PUBPOL 6090 Problem Set 3

Thelonious Goerz

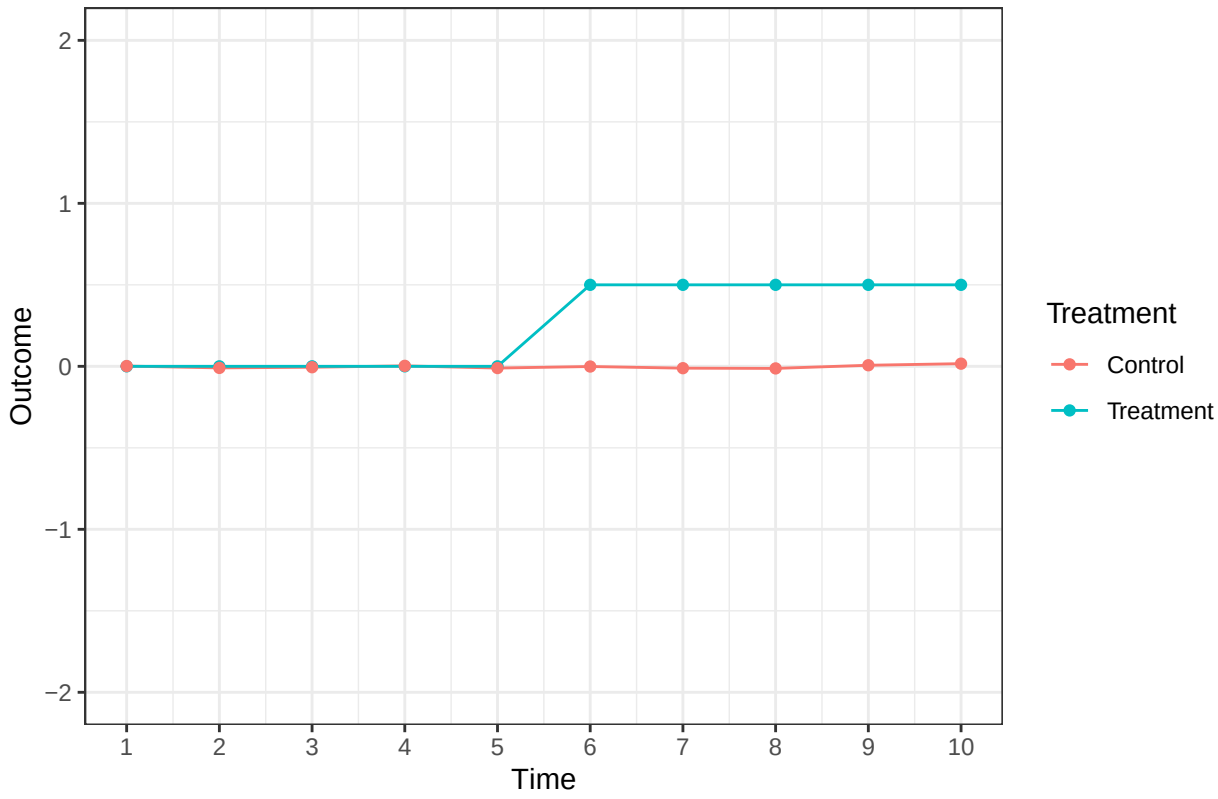
2023-10-31

Problem 1

1.1

In the first figure, I plot the raw data points for over time for treatment and control.

