ECON 471

Fall 2020

Problem Set 4

Due Monday November 9, by midnight

Name: Albert Wiryawan

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 $funds$?

An increase in risk tolerance will indicate that the worker has more will willingness to invest in the stock market. As such, β_2 > 0 and we can use:

$$\text{Plim}(\tilde{\beta}_1) = \beta_1 + \beta_2\delta_1 \text{ , where } \delta_1 > 0$$

$$\delta_1 = Cov(x_1, x_2)/Var(x_1)$$

Since *fund* and *risktol* are said to be positively correlated, $\delta_1 > 0$. In addition, x_2 has positive partial effect on the dependent variable *pctstck* so that β_2 > 0. As a result, the inconsistency in $\tilde{\beta}_1$ is positive.

Therefore, $\tilde{\beta}_1$ has positive inconsistency.

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

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

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

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

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

Marginal effect of *sales* on *rdintens* is given by

$$\frac{\Delta \widehat{rdintens}}{\Delta sales} = 0.00030 - 2(0.000000007) sales$$

In order for marginal effect to be negative $\frac{\Delta \widehat{rdintens}}{\Delta sales} < 0$

So,

Sales > 0.00030/0.000000014

Sales > 21428.5714

Therefore, at 21428.5714 million dollars of firm sales the marginal effect of sales on rdintens is negative

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

Calculate t-statistic for $sales^2$

$$t - statistic_{sales^2} = \frac{|\hat{\beta}_{sales^2}|}{SE_{sales^2}} = \frac{0.000000007}{0.0000000037} = 1.8919$$

At 29 degrees of freedom and 5% level of significance for a single tail, the t-statistic is 1.699 which is less than the calculated value above. As a result, the variable $sales^2$ is statistically significant and should be included in the model.

(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.]

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

$$\underset{(0.429)}{\widehat{\text{rdintens}} = 2.613} + \underset{(0.14)}{0.30 salesbil} - \underset{(0.0037)}{0.0070 sales^2}$$

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

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

For the purpose of reporting, I would use the formula created in part (iii) as there are less decimals minimizing complexity.

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?

$$\log(wage)_{educ1} - \log(wage)_{educ0} = \beta_1(educ_1 - educ_0) + \beta_2 paraeduc(educ_1 - educ_0)$$
$$\Delta \log(wage) = \beta_1 \Delta educ + \beta_2 pareduc * \Delta educ$$
$$\Delta \log(wage) = (\beta_1 + \beta_2 * pareduc) * \Delta educ$$
$$\Delta \log(wage)/\Delta educ = (\beta_1 + \beta_2 * pareduc)$$

The sign of β_2 is positive since the individual is expected to have higher wages with each additional year of education when they are from parents with high education qualification.

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

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

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

(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*.

The coefficient of the interaction term is given by 0.0078. If there is an increase in parent education and observer education by one year then the wage is supposed to increase by .78%.

$\beta_1 + \beta_2 * paraeduc$

0.047 + 0.0078paraeduc

At Paraeduc = 24 => 0.6572 ; paraeduc = 32 => 0.7196

0.7196 - 0.6572 = 0.0624

There is a 6.24% increase in return from education when parents go from being high school graduates to college ones

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

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

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

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

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.

Since the coefficient of the interaction term is negative (0.0016), the estimated return to education is now negatively dependent

$$H_0: \beta_{educ*pareduc} = 0$$

$$H_0: \beta_{educ*pareduc} \neq 0$$

$$t - statistic = \frac{|\hat{\beta}_{educ*pareduc}|}{SE_{educ*pareduc}} = \frac{0.0016}{0.0012} = 1.33$$

Since the predetermined t-statistic for 721 degrees of freedom at 5% significance level is 1.96, the found t-statistic is less than this. As a result, the null hypothesis is not rejected indicating that the interaction coefficient is not significant at the 5% level. This highlights that return to education does not 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.

Code

````
```{r}
setwd("/Users/albertwiryawan/Code/Class_Repos/Econometrics/Problem Set #4")
#load in Attend txt data set for the exercise
wage1 = read.table("wage1.txt")

attach(wage1)
wage = V1
educ = V2
exper = V3

model = lm(log(wage) ~ educ + exper + I(exper^2))
summary(model)
coefficients(model)
```
````

```
Call:
lm(formula = log(wage) ~ educ + exper + I(exper^2))

Residuals:
     Min       1Q   Median       3Q      Max 
-1.96387 -0.29375 -0.04009  0.29497  1.30216 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.1279975  0.1059323   1.208    0.227    
educ         0.0903658  0.0074680  12.100  < 2e-16 ***
exper        0.0410089  0.0051965   7.892 1.77e-14 ***
I(exper^2)  -0.0007136  0.0001158  -6.164 1.42e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4459 on 522 degrees of freedom
Multiple R-squared:  0.3003,	Adjusted R-squared:  0.2963 
F-statistic: 74.67 on 3 and 522 DF,  p-value: < 2.2e-16

  (Intercept)          educ         exper    I(exper^2) 
 0.1279975072  0.0903658158  0.0410088753 -0.0007135582 
```

$$\log(wage) = 0.128 + 0.0904educ + 0.0410exper - 0.000714exper^2$$

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

The t-statistic on exper^2 is -6.164. The p-value is extremely small indicating that exper^2 is significant at the 1% significance level.

