Final Exam

(June 21, 2021, **2 hours**)

Econometrics (Spring 2021)

Professor: Wonmun Shin

Part I

True or False

(Total 40 points) Read each statement below carefully. Write a **T** if you think the statement is True. Write an **F** if you think the statement is false. Do not forget to state a reason precisely if you answer F. (Grading guide: If your answer (with a reason in case F) is correct, you will get 5 points. If your answer is partially correct (*i.e.* F without a reason, F with an incorrect reason, etc.), you will get partial points.)

1. In the multiple regression models, testing the individual significance of all regression coefficients (using t-test) and testing the overall significance of all coefficients using R^2 are the same and hence will give the same conclusion.

2. In order to prove Gauss-Markov Theorem, normality of regression error should be assumed.

3. Since $E\left(\hat{\beta}_2\right) = \beta_2$ in the simple regression, the OLS estimator $\hat{\beta}_2$ is consistent when sample size gets large.

4. Under near multicollinearity, the classical assumptions are satisfied and hence an OLS estimator is the BLUE (best linear unbiased estimator) by Gauss-Markov theorem.

5. Heteroskedasticity occurs when the error term in a regression model is correlated with one of the explanatory variables.

6. In the simple regression model, suppose we multiply each X_i value by a constant, say 2. Then, it will not change the residuals ($\hat{e}_i$) and fitted values of Y ($\hat{Y}_i$).

7. The more observations you have, you will get more accurate estimates. Therefore, R^2 will always be higher if you have more observations.

8. The higher is the value of σ^2 (variance of regression error e_i), the larger is the variance of OLS estimator $\hat{\beta}_2$ in the simple regression.

Part II

Short Questions

1. **(10 points)** Consider a two-variable simple regression, *i.e.* $Y_i = \beta_1 + \beta_2 X_i + e_i$. *Student A* thinks that a slope parameter β_2 is more important and an intercept is not necessary, so she estimates $Y_i = \beta_2^A X_i + e_i$ without an intercept though the true model is $Y_i = \beta_1 + \beta_2 X_i + e_i$. In case, her OLS estimator $\hat{\beta}_2^A$ is an unbiased estimator of the true parameter β_2? (Yes or No without any explanation will be 0 points.)

2. **(10 points)** *Student B* is trying to analyze the presidential election of *Sejong University Student Association*. Assume that Sejong University has 4 schools (Economics School, Engineering School, Humanities School, and Arts School). She thinks that perhaps concern with tuition was relevant, as well as school differences. She decides to estimate a regression model where the dependent variable is the percentage of votes cast for *Candidate A* across departments (50 departments). The number of observation is 50. She has the following explanatory variables: tuition (per semester) by department, school dummy variables for whether a department is included in the Economics School, Engineering School, or Humanities School. (i) Write down a model including all of the variables (Be sure to define what dummy variables are, that is, when they take on what values). (ii) How would *Student B* test the hypothesis that "there is no school difference"? Write down the null hypothesis and (iii) clarify which test she would use (If an F-test, write down the restricted model).

3. **(10 points)** *Student C* obtained the following results from 10 observations :

$$\overline{X} = 5, \quad \overline{Y} = 6$$

$$\sum \left(X_i - \overline{X}\right)^2 = 100 \qquad \sum X_i Y_i = 100$$

Suppose that *Student C* wants to obtain the OLS estimators for a simple regression model, $Y_i = \beta_1 + \beta_2 X_i + e_i$. Using the above results, calculate $\hat{\beta}_1$ and $\hat{\beta}_2$.

