

## 1 Normalization of the Selection Equation

In class, we said that

$$ATT = \int_0^1 MTE(u) \frac{\mathbb{P}(p(Z) \geq u)}{\mathbb{P}(D = 1)} du$$

This result relies on the following normalization.

$$D = 1\{U \leq v(Z)\} = 1\{\tilde{U} \leq p(Z)\}$$

where  $p$  is the propensity score.

1. We write  $D = 1\{F_U(U) \leq F_U(v(Z))\}$ , where  $F_U$  is the CDF of the continuously distributed  $U$ . What is the distribution of  $F_U(U)$ ? Let  $\tilde{U} := F_U(U)$  from now on.

*Solution.* Given  $U$  is a continuously distributed random variable, its CDF is also a continuous distribution. From the Statistics lecture we learned that, the CDF of any random variable follows uniform  $[0, 1]$ , so  $\tilde{U} \sim \text{Uniform}[0, 1]$ . ■

2. Define  $p(z) := P(D = 1 | Z = z)$ , show that  $p(z) = F_U(v(z))$ .

*Solution.* Given the definition of  $p(z)$ , we can obtain

$$\begin{aligned} p(z) &\triangleq \mathbb{P}(D = 1 | Z = z) \\ &= \mathbb{P}(F_U(U) \leq F_U(v(z))) \quad (\text{Definition}) \\ &= F_U(v(z)) \quad (F_U(U) \sim \text{Uniform}[0, 1]) \end{aligned}$$

3. Show that  $D = 1\{\tilde{U} \leq p(Z)\}$ .

*Solution.* Starting from the definition of  $D$  in (a), note that  $\tilde{U} \triangleq F_U(U)$  and the result in (b) shows that  $p(z) = F_U(v(z))$ , so

$$\begin{aligned} D &= 1\{F_U(U) \leq F_U(v(z))\} \\ &= 1\{\tilde{U} \leq p(z)\} \end{aligned}$$

## 2 Derivation of the Weights for LATE

In this exercise, we try to show that LATE of instrument  $z$  and  $z'$  can be written as:

$$\text{LATE}_{z'}^z = \frac{E[Y|Z = z] - E[Y|Z = z']}{E[D|Z = z] - E[D|Z = z']} = \int_0^1 \text{MTE}(u) \times \frac{1\{u \in [p(z'), p(z)]\}}{p(z) - p(z')} du$$

1. Which part of the most right-hand side is  $E[D|Z = z] - E[D|Z = z']$  corresponding to?

*Solution.* The denominator  $p(z) - p(z')$ , which represents the change of potential treatment status. ■

2. Show that  $E[Y|Z = z] = E[Y_1|U \leq p(z)]p(z) + E[Y_0|U > p(z)](1 - p(z))$

*Solution.* Note that we implicitly assume  $U \sim \text{Uniform}[0, 1]$ , so we can derive the following result from definition:

$$\begin{aligned} & \mathbb{E}(Y | Z = z) \\ &= \mathbb{E}(Y | D = 1, Z = z) \cdot \mathbb{P}(D = 1 | Z = z) + \mathbb{E}(Y | D = 0, Z = z) \cdot \mathbb{P}(D = 0 | Z = z) \\ &= \mathbb{E}(Y | D = 1, Z = z) \cdot \mathbb{P}(U \leq p(z) | Z = z) + \mathbb{E}(Y | D = 0, Z = z) \cdot \mathbb{P}(U \geq p(z) | Z = z) \\ &= \mathbb{E}(Y_1 | U \leq p(z)) \cdot p(z) + \mathbb{E}(Y_0 | U \geq p(z)) \cdot (1 - p(z)) \end{aligned}$$

3. Show that  $E[Y|Z = z] = \int_0^{p(z)} E[Y_1|U = u]du + \int_{p(z)}^1 E[Y_0|U = u]du$

*Solution.* Given the preceding result, we further derive that:

$$\begin{aligned} & \mathbb{E}(Y_1 | U \leq p(z)) \cdot p(z) + \mathbb{E}(Y_0 | U \geq p(z)) \cdot (1 - p(z)) \\ &= \int_0^{p(z)} \mathbb{E}[Y_1 | U = u] du + \int_{p(z)}^1 \mathbb{E}[Y_0 | U = u] du \end{aligned}$$

where the former term capture the average  $Y_1$  for those with  $u \in [0, p(z)]$ ; while the latter term capture the average  $Y_0$  for those with  $u \in [p(z), 1]$  ■

