# Synthetic Difference-in-Differences with Time-Varying Covariates

Sebastian Kranz, Ulm University

2021-12-16

**Abstract**

Arkhangelsky et al. (2021) introduce the novel synthetic difference-in-difference (SDID) method. They briefly note a method to extend SDID for exogenous time-varying covariates. This paper illustrates shortfalls of the proposed method in a simple setting. We propose a simple alternative way to explicitly control for covariates and implement it in the R package `xsynthdid`. This paper illustrates the issue and shows in several Monte-Carlo simulations how the proposed method allows more precise estimation of causal effects.

# 1 Introduction

Arkhangelsky et al. (2021) introduce the novel synthetic difference-in-differences (SDID) method that is implemented in the R package synthdid. They briefly note how the approach could be extended to explicitly account for time-varying covariates (in Footnote 4):

*Some applications feature time-varying exogenous covariates $X_{it} \in \mathbb{R}^p$. We can incorporate adjustment for these covariates by applying SDID to the residuals*

$$Y_{it}^{res} = Y_{it} - X_{it}\hat{\beta}$$

*of the regression of $Y_{it}$ on $X_{it}$.*

In Section 2 we illustrate a structural problem of this approach in a simple setting. We propose a simple alternative approach, that slightly varies the procedure above. First, we estimate a fixed effects regression:

$$Y_{it} = X_{it}\beta + \gamma_t + \mu_i + u_i$$

where besides the covariates $X_{it}$, we also include period fixed effects $\gamma_t$ and unit fixed effects $\mu_i$. Importantly, this regression is only estimated on a subsample that omits any observation in which treatment takes place.

We then compute for every observation (also those in which treatment takes place) the following adjusted outcome:

$$Y_{it}^{adj} = Y_{it} - X_{it}\hat{\beta}$$

where $\hat{\beta}$ are the estimated coefficients from above's fixed effects regression. We then use these adjusted outcomes as input for the SDID method.

In Section 3 we compare the precision of our proposed estimator with alternative approaches using Monte-Carlo simulations for scenarios in which the SDID could be consistent even without explicitly controlling for the covariates. We show that our approach still generally performs better in terms of a lower RMSE of the estimated causal effect.

The adjustment approach is implemented in the function `adjust.outcome.for.x` in the R package xsynthdid.

Currently this paper is a mixture of a regular paper and a vignette for that package. In the future, I may write a separate package vignette and rewrite this paper in a more traditional fashion. Also note that this paper currently has no proven results about characteristics of the estimators like consistency, only intuitions and Monte-Carlo simulations.

