ECON471
Fall 2020
Problem Set 3
Due Tuesday October 27, by 11:59pm

Name: Wenxiao Yang

Section: B3

1. For the case of the multiple regression problem with two explanatory variables, $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i$, show that minimizing the sum of squared residuals results in three conditions:

$$\sum_{i=1}^{n} \hat{u}_i = 0; \sum_{i=1}^{n} \hat{u}_i X_{i1} = 0; \sum_{i=1}^{n} \hat{u}_i X_{i2} = 0$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}$$

$$\min \ SSR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left[(\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1) X_{i1} + (\beta_2 - \hat{\beta}_2) X_{i2} + u_i \right]^2$$

$$\frac{\partial SSR}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} -2 \hat{u}_i = 0 \implies \sum_{i=1}^{n} \hat{u}_i = 0$$

$$\frac{\partial SSR}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} - X_{i1} \cdot 2 \hat{u}_i = 0 \implies \sum_{i=1}^{n} \hat{u}_i X_{i1} = 0$$

$$\frac{\partial SSR}{\partial \hat{\beta}_2} = \sum_{i=1}^{n} - X_{i2} \cdot 2 \hat{u}_i = 0 \implies \sum_{i=1}^{n} \hat{u}_i X_{i2} = 0$$

2. The following model can be used to study whether campaign expenditures affect election outcomes:

$$VoteA = \beta_0 + \beta_1 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystrA + u,$$

where $VoteA$ is the percentage of the vote received by Candidate A, $expendA$ and $expendB$ are campaign expenditures by Candidates A and B, and $prtystrA$ is a measure of party strength for Candidate A (the percentage of the most recent presidential vote that went to A's party).

(i) What is the interpretation of β_1?

$\Delta VoteA = \approx \beta_1 \, r_{expandA} + \beta_2 \, r_{expandB} + \beta_3 \Delta prtystrA$

$= \frac{\beta_1}{100} \% expand A + \frac{\beta_2}{100} \% expand B + \beta_3 \Delta prtystrA$

interpretation of β_1: every 1% expend A inereases, VoteA increases $\frac{\beta_1}{100}$

Holding other factors fixed, $\frac{\beta_1}{100}$ is the percentage point change in voteA when expend A increases by 1%

(ii) In terms of the parameters, state the null hypothesis that a 1% increase in A's expenditures is offset by a 1% increase in B's expenditures.

$H_0: \beta_1 + \beta_2 = 0$

(iii) Estimate the given model using the data in VOTE1.TXT and report the results in usual forms. Do A's expenditures affect the outcome? What about B's expenditures? Can you use these results to test the hypothesis in part (ii)?

$VoteA = 45.079 + 6.03 \log(expendA) - 6.615 \log(expendB) + 0.152 prtystrA$

$n = 173 \quad R^2 = 0.793$

Yes, A's expenditures affect the outcome, so does B's expeditures

No, I can't use the result to test hypothesis in (ii).

(iv) Estimate a model that directly gives the t statistic for testing the hypothesis in part (ii). What do you conclude? (Use a two-sided test.)

$H_0: \beta_1+\beta_2=0$, Let $\beta_4=\beta_1+\beta_2$

$$VoteA = \beta_0 + \beta_1[\log(expendA) - \log(expendB)] + \beta_4 \log(expendB) + \beta_3\, prtystrA + u$$

$$VoteA = 45.0789 + 6.0833\,[\log(expendA) - \log(expendB)] - 0.5321 \log(expendB) + 0.1520\, prtystrA$$

$t_{\beta 4} = -0.998$ $\quad Pr(>|t|) = 0.3196$

accept H_0

3. The following model is used to study the tradeoff between time spent sleeping and working and to look at other factors affecting sleep:

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

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

(i) If adults trade off sleep for work, what is the sign of β_1?

negative

(ii) What signs do you expect for β_2 and β_3?

β_2: negative β_3: positive

(iii) The equation was estimated using the data in SLEEP75:

$$\widehat{sleep} = \underset{(112.28)}{3638.25} - \underset{(.017)}{0.148}\, totwork - \underset{(5.88)}{11.13}\, educ + \underset{(1.45)}{2.20}\, age$$

$n = 706,\ R^2 = 0.113.$

where the standard errors of the estimates are shown in parentheses. If someone works five more hours per week, by how many minutes is sleep predicted to fall? Is this a large trade off?

fall 0.74 hours it is not a large trade off

(iv) Discuss the sign and magnitude of the estimated coefficient on *educ*.

if other factors are same

one more year of education

sleep is predicted to fall 11.13 hours per week

(v) Would you say *totwrk*, *educ*, and *age* explain much of the variation in *sleep*?

No, R^2 is too low only 11.3% is explanied.

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

$$t_{educ} = \frac{-11.13}{5.88} = -1.892857143$$

$$t_{age} = \frac{2.20}{1.45} = 1.517241379$$

$t_{critical}$ is same for educ and age $\quad n-k-1 = 706-4 = 702$

$t_{critical} = 1.960$, $\quad t_{educ} \in [-1.960, 1.960]$
$\quad t_{age} \in [-1.960, 1.960]$

educ and age are significant

(vii) Dropping *educ* and *age* from the equation produces

$$\widehat{sleep} = 3586.38 - 0.151\, totwork$$

(38.91) (.017)

$n = 706,\ R^2 = 0.103.$

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

$$F = \frac{(0.113-0.103) \times \frac{1}{2}}{(1-0.113) \times \frac{1}{702}} = 3.957158963$$

$f_{critical} = 3.00 \qquad F > f_{critical}$

Hence educ and age aren't joiatly significant.

