International School of Economics at TSU

Econometrics 2

Lasha Chochua

Problem Set 5

Instructions: You are encouraged to solve the problems before the recitation. Additionally, you are encouraged to work in groups. It is **not mandatory** to submit solutions unless stated otherwise. However, if you would like to share your solution, I would be happy to review it.

Problem 1: Let X denote a 5×2 matrix and y a 5×1 vector:

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 3 \\ 1 & 5 \\ 1 & 2 \end{bmatrix}, \quad y = \begin{bmatrix} 14 \\ 17 \\ 8 \\ 16 \\ 3 \end{bmatrix}$$

a. Compute $Q = X'X$, $\det(Q)$, and Q^{-1}

b. Compute $A = (X'X)^{-1}X'$

Solution

a.

Notice that:

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 4 & 3 & 5 & 2 \end{bmatrix}$$

Then,

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 4 & 3 & 5 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 3 \\ 1 & 5 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 16 \\ 16 & 56 \end{bmatrix}$$

Determinant:

$$\det(Q) = (5)(56) - (16)^2 = 280 - 256 = 24$$

Inverse:

$$Q^{-1} = \frac{1}{\det(Q)} \begin{bmatrix} 56 & -16 \\ -16 & 5 \end{bmatrix} = \frac{1}{24} \begin{bmatrix} 56 & -16 \\ -16 & 5 \end{bmatrix} = \begin{bmatrix} \frac{7}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{5}{24} \end{bmatrix}$$

b.

We already found Q^{-1}:

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 4 & 3 & 5 & 2 \end{bmatrix}$$

Then:

$$A = Q^{-1}X' = \begin{bmatrix} \frac{7}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{5}{24} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 4 & 3 & 5 & 2 \end{bmatrix}$$

Compute:

$$A = \begin{bmatrix} \frac{7}{3} - \frac{4}{3} & \frac{7}{3} - \frac{8}{3} & \frac{7}{3} - \frac{6}{3} & \frac{7}{3} - \frac{10}{3} & \frac{7}{3} - \frac{4}{3} \\ -\frac{2}{3} + \frac{10}{24} & -\frac{2}{3} + \frac{20}{24} & -\frac{2}{3} + \frac{15}{24} & -\frac{2}{3} + \frac{25}{24} & -\frac{2}{3} + \frac{10}{24} \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{3} & \frac{1}{3} & -1 & 1 \\ -\frac{1}{20} & \frac{1}{10} & \frac{1}{40} & \frac{7}{120} & -\frac{1}{20} \end{bmatrix}$$

Problem 2 Consider the regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$$

Let $\hat{\beta}_0$ and $\hat{\beta}_1$ denote OLS estimators. Let ω_x and ω_y denote scaling factors for X and Y, respectively. Define:

- $Y^* = \omega_y Y$
- $X^* = \omega_x X$

What is the relationship between:

a. $\widehat{\beta}_1$ and $\widehat{\beta}_1^*$

b. $\widehat{\beta}_0$ and $\widehat{\beta}_0^*$

c. $\text{var}(\widehat{\beta}_0)$ and $\text{var}(\widehat{\beta}_0^*)$

d. $\text{var}(\widehat{\beta}_1)$ and $\text{var}(\widehat{\beta}_1^*)$

e. $\widehat{\sigma}^2$ and $(\widehat{\sigma}^*)^2$

f. R^2_{xy} and $R^2_{x^*y^*}$

Solution

a. Relationship between $\widehat{\beta}_1$ and $\widehat{\beta}_1^*$

Recall:

$$\widehat{\beta}_1 = \frac{s_{XY}}{s_X^2}$$

Now:

- $Y^* = \omega_y Y \Rightarrow s_{X^*Y^*} = \omega_x \omega_y s_{XY}$
- $X^* = \omega_x X \Rightarrow s^2_{X^*} = \omega_x^2 s_X^2$

