QUANTITATIVE ECONOMICS - SPRING 2019
Answer any FIVE questions.

1. Consider the single equation model
$$y_i = z_i\beta + e_i$$
where y_i and z_i are both real-valued (1×1) and $E(e_i|z_i) \neq 0$. Let $\widehat{\beta}$ denote the IV estimator of β using as an instrument a dummy variable d_i (takes only the values 0 and 1). Derive the explicit expression for the IV estimator.

2. Consider the following data generating process
$$Y_t = \varepsilon_t + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3}$$
where ε_t is a white noise process with $V(\varepsilon_t) = \sigma^2$. Assuming $-1 < \theta_2 < 0$ and $-1 < \theta_3 < 0$, 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 all j. Note that j represents the number of periods between Y_t and Y_{t-j}.
   (d) Derive the autocorrelation function for all j.
   (e) Plot the autocorrelation function with the information you derived above.

3. Consider a product market with a supply function $Q_i^s = \beta_0 + \beta_1 P_i + u_i^s$, a demand function $Q_i^d = \gamma_0 + u_i^d$, and a market equilibrium condition $Q_i^s = Q_i^d$, where u_i^s and u_i^d are mutually independent *i.i.d.* random variables, both with a mean of zero and variance σ_s^2 and σ_d^2, respectively.
   (a) Show that P_i and u_i^s are correlated.
   (b) Show that the OLS estimator of β_1 is inconsistent.
   (c) How would you estimate β_0, β_1, and γ_0?

4. Consider the model
$$y_t = \alpha x_t + u_t;\ t = 1, 2, \dots, T$$
where $E(u_t) = 0; E(u_t^2) = \sigma^2 x_t; E(u_s u_t) = 0$ if $s \neq t$, for all s and t. The density function for u_t is
$$f(u_t) = (2\pi\sigma^2 x_t^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}\left(\frac{u_t}{x_t}\right)^2\right].$$
Derive the maximum likelihood estimators of α and σ^2.

5. Consider the following data generating process
$$Y_t = Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \theta_3\varepsilon_{t-3}$$
where ε_t is a white noise process. Assuming $-1 < \theta_2 < 0$ (and without taking a first difference), answer the following:
   (a) Find the h-step-ahead forecast for Y for $h = 1, 2, ..., k$
   (b) Find the h-step-ahead forecast error for Y for $h = 1, 2, ..., k$
   (c) Find the h-step-ahead forecast error variance for Y for $h = 1, 2, ..., k$

6. (a) Suppose that $n = 100$ i.i.d observations for (Y_i, X_i) yield the following results where SER stands for the standard error of the regression:

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

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:

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

(b) Suppose that a random sample of 200 twenty-year-old-men is selected from a population and that these mens height and weight are recorded. A regression of weight on height yields $\widehat{Weight} = -99.41 + 3.94 \times Height$. Suppose further that instead of measuring weight and height in pounds and inches, these variables are measured in centimeters and kilograms. What are the regression estimates from this new centimeter-kilogram regression? (1 inch= 2.54 cm, 1 pound=0.45 kg)

7. A health economist plans to evaluate whether screening patients on arrival or spending extra money on cleaning is more effective in reducing the incidence of infections by the MRSA bacterium in hospitals. She hypothesizes the following model:

$$MRSA_i = \beta_1 + \beta_2 S_i + \beta_3 C_i + u_i$$

where in hospital i, MRSA is the number of infections per thousand patients, S is expenditure per patient on screening, and C is expenditure per patient on cleaning, u_i is a disturbance term that satisfies the usual regression model assumptions. In particular, u_i is drawn from a distribution with mean zero and constant variance σ^2. The researcher would like to fit the relationship using a sample of hospitals. Unfortunately, data for individual hospitals are not available. Instead she has to use regional data to fit

$$\overline{MRSA}_j = \beta_1 + \beta_2 \overline{S}_j + \beta_3 \overline{C}_j + \overline{u}_j$$

where $\overline{MRSA}_j, \overline{S}_j, \overline{C}_j$, and $\overline{u}_j$ are the averages of $MRSA, S, C$, and u for hospitals in region j. There were different number of hospitals in the regions, there being n_j hospitals in region j.

(a) Show that the variance of $\overline{u}_j$ is equal to $\frac{\sigma^2}{n_j}$ and that an OLS regression using the grouped regional data to fit the relationship will be subject to heteroscedasticity.

(b) Assuming that the researcher knows the value of n_j for each region, explain how she could re-specify the regression model to make it homoscedastic. State the revised specification and demonstrate mathematically that it is homoscedastic.

8. In a fixed effects model, we can consider several ways to control for the individual effects which may be correlated with the regressors. Three cases being the within estimator, least-squares dummy variable estimator (LSDV) and the first difference estimator. Consider the three methods, show that

(a) The within estimator and the LSDV estimator are equivalent.

(b) When T=2, the within estimator and the first difference estimator are equivalent.