Exercise session 5

Sulutions

Problem 1

Suppose that you have a sample of n individuals who apart from their mother tongue (Czech) can speak English, German, or are trilingual (i.e., all individuals in your sample speak in addition to their mother tongue at least one foreign language). You estimate the following model:

$$wage = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 exper + \beta_4 DM + \beta_5 Germ + \beta_6 Engl + \varepsilon \ ,$$

where

$educ$	...	years of education
IQ	...	IQ level
$exper$	...	years of on-the-job experience
DM	...	dummy, equal to one for males and zero for females
$Germ$	...	dummy, equal to one for German speakers and zero otherwise
$Engl$	...	dummy, equal to one for English speakers and zero otherwise

a. Explain why a dummy equal to one for trilingual people and zero otherwise is not included in the model.

If we included the dummy for people who are trilingual, we would have the complete set of dummies in the model (describing all three possible options - German speaker, English speaker, both foreign languages). Since we have the intercept in the model, this would lead to perfect multicollinearity.

b. Explain how you would test for discrimination against females (in the sense that *ceteris paribus* females earn less than males). Be specific: state the hypothesis, give the test statistic and its distribution.

For women, the dummy DM is equal to 0 and the model stands as follows:

$$\boldsymbol{wage = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 exper + \beta_5 Germ + \beta_6 Engl + \varepsilon} \ .$$

For men, the dummy DM is equal to 1 and the model stands as follows:

$$\boldsymbol{wage = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 exper + \beta_4 + \beta_5 Germ + \beta_6 Engl + \varepsilon} \ .$$

Therefore, ceteris paribus, the difference between the wage of men and the wage of women is equal to β_4. If this coefficient is positive, then men earn more than women. Hence, our hypothesis to be tested is

$$H_0 : \beta_4 \leq 0 \ \text{vs} \ H_A : \beta_4 > 0 \ .$$

This leads to a one-sided t-test with the test statistic

$$t = \frac{\widehat{\beta_4}}{SE(\widehat{\beta_4})} \sim t_{n-k}$$

where k = 7 in this case. When we compute this test statistic, we compare it to the critical value $t_{n-7,0.95}$. If the test statistic is larger than this critical value, then we reject the H_0 at 95% confidence level and we conclude that there is discrimination against females. **where k = 7 in this case. When we compute this test statistic, we compare it to the critical value $t_{n-7,0.95}$. If the test statistic is larger than this critical value, then we reject the H_0 at 95% confidence level and we conclude that there is discrimination against females.**

c. Explain how you would measure the payoff (in terms of wage) to someone of becoming trilingual given that he can already speak (i) English, (ii) German.

The payoff of a trilingual person is

$$wage = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 exper + \beta_4 DM + \beta_5 + \beta_6 + \varepsilon \ ,$$

the payoff of a German speaking person is

$$wage = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 exper + \beta_4 DM + \beta_5 + \varepsilon \ ,$$

and the payoff of an English speaking person is

$$wage = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 exper + \beta_4 DM + \beta_6 + \varepsilon \ .$$

Hence, by becoming trilingual, a person who can already speak English gains β_5 and a person who can already speak German gains β_6. If we assume that both coefficients are positive, this payoff should be positive.

d. Explain how you would test if the influence of on-the-job experience is greater for males than for females. Be specific: specify the model, state the hypothesis, give the test statistic and its distribution.

To allow the on-the-job experience to be greater for males than for females, we have to define a slope coefficient on *exper* that would be different for males and for females. We can do so using the following model:

$$wage = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 exper + \beta_4 DM + \beta_5 Germ + \beta_6 Engl + \beta_7 exper \cdot DM + \varepsilon \ .$$

Where we have created an interaction term exper*DM. In this case, the impact of on the on-the-job experience on wage would be β_3 for females and $\beta_3 + \beta_7$ for males. Hence, if β_7 is positive, then men gain more from experience than women. Hence, our hypothesis to be tested is

$$H_0 : \beta_7 \leq 0 \text{ vs } H_A : \beta_7 > 0 \ .$$

$$t = \frac{\widehat{\beta_7}}{SE(\widehat{\beta_7})} \sim t_{n-k}$$

where $\boldsymbol{k} =$ 8 in this case. When we compute this test statistic, we compare it to the critical value $t_{n-8,0.95}$. If the test statistic is larger than this critical value, then we reject the H_0 at 95% confidence level and we conclude that the influence of on-the-job experience is greater for males than for females.

Problem 2

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

$$\widehat{\log(rent)} = 0.43 + 0.066 \log(pop) + 0.507 \log(avginc) + 0.0056 pctstu + u$$

(0.844) (0 .039) (.081) (.0017)

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

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 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 the explanatory variables have no effect on wages. What do you conclude?

   Here we are testing a joint hypothesis that $\boldsymbol{\beta_1, \beta_2}$ *and* $\boldsymbol{\beta_3 = 0}$, 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 indicators 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

Given the following regression model:

$Inflation_i = \beta_0 + \beta_1 InterestRate_i + u_i$

Where both variables are measured in percentage points, a sample of 100 countries is used in order to estimate the above model and the following information is given:

$Var(Inflation_i) = 100$; $Var(InterestRate_i) = 50$; $Cov(Inflation_i, InterestRate_i) = -25$; $SSR = 49$

i) Find the OLS estimation of the effect of interest rates on inflation and the estimated standard error.

$$\beta_1 = -\frac{25}{50} = -0.5$$

$$\sigma^2 = VAR(u_i) = \frac{SSR}{n-2} = \frac{49}{98} \Longrightarrow \sigma = 0.71$$

$$SE(\beta_1) = \frac{\sigma}{\sqrt{n * var(Interest\ Rate)}} = \frac{0.71}{\sqrt{100 * 50}} = 0.01$$

ii) Interpret your estimation results. **If interest rate increases by 1%, then inflation reduces by 0.5 percentage points**

iii) Calculate a one-tailed t-test in order to validate the significance of the estimated slope coefficient at 1% significance level.

T=-0.5/0.01=-50 it is statistically significant at the 1% level

iv) What could you say about the explanatory power of the above model?

SST=n*Var(inflation)=100*100=10000

R^2= SSE/SST=(SST-SSR)/SST=(10000-49)/10000=0.99 very high explanatory power