Final Exam

(June 15, 2022, **2 hours**)

Econometrics (Spring 2022)

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 the case of 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. If you add one more additional independent variable to your linear regression model, the R^2 can increase.

2. One can introduce as many dummy variables as the number of categories of a qualitative variable if she/he omits the intercept in the regression model.

3. Under near multicollinearity, OLS estimator cannot be BLUE even under the classical assumptions, and therefore one should use GLS methodology to obtain the most efficient estimator.

4. In the multiple regression model, the variance of OLS estimator gets larger as the correlation between the regressors is higher.

5. At the same significance level, the probability of type I error tends to be smaller when doing a two-sided test than when doing a one-sided test.

6. When the regressors are not deterministic but random, the OLS estimator does not converge to the true parameter any more, even in a very large sample.

7. In the simple regression model, suppose we multiply each Y_i value by a constant, say 2. Then, it will not change the residuals and fitted values of Y.

8. Even if heteroskedasticity is present, we can still use OLS estimator because the OLS estimator is still consistent estimator.

Part II

Short Questions

1. **(10 points)** Suppose the true relationship of Y and X, Z is given by

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

under classical assumptions. Now, *Student A* estimates β_2 by regressing Y on X only (that is, omitting a relevant variable Z) as:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{e}_i$$

Then, calculate $Bias\left(\hat{\beta}_2\right)$.

2. **(10 points)** In the simple linear regression $Y_i = \beta_1 + \beta_2 X_i + e_i$, we can obtain the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. *Student B* considers the newly-defined dependent variable and independent variable by adding same constant to Y_i and X_i. Specifically, she defines $Y_i^* = Y_i + c$ and $X_i^* = X_i + c$ (where c is non-zero constant). As a result, she runs the following regression:

$$Y_i^* = \alpha_1 + \alpha_2 X_i^* + u_i$$

Express the OLS estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$ in terms of $\hat{\beta}_1$, $\hat{\beta}_2$ and c.

3. **(10 points)** In the following estimated model, Y_i is the expenditure on beers, X_i is the expenditures on fried chickens, and M_i is a dummy variable indicating 1 for males, 0 for females.

$$\hat{Y}_i = -0.6 + 1.5M_i + 0.5X_i + 0.2M_i X_i$$

Here, *Student C* uses a dummy variable F_i instead of M_i for regression, where F_i has the value 1 for females and 0 for males. That is, he estimates the following equation:

$$Y_i = \beta_1 + \beta_2 F_i + \beta_3 X_i + \beta_4 F_i X_i + e_i$$

Obtain the estimated values of $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$.

4. **(10 points)** From a sample of 100 observations, *Student D* obtained the following results:

$$\sum X_i = 400, \quad \sum Y_i = 700, \quad \sum X_i Y_i = 4,000, \quad \sum X_i^2 = 6,000$$

Suppose that *Student D* wants to obtain the OLS estimate for a two-variable regression through the origin, $Y_i = \beta_2 X_i + e_i$. Using the above results, calculate $\hat{\beta}_2$.

Part III

Long Questions

1. **(Total 20 points)** Suppose you have the following estimated models of relationships between consumption (C), income (INC), and time (t):

$$\textbf{Model A}: \widehat{C} = \underset{(2.73)}{0.2} + \underset{(3.44)}{0.5}\, INC \qquad R^2 = 0.5$$

$$\textbf{Model B}: \widehat{\ln C} = \underset{(2.66)}{0.01} + \underset{(4.50)}{0.3}\, \ln INC \qquad R^2 = 0.6$$

$$\textbf{Model C}: \widehat{\ln C} = \underset{(3.02)}{0.01} + \underset{(1.58)}{0.1}\, t \qquad R^2 = 0.6$$

Note that the values in parentheses are t-ratios.

(a) **(5 points)** Is the elasticity of consumption with respect to income in **Model A** constant? If yes, what is it?