4. Show that  $E[Y|Z = z] - E[Y|Z = z'] = \int_{p(z')}^{p(z)} \text{MTE}(u)du$

*Solution.* Given the preceding result, we can rearrange and obtain that

$$\begin{aligned}
& \mathbb{E}[Y|Z = z] - \mathbb{E}[Y|Z = z'] \\
&= \left( \int_0^{p(z)} E[Y_1|U = u]du + \int_{p(z)}^1 E[Y_0|U = u]du \right) \\
&\quad - \left( \int_0^{p(z')} E[Y_1|U = u]du + \int_{p(z')}^1 E[Y_0|U = u]du \right) \\
&= \int_{p(z')}^{p(z)} E[Y_1|U = u]du - \int_{p(z')}^{p(z)} E[Y_0|U = u]du \\
&= \int_{p(z')}^{p(z)} MTE(u)du
\end{aligned}$$

■

### 3 Policy Relevance Treatment Effect

We introduced LATE in class, but the “ideal” treatment effect depends on the research question. Let’s take attending college for example.

1. Let  $D \in \{0, 1\}$  be attending college or not and let the outcome  $Y$  being the future average earning. What is the ATE measuring?

*Solution.* The ATE measures the average effect of attending college on the sample, where the effect on each individual is:

$$\mathbb{E}[Y_{i1} - Y_{i0}]$$

in words, ATE captures the effect of attending college on future average earnings. ■

2. What is the ATT measuring?

*Solution.* For those choosing to attend the college, the average effect is summarized as ATT, that is:

$$\mathbb{E}[Y_{i1} - Y_{i0} | D_i = 1]$$

in words, ATE captures the effect of attending college on future average earnings only for those attending college. ■

3. Let  $Z$  be the tuition,  $Z^*$  be the tuition under the new policy,  $p^*$  be the propensity score under new policy, and  $D^* = 1\{U \leq p^*(Z^*)\}$ . Therefore  $Y^* = D^*Y_1 + (1 - D^*)Y_0$ . Define

$$\beta_{\text{PRTE}} = \frac{E(Y^*) - E(Y)}{E(D^*) - E(D)}.$$

Write down an argument why  $\beta_{\text{PRTE}}$  is more interesting than ATE or ATT.

*Solution.* Compared with ATE and ATT, where the former one measures the effect on all sample while the latter focuses on the effect in treatment group,  $\beta_{PTE}$  concerns more on those affected by the policy changes, which sheds the light for the policymakers since it reflects on the welfare changes resulting from the policy changes. ■

4. When will  $\beta_{PTE}$  equal to LATE?

*Solution.* If we assume those changes their treatment status (in this case, change their mind and attend the college) are all compliers, then  $\beta_{PTE}$  reflects on the LATE on those people. ■

## 4 Arellano-Bond Estimator

Consider the following model

$$Y_{it} = \rho Y_{it-1} + \delta_i + \varepsilon_{it}$$

Assume  $\text{Cov}(\varepsilon_{it}, Y_{is}) = 0$  for  $0 \leq s < t - 1$  (sequential exogeneity).

1. Show that the fixed effect estimator cannot recover  $\rho$  consistently.

*Solution.* In the given model, adding individual fixed effect  $\delta_i$  couldn't remedy the endogeneity problem, which arise from the correlation between  $Y_{it-1}$  and  $\delta_i$ ; as a consequence,  $\rho$  can not be estimated consistently. ■

2. Take the first difference to difference out the individual fixed effect.

*Solution.* After taking first difference, we can obtain

$$\underbrace{Y_{it} - Y_{it-1}}_{\Delta Y_{it}} = \rho \underbrace{(Y_{it-1} - Y_{it-2})}_{\Delta Y_{it-1}} + \underbrace{\varepsilon_{it} - \varepsilon_{it-1}}_{\Delta \varepsilon_{it}}$$

3. With the first difference, can OLS recover  $\rho$ ? If not, can you propose an instrument?

*Solution.* Although we did first difference, OLS might still not estimate  $\rho$  consistently due to the correlation between  $\Delta Y_{it-1}$  and  $\Delta \varepsilon_{it}$ , that is,  $\text{Cov}(Y_{it-1}, \varepsilon_{it-1}) \neq 0$

To solve the problem, we can utilize Arellano-Bond estimator, which use  $Y_{it-2}$  as the instrumental variable for  $Y_{it-1}$ . Recall the sequential exogeneity assumption,

