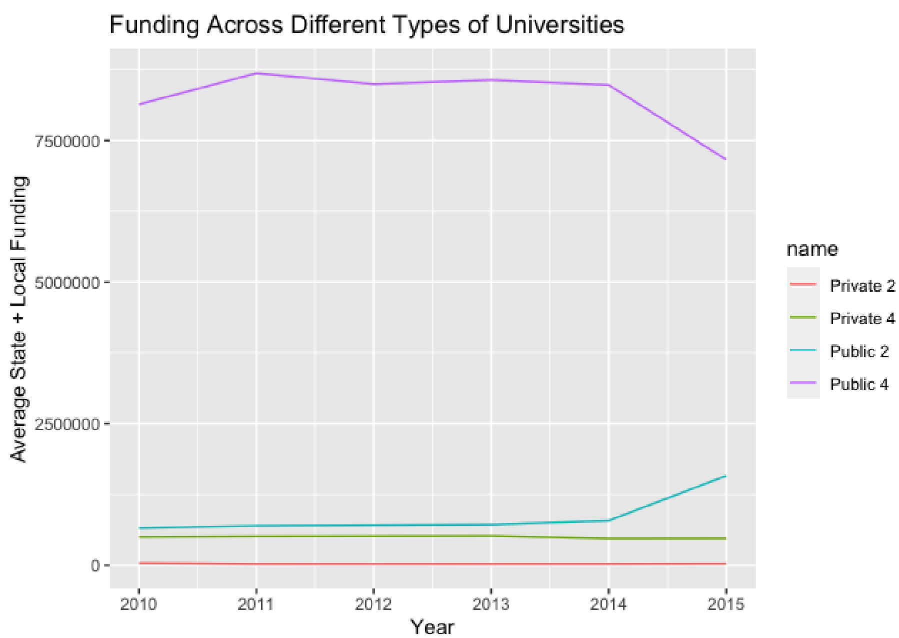
Honors Econometrics

Teddy Foley

September 12, 2023

Report

1. See attached code.
2. See attached code.



3. The graph shows relatively constant funding for Public 2-year and all Private colleges in Tennessee up until 2015, when Public 2-year colleges funding doubles. Public 4-year spending stays at around the same rate far above the other three up until 2015, when it takes a significant dip.

4. I propose using a difference in difference model for determining the causal effect of changes in funding between public and private 2-year colleges on enrollment. Since the two appear to satisfy parallel trends for funding and enrollment until the Tennessee Promise program started in 2015, we will be able to observe how the program affected enrollment. The regression described by the table below took the form $enroll_f tug\ pub2 + time + did$ where $time$ is a binary indicator for $year > 2014$ and did is the binary indicator for public schools in the year of treatment.

Table 1: Regression Results

	Dependent variable:
	enroll_ftug
pub2	385.927*** (56.986)
time	−26.733 (114.985)
did	202.312 (139.586)
Constant	204.178*** (46.942)
Observations	336
R^2	0.174
Adjusted R^2	0.167
Residual Std. Error	445.333 (df = 332)
F Statistic	23.351*** (df = 3; 332)
Note:	*p<0.1; **p<0.05; ***p<0.01

5. The parallel trends assumption we eyeballed from the graph above wasn't proven, so in order to fully recommend it to the Illinois State Board of Education, we would need to perform rigorous testing to ensure it was accurate. In addition, the p-value of the *did* coefficient was only 0.148, which is not particularly significant. As a result, we cannot say with complete confidence that the change in enrollment we are observing is the causal effect of the Tennessee program.

6. The table for the new regression is below:

	Robust SE	Clustering
(Intercept)	169.17*	169.17*
	(77.50)	(75.32)
pub2	409.10***	409.10***
	(105.13)	(102.89)
time	8.28	8.28
	(115.34)	(17.25)
did	179.14	179.14***
	(173.51)	(47.82)
R^2	0.20	0.20
Adj. R^2	0.17	0.17
Statistic	11.11	7.94
P Value	0.00	0.00
DF Resid.	108.00	55.00
nobs	112	112

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2: Statistical models

The standard error is far higher for the Robust SE model than the Clustering model. I would rather use clustering to estimate β to take into account school-specific variation.

7. Duplicated data regression results

	Robust	Clustering
(Intercept)	169.17**	169.17**
	(54.01)	(53.26)
pub2	409.10***	409.10***
	(73.53)	(72.75)
time	8.28	8.28
	(80.38)	(12.20)
did	179.14	179.14***
	(121.45)	(33.81)
R^2	0.20	0.20
Adj. R^2	0.18	0.18
Statistic	22.67	15.88
P Value	0.00	0.00
DF Resid.	220.00	111.00
nobs	224	224

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Statistical models

The ratio of the standard errors from part 6 is 173.51/47.82 = 3.628

The ratio of the standard errors from part 6 is 121.45/33.81 = 3.592

They are nearly the exact same ratio, which I am not particularly surprised by since doubling the data set should not have an impact on the standard errors.

8. Duplicate with original school IDs

	Robust	Clustering
(Intercept)	169.17^{**}	169.17^{*}
	(54.01)	(75.32)
pub2	409.10^{***}	409.10^{***}
	(73.53)	(102.89)
time	8.28	8.28
	(80.38)	(17.25)
did	179.14	179.14^{***}
	(121.45)	(47.82)
R^2	0.20	0.20
Adj. R^2	0.18	0.18
Statistic	22.67	7.94
P Value	0.00	0.00
DF Resid.	220.00	55.00
nobs	224	224

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 4: Statistical models

Now that the schools have the original IDs, the clustering regression standard error is the same as it was in the original regression (47.82 as opposed to 33.81).

9. Table of results after 500 simulations:

Significance level	Robust # rejected	Cluster # rejected	Pt 11 # rejected
1%	9	13	12
5%	48	53	49
10%	104	111	89
Mean t-value	0.01057837	0.01058321	0.01797227

The robust model rejects the null of $\beta = 0$ less often than the clustered model does.

10. The table for first-differences:

	First-Differences
(Intercept)	8.28
	(17.75)
public	179.14^{***}
	(48.55)
R^2	0.12
Adj. R^2	0.10
Num. obs.	56
RMSE	234.45

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 5: Statistical models

This estimate of the diff-in-diff (β) is the same as the non-first differenced regression, but it has a much smaller standard error (48.55 compared to 173.51 from the robust regression in pt 6).

11. From the table in Pt.9 we can see that the first differenced regression rejects the null less often than the other two, meaning its smaller standard errors result in the randomized regression correctly confirming the null hypothesis $\beta = 0$ more often.