{\beta}_0 X_i \\ + \hat{\delta}(X_i - \bar{X})D_i + \text{residual}.\end{aligned}$$

- The coefficient $\hat{\tau}$ on D_i directly estimates the ATE.
- This works because demeaning the interaction "absorbs" the $(\beta_1 - \beta_0)\mathrm{E}[X_i]$ part of the ATE into the coefficient on D_i.

Why not regress Y_i on D_i and X_i?

- A simpler regression omits the interaction:

$$Y_i = a + \tau D_i + b X_i + V_i.$$

- This is valid if $\beta_1 = \beta_0$ (the covariate has the same effect in both groups). Then $\delta = 0$, the interaction drops out, and the coefficient on D_i is the ATE.
- If $\beta_1 \neq \beta_0$, the omitted interaction creates bias. The regression with the demeaned interaction nests the simpler model as a special case.

Example: separate regressions

- Estimate separate regressions of `re78` on `educ` for the treatment and control groups:

```r
reg0 <- lm(re78 ~ educ, data = jtrain2, subset = (train == 0))
reg1 <- lm(re78 ~ educ, data = jtrain2, subset = (train == 1))
cbind(Control = coef(reg0), Treatment = coef(reg1))
```

```
                 Control  Treatment
(Intercept) 3.80301658 -0.7821703
educ        0.07451936  0.6892860
```

- Compute the estimated ATE:

```r
xbar <- mean(jtrain2$educ)
a0 <- coef(reg0)[1]; b0 <- coef(reg0)[2]
a1 <- coef(reg1)[1]; b1 <- coef(reg1)[2]
ATE_manual <- (a1 - a0) + (b1 - b0) * xbar
cat("Sample mean of educ:", round(xbar, 2), "\n")
```

```
Sample mean of educ: 10.2
```

```r
cat("Estimated ATE:", round(ATE_manual, 2), "\n")
```

```
Estimated ATE: 1.68
```

Example: demeaned regression

- Create the demeaned interaction and run the combined regression:

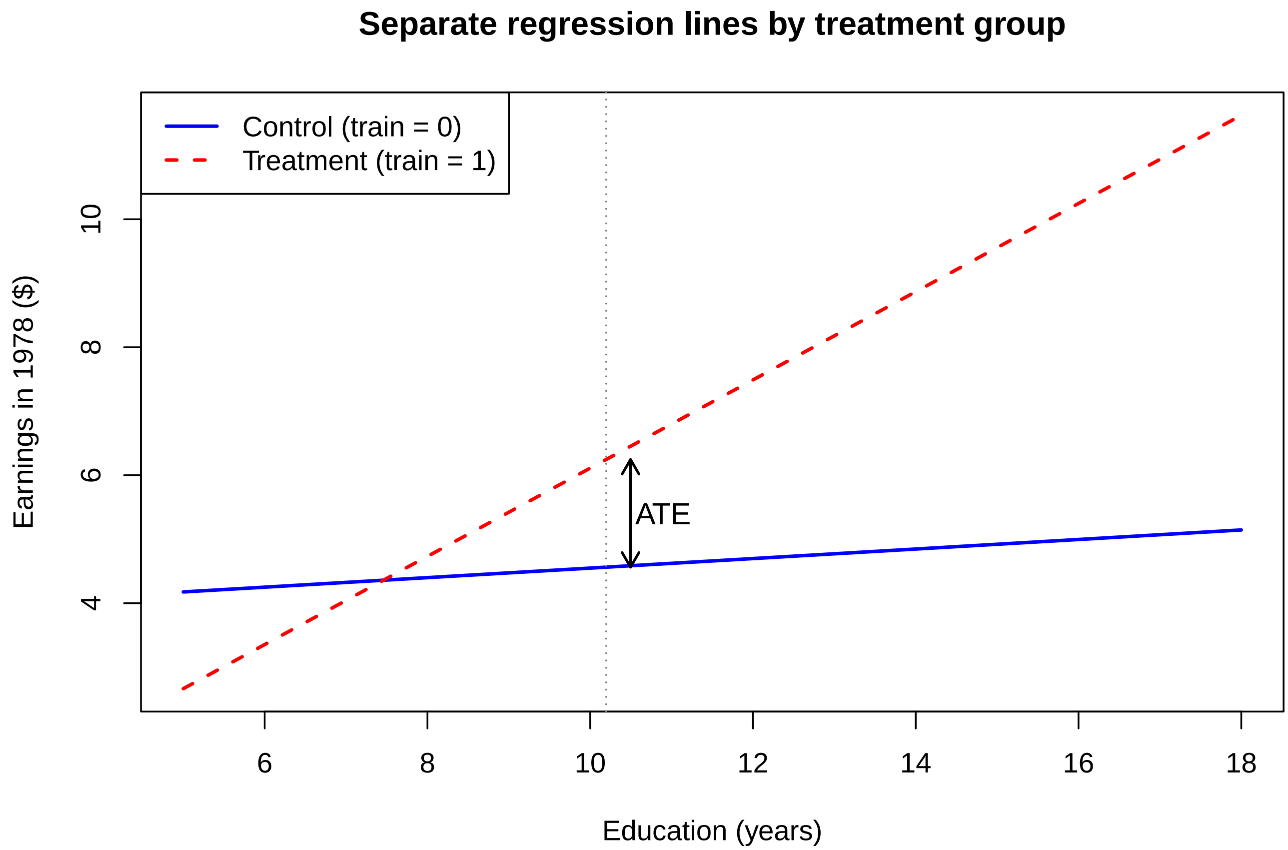
```r
jtrain2$educ_dm <- jtrain2$educ - mean(jtrain2$educ)
reg_dm <- lm(re78 ~ train + educ + I(train * educ_dm), data = jtrain2)
summary(reg_dm)$coefficients
```

```
                     Estimate Std. Error   t value   Pr(>|t|)
(Intercept)        3.80301658  2.5689136 1.4803988 0.13948090
train              1.68266981  0.6299604 2.6710725 0.00784062
educ               0.07451936  0.2514521 0.2963561 0.76709766
I(train * educ_dm) 0.61476663  0.3472755 1.7702560 0.07737534
```

- The coefficient on `train` directly estimates the ATE, matching the result from the separate regressions approach.

Example: regression lines

- The two regression lines, with a vertical line at $\bar{X}$ and the ATE marked as the gap:



Example: adding more covariates

- So far, we used only `educ` as the control. The dataset contains more pre-treatment characteristics: `age`, `black`, `married`.
- With random assignment, slopes are approximately equal across groups, so we can add controls directly without demeaned interactions:

```
reg_X <- lm(re78 ~ train + educ + age + black + married, data = jtrain2)
round(summary(reg_X)$coefficients, 1)
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.9        2.2     0.4      0.7
train            1.7        0.6     2.7      0.0
educ             0.4        0.2     2.4      0.0
age              0.1        0.0     1.2      0.2
black           -2.3        0.8    -2.7      0.0
married          0.2        0.8     0.2      0.9
```

Comparison: simple vs. controlled

- Compare the estimated ATE and its standard error across specifications:

```
m1 <- lm(re78 ~ train, data = jtrain2)
m2 <- lm(re78 ~ train + educ + age + black + married, data = jtrain2)
tab <- rbind(
  "No controls"   = c(Estimate = coef(m1)["train"],
                      SE       = summary(m1)$coef["train", "Std. Error"]),
  "With controls" = c(Estimate = coef(m2)["train"],
                      SE       = summary(m2)$coef["train", "Std. Error"])
)
round(tab, 0)
```

```
               Estimate.train SE
No controls                 2  1
With controls               2  1
```

- The estimated ATE barely changes. The standard error shrinks when controls are added.