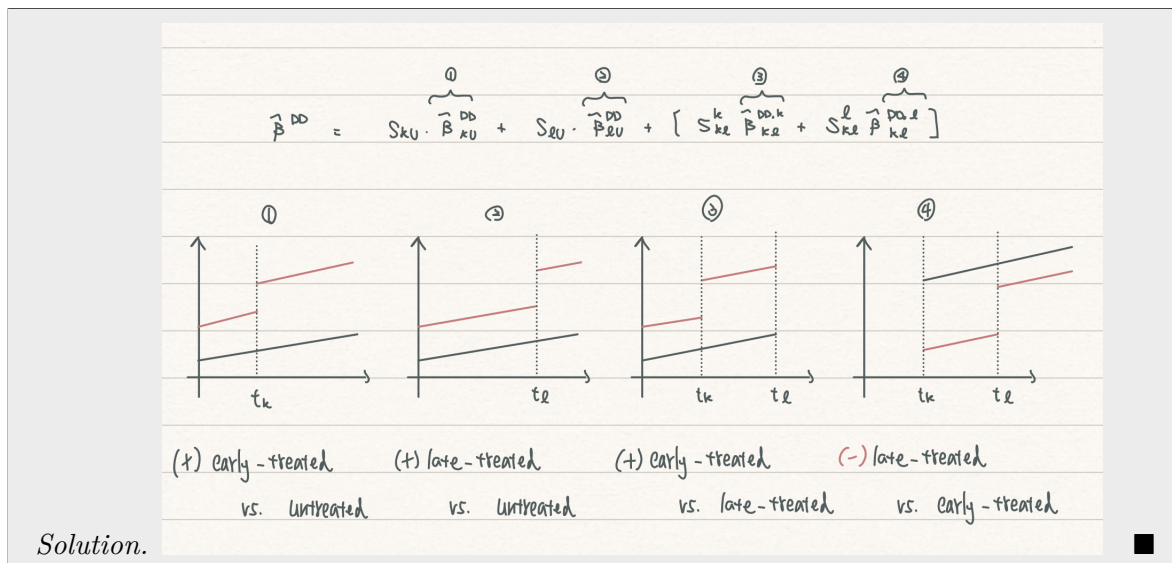
(a) *Relevance:*  $\text{Cov}(\Delta Y_{it-1}, Y_{it-2}) \neq 0$ , which is obvious.

(b) *Exogeneity:*  $Y_{it-2} \perp \Delta \varepsilon_{it}$  since  $Y_{it-2} \perp \{\varepsilon_{it-1}, \varepsilon_{it-2}\}$

hence, by two-stage GMM, we can estimate  $\rho$ . ■

## 5 Show that TWFE is Biased

1. Draw the figures in class to explain why TWFE cannot recover a positively weighted average of cohort specific ATT.



2. Simulate a data to illustrate your point above.

*Solution.* Follow the procedure of Pedro H. C. Sant'Anna and Brantly Callaway (2021), we use the code below to build data:

```

1  # Load required libraries
2  library(tidyverse)
3  library(lfe)
4  library(fastDummies)
5  library(ggthemes)
6  library(did)
7  theme_set(theme_clean() + theme(plot.background = element_blank()))
8
9  # random seed
10 iseed = 20201221
11 nrep <- 100
12 true_mu <- 1
13 set.seed(iseed)
14
15 # Data generating process
16 make_data <- function(nobs = 1000, nstates = 40) {
17
18   # unit fixed effects (unobserved heterogeneity)
19   unit <- tibble(
20     unit = 1:nobs,
21     state = sample(1:nstates, nobs, replace = TRUE),
22     unit_fe = rnorm(nobs, state/5, 1),
23     mu = true_mu
24   )
25

