# 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.

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

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?

- Yes. The law of demand states that quantity demanded has a reverse relation to price. The coefficient of the price term is negative, therefore, this result follows 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?

- The elasticity is 1.34, which means that when price increases for 1 percent, quantity demanded for cigarette decreases by 1.34 percent.

- It is **statistically significant** (different from zero) at the significance level of 0.05 since

$t_{cal} = \frac{-1.34-0}{0.32} = -4.1875, t_{cri} \approx \pm 2.021$ we **can** reject the null hypothesis. (n=46, df.=43)

- It is **also statistically different from 1** at the significance level of 0.05 since

$$t_{cal} = \frac{-1.34 - 1}{0.32} = -7.3125,$$

$$t_{cri} = [1 - 2.021 * 0.32\, , 1 + 2.021 * 0.32] = [-0.3533\, , 1.6467]$$

we can reject the null hypothesis.

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

- The coefficient representing income elasticity of demand is 0.17 but it is **not statistically significant** since

$t_{cal} = \frac{0.17-0}{0.2} = 0.85, t_{cri} \approx \pm 2.021$ we **cannot** reject the null hypothesis.

The reason might be that quantity demanded is not relevant to income. In other words, people at every level may have similar smoking pattern, at least from this dataset.

**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?

- Below is out null hypothesis (n=9,275, df.=9,272 or we can assume that it tends to infinity in this case.) for the level of significance at 0.05.

$H_0: \beta_3 \geq 1, H_a: \beta_3 < 1$

$$t_{cal} = \frac{1.0308 - 1}{0.0591} = 0.5212, t_{cri} \leq 1 - [0.0591 * 1.645] = -0.9028$$

We cannot reject the null hypothesis. Therefore, $\beta_3$ is not statistically less than 1.

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.

- For the t-test, we can see from the table that agesq is significantly different from zero.

- For the F-test, we compare the model with and without the agesq variable. Note that the degrees of freedom are 1 (number of new regressors) and 9,271 (number of observation minus number of all regressors in the new model)

$H_0$: $agesq$ has no marginal contribution.

$H_a$: $agesq$ has a marginal contribution.

$$F_{cal} = \frac{R^2_{new} - R^2_{old}/(\#\ of\ new\ regressors)}{1 - R^2_{new}/(n - k_{new})} = \frac{0.1731 - 0.1691/(1)}{1 - 0.1731/(9{,}275 - 4)}$$

$$= \frac{0.004}{0.000089} \approx 44.9438$$

$F_{cri,0.05,(1,\infty)} = 3.84$

Therefore, we can reject our null hypothesis. In other words, agesq has a contribution to our model and should be included in this model.

- The reason why agesq is significant because age and net financial wealth is not correlated linearly, but in a parabolic form. Net wealth may come down at first (as it decreases with age) when people are in their working age since they own a house or a car with mortgage, which is a very long-term payment. Once they pay off all those mortgages, their net wealth increases later in their life.

**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.

- $NOX_i$: **A percentage increases** in level of nitrous oxide in an area leads to a **decrease** of 0.9535 **percent** in median house price on average. The more a community is polluted, the cheaper house price.

- $DIST_i$: **A percentage increases** in distance between a community and a town leads to a **decrease** of 0.1343 **percent** in median house price on average. The farther away from a town, the cheaper house price.

- $ROOM_i$: **An increase** in average number of rooms leads to an **increase** of 0.2545 **percent** in median house price on average. The community with big houses is usually more expensive.

- $STRAT_i$: **An increase** in student-teacher ratio leads to a **decrease** of 0.05245 **percent** in median house price on average. This ratio may be a proxy for how concentrated population in a community is. Houses in a more concentrated area is likely to be cheaper.

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

- Testing the coefficient of $ln(NOX_i)$

$t_{cal} = \frac{-0.9535-0}{0.1167} = 8.1705$ ; $t_{cri,0.05,501} = \pm 1.96$ we can reject the null hypothesis.

- Testing the coefficient of $ROOM_i$

$t_{cal} = \frac{0.2545-0}{0.01853} = 13.7345$ ; $t_{cri,0.05,501} = \pm 1.96$ we can reject the null hypothesis.

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

- $R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{35.1835}{84.5822} = 0.5840$

- $\bar{R}^2 = 1 - \frac{RSS/(n-k)}{TSS(n-1)} = 1 - \frac{35.1835/(506-5)}{84.5822/(506-1)} \approx 0.5809$

- Overall significance: $F_{cal} = \frac{R^2/k-1}{1-R^2/n-k} = \frac{0.5840/5-1}{1-0.5840/506-5} = \frac{0.146}{0.0008} \approx 182.5$

$F_{cri,0.05,(4,\infty)} = 2.37$ so we can reject the null hypothesis and conclude that $R^2$ is significantly different from zero or the overall model is statistically significant.

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.

- Original model: $ln(P_i) = \beta_1 + \beta_2\, ln(NOX_i) + \beta_3\, ln(DIST_i) + \beta_4 ROOM_i + \beta_5 STRAT_i + u_i$

- New model assumes $\beta_2 = \beta_3$ then the model becomes

$ln(P_i) = \beta_1 + 2\beta_2\, ln(NOX_i) + \beta_4 ROOM_i + \beta_5 STRAT_i + u_i$ or

$ln(P_i) = \beta_1 + 2\beta_3\, ln(DIST_i) + \beta_4 ROOM_i + \beta_5 STRAT_i + u_i$

Note that these two models will result in different $R^2$ value, but we do not know which one that yields RSS = 41.9532.

- Testing restricted and unrestricted model

$H_0$: the marginal effect of $ln(NOX_i)$ equals the marginal effect of $ln(DIST_i)$

$H_a$: otherwise

$$F_{cal} = \frac{RSS_R - RSS_{UR}/m}{RSS_{UR}/n - k_{UR}} = \frac{41.9532 - 35.1835/1}{35.1835/506 - 5} = \frac{6.7697}{0.0702} \approx 96.4345$$

$F_{cri,0.05,(1,\infty)} = 3.84$ we can reject the null hypothesis, meaning that the restricted model is not valid. Note: look at the formula we will see that the $RSS_R - RSS_{UR}$ is high so that we can reject or $RSS_R$ is too high which means that the unexplained part of the restricted model is larger than the original model.

**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.

- The first model coefficients represent labor and capital elasticity to output. Increase in labor and capital by 1 percent leads to 1 percent increase in output by 0.536 and 0.024 respectively.

The second model coefficients, on the other hand, represent capital per labor elasticity to output. A percentage of capital per labor increase leads to 1.12 percent increase in output.

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.)

$H_0$: constant return to scale

$H_a$: otherwise

$F_{cal} = \frac{RSS_R - RSS_{UR}/m}{RSS_{UR}/n - k_{UR}} = \frac{0.0153 - 0.0124/1}{0.0124/31-3} = \frac{0.0029}{0.0004} \approx 7.25$

$F_{cri,0.05,(1,28)} = 4.2$ therefore, we can reject the null hypothesis or the industrial sector of Thailand is not characterized by constant return to scale.

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

- No, because the dependent variable in these models are different. $R^2$ is the coefficient of determination, calculated from explanatory power of a set of $X$ on $Y$. If $Y$ is different in two models, $R^2$ cannot be compared.