HE2003 Econometrics I

Tutorial 3 (Week 4) – Problems & Solutions

Academic Year 2026/2027, Semester 1

Quantitative Research Society @NTU

October 12, 2026

Overview

- **Lecture coverage:** Lecture 3 — the true-vs-estimated comparison (population objects vs. OLS objects), the algebraic properties of OLS residuals that follow from the first-order conditions, and the Least Squares Assumptions (with Assumption 1, zero conditional mean, as the conceptual centrepiece).
- **Questions:** 4 questions from Tutorial Question Set No. 3.
- **Notation:** as in lectures — u_i population error, $\hat{u}_i$ OLS residual; β_0, β_1 population (unknown) parameters, $\hat{\beta}_0, \hat{\beta}_1$ OLS estimators; "Least Squares Assumption 1/2/3" (never SLR.1–5).
- **Question themes:** Q1 population vs. sample objects; Q2 residual identities from the OLS first-order conditions; Q3 zero conditional mean in pictures and counterexamples; Q4 a binary regressor and group means.

For academic reference only. This document is provided for educational purposes and does not constitute an official question paper, answer key, or any formally authorised academic material. Its content is not guaranteed to be complete or accurate.

Question 1 (Population vs. OLS: five core objects)

Problem

Explain the following: population regression line, OLS regression line, error term, OLS residual, and OLS predicted value. *(Tutorial Question Set No. 3, Q1.)*

Solution

Method 1: Definitions with a numerical illustration

Let the population model be

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \qquad i = 1, \ldots, n,$$

where β_0, β_1 are fixed, unknown population parameters, and let $\hat{\beta}_0, \hat{\beta}_1$ be the OLS estimates, i.e. the values of b_0, b_1 that minimise

$$\sum_{i=1}^{n} \big(Y_i - b_0 - b_1 X_i\big)^2.$$

- **Population regression line.** The population regression line is $\beta_0 + \beta_1 X$. It describes the relationship between X and Y that holds on average in the whole population; under Least Squares Assumption 1 it equals the conditional mean $E[Y_i \,|\, X_i = X]$. Its intercept and slope are fixed numbers that no researcher observes.
- **Error term.** The error term is $u_i = Y_i - (\beta_0 + \beta_1 X_i)$, the vertical distance from observation i to the population line. It collects every factor other than X_i that affects Y_i. Since β_0, β_1 are unknown, u_i is never observed.
- **OLS regression line.** The OLS regression line is $\hat{\beta}_0 + \hat{\beta}_1 X$. It is computed from one sample, so a different sample gives a different line. It is the sample estimate of the population line.
- **OLS predicted value.** The OLS predicted value is $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$, the height of the OLS line at $X = X_i$. It is the sample counterpart of the systematic part $\beta_0 + \beta_1 X_i$ of Y_i.
- **OLS residual.** The OLS residual is $\hat{u}_i = Y_i - \hat{Y}_i$, the vertical distance from observation i to the OLS line. It is computable and is the sample counterpart of u_i, but $\hat{u}_i \neq u_i$ in general, because $\hat{\beta}_0 \neq \beta_0$ and $\hat{\beta}_1 \neq \beta_1$ in general.

For a numerical illustration, take the Lecture 3 regression of CEO salary (in \$1,000s) on return on equity (ROE, in percentage points), estimated on $n = 209$ CEOs:

$$\widehat{\text{salary}} = 963.191 + 18.501\,\text{roe}.$$

This is the OLS regression line; the population line $\beta_0 + \beta_1\,\text{roe}$ is unknown. The first CEO in the sample has $\text{roe}_1 = 14.1$ and $\text{salary}_1 = 1095$. The OLS predicted value is

$$\begin{aligned} \widehat{\text{salary}}_1 &= 963.191 + 18.501 \times 14.1 \\ &= 963.191 + 260.8641 \\ &= 1224.0551, \end{aligned}$$

i.e. about \$1,224,055. The OLS residual is

$$\begin{aligned}\hat{u}_1 &= \text{salary}_1 - \widehat{\text{salary}}_1 \\ &= 1095 - 1224.0551 \\ &= -129.0551,\end{aligned}$$

so this CEO earns about \$129,055 less than the OLS line predicts at an ROE of 14.1. The error term $u_1 = 1095 - (\beta_0 + \beta_1 \times 14.1)$ cannot be computed, because β_0 and β_1 are unknown.

