FLS 6415: Replication 3 - Natural Experiments

April 2019

To be submitted (code + answers) by midnight, Wednesday 24th April.

First read the paper by De La O (2013) on the class website.

The replication data is in the file *DelaO.csv*, and the important variables are described below. Each row of this dataset is one electoral precinct, some of which are considered treated because they received *Progresa*.

Table 1: Key Variables in De La O (2013)

Variable	Description
treatment	Whether the precinct received Progresa
numerotreated	Number of Treated Villages in Precinct
numerocontrol	Number of Control Villages in Precinct
avgpoverty	Poverty in 1995
pobtot1994	Population in 1994/5
pobelegiblep	Population Eligible
villages	Number of Villages in Precinct
t1994	Turnout % in 1994
pri1994s	PRI vote share 1994
pan1994s	PAN vote share 1994
prd1994s	PRD vote share 1994
votos_totales1994	Total Number of Votes in 1994
pri1994	Number of PRI votes in 1994
pan1994	Number of PAN votes in 1994
prd1994	Number of PRD votes in 1994
t2000	Turnout % in 2000
pri2000s	PRI Vote Share in 2000
pan2000s	PAN Vote Share in 2000
prd2000s	PRD Vote Share in 2000

1. First, what is treatment in this study? What is control? What is the outcome being measured?

Treatment is receiving Progresa 'early', 21 months before the election.

Control is receiving Progresa 'late', 6 months before the election.

The outcomes are turnout and party vote shares in the 2000 election.

2. To help assess the balance between treatment and control units, reproduce Table 2 in De La O (2013) (Don't worry about the standard errors in brackets in the 'Difference' column for now).

```r
tab2_vars <- names(d)[c(5,4,6,13,10,11,20,22,24,26)]

Q1 <- d %>% group_by(treatment) %>%
  summarise_at(tab2_vars,mean,na.rm=TRUE) %>%
  gather(key,value,-treatment) %>%
  spread(treatment,value) %>%
  rename(Variable="key",Early=`1`,Late=`0`) %>%
  mutate(Difference=Late-Early)
```

```r
kable(Q1,caption="Q2: Balance in the Data", digits=3)
```

Table 2: Q2: Balance in the Data

Variable	Late	Early	Difference
avgpoverty	4.593	4.576	0.017
one_random	0.909	0.906	0.003
pan1994s	0.062	0.051	0.012
pobelegiblep	0.849	0.887	-0.038
pobtot1994	1851.610	2040.107	-188.496
prd1994s	0.098	0.102	-0.004
pri1994s	0.410	0.435	-0.025
t1994	0.643	0.658	-0.015
two_random	0.084	0.091	-0.006
villages	6.377	6.084	0.292

3. Is the balance shown in this table (Table 2 in De La O) a necessary condition for causal inference? Is it a sufficient condition for causal inference?

Balance is a necessary condition for causal inference because systematic differences between the treatment and control groups on the covariates could be responsible for any diffences in the outcome variable and act as a confounder, preventing us from isolating the effect of treatment alone. However, balance is not a sufficient condition for causal inference. Even if we have balance on observed covariates, there may be imbalance on unobserved or unmeasurable covariates. For this reason it is also important to verify the *procedure* of treatment assignment could not have introduced any confounding. Other assumptions such as SUTVA (no spillovers) also matter.

4. The main analysis in De La O is conducted on a subset of the full dataset. Filter the data so that only precincts that have either one treatment village (`numerotreated`) or one control village (`numerocontrol`) inside them are included in your new dataset. What percentage of the original precincts are included in the new dataset?

```r
subset_d <- d %>% filter(numerotreated==1|numerocontrol==1)

pct_data <- round(dim(subset_d)[1]/dim(d)[1]*100,1)
```

After filtering for precincts that have only one treatment or control village in them, 90.7% of the data remains.

