620464127
Waritthha Sawangsang

EE325 Introductory Econometrics (Section 1 semester 1/2020)

Assignment 4

Instruction: Write your answer in either paper or digital paper. However, if you write on paper, please scan it and save as a PDF file. Submission is via BE-Moodle as a PDF file for both cases. (Please keep the file below 10MB as that is the maximum per file capacity for student.)

Due date: Friday, November 6,2020 (Before 10 P.M.)

1. From the data for 46 states in the United States for 1992, results of the regression are displayed as follows.

$$
\begin{aligned}
ln\, C_i &= 4\,.30 - 1.34\, ln\, P_i + 0.17\, ln\, Y_i \\
se \quad &= (0.91)\ \ (0.32) \qquad (0.20) \\
\bar{R}^2 &= 0.27
\end{aligned}
$$

where C_i = cigarette consumption, packs per year

P_i = real price per pack, \$ per pack

Y_i = real disposable income per capita, \$ per week

1.1) Do the estimation results follow the law of demand?

1.2) What is the elasticity of demand for cigarettes with respect to price? Is it statistically significant? If so, is it statistically different from 1?

1.3) What is the income elasticity of demand for cigarettes? Is it statistically significant? If not, what might be the reasons for it?

2. From estimating the regression equation on net financial wealth (nettfa), age of the survey respondent (age), and annual family income (inc) for people in the United States. The wealth and income variables are both recorded in thousands of dollars. The OLS estimation results for the model are given by

$$nettfa_i = \beta_1 + \beta_2 inc_i + \beta_3 age_i + u_i$$

```
reg nettfa inc age

      Source |       SS           df       MS      Number of obs   =     9,275
-------------+----------------------------------   F(2, 9272)      =    943.21
       Model |   6414618.8         2   3207309.4   Prob > F        =    0.0000
    Residual |  31528770.7     9,272  3400.42825   R-squared       =    0.1691
-------------+----------------------------------   Adj R-squared   =    0.1689
       Total |  37943389.5     9,274   4091.3726   Root MSE        =    58.313

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |   .9533566   .0252775    37.72   0.000     .9038072    1.002906
         age |   1.030777   .0591226    17.43   0.000     .9148838     1.14667
       _cons |  -60.69654   2.596333   -23.38   0.000    -65.78592   -55.60715
------------------------------------------------------------------------------
```

```
reg nettfa inc age agesq

      Source |       SS           df       MS      Number of obs   =     9,275
-------------+----------------------------------   F(3, 9271)      =    646.80
       Model |  6567017.15         3  2189005.72   Prob > F        =    0.0000
    Residual |  31376372.3     9,271  3384.35685   R-squared       =    0.1731
-------------+----------------------------------   Adj R-squared   =    0.1728
       Total |  37943389.5     9,274   4091.3726   Root MSE        =    58.175

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |   .9782522   .0254891    38.38   0.000      .928288    1.028216
         age |  -2.231489   .4897118    -4.56   0.000    -3.191432   -1.271547
       agesq |   .0377221   .0056214     6.71   0.000      .026703    .0487413
       _cons |   4.680388   10.08099     0.46   0.642    -15.08056    24.44134
------------------------------------------------------------------------------
```

2.1) Test the coefficient, in the first model, $\beta_3 < 1$ in the first model or not?

2.2) Due to estimation result by adding the age^2 variable or agesq. Perform the test whether we should include the quadratic term of the age variable or not? (Test for both t-test and F-test.) Also, interpret the meaning of this coefficient.

3. You are conducting an empirical investigation into the median prices of houses in 506 communities of a large metropolitan area. The sample data consist of 506 observations on the following observable variables:

P_i: the median house price in community i, in dollars;

NOX_i: the level of nitrous oxide in the air of community i, in parts per 100 million;

$DIST_i$: the weighted distance of community i from municipal area, in miles;

$ROOM_i$: the average number of rooms per house in community i;

$STRAT_i$: the average student-teacher ratio of schools in community i.

Researcher estimates the following model of median house price. The OLS estimation results for the model are given by

$$\begin{aligned} ln(P_i) &= 11.08 - 0.9535\, ln(NOX_i) - 0.1343\, ln(DIST_i) + 0.2545 ROOM_i - 0.05245 STRAT_i \\ se &= (0.3181)\ (0.1167) \qquad (0.04310) \qquad (0.01853) \qquad (0.005897) \\ RSS &= 35.1835 \quad TSS = 84.5822 \end{aligned}$$

3.1) Interpret each of the coefficient estimates in regression equation.