# 2 Illustration with a first example

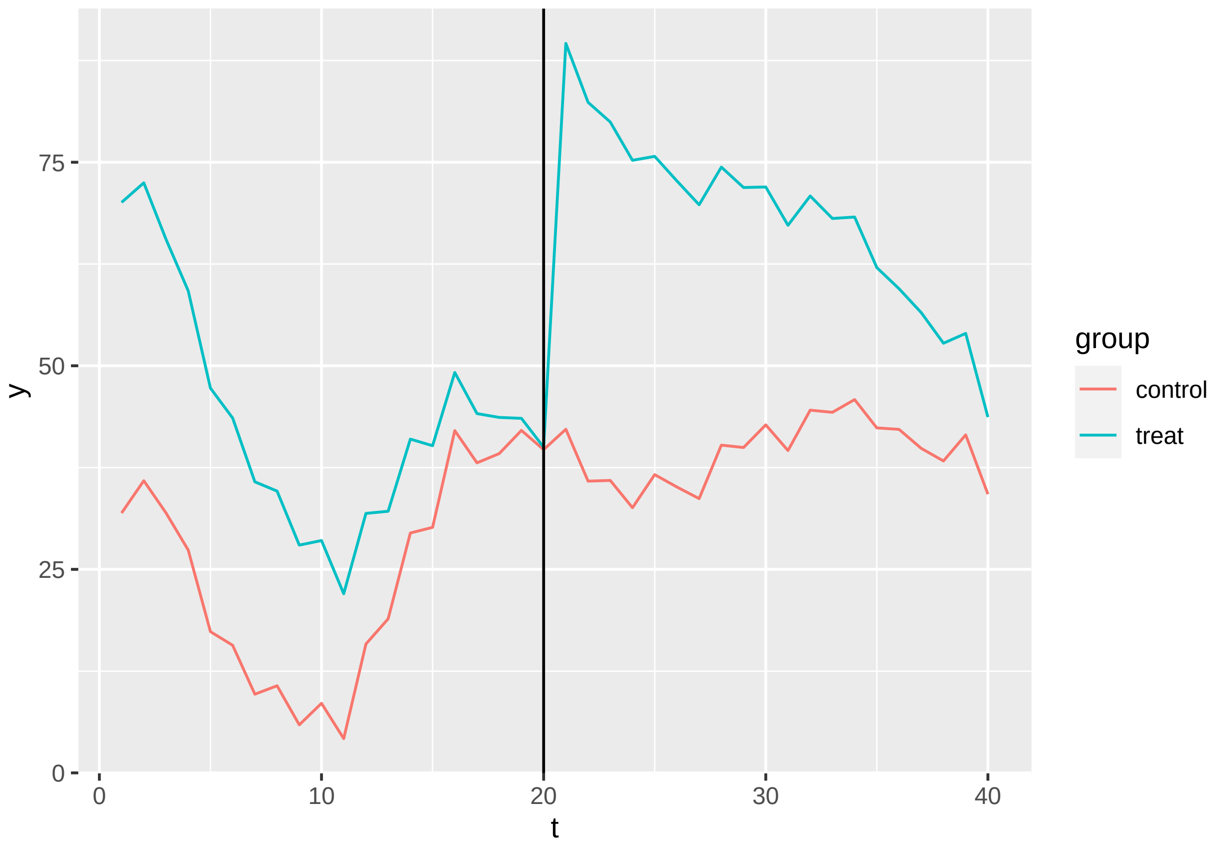
## 2.1 A simulated data set

We first simulate a panel data set that we will use for later estimations.

```r
library(dplyr)
set.seed(1)
N = 20 # no of subjects
T = 40 # no of periods
tau = 50
beta.x = 1

dat = expand.grid(i = 1:N, t = 1:T)

# Simulate a common AR(1) time trend
# will be used to create time varying x
# that affects treatment and control group differently
time.trend = as.numeric(arima.sim(n=T,list(ar = c(0.4,0.5), ma=c(0.6,0.5))))*5

dat = mutate(dat,
  group = ifelse(i > N/2,"treat","control"),
  treat = 1L*(group == "treat"),
  exp = 1L*(t > T/2),
  treat_exp = exp*treat,
  mu.t = time.trend[t],
  eps = rnorm(n()),
  x = ifelse(treat,-t, t)+runif(n())*3,
  y = mu.t + treat*40 + treat_exp*tau + beta.x*x + eps
) %>% arrange(treat_exp, treat, exp, i, t)
```

The `N` individuals are equally distributed among a treatment and control group, and the treatment starts in period `T/2=20`. The treatment increases the average value of the dependent variable `y` by `tau = 50` in the treatment group. We plot the average `y` values of the treatment and control group:

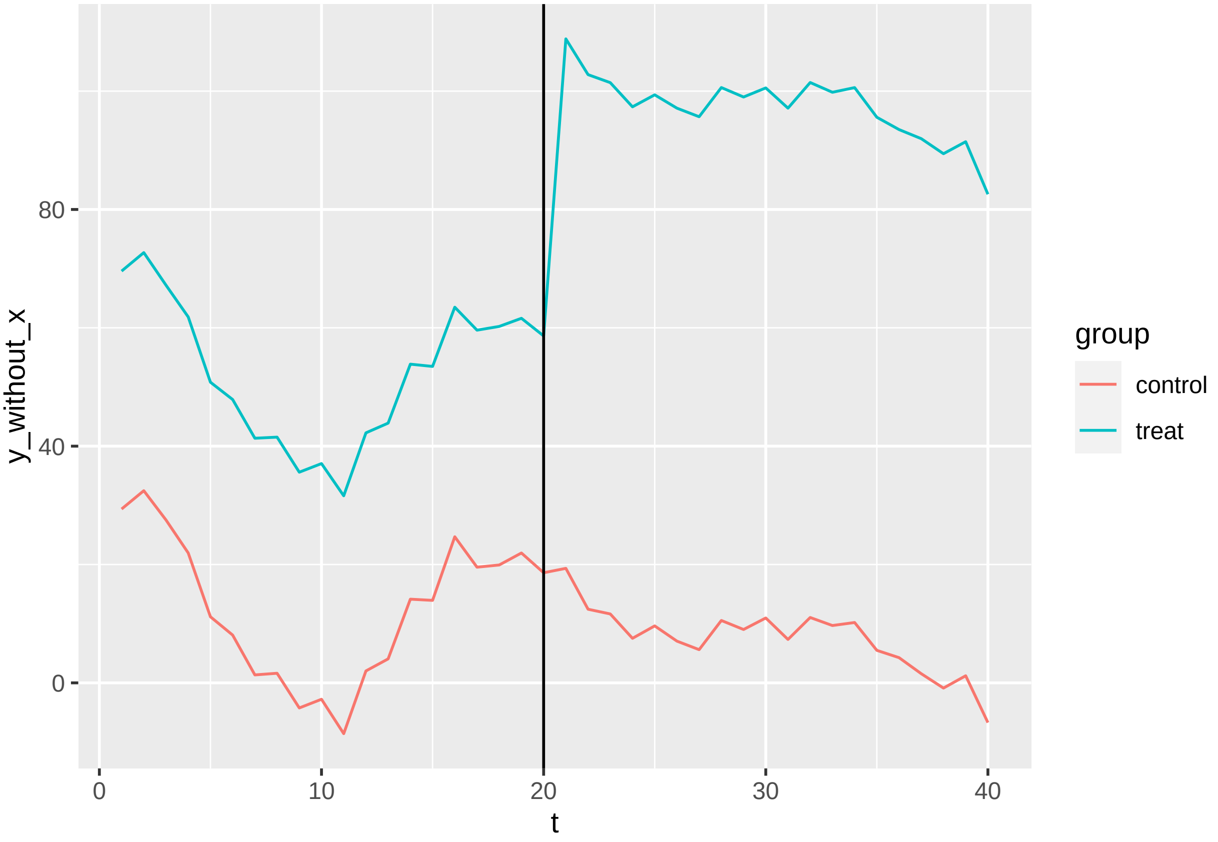
```r
library(ggplot2)
pdat = dat %>% group_by(group,t) %>% summarize(y=mean(y))
ggplot(pdat, aes(y=y,x=t, color=group)) +
  geom_line() + geom_vline(xintercept = T/2)
```



The parallel trends assumption for difference-in-differences (DID) estimation is not satisfied, as can be seen by the non-parallel pretrends. The reason is that the outcome `y` also depends on a time varying variable `x` that follows a different path in the control and treatment group.

If we remove the effect `beta.x*x` of `x` on `y`, we see nicely parallel pre-trends:

```
gdat = dat %>%
  group_by(group,t) %>%
  summarize(y_without_x=mean(y-beta.x*x))
ggplot(gdat,aes(y=y_without_x,x=t, color=group)) +
  geom_line() + geom_vline(xintercept = T/2)
```



For the correction, we used the true effect `beta.x=1` of `x` on `y`, but we would not know `beta.x` in a real world data set.

## 2.2 Illustrating that proposed method of Arkhangelsky et al. (2021) to control for x does not work well

As already noted, Arkhangelsky et al. (2021) propose to dealing with covariates like `x` in their Footnote 4 as following:

*Some applications feature time-varying exogenous covariates $X_{it} \in \mathbb{R}^p$. We can incorporate adjustment for these covariates by applying SDID to the residuals*

$$Y_{it}^{res} = Y_{it} - X_{it}\hat{\beta}$$

*of the regression of $Y_{it}$ on $X_{it}$.*

Let us run the proposed regression:

```
reg = lm(y~x, data=dat)
broom::tidy(reg)
```

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | 44.643 | 0.579 | 77.1 | 0 |
| x | -0.425 | 0.025 | -17.3 | 0 |

Unfortunately, this regression does not yield a consistent estimator for `beta.x=1`, because `x` is correlated with the time period and group. Using the residuals of that regression does thus not correctly remove the effect of x on y. Correspondingly, the resulting DID and SDID estimators don't perform well in our example. Let's start with the DID estimator.

```
dat$y.res = resid(reg)
broom::tidy(lm(y.res ~ treat_exp + exp + treat, data=dat))
```

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -14.86 | 0.884 | -16.81 | 0 |
| treat_exp | -6.96 | 1.768 | -3.94 | 0 |
| exp | 23.21 | 1.250 | 18.56 | 0 |
| treat | 10.00 | 1.250 | 8.00 | 0 |

We estimate a treatment effect of `-6.9` while the true effect is `tau = 50`.