5. One of De La O's conclusions is that treatment boosts turnout. Conduct a simple difference-in-means t-test on the filtered dataset from Q4 to assess this claim. What is the estimated difference-in-means and how statistically significant is the result?

```r
subset_d %>% t.test(t2000~treatment, data=.) %>%
  tidy() %>%
  kable(caption="Q5: Difference-in-Means")
```

Table 3: Q5: Difference-in-Means

estimate	estimate1	estimate2	statistic	p.value	parameter	conf.low	conf.high	method
-0.0456397	0.634774	0.6804137	-1.608248	0.1086198	376.1164	-0.1014401	0.0101607	Welch Two Sample t-te

The estimated difference in means is 4.56% points, which is not significant.

6. De La O's analysis of turnout is in the upper panel of Table 3, where she runs a regression, adding some controls. (We are going to focus on the 'ITT' estimates, we will talk about the 'IV' estimates next week). Replicate this turnout regression. The controls (listed under De La O's Table 3) are `avgpoverty,pobtot1994,votos_totales1994,pri1994,pan1994,prd1994` and there is a fixed effect for the *villages* variable. (Try to include the robust standard errors, but no problem if you cannot). Interpret the results.

Early receipt of Progresa is associated with a 5.3% points higher turnout, which is significant at the 10% but not the 5% level, holding all the other variables constant.

```
reg1 <- lm_robust(t2000~treatment + avgpoverty + pobtot1994 + votos_totales1994 + pri1994 + pan1994 + pr
texreg(reg1, include.ci=F, digits=3, caption="Q6: With Fixed Effects")
```

7. Now run the same regression but exclude the number-of-villages fixed effects (keep the other controls). How does this change the comparisons we are making between treated and control villages? How do the results change?

Before we were comparing between precincts that were treated or control but which could also vary in the number of villages in each precinct. To compare treated and control precincts only between precincts with the same number of villages she includes a fixed effect for the number of villages. The results become marginally weaker, with both the coefficient and the standard error changing.

```
reg2 <- lm_robust(t2000~treatment + avgpoverty + pobtot1994 + votos_totales1994 + pri1994 + pan1994 + pr
texreg(reg2, include.ci=F, digits=3, caption="Q7: Without village Fixed Effects")
```

8. Replicate all four columns of the upper panel of Table 3 in De La O (2013). Interpret the results.

Early receipt of Progresa boosts vote share for the PRI by 3.7% points, with a significant effect at the 1% level after controlling for the various variables. It has a positive but not significant effect on the other parties.

```
dvs <- c("t2000","pri2000s","pan2000s","prd2000s")

models <- dvs %>% map(~lm_robust(formula(paste(.x, "~treatment + avgpoverty + pobtot1994 + votos_totale
                                  data=subset_d,
                                  se_type="HC1"))
texreg(models, include.ci=F, digits=3, caption="Q8: De La O Table 3")
```

9.Now let's look at some critiques of the paper. Normally, we measure turnout percentages and vote shares as being naturally bounded between 0 and 100% (or 0 and 1). Other numbers don't make sense. Use a boxplot or similar graphic to assess the distribution of values on the four dependent variables. What do you find?

```
subset_d %>% select(dvs) %>%
  gather(key="Variable",value="Value") %>%
  ggplot() +
  geom_boxplot(aes(x=Variable, y=Value)) +
  theme_classic()
```

There are lots of 'non-feasible'values larger than one.

10. As a 'quick fix' replace all the unrealistic values above 100% (1) with `NA` for all the turnout percentage and vote share dependent variables. Re-run your regressions from question 8. Do your conclusions change? Why might this be?

The coefficient on treatment is no longer significant for both turnout and the PRI's vote share and the estimates are much lower. This may be because the extreme values exaggerate the turnout and vote share for the treatment units more than for control units, biasing the coefficient estimate upwards.