3.2) Test the individual significance of each of the slope coefficient estimates for $ln(NOX_i)$ and $ROOM_i$.

3.3) Find the R-squared, adjusted R-squared, and test the joint significance of all the slope coefficient estimates.

3.4) If researcher would like to test the proposition that the marginal effect of $ln(NOX_i)$ on $ln(P_i)$ equals the marginal effect of $ln(DIST_i)$ on $ln(P_i)$, write the restricted model and perform the test comparing restricted and unrestricted model, given that OLS estimation of this restricted regression equation yields a Residual Sum of Squares value = 41.9532.

4. Production function (Y) of the industrial sector in Thailand. It depends on the capital factor (K) and labor factor (L) in the years 1980-2010 with the following estimation.

Model 1:

$$ln\, Y_t = 18.27 + 0.536\, ln\, L_t + 0.024\, ln\, K_t$$

$$R^2 = 0.9389, RSS = 0.0124$$

Model 2:

$$ln\left(\frac{Y}{L}\right)_t = 2.13 + 1.12\, ln\left(\frac{K}{L}\right)_t$$

$$R^2 = 0.8087, RSS = 0.0153$$

4.1) Interpret the coefficients of the independent variables in models 1 and 2.

4.2) Test the hypothesis. Is the industrial production function characterized by constant return to scale? (Hint: you can perform any type of test that you see fit.)

4.3) Can we compare the R^2 value between the two regression models? Why?

6204641127
Waritthà Sawangsang

1. From the data for 46 states in the United States for 1992, results of the regression are displayed as follows.

$$\ln C_i = 4.30 - 1.34 \ln P_i + 0.17 \ln Y_i$$
$$se = (0.91) \quad (0.32) \qquad (0.20)$$
$$\bar{R}^2 = 0.27$$

where C_i = cigarette consumption, packs per year

P_i = real price per pack, \$ per pack

Y_i = real disposable income per capita, \$ per week

1.1) Do the estimation results follow the law of demand?

Yes, refer to given regression equation, Cigarette consumption has negative relationship with real price while it has positive relationship with real disposable income which all of it follow law of demand that if price is low, people want to buy more; high demand. However, if price is high, individuals will demand less. Moreover, if people receive higher income, they will consume more. Therefore, these relationship is follow law of demand.

1.2) What is the elasticity of demand for cigarettes with respect to price? Is it statistically significant? If so, is it statistically different from 1?

Price elasticity = −1.34

H_0 : elasticity of demand for cigarettes with respect to price is 1

H_a : otherwise

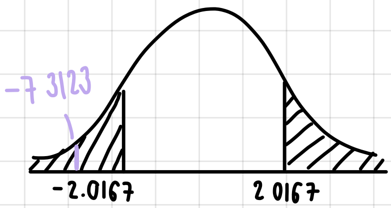
t-test

$$t_{cal} = \frac{-1.34-1}{0.32} = -7.3123 \sim t_{46-3}$$

Let $\alpha = 0.05$; 95% confident that estimator different from 1

$t_{lower} = -2.0167$

$t_{upper} = 2.0167$



∴ We reject H_0 ; elasticity of demand for cigarettes with respect to price for 95% is not 1, and it is statistically significant.

1.3) What is the income elasticity of demand for cigarettes? Is it statistically significant? If not, what might be the reasons for it?

H_0 : income elasticity of demand of cigarettes = 0

H_a : otherwise

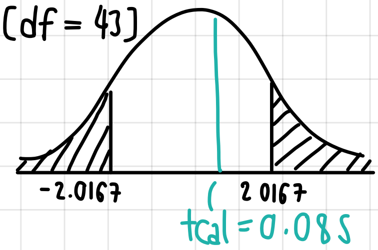
t-test

$$t_{cal} = \frac{0.17-0}{0.2} = 0.085 \sim t_{46-3}$$

Let $\alpha = 0.05$; 95% confident that estimator different from 0 (df = 43)

$t_{lower} = -2.0167$

$t_{upper} = 2.0167$



∴ Cannot reject H_0 ; income elasticity of demand for cigarettes is not 0 for 95% confident, and it is not statistically significant.