```

```

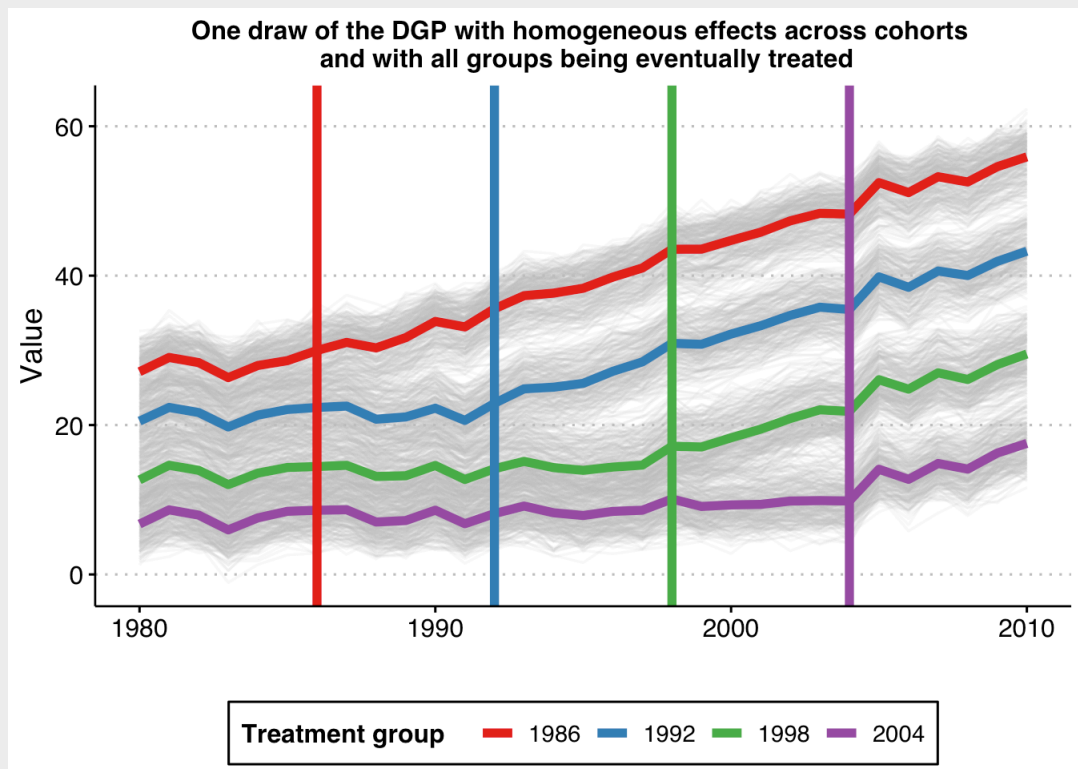
26 # year fixed effects (first part)
27 year <- tibble(
28   year = 1980:2010,
29   year_fe = rnorm(length(year), 0, 1)
30 )
31
32 # Put the states into treatment groups
33 treat_taus <- tibble(
34   state = sample(1:nstates, nstates, replace = FALSE),
35   cohort_year = sort(rep(c(1986, 1992, 1998, 2004), 10))
36 )
37
38 expand_grid(unit = 1:nobs, year = 1980:2010) %>%
39   left_join(., unit) %>%
40   left_join(., year) %>%
41   left_join(., treat_taus) %>%
42   mutate(error = rnorm(nobs*31, 0, 1),
43          treat = ifelse(year >= cohort_year, 1, 0),
44          tau = ifelse(treat == 1, mu, 0),
45          year_fe = year_fe + 0.1*(year - cohort_year)
46   ) %>%
47   group_by(unit) %>%
48   mutate(tau_cum = cumsum(tau)) %>%
49   ungroup() %>%
50   mutate(dep_var = (2010 - cohort_year) + unit_fe + year_fe + tau_cum +
51   error)
52 }
53
54 # Generate the data
55 data <- make_data()
56
57 # Draw the graph
58 plot1 <- data %>%
59   ggplot(aes(x = year, y = dep_var, group = unit)) +
60   geom_line(alpha = 1/8, color = "grey") +
61   geom_line(data = data %>%
62     group_by(cohort_year, year) %>%
63     summarize(dep_var = mean(dep_var)),
64     aes(x = year, y = dep_var, group = factor(cohort_year),
65        color = factor(cohort_year)),
66     size = 2) +
67   labs(x = "", y = "Value", color = "Treatment_group") +
68   geom_vline(xintercept = 1986, color = '#E41A1C', size = 2) +
69   geom_vline(xintercept = 1992, color = '#377EB8', size = 2) +
70   geom_vline(xintercept = 1998, color = '#4DAF4A', size = 2) +
71   geom_vline(xintercept = 2004, color = '#984EA3', size = 2) +
72   scale_color_brewer(palette = 'Set1') +
73   theme(legend.position = 'bottom',
74         axis.title = element_text(size = 14),
75         axis.text = element_text(size = 12)) +
76   ggtitle("One draw of the DGP with homogeneous effects across cohorts\nand with all groups being eventually treated") +
77   theme(plot.title = element_text(hjust = 0.5, size = 12))
78
79 # plot

