FLS 6415: Replication 7 - Controlling for Confounding

May 2019

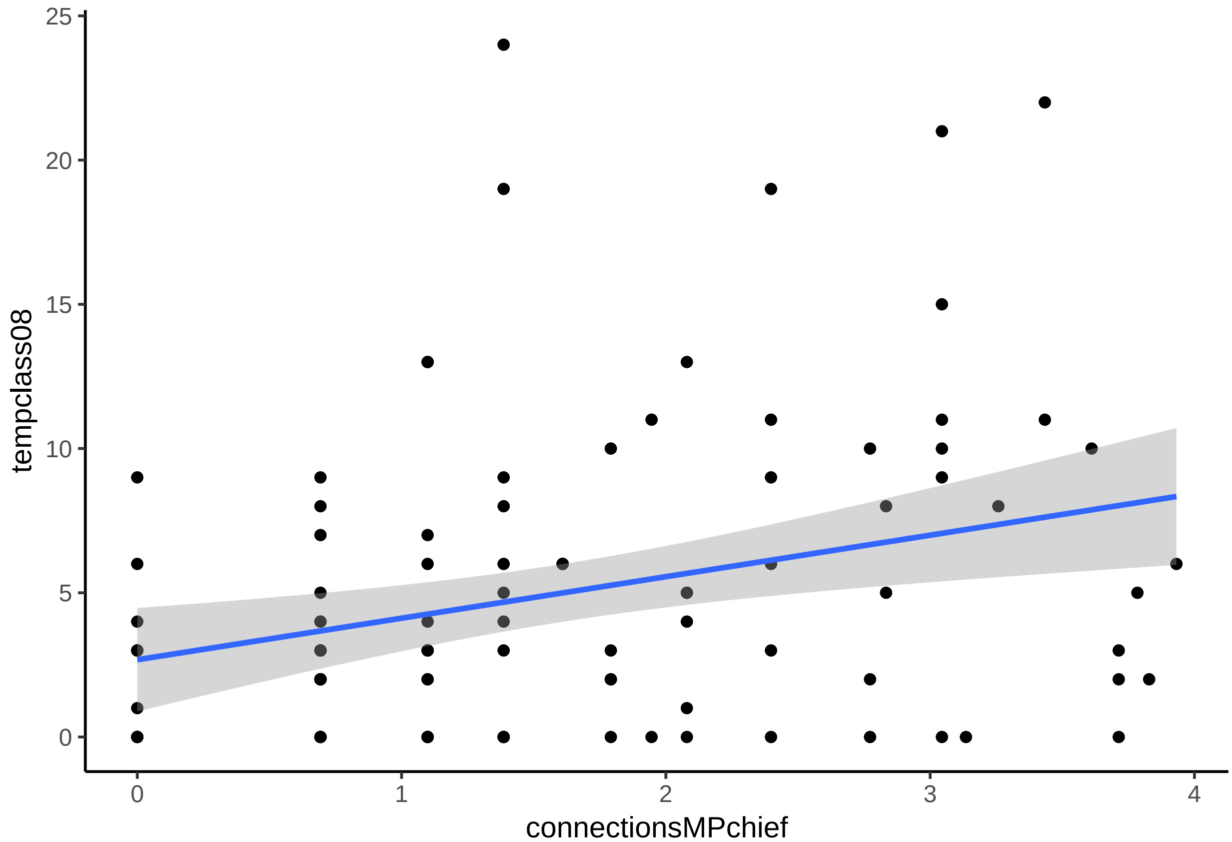
To be submitted (code + answers) by midnight, Wednesday 29th May.

First read the paper by Baldwin (2013) on the course website.

The replication data is in the file *Baldwin_adjusted.csv*. A list of available variables is also provided below.

Variable	Description
connectionsMPchief	Number of years the MP has known the chief (already in 'log' form) - **Treatment**
tempclass08	Number of temporary classrooms in 2007-2008 - **Outcome**
tempclass07	Number of temporary classrooms in 2006-2007
pop2000	Village population in 2000
pop2009	Village population in 2009
experienceMP	Years since MP first elected
experiencechief	Years since Chief installed
voteMP06	% Vote for MP
MMD06	MP form governing party
diffvoteconst06	Difference in vote share between top two candidates
univMP	MP went to University
cabinetMP	MP has ever been in the cabinet
localMP	MP is from the chiefdom
secondaryedchief	Chief completed secondary education
politicalexpchief	Chief has ever participated in politics
agechief	Age of the chief in years
constcode	Constituency code
classneedper100	Students per Classroom 2006-07
yearinstalledchief	Year became chief
percturnout06	Turnout 2006 election
numcandidates	Number of candidates in 2006 election

1. We will focus on assessing Baldwin's (2013) claim that "politicians with stronger relationships to chiefs actually do provide more local public goods". Create a plot of the treatment variable (`connectionsMPchief`) against the outcome variable, the number of temporary classrooms in 2008 (`tempclass08`). Add a linear line of best fit to assess the relationship.