4. **(10 points)** Consider the following Cobb-Douglas production function as

$$Y_i = \alpha L_i^{\beta} K_i^{\gamma} e^{e_i}$$
$$\Longrightarrow \quad \ln Y_i = \ln \alpha + \beta \ln L_i + \gamma \ln K_i + e_i$$

and now *Student D* wants to test the CRS (Constant Returns to Scale):

$$H_0: \ \beta + \gamma = 1 \ \Rightarrow \text{Constant Returns to Scale}$$

(i) Set up the restricted regression model, and (ii) test the null hypothesis of the CRS based on the following results ($n = 203$) ***(Note: please refer to the probability table on the last page.)***

Restricted Model: $R^2 = 0.94$

Unrestricted Model: $TSS = 100, \ \ RSS = 5$

Part III

Long Questions

1. **(Total 20 points)** Consider the simple regression model

$$Y_i = \beta_1 + \beta_2 X_i + e_i$$

Assume we have heteroskedasticity problem and the variance of regression errors are known as

$$Var\left(e_i\right) = \sigma^2 X_i$$

and X_i is deterministic, $E\left(e_i\right) = 0$ for all i, $Cov\left(e_i, e_j\right) = 0$ for $i \neq j$.

(a) **(5 points)** Is OLS estimator for β_2 unbiased and linear?

(b) **(5 points)** Transform the regression model for GLS estimation. What is the variance of transformed error e_i^*? Is e_i^* homoskedastic?

(c) **(5 points)** Is the GLS estimator for β_2 BLUE (best linear unbiased estimator)? Explain.

(d) **(5 points)** Explain FGLS (feasible GLS) estimation in the above case *(Tip: We have information about $Var\left(e_i\right)$, then the GLS estimation is infeasible?)* .

2. **(Total 20 points)** Consider the following regression models:

Model 1: $Y_i = \beta_1 + \beta_2 X_i + e_i$

Model 2: $Y_i^* = \beta_1^* + \beta_2^* X_i^* + e_i^*$

Model 3: $\ln Y_i = \alpha_1 + \alpha_2 \ln X_i + u_i$

where $Y_i^* = 2Y_i$ and $X_i^* = \frac{1}{2}X_i$. The **Model 3** is a log-linear model by taking log on both dependent variable and regressor in **Model 1**. Assume that the classical assumptions are satisfied .

(a) **(4 points)** Find the OLS estimators of β_1 and β_2 in **Model 1**.

(b) **(4 points)** Find the OLS estimators of β_1^* and β_2^* in **Model 2**. Are they identical to $\hat{\beta}_1$ and $\hat{\beta}_2$ you obtained in (a)?

(c) **(4 points)** Is the R^2 different between **Model 1** and **Model 2**?

(d) **(4 points)** Discuss the difference between the OLS estimator of β_2 in **Model 1** and that of α_2 in **Model 3**, in terms of interpretation.

(e) **(4 points)** Is the R^2 different between **Model 1** and **Model 3**?

3. **(Total 40 points)** Suppose you have 254 observations on house sales in Seoul. Consider the linear regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 D_i X_{2i} + e_i, \quad \text{for } i = 1, \cdots, 254$$

where

$$\begin{aligned} Y_i &: \text{ Sales Price (in million won)} \\ X_{2i} &: \text{ Size (in } m^2\text{)} \\ X_{3i} &: \text{ Number of rooms in a house} \\ D_i &: \text{ Location dummy} = \begin{cases} 1 & \text{if house is in the South} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Assume that the errors are homoskedastic and uncorrelated.

(a) **(5 points)** Compare the expected value of house sales price, $E(Y_i)$, in the South with one in elsewhere. Briefly interpret the difference.

(b) **(5 points)** You want to calculate the change of sales price of **a house in the South** when **a 10m^2-room** is added to a house. How much will you expect to get more on average (in terms of expected value), assuming everything else being held constant?

(c) **(5 points)** You estimate the regression equation as

$$\widehat{Y}_i = \underset{(0.10)}{12} + \underset{(1.5)}{20}\, X_{2i} + \underset{(3.5)}{10}\, X_{3i} + \underset{(0.5)}{5}\, D_i X_{2i}$$

$$n = 254 \quad RSS = 5,000 \quad ESS = 15,000$$

(standard errors in parentheses)

Now obtain the estimate of change the sales price of **a house in the South** when **a 10m^2-room** is added to a house.

