PS2

Mattar Klein (328780168)

I am trying out using markstat package for better reproducability, All code and asnwers will be in the same document, I will include the .stmd file and the boot file I use for running it (you will need to change the directories)

Also, all files can be found on my github at:

https://github.com/MattarKlein/Econometrics-B.git

Import Data

```
. clear all

. use "star_problem_set.dta"
```

Question 1

(a) Run Regression Basic Regression

```
. quietly gen sc = .

. quietly gen ra = .

. foreach v in k 1 2 3{
  2. quietly gen tss`v´ = treadss`v´ + tmathss`v´
  3. quietly replace sc = cltype`v´ == 1
  4. quietly replace ra = cltype`v´ == 3
  5. quietly regress tss`v´ sc ra, robust
  6. quietly eststo a_reg_`v´
  7. }
```

The regression model is

$$Y_i = \beta_0 + \beta_1 sc_i + \beta_2 ra_i + u_i$$

Where Y_i is the outcome variable done for 4 different grades (Kindergarden though 3rd), sc_i is a binary variable for being randomized into a small class, and ra_i is a binary variable for being randomized into a class with an assistant.

Therefore, the constant term is the average total score of the control group (i.e. no assistant assigned and reduction in class size), and the coefficients of sc_i and ra_i are the change in total test scores for those groups respectively.

As can be seen in the output table, being randomized into a small class has an effect across all classes and is largest in 1st grade.

```
. label variable tssk "K"
. label variable tss1 "1st"
. label variable tss2 "2nd"
. label variable tss3 "3rd"
. label variable sc "Small Class"
. label variable ra "Regular + Ass"
.
. esttab a_reg_k a_reg_1 a_reg_2 a_reg_3, label se ar2
```

	(1) K	(2) 1st	(3) 2nd	(4) 3rd
Small Class	13.90***	29.78***	19.39***	15.59***
	(2.454)	(2.831)	(2.712)	(2.396)
Regular + Ass	0.314	11.96***	3.479	-0.291
	(2.271)	(2.652)	(2.545)	(2.273)
Constant	918.0***	1039.4***	1157.8***	1228.5***
	(1.633)	(1.785)	(1.815)	(1.680)
Observations	5786	6379	6049	5967
Adjusted R-squared	0.007	0.017	0.009	0.010

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

(b) Including controls

```
. quietly gen totexp = .

. quietly gen ses = .

. foreach v in k 1 2 3{
  2. quietly replace tss`v´ = treadss`v´ + tmathss`v´
  3. quietly replace sc = cltype`v´ == 1
  4. quietly replace ra = cltype`v´ == 3
  5. quietly replace totexp = totexp`v´
  6. quietly replace ses = ses`v´
  7. quietly regress tss`v´ sc ra totexp ssex srace ses, robust
  8. quietly eststo b_reg_`v´
  9. }
```

The regression model is the same as before with added demographic controls for teacher experience and the gender race and socio-economic status of the student.

$$Y_i = \beta_0 + \beta_1 sc_i + \beta_2 ra_i + \beta_3 totexp_i + \beta_4 ssex_i + \beta_5 srace_i + \beta_6 ses_i + u_i$$

Comparing the size of the coefficients of small class to the coefficients of different demographics shows that the effects aren't very big. Take socio-economic status for instance this is measured by whether the sutdent is entitled to a free lunch or not, the coefficients on this are systematically larger and more than double the effect of small class.

The standard errors on sc_i went down for all coefficients, this is b/c by adding more controls the model is getting more exact. Also the adjusted R-Squared has gone up by quite a lot.

```
. label variable totexp "Exp Teacher"

. label variable ssex "Female"

. label variable srace "Black"

. label variable ses "Lower SE status"

.
. esttab b_reg_k b_reg_1 b_reg_2 b_reg_3, label se ar2
```

	(1) K	(2) 1st	(3) 2nd	(4) 3rd
Small Class	14.17***	24.80***	16.78***	11.99***
	(2.354)	(2.625)	(2.544)	(2.277)
Regular + Ass	0.776	8.250***	4.811*	-0.524
	(2.156)	(2.459)	(2.385)	(2.146)
Exp Teacher	1.117***	0.235*	0.124	0.0784
	(0.169)	(0.112)	(0.113)	(0.105)
Female	13.98***	11.91***	10.47***	7.654***
	(1.850)	(2.090)	(2.020)	(1.783)
Black	-10.36***	-27.97***	-31.55***	-20.06***
	(2.063)	(2.789)	(3.145)	(2.616)
Lower SE status	34.95***	56.98***	46.00***	42.02***
	(2.013)	(2.431)	(2.511)	(2.151)
Constant	847.5***	974.5***	1115.0***	1181.2***
	(6.044)	(7.459)	(8.124)	(6.940)

