Blueprint Labs Data Task

Joe Fast

May 12, 2024

A. Assembling the Data

This part of the task is straightforward and documented in the "Assembling the Data" do-file in the directory. There is one important thing to note. When deciding which universities to classify as granting bachelor degrees for the variable *degree_bach*, I chose to select all institutions for which the variables *instcat* had value 2. This is all institutions that were primarily granting bachelor degrees and I decided not to include those that offered "not-primarily" bachelor degrees.

Analysis

Question 1

Year	Total number of Institutions		Average School-Level Amount		
	Bachelor Granting	Public	Enrollment	State/Local Aid	Federal Aid
2010-2011	45	48	466	852,731	1,204,320
2011-2012	45	48	446	898,389	1,069,766
2012-2013	45	48	417	890,042	992,164
2013-2014	46	48	412	898,380	982,668
2014-2015	46	48	417	895,951	990,115
2015-2016	46	48	458	1,027,001	1,065,140

Table 1: Summary Statistics: Tennessee Higher Education Institutions 2010-2016

Question 2 on the next page.

Question 2

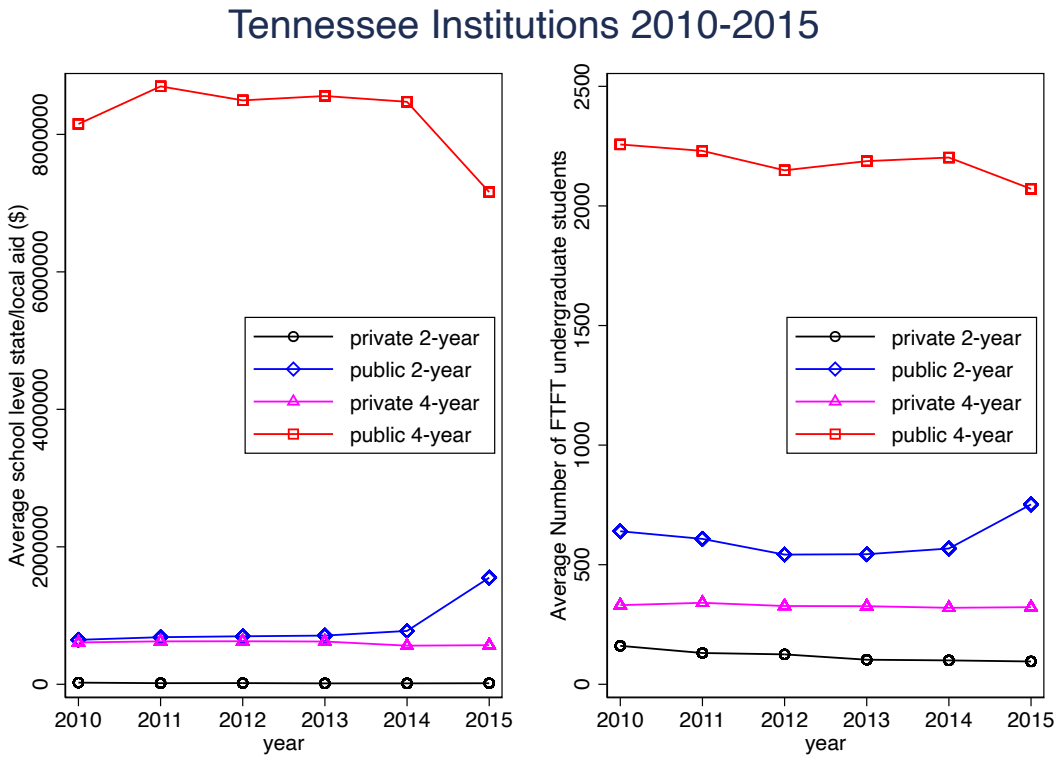

Figure 1: School-level state/local aid and first-time-full-time (FTFT) undergraduate enrollment

The findings for the first graph show that when the Tennessee Promise program was instituted there was significant increase in the average state and local aid provided at at 2-year public college. The magnitude of this increase is around $800,000 and represents about a doubling of the previous state/local aid provided.

