Replication 6 - Difference-in-Differences

Marina Merlo

22 de maio de 2019

1. What is treatment and control in this study? What is the treatment assignment mechanism?

The treatment is to have their District People's Council (DCP) abolished after 2009, the control is to still have it. The treatment assignment mechanism was designed by the Ministry Of Home Affairs through a stratified sample by region, sub-region and other characteristics.

2. Run the 'naive' cross-sectional OLS regression of the infrastructure index (one of the 6 presented in Table 3 of Malesky et al) on treatment. How do you interpret the results? Provide at least one specific reason why the treatment effect in your regression may be a biased estimate.

The results show there's no significant impact of the treatment on the infrastructure index. However, this is a biased estimate because we're comparing treated units across time, even when they weren't treated yet.

Table 1:

	Dependent variable:
	index_infra
treatment	−0.053
	(0.049)
Constant	3.205***
	(0.018)
Observations	4,129
R^2	0.0003
Adjusted R^2	0.00004
Residual Std. Error	1.059 (df = 4127)
F Statistic	1.173 (df = 1; 4127)
Note:	*p<0.1; **p<0.05; ***p<0.01

.

3. Run the 'naive' before-after OLS regression of the infrastructure index on the time variable(1 for 2010, 0 for 2008) for the treated units only. How do you interpret the results? Provide at least one specific reason why the treatment effect in your regression may be a biased estimate.

There's a positive and significant effect of 0.49 on the infrastructure index after the removal of the district council. This might be biased due other temporal changes and not only the treatment, such as an improvment on economy etc.

.

Table 2:

	Dependent variable:
	index_infra
time	0.490***
	(0.093)
Constant	2.898***
	(0.067)
Observations	528
R^2	0.050
Adjusted R^2	0.048
Residual Std. Error	1.070 (df = 526)
F Statistic	27.678*** (df = 1; 526)
Note:	*p<0.1; **p<0.05; ***p<0.01

4. Now perform the main Difference-in-differences analysis for the Infrastructure Index out-come. Don't cluster your standard errors or include any control variables yet. Interpret the results.

Our average treatment effect is the coefficient of the interaction term between time and treatment. This coefficient is positive and significant at 5% level, showing there's an increase of 0.25 of the infrastructure index after the removal of district councils on treated communes.

Table 3:

	Dependent variable:
	index_infra
time	0.241***
	(0.035)
treatment	−0.188***
	(0.070)
time:treatment	0.249**
	(0.098)
Constant	3.086***
	(0.025)
Observations	4,129
R^2	0.018
Adjusted R^2	0.018
Residual Std. Error	1.049 (df = 4125)
F Statistic	25.859*** (df = 3; 4125)
Note:	*p<0.1; **p<0.05; ***p<0.01

.

5. Repeat Q4 but now add the control variables (lnarea,lnpopden,city, and Region fixed effects)used in Table 3 of Malesky et al. Compare your answers to those in Table 3.

There're small differences from the results in Table 3 in the paper and the ones below, but the interpretation remains the same. Our ATE is 0.227 and in Malesky et al., 0.225, both significant at 5% level. (see Table 4 below)

.

6. Repeat Q5 but now with clustered standard errors at the District level. How does this alter your results?

There's an incrase in our standard errors and our ATE now is significant at 10%. (see Table 5 below)

Table 4:

	Dependent variable:
	index_infra
time	0.230***
	(0.034)
treatment	−0.240***
	(0.069)
lnarea	0.185***
	(0.038)
lnpopden	0.299***
	(0.030)
city	0.111*
	(0.065)
Region	0.018**
	(0.007)
time:treatment	0.227**
	(0.095)
Constant	1.119***
	(0.228)
Observations	4,126
R^2	0.069
Adjusted R^2	0.068
Residual Std. Error	1.022 (df = 4118)
F Statistic	43.692*** (df = 7; 4118)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5:

	Dependent variable:
	index_infra
time	0.230***
	(0.053)
treatment	−0.240
	(0.150)
lnarea	0.185***
	(0.063)
lnpopden	0.299***
	(0.055)
city	0.111
	(0.131)
Region	0.018
	(0.019)
time:treatment	0.227*
	(0.129)
Constant	1.119***
	(0.387)
Observations	4,126
R^2	0.069
Adjusted R^2	0.068
Residual Std. Error	1.022 (df = 4118)
F Statistic	43.692*** (df = 7; 4118)
Note:	*p<0.1; **p<0.05; ***p<0.01

7. Replicate all of the columns of Table 3 of Malesky et al.

.

