Econometrics 1 & 2
CEU MA

Sample Questions For Placement Exam

1. Consider the following regression on an *iid* sample:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 z + \beta_4 zx + u$$

   Derive (a) the partial effect and (b) the average partial effect of x on the expected value of y.

2. Suppose that you estimated a simple regression on an *iid* sample and you got the following results:

$$\hat{y}_t = 2 + x_t.$$

   If in the sample, $V(y_t) = 4V(x_t)$, (a) what is the (sample) correlation between the two variables? (b) What is the R^2 of the regression?

3. Is the White standard error estimator consistent when the error term of a linear regression estimated on an *iid* sample is homoskedastic?.

4. If the following regression model is estimated on a sample of size 50 of stationary variables:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + u_t,$$

   and the Durbin Watson statistic is 2.8, are the point estimates and conventional standard error estimates unbiased and consistent?

5. If the following regression model is estimated on a sample of size 50 of stationary variables:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + u_t,$$

   and the Durbin Watson statistic is 1.4, are the point estimates and conventional standard error estimates unbiased and consistent?

6. If the following regression model is estimated on a sample of size 50 on stationary variables:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + u_t,$$

   and the Breusch-Godfrey Serial Correlation LM test statistic (for 3 lagged variables) is 8, are the point estimates and conventional standard error estimates unbiased and consistent?

7. Consider the following regression on a sample of stationary variables:

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1, \varepsilon_t \sim WN\left(0, \sigma^2\right).$$

   If one estimates β_0 and β_1 by OLS but uses the transformed variables $y_t - y_{t-1}$ and $x_t - x_{t-1}$ instead of y_t and x_t, will the conventional standard error estimates be consistent (assume homoskedastic errors)?

8. Consider the following regression on a sample of stationary variables:

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1, \varepsilon_t \sim WN\left(0, \sigma^2\right).$$

If one estimates β_0 and β_1 by OLS but uses the transformed variables $y_t - \rho y_{t-1}$ and $x_t - \rho x_{t-1}$ instead of y_t and x_t, will the conventional standard error estimates be consistent (assume homoskedastic errors)?

9. Consider the following regression on an *iid* sample:

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

Suppose that there is some unmeasured variable w that affects x and also y (at any level of x). Will then x be exogenous? Will the OLS estimator for β_1 be unbiased and consistent?

10. If one were to predict y for different values of x from a simple regression model (where all classical assumptions hold), will the prediction error be the same regardless of the value of ?.

11. Is the confidence interval of a simple regression parameter smaller the larger the variance of the right-hand side variable (suppose that all classical assumptions hold)?

12. Consider a simple regression model where all the classical assumptions hold, but we can measure the left-hand side variable with a classical (i.e. the error is uncorrelated with the true value of y). Is the OLS estimator for the slope parameter going to be consistent (i.e. consistent for the slope parameter of the regression with the true left-hand side variable)?

13. Consider a simple regression model where all the classical assumptions hold, but we can measure the left-hand side variable with a classical (i.e. the error is uncorrelated with the true value of y). Is the OLS estimator for the intercept parameter going to be consistent (i.e. consistent for the intercept parameter of the regression with the true left-hand side variable)?

14. Consider the following regressions estimated on a sample of size 52 of *iid* variables. Suppose that all the classical assumptions hold (including non-stochastic RHS variables and normally distributed error term). The sample sum of squared deviations of y is $S_{yy} = 1000$:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \beta_2 z_i + u_t, \quad R^2 = 0.31 \\ y_i &= \gamma_0 + \gamma_1 x_i + u_t, \qquad\quad R^2 = 0.26 \end{aligned}$$

Can you use the above information to test whether $\beta_2 = 0$? If yes, carry out the test. Can you reject the hypothesis?

15. A regression of earnings on IQ gives you an R^2 of 0.16. (a) What is the implied correlation between the two variables? (b) Assume that IQ approximates intelligence with a classical measurement error. Is the true correlation between intelligence and earnings equal to, smaller or larger than the one between IQ and earnings? (c) Would you rather run IQ on earnings to get a better estimate for the correlation between intelligence and earnings?

16. The following production function was estimated on a cross-sectional sample of firms:

$$\log\left(Y_i\right) = \beta_0 + \beta_1 \log\left(L_i\right) + \beta_2 \log\left(K_i\right) + u_i,$$

where Y_i is output, L_i is labor input, and K_i is capital input. Assume that all classical assumptions hold.
Explain two different methods for testing whether there are constant returns to scale (one should involve estimating tow equations, and the other only one equation). State the adequate null hypotheses and the testing procedure step by step.

17. The following demand equation was estimated on a time-series of one good sold on one market ($n = 100$). The estimated parameters are in the equation, and their appropriately estimated standard error below in parentheses.

$$\log(Q_t) = \underset{(0.20)}{1.5} + \underset{(0.10)}{0.75} \log(P_t) + \underset{(0.40)}{0.6} \log(Y_t) + u_t,$$

where Q_i is quantity sold, P_i is price, and Y_i is income of the consumers. Assume that all classical assumptions hold. In particular, assume that price and income changes were exogenous to demand.
Test whether the demand is price-elastic. Test whether the good is inferior. In each case, state the null hypothesis and show the steps of the testing procedure.

18. Consider the following stochastic process.

$$Y_t = \varepsilon_t + 0.6\varepsilon_{t-1} - 0.1\varepsilon_{t-2}, \quad \varepsilon_t \sim WN\left(0, \sigma^2\right).$$

Draw the Impulse Response Function for $t = 1, 2, 3, 4$. Is the process (covariance) stationary? If yes, calculate its mean, variance, and 1st, 2nd, and 3rd autocorrelation coefficient. If not, can you define a new variable $\tilde{Y}_t$ that is stationary (do not consider the trivial transformation so that $\tilde{Y}_t = \varepsilon_t$).