Population: $\beta_0 + \beta_1 X_i, \quad u_i = Y_i - (\beta_0 + \beta_1 X_i)$ (unobserved)
Sample: $\hat{\beta}_0 + \hat{\beta}_1 X_i, \quad \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, \quad \hat{u}_i = Y_i - \hat{Y}_i$ (all computable)

Unofficial QRS@NTU solutions

Method 2: Population and sample decompositions of Y_i

Every observed Y_i can be split in two ways. Using the population line,

$$Y_i = \underbrace{(\beta_0 + \beta_1 X_i)}_{\text{systematic part (unobserved)}} + \underbrace{u_i}_{\text{error (unobserved)}}.$$

Using the OLS line,

$$Y_i = \underbrace{(\hat{\beta}_0 + \hat{\beta}_1 X_i)}_{\hat{Y}_i \text{ (computed)}} + \underbrace{\hat{u}_i}_{\text{residual (computed)}}.$$

The first split defines the population regression line $\beta_0 + \beta_1 X$ and the error term $u_i = Y_i - (\beta_0 + \beta_1 X_i)$. The second split defines the OLS regression line $\hat{\beta}_0 + \hat{\beta}_1 X$, the OLS predicted value $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ and the OLS residual $\hat{u}_i = Y_i - \hat{Y}_i$.

Both right-hand sides equal Y_i, so they equal each other:

$$\beta_0 + \beta_1 X_i + u_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i.$$

Solve for $\hat{u}_i$:

$$\begin{aligned} \hat{u}_i &= u_i + (\beta_0 + \beta_1 X_i) - (\hat{\beta}_0 + \hat{\beta}_1 X_i) \\ &= u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) X_i. \end{aligned}$$

Equivalently, the gap between the predicted value and the systematic part is

$$\hat{Y}_i - (\beta_0 + \beta_1 X_i) = (\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) X_i = u_i - \hat{u}_i.$$

Hence:

- $\hat{u}_i = u_i$ only at an X_i where the OLS line crosses the population line. This happens for every i only if $\hat{\beta}_0 = \beta_0$ and $\hat{\beta}_1 = \beta_1$, which a finite sample does not deliver in general.
- The difference $u_i - \hat{u}_i$ is the estimation error of the line at X_i, measured in the units of Y. It becomes small when $\hat{\beta}_0$ and $\hat{\beta}_1$ are close to β_0 and β_1.
- The residuals satisfy $\sum_{i=1}^{n} \hat{u}_i = 0$ in every sample (Question 2), whereas $\sum_{i=1}^{n} u_i$ is a sum of n random population deviations and is generally not zero.

Population: $\beta_0 + \beta_1 X_i, \quad u_i = Y_i - (\beta_0 + \beta_1 X_i)$ (unobserved)
Sample: $\hat{\beta}_0 + \hat{\beta}_1 X_i, \quad \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, \quad \hat{u}_i = Y_i - \hat{Y}_i$ (all computable)

Method 3: A constructed sample from a known population line

To see all five objects as numbers, suppose, for illustration only, that the population line is known to be

$$\beta_0 + \beta_1 X = 2 + 0.5X,$$

and that $n = 4$ draws have $X_i = 1, 2, 3, 4$ and errors $u_i = 1, -1, 0, 1$. Then $Y_i = 2 + 0.5X_i + u_i$:

$$Y_1 = 2 + 0.5(1) + 1 = 3.5, \qquad Y_2 = 2 + 0.5(2) - 1 = 2,$$
$$Y_3 = 2 + 0.5(3) + 0 = 3.5, \qquad Y_4 = 2 + 0.5(4) + 1 = 5.$$

A researcher sees only the pairs (X_i, Y_i) and computes OLS. The sample means are

$$\bar{X} = \frac{1+2+3+4}{4} = 2.5, \qquad \bar{Y} = \frac{3.5+2+3.5+5}{4} = \frac{14}{4} = 3.5.$$

The deviations are $X_i - \bar{X} = -1.5, -0.5, 0.5, 1.5$ and $Y_i - \bar{Y} = 0, -1.5, 0, 1.5$, so

$$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = (-1.5)(0) + (-0.5)(-1.5) + (0.5)(0) + (1.5)(1.5)$$
$$= 0 + 0.75 + 0 + 2.25 = 3,$$
$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = 2.25 + 0.25 + 0.25 + 2.25 = 5.$$

Hence

$$\hat{\beta}_1 = \frac{3}{5} = 0.6, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 3.5 - 0.6(2.5) = 3.5 - 1.5 = 2.$$

