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.

$$\boldsymbol{H_0: \beta_3 = 0, H_1: \beta_3 \neq 0}$$

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

Other things equal, a larger population increases the demand for rental housing, which should increase rents. The demand for overall housing is higher when average income is higher, pushing up the cost of housing, including rental rates. Therefore, we expect positive signs.

(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.

The coefficient on *log(pop)* is an elasticity. A correct statement is that “a 10% increase in population increases rent by .066*10 = .66%.”. Increasing the proportion of student population by one unit increases the rental rates by 0.56%.

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

Test statistic $\boldsymbol{t = \frac{0.0056}{.0017} = 3.29}$

Critical value at 1% given the degree of freedom =64-4=60 and two-tailed student distribution will be 2.660, so we reject the null hypothesis that $\boldsymbol{\beta_3 = 0}$

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:

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

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

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.

$$\boldsymbol{t_{educ} = \frac{11.13}{5.88} = 1.89, \qquad t_{age} = \frac{2.2}{1.45} = 1.52}$$

Critical value at 5% with two tails and df=702 is $t_{cr} = 1.96$, therefore both age and educ are individually insignificant

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

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

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

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

We know that $F = \frac{(SSR_r - SSR_{ur})/q}{\frac{SSR_{ur}}{df}}$, shere q is the number of restrictions. We also know that

$$\boldsymbol{R^2 = \frac{ESS}{TSS} = \frac{TSS-SSR}{TSS} = 1 - \frac{SSR}{TSS} \Rightarrow SSR = (1 - R^2) * TSS}$$

TSS will be the same for both restricted and the unrestricted models, therefore it will cancel out. We will have:

$$\boldsymbol{F = \frac{R_{ur}^2 - R_r^2}{1 - R_{ur}^2} * \frac{df}{q} = \frac{0.113 - 0.103}{1 - .113} * \frac{702}{2} = 3.96}$$

The 5% critical value in the F table at $F_{2,702} = 3$, Therefore, we reject the hypothesis that age and education are jointly insignificant at the 5% level (3.96 > 3.00). In fact, the p-value is about .019, and so educ and age are jointly significant at the 2% level.

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

Not really. These variables are jointly significant, but including them only changes the coefficient on totwrk from –.151 to –.148.

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

The standard t and F statistics that we used assume homoskedasticity. If there is heteroskedasticity in the equation, the tests are no longer valid. In fact, standard errors without controlling heteroskedasticity are smaller than what it should be - increasing the significance of the estimated parameters, which is wrong.

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? $\boldsymbol{\beta_2 + 2 * \beta_3 * Exper}$
b) What sign do you expect for each of the coefficients? Why? $\boldsymbol{\beta_2}$ **positive** $\boldsymbol{\beta_3}$ **negative, because there should be diminishing marginal increase in the wages with experience**
c) Estimate the model using data *cps_small.gdt*. Do the estimated coefficients have expecting signs?

   genr exp2=exper^2

   ols wage const educ exper exp2

   Output:

```
Model 1: OLS, using observations 1-1000
Dependent variable: wage

             coefficient    std. error    t-ratio    p-value
  -----------------------------------------------------------
  const      -9.81770       1.05496       -9.306     8.19e-020 ***
  educ        1.21007       0.0702378     17.23      2.04e-058 ***
  exper       0.340949      0.0514314      6.629     5.52e-011 ***
  exp2       -0.00509306    0.00119794    -4.252     2.32e-05  ***

Mean dependent var    10.21302   S.D. dependent var    6.246641
Sum squared resid     28420.08   S.E. of regression    5.341743
R-squared             0.270934   Adjusted R-squared    0.268738
F(3, 996)             123.3772   P-value(F)            5.98e-68
Log-likelihood       -3092.487   Akaike criterion      6192.973
Schwarz criterion     6212.604   Hannan-Quinn          6200.434
```

Yes

d) Test the hypothesis that education has no effect on wages. What do you conclude?

Test statistic for educ is very large 17.23, therefore we reject such hypothesis even without looking at critical values 😉

e) Test the hypothesis that education and experience have no effect on wages. What do you conclude?

Here we are testing a joint hypothesis that β_1, β_2 *and* $\beta_3 = 0$, which we already have in GRETL output. See red circle in the GRETL output. The p-value is very small, therefore we reject H_0

f) Include the dummy variable *black* in the regression. Interpret the coefficient and comment on its significance.

ols wage const educ exper exp2 black

	coefficient	std. error	t-ratio	p-value	
const	-9.55171	1.05516	-9.052	7.21e-019	***
educ	1.19881	0.0700907	17.10	1.08e-057	***
exper	0.346425	0.0512790	6.756	2.42e-011	***
exp2	-0.00523499	0.00119459	-4.382	1.30e-05	***
black	-1.71571	0.595372	-2.882	0.0040	***

Mean dependent var	10.21302	S.D. dependent var	6.246641
Sum squared resid	28184.85	S.E. of regression	5.322263
R-squared	0.276969	Adjusted R-squared	0.274062
F(4, 995)	95.28762	P-value(F)	1.18e-68
Log-likelihood	-3088.331	Akaike criterion	6186.662
Schwarz criterion	6211.200	Hannan-Quinn	6195.988