	Model 1
(Intercept)	0.619^{***}
	(0.171)
treatment	0.053
	(0.030)
avgpoverty	0.014
	(0.035)
pobtot1994	-0.000^{**}
	(0.000)
votos_totales1994	-0.000
	(0.000)
pri1994	0.000
	(0.000)
pan1994	0.000
	(0.001)
prd1994	0.000
	(0.000)
factor(villages)2	-0.058
	(0.046)
factor(villages)3	-0.066
	(0.040)
factor(villages)4	-0.029
	(0.055)
factor(villages)5	0.026
	(0.087)
factor(villages)6	-0.051
	(0.050)
factor(villages)7	-0.026
	(0.058)
factor(villages)8	-0.134^{*}
	(0.055)
factor(villages)9	0.082
	(0.127)
factor(villages)10	-0.134^{*}
	(0.064)
factor(villages)11	-0.115^{*}
	(0.053)
factor(villages)12	-0.032
	(0.067)
factor(villages)13	0.149
	(0.142)
factor(villages)14	-0.116
	(0.068)
R^2	0.116
Adj. R^2	0.071
Num. obs.	417
RMSE	0.298

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

Table 4: Q6: With Fixed Effects

	Model 1
(Intercept)	0.640***
	(0.147)
treatment	0.045
	(0.028)
avgpoverty	0.003
	(0.031)
pobtot1994	−0.000***
	(0.000)
votos_totales1994	−0.000
	(0.000)
pri1994	0.000
	(0.000)
pan1994	0.000
	(0.001)
prd1994	0.000
	(0.000)
R^2	0.079
Adj. R^2	0.063
Num. obs.	417
RMSE	0.299