The OLS line is $\hat{\beta}_0 + \hat{\beta}_1 X = 2 + 0.6X$. The predicted values and residuals are

$$\hat{Y}_1 = 2 + 0.6(1) = 2.6, \qquad \hat{u}_1 = 3.5 - 2.6 = 0.9,$$
$$\hat{Y}_2 = 2 + 0.6(2) = 3.2, \qquad \hat{u}_2 = 2 - 3.2 = -1.2,$$
$$\hat{Y}_3 = 2 + 0.6(3) = 3.8, \qquad \hat{u}_3 = 3.5 - 3.8 = -0.3,$$
$$\hat{Y}_4 = 2 + 0.6(4) = 4.4, \qquad \hat{u}_4 = 5 - 4.4 = 0.6.$$

i	X_i	$\beta_0 + \beta_1 X_i$	u_i	Y_i	$\hat{Y}_i$	$\hat{u}_i$
1	1	2.5	1	3.5	2.6	0.9
2	2	3.0	−1	2.0	3.2	−1.2
3	3	3.5	0	3.5	3.8	−0.3
4	4	4.0	1	5.0	4.4	0.6
Sum			1			0

- The columns $\beta_0 + \beta_1 X_i$ and u_i are population objects; in practice they are unknown, and they are visible here only because the population line was fixed in advance.
- The columns $\hat{Y}_i$ and $\hat{u}_i$ are OLS objects computed from the four pairs (X_i, Y_i). The OLS line $2 + 0.6X$ differs from the population line $2 + 0.5X$.

- The residual differs from the error: $\hat{u}_i - u_i = 0.9 - 1,\ -1.2 + 1,\ -0.3 - 0,\ 0.6 - 1 = -0.1,\ -0.2,\ -0.3,\ -0.4$, which is $-(\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)X_i = -0 - 0.1X_i$.
- The residuals sum to $0.9 - 1.2 - 0.3 + 0.6 = 0$, while the errors sum to $1 - 1 + 0 + 1 = 1$.

Population: $\beta_0 + \beta_1 X_i, \quad u_i = Y_i - (\beta_0 + \beta_1 X_i)$ (unobserved)
Sample: $\hat{\beta}_0 + \hat{\beta}_1 X_i, \quad \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, \quad \hat{u}_i = Y_i - \hat{Y}_i$ (all computable)

Unofficial - QRS@NTU - qrs.ntu.org

Question 2 (Residual properties from the FOCs)

Problem

(a) Show that $\sum_{i=1}^{n} \hat{u}_i = 0$ by using the first-order condition (FOC) of the OLS estimators.

(b) Show that $\sum_{i=1}^{n} \hat{u}_i X_i = 0$ by using the FOC of the OLS estimators.

(c) Show that $\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = 0$. *[Hint:* $\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = \sum_{i=1}^{n} \hat{u}_i(\hat{\beta}_0 + \hat{\beta}_1 X_i)$*.]*

(Tutorial Question Set No. 3, Q2.)

Solution

Method 1: First-order conditions

Let

$$S(b_0, b_1) = \sum_{i=1}^{n} \big(Y_i - b_0 - b_1 X_i\big)^2.$$

The OLS estimators $(\hat{\beta}_0, \hat{\beta}_1)$ minimise S. Differentiate term by term with the Chain Rule; the inner derivative of $Y_i - b_0 - b_1 X_i$ is -1 with respect to b_0 and $-X_i$ with respect to b_1:

$$\frac{\partial S}{\partial b_0} = \sum_{i=1}^{n} 2\big(Y_i - b_0 - b_1 X_i\big)(-1) = -2\sum_{i=1}^{n} \big(Y_i - b_0 - b_1 X_i\big),$$

$$\frac{\partial S}{\partial b_1} = \sum_{i=1}^{n} 2\big(Y_i - b_0 - b_1 X_i\big)(-X_i) = -2\sum_{i=1}^{n} \big(Y_i - b_0 - b_1 X_i\big) X_i.$$

At the minimum $(b_0, b_1) = (\hat{\beta}_0, \hat{\beta}_1)$ both partial derivatives are zero:

$$-2\sum_{i=1}^{n} \big(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\big) = 0, \qquad \text{(FOC } b_0\text{)}$$

$$-2\sum_{i=1}^{n} \big(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\big) X_i = 0. \qquad \text{(FOC } b_1\text{)}$$

