Department of Economics
University of Maryland

John C. Chao
Spring 2012

Economics 422

Final Examination Solution Sheet

QUESTION 1

(a) The slope coefficient estimate suggests that an increase in age of one year results in an increase in weekly earnings by \$5.20. The model does not fit the data very well, as $R^2 = 0.05$, so it only explains 5% of the variation in earnings. The t-statistics for testing the significance of the intercept and slope coefficients are as follows

$$
\begin{aligned}
t_{\widehat{\beta}_0} &= \frac{239.16}{20.24} = 11.82 > 1.96, \\
t_{\widehat{\beta}_1} &= \frac{5.20}{0.57} = 9.12 > 1.96.
\end{aligned}
$$

It follows that both estimated coefficients are significant at the 5% level.

(b) Let $\overline{E}_w$ denote average weekly earnings in the sample, and let $\overline{A}$ denote average age in the sample; then, the least squares estimate of the intercept $\widehat{\beta}_0$ is related to $\overline{E}$ and $\overline{A}$ by the well-known formula

$$\widehat{\beta}_0 = \overline{E}_w - \widehat{\beta}_1 \overline{A},$$

so that

$$\overline{E}_w = \widehat{\beta}_0 + \widehat{\beta}_1 \overline{A} = 239.16 + (5.2)(37.5) = 434.16$$

It follows that average annual earnings $\overline{E}_a$ is given by

$$\overline{E}_a = 52 \times \overline{E}_w = (52)(434.16) = 22576.32.$$

QUESTION 2

The problem is that the regression here contains an intercept and a dummy variables specification which exhausts all categories. Hence,

$$\sum_{j=1}^{4} Q_{jt} = 1 \text{ for each } t,$$

and we have perfect multicollinearity, and the OLS estimator cannot be computed.

QUESTION 3

(a) Note that in this case

$$E\left[GPA_i|PC_i = 0\right] = \beta_0.$$

It follows that we can reformulate the test of hypotheses in this case as

$$H_0 : \beta_0 = 1 \text{ versus } \ H_1 : \beta_0 > 1.$$

We can test this null hypothesis using the t-statistic, but because we do not assume that the errors are normally distributed, we must rely on large sample approximation. It follows that for testing against an one-side alternative at the 5% significance level, the critical value is 1.645. Since

$$t_{\widehat{\beta}_0} = \frac{\widehat{\beta}_0 - 1}{SE\left(\widehat{\beta}_0\right)} = \frac{1.30 - 1}{0.35} = \frac{0.3}{0.35} \approx 0.86 < 1.645,$$

we fail to reject H_0.

(b) Note that

$$E\left[GPA_i|PC_i = 1\right] - E\left[GPA_i|PC_i = 0\right] = \beta_0 + \beta_1 - \beta_0 = \beta_1.$$

It follows that we can reformulate the test of hypotheses in this case as

$$H_0 : \beta_1 = 0 \text{ versus } \ H_1 : \beta_0 \neq 0.$$

Again, because we do not assume that the errors are normally distributed, we must rely on large sample approximation. For testing against an one-side alternative at the 5% significance level, the critical values are ± 1.96. Since

$$t_{\widehat{\beta}_1} = \frac{\widehat{\beta}_1}{SE\left(\widehat{\beta}_1\right)} = \frac{0.16}{0.06} \approx 2.67 > 1.96,$$

we reject H_0.

(c) To see how we can transform the regression so that the null hypothesis of interest can be tested as a zero restriction, we start with the original regression

$$GPA_i = \beta_0 + \beta_1 PC_i + u_i. \tag{1}$$

Note first that

$$E\left[GPA_i|PC_i = 1\right] = \beta_0 + \beta_1,$$

so that we can reformulate the null hypothesis of interest as

$$H_0 : \beta_0 + \beta_1 = 2 \text{ or } H_0 : \beta_0 + \beta_1 - 2 = 0.$$

Now, adding and subtracting the term $(\beta_0 - 2)\, PC_i$ from the regression (1) and grouping, we obtain

$$GPA_i = \beta_0\left(1 - PC_i\right) + \left(\beta_0 + \beta_1 - 2\right) PC_i + 2PC_i + u_i.$$

Next, subtract $2PC_i$ from both sides of the regression above, we get

$$GPA_i - 2PC_i = \beta_0 (1 - PC_i) + (\beta_0 + \beta_1 - 2) PC_i + u_i,$$

or

$$Y_i = \beta_0 (1 - PC_i) + \gamma PC_i + u_i, \tag{2}$$

where $Y_i = GPA_i - 2PC_i$ and $\gamma = \beta_0 + \beta_1 - 2$. It is obvious that testing the null hypothesis $H_0 : \beta_0 + \beta_1 - 2 = 0$ is the same as testing the simple zero restriction

$$H_0 : \gamma = 0$$

in the transformed regression (2), which we can do using the t-statistic.