Table 6:

	index_infra	index_agric	index_health	index_education	index_comms	index_bus_dev
	Dependent variable:					
	(1)	(2)	(3)	(4)	(5)	(6)
time	0.230***	−0.149	−0.013	0.080***	−0.053**	−0.017
	(0.053)	(0.506)	(0.017)	(0.027)	(0.023)	(0.032)
treatment	−0.240	−1.722*	−0.040*	0.024	−0.112	0.016
	(0.150)	(1.043)	(0.022)	(0.099)	(0.108)	(0.186)
lnarea	0.185***	1.227*	−0.095***	0.197***	0.099**	0.399***
	(0.063)	(0.644)	(0.022)	(0.047)	(0.043)	(0.071)
lnpopden	0.299***	0.763	−0.170***	0.130***	0.223***	0.505***
	(0.055)	(0.606)	(0.022)	(0.041)	(0.044)	(0.063)
city	0.111	0.582	0.012	0.202**	0.067	0.178
	(0.131)	(3.530)	(0.019)	(0.086)	(0.086)	(0.195)
Region	0.018	0.466*	−0.009**	−0.044***	0.019**	−0.004
	(0.019)	(0.239)	(0.004)	(0.010)	(0.009)	(0.019)
time:treatment	0.227*	2.082	0.125***	0.094	0.142*	0.007
	(0.129)	(1.614)	(0.033)	(0.089)	(0.082)	(0.102)
Constant	1.119***	2.838	1.322***	0.562*	0.632*	−1.597***
	(0.387)	(4.292)	(0.160)	(0.298)	(0.334)	(0.422)
Observations	4,126	4,126	4,126	4,126	4,126	4,126
R^2	0.069	0.024	0.109	0.027	0.083	0.099
Adjusted R^2	0.068	0.022	0.107	0.026	0.081	0.097
Residual Std. Error (df = 4118)	1.022	11.055	0.393	0.857	0.682	1.030
F Statistic (df = 7; 4118)	43.692***	14.302***	71.729***	16.510***	53.162***	64.607***

Note: *$p<0.1$; **$p<0.05$; ***$p<0.01$

.

8. Assess the balance in land area (totalland) of the treated and control units in time t= 0.Focus on the substantive difference, not the p-value. Is there are any evidence of imbalance? Would this create a risk of bias for our difference-in-differences analysis?

Control units have, on average, 153.52 in land area than treated units. Larger areas might be harder to administer and require a different bureaucratic structure, so the impact of abolishing the council on efficiency can be different.

Table 7: Balance between treated and control in land area in t=0

treatment	mean	difference
1	2036.499	NA
0	2190.021	153.522

9. Difference-in-differences methodology cannot protect us against time-varying confounders.Provide an example of an omitted (confounding) variable that might create bias in our results.

The gain in the efficiency could have been driven by an increase in the economy performance, for example.

10. One way of testing for the presence of time-varying confounders is to check that there are parallel pre-treatment trends in the outcomes for treated and control units. Using the second dataset, Vietnam0608.csv, and your main difference-in-differences regression, assess if treated units had a different trend to control units before treatment, i.e. between 2006 and 2008, for each of the 6 outcome indices. This should replicate Panel 2 of Table 3 in Malesky et al.

There are some differences from this replication and the Panel 2 of Table 3 in the paper. In the paper, there seems to be no parallel pre-treatment trends for the infrastructure index (interaction term between treatment and time is significant at 1% level), while here it isn't significant. In this replication the only significant interaction term is for the agricultural index, at 10% level.

Table 8:

	index_infra	index_agric	index_health	index_education	index_comms	index_bus_dev
	Dependent variable:					
	(1)	(2)	(3)	(4)	(5)	(6)
time	0.0004	1.159***	−0.005	0.024	0.019	0.031
	(0.031)	(0.430)	(0.016)	(0.028)	(0.023)	(0.033)
treatment	−0.146	−0.437	−0.008	−0.030	−0.053	0.022
	(0.149)	(1.082)	(0.031)	(0.100)	(0.106)	(0.183)
lnarea	0.134**	1.031*	−0.106***	0.211***	0.056	0.433***
	(0.065)	(0.608)	(0.020)	(0.053)	(0.051)	(0.077)
lnpopden	0.204***	0.652	−0.181***	0.149***	0.196***	0.558***
	(0.060)	(0.579)	(0.024)	(0.047)	(0.049)	(0.070)
city	0.118	2.679	0.031	0.191	0.036	0.136
	(0.245)	(1.786)	(0.031)	(0.161)	(0.079)	(0.143)
Region	−0.003	0.363*	−0.009**	−0.054***	0.025**	−0.007
	(0.016)	(0.189)	(0.004)	(0.013)	(0.010)	(0.019)
time:treatment	−0.095	−1.659*	−0.034	0.040	−0.051	−0.018
	(0.071)	(0.953)	(0.032)	(0.052)	(0.050)	(0.076)
Constant	1.787***	3.162	1.411***	0.461	0.838**	−1.945***
	(0.441)	(4.328)	(0.171)	(0.326)	(0.374)	(0.477)
Observations	4,220	4,220	4,220	4,220	4,220	4,220
R^2	0.024	0.025	0.108	0.025	0.076	0.105
Adjusted R^2	0.023	0.024	0.107	0.024	0.075	0.104
Residual Std. Error (df = 4212)	0.989	9.767	0.400	0.850	0.703	1.053
F Statistic (df = 7; 4212)	15.007***	15.541***	73.113***	15.687***	49.572***	70.912***

Note: *p<0.1; **p<0.05; ***p<0.01

11. Create a Difference-in-differences chart showing the average Infrastructure Index outcome by treatment group between 2008 and 2010. Compare this to the same chart between 2006 and 2008. What do these charts suggest about the validity of our difference-in-differences methodology?

The charts show there's no parallel lines for both periods, meaning the estimates from the diff-in-diff might be questioned - even though the tests in Table 8 above shows they might be parallel 'enough' for the test.