Module 4: Difference-in-Differences and Effects of Medicaid Expansion

Part 2: Understanding Difference-in-Differences

Ian McCarthy | Emory University
Econ 470 & HLTH 470

Setup

- Denote by $Y_1(t)$ the (potential) outcome at time t with treatment
- Denote by $Y_0(t)$ the (potential) outcome at time t without treatment
- Consider $t = 0$ as the pre-period, $t = 1$ as the post-period
- Four potential outcomes: $Y_1(0)$, $Y_1(1)$, $Y_0(0)$, and $Y_0(1)$.

Setup

Want to estimate $ATT = E[Y_1(1) - Y_0(1)|D = 1]$

	Post-period	Pre-period
Treated	\`$E(Y_1(1)|D = 1)$\`	\`$E(Y_0(0)|D = 1)$\`
Control	\`$E(Y_0(1)|D = 0)$\`	\`$E(Y_0(0)|D = 0)$\`

Problem: We don't see $E[Y_0(1)|D = 1]$

Setup

Want to estimate $ATT = E[Y_1(1) - Y_0(1)|D = 1]$

	Post-period	Pre-period
Treated	`$E(Y_1(1)|D = 1)$`	`$E(Y_0(0)|D = 1)$`
Control	`$E(Y_0(1)|D = 0)$`	`$E(Y_0(0)|D = 0)$`

Strategy 1: Estimate $E[Y_0(1)|D = 1]$ using $E[Y_0(0)|D = 1]$ (before treatment outcome used to estimate post-treatment)

Setup

Want to estimate $ATT = E[Y_1(1) - Y_0(1)|D = 1]$

	Post-period	Pre-period
Treated	`` ` ``$E(Y_1(1)\|D = 1)$`` ` ``	`` ` ``$E(Y_0(0)\|D = 1)$`` ` ``
Control	`` ` ``$E(Y_0(1)\|D = 0)$`` ` ``	`` ` ``$E(Y_0(0)\|D = 0)$`` ` ``

Strategy 2: Estimate $E[Y_0(1)|D = 1]$ using $E[Y_0(1)|D = 0]$ (control group used to predict outcome for treatment)

Setup

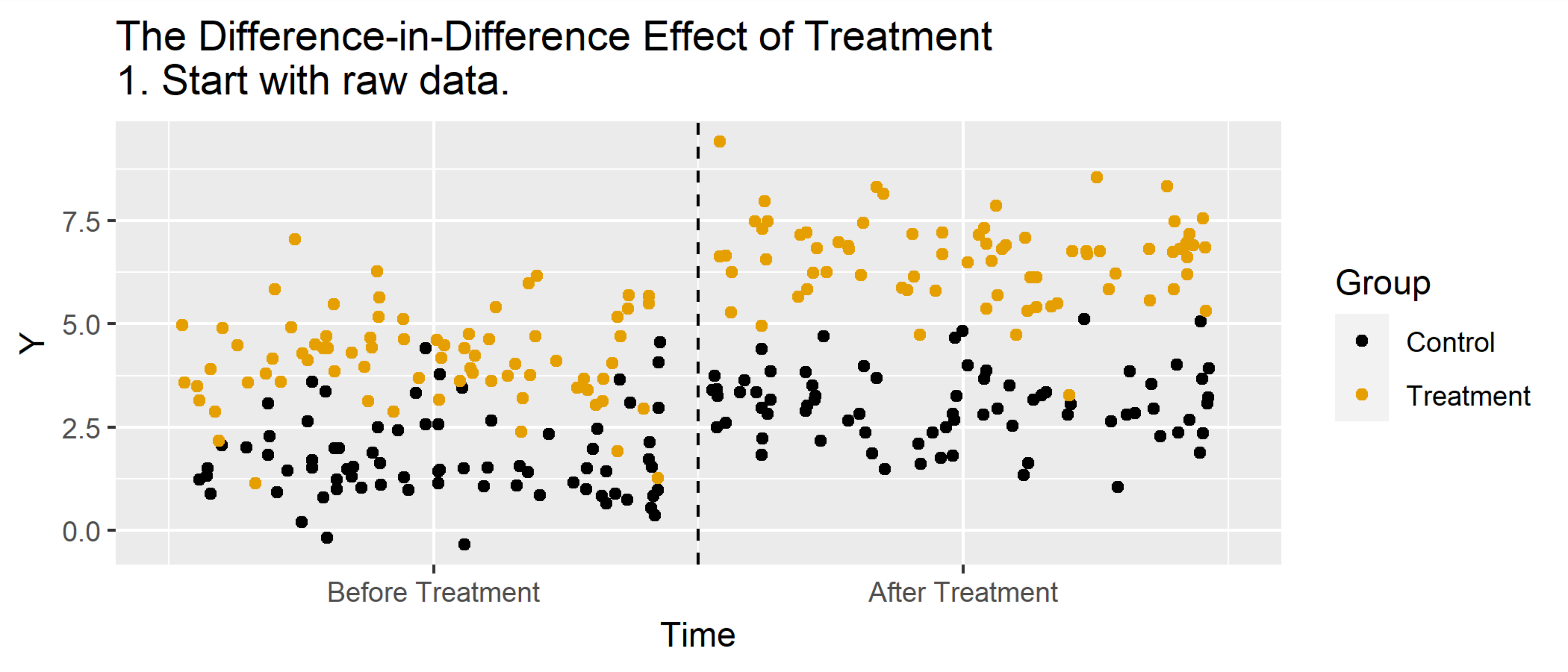
Want to estimate $ATT = E[Y_1(1) - Y_0(1)|D = 1]$