Let us now run the SDID estimator using Arkhangelsky et al. (2021)'s adjustment:

```
library(synthdid)
pm = panel.matrices(dat, unit="i",time = "t",outcome = "y.res",treatment = "treat_exp")
synthdid_estimate(Y=pm$Y,N0=pm$N0,T0 = pm$T0) # inconsistent
```

```
## synthdid: -5.868 +- 0.465. Effective N0/N0 = 9.9/10~1.0. Effective T0/T0 = 4.7/20~0.2. N1,T1 = 10,20
```

The estimated treatment effect `-5.868` of the SDID estimate is similarly far away the true effect `tau = 50`.

Not adjusting at all for x does also not yield a good estimate (even though it moves to the right direction):

```
# SDID without adjustment for x
pm = panel.matrices(dat, unit="i",time = "t",outcome = "y",treatment = "treat_exp")
synthdid_estimate(Y=pm$Y,N0=pm$N0,T0 = pm$T0) # inconsistent
```

```
## synthdid: 10.774 +- 0.363. Effective N0/N0 = 10.0/10~1.0. Effective T0/T0 = 4.9/20~0.2. N1,T1 = 10,2
```

Probably none of the estimators above is consistent for this DGP.

Yet, note that in other settings the SDID estimator could correct for confounding factors like x even without explicitly adjusting for them. Our simulation above is just structured in a way that makes it hard for SDID to work without explicitly dealing with x. In Section 3, we look at alternative DGPs that give SDID without controlling for the covariates better chance.

## 2.3 Our proposal to control for x

Our proposal to account for x is simple. We first estimate the effect x on y by adding time and unit fixed effects. We only use observations in which no treatment takes place to avoid confounding:

```
reg.adj = lm(y~x+as.factor(t) + as.factor(i),
             data=dat %>% filter(treat_exp==0))
beta.x.hat = coef(reg.adj)[2]
beta.x.hat
```

```
## x
## 1
```

Looks like a pretty consistent estimator for `beta.x=1`. We don't use the residuals of this regression to adjust y but subtract the effect of x on y using this estimate for `beta.x`.

```
dat$y.adj = dat$y - beta.x.hat * dat$x
```

Our function `adjust.outcome.for.x` in the package `xsynthdid` computes the same adjusted y. It is only a bit more general and works out of the box for multiple x and for x that are categorical variables. It is also faster, since it uses the `feols` function from the packages fixest for the fixed effects regression.

```r
library(xsynthdid)
dat$y.adj = adjust.outcome.for.x(dat,unit="i",time = "t",
  outcome = "y",treatment = "treat_exp", x="x")
```

We can now use this adjusted dependent variable in a DID or SDID regression.

```r
# DID regression with adjusted `y`
broom::tidy(lm(y.adj ~ treat_exp + treat + exp, data=dat))
```

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | 12.79 | 0.655 | 19.51 | 0 |
| treat_exp | 50.27 | 1.311 | 38.35 | 0 |
| treat | 39.97 | 0.927 | 43.12 | 0 |
| exp | -5.58 | 0.927 | -6.02 | 0 |

And here we run a SDID regression:

```r
pm = panel.matrices(dat, unit="i",time = "t",
  outcome = "y.adj",treatment = "treat_exp")
synthdid_estimate(Y=pm$Y,N0=pm$N0,T0 = pm$T0)
```

```
## synthdid: 50.188 +- 0.418. Effective N0/N0 = 10.0/10~1.0. Effective T0/T0 = 6.1/20~0.3. N1,T1 = 10,20
```

In both cases, it looks like we consistently estimate the treatment effect `tau=50`.

## 3. A wider range of Monte-Carlo studies

In the simulated DGP above, the covariate variable `x` has exactly the opposite sign for all control group members than for the treatment group members. That is an unrealistically extreme case that gives SDID little chance for a good estimate unless `x` is explicitly controlled for.

We now present the results of a broader range of Monte-Carlo studies where SDID would in principle be consistent even without controlling for `x`.