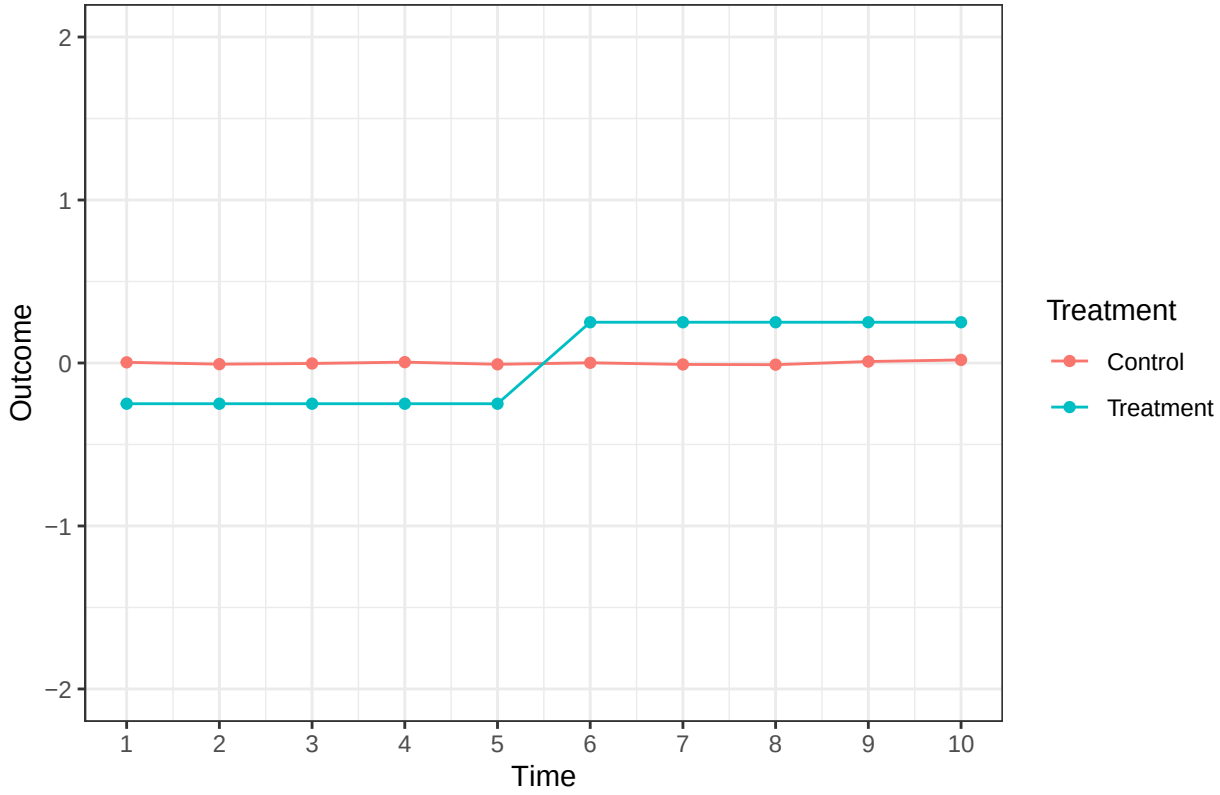
Fig. 1 : Treatment and Control Means of Simulated DID Data



1.2.1

In figure two, I plot the data over time and include a unit fixed effect (treatment status).