19. Consider the following stochastic process.

$$Y_t = 1.5Y_{t-1} - 0.5Y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim WN\left(0, \sigma^2\right).$$

Draw the Impulse Response Function for $t = 1, 2, 3, 4$. Is the process (covariance) stationary? If yes, calculate its mean, variance, and 1st, 2nd, and 3rd autocorrelation coefficient. If not, can you define a new variable $\tilde{Y}_t$ that is stationary (do not consider the trivial transformation so that $\tilde{Y}_t = \varepsilon_t$).

20. Consider the following simple regression model:

$$y_i = \alpha + \beta D_i + u_i. \quad D_i = 0 \text{ or } 1.$$

Show that the following estimator is consistent for β :

$$\tilde{\beta} = \bar{y}_1 - \bar{y}_0,$$

where $\bar{y}_1$ is the sample average of y when $D_i = 1$, and $\bar{y}_0$ is the sample average of y when $D_i = 0$.
Bonus problem: Show that $\tilde{\beta}$ is nothing else than the OLS estimator for β.

21. The following regression was estimated on an *iid* sample of employees who are all 20 to 50 years old (standard errors in parentheses):

$$\log(w_i) = \underset{(0.5)}{5} + \underset{(0.02)}{0.1}\, male_i + \underset{(0.01)}{0.05} age_i - \underset{(0.0001)}{0.0005} age_i^2 + u_i,$$

where w is wage, $male = 1$ if male and 0 if female, and age is age measured in years. Answer the following questions:
a) According to the point estimates, what is the average male log wage at age 20?
b) According to the point estimates, what is the average female log wage at age 30?
c) According to the point estimates, what is the male-female difference in earnings at age 20?
d) According to the point estimates, what is the male-female difference in earnings at age 50?
e) According to the point estimates, how much do wages grow at age 20 (expected earnings gains of one more year of age)?
f) According to the point estimates, how much do wages grow at age 50 (expected earnings gains of one more year of age)?
g) Does age have a negative expected effect on wages somewhere in the range of the sample? If yes, at what values?
h) Can you test the hypothesis that men and women earn the same? If yes, can you reject it?
i) Can you test the hypothesis that men earn at least 20% more than women? If yes, can you reject it?

22. Consider the following regression on a sample of countries:

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

where y is economic growth and x is the size of government budget relative to GDP. Suppose that we expect a negative effect of x on y ceteris paribus. Suppose that there is some unmeasured variable w, a measure of corruption that affects x and also y (at any level of x). Is the OLS estimator for β_1 will be consistent? If not, what do you think the direction of the (asymptotic) bias will be and why?

23. We want to model the *probability of employment among all men between 30 and 60,* conditional on age, race, education (in grades completed), and marital status (1 if married, 0 otherwise). The employment rate in the sample is 85%. The parameter estimates from a *probit* model are the following:

$$-2.5 + 0.11 age - 0.0015 age^2 - 0.31 black + 0.1 edu + 0.52 married$$

(a) What is the partial effect of education on the probability of employment for 30 years old unmarried white men with 12 grades of education? $\left(\hat{\beta}' x = 0.65\right)$
(b) What is the partial effect of age on the probability of employment for 30 years old unmarried white men with 12 grades of education? $\left(\hat{\beta}' x = 0.65\right)$
(c) What is the partial effect of education on the probability of employment at the sample mean? $\left(\hat{\beta}' \bar{x} = 1.1\right)$
(d) What is the partial effect of education on the probability of employment at the 90^{th} percentile of the estimated probability? $\left(\hat{\beta}' x = 1.65\right)$

Additional information to the next questions: the estimated probability is between 50% and 97%.
(e) Is the partial effect larger at the top of the predicted probability distribution than at the mean? Why?
(f) Is the partial effect larger at the bottom of the predicted probability distribution than at the mean? Why?

24. Consider the following simple regression model:

$$y_i = \alpha + \beta x_i + u_i, \quad Cov\left(x_i, u_i\right) = \gamma V\left(x_i\right), \gamma > 0$$

Derive the probability limit ($p\lim$) of the OLS estimator for β.
Suppose that you find a variable z such that

$$Cov\left(z_i, u_i\right) = 0.$$

Consider the following estimator for β (it is called the Instrumental Variables, or IV estimator):

$$\hat{\beta}_{IV} = \frac{\frac{1}{n}\sum\left(y_i - \bar{y}\right)\left(z_i - \bar{z}\right)}{\frac{1}{n}\sum\left(x_i - \bar{x}\right)\left(z_i - \bar{z}\right)}.$$

Is the IV estimator consistent for β ? If not, derive its asymptotic bias.
What if $Cov\left(x_i, u_i\right) > Cov\left(z_i, u_i\right) > 0$? Is the IV estimator consistent? If not, can you tell whether the IV or the OLS estimator is more biased (asymptotically)?

25. The following estimates correspond to polyethylene production for a typical manufacturer of the product (conventionally estimated t-statistics in parentheses):

$$\ln C_t = \underset{(8.55)}{1.00} - \underset{(-5.00)}{0.22} \ln\left(n_t\right) - \underset{(-1.20)}{0.06} \ln Y_t \quad \bar{R}^2 = 0.96, \quad DW = 0.84$$

where C_t is the average real cost of production in time period t, n_t is the cumulative number of units of output produced up to time t, and Y_t is period t output.
(a) Based on the point estimates, what is the implied elasticity of the learning curve?
(b) Are the standard error estimates consistent (assume homoskedastic errors)?
(c) Would you estimate anything in a different way? Why and how?