This large increase in state/local aid for public 2-year colleges coincides with a decrease in state/local aid for public 4-year colleges. The corresponding decreases is larger in magnitude than the original increase. Public 4-year colleges are provide about $1.5 million less in state/local aid in 2015 after the implementation of the Tennessee Promise Program.

These numbers represent the average total state/local aid that the school provides. This means that any increase and decrease could be due to a change in the total amount of students at any given university and not necessarily due to any change in financial aid policy. State/Local aid for private colleges is unaffected and constant throughout the years.

Question 3

Table 2: Two-Stage Least Squares Estimation

Variable	**Log FTFT studetns**
Promise program	0.291*
	(0.114)
Federal Aid	0.000000205***
	(5.08e-08)
Large City	0
	(.)
Medium City	0.0277
	(0.102)
Small City	0.0372
	(0.0880)
Large Suburb	0.110
	(0.0858)
Small Suburb	0
	(.)
Distant Town	-0.528***
	(0.115)
Remote Town	-0.283**
	(0.0952)
Rural Fringe	-0.121
	(0.0864)
Rural Distant	-0.333*
	(0.134)
Smallest	0
	(.)
Small	0.761***
	(0.100)
Medium	1.149***
	(0.139)
Large	1.047***
	(0.200)
%FTFT of Student Body	0.0102***
	(0.00186)
_cons	5.183***
	(0.143)
N	234

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Explaining my choice

Table 2 is a 2sls estimation of the effect of the Tennessee Promise program on FTFT enrollment at 2-year public institutions. A simple OLS regression of enrollment on whether the year is 2015 is not sufficient to find a causal result. That regression would fail to account for several

unobserved factors such as changes in popularity of a school, changes in tuition, closing and opening of departments etc. Since the provided resources focus a lot on IV estimation and professor Angrist is quite famous for IV estimation I decided to primarily do a 2sls estimation. At first, my idea was to use the decrease in state aid for 4-year colleges as an instrument for the introduction of the Tennessee Promise Program. However, since 4-year colleges were not experiencing any of the treatment, I didn't pursue this idea.

Instead, I noticed from the graphs that the amount of state/local aid at 2-year public institutions is extremely correlated with the introduction of the Tennessee Promise program and it could be used as an instrument for the introduction of the Tennessee Promise program. I limit the sample to public 2-year institutions and use the total state/local aid and the average amount of state/local aid per "FTFT receiving" student provide conditional on receiving aid as instruments. Since there is such high correlation between aid and FTFT enrollment we are confident the instruments are not weak. It is quite dubious to assume that amount of state financial aid is not correlated with the error term of regression of FTFT enrollment on the promise program. However, I could not think of a better instrument. To test the exclusion assumption I will conduct an over identification test.

In addition, I estimate a Difference-in Differences (my curiosity got the best of me) for the effect of the Tennessee Promise program on enrollment. I use 2-year public institutions as the treated group and 2-year private institutions as the control group. I present the results below. (I use all the same controls however they are not all shown to save space)

Table 3: Difference-in Differences Estimation

Variable	**ln_enroll**
2015-2016#2-year public	0.181
	(0.0990)
2015-2016	-0.262**
	(0.0576)
2 year public	1.261***
	(0.118)
Feederal Aid	0.000000729***
	(5.74e-08)
N	585
Standard errors in parentheses	
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$	

Below I present the results of the First-Stage of the the 2SLS as well.

