

Assignment 6

The assignment is due on Thursday June 23 in Learn DropBox.

Question 1: The data set 401ksubs.Rdata contains information on net financial wealth (*netffa*), age of the survey respondent (*age*), annual family income (*inc*), family size (*fsize*), and participation in certain pension plans for people in the United States. The wealth and income variables are both recorded in thousands of dollars. For this question, use only the data for single females and households of size greater than 1 (so *fsize* > 1, *marr* = 0, and *male* = 0).

- (i) Use OLS to estimate the following model:

$$netffa = \beta_0 + \beta_1 inc + \beta_2 age + u$$

and report the results using the usual format. Be sure to use only the single-person households in the sample. Interpret the slope coefficients. Are there any surprises in the slope estimates?

- (ii) Does the intercept from the regression in part (i) have an interesting meaning? Explain.
- (iii) Find the p-value for the test $H_0 : \beta_2 = 0.8$ against $H_1 : \beta_2 < 0.8$. Do you reject H_0 at the 1% significance level?
- (iv) If you do a simple regression of *netffa* on *inc*, is the estimated coefficient on *inc* much different from the estimate in part (ii)? Why or why not?
- (v) Perform an F-test for the significance of the whole model estimated in part (i) at 5%.

Question 2: Use the data in htv.RData to answer this question and restrict your sample to workers who live in western area (*west*=1).

- (i) Estimate the regression model

$$educ = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + \beta_3 abil + \beta_4 I(abil^2) + u$$

by OLS and report the results in the usual form. Test the null hypothesis that *educ* is linearly related to *abil* against the alternative that the relationship is quadratic.

- (ii) Using the equation in part (i), test $H_0 : \beta_1 = \beta_2$ against a two-sided alternative. What is the p-value of the test?
- (iii) Answer question (ii) using the indirect-t approach.
- (iv) Answer question (ii) using the F-test approach.
- (v) Add the two college tuition variables (*tuit17* and *tuit18*) to the regression from part (i) and determine whether they are jointly statistically significant.
- (vi) What is the correlation between *tuit17* and *tuit18*? Explain why using the average of the tuition over the two years might be preferred to adding each separately. What happens when you do use the average?
- (v) Do the findings for the average tuition variable in part (vi) make sense when interpreted causally? What might be going on?