QUANTITATIVE ECONOMICS - FALL 2017
Answer any FIVE questions.

1. Answer the following questions.

   (a) Show that the regression R^2 in the regression of Y on X is the squared value of the sample correlation between X and Y. That is, show that $R^2 = r_{xy}^2$.

   (b) Show that the R^2 from the regression of Y on X is the same as the R^2 from the regression of X on Y.

   (c) Show that $\widehat{\beta}_1 = r_{XY}(s_Y/s_X)$, where r_{XY} is the sample correlation between X and Y, and s_Y and s_X are the sample standard deviations of X and Y.

2. (a) In the following regression model

   $$wage = \beta_1 + \delta_1 female + u$$

   where female is a dummy variable, taking value 1 for female and value 0 for a male. Prove that applying the OLS formulas for simple regression you obtain that

   $$\begin{aligned} \widehat{\beta}_1 &= \overline{wage}_M \\ \widehat{\delta}_1 &= \overline{wage}_F - \overline{wage}_M \end{aligned}$$

   where $\overline{wage}_M$ is the average wage for men and $\overline{wage}_F$ is the average wage for women.

   (b) Suppose that $n = 100$ i.i.d observations for (Y_i, X_i) yield the following results:

   $$\begin{aligned} \widehat{Y} &= \underset{(15.1)}{32.1} + \underset{(12.2)}{66.8}X, \ SER = 15.1, R^2 = 0.81 \end{aligned}$$

   Another researcher is interested in the same regression, but he makes an error when he enters the data into his regression program: He enters each observation twice, so he has 200 observations. Using these 200 observations, what results will be produced by his program? Write it in the following format:

   $$\widehat{Y} = \underset{(\_\_\_)}{\_\_\_} + \underset{(\_\_\_)}{\_\_\_}X, \ SER = \_\_\_, R^2 = \_\_\_$$

3. Consider the following data generating process

   $$Y_t = c + \phi_3 Y_{t-3} + \varepsilon_t$$

   where ε_t is a white noise process. Assuming $0 < \phi_3 < 1$, answer the following:

   (a) Derive the expected value of this process.

   (b) Derive the variance of this process.

   (c) Derive the covariance of this process for $j = 1, 2, 3$.

   (d) Derive the autocorrelation function of this process for $j = 1, 2, 3$.

4. Consider the following model

   $$y_t = c + \phi y_{t-1} + \delta_1 D_{1t} + \delta_2 D_{2t} + \varepsilon_t, \qquad t = 1901, 1902, ..., 2000.$$

   where we have two breaks in the data.

   $$D_{1t} = \left\{ \begin{array}{l} 1 \text{ if } t \leq 1926 \\ 0 \text{ if } t > 1926 \end{array} \right\}$$

   $$D_{2t} = \left\{ \begin{array}{l} 1 \text{ if } t \leq 1975 \\ 0 \text{ if } t > 1975 \end{array} \right\}$$

   Considering the Eviews output below, answer the following questions.

Dependent Variable: Y
Method: Least Squares
Date: 04/22/15 Time: 10:19
Sample (adjusted): 1902 2000
Included observations: 99 after adjustments
Convergence achieved after 5 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.768326	0.056518	13.59436	0.0000
D1	0.939438	0.096269	9.758448	0.0000
D2	0.820134	0.094323	8.694947	0.0000
AR(1)	0.278380	0.097454	2.856532	0.0053

R-squared	0.725224	Mean dependent var	1.219040
Adjusted R-squared	0.716546	S.D. dependent var	0.535710
S.E. of regression	0.285214	Akaike info criterion	0.368412
Sum squared resid	7.727972	Schwarz criterion	0.473265
Log likelihood	-14.23638	Hannan-Quinn criter.	0.410836
F-statistic	83.57856	Durbin-Watson stat	1.783056
Prob(F-statistic)	0.000000		

Inverted AR Roots .28

(a) Explain why the sample in the regression table goes from 1902 to 2000.

(b) In the regression table, what econometric approach is used to estimate this model? Write down the objective function for this estimation method.

(c) If we know the true values of c, ϕ, δ_1 and δ_2 and know that $|\phi|$ is less than one, what is the expected value of the series from 1901 to 1926, from 1927 to 1974, from 1975 to 2000?

(d) Using the estimates from the regression table, give the estimated mean of the series from 1901 to 1926, from 1927 to 1974, from 1975 to 2000?

5. (a) Suppose the true relationship between dependent and the independent variable is

$$Y = X\beta + Z\delta + u$$

but the researcher thinks the correct model is

$$Y = X\beta + u$$

What is the problem that the researcher would run into? Explain in detail.

(b) Suppose now the true relationship is

$$Y = X\beta + u$$

but the researcher estimates the following regression

$$Y = X\beta + Z\delta + u$$

What would be the effect on coefficient estimates?

6. You are given the regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

You want to test the null hypothesis

$$\begin{aligned} H_0 &: \frac{\beta_1}{\beta_3} = \beta_2^2 \\ H_A &: \frac{\beta_1}{\beta_3} \neq \beta_2^2 \end{aligned}$$

Show how you would construct the test statistic.

7. In the least squares regression of $\mathbf{y}$ on a constant and $\mathbf{X}$, to compute the regression coefficients on $\mathbf{X}$, we can first transform $\mathbf{y}$ to deviations from the mean $\overline{y}$ and, likewise, transform each column of $\mathbf{X}$ to deviations from the respective column mean; second, regress the transformed $\mathbf{y}$ on the transformed $\mathbf{X}$ without a constant. Do we get the same result if we only transform $\mathbf{y}$? What if we only transform $\mathbf{X}$?

8. Consider observations (Y_{it}, X_{it}) from the linear panel data model

$$Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it}, \qquad i = 1, \ldots, N, t = 1, \ldots, T.$$

where α_i is an unobserved individual-specific fixed effects, X_{it} is a vector of exogenous regressors and u_{it} are zero mean $i.i.d$ innovations. Describe how you would estimate β_1?