The coefficient on black is -1.71, which means that being black rather than white reduces your wages by 1.71 dollars per hour. The coefficient on black is statistically significant at the 1% level since test statistic is -2.882 and the critical value in the student table is -2.57. Also P-Value=0.004<0.01, meaning statistically significant at 1% level. Three stars in the end of variables are also indicator of statistical significance at 1% level.

g) Include the interaction term of *black* and *educ*. Interpret the coefficient and comment on its significance.

genr bleduc=black*educ

```
Model 3: OLS, using observations 1-1000
Dependent variable: wage

              coefficient    std. error    t-ratio    p-value
  -----------------------------------------------------------
  const       -10.1179       1.08227       -9.349     5.68e-020 ***
  educ          1.23865      0.0721249     17.17      4.35e-058 ***
  exper         0.351995     0.0512321      6.871     1.13e-011 ***
  exp2         -0.00537840   0.00119380    -4.505     7.42e-06  ***
  black         6.30110      3.59031        1.755     0.0796    *
  bleduc       -0.620954     0.274259      -2.264     0.0238    **

Mean dependent var   10.21302   S.D. dependent var   6.246641
Sum squared resid    28040.24   S.E. of regression   5.311261
R-squared            0.280678   Adjusted R-squared   0.277060
F(5, 994)            77.57147   P-value(F)           9.75e-69
Log-likelihood      -3085.759   Akaike criterion     6183.518
Schwarz criterion    6212.964   Hannan-Quinn         6194.709

?
```

Coefficient on bleduc implies that for each extra year of education blacks receive less wages than whites by 0.62. It is statistically significant at the 5% level (2 stars). Including this term also reduces significance of the black variable alone and strangely, changes its sign to positive.

h) Transform dependent variable in logarithmic form and estimate the equation. Interpret the coefficients.

genr lwage=log(wage)

ols lwage const educ exper exp2 black bleduc

```
Model 4: OLS, using observations 1-1000
Dependent variable: lwage

              coefficient     std. error     t-ratio    p-value
  -------------------------------------------------------------
  const        0.298229       0.0938407       3.178     0.0015    ***
  educ         0.110994       0.00625375     17.75      1.96e-061 ***
  exper        0.0371932      0.00444220      8.373     1.90e-016 ***
  exp2        -0.000602239    0.000103511    -5.818     8.02e-09  ***
  black        0.289908       0.311306        0.9313    0.3519
  bleduc      -0.0356783      0.0237802      -1.500     0.1338

Mean dependent var   2.166837   S.D. dependent var   0.552806
Sum squared resid    210.8106   S.E. of regression   0.460525
R-squared            0.309472   Adjusted R-squared   0.305998
F(5, 994)            89.09560   P-value(F)           1.72e-77
Log-likelihood      -640.5409   Akaike criterion     1293.082
Schwarz criterion    1322.528   Hannan-Quinn         1304.274

Log-likelihood for wage = -2807.38
```

Increasing educ by one year increases the wage by 11%

Increasing exper by one year increases the wage by 100*(0.03-0.0006*exper) percent

Black and bleduc do not have significant impact on logarithmic wages

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:

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

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.
To test this hypothesis we instead want to test $\theta = \alpha_1 - \alpha_2 = 0$ and plug it in the original regression:

$$\log(wage) = \alpha_0 + (\theta + \alpha_2) jc + \alpha_2 univ + \alpha_3 exper + u$$
$$\log(wage) = \alpha_0 + \theta jc + \alpha_2 (univ + jc) + \alpha_3 exper + u \qquad (2)$$

Now run:

genr unjc=univ+jc

ols lwage const jc unjc exper

	coefficient	std. error	t-ratio	p-value	
const	1.47233	0.0210602	69.91	0.0000	***
jc	−0.0101795	0.00693591	−1.468	0.1422	
unjc	0.0768762	0.00230873	33.30	2.96e-225	***
exper	0.00494422	0.000157474	31.40	4.12e-202	***

Mean dependent var	2.248096	S.D. dependent var	0.487692
Sum squared resid	1250.544	S.E. of regression	0.430138
R-squared	0.222442	Adjusted R-squared	0.222097
F(3, 6759)	644.5330	P-value(F)	0.000000
Log-likelihood	−3888.687	Akaike criterion	7785.374
Schwarz criterion	7812.651	Hannan-Quinn	7794.789

$\widehat{\alpha_1} - \widehat{\alpha_2} = -0.0102$ **so the return to a year at a junior college is about one percentage point less than a year at a university.**

Test statistic on jc t=0.0102/.0069 =-1.48. We need to compare this with one sided alternative critical value. At 10% one-sided significance level, critical value is -1.282. Therefore, there is some but not strong evidence against the null hypothesis.

Check also command: ols lwage const jc univ exper. Make your own observation!

(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.

summary phsrank

(ii) 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?

ols lwage const jc unjc exper phsrank

***phsrank* has a *t* statistic equal to only 1.25; it is not statistically significant. If we increase *phsrank* by 10, log(*wage*) is predicted to increase by (.0003)10 = .003. This implies a .3% increase in *wage*, which seems a modest increase given a 10 percentage point increase in *phsrank*.**

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

Adding *phsrank* makes the *t* statistic on *jc* even smaller in absolute value, about 1.33, but the coefficient magnitude is similar to (2). Therefore, the base point remains unchanged: the return to a junior college is estimated to be somewhat smaller, but the difference is barely significant with one-sided test.