**ECON*3740**
**Fall 2018**
**Midterm**
**10/17/2018**
**Time Limit: 80 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 | 4 | |
| 2 | 4 | |
| 3 | 4 | |
| 4 | 4 | |
| 5 | 4 | |
| 6 | 40 | |
| 7 | 40 | |
| Total: | 100 | |

# 1 Multiple Choice Questions (20 marks)

1. (4 points) Let $Y = \beta_0 + \beta_1 X + \varepsilon$ where the distribution of $\varepsilon$ conditional on $X$ is normal with mean 0 and variance $\sigma^2$. Suppose that $\beta_0 = 1$, $\beta_1 = 2$, and $\sigma^2 = 4$. Further, suppose that $X$ is normally distributed with population mean 1 and population variance 9. Then $Var(Y|X)$ equal to:
   - ◯ 0
   - ◯ 22
   - ◯ 39
   - ◯ 4

2. (4 points) Which of following models is **not** a linear regression models?
   - ◯ $Y_i = \beta_0 + \beta_1(\frac{1}{X_i}) + \mu_i$
   - ◯ $Y_i = \beta_0 + \beta_1 ln(X_i) + \mu_i$
   - ◯ $ln(Y_i) = \beta_0 + \beta_1 ln(X_i) + \mu_i$
   - ◯ $ln(Y_i) = ln(\beta_0) + \beta_1 X_i + \mu_i$

3. (4 points) In the regression of y on x, the error term exhibits **Homoskedasticity** if
   - ◯ it has a constant variance
   - ◯ $Var(y|x)$ is a function of x
   - ◯ x is a function of y
   - ◯ y is a function of x

4. (4 points) A data set that consists of repeat observations on a specific agent over time, is called:
   - ◯ cross-sectional data set
   - ◯ longitudinal data set
   - ◯ time series data set
   - ◯ experimental data set

5. (4 points) If the variability in the $x_i$ increases, the variance of slope estimator will
   - ◯ decrease
   - ◯ increase
   - ◯ keep the same
   - ◯ it's hard to determine

## 2 Long Answer Questions: Theory Part (40 marks)

6. (40 points) Consider a simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

(a) (5 points) Show that $E(\mu_i|X_i) = 0$ implies $E(Y_i|X_i) = \beta_0 + \beta_1 X_i$

(b) (10 points) Write down five necessary OLS assumptions. Show that which assumptions needed to make sure OLS estimators $\widehat{\beta_0}$ and $\widehat{\beta_1}$ are unbiased estimators and which assumptions allow us to simplify the variance of OLS estimator.

(c) (10 points) Use the fact that $\widehat{\beta}_1$ is unbiased, show that $\widehat{\beta}_0$ is an unbiased estimator of $\beta_0$ (Hint: use $\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X}$, $E(\widehat{\beta}_1|X) = \beta_1$, and $E(\mu_i|X) = 0$)

(d) (10 points) Use the formulea for estimating $\beta_1$ and $\beta_0$ under OLS, show what will happen to these estimated coefficients if we **multiply** the **independent** variable by a constant, say 5.

(e) (5 points) Now, suppose you know that $\beta_0 = 0$. **Derive** a formula for the least squares estimator of $\beta_1$ (Hint: Write down the minimization problem first, then derive the FOC.)

# 3 Long Answer Questions: Applied Part (40 marks)

7. (40 points) One student would like to investigate the effect of the age on average weekly earning. Therefore, she considers a regression model

$$AWE_i = \beta_0 + \beta_1 Age_i + \mu_i,$$

where $AWE$ is average weekly earnings measured in dollars and $Age$ is measured in years

(a) (5 points) Explain what the term $\mu_i$ may represent. Why will different works have different value of $\mu_i$.

(b) (5 points) Do you agree that $E(\mu_i|X_i) = 0$ for this regression model? Why and Why not? (Hint: This is an open question. Just justify the economic intuition behind.)

Using a random sample of college-educated full-time workers aged 25–65, she estimates the model and yields the following:

$$\widehat{AWE} = 696.7 + 9.6 \times Age$$
$$R^2 = 0.023$$

(c) (5 points) Explain the economic meaning of the coefficients values 696.7 and 9.6.

(d) (5 points) What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?

(e) (10 points) The regression $R^2$ is 0.023. please interpret its meaning. Does the high $R^2$ value imply that higher *age* cause higher $AWE$? Explain. What are the units of measurement for the $R^2$? (Dollars? Years? Or is $R^2$ unit-free?)

(f) (5 points) The **average** *Age* in this sample is 41.6 years. What is the **average** value of $AWE$ in the sample?

Suppose now, one of her group member suggested that since the age is monotonically increasing, it might make more sense if they could use the deviations of age from the sample mean as the independent variable. That is, they consider a new formulation of regression model:

$$AWE_i = \alpha_0 + \alpha_1(Age_i - \bar{Age}) + \mu_i$$

,

(g) (5 points) Do you agree that $\widehat{\alpha}_0 = \widehat{\beta}_0$ and $\widehat{\alpha}_1 = \widehat{\beta}_1$, where $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are the OLS estimates of the original regression model? Use words, math to support your argument.

**END OF EXAM**

**End of examination**
**Total marks: 100**
**Total pages: 9**