Then:

$$\widehat{\beta}_1^* = \frac{s_{X^*Y^*}}{s^2_{X^*}} = \frac{\omega_x \omega_y s_{XY}}{\omega_x^2 s_X^2} = \frac{\omega_y}{\omega_x}\widehat{\beta}_1$$

b. Relationship between $\widehat{\beta}_0$ and $\widehat{\beta}_0^*$

Recall:

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X}$$

Then:

$$\widehat{\beta}_0^* = \bar{Y}^* - \widehat{\beta}_1^* \bar{X}^* = \omega_y \bar{Y} - \left(\frac{\omega_y}{\omega_x}\widehat{\beta}_1\right)(\omega_x \bar{X}) = \omega_y \bar{Y} - \omega_y \widehat{\beta}_1 \bar{X} = \omega_y(\bar{Y} - \widehat{\beta}_1 \bar{X}) = \omega_y \widehat{\beta}_0$$

c. In the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i \mid X_i] = 0, \quad \text{Var}(\varepsilon_i \mid X_i) = \sigma^2,$$

the OLS intercept estimator is:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

Taking variance:

$$\mathrm{Var}(\hat{\beta}_0) = \mathrm{Var}(\bar{Y} - \hat{\beta}_1 \bar{X}) = \mathrm{Var}(\bar{Y}) + \bar{X}^2\,\mathrm{Var}(\hat{\beta}_1),$$

since $\mathrm{Cov}(\bar{Y}, \hat{\beta}_1) = 0$ under OLS assumptions.

Now derive $\mathrm{Var}(\hat{\beta}_1)$:

We have

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X})\varepsilon_i}{\sum (X_i - \bar{X})^2},$$

since $Y_i - \bar{Y} = \beta_1 (X_i - \bar{X}) + \varepsilon_i$ and $\sum (X_i - \bar{X})^2$ is fixed given X.

Under $\mathbb{E}[\varepsilon_i \mid X_i] = 0$ and homoskedasticity, we compute the conditional variance:

$$\mathrm{Var}(\hat{\beta}_1 \mid X) = \frac{1}{\left[\sum (X_i - \bar{X})^2\right]^2} \cdot \sum (X_i - \bar{X})^2 \cdot \sigma^2 = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}.$$

Therefore,

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}.$$

Substitute:

$$\mathrm{Var}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right).$$

Relationship between $\mathrm{var}(\hat{\beta}_0)$ and $\mathrm{var}(\hat{\beta}_0^*)$

We use:

$$\mathrm{var}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right)$$

After scaling:

- $\bar{Y}^* = \omega_y \bar{Y}$
- $\bar{X}^* = \omega_x \bar{X}$

- $\sigma^{*2} = \omega_y^2 \sigma^2$

Thus:

$$\text{var}(\hat{\beta}_0^*) = \omega_y^2 \, \text{var}(\hat{\beta}_0)$$

d. Relationship between $\text{var}(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_1^*)$

We use:

$$\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$$

Under scaling: - Numerator scales as ω_y^2 - Denominator scales as ω_x^2

Hence:

$$\text{var}(\hat{\beta}_1^*) = \left(\frac{\omega_y}{\omega_x} \right)^2 \text{var}(\hat{\beta}_1)$$

e. Relationship between $\hat{\sigma}^2$ and $(\hat{\sigma}^*)^2$

Residuals scale as:

$$e_i^* = \omega_y e_i$$

So:

$$(\hat{\sigma}^*)^2 = \frac{1}{n-2} \sum (e_i^*)^2 = \omega_y^2 \hat{\sigma}^2$$

f. Relationship between R^2_{xy} and $R^2_{x^*y^*}$

Let $X^* = \omega_x X$ and $Y^* = \omega_y Y$ be scaled versions of the original variables. The coefficient of determination is defined as:

$$R^2 = (\text{Corr}(X,Y))^2 = \left(\frac{\text{Cov}(X,Y)}{s_X s_Y} \right)^2$$

Now consider the effect of scaling:

- $\text{Cov}(X^*, Y^*) = \text{Cov}(\omega_x X, \omega_y Y) = \omega_x \omega_y \text{Cov}(X,Y)$
- $s_{X^*} = |\omega_x| s_X, \quad s_{Y^*} = |\omega_y| s_Y$

Therefore:

$$\text{Corr}(X^*, Y^*) = \frac{\text{Cov}(X^*, Y^*)}{s_{X^*} s_{Y^*}} = \frac{\omega_x \omega_y \text{Cov}(X,Y)}{|\omega_x| s_X \cdot |\omega_y| s_Y} = \frac{\text{Cov}(X,Y)}{s_X s_Y} = \text{Corr}(X,Y)$$

Since correlation is unchanged by linear scaling of either variable, we have:

$$R^2_{x^*y^*} = \left(\text{Corr}(X^*, Y^*)\right)^2 = \left(\text{Corr}(X, Y)\right)^2 = R^2_{xy}$$

Problem 3 Given Regression Output

$$\begin{aligned}
\widehat{Price} = {} & \underset{(23.9)}{119.2} + \underset{(2.61)}{0.485\,BDR} + \underset{(8.94)}{23.4\,Bath} + \underset{(0.011)}{0.156\,Hsize} \\
& + \underset{(0.00048)}{0.002\,Lsize} + \underset{(0.311)}{0.090\,Age} - \underset{(10.5)}{48.8\,Poor}
\end{aligned}$$

$$\bar{R}^2 = 0.72, \quad SER = 41.5$$

a. Is the coefficient on BDR statistically significantly different from 0?

b. Typically, five-bedroom houses sell for much more than two-bedroom houses. Is this consistent with your answer to (a) and with the regression more generally?

c. A homeowner purchases 2000 square feet from an adjacent lot. Construct a 99% confidence interval for the change in the value of her house.

d. Lot size is measured in square feet. Do you think that another scale might be more appropriate? Why or why not?

e. The F-statistic for omitting BDR and Age from the regression is $F = 0.08$. Are the coefficients on BDR and Age statistically different from 0 at the 10% level?

Solution

a. The t-statistic is $\frac{0.485}{2.61} = 0.186 < 1.96$. Therefore, the coefficient on BDR is **not** statistically significantly different from zero.

b. The coefficient on BDR measures the *partial effect* of the number of bedrooms holding house size ($Hsize$) constant. Yet, the typical 5-bedroom house is much larger than the typical 2-bedroom house. Thus, the results in (a) say little about the conventional wisdom.

c. The 99% confidence interval for the effect of lot size on price is:

$$2000 \times [0.002 \pm 2.58 \times 0.00048] = 2000 \times [0.0007616,\ 0.0032384] = [1.52,\ 6.48]$$

This is measured in **thousands of dollars**.

d. Choosing the scale of the variables should be done to make the regression results easy to read and interpret. If lot size were measured in **thousands of square feet**, the coefficient would be 2 instead of 0.002.

This makes the result easier to interpret: on average, a one-thousand increase in lot size is associated with a **two-thousand dollar** increase in the price of a house.

e. The 10% critical value from the $F_{2,\infty}$ distribution is 2.30. Because $0.08 < 2.30$, the coefficients on BDR and Age are **not jointly significant** at the 10% level.

Problem 4 You are given the following data for $n = 5$ observations with Y_i representing log(wage) and $Educ_i$ representing years of education:

i	Y_i	$Educ_i$
1	2.5	12
2	3.0	16
3	2.2	10
4	2.8	14
5	2.6	13

a. Construct the $\mathbf{X}$ matrix including constants and the $\mathbf{Y}$ vector. Calculate $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{Y}$.

b. Compute the OLS estimator. Interpret the coefficient on education.

c. Calculate the fitted values and the residuals for each observation.

d. Verify that the orthogonality condition holds by directly computing $\mathbf{X}'\hat{\mathbf{e}}$. What does this result tell us about the relationship between the residuals and the regressors?