### Basic DGP

The first data generating processes is as follows:

$$y = \tau \cdot treat_i \cdot exp_t + \beta_{treat} \cdot treat_i + \beta_{tt} \cdot tt_t + beta_x \cdot x_{t,i} + u_{i,t}$$

where $tt_t$ is a stochastic time trend following an AR(1) process. The covariate $x_{it}$ satisfies

$$x_{it} = \phi_i \cdot tt_t^x$$

where $tt^x$ is another AR(1) time trend and $\phi_i \in [0,1]$ is a factor loading for unit $i$. We assume that in the treatment group the factor load for all subjects is independently drawn from a Beta-distribution with mean of 3/4 ($\alpha = 1.5$ and $\beta = 0.5$) while in the control group factor loadings are drawn from a Beta-distribution with mean 1/4 ($\alpha = 0.5$ and $\beta = 1.5$). So while the average factor loading differs between between control and treatment group, one should typically be able to construct a synthetic control group with the same average factor loading as the treatment group.

The parameter $\tau$ and all $\beta$ have a value of 50, both AR(1) processes have an AR(1) coefficient of 0.97. Finally, $u_{i_t}$ is just a idiosyncratic error term that is i.i.d. standard normally distributed.

The simulation code is specified in the function `xsdid.mc` in the xsynthdid package. We consider three different sample varying the number of periods and units from 20 to 200 and simulate 1000 samples for each specification. The following table shows the computed RMSE of the estimator of the causal effect `tau` for different estimation methods.

**RMSE**

| N | T | sdidx | didx | sdid | sdid.res | did | sc |
|---|---|---|---|---|---|---|---|
| 20 | 20 | 0.285 | 0.217 | 26.7 | 12.59 | 61.3 | 39.3 |
| 50 | 50 | 0.119 | 0.089 | 23.8 | 13.53 | 78.9 | 44.1 |
| 200 | 200 | 0.027 | 0.020 | 14.7 | 6.74 | 89.6 | 49.6 |

The columns `sdidx` shows the RMSE of the SDID estimator using our adjustment of the outcome to control for `x` and `didx` is the simple DID fixed effects estimator controlling for `x`. We see that both have a substantially lower RMSE than all other methods.

The worst performance has a simple fixed effects DID specification without controlling for x (column `did`). Using SDID without controlling for `x` (column `sdid`) substantially improves over the simple DID. The column `sdid.res` shows the results using adjustment for `x` proposed in Footnote 4 of Arkhangelsky et al. (2021). This adjustment improves over the SDID estimator without adjustment, but the RMSE still by a substantial factor larger than under our proposed adjustment. The final column `sc` shows the results of a synthetic control approach, which has quite poor performance.

### Variation 1: Correlated x and basic time trend

Our first variation assumes that the underlying time trend for `x` is not independent from all other variables anymore, but is now the sum of the basic time trend `tt` and an independent AR(1) process. Let's look at the resulting RMSE:

**RMSE**

| N | T | sdidx | didx | sdid | sdid.res | did | sc |
|---|---|---|---|---|---|---|---|
| 20 | 20 | 0.298 | 0.225 | 34.1 | 19.93 | 82.3 | 42.2 |
| 50 | 50 | 0.120 | 0.087 | 34.8 | 20.18 | 112.4 | 48.8 |
| 200 | 200 | 0.028 | 0.022 | 15.2 | 9.94 | 121.6 | 51.2 |

Not too much changes. Qualitatively the insights are the same as before.

### Variation 2: Correlated x and basic time trend

We now add an unobserved confounder `xc` that has a similar structure than `x` but cannot be added as control variable. The factor loadings are independently drawn to that of `x`. `xc` affect `y` also with a parameter `beta.xc=50`.

**RMSE**

| N | T | sdidx | didx | sdid | sdid.res | did | sc |
|---|---|---|---|---|---|---|---|
| 20 | 20 | 32.1 | 61.4 | 49.1 | 33.7 | 87 | 48.9 |
| 50 | 50 | 26.7 | 75.7 | 44.6 | 30.7 | 109 | 48.0 |
| 200 | 200 | 14.2 | 86.1 | 23.7 | 15.8 | 120 | 49.0 |

Now, generally all estimators perform substantially worse. The DID estimator that controls for `x` (`didx`) performs worse than all SDID methods.

Still our proposed SDID estimator (`sdidx`) performs best. The original adjustment proposed by Arkhangelsky et al. (2021) (`sdid.res`) follows closely behind, however.

## Concluding Remarks

We provided intuitive arguments for our proposed method to control for covariates in the SDID setting and confirmed more precise results with simulation studies. However, there is clearly scope for a comprehensive theoretical characterization.

Another point not touched in this paper is how to correctly compute standard errors so that the uncertainty from estimating the adjusted outcome is taken into account.

## References

- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic Difference-in-Differences. American Economic Review, 111(12), 4088-4118.