(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?

$$\%\Delta wage = 100(0.0410 - 2(0.000714)4) = 3.53\%$$

$\%\Delta wage = 100(0.0410 - 2(0.000714)20) = 1.39\%$

(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?

The point where the additional experience lowers predicted log(wage) is (0.041)/ ((2)(0.000714)) = 28.7 years. There are 121 samples with 29 years of experience.

5. Use the data in MEAP00_01 for this exercise.

(i) Estimate the model

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

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?

$$\underset{(17.9148)}{\mathrm{math4} = 46.1953} + \underset{(1.8365)}{8.5341 lexppp} - \underset{(0.9176)}{4.8396 lenroll} - \underset{(0.0143)}{0.4714 lunch}$$

Code:

```
attach(meap1)

lexppp = V11
lenroll = V9
lunch = V5
math4 = V3
mod = lm(math4 ~lexppp + lenroll +lunch)
summary(mod)
```

Each explanatory variable's coefficient has a small p-value (close to zero) this shows that each explanatory variable is statistically significant.

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

Code:

min = 46.1953 + 8.5341*min(lexppp)-4.8396*max(lenroll)-0.4714*max(lunch)

max = 46.1953 + 8.5341*max(lexppp)-4.8396*min(lenroll)-0.4714*min(lunch)

The range of the fitted values from the regression in part (i) is [24.231, 106.350]. The actual range of math4 is [0,100] as this is the percentage of students with satisfactory math. As such, the model has a possibility of producing predictions that are not possible

(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.

Code:

max(resid(mod)) = [1] 53.23674

match(max(resid(mod)), x) = 1269

The building code of the school that has the largest positive residual is 1269 with a residual value of 53.23674. School 1269 has a pass rate that is over 53 percent of what we expect based on lexppp, lenroll, and lunch. As such, the school shows a large proneness improving with the increase of these factors.

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

Code:

lexppp = V11
lenroll = V9
lunch = V5

```
math4 = V3
mod = lm(math4 ~lexppp + lenroll +lunch + I(V1^2) +I(V2^2) + I(V4^2) +I(V6^2) +I(V7^2) + I(V8^2) +
I(V10^2))
summary(mod)
```

The addition of quadratics of all explanatory variables prove to be significant as the F-value obtained from the fit results in a large F-statistic that obtains a small p-value close to zero to reject the null hypothesis. This would mean that the quadratics are jointly significant and should be added to the model.

(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?

The obtained explanatory variables are 1.989513, -0.5768535, and -0.0008942557 for lexppp, lenroll, and lunch respectively. This shows that lexppp has the largest effect on math pass rate in terms of standard deviation.

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?

Code

```
educ = V5
exper = V6
tenure =V7
married = V9
black =V10
south =V11
urban = V12
wage = V1
mod = lm(log(wage) ~ educ + exper + tenure + married +black + south + urban)
summary(mod)
```

$$\log(wage) = 5.395497 + 0.065431educ + 0.014043exper + 0.011747tenure + 0.199417married - 0.188350black - 0.090904\,south + 0.183912\,urban + u$$

Black is a binary option. Since being black decreases salary as indicated by the negative coefficient, it is said that black salary is about 18.84% lower than those that aren't black. The p-value for this calculated coefficient is extremely close to zero indicating that this wage difference is statistically significant.

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

The calculated F-statistic probability is 0.226 which is greater than the 20% level. Indicating that exper^2 and tenure^2 are statistically insignificant at the 20% significance level.

(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.

Code

```
educ = V5
exper = V6
tenure =V7
married = V9
black =V10
south =V11
urban = V12
wage = V1
mod = lm(log(wage) ~ educ + exper + tenure + married +black + south + urban + I(educ *black))
```

The found coefficient for educ *black is -0.02264 and the corresponding p-value is 0.2626. This p-value is larger than the critical p-value at 5% level of significance thus indicating that the return to education does on depend on race.

(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

The estimated wage differential between married*black and married*nonblack is the difference in their coefficients which is 0.188195-0.009448=0.179467

This means that married*nonblack earns 17.95% more in their monthly wages