Observations	5748	6227	5707	5733
Adjusted R-squared	0.101	0.184	0.171	0.144

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

(c) Test for randomization

```
. foreach v in k 1 2 3{
  2. quietly replace sc = cltype`v´ == 1
  3. quietly replace ra = cltype`v´ == 3
  4. quietly replace ses = ses`v´
  5. quietly regress totexp`v´ sc ra, robust
  6. quietly eststo c_reg_`v´_exp
  7. quietly regress ssex sc ra, robust
  8. quietly eststo c_reg_`v´_ssex
  9. quietly regress srace sc ra, robust
 10. quietly eststo c_reg_`v´_srace
 11. quietly regress ses sc ra, robust
 12. quietly eststo c_reg_`v´_ses
 13. }
```

The following tables check the distribution of observables among different test groups. It seems that while some of the randomizations (like gender) did actually randomize across demographics there were some demographics which weren't split randomly (specifically race and teacher experience).

```
. label variable totexpk "Exp K"

. label variable totexp1 "Exp 1st"

. label variable totexp2 "Exp 2nd"

. label variable totexp3 "Exp 3rd"

. esttab c_reg_k_exp c_reg_1_exp c_reg_2_exp c_reg_3_exp, label se
```

	(1) Exp K	(2) Exp 1st	(3) Exp 2nd	(4) Exp 3rd
Small Class	-0.149	1.899***	0.311	-0.172
	(0.181)	(0.262)	(0.268)	(0.265)
Regular + Ass	0.670***	2.353***	1.102***	1.335***
	(0.175)	(0.257)	(0.250)	(0.254)
Constant	9.068***	10.30***	12.66***	13.49***
	(0.122)	(0.171)	(0.182)	(0.189)
Observations	6304	6810	6738	6751

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

```
. esttab c_reg_k_ssex c_reg_1_ssex c_reg_2_ssex c_reg_3_ssex, label se
```

	(1) Female	(2) Female	(3) Female	(4) Female
Small Class	0.0217	0.0176	0.0250*	0.0232
	(0.0128)	(0.0128)	(0.0126)	(0.0123)
Regular + Ass	0.0187	0.00230	0.0108	0.00287
	(0.0121)	(0.0119)	(0.0116)	(0.0116)
Constant	1.464***	1.468***	1.465***	1.466***
	(0.00578)	(0.00582)	(0.00593)	(0.00602)
Observations	11578	11578	11578	11578

Standard errors in parentheses

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

. esttab c_reg_k_srace c_reg_1_srace c_reg_2_srace c_reg_3_srace, label se

	(1) Black	(2) Black	(3) Black	(4) Black
Small Class	-0.0803***	-0.0895***	-0.0665***	-0.0902***
	(0.0139)	(0.0139)	(0.0135)	(0.0130)
Regular + Ass	-0.0633***	-0.116***	-0.0386**	-0.0577***
	(0.0128)	(0.0128)	(0.0128)	(0.0123)
Constant	1.415***	1.427***	1.409***	1.419***
	(0.00658)	(0.00656)	(0.00671)	(0.00702)
Observations	11453	11453	11453	11453

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

. esttab c_reg_k_ses c_reg_1_ses c_reg_2_ses c_reg_3_ses, label se

	(1) Lower SE s~s	(2) Lower SE s~s	(3) Lower SE s~s	(4) Lower SE s~s
Small Class	0.00622	0.0418**	0.0451**	0.0334*
	(0.0157)	(0.0152)	(0.0155)	(0.0157)
Regular + Ass	-0.0256	0.00609	0.0130	-0.000146
	(0.0151)	(0.0145)	(0.0148)	(0.0151)
Constant	1.523***	1.470***	1.468***	1.484***
	(0.0107)	(0.00992)	(0.0106)	(0.0113)
Observations	6302	6651	6496	6520

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

(d) Treatment Heterogeneity

```
. quietly gen sc_totexp = .

. quietly gen sc_ssex = .

. quietly gen sc_srace = .

. quietly gen sc_ses = .

. foreach v in k 1 2 3{
  2. quietly replace tss`v´ = treadss`v´ + tmathss`v´
  3. quietly replace sc = cltype`v´ == 1
  4. quietly replace ra = cltype`v´ == 3
  5. quietly replace totexp = totexp`v´
  6. quietly replace ses = ses`v´
  7. quietly replace sc_totexp = sc*totexp
  8. quietly replace sc_ssex = sc*ssex
  9. quietly replace sc_srace = sc*srace
 10. quietly replace sc_ses = sc*ses
 11. quietly regress tss`v´ sc ra totexp sc_totexp ssex sc_ssex srace sc_srace ses sc_ses, robus
> t
 12. quietly eststo d_reg_`v´
 13. }
```

In order to account for heterogeneity across students with different co-variates I have added interaction terms for the 4 observed demographics with small class. It does not seem like there is much heterogeneity in the sample. However, as can be seen in the previous question, I am not very convinced about the randomization anymore.

```
. label variable sc_totexp "SC X Exp"

. label variable sc_ssex "SC X Female"

. label variable sc_srace "SC X Black"

. label variable sc_ses "SC X Lower SE status"

.
. esttab d_reg_k d_reg_1 d_reg_2 d_reg_3, label se ar2
```

