Final Exam, 2019-S1, Marking Guide

Question 1 (1 mark for each correct answer. No partial credit)

1. (c)
2. (d)
3. (d)
4. (b)
5. (c)
6. (e)
7. (a)
8. (d)
9. (e)
10. (a)
11. (a)
12. (d)
13. (d)
14. (a)
15. (b)

Question 2 (15 marks)

2.a. There are many ways to answer this question. One way is to say that:

$$\begin{aligned}\mathbf{X}'\hat{\mathbf{u}} &= \mathbf{0} \Rightarrow \sum_{i=1}^{n} \hat{u}_i = 0 \Rightarrow \\ \sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0\right) &= 0 \Rightarrow \sum_{i=1}^{n} y_i = n\widehat{\beta}_0 \Rightarrow \widehat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i.\end{aligned}$$

Another way is to use the OLS formula directly:

$$\begin{aligned}\mathbf{X} &= \begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix} \Rightarrow \\ \widehat{\beta}_0 &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\end{aligned}$$

(3 marks)

2.b.

$$\widehat{UNDER5} = \underset{(2.1)}{17.2} + \underset{(3.8)}{59.6}\ SUBSAHARA \quad (1)$$

$$\widehat{UNDER5} = \underset{(14.5)}{159.0} - \underset{(2.2)}{7.2}\ \log(GDPPC) - \underset{(0.1)}{0.6}\ SANITATION - \underset{(0.1)}{0.2}\ WATER \quad (2)$$

i. From the information provided, compute the average under-5 mortality rate (a) for the 35 sub-Saharan countries, (b) for the remaining 86 countries, and (c) for all 121 countries in this sample.

$$\begin{aligned} (a) &: 17.2 + 59.6 = 76.8 \text{ [1 mark]} \\ (b) &: 17.2 \text{ [1 mark]} \\ (c) &: \frac{1}{121}(35 \times 76.8 + 86 \times 17.2) = 34.44 \text{ [1 mark]} \end{aligned}$$

(3 marks)

ii. Keeping the proportions of population with access to drinking water and basic sanitation constant, a one percent increase in GDP per capita decreases under 5 mortality rate in 1000 live births by 0.072. (no marks if keeping access to water and sanitation constant is missing, even if the rest is correct. 0.5 mark off if it is said "all else constant" without specific reference to what "all else" are.

(2 marks)

iii. This can be answered in two ways, either using an F-test, or a t-test. For the F-test the restricted and the unrestricted models need to be mentioned, and the restricted model needs to be formulated (1 mark):

$$UNDER5 = \beta_0 + \beta_1 \log(GDPPC) + \beta_{WATER}(SANITATION + WATER) + u$$

For the t-test, the reparameterised model needs to be specified (1 mark)

$$\begin{aligned} \delta &= \beta_{SANITATION} - \beta_{WATER} \Rightarrow \beta_{SANITATION} = \delta + \beta_{WATER} \\ UNDER5 &= \beta_0 + \beta_1 \log(GDPPC) + \delta SANITATION + \beta_{WATER}(SANITATION + WATER) + u \end{aligned}$$

$$\left.\begin{array}{l} H_0 : \beta_{WATER} = \beta_{SANITATION} \quad (\Leftrightarrow \delta = 0) \\ H_1 : \beta_{WATER} \neq \beta_{SANITATION} \quad (\Leftrightarrow \delta \neq 0) \end{array}\right\} \text{(1 mark)}$$

$$\left.\begin{array}{l} F = \frac{(SSR_r - SSR_{ur})/1}{SSR_{ur}/(121-3-1)} \sim F_{1,117} \quad \text{under } H_0 \\ (\text{or } t_{\hat{\delta}} = \frac{\hat{\delta}}{se(\hat{\delta})} \sim t_{117} \quad \text{under } H_0) \end{array}\right\} \text{(1 mark)}$$

$$\left.\begin{array}{l} F_{crit} \approx 3.95 \text{ (OK if they use } F_{1,120} \text{ cv of 3.92)} \\ \text{Reject if } F_{calc} > F_{crit}, \text{ don't reject otherwise} \end{array}\right\} \text{(1 mark)}$$

$$\text{or} \quad \left.\begin{array}{l} t_{crit} \approx 1.987 \text{ (OK if they use } t_{120} \text{ cv of 1.980)} \\ \text{Reject if } t_{calc} > t_{crit} \text{ or } t_{calc} < -t_{crit}, \text{ don't reject otherwise} \end{array}\right\}$$

(4 marks)

iv.

$$\left.\begin{array}{l} H_0 : \beta_{SUBSAHARA} = 0 \\ H_1 : \beta_{SUBSAHARA} > 0 \end{array}\right\} \text{(1 mark)}$$

$$\left.\begin{array}{l} t_{\hat{\beta}_{SUBSAHARA}} = \frac{\hat{\beta}_{SUBSAHARA}}{se(\hat{\beta}_{SUBSAHARA})} \sim t_{116} \quad \text{under } H_0 \\ t_{calc} = \frac{18.2}{5.8} = 3.14 \\ t_{crit} \approx 1.662 \text{ (OK if they use } t_{120} \text{ cv of 1.658)} \end{array}\right\} \text{(1 mark)}$$

We reject H_0 and conclude that even after controlling for GDP per capita, and access to water and sanitation, there is still a statistically significant difference between the mortality rate in sub-Saharan Africa and the rest of the world. (1 mark).

