Seminar 2: OLS Properties

Unbiasedness, Variance, and Standard Errors

Giulio Rossetti*

giuliorossetti94.github.io

January 30, 2026

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

Roadmap

Part 1: OLS Unbiasedness

- Exercise 1: Unbiasedness
- Exercise 2: Variance of OLS Estimator

Part 2: Collinearity

- Exercise 2.2-2.3: Collinearity and Irrelevant Variables

Roadmap

Part 1: OLS Unbiasedness

- Exercise 1: Unbiasedness
- Exercise 2: Variance of OLS Estimator

Part 2: Collinearity

- Exercise 2.2-2.3: Collinearity and Irrelevant Variables

Unbiasedness

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:

$$\mathbb{E}[\hat{\theta}(\xi)] = \theta_0$$

$\hat{\beta} \equiv \hat{\theta}$

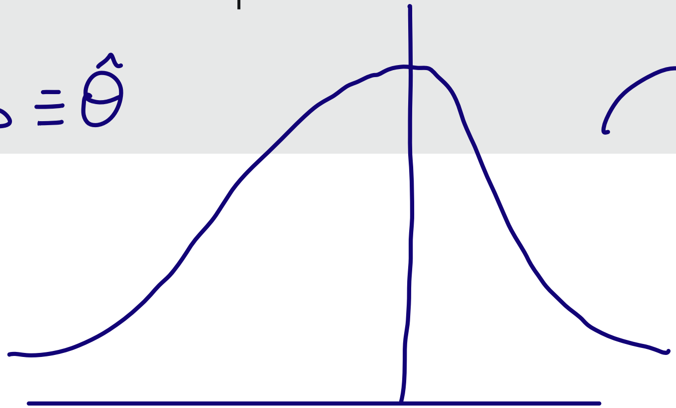
$Y = X\beta + u$ true pop model

$\Downarrow$

data $\Rightarrow$ y x

$\hat{\beta} = \underbrace{(X'X)^{-1}X'y}_{\text{estimate for } \beta}$

$\hat{\beta} \rightarrow$ RV



$\theta_0 \equiv \beta$

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

OLS unbiasedness (Exercise 1)

- OLS Assumptions:
  1. Linear in parameters $\Rightarrow y = x\beta + u$ $\quad y = x^2\beta + u$ $\quad \cancel{y = x\beta^2 + u}$
  2. Random sampling $\Rightarrow$
  3. No perfect collinearity $\Rightarrow r(X'X) = K$ $\quad \hat{\beta} = (X'X)^{-1}Xy$
  4. Zero conditional mean: $E[u_i|x_i] = E[u_i] = 0.$
- Under assumptions 1-4 the OLS estimator is unbiased

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

OLS unbiasedness: proof

$$E[\hat{\beta}] = \beta \qquad (1)\ y = X\beta + u$$

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

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

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

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

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

$$= \beta + \cancel{(X'X)^{-1}X'\underbrace{E[u|X]}_{0}} \quad (4)$$

$$= \beta$$

$$E[\hat{\beta}|X] \Rightarrow E[\hat{\beta}] = \beta$$

Variance of OLS estimator: derivation (Exercise 2)

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

$$
\begin{aligned}
var(\hat{\beta}|x) &= var(\beta + (x'x)^{-1}x'u \mid x) \\
&= \underbrace{var(\beta|x)}_{0} + var[(x'x)^{-1}x'u \mid x] \\
&= var(x'x^{-1}x'u \mid x) \\
&= (x'x)^{-1}x' \, var(u \overset{5}{|} x) \, x(x'x)^{-1} \\
&= (x'x)^{-1}x' \, \sigma^2 I_n \, x(x'x)^{-1} \\
&= \sigma^2 (\cancel{x'x})^{-1} \cancel{x'x} (x'x)^{-1} \\
&= \sigma^2 (x'x)^{-1}
\end{aligned}
$$

$var(ax) = a^2 var(x)$

$\rightarrow var(Ax) = A\,var(x)A'$

$\hat{\beta} \sim (\beta, \sigma^2(x'x)^{-1})$

Variance of OLS estimates

$y = x\beta + u$

variance of u → not observed

- The variance of the OLS estimates is given by:

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

$\text{var}\,\hat{\beta}_0$ $\text{var}\,\hat{\beta}_1$

- 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 $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\beta}$

$y, X \rightarrow \hat{\beta} \rightarrow \hat{u} = y - X\hat{\beta}$

$se(\hat{\beta}_0) = \hat{\sigma}\sqrt{}$

$se(\hat{\beta}_1) = \hat{\sigma}\sqrt{}$

Variance of OLS estimates

variance of residuals $\Rightarrow$ estimator for the variance of error term

- then

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

$$\frac{\sum_{i=1}^{n} u_i^2}{n-k-1} \leftarrow \text{variance of residuals}$$

- therefore,

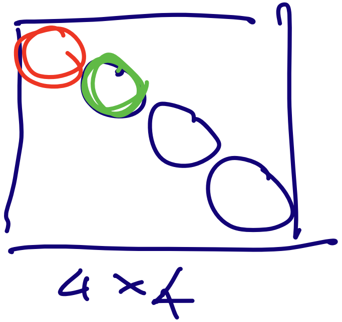
$$se\left(\hat{\beta}_j\right) = \hat{\sigma}\sqrt{\left(\mathbf{X}'\mathbf{X}\right)^{-1}_{jj}}.$$

$$var(x) = \frac{1}{N}\sum (x_i - \bar{x})^2$$

$$\bar{\hat{u}} = 0$$

- Increasing the sample size n reduces $\hat{\sigma}^2$ and hence the standard errors.

$$\underset{4\times 4}{(X'X)^{-1}} \rightarrow$$



$$var(\hat{u}) = \frac{1}{N}\sum(\hat{u}_i - \bar{\hat{u}})^2$$

$$= \frac{1}{N}\underbrace{\sum \hat{u}_i^2}_{\hat{u}'\hat{u}}$$

$$t\,stat(\beta_0) = \frac{\hat{\beta}_0 - \beta}{se(\hat{\beta}_0)}$$

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

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

$$\begin{array}{cc} & \begin{array}{cc} \text{intercept} & \text{indep var} \end{array} \\ \mathbf{X} = & \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \\ N \times 2 & \end{array}$$

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

- then $X'X$ is a 2×2 matrix:

$$\underset{2 \times N \times N \times 2}{X'X}$$

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

$$se(\hat{\beta}_0)$$

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

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

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:

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

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

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

Part 1: OLS Unbiasedness

- Exercise 1: Unbiasedness
- Exercise 2: Variance of OLS Estimator

Part 2: Collinearity

- Exercise 2.2-2.3: Collinearity and Irrelevant Variables

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

True model:

W E Ex

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

Estimated model:

W E Ex GPA

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

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

Roadmap

True model

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2i} + \beta_3 x_{3i} + v_i$$

Est model

Appendix

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

$$E[u_i | x_i] = 0 \Rightarrow cov(u_i, x_i) = 0 \quad \beta_3 x_{3i} + v_i$$

if x_{3i} is correlated with either x_{1i} or x_{2i}

$$E[u_i | x_i] \neq 0 \qquad E[\beta_3 x_{3i} + v_i | x_i]$$

Appendix: Proof of OLS Unbiasedness

1.1. EX1

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

② Random sampling

③ No multicollinearity

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

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

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

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

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

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

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

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

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

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

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

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

Appendix: Variance of OLS Estimator

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

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

$$E[\hat{\beta}_0] = E[\beta_0 + \bar{x}(\beta_1 - \hat{\beta}_1)] = E[\beta_0] + \bar{x}\underbrace{E[\beta_1 - \hat{\beta}_1]}_{0} = \beta_0$$

EX2

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

$$\underset{n\times1}{y} = \underset{n\times4}{X}\underset{4\times1}{\beta} + \underset{n\times1}{u}$$

OLS $\hat{\beta} = (X'X)^{-1}X'y$ ①

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

unbiasedness

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

$$\overset{!}{=} \beta + (X'X)^{-1}X'\underset{\llcorner 0}{E[u]}$$

$$E[\hat{\beta}] = \beta \qquad \overset{!}{=} \beta$$

Variance of $\hat{\beta}$

$$\sigma^2 I_n \; ⑤$$