	(1) K	(2) 1st	(3) 2nd	(4) 3rd
Small Class	39.53**	-5.587	5.243	-3.557
	(12.93)	(15.69)	(17.58)	(14.00)
Regular + Ass	0.423	8.341***	4.825*	-0.233
	(2.154)	(2.468)	(2.390)	(2.145)
Exp Teacher	1.607***	0.0948	0.0852	-0.181
	(0.198)	(0.132)	(0.134)	(0.123)
SC X Exp	-1.667***	0.462	0.111	0.791***
	(0.366)	(0.250)	(0.249)	(0.231)
Female	16.82***	11.37***	10.30***	8.242***
	(2.159)	(2.427)	(2.377)	(2.131)
SC X Female	-9.567*	1.785	0.572	-1.774
	(4.143)	(4.745)	(4.496)	(3.870)
Black	-10.89***	-30.63***	-33.50***	-21.49***
	(2.360)	(3.448)	(3.767)	(3.364)
SC X Black	1.066	8.155	6.771	4.937
	(4.610)	(5.910)	(6.901)	(5.258)
Lower SE status	34.81***	54.56***	46.00***	42.25***

	(2.373)	(2.886)	(2.978)	(2.678)
SC X Lower SE status	1.614	7.439	0.0482	0.589
	(4.419)	(5.383)	(5.553)	(4.544)
Constant	839.7***	984.1***	1118.4***	1185.5***
	(7.094)	(9.084)	(9.669)	(8.662)
Observations	5748	6227	5707	5733
Adjusted R-squared	0.105	0.184	0.171	0.146

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

(e) As I stated before I am not convinced about the randomization, it seems that there was some confounding factor which affected the randomization (or maybe just chance). My guess is that after the randomization there was some movement between treatments. This kind of movement will only be made worse in a real world environment if only some classes actually reduce size. Additionally if all classes reduce size then I am not sure how good the estimates actually help.

However, the experiment might be usefull for policy if researches try to go back and understand how such changes were done, and how they affected different demographic groups after the fact.

Question 2

$$Y_{0i} \perp D_i \rightarrow E[Y_{0i}|D_i = 0] = E[Y_{0i}|D_i = 1]$$

Adding this to the definition of ATT (1), we get:

$$ATT = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \overset{(1)}{=} E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$

Question 3

(a)

$$\tau_i = Y_{1i} - Y_{0i} = \beta_0 + \beta_1 + \beta_2 v_i + u_i - \beta_0 - u_i = \beta_1 + \beta_2 v_i$$

(b)

- $\tau_{ATE} = E[Y_{1i}] - E[Y_{0i}] = \beta_1 + \beta_2 E[v_i]$
- $\tau_{ATT} = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] = \beta_1 + \beta_2 E[v_i|D_i = 1]$
- $\tau_{ATU} = E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0] = \beta_1 + \beta_2 E[v_i|D_i = 0]$

(i)

$$E[Y_{0i}|D_i = 1] = \beta_0 + E[u_i|D_i = 1]$$

$$E[Y_{0i}|D_i = 0] = \beta_0 + E[u_i|D_i = 0]$$

This means that the assumption is equivelent to $E[u_i|D_i = 1] = E[u_i|D_i = 0]$, i.e. independance of u_i and D_i.

(ii)

$$E[Y_{1i}|D_i = 1]-E[Y_{0i}|D_i = 0] = \beta_0+\beta_1+\beta_2 E[v_i|D_i = 1]+E[u_i|D_i = 1]-\beta_0-E[u_i|D_i = 0] =$$

$$= \beta_1 + \beta_2 E[v_i|D_i = 1]$$

(iii)

Looking at the previous equation and the equation for the ATE - the assumption needed is that v_i is independant of D_i

(c)

$\gamma_1 = mean_i(Y|D_i = 1) - mean_i(Y|D_i = 0) \to^p E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$

Question 4

(a)

$$E[Y_{1i}] - E[Y_{0i}] = \beta_0 + \beta_1 E[I(\eta_i = \eta_2)] + \beta_2 E[\eta_i] + E[\epsilon_i] - \beta_0 - \beta_2 E[\eta_i] - E[\epsilon_i] =$$

$$\beta_1 PR(\eta_i = \eta_2)$$

The treatment D_i has no effect on W_i so ATE is 0.

(b)

$$E[Y_i|D_i = 1]-E[Y_i|D_i = 0] = \beta_1 PR(\eta_i = \eta_2|D_i = 1)+\beta_2 E[\eta_i|D_i = 1]-\beta_2 E[\eta_i|D_i = 0] =$$

$$= \beta_1 + \beta_2(\eta_2 - \eta_1)$$

This does not converge to ATE b/c the assignment to treatment is non-random.

(c)

(i) The first one without $Y_i > 20$ does give the ATE. Unfortunately we can only observe $Y_i > 20$ so we do not have a sample equivilent for the ATE.

(ii) This does not yeild the ATE for W_i seeing as getting treatment might push you over the hump of $Y_i > 20$ so it will look like there is a positive affect.