2. From estimating the regression equation on net financial wealth (nettfa), age of the survey respondent (age), and annual family income (inc) for people in the United States. The wealth and income variables are both recorded in thousands of dollars. The OLS estimation results for the model are given by

$$nettfa_i = \beta_1 + \beta_2 inc_i + \beta_3 age_i + u_i$$

2.1) Test the coefficient, in the first model, $\beta_3 < 1$ in the first model or not?

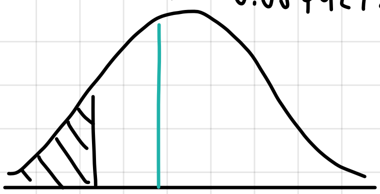
$H_0 : \beta_3 \geq 1$

$H_a : \beta_3 < 1$

<u>t-test</u> d.f = 9275 − 3 = 9272

$$t_{cal} = \frac{\hat{\beta}_3 - \beta_3}{\hat{\sigma}_{\hat{\beta}_3}} = \frac{1.030777 - 1}{0.0591226} = 0.5206 \sim t_{9272}$$

$\alpha = 0.05$; $t_{0.05, 9272} = -1.96$



−1.96 $t_{cal} = 0.5206$

∴ We can not reject H_0.

Therefore, we can say for sure that β_3 is less than one for 95% confident.

2.2) Due to estimation result by adding the age² variable or agesq. Perform the test whether we should include the quadratic term of the age variable or not? (Test for both t-test and F-test.) Also, interpret the meaning of this coefficient.

<u>F-test</u>

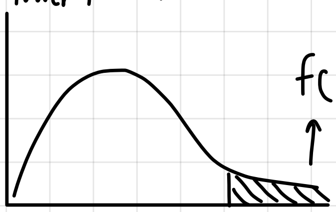
$H_0 : \beta_4 = 0$ [Agesq has contribution to the model]

$H_a : \beta_4 \neq 0$

$$F_{cal} = \frac{RSS_{new} - RSS_{old} / (\text{no. of new regressors})}{RSS_{new}/(n - K_{new})} = \frac{31528770.7 - 31376372.3 / 1}{31376372.3/(9275-4)} = 45.0302$$

$\alpha = 0.05$

$F_{lower, \alpha}(1, 9271) = 3.84$



$F_{cal} = 45.0302$

3.84

Therefore, we can reject H_0 because $F_{cal} > 3.84$. The additional of agesq does increase ESS and hence R^2 value at 95% confident. So, the estimator is significant and give more detail to the model

T-test

$H_0 : \beta_4 = 0$ [Agesq has contribution to the model]

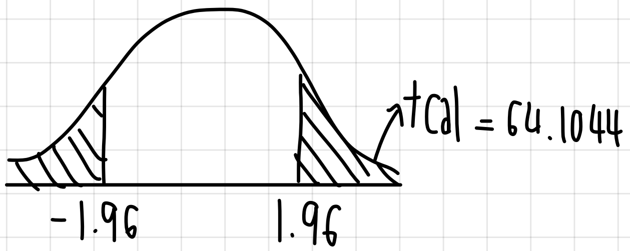
$H_a : \beta_4 \neq 0$

$t_{cal}(agesq) = \dfrac{0.377221 - 0}{0.0056214} = 67.1044$ ⇒ absolute t-value more than 1, which lead to higher $\bar{R}^2$

$\alpha = 0.05$ df $= 9271$

$t_{lower} = -1.96$

$t_{upper} = 1.96$



∴ We can reject H_0. The additional of agesq does increase ESS and hence R^2 value at 95% confident. This estimator is significant.

3.1) Interpret each of the coefficient estimates in regression equation.

$\hat{\beta}_1$ is 11.08; if $\ln(NOX_i) = 0$, $\ln(DIST_i) = 0$, and $STRAT_i = 0$, on average, the median house price $= e^{11.08} = 64{,}860.88349$ dollars.

