The International College of Economics and Finance
Econometrics – 2020-2021.
Midterm exam. 2020 October 22.
Part 2. Free Response Questions
(1 hour 50 minutes + 10 min reading time)

SECTION A

Answer **ALL** questions from this section (questions **1-3**).

Question 1. (17 marks)
Two students developed different models for their course papers: student (a) was convinced that the coefficient of variable Z_i is equal to 1, while student (b) believed that this coefficient is the opposite of the coefficient for the variable X_i

$$Y_i = \beta_1 + \beta_2 X_i + Z_i + u_i \qquad \textbf{(a)}$$
$$Y_i = \beta_1 + \beta_2 X_i - \beta_2 Z_i + v_i \qquad \textbf{(b)}$$

They think that in fact both models will give the same estimates of β_2, but both faced unexpected difficulties when trying to evaluate these models (what difficulties?). So they decided instead to calculate sample variances and covariances on the base of the same data (n observations): $\mathrm{Var}(Y) = 4$, $\mathrm{Var}(X) = 3$, $\mathrm{Var}(Z) = 5$, $\mathrm{Cov}(Y,X) = 6$, $\mathrm{Cov}(Y,Z) = 1$, $\mathrm{Cov}(X,Z) = 2$.

(a) □ Help the students to find the least squares estimates of β_2 for their models, indicating all necessary steps.

□ Are these estimates really the same as students think?

(b) The scientific advisor told the students that both their models are restricted versions of the more general model

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + w_i \qquad \textbf{(c)}.$$

□ What are restrictions in each case?

□ How to test the restriction for the model (a)? Indicate necessary steps.

□ What model should be chosen if both restrictions are invalid? What model (a), (b) or (c) should be chosen if only one restiction is valid?

□ Let both restrictions be valid. Which model out of (a) and (b) is preferable and why? Use numerical data above to choose between (a) and (b).

Question 2. (16 marks)

(a) □ What is R^2? What are its main properties and usage? What are its advantages and disadvantages in econometric analysis.

□ Why some disadvantages of R^2 can be overcomed by using instead adjusted $R^2{}_{adj} = \overline{R}^2 = 1 - \dfrac{RSS/(n-k)}{TSS/(n-1)}$?

□ What are its main properties and usage? What are disadvantages of $\overline{R}^2$ and how they can be overcomed?

(b) □ Prove the following properties of $\overline{R}^2$ (based on the definition of $\overline{R}^2$ in (a) above):

1) $\overline{R}^2 = 1 - (1 - R^2)\dfrac{n-1}{n-k}$, **2)** $\overline{R}^2 = R^2 - \dfrac{k-1}{n-k}(1 - R^2)$.

□ Investigate in what limits $\overline{R}^2$ can vary.

Question 3. (17 marks) The researcher investigates the relation between GDP – variable Y_t and money supply (M1) – variable M_t for Kingdom of Manama in millions of dirhams for a period 1994-2018 (in the beginnig of the year). She tries different models to fit the real data (all coefficients are significant)

$$\hat{Y}_t = 0.337 + 0.08 \cdot M_t \qquad R^2 = 0.89, \quad RSS = 7.0 \qquad (1)$$

$$\hat{Y}_t = -34.02 + 6.1 \cdot \log(M_t) \qquad R^2 = 0.85, \quad RSS = 8.43 \qquad \mathbf{(2)}$$

$$\widehat{\log(Y_t)} = 0.77 + 0.0014 \cdot M_t \qquad R^2 = 0.83, \quad RSS = 0.36 \qquad \mathbf{(3)}$$

(a) □ Give interpretation to the models and their coefficients. Are they significant?

□ Which regressions from (1) - (3) can be compared directly (without additional transformations) in their quality? Compare them and choose better model providing appropriate explanations. Can this comparision be done on the basis of their R^2? Explain.