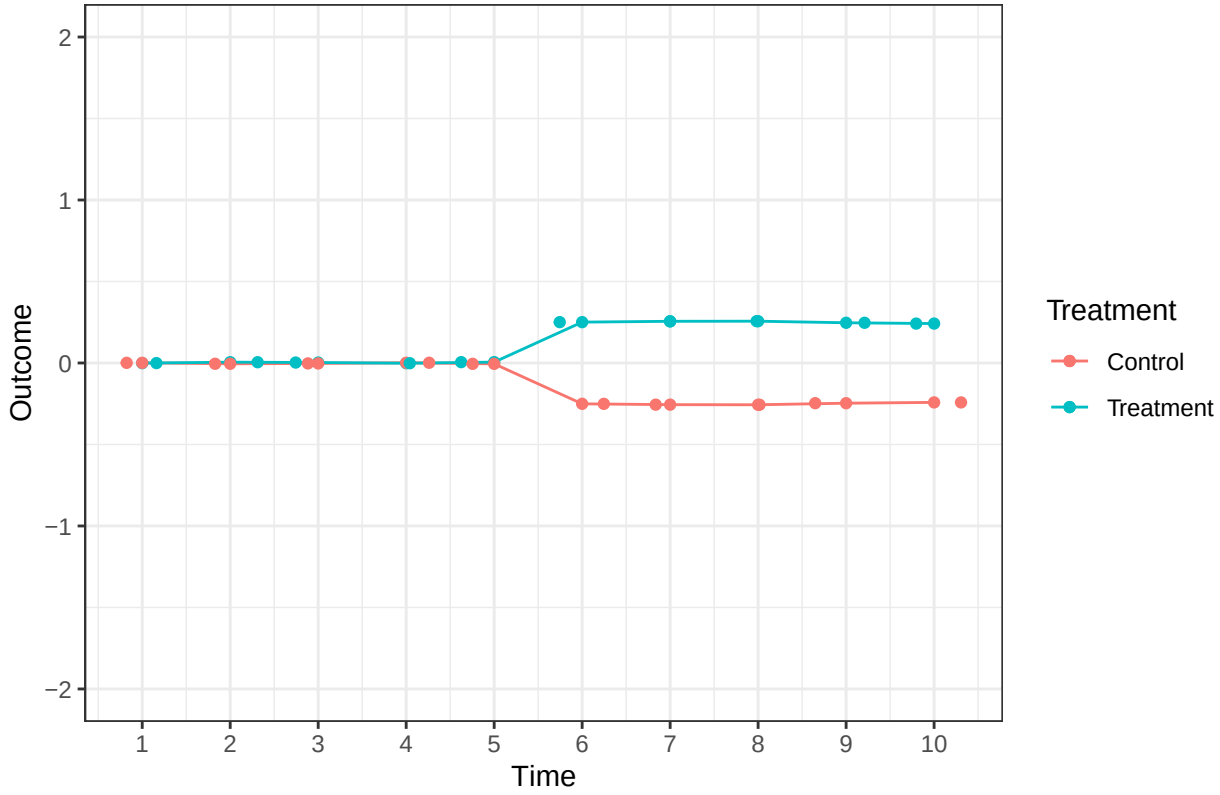
Fig. 2 :Treatment and Control Means of Simulated DID Data: Unit FE



1.2.2

In figure three, I add show the trends over time with a time fixed effect.

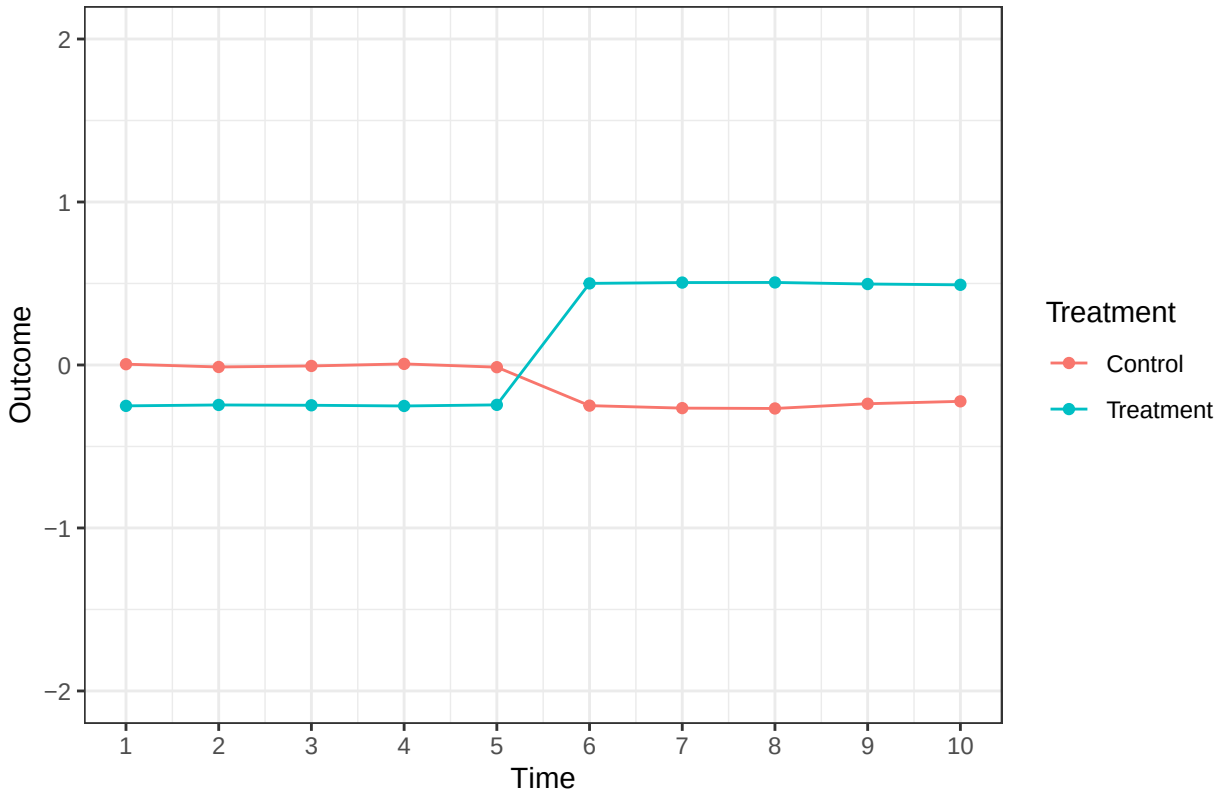
Fig. 3 :Treatment and Control Means of Simulated DID Data:Time FE