Table 4: First Stage of 2SLS

Variable	**Coefficient**
Federal Aid	-3.58e-08
	(4.58e-08)
Large City	0
	(.)
Medium City	-0.0192
	(0.0944)
Small City	-0.192*
	(0.0943)
Large Suburb	-0.393**
	(0.139)
Medium Suburb	0.0313
	(0.0999)
Small Suburb	-0.00903
	(0.0987)
Fringe Town	-0.134
	(0.0831)
Distant Town	-0.0633
	(0.105)
Remote Town	0
	(.)
Small	-0.0702
	(0.105)
Medium	-0.264*
	(0.113)
Large	-0.298
	(0.158)
% FTFT of Total	0.000769
	(0.00169)
Total State Aid	0.000000312***
	(3.38e-08)
Average State Aid for Student	0.0000608***
	(0.00000987)
_cons	-0.0614
	(0.129)
N	234

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Interpretation of the Results

My dependant variable for this is the the log enrollment of FTFT students. I use only FTFT students as they are the ones that are most affected by a change to financial aid policy since they are making their initial choice of institution. Since the distribution of FTFT enrollment is extremely right skewed, I use the log FTFT enrollment.

Therefore, the correct interpretation of the the 2SLS is that the Promise program increased FTFT enrollment at public 2 year colleges by a factor of e(0.291)=1.34. In other words, the introduction of the Promise program increased FTFT enrollment at 2-year public institutions by 34%. This result is statistically significant at a 5% level.

The correct interpretation of the the DiD is that the Promise program increased FTFT enrollment at public 2 year colleges by a factor of e(0.181)=1.20. In other words, the introduction of the Promise program increased FTFT enrollment at 2-year public institutions by 20%. This result is not statistically significant.

Question 4

There are four important assumptions of an IV estimation model that must be satisfied for the results to be unbiased:

1. Relevance: The amount of state/local aid and the average amount for FTFT student must be correlated with the Tennessee Promise program. $\text{Cov}(Z, X) \neq 0$

   This assumption is satisfied since an F test of *grant_state* and average state aid for student in the First Stage Regression gives a value of F-test of > 64. In general, an F-test > 10 is considered satisfactory.

2. Exogeneity: The amount of state/local aid and the average amount for FTFT student must not be correlated with the error term. $\text{Cov}(Z, \varepsilon) = 0$

   This assumption cannot be directly tested. However, as discussed before it is unlikely that state aid is uncorrelated with the error term of a regression of FTFT students on the promise program. We expect people more likely to enroll if the amount of aid they receive increases.

3. Exclusion Restriction: The amount of state/local aid and the average amount for FTFT student should only affect the number of FTFT enrolled students through its influence on the promise program variable. $\text{Cov}(Z, Y) = 0$

   This assumption is also not satisfied. Since our IV is over identified we can use an over-identification test to test for relevance. We use the "estat overid" command to test this. We get a p value of 0.045 for the chi square test. The null hypothesis that the exclusion restriction is satisfied, therefore we can reject that the exclusion restriction is satisfied with 95% confidence.

Since the IV model fails to satisfy two of the assumptions required for unbiased estimates I do not believe in the results of this regression. In addition to the problems of Instrumental Variables there are also problems of omitted variables in the simple regression. For example, the average income of a FTFT student, the proportion of students that have a job at the institution, the average age of FTFT students could all be examples of omitted variables that have significant effects on being part of the Promise program and the enrollment into 2-year public universities. We need more information on how the Promise program was administered to produce truly unbiased results.

Question 5

My perfect experiment on the effectiveness of no-cost tuition on college outcomes would implement the random assignment of no-cost tuition. In a perfect world, I would obtain a representative sample of thousands of high school students in the U.S. Then, I would randomly

provide a significant portion of them completely free college studies. Students would be blind to assignment and most outcomes would be collected from college databases to reduce to likelihood of interview bias. The outcomes of interest would be grades received, participation in extra curriculars, number of classes taken, happiness, type of degree pursued, after college plans, enrollment etc. In order to test this hypothesis we would need student-level data. Potential sources could be college administrative data for data on classes, grades, degree and anything directly related to educational attainment. In the same vein, MCAT, GRE, LSAT scores could be a source of such data. Perhaps learning management systems could provide data on engagement and activities. Lastly, the CPS or other government surveys might have several education questions which provide useful.