ECON 471

Fall 2020

Problem Set 4

Due Monday November 9, by midnight

Name: Wenxiao Yang

Section: B3.

1. Suppose that the model

$$pctstck = \beta_0 + \beta_1 funds + \beta_2 risktol + u$$

Satisfies the first four Gauss-Markov assumptions, where $pctstck$ is the percentage of a worker's pension invested in the stock market, $funds$ is the number of mutual funds that the worker can choose from, and $risktol$ is some measure of risk tolerance (larger $risktol$ means the person has a higher tolerance for risk). If $funds$ and $risktol$ are positively correlated what is the inconsistency in $\tilde{\beta}_1$, the slope coefficient in the simple regression of $pctstck$ on $funds$?

$$\frac{Cov(funds, risktol)}{Var(funds)} \beta_2$$

2. Using the data in RDCHEM.TXT, the following equation was obtained by OLS:

$$\widehat{\text{rdintens}} = 2.613 + 0.00030 sales - 0.0000000070 sales^2$$

$$(0.429) \quad (0.00014) \qquad (0.0000000037)$$

$$n = 32, \quad R^2 = 0.1484.$$

(i) At what points does the marginal effect of *sales* on *rdintens* become negative?

$$\frac{\Delta \text{ rdintens}}{\Delta \text{ sales}} = 0.0003 - 0.000000014 \text{ sales} = 0.$$

$$\text{sales} = 21428.57143.$$

(ii) Would you keep the quadratic term in the model? Explain.

keep.

$\left|\frac{\hat{\beta}_2}{se(\hat{\beta}_2)}\right| > 1$, add $sales^2$ will increase $\bar{R}^2$

$\hat{\beta}_2$ is significant

(iii) Define *salesbil* as sales measured in billions of dollars: $salesbil = sales/1{,}000$. Rewrite the estimated equation with *salesbil* and *salesbil*^{2} as the independent variables. Be sure to report standard errors and the *R*-squared. [Hint: Note that *salesbil*^{2} = *sales*^{2}/(1,000)^2.]

$$\widehat{\text{rdintens}} = 2.611554 + 0.300633 \text{ salesbil}$$

$$(0.4295) \qquad (0.139314)$$

$$- 0.006944 \text{ salesbil}^2$$

$$(0.003727)$$

$$R^2 = 0.1485.$$

(iv) For the purpose of reporting the results, which equation do you prefer?

the latter one

3. The following model allows the return to education to depend upon the total amount of both parents' education, called *pareduc*:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 educ \times pareduc + \beta_3 exper + \beta_4 tenure + u.$$

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

$$\frac{\Delta \log(wage)}{\Delta educ} = \beta_1 + \beta_2 pareduc.$$

What sign do you expect for β_2? Why?

positive parent's education may help educated people get higer wage.

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

$$\widehat{\log(wage)} = \underset{(.13)}{5.65} + \underset{(.010)}{.047} educ + \underset{(.00021)}{.0078}\, educ \times pareduc + \underset{(.004)}{.019}\, exper + \underset{(.003)}{.010}\, tenure$$

$$n = 722, \quad R^2 = 0.169.$$

(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 compute the estimated return to *educ*.

for people whose parents have a college education will have 6.24 % more wage than whose parents have a high school education.

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

$$\widehat{\log(wage)} = \underset{(.38)}{4.94} + \underset{(.027)}{.097}educ + \underset{(.017)}{.033}\,pareduc - \underset{(.0012)}{.0016}\,educ \times pareduc$$

$$+\underset{(.004)}{.020}\,exper + \underset{(.003)}{.010}\,tenure$$

$$n = 722, \quad R^2 = 0.174.$$

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.

$H_0: \beta_3 = 0.$

$$t = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{-0.0016}{0.0012} = -1.333333$$

$\Rightarrow$ reject H_0

$\Rightarrow$ depend on parent education.

4. Use the data in WAGE1.TXT for this exercise.

(i) Use OLS to estimate the equation

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u$$

and report the results using the usual format.

$$\widehat{l\,wage} = \underset{(0.1059323)}{0.1279975} + \underset{(0.0074680)}{0.0903658}\,educ$$

$$+ \underset{(0.0051965)}{0.0410089}\,exper - \underset{(0.0001158)}{0.0007136}\,exper^2$$

(ii) Is $exper^2$ statistically significant at the 1% level?

$$t = \frac{-0.0007136}{0.0001158} = -6.162349$$

$\Rightarrow$ significant

(iii) Using the approximation

$$\%\Delta\widehat{wage} \approx 100(\hat{\beta}_2 + 2\hat{\beta}_3 exper)\Delta exper,$$

find the approximate return to the fifth year of experience. What is the approximate return to the twentieth year of experience?

$\rightarrow exper = 4$

$\rightarrow exper = 19$

5^{th}: 3.53 %

20^{th}: 1.38 %

(iv) At what value of $exper$ does additional experience actually lower predicted $\log(wage)$? How many people in the sample have more experience than the calculated turning point?