(viii) Suppose that the sleep equation contains heteroskedasticity. What does this mean about the tests computed in parts (vi) and (vii)?

the t statistics and F statistics will be invalid.

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

(i) Consider the standard wage equation

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 \text{exper} + \beta_3 tenure + u.$$

State the null hypothesis that another year of general workforce experience has the same effect on *log(wage)* as another year of tenure with the current employer.

$H_0: \beta_2 = \beta_3$

(ii) Test the null hypothesis in part (i) against a two-sided alternative, at the 5% significance level, by constructing a 95% confidence interval. State your conclusion.

$H_0: \beta_3 - \beta_2 = 0$ Set $\beta_4 = \beta_3 - \beta_2$

$Log(wage) = \beta_0 + \beta_1 educ + \beta_2(exper + tenure) + \beta_4 tenure + u.$

$$P(-c < \frac{\hat{\beta}_4 - \beta_4}{se(\hat{\beta}_4)} < c) = 95\%$$

$$P(-1.960 < \frac{-0.001954 - \beta_4}{0.004743} < 1.960) = 95\%$$

$\beta_4 \in [-0.01125028, 0.00734228]$

confidence interval.

accept H_0

5. The data set 401KSUBS.TXT contains information on net financial wealth (*nettfa*), age of the survey respondent (*age*), annual family income (*inc*), family size (*fsize*), and participation in certain pension plans for people in the United States. The wealth and income variables are both recorded in thousands of dollars. For this question use only the data for single-person households (so *fsize*=1).

(i) How many single-person households are there in the data set?

2017

(ii) Use OLS to estimate the model

$$nettfa = \beta_0 + \beta_1 inc + \beta_2 \text{age} + u,$$

and report the results using the usual format. Be sure to use only the single-person households in the sample. Interpret the slope coefficients. Are there any surprises in the slope estimates?

$$nettfa = -43.0398 + 0.7993\, inc + 0.8427\, age$$

no surprises.

(iii) Does the intercept from the regression in part (ii) have an interesting meaning? Explain.

not interesting

the mean net financial wealth when income is zero and age is zero.

no one close to these values in relevant population.

(iv) Find the p-value for the test $H_0: \beta_2 = 1$ against $H_1: \beta_2 < 1$. Do you reject H_0 at the 1% significance level? Explain.

$$t = \frac{\beta_2 - 1}{se(\hat{\beta}_2)} = \frac{0.84266-1}{0.09202} = -1.709845686$$

$$p\text{-value} = Pr(T < -1.709845686) \in [0.025, 0.05] = 0.044 > 1\%$$

$\Rightarrow$ reject H_0

(v) If you do a simple regression of *nettfa* on *inc*, is the estimated coefficient on *inc* much different from the estimate in part (ii)? Why or why not?

$\tilde{\beta}_0 = -10.5709$, $\tilde{\beta}_1 = 0.8207$

change a few but not too much

inc and age have linear relationship but not perfect collinearity.

6. Regression analysis can be used to test whether the market efficiently uses information in valuing stocks. For concreteness, let *return* be the total return from holding a firm's stock over the four-year period from the end of 1990 to the end of 1994. The *efficient market hypothesis* says that these returns should not be systematically related to information known in 1990. If firm characteristic known at the beginning of the period help to predict stock returns, then we could use this information in choosing stocks.

For 1990, let *dkr* be a firm's debt to capital ratio, let *eps* denote the earnings per share, let *netinc* denote net income, and let *salary* denote total compensation for the CEO.

(i) Using the data in RETURN, the following equation was estimated:

$$\widehat{return} = -14.37 + .321\, dkr + .043\, eps - .0051\, netinc + .0035\, salary$$

$$(6.89) \quad (.201) \qquad (.078) \qquad (.0047) \qquad\qquad (.0022)$$

$$n = 142, \quad R^2 = .0395.$$

Test whether the explanatory variables are jointly significant at the 5% level. Is any explanatory variable individually significant?

$$F = \frac{R^2/4}{(1-R^2)/137} = 1.408 < f_{critical} = 2.37$$

reject => jointly significant.

p-value of dkr : 0.113 > 0.05

eps : 0.586 > 0.05

netinc : 0.276 > 0.05

salary : 0.113 > 0.05

accept => all explanatory vars are not individually significant

(ii) Now, reestimate the model using the log form for *netinc* and *salary*:

$$\widehat{return} = -36.30 + .327\, dkr + .069\, eps - 4.74 \log(netinc) + 7.24 \log(salary)$$

$$(39.37) \quad (.203) \qquad (.080) \qquad (3.39) \qquad\qquad (6.31)$$

$$n = 142, \quad R^2 = .0330.$$

Do any of your conclusion from part (i) change?

$F = \frac{0.033/4}{0.967/137} = 1.16882 1096 < f_{critical} = 2.37$

reject ⇒ still jointly significant.

p-value of dkr : 0.109
eps : 0.395
l netinc : 0.163
l salary : 0.253

None changes.

accept ⇒ still all not individually significant

(iii) In this sample, some firms have zero debt and some have negative earnings. Should we try to use log(*dkr*) or log(*eps*) in the model to see if these improve the fit? Explain.

No, if dkr = 0 or eps
log(dkr) or log(eps) will not exist.

(iv) Overall, is the evidence for predictability of stock returns strong or weak?

weak, R^2 is really low
no matter in (i) or (ii).