2. Implement the basic linear regression of the outcome on treatment with no controls/covariates. Interpret what you can conclude from this regression. *Note:* The `connectionsMPchief` variable is already in 'log' form (see Class 1 for guidance on how to interpret logged explanatory variables).

Table 2:

	Dependent variable:
	tempclass08
connectionsMPchief	1.439***
	(0.466)
Constant	2.679***
	(0.903)
Observations	101
R^2	0.088
Adjusted R^2	0.079
Residual Std. Error	5.073 (df = 99)
F Statistic	9.530*** (df = 1; 99)
Note:	*p<0.1; **p<0.05; ***p<0.01

A 1% increase in the number of years the chief has known the MP is associated with a 0.014 ($log(\frac{101}{100}) * 1.439$) increase in the number of temporary classrooms.

3. Provide two concrete reasons of how our estimate in Q2 might be biased. In which direction would the bias be?

4. Describe the Treatment Assignment Mechanism for the length of the relationship between MP and Chief

5. Draw (by hand) the causal diagram (DAG) for our study, including the treatment effect of interest, the treatment assignment mechanism, and the threats to causal inference you described above. (Don't make it too complicated!)

6. Based on your causal diagram (DAG), describe one set of control variables which would be sufficient to provide an unbiased estimate of the causal effect of treatment (if the DAG were correct).

7. One potential omitted variable (confounder) is population - in larger villages the MP and Chief are less likely to know each other personally, and village size might also affect the resources/demand for temporary classrooms. There are two potential control variables in the dataset we could use, a measure of population in 2000 (`pop2000`) and a measure of population in 2009 ('pop2009'). Which should we use, and why?

We have to use `pop2000` to avoid post-treatment bias. It could be that better relations between chief and MP attract more people, increasing population, and therefore more classrooms are built. So controlling for population in 2009 might remove part of the real treatment effect.

8. Run the simple linear regression of the outcome on treatment, controlling for any variables you identified as appropriate in Q6 and Q7 above. How do your results compare to the results of the regresssion in Q2?

Table 3:

	Dependent variable:
	tempclass08
connectionsMPchief	1.201^{***}
	(0.437)
univMP	-0.092
	(0.975)
secondaryedchief	-0.013
	(0.956)
numcandidates	-0.002
	(0.179)
Constant	2.842^{**}
	(1.411)
Observations	95
R^2	0.079
Adjusted R^2	0.038
Residual Std. Error	4.574 (df = 90)
F Statistic	1.940 (df = 4; 90)
Note:	*p<0.1; **p<0.05; ***p<0.01

9. Baldwin (2013) runs her regression using an ordered multinomial (ordered logit) model,

reflecting the fact that the outcome variable (number of temporary classrooms) can only take on a fixed set of integer values. Repeat your regression from Q8 but with an ordered logit model and interpret the results. (Note that Baldwin also clusters standard errors at the constituency (`constcode`) level, but don't worry about replicating this, just focus on the coefficient).

Table 4:

	Dependent variable:
	tempclass08
connectionsMPchief	0.407**
	(0.172)
univMP	−0.248
	(0.380)
secondaryedchief	−0.040
	(0.369)
numcandidates	−0.008
	(0.066)
Observations	95
Note:	*p<0.1; **p<0.05; ***p<0.01

A 1% increase in the number of years the chief has known the MP is associated with a 0.004 ($log(\frac{101}{100}) * 0.407$) increase in the log-odds of having one more temporary classroom.

10. Now replicate the results from column (1) of Baldwin's Table 1, i.e. only include the control variables that she includes in Table 1. Compare the estimated treatment effect with your own model in Q9. (Again, don't worry about the standard errors).

11. Replicate all three columns of Table 1 in Baldwin (2013). How stable is the estimate of the treatment effect to alternative specifications of the covariates?

Table 5:

	Dependent variable:
	tempclass08
connectionsMPchief	0.352^{**}
	(0.175)
tempclass07	0.385^{***}
	(0.057)
pop2000	0.043^{***}
	(0.015)
experienceMP	-0.057
	(0.042)
experiencechief	0.027^{**}
	(0.013)
Observations	99
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6:

	Dependent variable:		
	tempclass08		
	(1)	(2)	(3)
connectionsMPchief	0.352**	0.378**	0.331*
	(0.175)	(0.181)	(0.184)
tempclass07	0.385***	0.380***	0.440***
	(0.057)	(0.058)	(0.064)
pop2000	0.043***	0.053***	0.033*
	(0.015)	(0.017)	(0.017)
experienceMP	−0.057	−0.054	−0.045
	(0.042)	(0.045)	(0.056)
experiencechief	0.027**	0.023*	0.009
	(0.013)	(0.013)	(0.018)
MMD06		−0.480	
		(0.563)	
voteMP06		0.932	
		(1.142)	
diffvoteconst06		−1.934	
		(1.371)	
univMP			−0.129
			(0.464)
cabinetMP			0.399
			(0.748)
localMP			0.482
			(0.478)
agechief			0.030*
			(0.017)
secondaryedchief			−0.012
			(0.447)
politicalexpchief			0.617
			(0.508)
Observations	99	96	94
Note:	*p<0.1; **p<0.05; ***p<0.01		