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

Table 5: Q7: Without village Fixed Effects

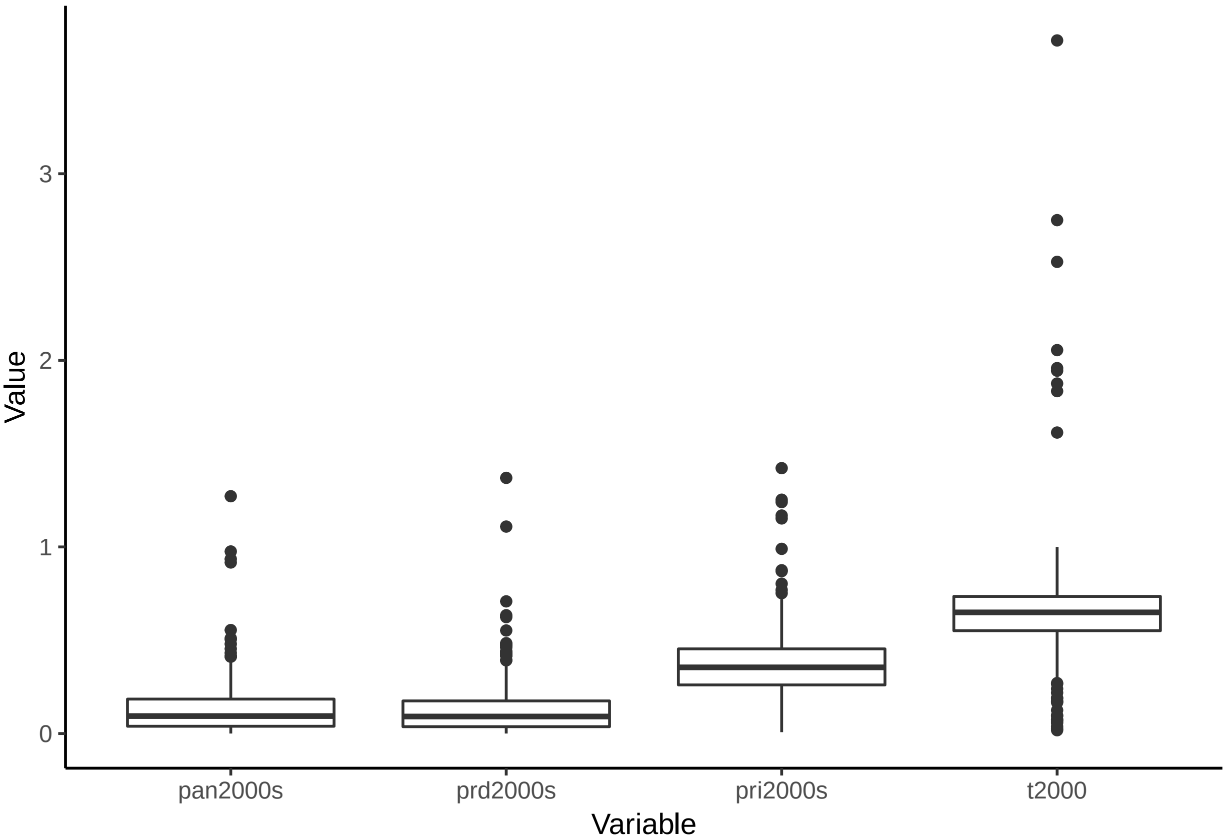

Figure 1: Non-feasible values in the Dependent Variables

	Model 1	Model 2	Model 3	Model 4
(Intercept)	0.619^{***}	0.357^{***}	0.135	0.137
	(0.171)	(0.083)	(0.072)	(0.075)
treatment	0.053	0.037^{*}	0.007	0.002
	(0.030)	(0.015)	(0.012)	(0.014)
avgpoverty	0.014	0.029	-0.013	-0.009
	(0.035)	(0.017)	(0.015)	(0.015)
pobtot1994	-0.000^{**}	-0.000^{***}	-0.000^{***}	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
votos_totales1994	-0.000	-0.000^{*}	-0.000^{*}	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
pri1994	0.000	0.001^{**}	0.000	-0.000^{*}
	(0.000)	(0.000)	(0.000)	(0.000)
pan1994	0.000	-0.000	0.001^{***}	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)
prd1994	0.000	-0.000	-0.000	0.001^{***}
	(0.000)	(0.000)	(0.000)	(0.000)
factor(villages)2	-0.058	-0.094^{*}	0.023	0.016
	(0.046)	(0.039)	(0.021)	(0.022)
factor(villages)3	-0.066	-0.103^{**}	0.040	0.007
	(0.040)	(0.038)	(0.021)	(0.017)
factor(villages)4	-0.029	-0.115^{**}	0.052	0.029
	(0.055)	(0.038)	(0.026)	(0.021)
factor(villages)5	0.026	-0.116^{**}	0.068^{*}	0.061
	(0.087)	(0.045)	(0.031)	(0.035)
factor(villages)6	-0.051	-0.183^{***}	0.098^{***}	0.034
	(0.050)	(0.037)	(0.025)	(0.029)
factor(villages)7	-0.026	-0.118^{**}	0.059^{*}	0.047
	(0.058)	(0.045)	(0.023)	(0.027)
factor(villages)8	-0.134^{*}	-0.153^{***}	0.010	0.017
	(0.055)	(0.043)	(0.021)	(0.030)
factor(villages)9	0.082	-0.115^{*}	0.142^{*}	0.044
	(0.127)	(0.054)	(0.069)	(0.026)
factor(villages)10	-0.134^{*}	-0.204^{***}	0.042	0.029
	(0.064)	(0.047)	(0.035)	(0.033)
factor(villages)11	-0.115^{*}	-0.211^{***}	0.072^{*}	0.027
	(0.053)	(0.042)	(0.031)	(0.024)
factor(villages)12	-0.032	-0.175^{***}	0.033	0.106^{**}
	(0.067)	(0.051)	(0.026)	(0.040)
factor(villages)13	0.149	-0.032	0.110^{*}	0.059
	(0.142)	(0.093)	(0.046)	(0.033)
factor(villages)14	-0.116	-0.188^{***}	0.061	0.016
	(0.068)	(0.045)	(0.031)	(0.023)
R^2	0.116	0.288	0.197	0.318
Adj. R^2	0.071	0.252	0.157	0.284
Num. obs.	417	417	417	417
RMSE	0.298	0.158	0.126	0.122