By definition of the OLS residual,

$$\hat{u}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i.$$

(a) Replace $Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$ by $\hat{u}_i$ in (FOC b_0):

$$-2\sum_{i=1}^{n} \hat{u}_i = 0.$$

Divide both sides by -2:

$$\boxed{\sum_{i=1}^{n} \hat{u}_i = 0.}$$

(b) Replace $Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$ by $\hat{u}_i$ in (FOC b_1):

$$-2\sum_{i=1}^{n} \hat{u}_i X_i = 0.$$

Divide both sides by -2:

$$\boxed{\sum_{i=1}^{n} \hat{u}_i X_i = 0.}$$

(c) Following the hint, write $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ and expand the sum:

$$\begin{aligned}
\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i &= \sum_{i=1}^{n} \hat{u}_i \big(\hat{\beta}_0 + \hat{\beta}_1 X_i\big) \\
&= \sum_{i=1}^{n} \big(\hat{\beta}_0 \hat{u}_i + \hat{\beta}_1 \hat{u}_i X_i\big) \\
&= \sum_{i=1}^{n} \hat{\beta}_0 \hat{u}_i + \sum_{i=1}^{n} \hat{\beta}_1 \hat{u}_i X_i \\
&= \hat{\beta}_0 \sum_{i=1}^{n} \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^{n} \hat{u}_i X_i.
\end{aligned}$$

The constants $\hat{\beta}_0, \hat{\beta}_1$ come out of the sums because they do not depend on i. Substitute the results of (a) and (b):

$$\begin{aligned}
\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i &= \hat{\beta}_0(0) + \hat{\beta}_1(0) \\
&= 0.
\end{aligned}$$

Hence

$$\boxed{\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = 0.}$$

Unofficial Tutorial Solutions

Method 2: Closed-form estimators

Recall the OLS estimators

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X},$$

which require $\sum_{i=1}^{n}(X_i - \bar{X})^2 > 0$, i.e. the X_i are not all equal. Substitute $\hat{\beta}_0$ into the residual:

$$\begin{aligned}
\hat{u}_i &= Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \\
&= Y_i - (\bar{Y} - \hat{\beta}_1 \bar{X}) - \hat{\beta}_1 X_i \\
&= (Y_i - \bar{Y}) - \hat{\beta}_1 (X_i - \bar{X}).
\end{aligned}$$

Deviations from a sample mean sum to zero, because $\sum_{i=1}^{n} Y_i = n\bar{Y}$ and $\sum_{i=1}^{n} X_i = n\bar{X}$:

$$\sum_{i=1}^{n}(Y_i - \bar{Y}) = \sum_{i=1}^{n} Y_i - n\bar{Y} = n\bar{Y} - n\bar{Y} = 0,$$
$$\sum_{i=1}^{n}(X_i - \bar{X}) = \sum_{i=1}^{n} X_i - n\bar{X} = n\bar{X} - n\bar{X} = 0.$$

(a) Sum the centred form of the residual and split the sum:

$$\begin{aligned}
\sum_{i=1}^{n} \hat{u}_i &= \sum_{i=1}^{n} \big[(Y_i - \bar{Y}) - \hat{\beta}_1 (X_i - \bar{X})\big] \\
&= \sum_{i=1}^{n}(Y_i - \bar{Y}) - \hat{\beta}_1 \sum_{i=1}^{n}(X_i - \bar{X}) \\
&= 0 - \hat{\beta}_1 (0) \\
&= 0.
\end{aligned}$$

Hence

$$\boxed{\sum_{i=1}^{n} \hat{u}_i = 0.}$$

(b) Write $X_i = (X_i - \bar{X}) + \bar{X}$ and split the sum:

$$\begin{aligned}
\sum_{i=1}^{n} \hat{u}_i X_i &= \sum_{i=1}^{n} \hat{u}_i (X_i - \bar{X}) + \bar{X} \sum_{i=1}^{n} \hat{u}_i \\
&= \sum_{i=1}^{n} \hat{u}_i (X_i - \bar{X}) + \bar{X}(0) \qquad \text{(by (a))} \\
&= \sum_{i=1}^{n} \hat{u}_i (X_i - \bar{X}).
\end{aligned}$$

Unofficial Notes, Not Guaranteed

Substitute the centred form of $\hat{u}_i$ and then the formula for $\hat{\beta}_1$:

$$\begin{aligned}
\sum_{i=1}^{n} \hat{u}_i(X_i - \bar{X}) &= \sum_{i=1}^{n} \big[(Y_i - \bar{Y}) - \hat{\beta}_1(X_i - \bar{X})\big](X_i - \bar{X}) \\
&= \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) - \hat{\beta}_1 \sum_{i=1}^{n} (X_i - \bar{X})^2 \\
&= \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) - \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 \\
&= \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) - \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) \\
&= 0.
\end{aligned}$$

Hence

$$\boxed{\sum_{i=1}^{n} \hat{u}_i X_i = 0.}$$

(c) Substitute $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ into the predicted value:

$$\hat{Y}_i = \bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X_i = \bar{Y} + \hat{\beta}_1 (X_i - \bar{X}).$$

