ECON*3740
Fall 2018
Final Exam

Time Limit: 120 Minutes

Instructor: **Chaoyi Chen**

Name: ____________________

ID: ____________________

- There are THREE separate parts in this exam. You must answer ALL questions.
- There are a total of 100 points.
- Do not hesitate to raise your hand if you have a question.

Good Luck!

Grade Table (for marker use only)

Question	Points	Score
1	20	
2	30	
3	25	
4	25	
Total:	100	

Multiple Choice Questions (20 marks)

1. (20 points) All following questions only have one correct answer. Please record your response by filling the circle.

   (a) (2 points) Which of the following is true of the White test?

   - ◯ The White test is used to detect the presence of multicollinearity in a linear regression model.
   - ◯ The White test cannot detect forms of heteroskedasticity that invalidate the usual Ordinary Least Squares standard errors.
   - ◯ The White test can detect the presence of heteroskedasticty in a linear regression model even if the functional form is misspecified.
   - ◯ The White test assumes that the square of the error term in a regression model is uncorrelated with all the independent variables, their squares and cross products.

   (b) (2 points) Which of following models is **not** a linear regression models?

   - ◯ $Y_i = \beta_0 + \beta_1(\frac{1}{X_i}) + \mu_i$
   - ◯ $Y_i = \beta_0 + ln(\beta_1)X_i + \mu_i$
   - ◯ $ln(Y_i) = \beta_0 + \beta_1 ln(X_i) + \mu_i$
   - ◯ $ln(Y_i)^2 = \beta_0 + \beta_1 X_i^{1/2} + \mu_i$

   (c) (2 points) Which of the following is true of dummy variables?

   - ◯ A dummy variable always takes a value less than 1.
   - ◯ A dummy variable always takes a value higher than 1.
   - ◯ A dummy variable takes a value of 0 or 1.
   - ◯ A dummy variable takes a value of 1 or 10.

   (d) (2 points) In the following equation, gdp refers to gross domestic product, and FDI refers to foreign direct investment.

   $$log(gdp) = 2.65 + 0.527log(bank\ credit) + 0.222FDI$$

   Which of the following statements is then true?

   - ◯ If gdp increases by 1%, bank credit increases by 0.527%, the level of FDI remaining constant.
   - ◯ If bank credit increases by 1%, gdp increases by 0.527%, the level of FDI remaining constant.

- ◯ If gdp increases by 1%, bank credit increases by log(0.527)%, the level of FDI remaining constant.
- ◯ If bank credit increases by 1%, gdp increases by log(0.527)%, the level of FDI remaining constant.

(e) (2 points) Refer to the model above. Which of the following statements is then true?

- ◯ If FDI increases by 1 unit, gdp increases by approximately 22.2%, the amount of bank credit remaining constant.
- ◯ If FDI increases by 1%, gdp increases by approximately 22.2%, the amount of bank credit remaining constant.
- ◯ If FDI increases by 1 unit, gdp increases by approximately 24.8%, the amount of bank credit remaining constant.
- ◯ If FDI increases by 1%, gdp increases by approximately 24.8%, the amount of bank credit remaining constant.

(f) (2 points) Which of the following correctly identifies an advantage of using adjusted R^2 over R^2?

- ◯ Adjusted R^2 corrects the bias in R^2.
- ◯ Adjusted R^2 is easier to calculate than R^2.
- ◯ The penalty of adding new independent variables is better understood through adjusted R^2 than R^2.
- ◯ The adjusted R^2 can be calculated for models having logarithmic functions while R^2 cannot be calculated for such models.

(g) (2 points) Suppose the variable x_2 has been omitted from the following regression equation, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$. $\tilde{\beta}_1$ is the estimator obtained when x_2 is omitted from the equation. The bias in $\tilde{\beta}_1$ is **positive** if

- ◯ $\beta_2 > 0$ and x_1 and x_2 are positively correlated
- ◯ $\beta_2 < 0$ and x_1 and x_2 are positively correlated
- ◯ $\beta_2 > 0$ and x_1 and x_2 are negatively correlated
- ◯ $\beta_2 = 0$ and x_1 and x_2 are negatively correlated

(h) (2 points) Exclusion of a relevant variable from a multiple linear regression model leads to the problem of

- ◯ misspecification of the model

- ◯ multicollinearity
- ◯ perfect collinearity
- ◯ homoskedasticity

(i) (2 points) A data set that has repeat observations for the same agents over a given period is called:

- ◯ cross-sectional data set
- ◯ time series data set
- ◯ pooled cross-sectional data set
- ◯ panel data set

(j) (2 points) Consider a log-level model

$$log(y) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \mu$$

Let $\widehat{log(y)}$ denote the fitted value of above regression model. Hence, $\tilde{y} = e^{\widehat{log(y)}}$ will

- ◯ always under estimate $E[y|\mathbf{x}]$
- ◯ always over estimate $E[y|\mathbf{x}]$
- ◯ always equally estimate $E[y|\mathbf{x}]$
- ◯ sometimes under, sometimes over, and sometimes equally estimate $E[y|\mathbf{x}]$

Long Answer Questions: Theory Part (30 marks)

2. (30 points) Consider the regression model

$$Y_i = \beta_1 X_{1i} + \mu_i$$

for $i = 1, ..., n$. Notice that there is **no** constant term in the regression.

