Econ 120B **Ramu Ramanathan**
Fall 2003 **Final Exam Answers**

The campus grading regulations prohibit regrading of final exams. Therefore, read all parts of questions carefully and answer fully. Maximum score is 50.

I. A labor economist wished to examine the effects of schooling and experience on earnings. Using cross-section data, she obtained the following relationships:

$$\widehat{\ln E} = \underset{(0.113)}{7.71} + \underset{(0.005)}{0.094}S + \underset{(0.009)}{0.023}N - \underset{(0.000187)}{0.000325}N^2$$

$$R^2 = 0.337 \qquad n = 60$$

where ln *E* is the natural logarithm of earnings, *S* is the number of years of schooling, and *N* is the number of years of experience. R^2 is unadjusted and the <u>values in parentheses are standard errors</u>. *In all the tests below, use the 5% level of significance.*

Ia (5 points) Test the hypothesis (state the null and alternate hypotheses in terms of b s) "Schooling has no effect on earnings." Show all your work. Do you agree with the statement? Why or why not? (use a two-sided test.)

If schooling has no effect on earnings, the coefficient for S would be zero. H_0: $b_2=0$ and H_1: $b_2 \neq 0$. The test is a two-tailed t-test with 56 d.f., and t_c= 0.094/0.005 = 18.8. We would expect such a high t-statistic to be very significant. At the 5 percent level, t is between 2.0 and 2.021. Since $t_c > t^*$, we conclude that schooling is significant at the 5 percent level. Statement is therefore wrong.*

Ib (5 points) Test the single <u>joint hypothesis</u> (state the null and alternate hypotheses in terms of b s) "Neither schooling nor experience has any effect on earnings." (You have all the information needed to perform this test.) Show all your work. Do you agree with the statement? Why or why not?

If neither schooling nor experience have any effect on earnings, the coefficients for all the explanatory variables (except the constant) will be zero. Thus, $b_2 = b_3 = b_4 = 0$ is the null hypothesis. The alternate is that at least one of them is nonzero. The test is an F-test on the overall significance of the model. The test statistic is given by equation (4.4) as

F_c = {0.337/3} ¸ {0.663/56} = 9.5

Under the null hypothesis, this has the F distribution with 3 d.f. for the numerator and 56 d.f. for the denominator. $F^*_{3,56}(0.05)$ is in (2.76, 2.84) which *is well below F_c. Therefore, we reject the null hypothesis and conclude that the model is significant overall. The statement is therefore false.*

Ic (10 points) Describe how you would use the Wald test (not LM test) to test the hypothesis "Experience has no effect on earnings." More specifically, state the null and alternative hypotheses; describe what additional regression(s), if any, you would run; write an expression for the test statistic; state its distribution, degrees of freedom, and the acceptance/rejection criterion. Where available you should provide actual numbers, not symbols.

The null hypothesis is that the coefficients for N and N^2 are both zero, that is, $b_3 = b_4 = 0$. The alternative is that at least one of the coefficients is nonzero. Estimate the restricted model by regressing lnE against a constant and S. Then compute the F-statistic [as in equation (4.3)].

$$F_c = \frac{(ESS_R - ESS_U)/2}{ESS_U/56},$$ *ESS_R and ESS_U are error sum of squares for models R and U. F_c has the F-distribution with 2 and 56 d.f. Reject H_0 if $F_c >$ $F^*_{2,56}(0.05)$ which is in the range (3.15, 3.23).*

Id (5 points) Derive the expression for the elasticity of earnings with respect to schooling. Compute the numerical value of this elasticity when S = 8. (You don't need the value of N for this.)

Elasticity of E with respect to S is (DE/DS)(S/E) = 0.094S. The elasticity is not constant but varies with S. For S = 8, the value is 0.752.

Ie (5 points) In the next page, derive the expression for the elasticity of earnings with respect to experience. Compute the numerical value of this elasticity when N=5. (You don't need the value of S for this.)

Elasticity with respect to experience is (DE/DN)(N/E) = N(0.023 - 0.00065N). This varies with N. At N=5, we have, 5(0.023 – 5 ´ 0.00065) = 0.09875.

II.

You have data on the sale price (PRICE) in thousands of dollars, square feet of living area (SQFT), and square feet of the yard size (YARD) for a sample of 59 single-family homes sold recently. The basic model first estimated was

$$\text{PRICE} = b_1 + b_2\,\text{SQFT} + b_3\,\text{YARD} + u$$

I suspect that the terms ln(SQFT) and ln(YARD) should be added to the model.

IIa (3 points) Write down a general model that adds these variables.

$$\text{PRICE} = b_1 + b_2\,\text{SQFT} + b_3\,\text{YARD} + b_4 \ln(\text{SQFT}) + b_5 \ln(\text{YARD}) + \text{u}$$

To perform an LM test for the addition of these variables, I obtained the following auxiliary regression:

$$\hat{u} = \underset{(0.01)}{3265} + \underset{(0.012)}{0.255}\,\text{SQFT} - \underset{(0.917)}{0.000485}\,\text{YARD} - \underset{(0.011)}{507}\ln(\text{SQFT}) + \underset{(0.886)}{6.765}\ln(\text{YARD})$$

IIb (5 points) Carefully describe how I must have obtained $\hat{u}_t$.

First regress PRICE against a constant, SQFT, and YARD and obtain $\hat{b}_1$, $\hat{b}_2$, and $\hat{b}_3$. Then compute $\hat{u} = PRICE - \hat{b}_1 - \hat{b}_2 SQFT - \hat{b}_3 YARD$.

IIc (3 points) For this regression, unadjusted R^2 = 0.115, n = 59. Compute the test statistic and state its distribution and d.f. under the null.

The test statistic is LM = nR^2 = 59 ´ 0.115 = 6.785. It is distributed as chi-square with 2 d.f.

IId (3 points) Use a 5 percent level of significance and actually carry out the test. What do you conclude?

From the chi-square table we have LM = 5.99146. Since LM > LM*, we reject the null hypothesis and conclude that either ln(SQFT), or ln(YARD), or both belong in the model.*

IIe (6 points) In the above regression, the <u>values in parentheses are *p*-values</u> for a two-tailed test. From the information given, write down a model you should estimate. Carefully justify your choice. [Note: This should not be the "kitchen sink" model from IIa.]

The rule of thumb for inclusion is any new variable with p-value less than 0.50 (or less than 0.25). By this rule, only ln(SQFT) should be added to the original model because its coefficient has a p-value < 0.05.

$$\text{PRICE} = b_1 + b_2\,\text{SQFT} + b_3\,\text{YARD} + b_4\ln(\text{SQFT}) + v$$