Then

$$\begin{aligned}
\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i &= \sum_{i=1}^{n} \hat{u}_i \big[\bar{Y} + \hat{\beta}_1 (X_i - \bar{X})\big] \\
&= \bar{Y} \sum_{i=1}^{n} \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^{n} \hat{u}_i (X_i - \bar{X}) \\
&= \bar{Y}(0) + \hat{\beta}_1 (0) && \text{(by (a), (b))} \\
&= 0.
\end{aligned}$$

Hence

$$\boxed{\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = 0.}$$

Method 3: Perturbing the least-squares minimiser

This method uses only the fact that $(\hat{\beta}_0, \hat{\beta}_1)$ gives the smallest possible value of

$$S(b_0, b_1) = \sum_{i=1}^{n} \big(Y_i - b_0 - b_1 X_i\big)^2,$$

and no derivatives. Each part moves the OLS line in one direction by an amount t; since the moved line cannot have a smaller sum of squares, a certain inner product must vanish. Geometrically, the three results say that the residual vector $(\hat{u}_1, \ldots, \hat{u}_n)$ is perpendicular to $(1, \ldots, 1)$, to $(X_1, \ldots, X_n)$ and to $(\hat{Y}_1, \ldots, \hat{Y}_n)$: the OLS fit is the orthogonal projection of $(Y_1, \ldots, Y_n)$ onto the set of straight-line fits.

We use one algebraic fact. Let A and $C > 0$ be numbers with

$$-2tA + t^2 C \geq 0 \quad \text{for every real } t.$$

Then $A = 0$. Indeed, if $A \neq 0$, take $t = A/C$:

$$-2\frac{A}{C}A + \frac{A^2}{C^2}C = -\frac{2A^2}{C} + \frac{A^2}{C} = -\frac{A^2}{C} < 0,$$

which contradicts the inequality.

(a) Shift the intercept: for every real t the line $(\hat{\beta}_0 + t) + \hat{\beta}_1 X$ is a candidate, so

$$S(\hat{\beta}_0 + t, \hat{\beta}_1) \geq S(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} \hat{u}_i^2.$$

The residual of the shifted line at observation i is

$$Y_i - (\hat{\beta}_0 + t) - \hat{\beta}_1 X_i = \hat{u}_i - t.$$

Expand:

$$\begin{aligned} S(\hat{\beta}_0 + t, \hat{\beta}_1) &= \sum_{i=1}^{n} (\hat{u}_i - t)^2 \\ &= \sum_{i=1}^{n} \big(\hat{u}_i^2 - 2t\hat{u}_i + t^2\big) \\ &= \sum_{i=1}^{n} \hat{u}_i^2 - 2t \sum_{i=1}^{n} \hat{u}_i + nt^2. \end{aligned}$$

Subtract $\sum_{i=1}^{n} \hat{u}_i^2$ from both sides of the inequality:

$$-2t \sum_{i=1}^{n} \hat{u}_i + nt^2 \geq 0 \quad \text{for every real } t.$$

Apply the fact with $A = \sum_{i=1}^{n} \hat{u}_i$ and $C = n > 0$:

$$\boxed{\sum_{i=1}^{n} \hat{u}_i = 0.}$$

(b) Tilt the slope: for every real t the line $\hat{\beta}_0+(\hat{\beta}_1+t)X$ is a candidate, so $S(\hat{\beta}_0, \hat{\beta}_1+t) \geq \sum_{i=1}^n \hat{u}_i^2$. Its residual at observation i is

$$Y_i - \hat{\beta}_0 - (\hat{\beta}_1 + t)X_i = \hat{u}_i - tX_i.$$

Expand:

$$\begin{aligned} S(\hat{\beta}_0, \hat{\beta}_1 + t) &= \sum_{i=1}^n (\hat{u}_i - tX_i)^2 \\ &= \sum_{i=1}^n \hat{u}_i^2 - 2t\sum_{i=1}^n \hat{u}_i X_i + t^2 \sum_{i=1}^n X_i^2. \end{aligned}$$

Hence

$$-2t\sum_{i=1}^n \hat{u}_i X_i + t^2 \sum_{i=1}^n X_i^2 \geq 0 \quad \text{for every real } t.$$

If $\sum_{i=1}^n X_i^2 = 0$, every $X_i = 0$ and $\sum_{i=1}^n \hat{u}_i X_i = 0$ at once. Otherwise apply the fact with $A = \sum_{i=1}^n \hat{u}_i X_i$ and $C = \sum_{i=1}^n X_i^2 > 0$. In both cases

$$\boxed{\sum_{i=1}^n \hat{u}_i X_i = 0.}$$

(c) Scale the whole line: for every real t the line $(1+t)\hat{\beta}_0 + (1+t)\hat{\beta}_1 X$ is a candidate, and at X_i it equals $(1+t)\hat{Y}_i$. Its residual at observation i is

$$Y_i - (1+t)\hat{Y}_i = (Y_i - \hat{Y}_i) - t\hat{Y}_i = \hat{u}_i - t\hat{Y}_i.$$