(b) **(5 points)** Is the elasticity of consumption with respect to income in **Model B** constant? If yes, what is it?

(c) **(5 points)** How would you interpret 0.1 (coefficient of time variable) in **Model C**?

(d) **(5 points)** Is the slope coefficient in **Model A** significant at 1% significance level? ***(Note: If you need, please refer to the probability table on the last page.)***

2. **(Total 35 points)** A *Student E* has data for the year 2020 from US census data on the following characteristics of the respondents: weekly earnings, $EARNINGS$, measured in thousand dollars; years of schooling, SCH; years of work experience, EXP; sex (males and females); and ethnicity (whites and non-whites). *Student E* defines dummy variables $MALE$ and NWH. $MALE$ is defined to be 1 for males and 0 for females. NWH is defined to be 1 for non-whites and 0 for whites. She also defines interactive term, $MNW = MALE * NWH$. She fits the following ordinary least squares regressions, each with $EARNINGS$ as the dependent variable:

Variable	**Model (1)**	**Model (2)**	**Model (3)**
Constant	0.390	0.459	0.411
	(0.075)	(0.076)	(0.082)
SCH	0.124	0.121	0.122
	(0.004)	(0.004)	(0.004)
EXP	0.033	0.032	0.033
	(0.002)	(0.002)	(0.003)
$MALE$	0.278	0.277	0.306
	(0.0003)	(0.0003)	(0.0003)
NWH		-0.144	0.205
		(0.022)	(0.099)
MNW			-0.280
			(0.065)
R^2	0.300	?	0.376

The numbers in parenthesis are standard deviations of the least squares estimates. Note that the total sum of squares (TSS) is 350 and the number of observations is 214.

(a) **(5 points)** What is RSS (residual sum of squares) in **Model (1)**?

(b) **(5 points)** Test the overall significance of **Model (1)** at 1% significance level. ***(Note: If you need, please refer to the probability table on the last page.)***

(c) **(5 points)** Give an interpretation of the coefficient of NWH in **Model (2)**.

(d) **(5 points)** Specify the interval for possible values of R^2 in **Model (2)**. In other words, find the lower bound and the upper bound of R^2 in **Model (2)**, using the above results.

(e) **(5 points)** Give an interpretation of the coefficient of MNW in **Model (3)**.

(f) **(5 points)** Perform the test of the joint significance of the coefficients of NWH and MNW in **Model (3)** at 5% significance level. ***(Note: If you need, please refer to the probability table on the last page.)***

(g) **(5 points)** Explain whether a simple t-test on the coefficient of NWH in **Model (2)** is sufficient to show that the wage equations are different for whites and non-whites.

3. **(Total 25 points)** There exists an exact linear relationship between two variables X and Y as:

$$Y_i = \beta_1 + \beta_{2i} X_i \tag{1}$$

Assume that X_i is deterministic. However, in fact, β_{2i} is not a fixed constant, but a random variable. To be specific, suppose that β_{2i} is expressed as:

$$\beta_{2i} = \beta_2 + \nu_i \tag{2}$$

where β_2 is a parameter (*i.e.* unknown but fixed constant), and ν_i is an error term with mean 0 and constant variance σ_ν^2. Also, ν_i is independent for different i.

(a) **(5 points)** Using the equation (2), express the equation (1) as the standard model:

$$Y_i = \beta_1 + \beta_2 X_i + e_i \tag{3}$$

What is e_i in the equation (3)?

(b) **(5 points)** What is $E(e_i)$ and $Var(e_i)$ in the equation (3)?

(c) **(5 points)** Is OLS estimator for β_2 in the equation (3) unbiased and linear? Why or why not?

(d) **(5 points)** Is OLS estimator for β_2 in the equation (3) BLUE? Why or why not?

(e) **(5 points)** If you answered "NO" in the above question, transform the equation (3) appropriately to obtain BLUE. If you answered "YES" in the above question, write down "Not Applicable".

- **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.

(End of Exam, Total 5 Pages)