$\hat{\beta}_2$ is -0.9535, so if level of nitrous oxide in the air of community increase 1 unit, median house price decrease by 0.9535 units on average, keeping other variables constant. $\hat{\beta}_3$ is -0.1343, so if the weighted distance of community increase 1 unit, median house price decrease by 0.1343 units on average, keeping other variables constant. $\hat{\beta}_4$ is 0.2545, so if average number of rooms per house in community increase 1 unit, median house price increase by 0.2545 units on average, keeping other variables constant. $\hat{\beta}_5 = 0.05245$, so if average student-teacher ratio of schools in community increase by 1 unit, median house price decrease by 0.0525 units on average, keeping other variables constant.

3.2) Test the individual significance of each of the slope coefficient estimates for $ln(NOX_i)$ and $ROOM_i$.

$\ln(NOX_i)$

$H_0: \beta_2 = 0$

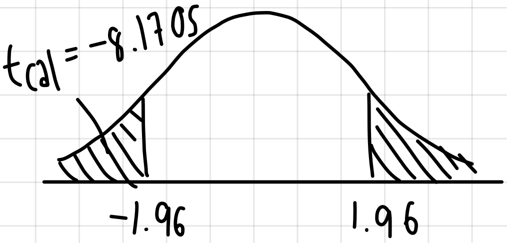
$H_a: \beta_2 \neq 0$

$t_{cal} = \dfrac{-0.9535-0}{0.1167} = -8.1705 \sim t_{506-5}$

$\alpha = 0.05$, d.f. $= 501$

$t_{upper} = 1.96$

$t_{lower} = -1.96$



∴ we can reject H_0 and say that β_2 is not zero with 95% confident and β_2 is significant.

$ROOM_i$

$H_0: \beta_5 = 0$

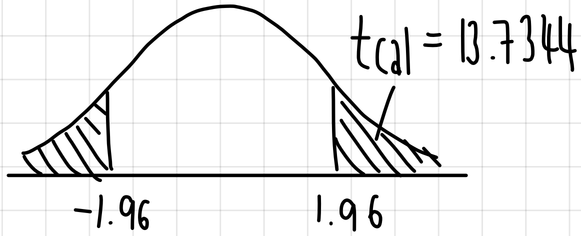
$H_a: \beta_5 \neq 0$

$t_{cal} = \dfrac{0.2545-0}{0.01853} = 13.7344 \sim t_{506-5}$

Let $\alpha = 0.05$, d.f $= 501$

$t_{upper} = 1.96$

$t_{lower} = -1.96$



∴ we can reject H_0 and say that β_5 is not zero with 95% confident and β_5 is significant.

3.3) Find the R-squared, adjusted R-squared, and test the joint significance of all the slope coefficient estimates.

$ESS = TSS - RSS$
$= 84.5822 - 35.183$
$= 49.3987$

$R^2 = \dfrac{ESS}{TSS} = \dfrac{84.5822}{49.3987} = 1.7122$

Adjust $R^2 = \dfrac{1 - RSS/(n-k)}{TSS/(n-1)}$
$= \dfrac{1 - 35.183/501}{84.5822/505} = -0.4081$

__Joint-test__

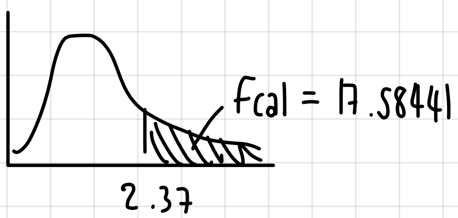
$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_1: not all slope coefficient are simultaneously zero

$F_{cal} = \dfrac{ESS/(k-1)}{RSS/(n-k)} = \dfrac{R^2/(k-1)}{(1-R^2)/(n-k)}$

$= \dfrac{49.3987/4}{35.1835/501} = 175.8441$

$\alpha = 0.05$, d.f. $= 501 \Rightarrow F_{table} = 2.37$



∴ we can reject H_0, so we can say that at least one parameter is not equal to zero with 95% confident.

3.4) If researcher would like to test the proposition that the marginal effect of $ln(NOX_i)$ on $ln(P_i)$ equals the marginal effect of $ln(DIST_i)$ on $ln(P_i)$, write the restricted model and perform the test comparing restricted and unrestricted model, given that OLS estimation of this restricted regression equation yields a Residual Sum of Squares value = 41.9532.