Why do controls change little here?

- Under **random assignment**, D_i is independent of all covariates: the treatment and control groups are balanced in expectation.
- Because of balance, including X_i does not change the coefficient on D_i: there is no selection bias to remove.

Why does the standard error shrink?

- Recall that the variance of the OLS estimator depends on the residual variance $\hat{\sigma}^2$:

$$\widehat{\mathrm{Var}}\,(\hat{\tau}) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(D_i - \bar{D})^2(1 - R_D^2)},$$

  where R_D^2 is the R^2 from regressing D_i on the other regressors.

- With randomization, D_i is nearly uncorrelated with covariates, so $R_D^2 \approx 0$: the denominator stays roughly the same.
- However, covariates that predict earnings absorb variation in Y_i, reducing $\hat{\sigma}^2$. Since $\mathrm{se}\,(\hat{\tau}) \propto \hat{\sigma}$, the standard error shrinks.
- In short: **adding covariates under randomization buys precision without changing the point estimate**.

Example: observational data (jtrain3)

- Lalonde (1986) asked: what if we had no experiment and instead compared the NSW trainees to ordinary workers?
- The `jtrain3` dataset keeps the **same 185 NSW trainees** as the treated group but replaces the experimental control group with **2,490 respondents from the Current Population Survey (CPS)**, a nationally representative survey of American workers.

```r
data(jtrain3)
cat("n =", nrow(jtrain3), " (NSW trainees:", sum(jtrain3$train == 1),
    ", CPS controls:", sum(jtrain3$train == 0), ")\n")
```

```
n = 2675  (NSW trainees: 185 , CPS controls: 2490 )
```

- CPS respondents were not selected for being disadvantaged. They are typical American workers with higher education, higher earnings, and more stable employment than NSW participants.

The selection problem in jtrain3

- The two groups have very different baseline characteristics:

```r
grp <- split(jtrain3, jtrain3$train)
tab <- rbind(
  "NSW trainees (train=1)" = colMeans(grp[["1"]][, c("re78","re75","educ","age","black","married")]),
  "CPS controls (train=0)" = colMeans(grp[["0"]][, c("re78","re75","educ","age","black","married")])
)
round(tab, 1)
```

```
                          re78 re75 educ  age black married
NSW trainees (train=1)  6.3  1.5 10.3 25.8   0.8     0.2
CPS controls (train=0) 21.6 19.1 12.1 34.9   0.3     0.9
```

- CPS workers earn about $21,500 in 1978; NSW trainees earn about $6,300. This gap is not a treatment effect — it reflects the very different populations.

Observational: without controls

- A simple regression of earnings on the training dummy:

```r
obs1 <- lm(re78 ~ train, data = jtrain3)
round(summary(obs1)$coefficients, 3)
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   21.554      0.304  70.985        0
train        -15.205      1.155 -13.169        0
```

- The estimated "effect" is **large and negative**: trainees appear to earn about $15,200 less than CPS workers.
- This is pure **selection bias**: the NSW program recruited disadvantaged workers who would have earned less than CPS respondents regardless of training.

Observational: with controls

- Adding covariates dramatically changes the estimate:

```r
obs2 <- lm(re78 ~ train + educ + age + black + hisp + married + re74 + re75,
           data = jtrain3)
round(summary(obs2)$coefficients, 3)
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.777      1.366   0.569    0.570
train          0.860      0.908   0.947    0.344
educ           0.528      0.075   7.015    0.000
age           -0.082      0.021  -3.940    0.000
black         -0.543      0.494  -1.098    0.272
hisp           2.166      1.092   1.983    0.048
married        1.220      0.586   2.083    0.037
re74           0.278      0.028   9.952    0.000
re75           0.568      0.028  20.615    0.000
```

- The coefficient on `train` shifts from about $-$15.2 to about $+$0.9 (thousands), much closer to the experimental benchmark of $+$1.8.

