

Econ 312: Problem Set #3

Solutions

Question.1 A researcher plans to study the causal effect of police on crime using data from a random sample of U.S. counties. He plans to regress the county's crime rate on the (per capita) size of the county's police forces.

- a Explain why this regression is likely to suffer from omitted variable bias. Which variables would you add to the regression to control for important variables?

Answer: There are other important determinants of a country's crime rate, including demographic characteristics of the population.

- b Use you answer to (a) and the expression for omitted variable bias to determine whether the regression will likely over- or underestimate the effect of police on the the crime rate. (That is, do you think that $\hat{\beta}_1 > \beta_1$ or $\hat{\beta}_1 < \beta_1$?)

Answer: Suppose that the crime rate is positively affected by the fraction of young males in the population, and that counties with high crime rates tend to hire more police. In this case, the size of the police force is likely to be positively correlated with the fraction of young males in the population leading to a positive value for the omitted variable bias so that $\hat{\beta}_1 > \beta_1$

Question.2 Using the data set CollegeDistance, carry out following exercises.

```
. regress yrsed dist
```

Source	SS	df	MS	Number of obs =	943
Model	34.8194696	1	34.8194696	F(1, 941) =	12.44
Residual	2634.83058	941	2.8000325	Prob > F =	0.0004
				R-squared =	0.0130
				Adj R-squared =	0.0120
Total	2669.65005	942	2.83402341	Root MSE =	1.6733

yrsed	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
dist	-.0676604	.0191869	-3.53	0.000	-.1053144	-.0300063
_cons	13.86432	.0679604	204.01	0.000	13.73094	13.99769

```
. regress yrsed dist bytest female black incomehi ownhome dadcoll cue80 stwmfg8 > 0
```

Source	SS	df	MS	Number of obs =	943
--------	----	----	----	-----------------	-----

-----+-----				F(9, 933) = 29.08		
Model		584.86284	9	64.9847601	Prob > F = 0.0000	
Residual		2084.78721	933	2.23449862	R-squared = 0.2191	
-----+-----				Adj R-squared = 0.2115		
Total		2669.65005	942	2.83402341	Root MSE = 1.4948	
-----+-----						
yrsed		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----						
dist		-.0517721	.0200209	-2.59	0.010	-.0910632 -.0124809
bytest		.0735757	.0063303	11.62	0.000	.0611524 .0859989
female		.0556881	.0991044	0.56	0.574	-.1388052 .2501814
black		.0946566	.2119147	0.45	0.655	-.3212281 .5105412
incomehi		.4313288	.1159349	3.72	0.000	.2038054 .6588523
ownhome		.1868644	.1302678	1.43	0.152	-.0687875 .4425163
dadcoll		.530195	.1251433	4.24	0.000	.2846 .77579
cue80		.0439668	.0249415	1.76	0.078	-.0049812 .0929148
stwmfg80		-.0424569	.0401261	-1.06	0.290	-.1212047 .036291
_cons		9.748824	.5055287	19.28	0.000	8.756718 10.74093
-----+-----						

a Run a regression of years of completed education (*ED*) on distance to the nearest college (*Dist*). What is the estimated slope?

Answer: -0.068

b Run a regression of *ED* on *Dist*, but include some additional regressors to control characteristics of the students, the student's family, and the local labor market. In particular, includes as additional regressors *Bytest*, *Female*, *Blace*, *Incomehi*, *Ownhome*, *DadColl*, *Cue80*, and *Stwmfg80*. What is the estimated effect of *Dist* on *ED*?

Answer: -0.052

c Is the estimated effect of *Dist* on *ED* in the regression in (b) substantively different from the regression in (a)? Based on this, does the regression in (a) seem to suffer from important omitted variable bias?

Answer: The coefficient has fallen by more than 24%. Thus, it seems that result in (a) did suffer from omitted variable bias.

d Compare the fit of the regression in (a) and (b) using the regression standard errors, R^2 and \bar{R}^2 . Why are the R^2 and \bar{R}^2 so similar in regression (b)?

Answer: The regression in (b) fits the data much better as indicated by the R^2, \bar{R}^2 and SER. The R^2 and \bar{R}^2 are similar because the number of observations is large (n= 943).

e The value of the coefficient on *DadColl* is positive. What does this coefficient measure?

Answer: Students with a "dadcoll= 1" (so that the student's father went to college) complete

0.530 more years of education, on average, than students with “dadcoll=0” (so that the student’s father did not go to college).

f Explain why *Cue80*, and *Stwmfg80* appear in the regression. Are the signs of their estimated coefficients (+ or -) what you would have believed? Interpret the magnitudes of these coefficients.

Answer: These terms capture the opportunity cost of attending college. As *STWCFG* increases, forgone wages increase, so that, on average, college attendance declines. The negative sign on the coefficient is consistent with this. As *CUE80* increases, it is more difficult to find a job, which lowers the opportunity cost of attending college, so that college attendance increases. The positive sign on the coefficient is consistent with this.

g Bob is a black male. His high school was 20 miles from the nearest college. His base-year composite test score (*Bytest*) was 58. His family income in 1980 was \$26,000, and his family owned a home. His mother attended college, but his father did not. The unemployment rate in his county was 7.5%, and the state average manufacturing hourly was \$9.75. Predict Bob’s years of completed schooling using the regression in (b).

Answer: Bob’s predicted years of education = $9.749 - 0.0518 \times 2 + 0.0736 \times 58 + 0.0557 \times 0 + 0.0947 \times 1 + 0.4313 \times 1 + 0.1867 \times 1 + 0.53 \times 0 + 0.044 \times 7.5 - 0.0425 \times 9.75 = 14.543$

h Jim has the same characteristics as Bob except that his high school was 40 miles from the nearest college. Predict Jim’s years of completed schooling using the regression in (b)?

Answer: Jim’s expected years of education is $2 \times 0.0518 = 0.1036$ less than Bob’s. Thus, Jim’s expected years of education is 14.439