Supplemental Materials for “Confounding and Regression Adjustment in Difference-in-Differences Studies”

# Appendix A: Adjusting for Time-Varying Covariates

In this section of the appendix, we discuss of the problems of adjusting for time-varying confounders as described in Section 3.2 in the main paper. A time-varying covariate that is *affected* by treatment and *affects* the outcome makes recovering the causal effect difficult. On one hand, failing to adjust for the time-varying covariate will result in failures of parallel trends. On the other hand, adjusting for the time-varying covariate, since it is on the pathway between treatment and the outcome, will adjust away some of the effect of treatment on the outcome, resulting in biased estimates.

We begin with notation that should be familiar to those who read our paper. is the continuous outcome measured at time . For simplicity, we assume thatwhere is the pre-treatment period and the post-treatment period. Treatment is binary and represented by . Finally, we have a time-varying covariate where in an index for a unit (e.g., a state or an individual). Let be the covariate group-time mean. We also introduce counterfactual notation for the covariate so that is the (possibly counterfactual) value of for individual and time under treatment . Since we assume that treatment directly affects , we may have that .

Let's extend the notation for the covariate means to counterfactual world so that and . We assume that treatment (which occurs between times 0 and 1) does not affect past versions of so that . We also assume that the covariate evolves differently in the two groups even absent treatment, leading to the failure of parallel trends. That is, .

Suppose we have the same model for untreated outcomes as the main text:

For simplicity, let . We can connect the untreated outcomes to the treated outcomes with a fixed treatment effect, : .

Recall that the average treatment effect on the treated (ATT) is

Now, we have:

and

Plugging into the ATT:

The ATT is what we want to calculate, but what is our estimate for an unadjusted model versus one from a regression model that correctly adjusts for .

**Unadjusted Estimator:**

Without significant restrictions on the and values, this does not equal the ATT.

**Adjusted Estimator:**

Now, imagine we know which regression model to fit. In R, we can fit the model

lm(y ~ a\*t + x\*t), which is correctly specified. The estimate of the treatment effect will be the coefficient on the interaction between a (treatment indicator) and t (time). However, when we fit the model, we will get:

which is biased for the true ATT.

# Appendix B: Calculation of ATT for Simulation Scenario 6

In the main paper, we state that the average treatment effect on the treated (ATT) in Scenario 6 is different than in the other scenarios. Here, we show our calculations for the ATT using our data-generating example. Below is the code used to generate data, using the dplyr R package.

dat <- expand.grid(id = 1:n, tp = 1:max.time) %>% arrange(id,tp) %>% group\_by(id) %>%

mutate(int=rnorm(1,0,sd=0.25), # random intercept

p.trt=0.5, # probability of treatment

trt=rbinom(1, 1, p.trt), # treatment

x=rnorm(1, mean = 1.5 - 0.5\*trt, sd = 1.5 - 0.5\*trt),

post=I(tp >= trt.time), # indicator of post-treatment period

treated=I(post == 1 & trt == 1), # time-varying indicator if treated or not

x=ifelse(tp>=2, lag(x, 1) + (tp-1)/10 \*

rnorm(1, mean = 1, sd = 0.1) - I(trt == 1) \* I(tp>6)\*(tp)/20, x)

) %>%

ungroup()

dat <- dat %>% mutate(err=rnorm(n\*max.time),

y = 1 + x + trt + int + err + treated + ((tp - 2.5)^2)/10,

y.t = 1 + x \* tp / 10 + trt +

int + err + treated + ((tp - 2.5)^2)/10) %>%

group\_by(id) %>% mutate(y.diff = y - lag(y), y.diff2 = y.t - lag(y.t)) %>% ungroup()

To begin, we *only* need to look at the treated group since the ATT is defined on the treated population. The setup is relatively simple. We set to be the total number of units followed over 10 max.time time points. Units were assigned to the treatment group with probability 0.5. The treated units were given treatment beginning at ; thus, we had five pre-treatment time points and five post-treatment time points. The covariate at baseline was drawn from a Normal distribution, from the treated population. During the pre-treatment period, the means of the covariate increased by about cumulatively from . However, the mean of the covariate was affected by treatment too, so that for the treated group when the mean went down by an average of per time point.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Time** |  |  |  |  |  |  |  |  |  |  |
| Mean() | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 |
| Mean() | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.45 | 1.5 | 1.55 | 1.6 | 1.65 |

Table: Evolution of counterfactual means of covariate X for the treated group.

