

Prof. C. M. Dalton
ECN 209A
Spring 2015
Homework 3

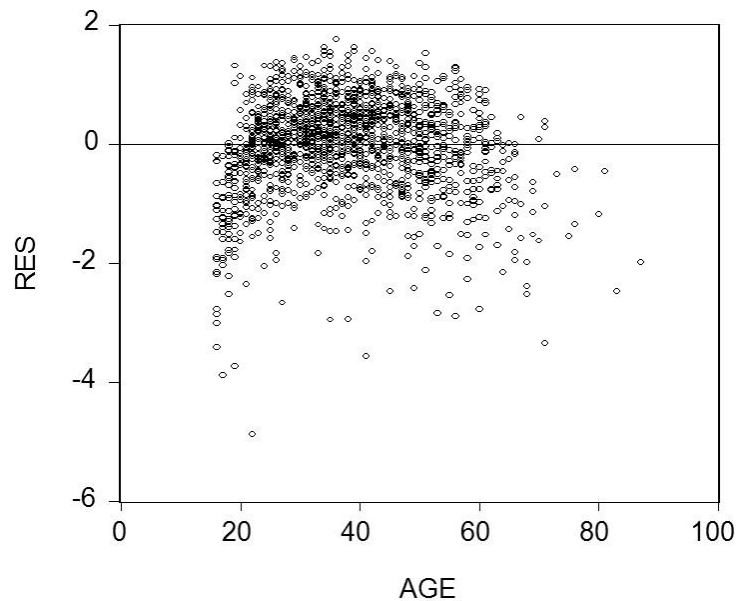
Due Thursday, April 23rd

For full credit: Questions must be solved completely on separate, clean papers and stapled. You may work in groups, but make sure to list collaborators.

Question 1 Earnings functions are one of the most investigated relationships in labor economics. These typically relate the logarithm of earnings to a series of explanatory variables such as education, work experience, gender, race, etc.

a) Why do you think that researchers have preferred a log-linear specification over a linear specification? Think about the dependent and independent variables and resulting coefficient interpretation.

b) To establish age-earnings profiles, you regress $\ln(Earn)$ on Age , where $Earn$ is weekly earnings in dollars, and Age is in years. Plotting the resulting unexplained residuals (u_i) of the regression against age for 1,744 individuals looks as shown in the figure:



Do you sense a problem? Why do you think that?

c) You decide, given your knowledge of age-earning profiles, to allow the regression line to differ for the below and above 40 years age category. Accordingly you create a binary variable, D_{young} , that takes the value one for age 39 and below, and is zero otherwise. Estimating the earnings equation results in the following output (using heteroskedasticity-robust standard errors):

$$\widehat{\ln Eam} = 6.92 - 3.13 \times D_{young} - 0.019 \times Age + 0.085 \times (D_{young} \times Age)$$

(38.33) (0.22) (0.004) (0.005)

Where $R^2=0.20$, $SER=0.721$

d) Interpret the coefficients. How well does the regression explain the data?

e) Sketch a diagram with both regression lines: one for the age category 39 years and under, and one for 40 and above. Does it make sense to have a negative sign on the Age coefficient? Predict the $\ln(\text{earnings})$ for a 30 year old and a 50 year old. What is the percentage difference between these two?

Question 2 External Validity

- a. What does it mean for an analysis to be “externally valid?”
- b. Discuss the external validity (or invalidity) of a study on Wake Forest students’ candy bar preferences.

Question 3: For each of the following, discuss whether or not multiple OLS regression analysis would be invalid, and if so how, in the presence of:

- a. Omitted variables bias
- b. Measurement error in variables
- c. Heteroskedasticity of the error term
- d. Externally invalid data