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

Table 6: Q8: De La O Table 3

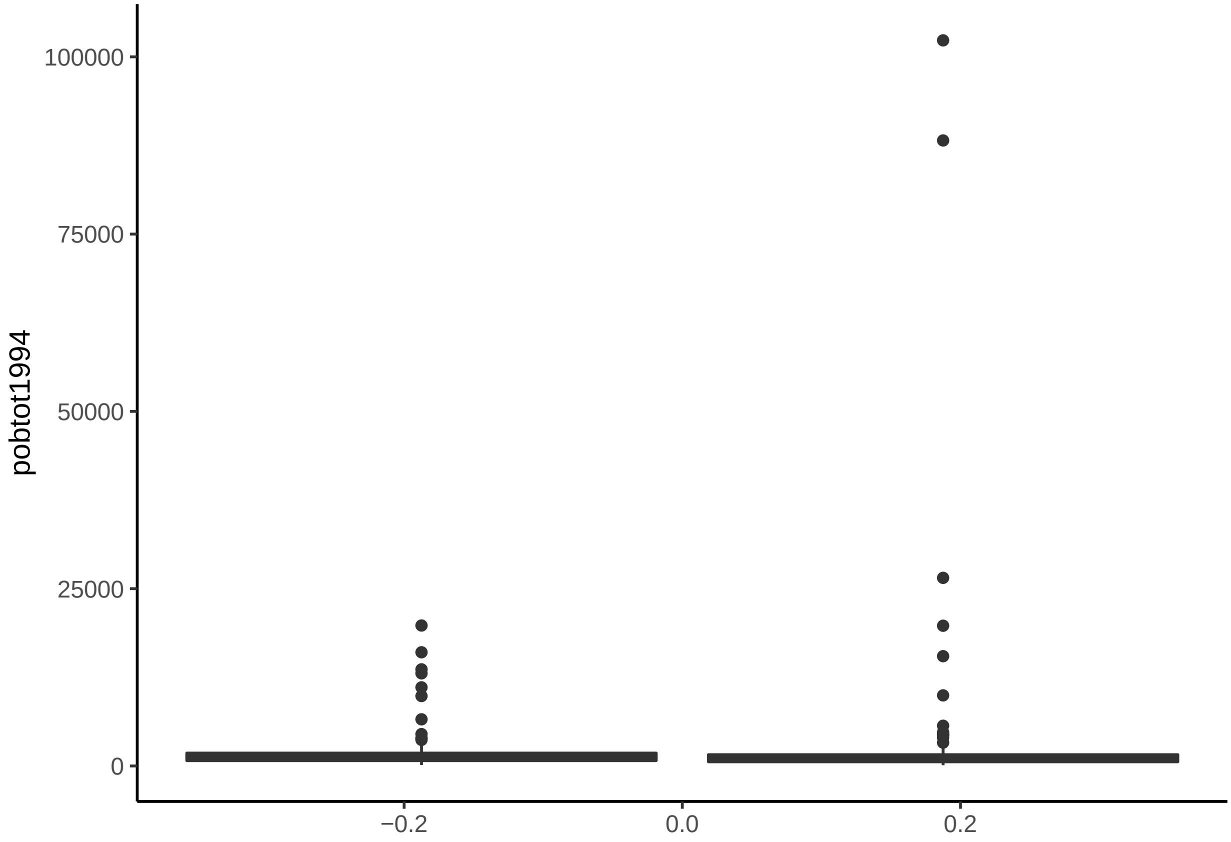

Figure 2: Q11: Outliers in Population Data

```
corrected_d <- subset_d %>% mutate_at(dvs, function(x) {ifelse(x>1, NA,x)})

models2 <- dvs %>% map(~lm_robust(formula(paste(.x, "~treatment + avgpoverty + pobtot1994 + votos_total
                                  data=corrected_d,
                                  se_type="HC1"))
texreg(models2, include.ci=F, digits=3, caption="Q10: Feasible values for Dependent Variables")
```

11. Next, examine the control variable for population in 1994 (pobtot1994). Use a graph or other method to identify any extreme outliers. Extreme values of control variables are not a problem if they are balanced across treatement and control groups. But are they in this case? Identify whether the extreme outliers are in the control or treatment group.