Expand:

$$\begin{aligned} S\big((1+t)\hat{\beta}_0, (1+t)\hat{\beta}_1\big) &= \sum_{i=1}^n (\hat{u}_i - t\hat{Y}_i)^2 \\ &= \sum_{i=1}^n \hat{u}_i^2 - 2t\sum_{i=1}^n \hat{u}_i \hat{Y}_i + t^2 \sum_{i=1}^n \hat{Y}_i^2. \end{aligned}$$

Since this is at least $\sum_{i=1}^n \hat{u}_i^2$,

$$-2t\sum_{i=1}^n \hat{u}_i \hat{Y}_i + t^2 \sum_{i=1}^n \hat{Y}_i^2 \geq 0 \quad \text{for every real } t.$$

If $\sum_{i=1}^n \hat{Y}_i^2 = 0$, every $\hat{Y}_i = 0$ and $\sum_{i=1}^n \hat{u}_i \hat{Y}_i = 0$ at once. Otherwise apply the fact with $A = \sum_{i=1}^n \hat{u}_i \hat{Y}_i$ and $C = \sum_{i=1}^n \hat{Y}_i^2 > 0$. This route does not use (a) or (b). In both cases

$$\boxed{\sum_{i=1}^n \hat{u}_i \hat{Y}_i = 0.}$$

Question 3 (Assumption 1 by diagrams, and a counterexample)

Problem

(a) Explain the meaning of the OLS Assumption 1 by using diagrams.

(b) Explain how we can obtain the figure shown on slide No. 11 for Lecture 3.

(c) Give an example where $E[u] = 0$ but $E[u \mid X] \neq 0$ by using diagrams.

(Tutorial Question Set No. 3, Q3.)

Solution

Method 1: Conditional-mean figures and a discrete counterexample

(a) Least Squares Assumption 1 states

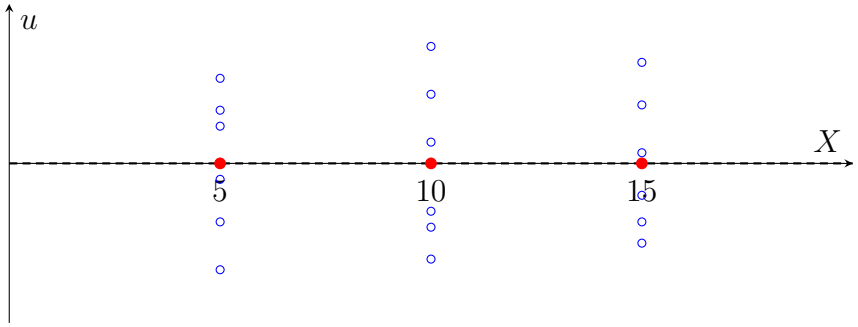
$$E[u_i \mid X_i] = 0.$$

It is a statement about the population and must hold at every value of X_i; for the values plotted in Figure 1,

$$E[u_i \mid X_i = 5] = 0, \qquad E[u_i \mid X_i = 10] = 0, \qquad E[u_i \mid X_i = 15] = 0.$$

- Figure 1 plots draws of the error u against X. At each fixed X the draws (blue circles) lie both above and below the horizontal axis, and their population mean (red dot) sits exactly on $u = 0$.
- In words, the factors other than X that are collected in u have the same average, zero, whatever the value of X. For example, if Y is the hourly wage (\$), X is years of schooling and u contains ability, Assumption 1 says that average ability is the same among people with 5, 10 and 15 years of schooling.
- The red dots are population conditional means, not sample averages. The six plotted draws at $X = 5$ sum to $1.6 - 1.1 + 0.7 - 2.0 + 1.0 - 0.3 = -0.1$, so their sample average is $-0.1/6 \approx -0.017$: a finite sample scatters around the population value 0 without hitting it exactly.
- Because the conditional mean of u is zero at every X, knowing X does not help to predict u. This is what allows β_1 to be read as the effect of X on Y rather than the effect of the omitted factors.

Figure 1: u_i vs. X_i – zero conditional mean at $X = 5, 10, 15$



Assumption 1: $E[u_i \mid X_i] = 0$ for every value of X_i.