(d) **(5 points)** Based on the estimates, test the significance of $\hat{\beta}_4$ with 5% significance level. ***(Note: please refer to the probability table on the last page.)***

(e) **(5 points)** Report R^2 of this regression.

(f) **(5 points)** Report the adjusted R^2 $\left(= \overline{R}^2\right)$ of this regression. Is it greater than R^2?

(g) **(5 points)** Test the overall significance of the model based on the estimates. You should clarify what is a test statistic and its distribution. ***(Note: please refer to the probability table on the last page.)***

(h) **(5 points)** Calculate the OLS estimator of σ^2 using the above regression output.

4. **(Total 10 points)** Suppose you have a cross-sectional data set of workers from census data. You want to run a regression of their salaries (Y_i) on the years of education (X_i) and the years of working experience (Z_i). Now you want to compare the wage level of workers of the same age, so **you have chosen 50 workers who were born in 1986** from your census data. However, years of working experience (Z_i) is not available in the

data set, so you decided to use

$$\text{years of working experience}\,(Z_i) = \text{age} - \text{years of education}\,(X_i) - \text{pre-education years}$$

Since all the 50 persons in your observations are 36 years old (and we can assume that pre-education years is 5), this becomes

$$Z_i = 36 - X_i - 5 = 31 - X_i$$

You have run the GRETL program based on the following regression model

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + e_i$$

and you encountered the following result and it showed an error message "Omitted due to (?): Z_i". There were only the values of $\hat{\beta}_1$ and $\hat{\beta}_2$.

Model 1: OLS, using observations 1–50
Dependent variable: salaries
Omitted due to (**?**): Z_i

	Coefficient	Std. Error	t-ratio	p-value
const	−6.71033	1.91416	−3.506	0.0005
X_i	1.98029	0.136117	14.55	0.0000

(a) **(5 points)** Fill in the blank (?).

(b) **(5 points)** Explain why the program omitted the variable of years of working experience (Z_i). If the model maintains Z_i in the regression model, what happens?

- **Probability table for Final Exam**:
  - $P\left(\chi^2(1) \geq 3.841\right) = 0.05$, $P\left(\chi^2(1) \geq 5.023\right) = 0.025$, $P\left(\chi^2(1) \geq 7.879\right) = 0.01$ for χ^2 distribution with degrees of freedom 1.
  - $P\left(\chi^2(2) \geq 5.991\right) = 0.05$, $P\left(\chi^2(2) \geq 7.377\right) = 0.025$, $P\left(\chi^2(2) \geq 9.210\right) = 0.01$ for χ^2 distribution with degrees of freedom 2.
  - $P\left(\chi^2(3) \geq 7.814\right) = 0.05$, $P\left(\chi^2(3) \geq 9.348\right) = 0.025$, $P\left(\chi^2(3) \geq 11.34\right) = 0.01$ for χ^2 distribution with degrees of freedom 3.
  - $P(F(1, +200) \geq 3.84) = 0.05$, $P(F(1, +200) \geq 6.63) = 0.01$ for F distribution with degrees of freedom $(1, +200)$. (+200 indicates for more than 200 degrees of freedom.)
  - $P(F(2, +200) \geq 3.00) = 0.05$, $P(F(2, +200) \geq 4.61) = 0.01$ for F distribution with degrees of freedom $(2, +200)$. (+200 indicates for more than 200 degrees of freedom.)
  - $P(F(3, +200) \geq 2.60) = 0.05$, $P(F(3, +200) \geq 3.78) = 0.01$ for F distribution with degrees of freedom $(3, +200)$. (+200 indicates for more than 200 degrees of freedom.)
  - $P(Z \geq 1.28) = 0.1$, $P(Z \geq 1.645) = 0.05$, $P(Z \geq 1.96) = 0.025$, $P(Z \geq 2.58) = 0.005$ where Z is a standard normal random variable.