Exam

Name: Quasi Final

ESSAY. Write your answer in the space provided or on a separate sheet of paper.

1) (Continuation from Chapter 4) Sir Francis Galton, a cousin of James Darwin, examined the relationship between the height of children and their parents towards the end of the 19th century. It is from this study that the name “regression” originated. You decide to update his findings by collecting data from 110 college students, and estimate the following relationship:

$$\widehat{Studenth} = \underset{(7.2)}{19.6} + \underset{(0.10)}{0.73} \times Midparh, \quad R^2 = 0.45, \; SER = 2.0$$

where *Studenth* is the height of students in inches, and *Midparh* is the average of the parental heights. Values in parentheses are heteroskedasticity robust standard errors. (Following Galton’s methodology, both variables were adjusted so that the average female height was equal to the average male height.)
(a) Test for the statistical significance of the slope coefficient.
(b) If children, on average, were expected to be of the same height as their parents, then this would imply two hypotheses, one for the slope and one for the intercept.
(i) What should the null hypothesis be for the intercept? Calculate the relevant t-statistic and carry out the hypothesis test at the 1% level.
(ii) What should the null hypothesis be for the slope? Calculate the relevant t-statistic and carry out the hypothesis test at the 5% level.
(c) Can you reject the null hypothesis that the regression R^2 is zero?
(d) Construct a 95% confidence interval for a one inch increase in the average of parental height.

2) Females, on average, are shorter and weigh less than males. One of your friends, who is a pre-med student, tells you that in addition, females will weigh less for a given height. To test this hypothesis, you collect height and weight of 29 female and 81 male students at your university. A regression of the weight on a constant, height, and a binary variable, which takes a value of one for females and is zero otherwise, yields the following result:

$$\widehat{Studentw} = -229.21 - 6.36 \times Female + 5.58 \times Height, \; R^2 = 0.50, \; SER = 20.99$$

where *Studentw* is weight measured in pounds and *Height* is measured in inches.

(a) Interpret the results. Does it make sense to have a negative intercept?
(b) You decide that in order to give an interpretation to the intercept you should rescale the height variable. One possibility is to subtract 5 ft. or 60 inches from your *Height*, because the minimum height in your data set is 62 inches. The resulting new intercept is now 105.58. Can you interpret this number now? Do you thing that the regression R^2 has changed? What about the standard error of the regression?
(c) You have learned that correlation does not imply causation. Although this is true mathematically, does this always apply?

3) The F-statistic with q = 2 restrictions when testing for the restrictions $\beta_1 = 0$ and $\beta_2 = 0$ is given by the following formula:

$$F = \frac{1}{2}\left(\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1,t_2} t_1 t}{1 - \hat{\rho}^2_{t_1,t_2}}\right)$$

Discuss how this formula can be understood intuitively.

4) Females, it is said, make 70 cents to the dollar in the United States. To investigate this phenomenon, you collect data on weekly earnings from 1,744 individuals, 850 females and 894 males. Next, you calculate their average weekly earnings and find that the females in your sample earned \$346.98, while the males made \$517.70.

(a) Calculate the female earnings in percent of the male earnings. How would you test whether or not this difference is statistically significant? Give two approaches.

(b) A peer suggests that this is consistent with the idea that there is discrimination against females in the labor market. What is your response?

(c) You recall from your textbook that additional years of experience are supposed to result in higher earnings. You reason that this is because experience is related to “on the job training.” One frequently used measure for (potential) experience is “Age-Education-6.” Explain the underlying rationale. Assuming, heroically, that education is constant across the 1,744 individuals, you consider regressing earnings on age and a binary variable for gender. You estimate two specifications initially:

$$\widehat{Earn} = \underset{(21.18)}{323.70} + \underset{(0.55)}{5.15} \times Age - \underset{(13.06)}{169.78} \times Female,\ R^2=0.13,\ SER=274.75$$

$$\widehat{Ln(Earn)} = \underset{(0.08)}{5.44} + \underset{(0.002)}{0.015} \times Age - \underset{(0.036)}{0.421} \times Female,\ R^2=0.17,\ SER=0.75$$

where *Earn* are weekly earnings in dollars, *Age* is measured in years, and *Female* is a binary variable, which takes on the value of one if the individual is a female and is zero otherwise. Interpret each regression carefully. For a given age, how much less do females earn on average? Should you choose the second specification on grounds of the higher regression R^2?

(d) Your peer points out to you that age-earning profiles typically take on an inverted U-shape. To test this idea, you add the square of age to your log-linear regression.

$$\widehat{Ln(Earn)} = \underset{(0.18)}{3.04} + \underset{(0.009)}{0.147} \times Age - \underset{(0.033)}{0.421} \times Female - \underset{(0.0001)}{0.0016}\, Age^2,$$

R^2 =0.28, SER=0.68

Interpret the results again. Are there strong reasons to assume that this specification is superior to the previous one? Why is the increase of the *Age* coefficient so large relative to its value in (c)?

(e) What other factors may play a role in earnings determination?