1.2.3

In figure four, I show the trends over time by treatment status with both time and unit fixed effects, which shows the variation that we use to identify out DID estimate.

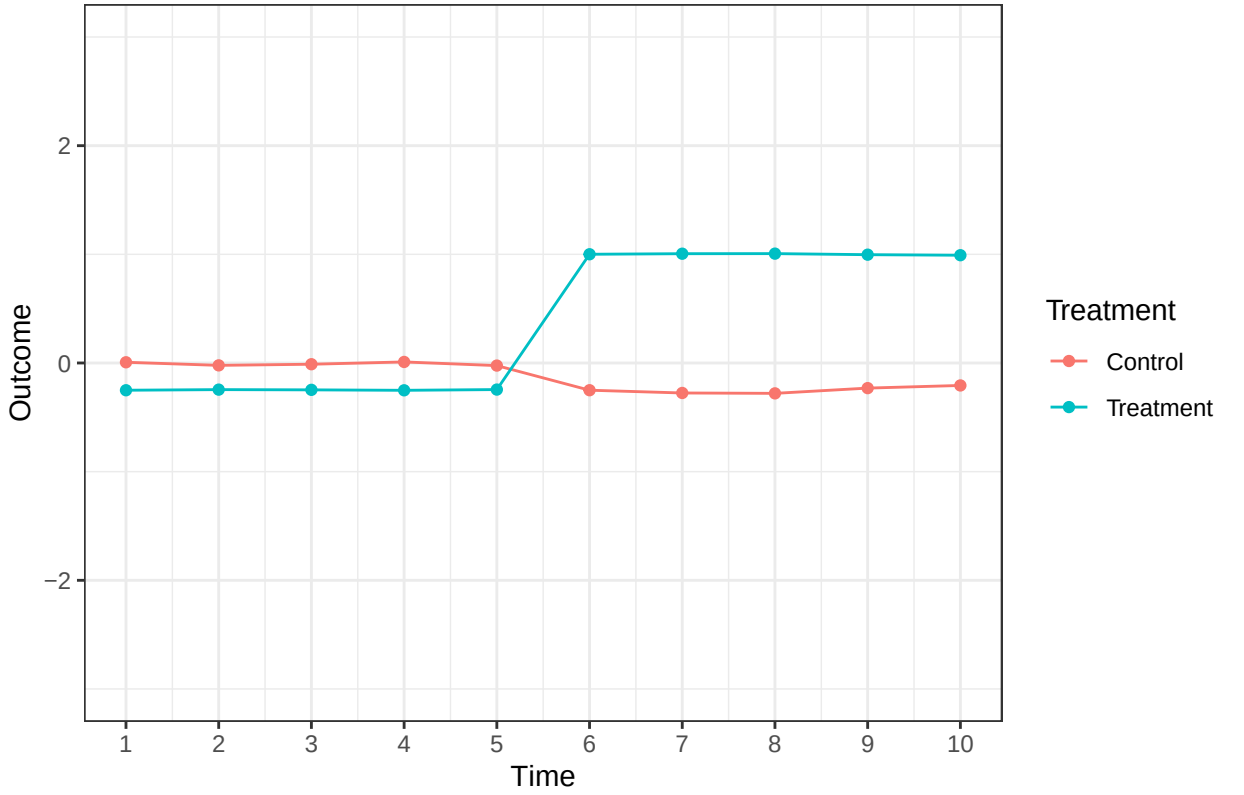
Fig. 4 :Treatment and Control Means of Simulated DID Data: Unit and Time



1.2.4

In figure five, I display the same plot as above while also adding unit-specific time trends. We can see that adding the unit-specific time trend increases the amount of variation from treatment cases that we are using for our DID estimate. Additionally, the slope parameter for the post-treatment time period is double that of our standard two-way fixed effects estimate.

