Practice Exam 1
Econ471
Fall 2020
Solution

Part I. Multiple Choice Questions

1. The regression model includes a random error or disturbance term for a variety of reasons. Which of the following is NOT one of them?
   a. measurement errors in the observed variables
   b. omitted influences on Y (other than X)
   c. linear functional form is only an approximation
   d. the observable variables do not exactly correspond with their theoretical counterparts
   e. there may be approximation errors in the calculation of the least squares estimates

2. Which of the following is NOT true?
   a. the point $\bar{x}$, $\bar{y}$ always lies on the regression line
   b. the sum of the residuals is always zero
   c. the mean of the fitted values of Y is the same as the mean of the observed values of Y
   d. there are always as many points above the fitted line as there are below it
   e. the regression line minimizes the sum of the squared residuals

3. In a simple linear regression model the slope coefficient measures
   a. the elasticity of Y with respect to X
   b. the change in Y which the model predicts for a unit change in X
   c. the change in X which the model predicts for a unit change in Y
   d. the ratio Y/X
   e. the value of Y for any given value of X

4. Changing the units of measurement of the Y variable will affect all but which one of the following?
   a. the estimated intercept parameter
   b. the estimated slope parameter
   c. the Total Sum of Squares for the regression
   d. R squared for the regression
   e. the estimated standard errors

5. A fitted regression equation is given by Yhat = 20 + 0.75X. What is the value of the residual at the point X=100, Y=90?
   a. 5
   b. -5
   c. 0
   d. 15
   e. 25

6. R squared measures
   a. the correlation between X and Y
   b. the amount of variation in Y
   c. the covariance between X and Y
   d. the residual sum of squares as a proportion of the total sum of squares
   e. the explained sum of squares as a proportion of the total sum of squares

7. In which of the following relationships does the intercept have a real-world interpretation?
   a. the relationship between the change in the unemployment rate and the growth rate of real GDP ("Okun's Law")
   b. the demand for coffee and its price
   c. test scores and class-size
   d. weight and height of individuals

Part II

1. The sample correlation coefficient between two variables X and Y is given by

$$r_{xy} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

a. Show that if one estimates the regressions

$$Y = \beta_1 + \beta_2 X + u$$

and

$$X = \alpha_1 + \alpha_2 Y + v,$$

the product of the estimators for β_2 and α_2 will equal r_{xy}^2.

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{\alpha}_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

$$\Rightarrow \hat{\beta}_2\hat{\alpha}_2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}$$

$$= \frac{[\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})]^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2}$$

$$= \frac{\left[\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})\right]^2}{\left[\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}\right]^2} = \left[\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}\right]^2 = r_{xy}^2$$

b. For a sample of 12 observations on *X* and *Y* the following quantities were calculated

$\overline{X}$ **= 14.25** $\overline{Y}$ **= 79.5** ΣX^2 **= 2501** ΣY^2 **= 79394** ΣXY **= 14007**

Estimate both regression slopes β_2 and α_2, compute r_{xy}^2, and confirm the statement in part a.

$$\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}) = \sum_{i=1}^{n}X_iY_i - n\bar{X}\bar{Y} = 14007-12(14.25)(79.5) = 412.5$$

$$\sum_{i=1}^{n}(X_i-\bar{X})^2 = \sum_{i=1}^{n}X_i^2 - n\bar{X}^2 = 2501-12(14.25)^2 = 64.25$$

$$\sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \sum_{i=1}^{n}Y_i^2 - n\bar{Y}^2 = 79394-12(79.5)^2 = 3551$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} = \frac{412.5}{64.25} = 6.42$$

$$\hat{\alpha}_2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2} = \frac{412.5}{3551} = 0.116$$

$$r_{xy} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} = \frac{412.5}{\sqrt{64.25}\sqrt{3551}} = 0.864$$

Now

$\hat{\beta}_2 \hat{\alpha}_2 = 6.42 \times 0.116 = 0.745$

and

$r_{xy}^2 = (0.864)^2 = 0.745$

$\Rightarrow \hat{\beta}_2 \hat{\alpha}_2 = 0.745 = r_{xy}^2$.

