

First Midterm (practice with solution)

Econometrics 410

Thursday, **Oct. 7**

1. True or False (15min) (Answers without reasoning receive no credit.)
 - (a) If Assumption MLR.5 (Homoskedasticity) does not hold, then the ordinary least square estimators are biased.
[A:] False. The OLS estimators are unbiased under Assumptions MLR.1-4 (linearity, random sampling, no perfect multicollinearity and zero-conditional mean). Homoskedasticity is not necessary for unbiasedness to hold.
 - (b) Because omitted variables cause bias, it is always recommended to include all available explanatory variables in a regression.
[B:] First of all, omitted variables do not always cause bias. Omitting a variable only causes bias when the omitted variable is correlated with the included variables. Second, there are reasons that some variables should not be included in a regression. One reason is that some variable may be an intermediate variable between the important explanatory variable and the dependent variable.
 - (c) R -squared always increases when we add variables to a regression.
[A:] Strictly speaking, it is true if "increases" is replaced with "does not decrease". R -squared is the proportion of variation in the dependent variable that is explained by the model. When we add variables to regression, at least the same amount of variations is explained as before.
2. (10min) Consider the linear regression model:

$$\log(wage) = \beta_0 + \beta_1 Math + \beta_2 Writing + \varepsilon$$

And you would like to test the hypothesis that math score and writing score have the same effect on $\log(wage)$. That is, you want to test $H_0 : \beta_1 = \beta_2$. We discussed a way to do this in STATA by running a modified regression.

- (a) What is the modified regression to run?

[A:]

$$\begin{aligned}\log(\text{wage}) &= \beta_0 + \beta_1 \text{Math} + \beta_2 \text{Writing} + \varepsilon \\ &= \beta_0 + (\beta_1 - \beta_2) \text{Math} + \beta_2 (\text{Writing} - \text{Math}) + \varepsilon\end{aligned}$$

We can regress $\log(\text{wage})$ on Math and $(\text{Writing} - \text{Math})$.

- (b) How do you use the modified regression to test the hypothesis?

[A:] We can test H_0 by checking the significance of the coefficient on Math in the modified regression.

3. The econometrician estimated the following regression model:

$$lwage_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 expersq_i + U_i, \quad (1)$$

where $lwage$ denotes the natural logarithm of wage, $educ$ is education, $exper$ is working experience in years.

- (a) (5 min) What is the meaning of β_1 in equation (1)?

[A:] If Assumptions MLR.1-4 are satisfied for the regression model (1), then β_1 can be interpreted as the effect of education on $\log(\text{wage})$: one more year of education increases wage by $\beta \times 100\% = 100\beta\%$.

- (b) (10 min) What is the marginal effect of experience on wage in equation (1)?

[A:] $\frac{\partial lwage}{\partial exper} = \beta_2 + 2\beta_3 exper$. The marginal effect of experience on wage when experience is $exper$ is $\beta_2 + 2\beta_3 exper$.

- (c) (20 min) The following output for equation (1) was obtained in Stata, where several entries were replaced with letters. Find A-G. For example:

$$\begin{aligned}H &= \frac{\text{lower end of 95\%confidence interval} + \text{upper end of confidence interval}}{2} \\ &= \frac{0.0308002 + 0.0512175}{2} = 0.041009.\end{aligned}$$

If you don't have a calculator, show all but the final calculating step. For example, in the above equation, you get full credit if you show all steps but " $=0.041009$ ".

Source	SS	df	MS	Number of obs = 526		
-----+-----				F(3, 522) = D		
Model	44.5393713	A	14.8464571	Prob > F	=	0.0000
Residual	103.79038	B	.198832146	R-squared	=	E
-----+-----				Adj R-squared = 0.2963		
Total	C	525	.28253286	Root MSE	=	.44591

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
educ	.0903658	.007468	F	0.000	G	.1050368
exper	H	.0051965	7.89	0.000	.0308002	.0512175
exper2	-.0007136	.0001158	-6.16	0.000	-.000941	-.0004861
_cons	.1279975	.1059323	1.21	0.227	-.0801085	.3361035

[A:]

$$A = \#regressors = 3;$$

$$B = n - A - 1 = 526 - 3 - 1 = 522$$

$$C = SST = SSE + SSR = 44.5393713 + 103.79038 = 148.33$$

$$D = F = \frac{SSE/A}{SSR/B} = \frac{44.5393713/3}{103.79038/522} = 74.668$$

$$E = \frac{SSE}{SST} = \frac{44.5393713}{148.33} = 0.30027$$

$$F = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{.0903658}{.007468} = 12.1$$

$$G = \hat{\beta}_1 - se(\hat{\beta}_1) \cdot 1.96 = .0903658 - .007468 \times 1.96 = 0.075729$$

Note that when computing G we use the critical value from the standard normal distribution. This is OK because $n = 526$ is large.