

## 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.

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

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

$$\text{nettfa}_i = \beta_1 + \beta_2 \text{inc}_i + \beta_3 \text{age}_i + u_i$$

reg nettfa inc age

Source	SS	df	MS	Number of obs	=	9,275
Model	6414618.8	2	3207309.4	F(2, 9272)	=	943.21
Residual	31528770.7	9,272	3400.42825	Prob > F	=	0.0000
				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
Model	6567017.15	3	2189005.72	F(3, 9271)	=	646.80
Residual	31376372.3	9,271	3384.35685	Prob > F	=	0.0000
				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<sup>2</sup> 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

$$\ln(P_i) = 11.08 - 0.9535 \ln(NOX_i) - 0.1343 \ln(DIST_i) + 0.2545 ROOM_i - 0.05245 STRAT_i$$

$$se = (0.3181) \quad (0.1167) \quad (0.04310) \quad (0.01853) \quad (0.005897)$$

$$RSS = 35.1835 \quad TSS = 84.5822$$

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:**

$$n = 2010 - 1980 + 1$$

$$n = 31$$

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

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

**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?

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

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

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?

According to the estimation result, the consumption and price is negatively related. The law of demand is that the quantity demanded and price is negatively related, thus the increase in price will make people demand less, finally consumption falls. So, this estimation goes with the law of demand.

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?  
→ the call demand

Elasticity of demand with respect to price  $\beta_2$  which is  $-1.34$ .

$$H_0: \beta_2 = 1$$

$$H_1: \beta_2 \neq 1$$

$$t(\alpha|\beta_2) = \frac{\hat{\beta}_2 - \beta_2}{\delta \hat{\beta}_2} \sim t_{n-3}$$
$$= \frac{-1.34 - 1}{0.32} = -7.3125$$

$$\alpha = 0.05; t_{0.025, 43} = 2.021$$

$n = 48$   
 $n - k = 43$

$$t_{\text{upper}} = 2.021$$

$$t_{\text{lower}} = -2.021$$

$t_{\beta_2}$  lies beyond the boundaries so reject  $\beta_2$ , so it is different at the 95% confidence interval

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

$$1. H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

$$2. t(\text{cal}(\beta)) = \frac{\hat{\beta}_3 - \beta_3}{\delta \beta_3} = \frac{0.17 - 0}{0.2} = 0.85$$

$$3. \alpha = 0.05 \quad i t_{0.0025, 43} = 2.021$$

$$n = 46$$

$$n - k = 43$$

$$t_{\text{upper}} = 2.021$$

$$t_{\text{lower}} = -2.021$$

4.  $t_{\beta_3}$  lies within the boundaries, so we cannot reject  $H_0$ ,  
as the result, we cannot calculate that  $\beta_3$  is significant  
at 95% confidence interval.

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

$$\text{nettfa}_i = \beta_1 + \beta_2 \text{inc}_i + \beta_3 \text{age}_i + u_i$$

reg nettfa inc age

Source	SS	df	MS	Number of obs	=	9,275
Model	6414618.8	2	3207309.4	F(2, 9272)	=	943.21
Residual	31528770.7	9,272	3400.42825	Prob > F	=	0.0000
				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
Model	6567017.15	3	2189005.72	F(3, 9271)	=	646.80
Residual	31376372.3	9,271	3384.35685	Prob > F	=	0.0000
				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

h

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

$$\text{nettf}_i = \beta_1 + \beta_2 \text{inc}_i + \beta_3 \text{age}_i + u_i$$

$$\widehat{\text{nettf}_i} = -60.696 + 0.9533 \text{inc}_i + 1.3077 \text{age}_i$$
$$\text{se} = (2.596) \quad (0.0252) \quad (0.0591)$$

$$2.1 H_0: \beta_3 \geq 1$$

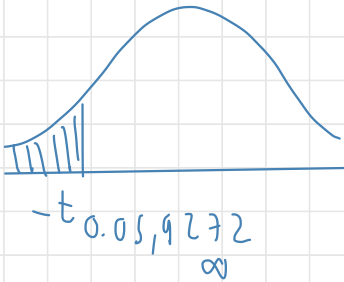
$$H_1: \beta_3 < 1$$

$$\lambda = 0.05$$

$$t_{\text{cal}} = \frac{\hat{\beta}_3 - \beta}{\text{se}(\hat{\beta}_3)} = \frac{1.0307 - 1}{0.0591} = 0.520512$$

$$df = n - k = 9275 - 3 = 9272$$

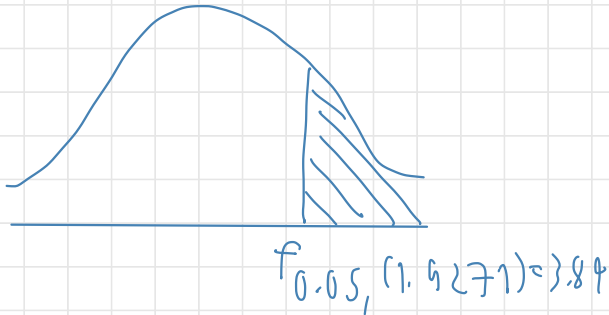
critical value



2.2) Due to estimation result by adding the age<sup>2</sup> 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.