(3 marks)

Question 3 (15 marks)

3.a. Mathematically R^2 will never decrease (OK if they say it will always increase. Although it is not technically correct, but in practice it R^2 will never stays the same) as we add explanatory variables, even if those explanatory variables are completely unrelated to y. Hence R^2 cannot be used for model selection. (1 mark). The alternative is to use one of the adjusted R^2, AIC, BIC, HQ. They need to write the formula, and explain how it is used (choose model with the largest for the adjusted R^2, or choose the model with the smallest for the other three) (1 mark). They also need to refer to how their criterion takes into account goodness of fit, and penalises larger models.(1 mark)

(3 marks)

3.b.

$$\log(WAGE) = \beta_0 + \beta_1\ (EDUC - 12) + \beta_2\ EXPER + \beta_3\ EXPER^2 + u \tag{3}$$

We have estimated the following regression using OLS:

$$\begin{aligned} \log(\widehat{WAGE}) &= \underset{(0.066)}{2.837} + \underset{(0.010)}{0.095}\ (EDUC - 12) + \underset{(0.009)}{0.055}\ EXPER - \underset{(0.0003)}{0.001}\ EXPER^2 \quad (4) \\ R^2 &= 0.394,\ \text{standard error of the regression} = 0.420,\ n = 249 \end{aligned}$$

i. 2.837 is the predicted $\log(WAGE)$ for a person with 12 years of education and no experience. This translates to $WAGE = e^{2.837} = \$17.064$ (1 mark)
0.095 : keeping experience constant, an extra year of education increases the predicted wage by $100 * \left(e^{0.095} - 1\right) = 9.97$ percent. OK if they say $100 * 0.095 = 9.5$ percent.(1 mark)
0.055 and 0.001 are not interpretable by themselves, but together they tell us that keeping education constant

$$\frac{\partial \log(wage)}{\partial EXPER} = 0.055 - 0.001 * 2EXPER,$$

$\log(wage)$ reaches a maximum when a person gains

$$\frac{0.055}{2 * 0.001} = 27.5$$

years of experience, i.e. % change in wage as experience increases in positive before this age and then turns negative after this age.(2 marks - no need to have all points explained above in their answer to get the full 2 marks)

(4 marks)

ii. No. In order to do so, we need to be confident that all factors that can affect both education and wage are controlled for. Intellectual ability affects both wage and education, and experience cannot control for intellectual ability. Hence the coefficient of $(EDUC - 12)$ may reflect the effect of intellectual ability as well as education (i.e. suffer from omitted variable bias) - no need to mention the phrase "omitted variable bias" to get full marks if the explanation is clear.

(2 marks)

iii.

$$\left.\begin{array}{l} Var(u \mid EDUC, EXPER) = \gamma_0 + \gamma_1 (EDUC - 12) + \gamma_2\ EXPER \\ H_0 : \gamma_1 = \gamma_2 = 0 \\ H_1 : \text{ at least one of the above is not zero} \end{array}\right\} \text{(1.5 mark)}$$

$$\left.\begin{array}{l} BP = n \times R^2_{aux} \overset{a}{\sim} \chi^2_2 \quad \text{under } H_0 \\ BP_{calc} = 249 \times 0.039 = 9.711 \\ BP_{crit} = 5.99 \end{array}\right\} \text{(1.5 mark)}$$

$$\left.\begin{array}{l} BG_{calc} > BG_{crit} \Rightarrow \text{ We reject the null} \\ \text{We conclude that the errors are heteroskedastic...} \end{array}\right\} \text{(1 mark)}$$

(4 marks)

iv. Dividing both side of the model by $\sqrt{EXPER}$ (i.e. multiplying both sides by $\frac{1}{\sqrt{EXPER}}$) does it. The transformed model will be

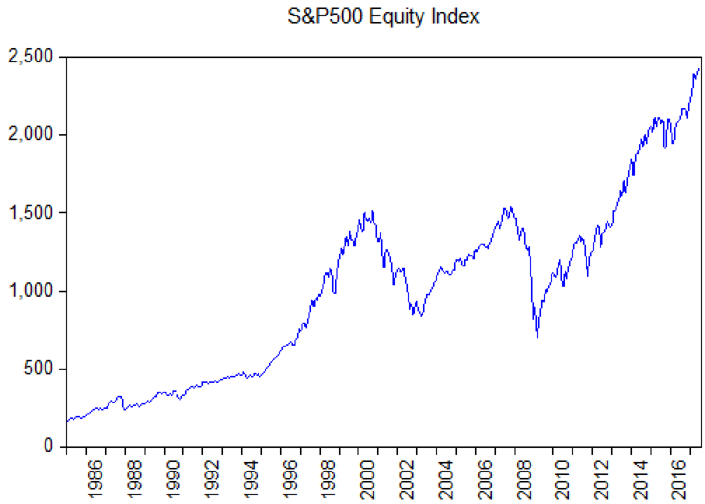
$$\frac{\log(WAGE)}{\sqrt{EXPER}} = \beta_0 \frac{1}{\sqrt{EXPER}} + \beta_1 \frac{(EDUC-12)}{\sqrt{EXPER}} + \beta_2 \frac{EXPER}{\sqrt{EXPER}} + \beta_3 \frac{EXPER^2}{\sqrt{EXPER}} + \frac{u}{\sqrt{EXPER}} (1$$

$$Var\left(\frac{u}{\sqrt{EXPER}} \mid EDUC, EXPER\right) = \sigma^2 \text{ (1 mark)}$$

