Assignment 3: Multiple and Nonlinear Regression Models

ECO 321

DUE: Thursday, April 12 in class

Instructions: You need to write a Stata do-file to answer these questions. Your do-file must produce a log file that contains your Stata output. You must submit your log file along with your answers to the assignment. Assignments that do not contain log files will be marked down.

1. In this question, we are interested in the determinants of years of education. Use the data set HTV.dta for this question. The data set includes information on education, ability, parents' education, and several other variables for men in 1991.

a) Load the data set in Stata (via the `use` command). First, use the `sum, detail` command to show the range of values that the education variable takes in the sample. Second, use the `gen` command to create a dummy variable that equals 1 if men completed exactly 12 years of education and equals 0 otherwise. Third, use the `tab` command to show what percentage of men completed exactly 12 years of education. Fourth, use the `sum` command to determine whether the men or their parents have higher levels of education.

b) Use the `reg` command to estimate the regression model:

$$educ_i = \beta_0 + \beta_1 motheduc_i + \beta_2 fatheduc_i + u_i$$

Make sure to use heteroskedasticity-robust standard errors. How much of the variation in *educ* is explained by parents' education? Interpret the coefficient on *motheduc*.

c) Add the variable *abil* (a measure of cognitive ability) to the model in part (b). Does ability help to explain variation in education, even after controlling for parents' education? Use at least one test statistic to justify your answer.

d) Compare the coefficient estimates on *motheduc* and $fatheduc$ from the model in part (b) to the model in part (c). What do the differences tell us about omitted variable bias in the model in part (b)? Be specific.

e) Use the results from part (c) to test the hypothesis that the effect of mother's education on men's education equals two times the effect of father's education on men's education. First, write down the null hypothesis and the alternative hypothesis. Second, evaluate the hypothesis using the F-statistic and the `test` command. Third, rearrange the regression model and evaluate the hypothesis using a t-statistic. Do you arrive at the same conclusion using the F- and t-statistics?

2. Next we are interested in the determinants of wages. Use the data set wage1.dta to answer this question.

a) Use OLS to estimate the equation:

$$log(wage_i) = \beta_0 + \beta_1 edu_i + \beta_2 exper_i + \beta_3 exper_i^2 + u_i$$

Write the regression results in standard form (i.e., plug in the correct values for the parameters estimates and their standard errors):

$$\hat{log(wage_i)} = \underset{(SE(\hat{\beta}_0))}{\hat{\beta}_0} + \underset{(SE(\hat{\beta}_1))}{\hat{\beta}_1} educ_i + \underset{(SE(\hat{\beta}_2))}{\hat{\beta}_2} exper_i + \underset{(SE(\hat{\beta}_3))}{\hat{\beta}_3} exper_i^2$$

b) Interpret β_1 and test its statistical significance at the 1% level. What can we say about the relationship between education and $log(wage)$?

c) Test whether experience has a statistically significant effect on $log(wage)$.

d) Is $exper^2$ statistically significant at the 5% level? What does this hypothesis test say about the relationship between $log(wage)$ and experience?

e) Use the approximation of the marginal effect of experience:

$$\%\Delta\hat{wage_i} = \frac{dlog(wage_i)}{dexper_i} = (\hat{\beta}_2 + 2\hat{\beta}_3 exper_i) \times 100\%$$

to find the approximate return to going from 5 to 6 years of experience. What is the approximate return to going from 20 to 21 years of experience? Do you think experience exhibits diminishing marginal returns?

f) Find the predicted difference in $log(wage)$ between someone with 10 years of experience and someone with 20 years of experience. Is the difference statistically significant? (Note: Do not use the approximation from part (e). We only use the approximation in part (e) when $\Delta X = 1$. In this case, $\Delta X = 10$.)

g) At what value of *exper* does an additional year of experience reduce predicted $log(wage)$? How many people have experience beyond that turning point in the sample? Do you think the turning point is a problem or do you think the results make sense?

h) What if you think that there should not be a turning point for experience, but you still think there are diminishing marginal returns to experience. How might you change the model? (Hint: You need to change the functional form of experience.)