The regressions (1) and (3) cannot be compared directly (why?). The researcher used Zarembka scaling for the linear regression. After this transformation new variable $YZ = \dfrac{Y}{geometric\ mean} = \dfrac{Y}{(Y_1 \cdot Y_2 \cdot \ldots \cdot Y_n)^{1/n}}$ was regressed on M_t with the following results

$$\widehat{YZ}_t = 0.05 + 0.001 \cdot M_t \qquad R^2 = 0.89, \quad RSS = 0.18 \qquad (4)$$

□ Do Box-Cox test to compare regressions (1) and (3) and choose the better one on the basis of their RSS. Can this comparision be done on the basis of their R^2?

(b) Demonstrate that Zarembka transformation provides approximate comparability of RSS values for the model with the natural logarithm on the left and for the model in which the dependent variable is subject to the Zarembka scaling.

SECTION B.

Answer **ONE** question from this section (**4 OR 5**).

Problem 4. [25 marks]. Consider the following model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad i = 1, \ldots, n$$

together with the fitted model

$$\hat{Y} = b_1 + b_2 X_2 + b_3 X_3$$

Variables X_2, X_3 are assumed to be fixed, assumptions of model A are satisfied.

(a) Show that OLS estimator of b_3 is given by expression

$$b_3 = \frac{\mathrm{Cov}(X_3, Y)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2, Y)\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - [\mathrm{Cov}(X_2, X_3)]^2}$$

(b) Derive the formula of decomposition of the coefficient's estimator into fixed and random components.

$$b_3 = \beta_3 + \left(\frac{\mathrm{Cov}(X_3, u)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2, u)\mathrm{Cov}(X_3, X_2)}{\Delta} \right) \text{ where } \Delta = \mathrm{Var}(X_2)\mathrm{Var}(X_3) - [\mathrm{Cov}(X_2, X_3)]^2$$

(c) □ Prove that OLS estimator b_3 of the coefficient β_3 in multiple regression is unbiased.

Question 5. [25 marks] A lazy student found an interesting paper with econometric models describing how $COVID$ - aggregate anti COVID expendidtures, depend on GDP, aggregate gross national product, and POP, total population, for a sample of 70 countries in the second quarter of 2020. $COVID$ and GDP are both measured in US$ billion. POP is measured in million. *RSS* – Residual Sum of Squares. He decided to use this article in his course paper pretending that he got all equations himself using original data (what in fact was not true). He wrote the equations on a paper:

$$\log\frac{\widehat{COVID}}{POP} = -3.74 + 1.27\log\frac{GDP}{POP} \qquad R^2 = 0.90 \quad RSS = 15.45$$

(1)

$$\log\widehat{COVID} = -3.60 + 1.27\log GDP - 0.33\log POP \qquad R^2 = 0.95 \quad RSS = 13.90 \qquad \textbf{(2)}$$

$$\log\frac{\widehat{COVID}}{POP} = -3.60 + 1.27\log\frac{GDP}{POP} - 0.06\log POP \qquad R^2 = 0.91 \quad RSS = 13.90 \qquad \textbf{(3)}$$

But when he wrote his coursework some details seemed a little strange to him and he began to doubt that he had correctly rewritten the equations on paper. He asked your advice.

(a) □ The student now believes that by mistake he repeated the same coefficient 1.27 in equations (1), (2) and (3), as well as he repeated the intercept -3.60 in equations (2) and (3), but he does not remember the correct values. Are these coincedences really happened by mistake?

□ It seems strange to him that both coefficietns of the variable $\log POP$ in equations (2) and (3) are negative but different in absolute value. Help the student to understand these.

□ The scietific advisor asked the student to add standard errors to all coefficients in equations. Having no records the student decided to write arbitrary numbers as standard errors, but he is afraid that some of them must match, and some not, and if he makes a mistake, then his supervisor immediately recognizes the cheat. Can't you help the student (giving necessary explanations)?

(b) □ *RSS'* are the same in models (2) and (3) while their R^2 are different? Explain.

□ In equations (1) and (3) both R^2 and *RSS'* are different. Explain.

(c) □ The supervisor made it clear to the student that equation (1) is a restricted version of equation (2) and asked the student to find the restriction and test it using an F test, and on this basis to choose the best equation. Can you help?

□ Show the student that the F-test in (b) may well be replaced by the t-test. Is there any advantage to this approach?