5304 Econometrics Autumn 2023[1]

Problem set 2
OLS and bad controls

Part 1: OLS and bad controls in practice

For this assignment you will be using the same data as in PS1, i.e. the excerpt from the American National Election Studies survey (ANES 2020) available in *PS1_ANES2020.dta*. This time we will dig deeper into the demographics of Trump voters in 2020 using regression.

Trump voting and unemployment

1. Clean the data and generate variables in exactly the same way as you did in PS1.
2. Regress the Trump voter dummy on not working in the last week. Interpret the coefficient.
3. Add age as a control variable. Why do you think the coefficient on not working changes the way it does?
4. In PS1 you looked at the relationship between Trump voting and age, which turned out not to be linear. Report specifications that control for 1) a quadratic in age and 2) quintiles of age. Interpret what you find in terms of the functional form of age. *Hint: for quintiles, use "xtile"*
5. How many percentage points more likely are the second oldest quintile (4th) of respondents to vote Trump compared to the second youngest (2nd), conditional on the not working dummy? Test the hypothesis that the share of Trump voters in these age groups are equal and report the p-value (use robust standard errors). Do the same for the 5th versus the 1st age quintiles (for very small p-values, interpret the value as virtually zero).
6. Choose one of the specifications including a non-linear function of age and include the dummy for being white. Write down the regression specification, and run the regression in Stata. Does the estimated coefficient on not working change, and if so, what does the change imply for the correlation between being white and not working (conditional on age)? It is enough to comment on the sign of the correlation and interpreting what this means.

Trump voting and education

7. Regress Trump voting on the higher education dummy. Why is the estimated coefficient unlikely to have a causal interpretation? Mention several possible sources of bias.
8. Let's disregard the problems with omitted variable bias you have just brought up. A friend claims that an interesting link between higher education and Trump voting is that education makes people more interested in factual policy, which makes them less likely to vote for Trump no matter their political views.[2] However, higher education may also affect people's political views directly due to them being in a particular social milieu. Your friend asks you to isolate the 'factual policy' channel between education and Trump voting, holding any potential effect via political ideology constant. Regress Trump voting on high education and a dummy variable that equals one if feeling towards conservatives exceeds 50 and zero otherwise, as a proxy for being conservative.

---

[1] TA: Petter Berg, petter.berg@phdstudent.hhs.se

[2] This is just an example that may or may not be true in reality.

9. Why is the specification estimated in the previous question problematic (again, disregarding the problems with omitted factors that you have already discussed)? Explain. *Hint: is the Conditional Independence Assumption likely to hold?*

Part 2: Interpreting published results

In this part of the assignment, we will be looking at an experiment on gender differences in willingness to lead a group exercise conducted in Born et al. (2022). In the experiment, participants get randomly allocated to male- or female-majority teams. The participants get to do two similar tasks, one individually and one in group. Before the group exercise, they get to choose a group leader and willingness to lead is elicited. The main results are shown in Table 1 of the paper, reproduced below.

A friend needs your help in interpreting the table – your job is to find the estimates answering the questions. Disregard whether or not they are statistically significant.

1. What is the gender gap in willingness to lead the group exercise among men and women of equal performance in the individual (first) task, unconditional on team gender composition?

2. How much more are men willing to lead than women when assigned to male-majority teams, unconditional on performance in the individual (first) task?

3. Conditioning on the full set of variables available, are men who perform the *worst* on the individual task more willing to lead than the women who perform the *best*? Is this true for both male- and female-majority teams?

Table 1: Differences in willingness to lead across gender and team composition

	Dependent variable: Willingness to lead (1-10)			
	(1)	(2)	(3)	(4)
Male	1.633***	1.584***	1.778***	1.417*
	(0.261)	(0.262)	(0.361)	(0.600)
Male-majority team			-1.386***	-1.355***
			(0.403)	(0.402)
Male X Male-majority team			0.713	0.706
			(0.593)	(0.588)
Relative performance first task		-0.269*		-0.319
(1=best, 4=worst)		(0.117)		(0.171)
Male X Relative performance first task				0.127
				(0.203)
Constant	5.633***	6.303***	6.005***	6.792***
	(0.174)	(0.345)	(0.188)	(0.465)
N	580	580	580	580
***F*-test:** Male-majority team +			-0.673	-0.650
'Male X Male-majority team':			(*F*=3.586)	(*F*=3.380)

Note: OLS regressions. Standard errors clustered at the team level. The coefficient for *Male-majority team* indicates the treatment effect on women, while the final row shows the treatment effect on men. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

References

ANES (2020): "ANES 2020 Time Series Study," *American National Election Studies. University of Michigan, Stanford University.*

Born, A., E. Ranehill, and A. Sandberg (2022): "Gender and Willingness to Lead: Does the Gender Composition of Teams Matter?" *The Review of Economics and Statistics*, 104, 259–275.