```

79 plot1

80

and the data looks like



Next, we utilize two-way fixed effect model to estimate

```
1 # variables we will use
2 keepvars <- c("`rel_year_-5`", "`rel_year_-4`", "`rel_year_-3`", "`rel_year_-2`",
3               "rel_year_0", "rel_year_1", "rel_year_2", "rel_year_3", "rel_year_4", "rel_year_5")
4
5 run_ES_DiD <- function(...) {
6
7   # resimulate the data
8   data <- make_data()
9
10  # make dummy columns
11  data <- data %>%
12    # make dummies
13    mutate(rel_year = year - cohort_year) %>%
14    dummy_cols(select_columns = "rel_year") %>%
15    # generate pre and post dummies
16    mutate(Pre = ifelse(rel_year < -5, 1, 0),
17           Post = ifelse(rel_year > 5, 1, 0))
18
19  # estimate the model
```

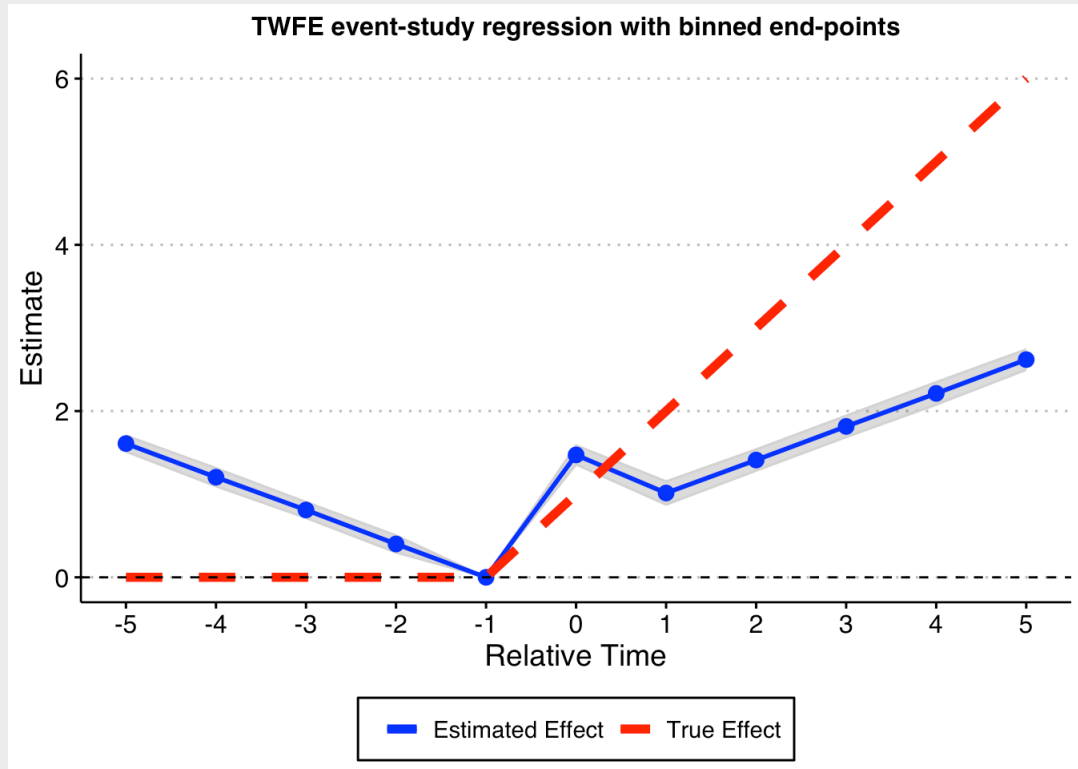
```

20   mod <- lfe::felm(dep_var ~ Pre + `rel_year_-5` + `rel_year_-4` + `
rel_year_-3` + `rel_year_-2` +
21       `rel_year_0` + `rel_year_1` + `rel_year_2` + `rel_year_3`
+ `rel_year_4` +
22       `rel_year_5` + Post | unit + year | 0 | state, data = data
, exactDOF = TRUE)
23
24   # grab the obs we need
25   mod2 <- tibble(
26     estimate = mod$coefficients,
27     term1 = rownames(mod$coefficients)
28   )
29
30   es <-
31   mod2 %>%
32     filter(term1 %in% keepvars) %>%
33     mutate(t = c(-5:-2, 0:5)) %>%
34     select(t, estimate)
35
36   es
37
38   # Run ES DID multiple times
39   data_classical <- map_dfr(1:nrep, run_ES_DiD)
40
41   # Draw the graph
42   colors <- c("True Effect" = "red", "Estimated Effect" = "blue")
43
44   ES_plot_classical <- data_classical %>%
45     group_by(t) %>%
46     summarize(avg = mean(estimate),
47               sd = sd(estimate),
48               lower.ci = avg - 1.96*sd,
49               upper.ci = avg + 1.96*sd) %>%
50     bind_rows(tibble(t = -1, avg = 0, sd = 0, lower.ci = 0, upper.ci = 0)) %>%
51     mutate(true_tau = ifelse(t >= 0, (t + 1)* true_mu, 0)) %>%
52     ggplot(aes(x = t, y = avg)) +
53     geom_ribbon(aes(ymin = lower.ci, ymax = upper.ci), color = "lightgrey",
alpha = 0.2) +
54     geom_point(color = 'blue', size = 3) +
55     geom_line(aes(color = 'Estimated Effect'), size = 1) +
56     geom_line(aes(x = t, y = true_tau, color = 'True Effect'), linetype = "
dashed", size = 2) +
57     geom_hline(yintercept = 0, linetype = "dashed") +
58     scale_x_continuous(breaks = -5:5) +
59     labs(x = "Relative Time", y = "Estimate") +
60     theme(axis.title = element_text(size = 14),
61           axis.text = element_text(size = 12)) +
62     ggtitle("TWFE event-study regression with binned end-points")+
63     scale_color_manual(values = colors) +
64     theme(plot.title = element_text(hjust = 0.5, size = 12),
65           legend.position = "bottom",
66           legend.title = element_blank())
67
68   ES_plot_classical
69