The plot identifies two extreme outliers far from the rest of the data. These two extreme values are both treatment values, which means neither can have a suitable counterfactual that is a control unit.

```
subset_d %>%
  select(treatment, pobtot1994) %>%
  ggplot() +
  geom_boxplot(aes(y=pobtot1994, group=treatment)) +
  theme_classic()
```

12. Remove the extreme outliers you identified in Q9 from the dataset (the dataset before you removed the infeasible values of the dependent variables). Re-run your regressions. Do your conclusions change? Why might this be?

	Model 1	Model 2	Model 3	Model 4
(Intercept)	0.783^{***}	0.451^{***}	0.106	0.155^{**}
	(0.089)	(0.070)	(0.067)	(0.056)
treatment	0.016	0.018	0.004	0.003
	(0.017)	(0.013)	(0.012)	(0.011)
avgpoverty	-0.016	0.012	-0.006	-0.013
	(0.019)	(0.015)	(0.013)	(0.012)
pobtot1994	-0.000^{***}	-0.000^{***}	-0.000^{***}	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
votos_totales1994	-0.000	-0.000	-0.000^{*}	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
pri1994	0.000	0.000^{**}	0.000^{*}	-0.000^{*}
	(0.000)	(0.000)	(0.000)	(0.000)
pan1994	-0.000	-0.001^{*}	0.001^{***}	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
prd1994	-0.000	-0.000^{*}	-0.000	0.001^{***}
	(0.000)	(0.000)	(0.000)	(0.000)
factor(villages)2	-0.075	-0.109^{**}	0.023	0.016
	(0.039)	(0.035)	(0.021)	(0.022)
factor(villages)3	-0.059	-0.095^{*}	0.038	0.009
	(0.038)	(0.038)	(0.021)	(0.017)
factor(villages)4	-0.053	-0.118^{***}	0.049	0.032
	(0.037)	(0.035)	(0.026)	(0.021)
factor(villages)5	-0.040	-0.131^{***}	0.064^{*}	0.036
	(0.039)	(0.038)	(0.031)	(0.021)
factor(villages)6	-0.064	-0.172^{***}	0.095^{***}	0.022
	(0.041)	(0.036)	(0.025)	(0.025)
factor(villages)7	-0.063	-0.135^{***}	0.056^{*}	0.051
	(0.044)	(0.039)	(0.023)	(0.027)
factor(villages)8	-0.116^{*}	-0.139^{**}	0.004	0.026
	(0.051)	(0.042)	(0.020)	(0.029)
factor(villages)9	-0.012	-0.099	0.077^{**}	0.051^{*}
	(0.048)	(0.052)	(0.026)	(0.025)
factor(villages)10	-0.118	-0.193^{***}	0.039	0.032
	(0.062)	(0.049)	(0.035)	(0.032)
factor(villages)11	-0.103^{*}	-0.197^{***}	0.067^{*}	0.032
	(0.050)	(0.041)	(0.031)	(0.024)
factor(villages)12	-0.022	-0.158^{**}	0.028	0.111^{**}
	(0.063)	(0.051)	(0.026)	(0.040)
factor(villages)13	-0.060	-0.100	0.105^{*}	0.064^{*}
	(0.041)	(0.062)	(0.046)	(0.033)
factor(villages)14	-0.139^{**}	-0.176^{***}	0.055	0.025
	(0.051)	(0.044)	(0.031)	(0.022)
R^2	0.233	0.353	0.221	0.360
Adj. R^2	0.194	0.320	0.182	0.328
Num. obs.	408	412	416	415
RMSE	0.162	0.129	0.113	0.099