<u>f-test</u>

$H_0: \beta_2 = \beta_3$ or $\beta_2 - \beta_3 = 0$

$H_a: \beta_2 \neq \beta_3$ or $\beta_2 - \beta_3 \neq 0$

unrestricted

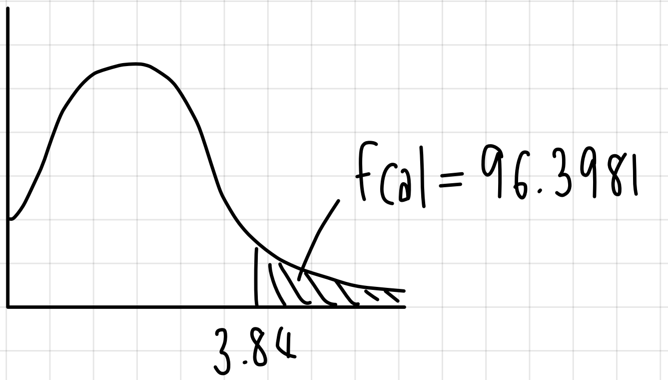
$\Rightarrow \ln(p_i) = \beta_1 + \beta_2 \ln(NOX_i) + \beta_3 \ln(DIST_i) + \beta_4 Room_i + \beta_5 STRAT_i + u_i$

$= \beta_1 + \beta_2 \ln(NOX_i) + \beta_4 Room_i + \beta_5 STRAT_i + u_i$

restricted $= \beta_1 + \beta_2 \ln(NOX_i) + (DIST_i) + \beta_4 Room_i + \beta_5 STRAT_i + u_i$

$$F_{cal} = \frac{RSS_R - RSS_{UR}/m}{RSS_{UR}/(n-k)} = \frac{41.9532 - 35.1835/1}{35.1835/(506-5)} = 96.3980$$

$\alpha = 0.05$, d.f. $= 501$; $F_{table} = 3.84$



∴ We can reject H_0. We can say that $\beta_2 \neq \beta_3$ or $\beta_2 - \beta_3 = 0$ with 95 % confident.

4.1) Interpret the coefficients of the independent variables in models 1 and 2.

Model 1

$$\ln Y_t = 18.27 + 0.536 \ln L_t + 0.024 \ln K_t$$

from the equation, elasticity of output depend on labor and capital. $\ln L_t$ is output elasticity of labor, and $\ln K_t$ is output elasticity of capital. $\hat{\beta}_1$ is 18.27. $\hat{\beta}_2$ is 0.5631, coefficient of output elasticity of labor; as labor increase by 1 unit, output rise by 0.536 units; keep other independent variables constant $\hat{\beta}_3$ is 0.1024, coefficient of output elasticity of capital; as capital increase by 1 unit, output rise by 0.024 units; keep other independent constant. If $\ln L_t$ and $\ln K_t = 0$, output $= e^{18.27}$ on average.

4.2) Test the hypothesis. Is the industrial production function characterized by constant return to scale? (Hint: you can perform any type of test that you see fit.)

Model 1

$$\ln Y_t = \beta_1 + \beta_2 \ln L_t + \beta_3 \ln K_t + u_t \Rightarrow \text{unrestricted regression}$$

Model 2

$$\ln Y_t = \beta_1 + \beta_3 \ln\left(\frac{K}{L}\right) + V_t \Rightarrow \text{restricted regression}$$

test that $\beta_2 + \beta_3 = 1$

$H_0: \beta_2 + \beta_3 = 1$ or the restriction is valid

$H_a: \beta_2 + \beta_3 \neq 1$ or the restriction is not valid

$$F_{cal} = \frac{(R^2_{UR} - R^2_R)/m}{(1 - R^2_{UR})/(n - k_{UR})} = \frac{(0.9389 - 0.8087)/1}{(1 - 0.9389)/27} = 57.5351 \sim F_{(1,27)}$$

$\alpha = 0.05$

$F_{0.05, 1.27} \approx F_{0.05, 1.26} = 4.23$

$F_{cal} = 57.5351 > F_{0.05, 1.27} = 4.23$

$\therefore$ We can reject H_0 because $F_{cal} > F_{0.05, 1.27}$. This industrial production function is not a constant return to scale at 95% confident.

4.3) Can we compare the R^2 value between the two regression models? Why?

Model 1 use $\ln Y_t$ to be dependent variable while model 2 use $\ln (Y/L)_t$ to be dependent variable. So, these 2 models cannot be compare R^2 value between two resgression models.