Fig. 5 :Treatment and Control Means of Simulated DID Data: Unit and Time



2.3

In graph 4, we see that the treatment and control units are contributing symmetrically to the treatment effect estimate pre and post treatment. After controlling for panel fixed effects, the treatment and control units are contributing about equal variation.

Problem 2

a.

If we set our null hypothesis reject rate to an $\alpha = .05$ then we should expect that if we were to simulate the sampling process many times, five percent of the time we should be rejecting the null hypothesis as opposed to failing to reject it. If our estimate has have an incorrect rejection rate, then we end up discarding results that have actually statistically meaningful effects. Table 8 in Bertrand, Duflo, and Mullinathan suggests that DID estimates reject the null hypothesis anywhere from 11-50% with the CPS data. This indicates that a large number of effects that were significant assuming correct inference at the .05 level are actually vastly inflated compared to the true result. In the models with the AR(1) distribution, hypothesis reject rates are more sensible, with the second of the two models under-rejecting .035 the null hypothesis. The main takeaway is that the DID estimates may be rejecting the null hypothesis of no effect much more frequently than it should, suggesting that DID point estimates may be unreliable.

b.

In comparison to Table 8 in BDM, I replicate the simulated rejection rates net of some rounding error. Per their punch line, the cluster bootstrap adjusts for the SEs such that the rejection rate corresponds to the true rejection rate of .05, however performance is less strong with few clusters.

Of note, the OLS simulations with 20 and 6 clusters diverge by about 10% from their counterparts in BDM.

Table 1: Replication of BDM Table 8

N States	Rejection Rate	SE	Cluster
50.00	0.51	0.50	OLS
50.00	0.07	0.25	Cluster
20.00	0.48	0.50	OLS
20.00	0.05	0.22	Cluster
10.00	0.52	0.50	OLS
10.00	0.09	0.29	Cluster
6.00	0.52	0.50	OLS
6.00	0.11	0.31	Cluster

Asymptotic rejection rate results from 1000 monte carlo simulations of difference in difference models with varying N.

Table 2: Replication of BDM Table 8 (Adjusted T-stat)

N States	Rejection Rate	SE	Cluster
6.00	0.53	0.50	OLS
6.00	0.16	0.37	Cluster
10.00	0.53	0.50	OLS
10.00	0.11	0.31	Cluster

Asymptotic rejection rate results from 1000 monte carlo simulations of difference in difference models with varying N.

This could be due to the much larger simulation size or it could suggest that rejection rates are actually understated in BDM. Nevertheless, clustering the SEs overcomes these issues and performs at the levels presented in Table 8 of BDM.

c.

In comparison to non-adjusted standard errors. The correction for the t-distribution with $N = 1$ degrees of freedom makes no difference in the estimates.

d.

Table 3: Replication of BDM Table 5 (Cluster Bootstrapped SEs)

N States	Rejection Rate	Cluster
6.00	0.56	OLS
6.00	0.06	Cluster
10.00	0.52	OLS
10.00	0.08	Cluster

Asymptotic rejection rate results from 200 monte carlo simulations of difference in difference models with varying N.

Table 4: Replication of CGM (Wild Cluster Bootstrap)

N States	Rejection Rate	SE
50.00	0.03	0.16
20.00	0.04	0.21
10.00	0.04	0.20
6.00	0.07	0.26

Asymptotic rejection rate results from 200 monte carlo simulations of difference in difference models with varying N.