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

Table 7: Q10: Feasible values for Dependent Variables

The coefficients are again smaller and no longer significant. One possible reason could be because the outliers on high population were both treated and had unusually high levels of turnout and PRI vote share (perhaps because urban areas are confounded and systematically differnet in these ways). So their inclusion in the original regression biases the coefficients upwards. In fact, that's not actually happening here - the outliers have lower than normal turnout and PRI vote share. But because we are controlling for population, when we include these outliers in the regression the linear effect of population looks very different and changes our estimates of the 'controlled' effect of treatment for all the other units.

```
popn_corrected_d <- subset_d %>% mutate(rank_popn=rank(-pobtot1994))  %>%
  filter(rank_popn>2)

models3 <- dvs %>% map(~lm_robust(formula(paste(.x, "~treatment + avgpoverty + pobtot1994 + votos_total
                                 data=popn_corrected_d,
                                 se_type="HC1"))
texreg(models3, include.ci=F, digits=3, caption="Q12: Removing Population Outliers")
```

13. One more issue. The controls for the regressions you have conducted so far are the *absolute number* of votes for turnout, PRI, PAN and the PRD. But for the dependent variable, De La O is using the *percentage vote share of the population.* Arguably it might be more consistent to use the same measurement approach on both the left and right-hand sides of the regression. Try implementing the regressions using the controls `t1994`, `pri1994s`, `pan1994s`, `prd1994s` in place of `votos_totales1994`, `pri1994`, `pan1994`, `prd1994`. Ignore the other corrections you made in previous questions. Does this change your conclusions? Why might this be?

The Table shows the coefficients are again much lower, and no longer significant. This may be because the controls in absolute numbers controlled more for the size of the precinct than for the relative strength of the parties. So when using the relative measure of support for each party we are controlling for potential confounders and imbalance that reduces the treatment effect. Moreover, the interpretation of the regression is much more natural with the same format for vote share on the left hand side and the right hand side, because we can interpret the outcome now as the change in vote share between 1994 and 2000.

```
models4 <- dvs %>% map(~lm_robust(formula(paste(.x, "~treatment + avgpoverty + pobtot1994 + t1994 + pri
                                 data=popn_corrected_d,
                                 se_type="HC1"))
texreg(models4, include.ci=F, digits=3, caption="Q13: Using Percentage Controls")
```

	Model 1	Model 2	Model 3	Model 4
(Intercept)	0.636***	0.363***	0.140*	0.141
	(0.162)	(0.081)	(0.069)	(0.074)
treatment	0.039	0.031*	0.003	−0.002
	(0.028)	(0.015)	(0.012)	(0.014)
avgpoverty	0.012	0.028	−0.014	−0.010
	(0.033)	(0.017)	(0.014)	(0.015)
pobtot1994	−0.000***	−0.000***	−0.000***	−0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
votos_totales1994	−0.000	−0.000*	−0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
pri1994	0.000	0.001***	0.000*	−0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
pan1994	0.001	−0.000	0.002***	−0.000
	(0.001)	(0.000)	(0.000)	(0.000)
prd1994	0.001	−0.000	−0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
factor(villages)2	−0.041	−0.087*	0.027	0.022
	(0.043)	(0.039)	(0.020)	(0.022)
factor(villages)3	−0.071	−0.105**	0.039	0.006
	(0.038)	(0.038)	(0.020)	(0.017)
factor(villages)4	−0.042	−0.120**	0.048	0.025
	(0.054)	(0.038)	(0.026)	(0.020)
factor(villages)5	0.016	−0.119**	0.064*	0.058
	(0.085)	(0.044)	(0.031)	(0.035)
factor(villages)6	−0.068	−0.189***	0.092***	0.029
	(0.047)	(0.036)	(0.025)	(0.028)
factor(villages)7	0.001	−0.108*	0.068**	0.055*
	(0.055)	(0.045)	(0.023)	(0.026)
factor(villages)8	−0.130**	−0.150***	0.011	0.019
	(0.049)	(0.042)	(0.020)	(0.028)
factor(villages)9	0.069	−0.120*	0.138*	0.040
	(0.126)	(0.053)	(0.069)	(0.025)
factor(villages)10	−0.112*	−0.196***	0.048	0.036
	(0.053)	(0.045)	(0.032)	(0.033)
factor(villages)11	−0.119*	−0.212***	0.070*	0.026
	(0.050)	(0.042)	(0.031)	(0.023)
factor(villages)12	−0.037	−0.177***	0.030	0.105**
	(0.060)	(0.049)	(0.025)	(0.039)
factor(villages)13	0.142	−0.034	0.107*	0.057
	(0.137)	(0.092)	(0.045)	(0.032)
factor(villages)14	−0.095	−0.180***	0.067*	0.022
	(0.061)	(0.043)	(0.031)	(0.022)
R^2	0.197	0.319	0.236	0.356
Adj. R^2	0.156	0.285	0.198	0.323
Num. obs.	415	415	415	415
RMSE	0.282	0.154	0.123	0.118

