Seminar 2 Solutions

Giulio Rossetti*

giuliorossetti94.github.io

January 24, 2025

* email: giulio.rossetti.1@wbs.ac.uk

Roadmap

Exercise 1

Exercise 2

true

Unbiasedness

$$Y = X\beta + u$$

$$RV \rightarrow \hat{\beta} = (X'X)^{-1}X'y$$

Definition

An estimator of a given paramater is said to be unbiased if its expected value is equal to the true value of the parameter:

$$\mathbb{E}[\widetilde{\hat{\theta}(\xi)}] = \overset{\beta}{\overset{\shortparallel}{\theta_0}}$$

sample

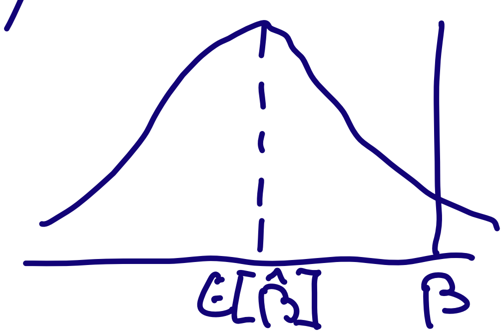
true estimator

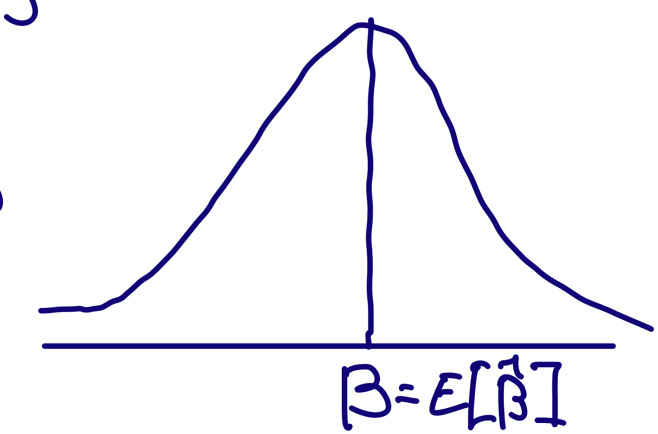
$$\mu \rightarrow \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\beta \rightarrow \hat{\beta} = (X'X)^{-1}X'y$$

$$E[\hat{\beta}] = \beta$$

$$E[\hat{\beta}] \neq \beta$$





OLS unbiasedness (Exercise 1)

- OLS Assumptions:
  1. Linear in parameters
  2. Random sampling
  3. No perfect collinearity
  4. Zero conditional mean: $E[u_i|x_i] = E[u_i] = 0.$
- Under assumptions 1-4 the OLS estimator is unbiased

$y = X\beta + u$

$y = \beta_0 + \beta_1 x_1 + u$

$y = \beta_0 + \beta_1^2 x_1 + u$

$y = \beta_0 + \beta_1 x_1^2 + u$

$X \to \text{rank}(X'X) = K$

$\hat{\beta} = \underbrace{(X'X)}^{-1} X'Y$

$$E[\hat{\beta}] = \beta$$

OLS unbiasedness: proof

$$E[\hat\beta] = \beta \qquad y = X\beta + u$$

$$\hat\beta = (X'X)^{-1}X'y \overset{1}{=} (X'X)^{-1}X'(X\beta + u)$$

$$= \cancel{(X'X)^{-1}X'X}\beta + (X'X)^{-1}X'u \qquad y = X\underset{\downarrow\,4}{\beta} + u$$

$$= \underbrace{\beta + (X'X)^{-1}X'u}$$

$$E[\hat\beta \mid X] = E[\beta + (X'X)^{-1}X'u \mid X]$$

$$= E[\beta \mid X] + E[(X'X)^{-1}X'u \mid X]$$

$$= \beta + (X'X)^{-1}X'\underbrace{E[u \mid X]}_{0}$$

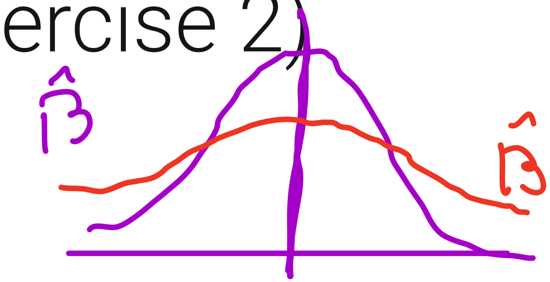
$$\overset{2}{=} \beta$$

$$E[X+Y] = E[X] + E[Y]$$

$$E[\hat\beta \mid X] = \beta$$

$$E[\hat\beta] = \beta$$

Variance of OLS estimator: derivation (Exercise 2)