	Post-period	Pre-period
Treated	\`$E(Y_1(1)|D = 1)$\`	\`$E(Y_0(0)|D = 1)$\`
Control	\`$E(Y_0(1)|D = 0)$\`	\`$E(Y_0(0)|D = 0)$\`

Strategy 3: DD estimate...

Estimate $E[Y_1(1)|D = 1] - E[Y_0(1)|D = 1]$ using $E[Y_0(1)|D = 0] - E[Y_0(0)|D = 0]$ (pre-post difference in control group used to predict difference for treatment group)

Animations!



The Difference-in-Difference Effect of Treatment
1. Start with raw data.
Y
7.5
5.0
2.5
0.0
Before Treatment
After Treatment
Time
Group
Control
Treatment

Estimation

Key identifying assumption is that of *parallel trends*

$$E[Y_0(1) - Y_0(0)|D = 1] = E[Y_0(1) - Y_0(0)|D = 0]$$

Estimation

Sample means:

$$
\begin{aligned}
E[Y_1(1) - Y_0(1)|D = 1] = \quad & (E[Y(1)|D = 1] - E[Y(1)|D = 0]) \\
& - (E[Y(0)|D = 1] - E[Y(0)|D = 0])
\end{aligned}
$$

Estimation

Regression:

$$Y_i = \alpha + \beta D_i + \lambda 1(Post) + \delta D_i \times 1(Post) + \varepsilon$$

	After	Before	After - Before
Treated	`$\alpha + \beta + \lambda + \delta$`	`$\alpha + \beta$`	`$\lambda + \delta$`
Control	`$\alpha + \lambda$`	`$\alpha$`	`$\lambda$`
Treated - Control	`$\beta + \delta$`	`$\beta$`	`$\delta$`

Simulated data

```r
N <- 5000
dd.dat <- tibble(
  d = (runif(N, 0, 1)>0.5),
  time_pre = "pre",
  time_post = "post"
)

dd.dat <- pivot_longer(dd.dat, c("time_pre","time_post"), values_to="time") %>%
  select(d, time) %>%
  mutate(t=(time=="post"),
         y.out=1.5+3*d + 1.5*t + 6*d*t + rnorm(N*2,0,1))
```

Mean differences

```
dd.means <- dd.dat %>% group_by(d, t) %>% summarize(mean_y = mean(y.out))
knitr::kable(dd.means, col.names=c("Treated","Post","Mean"), format="html")
```

Treated	Post	Mean
FALSE	FALSE	1.522635
FALSE	TRUE	3.002374
TRUE	FALSE	4.515027
TRUE	TRUE	12.004623

Mean differences

In this example:

- $E[Y(1)|D = 1] - E[Y(1)| = 0]$ is 9.0022495
- $E[Y(0)|D = 1] - E[Y(0)|D = 0]$ is 2.9923925

So the ATT is 6.0098571

Regression estimator

```
dd.est <- lm(y.out ~ d + t + d*t, data=dd.dat)
summary(dd.est)
```

```
##
## Call:
## lm(formula = y.out ~ d + t + d * t, data = dd.dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0038 -0.6674  0.0047  0.6609  3.6135
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.52263    0.01970   77.28   <2e-16 ***
## dTRUE        2.99239    0.02795  107.07   <2e-16 ***
## tTRUE        1.47974    0.02786   53.10   <2e-16 ***
## dTRUE:tTRUE  6.00986    0.03953  152.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9881 on 9996 degrees of freedom
## Multiple R-squared:  0.9433,    Adjusted R-squared:  0.9433
## F-statistic: 5.543e+04 on 3 and 9996 DF,  p-value: < 2.2e-16
```