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

Table 8: Q12: Removing Population Outliers

	Model 1	Model 2	Model 3	Model 4
(Intercept)	−0.022	0.046	−0.062	0.030
	(0.162)	(0.085)	(0.075)	(0.057)
treatment	0.008	0.011	−0.004	−0.005
	(0.015)	(0.011)	(0.009)	(0.010)
avgpoverty	0.040	0.042**	−0.004	−0.006
	(0.023)	(0.014)	(0.012)	(0.010)
pobtot1994	−0.000	−0.000*	0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)
t1994	0.681**	0.072	0.153	0.314**
	(0.215)	(0.129)	(0.093)	(0.101)
pri1994s	0.201	0.500***	0.056	−0.271**
	(0.216)	(0.140)	(0.072)	(0.096)
pan1994s	0.697**	−0.113	1.121***	−0.194
	(0.249)	(0.194)	(0.166)	(0.161)
prd1994s	0.315	−0.040	−0.054	0.574***
	(0.199)	(0.137)	(0.079)	(0.099)
factor(villages)2	−0.091*	−0.094**	0.020	−0.013
	(0.042)	(0.034)	(0.018)	(0.019)
factor(villages)3	−0.087*	−0.089**	0.030	−0.020
	(0.036)	(0.033)	(0.017)	(0.016)
factor(villages)4	−0.102**	−0.114***	0.025	−0.018
	(0.038)	(0.032)	(0.017)	(0.018)
factor(villages)5	−0.055	−0.123***	0.042*	0.012
	(0.037)	(0.035)	(0.019)	(0.019)
factor(villages)6	−0.090*	−0.164***	0.065***	0.009
	(0.040)	(0.033)	(0.019)	(0.019)
factor(villages)7	−0.113**	−0.127***	0.021	0.008
	(0.042)	(0.036)	(0.023)	(0.022)
factor(villages)8	−0.120**	−0.116**	−0.001	0.000
	(0.042)	(0.039)	(0.018)	(0.022)
factor(villages)9	0.009	−0.114*	0.114	−0.004
	(0.105)	(0.044)	(0.060)	(0.022)
factor(villages)10	−0.099*	−0.146***	0.039	0.004
	(0.045)	(0.043)	(0.027)	(0.028)
factor(villages)11	−0.095*	−0.162***	0.063*	0.000
	(0.045)	(0.041)	(0.025)	(0.020)
factor(villages)12	−0.052	−0.147***	0.015	0.069*
	(0.049)	(0.042)	(0.022)	(0.032)
factor(villages)13	0.102	−0.014	0.086**	0.011
	(0.092)	(0.069)	(0.031)	(0.035)
factor(villages)14	−0.136**	−0.157***	0.032	−0.010
	(0.044)	(0.036)	(0.019)	(0.019)
R^2	0.704	0.600	0.505	0.665
Adj. R^2	0.689	0.580	0.480	0.648
Num. obs.	415	415	415	415
RMSE	0.171	0.118	0.099	0.085

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

Table 9: Q13: Using Percentage Controls