Econometrics Exercise session 5

Problem 1

Are rent rates influenced by the student population in a college town? Let *rent* be the average monthly rent paid on rental units in a college town in the United States. Let *pop* denote the total city population, *avginc* the average city income, and *pctstu* the student population as a percentage of the total population. One model to test for a relationship is

$$\log(rent) = \beta_0 + \beta_1 \log(pop) + \beta_2 \log(avginc) + \beta_3 pctstu + u$$

(i) State the null hypothesis that size of the student body relative to the population has no ceteris paribus effect on monthly rents. State the alternative that there is an effect.

(ii) What signs do you expect for β_1 and β_2?

(iii) The equation estimated using 1990 data from RENTAL.RAW for 64 college towns is

$$\widehat{\log(rent)} = \underset{(0.844)}{0.43} + \underset{(0\ .039)}{0.066} \log(pop) + \underset{(.081)}{0.507} \log(avginc) + \underset{(.0017)}{0.0056} pctstu + u$$

$n = 64, R^2 = .458$

What is wrong with the statement: "A 10% increase in population is associated with about a 6.6% increase in rent"? Interpret the coefficient on pctstu.

(iv) Test the hypothesis stated in part (i) at the 1% level.

Problem 2

Suppose you are interested in studying the tradeoff between time spent sleeping and working and to look at other factors affecting sleep. You specify the following model:

$$sleep = \beta_0 + \beta_1 * totwrk + \beta_2 * educ + \beta_3 * age + u$$

where $sleep$ and $totwrk$ (total work) are measured in minutes per week and $educ$ and age are measured in years.

Suppose we estimated the following regression:

$$\widehat{sleep} = \underset{(112.28)}{3638.25} + \underset{(.017)}{0.148} * totwrk - \underset{(5.88)}{11.13} * educ + \underset{(1.45)}{2.2} * age$$

$n = 706,\ R^2 = .113$

where we report standard errors along with the estimates.

(i) Is either *educ* or *age* individually significant at the 5% level against a two-sided alternative? Show your work.

(ii) Dropping *educ* and *age* from the equation gives

$$\widehat{sleep} = \underset{(38.91)}{3586.38} + \underset{(.017)}{0.151} * totwrk$$

$$n = 706,\ R^2 = .103$$

Are *educ* and *age* jointly significant in the original equation at the 5% level? Justify your answer.

(iii) Does including *educ* and *age* in the model greatly affect the estimated tradeoff between sleeping and working?

(iv) Suppose that the sleep equation contains heteroskedasticity. What does this mean about the tests computed in parts (i) and (ii)?

Problem 3

When estimating wage equations, we expect that young, inexperienced workers will have relatively low wages and that with additional experience their wages will rise, but then begin to decline after middle age, as the worker nears retirement. This lifecycle pattern of wages can be captured by introducing experience and experience squared to explain the level of wages. If we also include years of education, we have the equation:

$$Wage = \beta_0 + \beta_1 * Educ + \beta_2 * Exper + \beta_3 Exper^2 + u$$

a) What is the marginal effect of experience on wages?
b) What sign do you expect for each of the coefficients? Why?
c) Estimate the model using data *cps_small.gdt*. Do the estimated coefficients have expecting signs?
d) Test the hypothesis that education has no effect on wages. What do you conclude?
e) Test the hypothesis that education and experience have no effect on wages. What do you conclude?
f) Include the dummy variable *black* in the regression. Interpret the coefficient and comment on its significance.
g) Include the interaction term of *black* and *educ*. Interpret the coefficient and comment on its significance.
h) Transform dependent variable in logarithmic form and estimate the equation. Interpret the coefficients.

Problem 4

consider a simple model to compare the returns to education at junior colleges and four-year colleges; for simplicity, we refer to the latter as "universities." The population includes working people with a high school degree, and the model is:

$$\log(wage) = \alpha_0 + \alpha_1 jc + \alpha_2 univ + \alpha_3 exper + u \quad (1)$$

where

jc is number of years attending a two-year college, *univ* is number of years at a four-year college. *exper* is months in the workforce.

Note that any combination of junior college and four-year college is allowed, including *jc* =0 and *univ* = 0. Use the data ***twoyear.dta***

i) Test the hypothesis that $\alpha_1 = \alpha_2$. The hypothesis of interest is whether one year at a junior college is worth one year at a university.

(ii) The variable phsrank is the person's high school percentile. (A higher number is better. For example, 90 means you are ranked better than 90 percent of your graduating class.) Find the smallest, largest, and average phsrank in the sample.

(iii) Add phsrank to regression (2) and report the OLS estimates in the usual form. Is phsrank statistically significant? How much is 10 percentage points of high school rank worth in terms of wage?

(iv) Does adding phsrank to regression (2) substantively change the conclusions on the returns to two- and four-year colleges? Explain.