Solutions

a. Construct $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{Y}$

Matrices:

$$\mathbf{X} = \begin{bmatrix} 12 & 1 \\ 16 & 1 \\ 10 & 1 \\ 14 & 1 \\ 13 & 1 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} 2.5 \\ 3.0 \\ 2.2 \\ 2.8 \\ 2.6 \end{bmatrix}$$

Calculate $\mathbf{X'X}$:

$$\mathbf{X'X} = \begin{bmatrix} \sum X_i^2 & \sum X_i \\ \sum X_i & n \end{bmatrix} = \begin{bmatrix} 865 & 65 \\ 65 & 5 \end{bmatrix}$$

Where: $\sum X_i^2 = 144 + 256 + 100 + 196 + 169 = 865$ and $\sum X_i = 65$

Calculate $\mathbf{X'Y}$:

$$\mathbf{X'Y} = \begin{bmatrix} \sum X_i Y_i \\ \sum Y_i \end{bmatrix} = \begin{bmatrix} 173.0 \\ 13.1 \end{bmatrix}$$

Where: $\sum X_i Y_i = 30 + 48 + 22 + 39.2 + 33.8 = 173$

b. Compute $\hat{\beta}$

Inverse:

$$\det(\mathbf{X'X}) = 865(5) - 65^2 = 100$$

$$(\mathbf{X'X})^{-1} = \frac{1}{100} \begin{bmatrix} 5 & -65 \\ -65 & 865 \end{bmatrix} = \begin{bmatrix} 0.05 & -0.65 \\ -0.65 & 8.65 \end{bmatrix}$$

OLS Estimator:

$$\hat{\beta} = (\mathbf{X'X})^{-1}(\mathbf{X'Y}) = \begin{bmatrix} 0.05 & -0.65 \\ -0.65 & 8.65 \end{bmatrix} \begin{bmatrix} 173 \\ 13.1 \end{bmatrix} = \begin{bmatrix} 0.135 \\ 0.865 \end{bmatrix}$$

Regression: $\log(\text{wage}) = 0.135 \times \text{education} + 0.865$

Interpretation: Each additional year of education is associated with a 13.5% increase in wages (exact: 14.45%).

i	Y_i	$\hat{Y}_i$	$\hat{e}_i$

c. Fitted Values and Residuals

Formula: $\hat{Y}_i = 0.135 \times \text{Education}_i + 0.865$, $\hat{e}_i = Y_i - \hat{Y}_i$

i	Y_i	$\hat{Y}_i$	$\hat{e}_i$
1	2.5	2.485	0.015
2	3.0	3.025	-0.025
3	2.2	2.215	-0.015
4	2.8	2.755	0.045
5	2.6	2.620	-0.020

Vectors:

$$\hat{\mathbf{Y}} = \begin{bmatrix} 2.485 \\ 3.025 \\ 2.215 \\ 2.755 \\ 2.620 \end{bmatrix}, \quad \hat{\mathbf{e}} = \begin{bmatrix} 0.015 \\ -0.025 \\ -0.015 \\ 0.045 \\ -0.020 \end{bmatrix}$$

d. Orthogonality Condition

Verify: $\mathbf{X}'\hat{\mathbf{e}} = 0$

Education orthogonality:

$$\sum \text{Education}_i \cdot \hat{e}_i = 12(0.015) + 16(-0.025) + 10(-0.015) + 14(0.045) + 13(-0.020)$$
$$= 0.18 - 0.40 - 0.15 + 0.63 - 0.26 = 0 \quad \checkmark$$

Constant orthogonality:

$$\sum \hat{e}_i = 0.015 - 0.025 - 0.015 + 0.045 - 0.020 = 0 \quad \checkmark$$

Result: $\mathbf{X}'\hat{\mathbf{e}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Interpretation: Residuals are uncorrelated with all regressors. OLS has found the unique minimum of SSE.