QUESTION 4

Note that for this problem we are testing three restrictions, so that $q = 3$. Moreover, there are five regressors in the unrestricted model (not counting the "regressor" associated with the intercept). Hence, $k = 5$. Given that we assume the errors to be homoskedastic and normally distributed, the F-statistic will have an F-distribution with 3 and $n - k - 1 = 30 - 5 - 1 = 24$ degrees of freedom under H_0 in the present case, so that, at 5% level of significance, the critical value will be 3.01.

Since

$$\begin{aligned} F &= \frac{(RSS_{\text{restricted}} - RSS_{\text{unrestricted}})/q}{RSS_{\text{unrestricted}}/(n-k-1)} \\ &= \frac{(351 - 274)/3}{274/(30-5-1)} \\ &= \frac{1848}{822} \\ &\approx 2.25 < 3.01, \end{aligned}$$

we fail to reject H_0.

QUESTION 5

(a) To be a valid instrumental variable, Z_i must be correlated with X_i, which is the average student-teacher ratio in district i. Since being hit by an earthquake may result in damage to school facilities, leading to fewer classrooms and larger classes, it is plausible that Z_i will be correlated with X_i, as required. Moreover, a valid instrument Z_i must also be uncorrelated with u_i, the error term which includes all other factors besides X_i which may affect test scores. Since the occurrence of an earthquake should be a totally exogenous event, it is also reasonable to believe that it will be uncorrelated with u_i. Since Z_i, as defined, seems to satisfy both conditions for a valid instrument.

(b) The two-stage least squares estimator is given by the formula

$$\begin{aligned}\widehat{\beta}^{TSLS} &= \frac{\widehat{Cov\left(Z_i, Y_i\right)}}{\widehat{Cov\left(Z_i, X_i\right)}} \\ &= \frac{\frac{1}{n}\sum_{i=1}^{n} Z_i Y_i}{\frac{1}{n}\sum_{i=1}^{n} Z_i X_i} \\ &= \frac{\frac{1}{n}\sum_{i=1}^{n} Z_i Y_i}{\frac{1}{n}\sum_{i=1}^{n} Z_i X_i} \qquad (3)\end{aligned}$$

Now, since $Z_i = 0$ for any district not hit by an earthquake in the sample, it follows that the sum in both the numerator and the denominator of (3) will only involve those districts which have been hit by an earthquake. More formally, let

$$S = \{i : \text{district } i \text{ was hit by earthquake}\}$$

i.e., the set of indices of those districts which were hit by an earthquake in 1998; then,

$$\frac{1}{n}\sum_{i=1}^{n} Z_i Y_i = \frac{1}{n}\sum_{i\in S} Y_i = \left(\frac{n_S}{n}\right)\frac{1}{n_S}\sum_{i\in S} Y_i = \left(\frac{n_S}{n}\right)\overline{Y}_Q$$

and

$$\frac{1}{n}\sum_{i=1}^{n} Z_i Y_i = \left(\frac{n_S}{n}\right)\frac{1}{n_S}\sum_{i\in S} X_i = \left(\frac{n_S}{n}\right)\overline{X}_Q.$$

where n_S denotes the number of elements in S. It follows immediately that

$$\widehat{\beta}^{TSLS} = \frac{\frac{1}{n}\sum_{i=1}^{n} Z_i Y_i}{\frac{1}{n}\sum_{i=1}^{n} Z_i X_i} = \frac{\left(\frac{n_S}{n}\right)\overline{Y}_Q}{\left(\frac{n_S}{n}\right)\overline{X}_Q} = \frac{\overline{Y}_Q}{\overline{X}_Q}.$$