(b) The figure on slide 11 (Figure 2 below) plots Y instead of u. Start from the model

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

Take the conditional expectation of both sides given X_i:

$$E[Y_i \mid X_i] = E[\beta_0 + \beta_1 X_i + u_i \mid X_i].$$

Conditional expectation is linear, so it splits across the sum:

$$E[Y_i \mid X_i] = E[\beta_0 \mid X_i] + E[\beta_1 X_i \mid X_i] + E[u_i \mid X_i].$$

β_0 is a constant, and $\beta_1 X_i$ is a known constant once X_i is given, so

$$E[\beta_0 \mid X_i] = \beta_0, \qquad E[\beta_1 X_i \mid X_i] = \beta_1 X_i.$$

By Least Squares Assumption 1, $E[u_i \mid X_i] = 0$. Substituting,

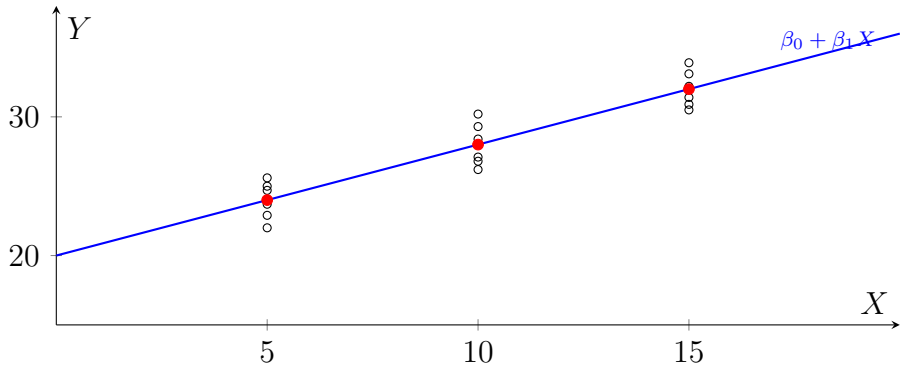
$$\begin{aligned} E[Y_i \mid X_i] &= \beta_0 + \beta_1 X_i + 0 \\ &= \beta_0 + \beta_1 X_i. \end{aligned}$$

So Figure 2 is obtained from Figure 1 by lifting every point at X by the height of the population line, $\beta_0 + \beta_1 X$. In Figure 2 the line is $\beta_0 + \beta_1 X = 20 + 0.8X$, so the lifts are

$$\begin{aligned} X = 5: &\quad 20 + 0.8(5) = 24, \\ X = 10: &\quad 20 + 0.8(10) = 28, \\ X = 15: &\quad 20 + 0.8(15) = 32. \end{aligned}$$

For example, the draw $u = 1.6$ at $X = 5$ in Figure 1 becomes $Y = 24 + 1.6 = 25.6$ in Figure 2, and the red dot $(5, 0)$ becomes $(5, 24)$. Each red dot, a conditional mean $E[Y \mid X]$, lands on the line, while the individual values of Y scatter above and below it by u.

Figure 2: Y_i vs. X_i – data centred on the population line at $X = 5, 10, 15$



$$\boxed{E[Y_i \mid X_i] = \beta_0 + \beta_1 X_i.}$$

(c) Let X take two values with

$$\Pr(X = 0) = \Pr(X = 1) = \tfrac{1}{2},$$

and let the error have conditional means

$$E[u \mid X = 0] = -1, \qquad E[u \mid X = 1] = 1.$$

One way to generate this is $u = 2X - 1 + v$ with v independent of X and $E[v] = 0$; then $E[u \mid X = x] = 2x - 1$, which is -1 at $x = 0$ and 1 at $x = 1$. The unconditional mean follows from the Law of Total Expectation, which weights each conditional mean by its probability:

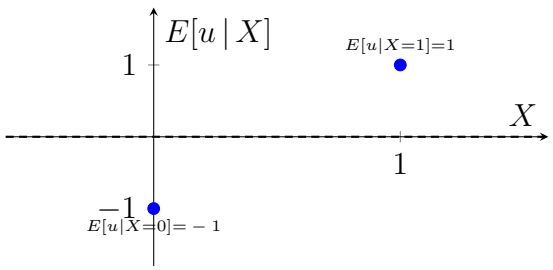
$$\begin{aligned} E[u] &= \Pr(X = 0)\, E[u \mid X = 0] + \Pr(X = 1)\, E[u \mid X = 1] \\ &= \tfrac{1}{2}(-1) + \tfrac{1}{2}(1) \\ &= -\tfrac{1}{2} + \tfrac{1}{2} \\ &= 0. \end{aligned}$$

But the conditional means are

$$E[u \mid X = 0] = -1 \neq 0, \qquad E[u \mid X = 1] = 1 \neq 0,$$

so Least Squares Assumption 1 fails at both values of X. In Figure 3 the two conditional means lie on opposite sides of the axis and cancel only when averaged. Economically, if $X = 1$ indicates attending a selective programme and u contains ability, higher-ability people select into $X = 1$: their other factors average $+1$ while those of the $X = 0$ group average -1, even though the population as a whole averages zero.

