Department of Economics
University of Maryland

John C. Chao
Spring 2012

Economics 422

Final Examination

This exam contains 5 questions. The total time allowed for the exam is 120 minutes. The total number of points awarded is 100. You should provide explanations for your answers. Right answers with no explanation will receive no credit. On the other hand, wrong answers with thoughtful explanations may receive some credit. Partial credit will be given. **GOOD LUCK!**

QUESTION 1 (20 POINTS)

You have obtained a sub-sample of 1744 individuals from the Current Population Survey (CPS) and are interested in the relationship between weekly earnings and age. The regression, using heteroskedasticity-robust standard errors, yielded the following result:

$$\widehat{Earn}_i = \underset{(20.24)}{239.16} + \underset{(0.57)}{5.20} \times Age_i, \quad R^2 = 0.05, \ SER = 287.21;$$

where $Earn_i$ and Age_i are measured in dollars and years, respectively, and where the numbers in parentheses are standard errors. For this problem, you should consider the case where an assumption that the errors of the regression are normally distributed might not be appropriate.

(a) Interpret the slope coefficient estimate, and discuss how well the model fits the data. Indicate whether or not the estimated coefficients are significantly different from zero.

(b) The average age in this sample is 37.5 years. What is the average annual earnings in the sample (assuming that there are 52 weeks in a year)?

QUESTION 2 (10 POINTS)

Suppose that you have quarterly data on aggregate consumption C_t and disposable income X_t, and you specify the following regression model

$$C_t = \beta_0 + \beta_1 X_t + \beta_2 Q_{1t} + \beta_3 Q_{2t} + \beta_4 Q_{3t} + \beta_5 Q_{4t} + u_t,$$

where

$$Q_{1t} = \begin{cases} 1 & \text{if } t^{th} \text{ observation belongs to the } 1^{st} \text{ quarter (Jan-March)} \\ 0 & \text{otherwise} \end{cases},$$

$$Q_{2t} = \begin{cases} 1 & \text{if } t^{th} \text{ observation belongs to the } 2^{nd} \text{ quarter (April-June)} \\ 0 & \text{otherwise} \end{cases},$$

$$Q_{3t} = \begin{cases} 1 & \text{if } t^{th} \text{ observation belongs to the } 3^{rd} \text{ quarter (July-Sept)} \\ 0 & \text{otherwise} \end{cases},$$

$$Q_{4t} = \begin{cases} 1 & \text{if } t^{th} \text{ observation belongs to the } 4^{th} \text{ quarter (Oct-Dec)} \\ 0 & \text{otherwise} \end{cases}.$$

Discuss if there is any problem with this specification and whether you will be able to compute the ordinary least squares (OLS) estimator.

QUESTION 3 (30 POINTS)

Suppose you are interested in the effects of computer ownership on college grade point average. Suppose further that based on data you have, you obtained the following regression result

$$\widehat{GPA}_i = \underset{(.35)}{1.30} + \underset{(.06)}{0.16} \times PC_i, \;\; n = 150, \;\; R^2 = 0.2; \tag{1}$$

where PC_i is a binary regressor which equals one if student i owns a personal computer and zero otherwise and where the number given inside a parenthesis denotes a standard error. For this problem, you should again consider the case where an assumption that the errors of the regression are normally distributed might not be appropriate. You should also assume that the regression above is properly specified, so that it satisfies all the assumptions needed to obtain unbiased and consistent estimates of the coefficients, as discussed in lectures.

(a) Test, at the 5% significance level, the null hypothesis

$$H_0 : E\left[GPA_i | PC_i = 0\right] = 1$$

versus the alternative hypothesis

$$H_1 : E\left[GPA_i | PC_i = 0\right] > 1.$$

Show your work.

(b) Test, at the 5% significance level, the null hypothesis

$$H_0 : E\left[GPA_i | PC_i = 1\right] - E\left[GPA_i | PC_i = 0\right] = 0$$

versus the alternative hypothesis

$$H_1 : E\left[GPA_i | PC_i = 1\right] - E\left[GPA_i | PC_i = 0\right] \neq 0.$$

Show your work.

(c) Consider testing the null hypothesis

$$H_0 : E\left[GPA_i | PC_i = 1\right] = 2$$

versus the alternative hypothesis

$$H_1 : E\left[GPA_i|PC_i = 1\right] < 2$$

Show how you can transform the regression model given by (1) above so that you can test this null hypothesis as a simple zero restriction using the t-statistic. Do not attempt to implement the test in this case. Show your work.

QUESTION 4 (16 POINTS)

Consider the linear regression

$$Y_i = \beta_0 + \sum_{j=1}^{5} \beta_j X_{j,i} + u_i. \tag{2}$$

Suppose that you are able to collect a sample of 30 observations on each of the variables (i.e., $Y_i, X_{1,i}, X_{2,i}, X_{3,i}, X_{4,i}$, and $X_{5,i}$) in equation (2) above. Now, your objective is to test the null hypothesis

$$H_0 : \beta_1 = 0,\ \beta_3 = 0, \text{ and } \beta_5 = 0 \tag{3}$$

versus the alternative hypothesis

$$H_1 : \text{ null hypothesis is false.}$$

Using your data, you estimate the regression given by (2) and obtain

$$RSS_{unrestricted} = 274.$$

(Note that RSS stands for Residual Sum of Squares.) In addition, you have also estimated the restricted model implied by the null hypothesis (3), i.e.,

$$Y_i = \beta_0 + \beta_2 X_{2,i} + \beta_4 X_{4,i} + u_i$$

and obtain

$$RSS_{restricted} = 351.$$

Consider the case where it is reasonable to assume homoskedasticity and normally-distributed errors, and test at the 5% significance level the null hypothesis given by (3) using an exact finite sample test (i.e., do not use large sample approximation). Show your work.

QUESTION 5 (24 POINTS)

Consider the model

$$Y_i = \beta X_i + u_i, \tag{4}$$

where Y_i denotes average test score on the Standard 9 achievement test for school district i and X_i denotes the average student-teacher ratio for school district i and where the data for this model are for K-8 California school districts for the year 1998 and, hence, the index i ranges over these districts. Because of various possible omitted variable problems, we suspect that equation (4) cannot be estimated consistently by ordinary least squares (OLS). Suppose further that there are data on the occurrence of earthquakes in these districts in 1998, so that we can define the binary regressor

$$Z_i = \begin{cases} 1 & \text{if district } i \text{ was hit by an earthquake in 1998} \\ 0 & \text{otherwise} \end{cases}$$

(a) Discuss why Z_i might plausibly be considered a valid instrument for the purpose of estimating β in equation (4) using the method of instrumental variable.

(b) Now, assume that Z_i is a valid instrument and that we can define the first-stage regression

$$X_i = \pi Z_i + v_i,$$

where $\pi \neq 0$. Show that the two-stage least squares (TSLS) estimator in this case is given by the formula

$$\widehat{\beta}^{TSLS} = \frac{\overline{Y}_Q}{\overline{X}_Q},$$

where $\overline{Y}_Q$ denotes the sample average of the Y_i's across only those districts which were hit by an earthquake in 1998 whereas $\overline{X}_Q$ is the sample average of the X_i's again across only those districts which were hit by an earthquake in 1998. Justify your answer carefully.