$E[\hat{\beta}] = \beta$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

5. Homoskedasticity: $Var(\mathbf{u}|\mathbf{X}) = \sigma^2 \mathbf{I}_n \quad \sigma^2 > 0$

$$\hat{\beta} = \beta + (X'X)^{-1}X'u$$

$var(a + X) = var(X)$

$var(a + bX) = b^2 var(X)$

$$var(\hat{\beta}|X) = var(\beta + (X'X)^{-1}X'u|X)$$

$$= var((X'X)^{-1}X'u|X)$$

RV

$var(AX) = A\, var(X) A'$

$$= (X'X)^{-1}X'\, var(u|X)\, X(X'X)^{-1}$$

$$\overset{5}{=} (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1}$$

$$= \sigma^2 \cancel{(X'X)^{-1}X'X}(X'X)^{-1}$$

$$= \sigma^2 \underbrace{(X'X)^{-1}}_{K \times K}$$

$$se(\hat{\beta}_1) = \sqrt{\sigma^2 (X'X)^{-1}_{11}}$$

$$\sigma^2 \begin{bmatrix} var(\hat{\beta}_1) & \cdots & \\ cov(\hat{\beta}_1, \hat{\beta}_2) & var(\hat{\beta}_2) & \\ \vdots & & \ddots \end{bmatrix}$$

Variance of OLS estimates

- The variance of the OLS estimates is given by:

$$\text{Var}(\hat{\beta}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

- The standard errors are given by:

$$se\left(\hat{\beta}_j\right) = \sqrt{\text{Var}\left(\hat{\beta}_j \mid \mathbf{X}\right)} = \sqrt{\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}_{jj}} = \sigma\sqrt{\left(\mathbf{X}'\mathbf{X}\right)^{-1}_{jj}}.$$

- σ^2 is not observed. Obtain an unbiased estimate through the OLS residuals $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\beta}$

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-k-1}$$

Variance of OLS estimates

- then
$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k-1},$$
- therefore,
$$se\left(\hat{\beta}_j\right) = \hat{\sigma}\sqrt{(\mathbf{X}'\mathbf{X})^{-1}_{jj}}.$$
- Increasing the sample size n reduces $\hat{\sigma}^2$ and hence the standard errors.

SE of OLS estimates: SLR and matrices (Exercise 5)

$y_i = \beta_0 + \beta_1 \underline{x_i} + u_i$

- In the SLR case, X is a $n \times 2$ matrix:

$var(\hat{\beta} | X) = \sigma^2 (X'X)^{-1}$

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$X' = \begin{bmatrix} 1 & 1 & 1 & \cdots \\ x_1 & \cdots & & x_n \end{bmatrix}$

$A = \begin{bmatrix} a & c \\ d & b \end{bmatrix}$

$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} b & -c \\ -d & a \end{bmatrix}$

- then $X'X$ is a 2×2 matrix:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

- and its inverse is (see matrix algebra slides):

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n \sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}$$

SE of OLS estimates: SLR and matrices (Exercise 5)

- to compute the standard errors we use the formula:

$$se\left(\hat{\beta}_j\right) = \hat{\sigma}\sqrt{(\mathbf{X}'\mathbf{X})^{-1}_{jj}}$$

- therefore, the standard errors for $\hat{\beta}_0$ and $\hat{\beta}_1$ are:

position 1
X

$$se\left(\hat{\beta}_0\right) = \hat{\sigma}\sqrt{\frac{\sum x_i^2}{n\sum x_i^2 - \left(\sum x_i\right)^2}}$$

$$se\left(\hat{\beta}_1\right) = \hat{\sigma}\sqrt{\frac{n}{n\sum x_i^2 - \left(\sum x_i\right)^2}}$$

Exercise 1

Standard Errors of OLS Estimates

Model:

$$Wage = \beta_0 + \beta_1 Educ + u$$

- $\hat{\beta}_1 = 1.04, \quad \hat{\beta}_0 = -6.90.$
- $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2 = 10.31.$

$$\text{Var}(\hat{\beta}_0) = 19.54 \implies \text{se}(\hat{\beta}_0) = 4.42.$$

$$\text{Var}(\hat{\beta}_1) = 0.108 \implies \text{se}(\hat{\beta}_1) = 0.33.$$

Roadmap

Exercise 1

Exercise 2

Exercise 2.2-2.3

Collinearity

Model:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + u_i \quad i = 1, \ldots, n$$

- Collinearity: If $x_3 = x_1 + x_2 + 6$, perfect collinearity exists, making OLS infeasible.
- assumption 3 is violated.
- rank of $\mathbf{X}'\mathbf{X}$ is 2, not 3. Cannot invert the matrix.