Experimental vs. observational

- Side-by-side comparison:

```r
make_row <- function(m) c(Estimate = coef(m)["train"],
                          SE = summary(m)$coef["train", "Std. Error"])
tab <- rbind(
  "Experimental: no controls"   = make_row(m1),
  "Experimental: with controls" = make_row(m2),
  "Observational: no controls"  = make_row(obs1),
  "Observational: with controls"= make_row(obs2)
)
round(tab, 3)
```

```
                             Estimate.train    SE
Experimental: no controls             1.794 0.633
Experimental: with controls           1.679 0.629
```

```
Observational: no controls          -15.205 1.155
Observational: with controls          0.860 0.908
```

- **Experimental data:** controls barely change the estimate (1.79 vs. 1.68); standard errors shrink.
- **Observational data:** controls remove most of the selection bias ($-$15.2 → $+$0.9); covariates are essential for a credible estimate.

From cross-sections to panel data

- So far, we used **cross-sectional** data and assumed selection on observables: after controlling for X_i, treatment is as good as random.
- In some settings, this assumption is hard to justify. An alternative approach exploits **panel data** (repeated observations on the same units over time).
- The **difference-in-differences (DID)** method compares changes over time between a treatment group and a control group.

DID 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, and only the treatment group ($D_i = 1$) is affected.
- We observe Y_{it}: the outcome for individual i at time t.

DID regression model

- The DID regression is:

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

  where $\mathrm{E}\left[U_{it} \mid D_i\right] = 0$.

- The 2×2 table of conditional means:

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

Interpreting the coefficients

- α: baseline expected outcome (control group, before treatment).
- δ: **time effect** — the change in the control group's outcome from $t = 0$ to $t = 1$. This captures common trends (e.g., inflation, economic growth).
- γ: **group difference** at baseline — the pre-existing difference between treatment and control groups at $t = 0$.
- β: **DID estimand** — the additional change in the treatment group's outcome, beyond what the control group experienced.

Deriving the DID estimand

- For each combination of t and D_i, take expectations using $\mathrm{E}[U_{it} \mid D_i] = 0$:

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

- Subtracting the control group change from the treatment group change:

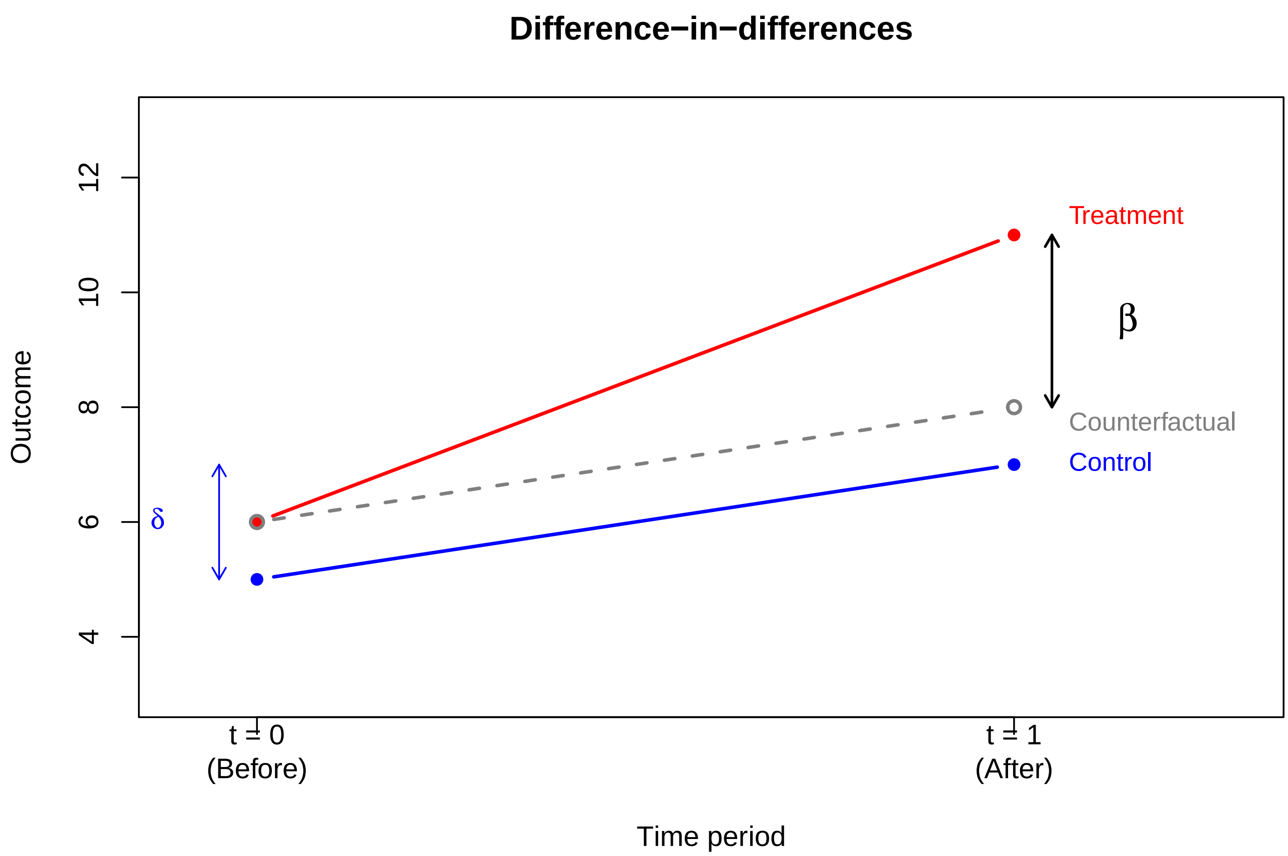
$$
\begin{aligned}
\beta = &\Big(\mathrm{E}[Y_{i1} \mid D_i = 1] - \mathrm{E}[Y_{i0} \mid D_i = 1]\Big)\\
&- \Big(\mathrm{E}[Y_{i1} \mid D_i = 0] - \mathrm{E}[Y_{i0} \mid D_i = 0]\Big).
\end{aligned}
$$

- The DID estimand is the **difference in within-group changes** over time:

$$
\begin{aligned}
\beta = &\,\mathrm{E}[Y_{i1} - Y_{i0} \mid D_i = 1]\\
&- \mathrm{E}[Y_{i1} - Y_{i0} \mid D_i = 0]\,.
\end{aligned}
$$

DID diagram

- The classic DID diagram shows the control group, treatment group, and counterfactual:



Example: incinerator and house prices

- Kiel and McClain (1995) studied how the construction of a garbage incinerator affected nearby house prices in North Andover, Massachusetts.
- We use the `kielmc` dataset from the `wooldridge` package:

```
data(kielmc)
head(kielmc[, c("rprice", "y81", "nearinc", "y81nrinc", "age")], n = 10)
```

```
   rprice y81 nearinc y81nrinc age
1   60000   0       1        0  48
2   40000   0       1        0  83
3   34000   0       1        0  58
4   63900   0       1        0  11
5   44000   0       1        0  48
6   46000   0       1        0  78
7   56000   0       1        0  22
8   38500   0       1        0  78
9   60500   0       1        0  42
10  55000   0       1        0  41
```

- `rprice`: house price in 1978 dollars.
- `y81`: 1 if year is 1981 (after incinerator announced), 0 if 1978.
- `nearinc`: 1 if house is near the incinerator site.
- `y81nrinc`: interaction `y81 * nearinc`.