Table 5: Replication of Table 1 Dehejia and Wahaba

Treatment		Age	Education	Black	Hispanic	No degree	Married	RE 74	RE 75
Control	N	260	260	260	260	260	260	260	260
	Mean	25.05	10.09	0.83	0.11	0.83	0.15	2107.03	1266.91
Treatment	N	185	185	185	185	185	185	185	185
	Mean	25.82	10.35	0.84	0.06	0.71	0.19	2095.57	1532.06

In comparison to Table five in BDM, the cluster bootstrapped SEs are much more accurate than they report. For both 10 and 6 clusters, there is a significant improvement on rejection rates. They report .225 and .435 for 10 and 6 respectively, where as I report .08 and .06 which are both significant improvements on accuracy.

Note: I had to reduce the number of replications for the last two examples due to computational burden.

e.

In comparison to CGM, I replicate the main results of their null hypothesis rejection rate in Table 5, which suggests that the wild cluster bootstrap overcomes the issues of overrejction noted in BDM. The only small divergence in my estimates is the rejection rate for 20 clusters which is actually smaller than CGM.

Problem 3

a.

Overall the means of the key variables are very close to those in Table 1 od Dehejia and Wahaba despite slight discrepancies in the decimal points. For the 74 earnings there is a slight difference in the treatment and control earnings which drops sharply in 75. Since both earnings for treatment and control follow a similar downward pattern pre-treatment this suggests an Ashenfelter's dip scenario.

b.

Table 6: Replication of Table 2 Row 1 Panel B and C (Dehejia and Wahaba)

	B					C				
	All Covariates	Quasi DID		Treatment - Control		All Covariates	Quasi DID		Treatment - Control	
	.	Adjusted	Unadjusted	Adjusted	Unadjusted	.	Adjusted	Unadjusted	Adjusted	Unadjusted
Point Estimate	1560.95	1631.41	1750.15	1672.43	1794.34	1601.31	1672.20	1750.15	1688.36	1794.34

1. The all covariates model, Panel B and C, includes all covariates in addition to an education squared term and the re74.

Table 7: Replication of Table 1 Dehejia and Wahaba (PSID)

Treatment		Age	Education	Black	Hispanic	No degree	Married	RE 74	RE 75
Control	N	2490	2490	2490	2490	2490	2490	2490	2490
	Mean	34.85	12.12	0.25	0.03	0.31	0.87	19 428.75	19 063.34
Treatment	N	184	184	184	184	184	184	184	184
	Mean	25.76	10.34	0.84	0.06	0.71	0.18	2106.96	1540.38

Overall, I am able to replicate the coefficients, however I was not able to replicate unadjusted coefficient for Quasi-DID for Panel C, despite using the core variables.

Per the problem set, I was unable to replicate column 5 exactly, however I am able to get a very close point estimate by controlling for education squared which is described in the table note. Overall, my point estimates match closely with table 2.

c.

In the following table, I present descriptive statistics for the PSID sample

The replicated means of the variables in the PSID control group match closely to the results that Dehejia and Wahaba report, net of some rounding error.

d.

Table 8: PSID-1 Replication of Table 2 Row 1 Panel B and C

	B					C				
	All Covariates	Quasi DID		Treatment - Control		All Covariates	Quasi DID		Treatment - Control	
	.	Adjusted	Unadjusted	Adjusted	Unadjusted	.	Adjusted	Unadjusted	Adjusted	Unadjusted
Point Estimate	1453.37	−289.80	−607.96	−7731.90	−15 224.24	909.63	189.82	−607.96	−904.94	−15 224.24

1. The all covariates model, Panel B and C, includes all covariates in addition to an education squared term and the re74.

For regression results, the results are very close to those reported in Dehejia and Wahaba, however I am not able to replicate them exactly, despite using their specifications. I tried alternate specifications, but was ultimately unable to recover exact point estimates. Nevertheless, I replicate estimates with the same sign and poitn estimate as the paper net of some error.

e.

In the PSID-1 and NSW samples, the NSW sample suggests that the growth in income for treatment is positive and significant while the PSID-1 sample suggests the complete opposite, flipping the point estimate negative. It think that this control sample points to the sensitivity of the estimates to sample size and potential selection that resulted from that specific treatment sample. The PSID sample is significantly larger than the entire sample for the NSW experiment. Furthermore, while the treatment and control groups were randomly assigned in the experiment, the PSID were not. It is possible that the discrepancy in the treatment effects could be due to the observational nature of the data which underscores Lalonde's original point.