

Problem Set VI

Question 1

Consider the simple linear regression model as follows $Y_t = \beta_0 + \beta_1 (X_t - \bar{X}) + u_t$, $t=1, 2, \dots, n$ where the error terms are independently distributed with zero mean and constant variance, σ^2 . Find the least square estimates of β_0 and β_1 .

Solution:

$$\text{Let } \sum \hat{u}_t^2 = \text{SSR} = \sum \hat{u}_t^2 = \sum (Y_t - \hat{Y}_t)^2 = \sum [Y_t - \beta_0 - \beta_1 (X_t - \bar{X})]^2$$

$$\text{a. } \frac{\partial S}{\partial \beta_0} = -2 \sum [Y_t - \beta_0 - \beta_1 (X_t - \bar{X})] = 0$$

$$\text{b. } \frac{\partial S}{\partial \beta_1} = -2 \sum \{ [Y_t - \beta_0 - \beta_1 (X_t - \bar{X})] (X_t - \bar{X}) \} = 0$$

From (a) we have;

$$\sum [Y_t - \beta_0 - \beta_1 (X_t - \bar{X})] = 0$$

$$\sum Y_t - \sum \beta_0 - \beta_1 \sum (X_t - \bar{X}) = 0$$

$$\sum Y_t - \beta_1 \sum (X_t - \bar{X}) = \sum \beta_0$$

$$\sum Y_t - \beta_1 [\sum X_t - T \cdot \bar{X}] = \sum \beta_0$$

$$\frac{\sum Y_t - \beta_1 [\sum X_t - T \cdot \bar{X}]}{T} = \frac{\sum \beta_0}{T}$$

$$\hat{\beta}_0 = \bar{Y} - \beta_1 (T \cdot \bar{X} - T \cdot \bar{X})$$

$$\hat{\beta}_0 = \bar{Y}$$

From (b) we have;

$$\sum \{ [Y_t - \beta_0 - \beta_1 (X_t - \bar{X})] (X_t - \bar{X}) \} = 0$$

$$\sum Y_t (X_t - \bar{X}) - \beta_0 \sum (X_t - \bar{X}) - \beta_1 \sum (X_t - \bar{X})^2 = 0$$

$$\sum Y_t (X_t - \bar{X}) - \beta_0 \sum (X_t - \bar{X}) = \beta_1 \sum (X_t - \bar{X})^2$$

$$\sum Y_t (X_t - \bar{X}) = \beta_1 \sum (X_t - \bar{X})^2$$

$$\hat{\beta}_1 = \frac{\sum Y_t (X_t - \bar{X})}{\sum (X_t - \bar{X})^2}$$

Question 2

Consider the following regression model: $Y_t = \beta_0 + \beta_1 X_t + u_t$. Show that $R^2 = 1 - \frac{\sum \tilde{u}_t}{\sum y_t}$.

Solution:

Define $Y_t - \bar{Y} = y_t$ then add and subtract $\hat{Y}_t$ then we are left with

$$Y_t - \bar{Y} = (\hat{Y}_t - \bar{Y}) + (Y_t - \hat{Y}_t) = (\hat{Y}_t - \bar{Y}) + \hat{u}_t$$

Now square and sum w.r.t the index

$$\sum (Y_t - \bar{Y})^2 = \sum [(\hat{Y}_t - \bar{Y}) + (\hat{u}_t)]^2$$

Then we use the normal equations to conclude $2\Sigma (\hat{Y}_t - \bar{Y}) (\hat{u}_t) = 0$ this shows that when we expand the square we are left with

$$\sum y_t^2 = \sum (\hat{Y}_t - \bar{Y})^2 + \sum (\hat{u}_t)^2$$

From here we divide by $\sum y_t^2$ to get

$$1 = \left(\frac{\sum \hat{y}_t^2}{\sum y_t^2} \right) + \frac{\sum \hat{u}_t^2}{\sum y_t^2}$$

And from here using the definition of R^2 to conclude

$$R^2 = 1 - \frac{\sum \hat{u}_t^2}{\sum y_t^2}$$

Question 3

Show that $r_{xy} = \sqrt{R^2}$ in simple linear regression model.