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)
```

```
             Far (nearinc=0) Near (nearinc=1)
1978 (y81=0)           82517            63693
1981 (y81=1)          101308            70619
```

- 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), "\n")
```

```
Change (near): 6926
```

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

```
Change (far):  18790
```

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

```
DID:           -11864
```

DID regression

- The DID regression:

```r
reg_did <- lm(rprice ~ y81 + nearinc + y81nrinc, data = kielmc)
summary(reg_did)$coefficients
```

```
             Estimate Std. Error  t value     Pr(>|t|)
(Intercept)  82517.23   2726.910 30.260341 1.709246e-95
y81          18790.29   4050.065  4.639502 5.116892e-06
nearinc     -18824.37   4875.322 -3.861154 1.368017e-04
y81nrinc    -11863.90   7456.646 -1.591051 1.125948e-01
```

- The coefficient on `y81nrinc` matches the DID computed from the 2×2 table.
- The estimated effect is negative (incinerator reduced nearby prices), but the p-value is around 0.11, so it is not statistically significant at the 5% level.

DID with covariates

- Adding house characteristics as controls can improve precision:

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

```
                Estimate   Std. Error    t value      Pr(>|t|)
(Intercept)  89116.535375 2406.0510717  37.038505 8.247920e-117
y81          21321.041753 3443.6310979   6.191442  1.857145e-09
nearinc       9397.935862 4812.2218389   1.952931  5.171307e-02
y81nrinc    -21920.269951 6359.7453905  -3.446721  6.444775e-04
age          -1494.424046  131.8603155 -11.333387  3.347719e-25
I(age^2)         8.691277    0.8481268  10.247615  1.859361e-21
```

- After controlling for house age, the DID estimate becomes larger in magnitude and statistically significant.
- Controlling for covariates reduces residual variance, leading to more precise estimates.

DID and potential outcomes

- To connect DID with the potential outcomes framework, define **panel potential outcomes**: $Y_{it}(d)$ is the outcome for individual i at time t if assigned to group $d \in \{0, 1\}$.
- 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 ($D_i = 1$)
$t = 0$	$Y_{i0}(0)$	$Y_{i0}(1)$
$t = 1$	$Y_{i1}(0)$	$Y_{i1}(1)$

- The treatment effect at time $t = 1$ for the treated group is:

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

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

DID as a treatment effect

- Recall that $\beta = \mathrm{E}[Y_{i1} - Y_{i0} \mid D_i = 1] - \mathrm{E}[Y_{i1} - Y_{i0} \mid D_i = 0]$.
- Substituting observed outcomes with potential outcomes:

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

- To relate β to the ATT, add and subtract $\mathrm{E}[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 1]$:

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

- For β to equal the ATT, we need two additional assumptions.

Assumption 1: no anticipation

- **No anticipation:** the outcome at $t = 0$ (before treatment) is not affected by future treatment group assignment:

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

- Being assigned to the treatment group does not change pre-treatment outcomes in expectation.
- In the incinerator example: before the incinerator was announced, living near the future site did not affect house prices (relative to what they would have been otherwise).
- Under no anticipation, we can replace $Y_{i0}(1)$ with $Y_{i0}(0)$ in the expression for β:

$$\begin{aligned}\beta = {} & \text{ATT} + \mathrm{E}[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 1] \\ & - \mathrm{E}[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0].\end{aligned}$$

Assumption 2: parallel trends

- **Parallel trends:** in the absence of treatment, both groups would have experienced the same change over time:

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

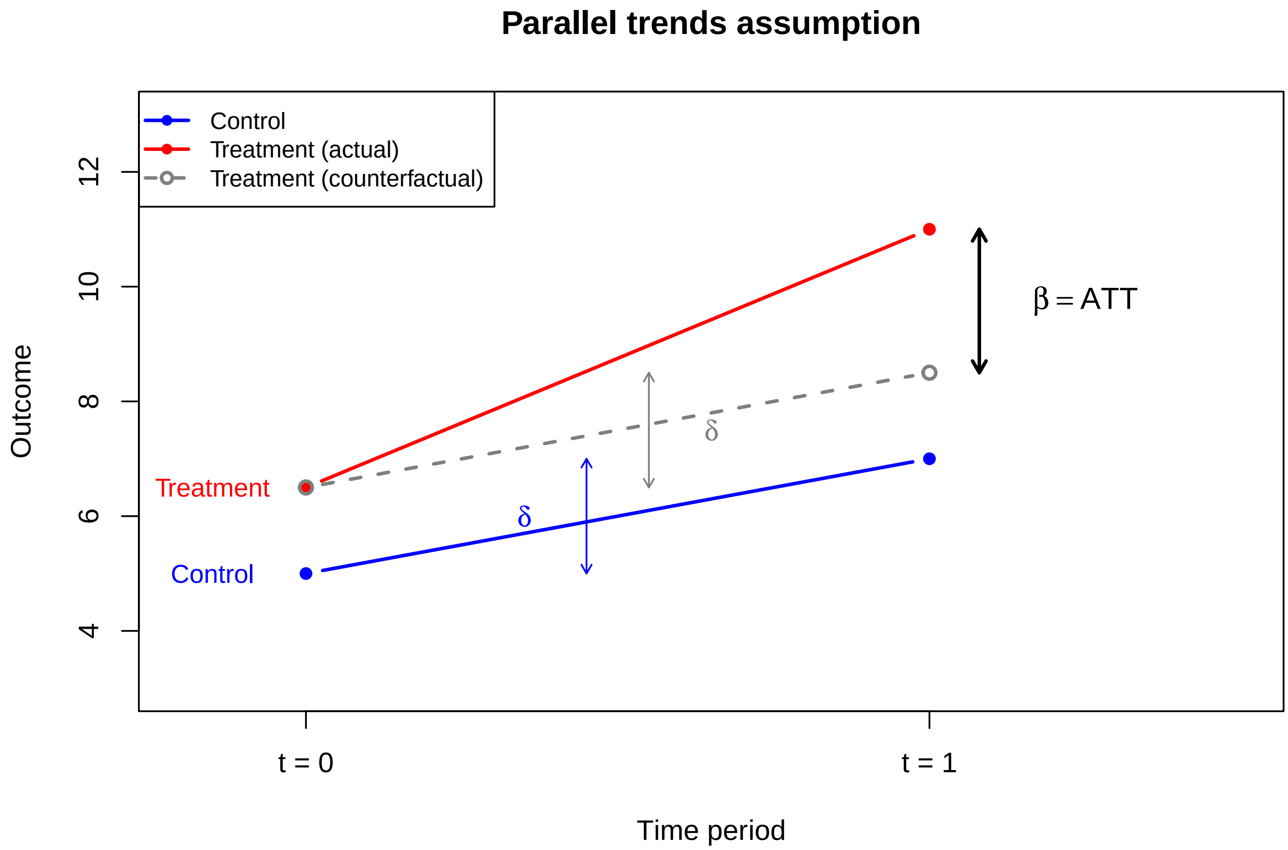
- Under both no anticipation and parallel trends:

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

- The parallel trends assumption cannot be directly tested because $Y_{i1}(0)$ is unobserved for the treated group. However, if pre-treatment data for multiple periods exist, one can check whether trends were parallel before treatment.

Parallel trends diagram

- Illustrating the parallel trends assumption:



Summary

- **Potential outcomes** $Y_i(1)$ and $Y_i(0)$ formalize causal effects. The individual treatment effect $Y_i(1) - Y_i(0)$ is never fully observed.
- The **ATE** and **ATT** are population-level summaries of treatment effects.
- With **random assignment**, a simple regression of Y_i on D_i estimates the ATE.
- With **observational data**, controlling for covariates and using the **demeaning trick** (interacting D_i with $X_i - \bar{X}$) allows the coefficient on D_i to estimate the ATE.
- **Difference-in-differences** uses panel data to compare changes over time between groups:

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

- Under **no anticipation** and **parallel trends**, the DID estimand β equals the ATT.