2. Explain why in the simple linear regression model, the regressor *X* must take at least two different values.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

For $\hat{\beta}_1$ to be defined, we need the denominator to be different from zero. This will be the case if the regressor X takes at least two different values.

3. Explain the role of the error term *u* in the simple linear regression model. In the context of an example, explain what factors cause u to exist.

$Y = \alpha + \beta X + u$

There are several reasons for the existence of the error term u. Some of the main reasons are:

1. Omitted variables. In addition to X there are many other factors that affect Y. The error term u takes care of the effects of all omitted variables.

2. Nonlinearity. The true relationship between Y and X may be nonlinear. So if we use a linear regression model, the error term represents nonlinearities in the model.

3. Measurement errors. Errors in measuring Y and X are captured by u.

4. Suggest a transformation in the variables that will linearize the population regression functions below. Write the resulting regression function in a form that can be estimated by using OLS.

(a) $\boldsymbol{Y_i = \beta_0 {X_{1i}}^{\beta_1} {X_{2i}}^{\beta_2}}$

(b) $\boldsymbol{Y_i = \frac{X_i}{\beta_0 + \beta_1 X_i}}$

(c) $\boldsymbol{Y_i = \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}}}$

(d) $\boldsymbol{Y_i = \beta_0 {X_{1i}}^{\beta_1} e^{\beta_2 X_{2i}}}$

(a)

$$\ln(Y_i) = \ln(\beta_0) + \beta_1 \ln(X_{1i}) + \beta_2 \ln(X_{2i})$$

(b)

$$\frac{1}{Y_i} = \beta_0 \frac{1}{X_i} + \beta_1$$

(c)

$$\ln\left(\frac{Y_i}{1 - Y_i}\right) = \beta_0 + \beta_1 X_i$$

(d)

$$\ln(Y_i) = \ln(\beta_0) + \beta_1 \ln(X_{1i}) + \beta_2 X_{2i}$$

5. In the case of perfect multicollinearity, OLS is unable to estimate the slope coefficients of the variables involved. Assume that you have included both $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ as explanatory variables, and that $\boldsymbol{X_2 = X_1^2}$, so that there is an exact relationship between two explanatory variables. Does this pose a problem for estimation? Explain.

There is no problem for estimation, since the second explanatory variable is not linearly related to the first. This is an example of a polynomial regression model of degree 2, which is frequently estimated in econometrics.

6. Consider the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

and suppose that application of least squares to 20 observations on these variables yields the following results

$$\hat{y}_i = \underset{(0.467)}{0.96587} + \underset{(0.2203)}{0.69914} x_{i1} + \underset{(0.1927)}{1.7769} x_{i2}$$

$$\hat{\sigma}^2 = 2.5193 \qquad\qquad R^2 = 0.9466$$

where values inside the parentheses are standard errors of the estimates. Find the total sum of squares, regression (explained) sum of squares, and residual sum of squares.

$$\hat{\sigma}^2 = \frac{SSR}{n-k-1}$$
$$\rightarrow SSR = \hat{\sigma}^2 (n-k-1)$$
$$= 2.5193(20-2-1) = 42.8281$$

$$R^2 = 1 - \frac{SSR}{SST}$$
$$\rightarrow SST = \frac{SSR}{1-R^2}$$
$$= \frac{42.8281}{1-0.9466}$$
$$= 802.02$$

$$SSE = SST - SSR$$
$$= 802.02 - 42.8281 = 759.20$$

7.

a. The estimated coefficient is $\hat{\beta} = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$. Substitute from the true model to obtain

$$\hat{\beta} = \frac{\Sigma x_i(\alpha + \beta x_i + u_i)}{\Sigma x_i^2} = \frac{\alpha \Sigma x_i}{\Sigma x_i^2} + \beta + \frac{\Sigma x_i u_i}{\Sigma x_i^2}$$

The expected value of the third term is zero because $E(u_i)$=0. But $\hat{\beta}$ will be biased because of the first term.

b. The required condition for $\hat{\beta}$ to be unbiased is that $\Sigma x_i = 0$ or that the sample mean of *x* is zero.