WANG Liangjiayi - ECO372 Assignment 3

Student Nb: [1004405789]

a.

i. The unit of observation in dataset is the students who apply the PACES program in Bogota 1995

ii. There are 283 observations

iii. The 'vouch0' records whether a student was allocated a voucher

iv. The 'math' records their math score, 'reading' records reading score, 'writing' records writing score, and 'totalpts' records total score

b.

```
. summarize math reading writing totalpts
```

Variable	Obs	Mean	Std. Dev.	Min	Max
math	282	.0028718	1.002821	-2.024617	3.239387
reading	283	.0006201	1.003512	-3.972432	2.097389
writing	283	.0049475	1.001712	-3.295184	2.20036
totalpts	282	.0033445	1.002749	-2.448572	2.924793

We use the tabstat() to find the sample average and standard deviation for each variable. The average value(mean) for each variable is close to 0. And the relatively large standard deviation represents that the values in the data set are farther away from the mean for each variable, on average.

c.

	(1) Total Points OLS	(2) Math Score OLS	(3) Reading Score OLS	(4) Writing Score OLS
vouch0	0.217	0.178	0.204	0.126
	(0.121)	(0.118)	(0.123)	(0.122)
Obs.	282	282	283	283
R-squared	0.0415	0.0480	0.0265	0.00918
F-statistic	4.081	4.844	2.819	0.820

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Based on our regression, we gain the similar results for each score in column 1 table 5 Panel A. The regression estimates are the same in my regression and paper. By contrast, the standard error for each score in my regression are different from paper's standard error. The standard errors in paper are corrected for within-school-of-application clustering, but it not the case in my regression.

d.

	(1) Total Points OLS results with covariates
vouch0	0.0734
	(0.136)
Obs.	189
R-squared	0.174
F-statistic	4.568

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Based on our regression, the results for total score have large difference from the results in column 2 table 5 Panel A. The regression estimates are the smaller in my regression (0.0734) compared to the estimate in paper (0.205). Also, the standard error for each score in my regression (0.136) are larger than paper's standard error (0.108). The standard errors in paper are corrected for within-school-of-application clustering, but it not the case in my regression. The estimate be so different is may because the smaller sample size in our cases.

e.

The number 0.217 represents that the students who gain the voucher are expected to have 0.217 points higher in total score compared to the students who do not gain the voucher on average. However, 0.217 is not statistically significant at 5% significance level and we do not have evidence to reject the null hypothesis that there is no difference in total scores between the students who win the voucher and who not win the voucher. Thus, the effect on wining the voucher to student's total score is not remarkable.

f.

i. The unit of observation in dataset is the students who apply the PACES program in Bogota 1995 and there are 1135 observations.
ii. The 'vouch0' records whether a student was allocated a voucher
iii. The 'usesch' records whether a student used the voucher
iv. The 'scyfnsh' records the highest grade achieved by the student, and 'inschl' whether they are still in school by the time of the survey

g.

	(1) Highest_Grade_OLS	(2) Still_in_Scl_OLS
vouch0	0.126*	0.00764
	(0.0516)	(0.0203)
Obs.	1135	1135
R-squared	0.108	0.165
F-statistic	.	.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For both estimation results, we need to compare the result in table 3 column 3. The estimates in column 3 from model that controls for the city, year of application, phone access, age, type of survey and instrument, strata of residence, and month of survey. Most of control variables are identical in our regression. Thus, the comparison between our estimate and column 3 estimate will be meaningful.

Compare the result in table 3 column 3(basic controls) and value for highest grade completed, the estimate for highest grade in our regression (0.126) is bit lower to the estimate in paper (0.13). The standard error for highest grade in our regression (0.0516) is close to the standard error in paper (0.051).

Compare the result in table 3 column 3 (basic controls) and value for currently in school, the estimate for in school by time of survey in our regression (0.00764) is bit higher to the estimate in paper (0.007). The standard error for in school by time of survey in our regression (0.0203) is close to the standard error in paper (0.0203).

h.