Solution:

$$\begin{aligned} R^2 &= \frac{\sum \hat{y}_t^2}{\sum y_t^2} = \frac{\sum \hat{y}_t \hat{y}_t}{\sum y_t^2} \text{ since } \hat{y}_t = \hat{\beta}_1 x_t \\ R^2 &= \hat{\beta}_1 \frac{\sum x_t \hat{y}_t}{\sum y_t^2} \text{ and we know } \hat{\beta}_1 = \frac{\sum x_t y_t}{\sum x_t^2}; \text{ so} \\ R^2 &= \frac{\sum x_t y_t}{\sum y_t^2} \frac{\sum x_t y_t}{\sum x_t^2} = \frac{(\sum x_t y_t)^2}{\sum x_t^2 \sum y_t^2} \\ \sqrt{R^2} &= \frac{\sum x_t y_t}{\sqrt{\sum x_t^2 \sum y_t^2}} = r_{xy} \end{aligned}$$

Question 4

Assume the population regression function is as follows and satisfies the assumptions of CLRM:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

By stating the problem explain the properties of the OLS estimators if:

- $cov(X_i, u_i) \neq 0$
- $var(u_i) = \sigma_i^2$
- $cov(u_i, u_j) \neq 0$ where $i \neq j$

Solution:

a. $Cov(X_i, u_i) \neq 0$

Endogeneity problem. Explanatory variable is correlated with the disturbance term. This problem causes the OLS estimator to be biased and inconsistent.

b. $var(u_i) = \sigma_i^2$

Heteroscedasticity problem. Variance of the disturbance term varies in each observation. This problem causes inefficiency in OLS estimator. Variance of the OLS estimator is not minimum.

c. $cov(u_i, u_j) \neq 0$ where $i \neq j$

Autocorrelation problem. Disturbance term is correlated with its lagged observations. This problem causes inefficiency in OLS estimator.

Question 5

Assume the population follows the regression function:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

where $\hat{u}_i$ denotes the residuals and $\hat{Y}_i$ is the fitted values show that $\sum y_i = \sum \hat{y}_i$ and $\sum \hat{y}_i \hat{u}_i = 0$

Solution:

First note that we have the following normal equations

$$\sum \hat{u}_i = 0 \quad \text{and} \quad \sum x_i \hat{u}_i = 0$$

Observe that $0 = \sum \hat{u}_i = \sum (y_i - \hat{y}_i) = \sum y_i - \sum \hat{y}_i$ then $\sum y_i = \sum \hat{y}_i$ and we are done.

For the second proposition from second normal equation, we have

$$\sum \hat{y}_i \hat{u}_i = \sum \hat{u}_i (\hat{\beta}_0 + \hat{\beta}_1 x_i) = \sum \hat{u}_i \hat{\beta}_0 + \sum \hat{u}_i \hat{\beta}_1 x_i = \hat{\beta}_0 \sum \hat{u}_i + \hat{\beta}_1 \sum x_i \hat{u}_i$$

Then by first normal equation $\hat{\beta}_0 \sum \hat{u}_i = \hat{\beta}_0 \cdot 0 = 0$ and by second normal equation $\hat{\beta}_1 \sum x_i \hat{u}_i = \hat{\beta}_1 \cdot 0 = 0$ then $\sum \hat{y}_i \hat{u}_i = 0$

Question 6

Given the following simple linear regression model without intercept

$$Y_i = \beta_1 X_i + u_i$$

a. Derive the OLS estimator of the slope.

b. Does $\sum \hat{u}_i = 0$ hold?

Solution:

Note that in this case we will have only 1 normal equation to see that observe

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 X_i)^2$$

and once we derivate it with respect to the slope term we will have the equation:

- a. $\Sigma (Y_i - \hat{\beta}_1 X_i) X_i = \sum \hat{u}_i X_i = 0$ not that this is the only normal equation. So by using it we can compute the OLS estimator for the slope term as $\sum (Y_i X_i) - \hat{\beta}_1 \Sigma X_i^2 = 0$ from this clearly we are left with

$$\hat{\beta}_1 = \frac{\sum Y_i X_i}{\sum X_i^2}$$

- b. Since we do not have the normal equation $\sum \hat{u}_i = 0$ like we did in the model with the intercept term this will not hold.

