Econ 300 Spring 2020
Problem set 2
Due 2/19

Instructions: Please write/type your answers NEATLY and hand in a hard copy before class starts. If you choose to write the solution by hand, please write in legible writing and do not try to fit answers in weird spaces.

The total points for this problem set is 100 and a further 10 points are available as a bonus. For STATA questions, paste codes/screenshot (It is very important to show your STATA work).

Please do not copy answers from each other or directly from the internet. It is okay to discuss the problems with classmates but answers must be written in your own words.

Question 1 (10 points). In order to improve performance, Chicago public schools decide to give bonus salary to teachers whose students achieve top 20 percentile of a standardized test. In order to evaluate if this bonus salary is effective in raising students' performance, one researcher runs the following regression:

Avg_student_score=β0+β1Receive_Bonus+ε

a. Give two examples of omitted variables that may bias $\widehat{\beta_1}$

b. Explain the direction of bias and the plausible stories leading to the bias

Question 2 (30 points (+5 bonus)). You are studying the causal effect of experience on wages (i.e. return to experience). ttl_exp is total work experience. You run the following model:

Wage=β0+β1total_work_experience+ε (1)

a. Interpret the coefficient for total_work_experience in this model 1.

Suppose instead that you run the following model 2 now:

Wage=β0+β1 total_work_experience
+ β2age+ β3race+ β4south+ β5industry+ β6grade+ε (2)

where race is a dummy variable equal to 1 for Whites, south refers to a dummy variable equals to 1 for individuals living in southern states, industry is a dummy variable for white-collar jobs, and grade picks up the highest year of schooling for the individual. Note: Wage is measured as hourly wage rate, and experience is measured in years.

The Stata output is shown below:

```
. reg wage ttl_exp age race south industry grade
```

Source	SS	df	MS		
Model	6985.43122	6	1164.23854	Number of obs =	1191
Residual	38033.0728	1184	32.1225277	F(6, 1184) =	36.24
				Prob > F =	0.0000
Total	45018.504	1190	37.8306756	R-squared =	0.1552
				Adj R-squared =	0.1509
				Root MSE =	5.6677

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ttl_exp	.2845358	.0366744	7.76	0.000	.2125817	.3564899
age	-.1219587	.054548	-2.24	0.026	-.2289803	-.0149371
race	-.4190111	.352757	-1.19	0.235	-1.11111	.2730875
south	-1.176419	.3408898	-3.45	0.001	-1.845234	-.5076036
industry	-.1641742	.0581562	-2.82	0.005	-.278275	-.0500735
grade	.6488261	.0724276	8.96	0.000	.5067254	.7909268
_cons	3.035072	2.443292	1.24	0.214	-1.758594	7.828737

b. Interpret the estimate for total_work_experience in this model. How is the interpretation here different from model in part a?

c. Write out the null and alternative hypotheses to test whether there is any relationship between wage and experience.

d. Use any method explain whether you reject the null hypothesis specified in c. ($t\alpha/2$ for $\alpha=0.01$, n=1184 is 2.58)

e. What is the null hypothesis for the F-test that all independent variables have no explanatory power in wage? Do you reject or fail to reject the null based on the Stata output? Explain.

f. Explain in words what R-square means in this example. Does big R-square necessary mean better model for this research project?

g. (Bonus worth 10 points) Explain why adding additional variables such as *race*, *south*, *industry* and *grade* in model 2 might be a good idea as compared to model 1?

Question 3 (50 points). (Stata application)
Load the data nlsw_ps2.dta. Our research question is to examine whether being in the union affects one's wage. For the following questions that require STATA commands, you can either paste the STATA output or write/type the key results.

a. Before getting to the data, what is your prior expectation? Do you think being in the union has a positive, negative or no effect on one's wage?

b. Run a simple regression of hourly wage on one's union status. Write down/paste your codes.

c. What is the coefficient and standard error of union status? Interpret the meaning of the coefficient and its standard error.

d. Look at its t-stat or p-value, do you think there is a relationship between union status and wage? Explain.

e. Now based on the confidence interval, do you think there is a relationship between union status and wage? Explain

f. Using the estimated regression, let's make predictions for wage given each person's union status. What is the predicted wage for someone in the union?

g. Now include two additional variables race and age in your regression. Write down/paste the codes. What is the coefficient and standard error for union?

h. What values are the ESS and TSS from the STATA output? Explain what they mean in this example.

i. Compare the R-square with adjusted R-square, which one do you trust more and why?

j. Compare the simple regression with the multiple regression. Which model have more explanatory power in explaining the variation in wage? Why?