The paper indicates that "only about 90 percent of lottery winners had ever used the voucher or any other type of scholarship, while 24 percent of losers received scholarships from other sources." (p1356) There must have some unobservable factors that influence the students' grades and hide in the error term. For instance, whether the student is supported by other scholarships will influence the usage of voucher and longer schooling (outcome). Thus, we cannot use $Outcome_i = \alpha + \beta * UsedVoucher_i + \varepsilon_i$ to estimate the causal effect of actually using the voucher.

i.

The instrumental variable here is win voucher/not win voucher.

```
. reg usesch vouch0,robust

Linear regression                               Number of obs     =      1,135
                                                F(1, 1133)        =     971.66
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4663
                                                Root MSE          =     .35983

------------------------------------------------------------------------------
             |               Robust
     usesch |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------+----------------------------------------------------------------
     vouch0 |   .6721963   .0215645    31.17   0.000     .6298855    .7145072
      _cons |    .245045   .0182734    13.41   0.000     .2091915    .2808986
------------------------------------------------------------------------------
```

For first stage, we need to test whether IV have a causal impact on treatment variable (Used School Voucher). From the regression results above, we noticed that the 0.672 (estimate for voucher) is statistically significant and p-value <0.05, which represent that win voucher/not win voucher have strong relationship with usage of voucher. The IV have causal impact on our treatment variable and the requirement 1 hold.

For independence/exogeneity, we need to test whether IV are uncorrelated with unobservable that might affect outcomes. Since the allocation of voucher decided by the result of lottery and the lottery result is completely randomized, win voucher/not win voucher must independent to other unobservable factors that influence our outcome.
(Lottery winners received the voucher, while lottery losers did not.)

For exclusion restriction, we need to test whether IV impact on two outcomes: the highest grade achieved by the student and whether still in school only through our treatment variable. If the students want to use voucher, he must win the voucher at first. Also, the allocation of voucher is independent to other unobservable factors that influence the outcome, the allocation of voucher can only influence the two outcomes through the usage of voucher.

j.

	(1) High_Score_OLS_estimate	(2) High_Score_IV_estimate
usesch	0.171**	0.191*
	(0.0544)	(0.0770)
Obs.	1135	1135
R-squared	0.111	0.111
F-statistic	.	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For our estimation results, we need to compare the results in table 7 Bogota 1995 OLS and 2SLS estimation for highest grade completed. The estimates in paper from model that controls for the city,

year of application, phone access, age, type of survey and instrument, strata of residence, and month of survey. The paper uses OLS and 2SLS estimates to measure the effect of use voucher. Most of control variables and estimation methods are identical to our regression. Thus, the comparison between our estimate and table 7 Bogota 1995 OLS/2SLS results will be meaningful.

Compare the result in table 7(Bogota 1995) OLS estimate for highest grade completed, the estimate for highest grade in our regression (0.171) is bit higher to the estimate in paper (0.167). The standard error for highest grade in our regression (0.0544) is bit higher to the standard error in paper (0.053).

Compare the result in table 7(Bogota 1995) 2SLS estimate for highest grade completed, the estimate for highest grade in our regression (0.191) is bit lower to the estimate in paper (0.196). The standard error for highest grade in our regression (0.0770) is bit lower to the standard error in paper (0.078).

k.

From the result in question j, the 2sls estimate 0.191 is statistically significant at 5% level since P-value <0.05. The number 0.191 represents that the students who use the voucher are expected to have 0.191 points higher in highest score compared to the students who do not use the voucher on average. Thus, there is an effect of school vouchers on the highest grade achieved.

l.

	(1) In_Scl_OLS_estimate	(2) In_Scl_IV_estimate
usesch	0.0223	0.0115
	(0.0214)	(0.0304)
Obs.	1135	1135
R-squared	0.165	0.165
F-statistic	.	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The 2sls estimate 0.0115 is not statistically significant at 5% level. We fail to reject the null hypothesis that the usage of the voucher will not influence whether students still in school at the time of survey. Thus, we do not have evidence to prove the effect of school vouchers on the likelihood of being in school by the time of survey.