(a) (5 points) Write down assumptions of the Gauss-Markov theorem.

(b) (5 points) Assuming all assumptions in (a) hold, specify the least squares function that is minimized by OLS and solve for $\widehat{\beta}_1$.

(c) (4 points) Use the result in (b), show what will happen to $\widehat{\beta}_1$ if we **minus** the **independent** variable by a constant, say 5.

(d) (4 points) Now, suppose all conditions in (a) hold **except** that $Var(\mu_i) = \sigma^2 X_{1i}$. Show that the OLS estimate $\widehat{\beta}_1$ is still unbiased.

(e) (4 points) Continue to (d), describe how you can transform the model and obtain the best linear unbiased estimator (BLUE).

Now, suppose we introduce one more variable (X_2) into the model,

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \mu_i$$

(f) (4 points) If X_2 is an **irrelevant** variable, show that the OLS estimators are still unbiased ($E(\widehat{\beta}_1) = \beta_1$, $E(\widehat{\beta}_2) = 0$).

(g) (4 points) If $X_2 = 2X_1$, show that we cannot obtain OLS estimates for β_1 and β_2 and explain why. (Hint: $Corr(X_1, X_2) = 1$)

Long Answer Questions: Applied Part (50 marks)

3. (25 points) Consider a regression model that relates the proportion of a household's budget spent on alcohol $WALC$ to total expenditure $TOTEXP$, age of the household head AGE, and the number of children in the household NK. Followings are the regression output after the estimation,

$$\begin{array}{lcccc} \widehat{WALC} = & 0.0091 & +0.0276 ln(TOTEXP) & +\mathbf{a} \times AGE & -0.0133NK \\ & (0.019) & (\mathbf{b}) & (0.0002) & (0.0033) \\ & \{\mathbf{c}\} & \{6.6086\} & \{-6.9624\} & \{-4.0750\} \end{array}$$

where $n = 2000$, $SSR = 5.7529$, (.) are the standard errors, and $\{.\}$ are the corresponding t-statistics. $WALC$ is measured by %, $TOTEXP$ is measured by CAD\$, AGE is measured by year. Note that only households with one or two children are being considered. Thus, NK takes only the values **one** or **two**.

(a) (5 points) Find the values of a, b, and c.

(b) (5 points) Interpret each of the coefficient estimates.

(c) (5 points) Now, suppose we know that the standard deviation of the dependent variable = 0.0633, find the R^2 and interpret its meaning. Specify the hypothesis to test the overall significance of the regression model. Use the value of R^2 to test this hypothesis.

(d) (5 points) Compute a 95% interval coefficient estimate of AGE . What does this interval tell you?

(e) (5 points) Suppose now, we consider a new fomulation of the regression model,

$$WALC = \beta_0 + \beta_1 ln(TOTEXP) + \beta_2 AGE + \beta_3 (NK - 1) + \mu$$

If we estimate the model using the same dataset, find values of $\widehat{\beta}_1$, $\widehat{\beta}_2$, and $\widehat{\beta}_3$. Then, interpret the meaning of $\widehat{\beta}_3$.

4. (25 points) Consider the following regression model,

$$WAGE_i = \beta_0 + \beta_1 EDU + \beta_2 EXPER + \beta_3 EXPER^2 + \mu_i$$

where $WAGE$ is the monthly wage measured by thousand dollars, $EDUC$ is the education measured by years, and $EXPER$ is the experience measured by years.

(a) (5 points) What is the marginal effect of experience on wages?

(b) (5 points) Assuming β_3 is negative, explain what signs do you expect for the coefficients β_1, β_2. After how many years of experience do wages start to decline? (Express your answer in terms of β's.)

(c) (5 points) One student believes the turning point should be 20 years working experience. Describe (i) how you can form a hypothesis test of above argument (ii) how you can transform the regression model so that you can use a t-statistic to test (i).

(d) (5 points) After estimation, you have the following fitted regression model,

$$\widehat{WAGE}_i = \underset{(2.02)}{-13.43} \quad \underset{(0.14)}{+2.28EDUC} \quad \underset{(0.1)}{+0.68EXPER} \quad \underset{(0.002)}{-0.01EXPER^2}$$

Test the null hypothesis that $WAGE$ is linearly related to $EXPER$ against the alternative that the relationship is quadratic.

(e) (5 points) Given the following Covariance Matrix for Least Squares Estimates

Covariance Matrix for Least Squares Estimates

	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$
$\widehat{\beta}_0$	4.11			
$\widehat{\beta}_1$	-0.22	0.02		
$\widehat{\beta}_2$	-0.12	-0.002	0.01	
$\widehat{\beta}_3$	0.002	0.002	-0.002	0.003

Find 95% interval estimates for (i) The marginal effect of education on wages (ii) The marginal effect of experience on wages when $EXPER = 4$

END OF EXAM

End of examination
Total marks: 100
Total pages: 14