$\hat{\beta}_2 + 2\hat{\beta}_3 \, exper < 0.$

$$exper > \frac{-\hat{\beta}_2}{2\hat{\beta}_3} = 28.73381$$

121 peple

5.Use the data in MEAP00_01 for this exercise.

(i) Estimate the model

$$\text{math4} = \beta_0 + \beta_1 lexppp + \beta_2 lenroll + \beta_3 lunch + u$$

by OLS, and report the results in usual form, including the standard error of the regression. Is each explanatory variable statistically significant at the 5% level?

$$\widehat{math4} = 91.9324 + 3.52474\, lexppp - 5.39915\, lenroll - 0.44874\, lunch$$

$$n = 1692 \qquad R^2 = 0.3729$$

lenroll lunch are significant.

lexppp isn't significant.

(ii) Obtain the fitted values from the regression in part (i). What is the range of the fitted values? [0, 100]. How does it compare with the range of the actual data on *math4*?

range: [42.41416, 92.67098].

~~fit not well~~ much narrower.

(iii) Obtain the residuals from the regression in part (i). What is the building code of the school that has the largest (positive) residual? Provide an interpretation of this residual.

1141.

(iii) The largest residual is about 51.42, and it belongs to building code 1141. This residual is the difference between the actual pass rate and our best prediction of the pass rate, given the values of spending, enrollment, and the free lunch variable. If we think that per pupil spending, enrollment, and the poverty rate are sufficient controls, the residual can be interpreted as a "value added" for the school. That is, for school 1141, its pass rate is over 51 points higher than we would expect, based on its spending, size, and student poverty.

(iv) Add quadratics of all explanatory variables to the equation, and test them for joint significance. Would you leave them in the model? Explain.

$$F = \frac{(0.3735-0.3729)/3}{(1-0.3735)/1685} = 0.5379010184.$$

$\Rightarrow$ joint insignificant

$\Rightarrow$ wouldn't leave them

(v) Returning to the model in part (i), divide the dependent variable, and each explanatory variable by its sample standard deviation, and rerun the regression. In terms of standard deviation units, which explanatory variable has the largest effect on the math pass rate?

$$\widehat{\frac{math4}{sd(math4)}} = 4.76176 + 0.03474\frac{lexppp}{sd(lexppp)} - 0.11462\frac{lenroll}{sd(lenroll)} - 0.61285\frac{lunch}{sd(lunch)}$$

lunch.

6. Use the data in WAGE2.TXT for this exercise.

(i) Estimate the model

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 married$$

$$+\beta_5 black + \beta_6 south + \beta_7 urban + u$$

And report the results in the usual form. Holding other factors fixed, what is the approximate difference in monthly salary between blacks and nonblacks? Is this difference statistically significant?

$$\widehat{\log(wage)} = 5.395497 + 0.065431\, educ + 0.014043\, exper$$
$$+ 0.011747\, tenure + 0.199417\, married$$
$$- 0.188350\, black - 0.090904\, south$$
$$+ 0.183912\, urban.$$
$$n = 925 \quad R^2 = 0.255$$

0.188350 higher for nonblacks.

Yes.

(ii) Add the variable $exper^2$ and $tenure^2$ to the equation and show that they are jointly insignificant at even the 20% level.

$$F = \frac{(0.255 - 0.2526)/2}{(1-0.255)/925} = 1.489932886.$$

$$< F_{\text{critical } 20\%}$$

+ β_8 black·educ.

(iii) Extend the original model to allow the return to education to depend on race and test whether the return to education does depend on race.

add interaction: black·educ.

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 married$$

$$+\beta_5 black + \beta_6 south + \beta_7 urban + \beta_8\, black\cdot educ + u.$$

$\beta_8 = -0.022624.$ p-value $= 0.26260$

doesn't depend.

(iv) Again, start with the original model, but now allow wages to differ across four groups of people: married and black, married and nonblack, single and black, single and nonblack. What is the estimated wage differential between married blacks and married nonblacks?

$$\widehat{\log(wage)} = 5.56730 + 0.064741\, educ + 0.015248\, exper + 0.012318\, tenure - 0.200392\, black - 0.086433\, south + 0.180590\, urban.$$

difference is $\underline{0.200392}$ (blacks are lower).

$$\log(\widehat{wage}) = \underset{(0.11)}{5.40} + \underset{(.0063)}{.0655}\, educ + \underset{(.0032)}{.0141}\, exper + \underset{(.0025)}{.0117}\, tenure$$

$$- \underset{(.026)}{.092}\, south + \underset{(.027)}{.184}\, urban + \underset{(.043)}{.189}\, marrnonblck$$

$$- \underset{(.096)}{.241}\, singblck + \underset{(.0560)}{.0094}\, marrblck$$

$n = 935, \quad R^2 = .253.$

We obtain the ceteris paribus differential between married blacks and married nonblacks by taking the difference of their coefficients: $.0094 - .189 = -.1796$, or about $-.18$. That is, a married black man earns about 18% less than a comparable, married nonblack man.