$$\begin{aligned}
 F_{cal} &= \frac{RSS_R - RSS_{UR} / m}{RSS_{UR} / n - k} \\
 &= \frac{R^2_{UR} - R^2_R / m}{(1 - R^2_{UR}) / n - k} \\
 &= \frac{(31528990.7 - 31376372.3) / 2}{313763.3 / 9275 - 4} \\
 &= 45.03024
 \end{aligned}$$

critical value



∴ reject  $H_0$  if  $F_{cal} > 3.84$

find that  $45.03 > 3.82$

$\therefore$  reject  $H_0$

conclusion: There is enough evidence to say that the addition of  $\text{age}^2$  to the model has significant, at  $\alpha = 0.05$

6-201 (81)  
**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

$$\ln(P_i) = 11.08 - 0.9535 \ln(NOX_i) - 0.1343 \ln(DIST_i) + 0.2545 ROOM_i - 0.05245 STRAT_i$$

$se$	(0.3181)	(0.1167)	(0.04310)	(0.01853)	(0.005897)
$RSS = 35.1835$	$TSS = 84.5822$				

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

$\hat{\beta}_0 = 11.08$ ; If  $\ln(NOX_i) = 0$ ,  $\ln(DIST_i) = 0$ ,  $ROOM_i = 0$ , and  $STRAT_i = 0$  on average the median house price would be equal to  $8e^{11.08} = 89,860.883498$

$\hat{\beta}_1 = -0.9535$ ; as the level of nitrous oxide increase by 1%, on average, the median house price decrease by about 0.9535%, holding another variable constant.

$\hat{\beta}_2 = -0.1343$ ; As the weighted distance from municipal area increase by 1%, on average, the median house price decrease by about 0.1343%, holding other variable constant.

$\hat{\beta}_3 = 0.2545$ ; as the average number of rooms per house increases by 1 room per house on average the median house price increase by about  $0.2545 \times 100 = 25.45\%$ , holding other variable constant.

$\hat{\beta}_4 = -0.05245$ ; as the average student-teacher ratio of school increase by 1 student per teacher, on average, the median house price decrease by about  $0.05245 \times 100 = 5.245\%$ , holding other variable constant.

3.2) Test the individual significance of each of the slope coefficient estimates for  $\ln(\text{NOX}_i)$  and  $\text{ROOM}_i$ .

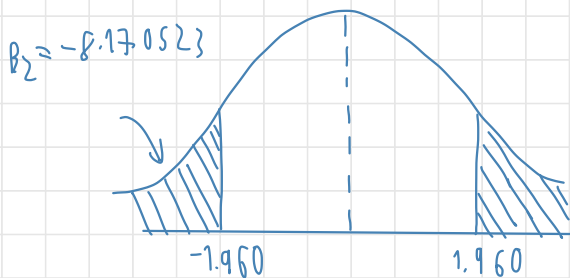
$\ln(\text{NOX}_i)$

$$H_0: \beta_2 = 0$$

$$H_a: \beta_2 \neq 0$$

$$-t(\text{cal}(\beta_2)) = \frac{\hat{\beta}_2 - \beta_2}{\hat{b}_{\beta_2}} = \frac{-0.4535 - 0}{0.1167} = -8.970523$$

$$\alpha = 0.05 \quad \text{d.f.} = 501 (506 - 5) = \infty$$



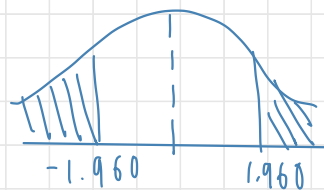
$\beta_2$  is significant

$\text{ROOM}_i$

$$H_0: \beta_5 = 0$$

$$H_1: \beta_5 \neq 0$$

$$\alpha = 0.05 \quad \text{d.f.} = 501$$



$\beta_5$  is significant

$$t(\text{cal}(\beta_5)) = \frac{\hat{\beta}_5 - \beta_5}{\hat{b}_{\beta_5}} = \frac{0.2545 - 0}{0.0155} = 16.42$$

3.3

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

$$R^2 = ESS/TSS = 84.5822/99.3987 = 0.8511$$

$$\text{adjusted } R^2 = \frac{1 - RSS/n - k}{TSS/n - 1} = \frac{1 - 35.183/501}{84.5822/501} = \frac{0.06822}{0.167158}$$

$$= 0.408176$$

Joint

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

$H_1$ : Not all slope coefficients are simultaneously zero

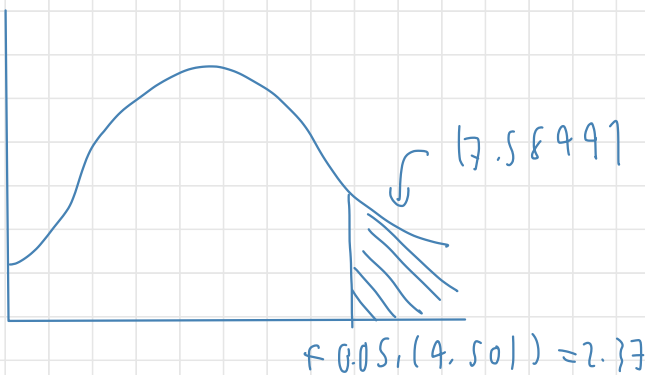
$$\alpha = 0.05$$

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

$$= \frac{49.3987/4}{35.1835/501} = \frac{12.3497}{0.070225} = 175.3999$$

$$= 175.3999$$

critical value



$\therefore$  reject  $H_0$

$\hookrightarrow$  This is enough evidence to say that at least one parameter is not equal to zero, at  $\alpha = 5\%$ .