(2 marks)

Question 4 (15 marks)

4.a. In finance, it is generally believed that asset prices are non-stationary time series. The following graph shows a time series plot of the Standard and Poor's U.S. Composite Stock Price Index (S&P500) from January 1985 to July 2017.



i. $\{y_t\}$ is covariance stationary if its mean (1 mark) and variance (1 mark) are finite and do not depend on t, and the covariance between y_t and any of its lags only depend on the lag and does not depend on t.(1 mark)

(3 marks)

ii. It is clear from the graph that the mean depends on t. It is very unlikely that the mean in the first half of the sample is the same as the mean in the second half.

(1 mark)

4.b. In order to forecast the number of international visitors to Victoria, we postulated the following model:

$$\log(VIC_t) = \beta_0 + \beta_1 \ t + \beta_2 \ Q1_t + \beta_3 \ Q2_t + \beta_4 \ Q3_t + \beta_5 \ AUD_{t-1} + u_t \tag{5}$$

$$\begin{aligned}\log(\widehat{VIC_t}) &= \underset{(0.037)}{10.71} + \underset{(0.0002)}{0.017} \ t - \underset{(0.017)}{0.027} \ Q1_t - \underset{(0.017)}{0.364} \ Q2_t - \underset{(0.017)}{0.300} \ Q3_t - \underset{(0.050)}{0.370} \ AUD_{t-1} \\ n &= 109, \ R^2 = 0.987\end{aligned}$$

i. The intercepts in quarters 1, 2 and 3 are smaller than the intercept in the 4th quarter (they may say that all else equal, the tourist arrivals in Q1, Q2 and Q3 are lower than visitor arrivals in the 4th quarter). The sign of coefficient of AUD shows that after taking the influence of trend and seasonal variations away, tourist arrivals are predicted to decrease as the value of Australian dollar in the previous quarter increases.

(2 marks)

ii. This hypothesis implies that the intercept in Q1 and Q4 must be equal, and the intercept in Q2 and Q3 must be equal. In the current model, this translates to

$$\left.\begin{array}{l} H_0 : \beta_2 = 0 \text{ and } \beta_3 = \beta_4 \\ H_1 : \text{at least one of these restrictions is false} \\ F = \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(109-6)} \sim F_{2,103} \qquad \text{under } H_0 \end{array}\right\} \text{(1.5 marks)}$$

The restricted model is:

$$\log(VIC_t) = \beta_0 + \beta_1\, t + \beta_3\, (Q2_t + \ Q3_t) + \beta_5\, AUD_{t-1} + u_t \text{ (1.5 marks)}$$

Estimating the restricted model gives us SSR_r, and we get the SSR_{ur} from the estimated model in equation (5). Based on these we get F_{calc}. We reject the null if $F_{calc} > F_{crit}$, where F_{crit} is the 95th percentile of the $F_{2,103}$ degrees of freedom (or alternatively they may express this using the F_{crit} from the F table. (1 mark)

(4 marks)

iii.

Correlogram of residuals of the estimated model
Sample: 1991Q1 2018Q2
Included observations: 109

Autocorrelation	Partial Correlation		AC	PAC
		1	0.583	0.583
		2	0.441	0.153
		3	0.358	0.079
		4	0.328	0.093
		5	0.223	-0.061
		6	0.122	-0.083
		7	0.012	-0.120
		8	-0.003	0.005
		9	-0.005	0.033
		10	-0.097	-0.112
		11	-0.127	-0.020
		12	-0.049	0.122

iii.a. The OLS estimator will be unbiased, but will no longer be BLUE (1 mark). The standard errors estimated using the usual formula for the OLS standard errors will be incorrect (1 mark)

(2 marks)

iii.b. It appears that the errors are serially correlated and it also appears that they are an AR(1) process because only the first order partial autocorrelation seems significant (2 marks). We can improve by adding an equation to the model for its AR(1) errors:

$$\begin{aligned} \log(VIC_t) &= \beta_0 + \beta_1\, t + \beta_2\, Q1_t + \beta_3\, Q2_t + \beta_4\, Q3_t + \beta_5\, AUD_{t-1} + u_t \\ u_t &= \rho u_{t-1} + e_t \end{aligned}$$

and estimating it using feasible GLS [it is OK if they don't name feasible GLS. They may also say by adding AR(1) in EViews equation command, which is OK also] (1 mark)

(3 marks)

END OF EXAMINATION