Exercise 2.2-2.3

~~Irrelevant~~ Variables

omit.

EST.

~~True~~ model:

$\beta_3 x_{3,i} + u_i$

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i \quad i = 1, \ldots, n$$

TRUE

~~Estimated~~ model:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + u_i \quad i = 1, \ldots, n$$

- Irrelevant Variables: Including irrelevant variables (e.g., x_3) does not affect the unbiasedness but reduces efficiency.

Roadmap

Appendix

Appendix: Proof of OLS Unbiasedness

1.1. EX1

① Linear in parameters $y = \beta_0 + \beta_1 x + u$

② Random sampling

③ No multicollinearity

④ Errors have zero conditional means $E[u_i|x_i] = 0 \Rightarrow E[u_i] = 0$

under 1-4 OLS estimates are unbiased: $E[\hat{\beta}_0] = \beta_0$, $E[\hat{\beta}_1] = \beta_1$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{n}(x_i y_i - \bar{y}x_i - \bar{x}y_i + \bar{x}\bar{y})$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i y_i) - \bar{y}\sum_{i=1}^{n}\frac{x_i}{n} - \bar{x}\sum_{i=1}^{n}\frac{y_i}{n} + \frac{n}{n}\bar{x}\bar{y} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n}\frac{x_i}{n}\sum_{i=1}^{n}\frac{y_i}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \frac{1}{n^2}\sum_{i=1}^{n} x_i y_i$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \bar{x}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})y_i$$

$$\bar{x} = \sum_{i=1}^{n}\frac{x_i}{n}$$

$$* \sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x} = \bar{x}n - \bar{x}n = 0$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

using ① $y_i = \beta_0 + \beta_1 x_i + u_i$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\beta_0\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})}_{*\,0} + \beta_1\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})x_i}_{\sum_{i=1}^{n}(x_i - \bar{x})^2} + \sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \frac{\beta_0 \times 0}{\sum_{i=1}^{n}(x_i - \bar{x})^2} + \beta_1 \times \frac{\cancel{\sum_{i=1}^{n}(x_i - \bar{x})^2}}{\cancel{\sum_{i=1}^{n}(x_i - \bar{x})^2}} + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$E[\hat{\beta}_1] = E\left[\beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum^{n}(x_i - \bar{x})^2}\right] = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{\sum^{n}(x_i - \bar{x})^2}\underbrace{E[u_i]}_{0 \text{ Using } 4} \Rightarrow E[\hat{\beta}_1] = \beta_1$$

Appendix: Variance of OLS Estimator

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{using } ① \quad y_i = \beta_0 + \beta_1 x_i + u_i \quad \bar{y} = E[y_i] = \beta_0 + E[\beta_1 x_i] + E[\overset{=0}{u_i}] \quad E[\cancel{\beta_1}] =$$

$$= \beta_0 + \beta_1 \bar{x} - \hat{\beta}_1 \bar{x} = \beta_0 + \bar{x}(\beta_1 - \hat{\beta}_1) \qquad = \beta_0 + E[\beta_1]\bar{x} \qquad = \beta_0 + \beta_1 \bar{x}$$

$$E[\beta_0] = E[\beta_0 + \bar{x}(\beta_1 - \hat{\beta}_1)] = E[\beta_0] + \bar{x}\underbrace{E[\beta_1 - \hat{\beta}_1]}_{0} = \beta_0$$

EX2

⑤ Homoskedasticity: $Var(u|X) = \sigma^2 I_n \quad \sigma^2 > 0$

$$\underset{n\times 1}{y} = \underset{n\times 4}{X}\,\underset{4\times 1}{\beta} + \underset{n\times 1}{u}$$

OLS $\hat{\beta} = (X'X)^{-1}X'y$ ①

$$= (X'X)^{-1}X'(X\beta + u) = \cancel{(X'X)^{-1}X'X}\beta + (X'X)^{-1}X'u$$

unbiasedness

$$E[\hat{\beta}] = E[\cancel{(X'X)^{-1}X'X}\beta + (X'X)^{-1}X'u] = E[\beta] + E[(X'X)^{-1}X'u]$$

$$= \beta + (X'X)^{-1}X'\underbrace{E[u]}_{0}$$

$E[\hat{\beta}] = \beta$

$$= \beta$$

variance of $\hat{\beta}$

$$var(\hat{\beta}|X) = var[(X'X)^{-1}X'y|X] = var[(\beta + (X'X)^{-1}X'u)|X] = (X'X)^{-1}X'\overbrace{var(u|X)}^{⑤\ \sigma^2 I_n}X(X'X)^{-1}$$

$$= \sigma^2 \cancel{(X'X)^{-1}X'X}(X'X)^{-1} = \sigma^2(X'X)^{-1}$$