3.9

$$H_0: \beta_2 = \beta_3 \text{ or } \beta_2 - \beta_3 = 0$$

$$H_1: \beta_2 \neq \beta_3 \text{ or } \beta_2 - \beta_3 \neq 0$$

$$\alpha = 0.05$$

$$\ln(p_i) = \beta_1 + \beta_2 \ln(\text{NOX}_i) + \beta_3 \ln(\text{DIST}_i) + \beta_4 \text{ROOM}_i + \beta_5 \text{STRAT}_i + u_i$$

(unrestricted)

$$= \beta_1 + \beta_2 \ln(\text{NOX}_i) + \beta_2 \ln(\text{DIST}_i) + \beta_4 \text{ROOM}_i + \beta_5 \text{STRAT}_i + u_i$$

$$= \beta_1 + \beta_2 [\ln(\text{NOX}_i) + \ln(\text{DIST}_i) + \beta_4 \text{ROOM}_i + \beta_5 \text{STRAT}_i]$$

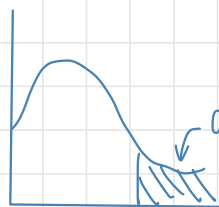
$$= \beta_1 + \beta_2 \ln(\text{NOX}_i) \cdot (\text{DIST}_i) + \beta_4 \text{ROOM}_i + \beta_5 \text{STRAT}_i + u_i$$

(restricted)

$$F_{\text{cal}} = \frac{RSS_R - RSS_{UR}}{RSS_{UR} / n - k}$$

$$= \frac{41.9532 - 35.1835 / 1}{35.1835 / 506 - 5} = 96.39801$$

critical value



$\therefore$  reject  $H_0 \therefore \beta_2 \neq \beta_3$

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:**

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

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

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

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

model 1:  $\ln \hat{Y}_t = 18.27 + 0.536 \ln L_t + 0.024 \ln K_t$

output elasticity of capital

output elasticity of labor

$\frac{1.04}{1.0L}$ 
 $\frac{1.04}{1.0K}$

$$\epsilon_x^y = \frac{1.04}{1.0X}$$

$$\hat{\beta}_1 = 18.27$$

$\ln L_t = 0$ , and  $\ln K_t = 0$  on average  
 output would be equal to  $e^{18.27}$

$\hat{\beta}_2 = 0.536$ , As Labor increase by 1%, on average,

output increase by about 0.536%  
 holding other variables constant

$\hat{\beta}_3 = 0.1024$ , As capital increase by 1%, on average, output increases by about 0.1024%, holding other variables constant

model 2:

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

assume that

$$\hat{\alpha}_1 = 2.13$$

$$\hat{\alpha}_2 = 1.12$$

$\hat{\alpha}_1 = 2.13$  if  $\ln(K/L) = \delta$ , on average, output per labor ratio (labor productivity) would be equal to  $e^{2.13}$

$$\hat{\alpha}_2 = 1.12$$

As capital per labor ratio increase by 1%, on average labor productivity increase by about 1.12%, holding other variable constant.

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.)

Sol

$$\text{model 1: } \ln Y_t = \beta_1 + \beta_2 \ln L_t + \beta_3 \ln K_t + U_t \\ = \text{unrestricted regression}$$

$$\text{model 2: } \ln Y_t = \beta_1 + \beta_3 \ln \left(\frac{K}{L}\right)_t + U_t \text{ is restricted regression}$$

$$\text{condition is } \beta_2 + \beta_3 = 1$$

$$H_0: \beta_2 + \beta_3 = 1$$

$$H_1: \beta_2 + \beta_3 \neq 1$$

$$\alpha = 0.05$$

$$F = \frac{(R^2_{UR} - R^2_R) / m}{(1 - R^2_{UR}) / (n - k_{UR})} = \frac{(0.9389 - 0.8087) / 1}{(1 - 0.9389) / 27} \\ = 57.535$$

$$\text{find critical value } F_{0.05, 1, 27} \approx F_{0.05, 1, 26} = 4.23$$

reject  $H_0$  when  $F > F_{0.05, 1, 27}$  because

$$F = 57.535 > F_{0.05, 1, 26} = 4.23$$

So, reject  $H_0$

To conclude, production function in Thailand doesn't be a constant return to scale at significant level 0.05

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

We cannot compare  $R^2$  value between two regression models because model 1 and 2 has different variables. Model 1 has  $\ln y_t$  as dependent variable while model 2 has  $\ln(\frac{y}{L})$  as dependent variable although both models have equal number of information.