Question 7

Consider the following regression equations:

$$Y_t = \beta_0 + \beta_1 X_t + u_t \quad \text{and} \quad X_t = \alpha_0 + \alpha_1 Y_t + v_t$$

Find the relation between $\hat{\beta}_1$ and $\hat{\alpha}_1$.

Solution:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2} \\ \hat{\alpha}_1 &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2} = \frac{\sum x_i y_i}{\sum y_i^2} \\ \hat{\beta}_1 \sum x_i^2 &= \hat{\alpha}_1 \sum y_i^2 \\ \hat{\beta}_1 &= \hat{\alpha}_1 \frac{S_{yy}}{S_{xx}} \end{aligned}$$

Question 8

A researcher is using a sample of 274 male employees to investigate the relationship between hourly wage rates Y_i (measured in dollars per hour) and firm tenure X_i (measured in years). Here are the sample information when $N = 274$.

$$\begin{aligned} \sum_{i=1}^N Y_i &= 1945.26 & \sum_{i=1}^N X_i^2 &= 30608.00 & \sum_{i=1}^N y_i^2 &= 4726.377 \\ \sum_{i=1}^N X_i &= 1774.00 & \sum_{i=1}^N X_i Y_i &= 16040.72 & \sum_{i=1}^N x_i^2 &= 19122.32 \\ \sum_{i=1}^N Y_i^2 &= 18536.73 & \sum_{i=1}^N x_i y_i &= 3446.226 & \sum_{i=1}^N \hat{u}_i^2 &= 4105.297 \end{aligned}$$

- Compute OLS estimates of the intercept coefficient β_1 and the slope coefficient β_2 .
- Interpret the slope coefficient estimate you find in part a.
- Calculate an estimate of σ^2 , the error variance.
- Calculate an estimate of $Var(\hat{\beta}_2)$, variance of $\hat{\beta}_2$; $Var(\hat{\beta}_1)$ and $Cov(\hat{\beta}_1, \hat{\beta}_2)$.
- Compute the value of R^2 , the coefficient of determination for estimated OLS sample regression equation. Explain what R^2 means.
- Calculate a sample value of t- statistic for testing the null hypothesis $H_0 : \beta_2 = 0$ against the alternative hypothesis $H_1 : \beta_2 \neq 0$.

Solution:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{S_{yy}}{S_{xx}}$$

$$\hat{\beta}_2 = \frac{3446.226}{19122.32} = 0.18022$$

a.

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

$$\hat{\beta}_1 = \frac{1945.26}{274} - 0.18022 \frac{1774.00}{274} = 7.0995 - 0.18022 \times 6.475 = 5.9326$$

$$\hat{Y} = 5.9326 + 0.18022X$$

- If the period of the work (tenure) increased by one hour; hourly wage will increase by \$0.18.

$$s^2 = \frac{SSE}{n-2}$$

c.

$$s^2 = \frac{4105.297}{272} = 15.093$$

$$Var(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{S_{xx}}$$

$$Var(\hat{\beta}_2) = \frac{15.093}{19122.32} = 0.00078929$$

d. $Var(\hat{\beta}_1) = \frac{\hat{\sigma}^2 \sum X_i^2}{n \sum x_i^2}$

$$Var(\hat{\beta}_1) = \frac{15.093 \times 30608}{274 \times 19122.32} = 0.088169$$

$$\begin{aligned} Cov(\hat{\beta}_1, \hat{\beta}_2) &= -\bar{X} Var(\hat{\beta}_2) \\ &= -\frac{1774}{274} \times 0.00078929 = -0.00511 \end{aligned}$$

e. $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$
 $R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$
 $R^2 = 1 - \frac{4105.297}{4726.377} = 0.1314$

The model explains only %13.14 of the total change in Y.

f. $H_0 : \beta_2 = 0$

$$H_1 : \beta_2 \neq 0$$

$$t = \frac{\hat{\beta}_2 - \beta_2}{sh(\hat{\beta}_2)}$$

$$t = \frac{0.18022 - 0}{\sqrt{0.00078929}} = 6.414$$

$$t_{\alpha/2} < 2.576$$

So reject $H_0 : \beta_2 = 0$; we could not say there is not linear relationship between X_i and Y_i .