Note that for this simulation scenario, we have two different outcomes. In the first, denoted y, the effect of on the outcome is the same at every time point. For the second outcome, denoted

y.t, the covariate has a time-varying effect on the outcome. The two outcome processes are detailed below:

So this difference is that in the second equation, interacts with time. Note that both and are mean zero normal random variables and whenever . (We are only

considering the treated group. This would not be true for the comparison group.) Like we did for the mean of , we can build a table for the means of using the above equations.

For y, we get the following results:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Time** |  |  |  |  |  |  |  |  |  |  | Avg pre | Avg post |
| Mean() | 3.225 | 3.125 | 3.225 | 3.525 | 4.025 | 4.725 | 5.625 | 6.725 | 8.025 | 9.525 | 3.425 | 6.925 |
| Mean() | 3.225 | 3.125 | 3.225 | 3.525 | 4.025 | 5.675 | 6.525 | 7.575 | 8.825 | 10.275 | 3.425 | 7.775 |

Table: Evolution of counterfactual means of outcome $Y$ for the treated group.

We'll calculate a few of these by hand to give an idea of what we're doing. Take the mean of at :

Here, we plugged in 1.6 for since it equals the untreated mean of the covariate (see Table A1). Both and are independent mean zero random variables so we plug in 0.

Following similar calculations, the mean of at is:

The ATT here is , which is calculated by taking the mean of the last 5 columns (the post-treatment time points) for each row and subtracting them.

And for y.t, we get the following results:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Time** |  |  |  |  |  |  |  |  |  |  | Avg pre | Avg post |
| Mean() | 2.325 | 2.245 | 2.385 | 2.745 | 3.325 | 4.125 | 5.145 | 6.385 | 7.845 | 9.525 | 2.605 | 6.605 |
| Mean() | 2.325 | 2.245 | 2.385 | 2.745 | 3.325 | 5.095 | 6.075 | 7.265 | 8.665 | 10.275 | 2.605 | 7.475 |

Table: Evolution of counterfactual means of outcome Y for the treated group.

The ATT here equals 0.87.

# Appendix C: Data-Generating Distributions for Simulations

|  |  |
| --- | --- |
| Scenario | Data-Generating Model |
| 1: Time-invariant covariate effect |  |
| 2: Time-varying covariate effect |  |
| 3: Treatment-independent covariate |  |

**Table C1**: Data-generating models for simulations with a time-invariant covariate, . : outcome for subject at time . : group indicator. : random intercept. : indicator of post-treatment time point. The treatment assignment for all scenarios is and the unit-level intercepts are all . Lastly, , , , and .

|  |  |
| --- | --- |
| Scenario | Data-Generating Model |
| 4: Parallel evolution |  |
| 5: Evolution differs by group |  |
| 6: Evolution diverges in post-treatment period |  |

**Table C2**: Data-generating models for simulations with a time-invariant covariate, . : outcome for subject at time . : group indicator. : random intercept. : indicator of post-treatment time point. The treatment assignment for all scenarios is . The unit-level intercepts are all . The covariate value at the first time point is . And . The functions that govern the evolution of the covariate are , , and . Lastly, the outcome process (a) refers to those where the covariate has a time-invariant effect on the outcome and uses the functions and . The outcome process (b) refers to the simulations where the covariate has a time-varying effect on the outcome and uses the functions and .