Tutorial Exercises

Question 1 (Textbook 6.3):
The following model allows the return to education to depend upon the total amount of both parents’ education, called pareduc:

\[ \log (wage) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{educ} \cdot \text{pareduc} + \beta_3 \text{exper} + \beta_4 \text{tenure} + u \]

(i) Show that, in decimal form, the return to another year of education in this model is

\[ \frac{\Delta \log (wage)}{\Delta \text{educ}} = \beta_1 + \beta_2 \text{pareduc} \]

What sign do you expect for \( \beta_2 \)? Why?

(ii) Using the data in WAGE2.RAW, the estimated equation is

\[ \Delta \log (\text{wage}) = 5.65 + 0.047 \text{educ} + 0.0078 \text{educ} \cdot \text{pareduc} + 0.019 \text{exper} + 0.010 \text{tenure} \]

(Only 722 observations contain full information on parents’ education.) Interpret the coefficient on the interaction term. It might help to choose two specific values for pareduc—for example, pareduc=32 if both parents have a college education, or pareduc=24 if both parents have a high school education—and to compare the estimated return to educ.

(iii) When pareduc is added as a separate variable to the equation, we get:

\[ \Delta \log (\text{wage}) = 4.94 + 0.097 \text{educ} + 0.033 \text{pareduc} - 0.0016 \text{educ} \cdot \text{pareduc} + 0.020 \text{exper} + 0.010 \text{tenure} \]

Does the estimated return to education now depend positively on parent education? Test the null hypothesis that the return to education does not depend on parent education.

Question 2 (Textbook 6.6):
The following three equations were estimated using the 1,534 observations in 401K.RAW:

\[ \text{prate} = 80.29 + 5.44 \text{mrate} + 0.269 \text{age} - 0.00013 \text{ttemp} \]

(0.78) (0.52) (0.045) (0.00004)

\[ R^2 = .100, \quad R^2 = .098 \]

\[ \text{prate} = 97.32 + 5.02 \text{mrate} + 0.314 \text{age} - 2.66 \log(\text{ttemp}) \]

(1.95) (0.51) (0.044) (0.28)

\[ R^2 = .144, \quad R^2 = .142 \]
The equation estimated using the data in CEOSAL1.RAW:

\[ \text{prâte} = 80.62 + 5.34 \text{ mrate} + .290 \text{ age} - .00043 \text{ totemp} + .0000000039 \text{ totemp}^2 \]

\[ (0.78) \quad (0.52) \quad (.45) \quad (.00009) \]

\[ + .00000000010 \text{ totemp}^2 \]

\[ R^2 = .108, \; R^2 = .106 \]

Which of these three models do you prefer? Why?

**Question 3 (textbook 6.9)**

The following equation was estimated using the data in CEOSAL1.RAW:

\[ \text{log(salary)} = 4.322 + .276 \text{ log(sales)} + .0215 \text{ roe} - .00008 \text{ roe}^2 \]

\[ (.324) \quad (.033) \quad (.0219) \quad (.00026) \]

\[ n= 209, \; R^2 = .282 \]

This equation allows roe to have a diminishing effect on log(salary). Is this generality necessary? Explain why or why not.

**Question 4 (textbook 7.3)**

Using the data in GPA2.RAW, the following equation was estimated:

\[ \text{sat} = 1028.10 + 19.30 \text{ hsize} - 2.19 \text{ hsize}^2 - 45.09 \text{ female} \]

\[ (6.29) \quad (3.83) \quad (0.53) \quad (4.29) \]

\[ - 169.81 \text{ black} + 62.31 \text{ female-black} \]

\[ (12.71) \quad (18.15) \]

\[ n= 4137, \; R^2 = .0858 \]

The variable sat is the combined SAT score, hsize is size of the student’s high school graduating class, in hundreds, female is a gender dummy variable, and black is a race dummy variable equal to one for blacks, and zero otherwise.

(i) Is there strong evidence that hsize2 should be included in the model?

From this equation, what is the optimal high school size?

(ii) Holding hsize fixed, what is the estimated difference in SAT score between nonblack females and nonblack males? How statistically significant is this estimated difference?

(iii) What is the estimated difference in SAT score between nonblack males and black males? Test the null hypothesis that there is no difference between their scores, against the alternative that there is a difference.

(iv) What is the estimated difference in SAT score between black females and nonblack females? What would you need to do to test whether the difference is statistically significant?