```



where the estimation result is Biased.



- Look up the paper “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature.” Learn to use one of the packages listed in the paper’s appendix. Show that you can recover one of the treatment effect from your simulation.

*Solution.* We use untreated group as control group and do the estimation again

```

1  # function to run ES DID
2  run_CS <- function(...) {
3
4  # resimulate the data
5  data <- make_data()
6
7  mod <- did::att_gt(ymame = "dep_var",
8                    tname = "year",
9                    idname = "unit",
10                   gname = "cohort_year",
11                   control_group= "notyettreated",
12                   bstrap = FALSE,
13                   data = data,
14                   print_details = FALSE)
15  event_std <- did::aggte(mod, type = "dynamic")
16
17  att.egt <- event_std$att.egt

```

```

18 names(att.egt) <- event_std$egt
19
20 # grab the obs we need
21 broom::tidy(att.egt) %>%
22   filter(names %in% -5:5) %>%
23   mutate(t = -5:5, estimate = x) %>%
24   select(t, estimate)
25 }
26
27 data_CS <- map_dfr(1:nrep, run_CS)
28
29 ES_plot_CS <- data_CS %>%
30   group_by(t) %>%
31   summarize(avg = mean(estimate),
32             sd = sd(estimate),
33             lower.ci = avg - 1.96*sd,
34             upper.ci = avg + 1.96*sd) %>%
35   mutate(true_tau = ifelse(t >= 0, (t + 1)* true_mu, 0)) %>%
36   ggplot(aes(x = t, y = avg)) +
37   #geom_linerange(aes(ymin = lower.ci, ymax = upper.ci), color = 'darkgrey',
38     size = 2) +
39   geom_ribbon(aes(ymin = lower.ci, ymax = upper.ci), color = "lightgrey",
40     alpha = 0.2) +
41   geom_point(color = 'blue', size = 3) +
42   geom_line(aes(color = 'Estimated Effect'), size = 1) +
43   geom_line(aes(x = t, y = true_tau, color = 'True Effect'), linetype = "
44     dashed", size = 2) +
45   geom_hline(yintercept = 0, linetype = "dashed") +
46   scale_x_continuous(breaks = -5:5) +
47   labs(x = "Relative Time", y = "Estimate") +
48   theme(axis.title = element_text(size = 14),
49     axis.text = element_text(size = 12)) +
50   ggtitle("Event-study-parameters estimated using Callaway and Sant'Anna
51     (2020)\nComparison group: Not-yet-treated")+
52   scale_color_manual(values = colors) +
53   theme(plot.title = element_text(hjust = 0.5, size=12),
54     legend.position = "bottom",
55     legend.title = element_blank())

```

which gives us the correct estimation of treatment effect