Figure 3: binary X – $E[u] = 0$ but $E[u \mid X] \neq 0$



Counterexample: binary X with $E[u|X{=}0]{=}-1$, $E[u|X{=}1]{=}1$: $E[u] = 0$ but $E[u|X] \neq 0$.

Unofficial - QRS@NTU - qrsntu.org

Method 2: Conditional densities and a functional-form counterexample

(a) Write $f_{u\,|\,X}(u\,|\,x)$ for the density of u among population members with $X = x$. Least Squares Assumption 1 states that the balance point of every such density is zero:

$$E[u\,|\,X = x] = \int_{-\infty}^{\infty} u\, f_{u\,|\,X}(u\,|\,x)\, du = 0 \quad \text{for every } x.$$

The unconditional mean is an average of these balance points, weighted by the density f_X of X:

$$E[u] = \int_{-\infty}^{\infty} E[u\,|\,X = x]\, f_X(x)\, dx.$$

- In Figure 1, each vertical slice at a fixed X is one conditional distribution of u. Assumption 1 requires every slice to balance at $u = 0$, not merely the cloud as a whole.
- $E[u] = 0$ allows slices that balance above zero to offset slices that balance below zero. Assumption 1 rules this out, because it requires the integrand $E[u\,|\,X = x]$ itself to be zero, not only its weighted average.
- The six draws at $X = 10$ in Figure 1 sum to $2.2 - 1.8 + 0.4 - 0.9 + 1.3 - 1.2 = 0$, so their sample average is 0. Such a sample average is only an estimate of the balance point; Assumption 1 is a statement about the population density of each slice.

Assumption 1: $E[u_i\,|\,X_i] = 0$ for every value of X_i.

(b) Fix $X = x$ and let $m(x) = \beta_0 + \beta_1 x$, a constant once x is fixed. Then $Y = m(x) + u$, so $Y \le y$ exactly when $u \le y - m(x)$, and the conditional density of Y is the density of u shifted by $m(x)$:

$$f_{Y\,|\,X}(y\,|\,x) = f_{u\,|\,X}\big(y - m(x)\,|\,x\big).$$

Hence

$$E[Y\,|\,X = x] = \int_{-\infty}^{\infty} y\, f_{u\,|\,X}\big(y - m(x)\,|\,x\big)\, dy.$$

Substitute $v = y - m(x)$, so $y = m(x) + v$ and $dy = dv$; as y runs over $(-\infty, \infty)$, so does v. A density integrates to one, so the first integral below equals 1:

$$\begin{aligned}
E[Y\,|\,X = x] &= \int_{-\infty}^{\infty} \big(m(x) + v\big) f_{u\,|\,X}(v\,|\,x)\, dv \\
&= m(x) \int_{-\infty}^{\infty} f_{u\,|\,X}(v\,|\,x)\, dv + \int_{-\infty}^{\infty} v\, f_{u\,|\,X}(v\,|\,x)\, dv \\
&= m(x) \cdot 1 + E[u\,|\,X = x] \\
&= m(x) + 0 \qquad \text{(Assumption 1)} \\
&= \beta_0 + \beta_1 x.
\end{aligned}$$

So the picture on slide 11 is obtained by sliding each conditional distribution of Figure 1 upward by $\beta_0 + \beta_1 x$: the shape of each slice is unchanged and its balance point moves from 0 to the population line. The same shift applies to sample draws: at $X = 5$ the six values of Y in Figure 2 average

$$\frac{25.6 + 22.9 + 24.7 + 22.0 + 25.0 + 23.7}{6} = \frac{143.9}{6} \approx 23.983 = 24 + (-0.017),$$

which is the Figure 1 average -0.017 moved up by exactly $m(5) = 24$.

$$\boxed{E[Y_i \mid X_i] = \beta_0 + \beta_1 X_i.}$$

(c) Let $X \sim N(0,1)$ and let the error be a nonlinear function of X:

$$u = X^2 - 1.$$

For the unconditional mean, use $\mathrm{Var}(X) = E[X^2] - (E[X])^2$:

$$E[X^2] = \mathrm{Var}(X) + (E[X])^2 = 1 + 0^2 = 1,$$
$$E[u] = E[X^2] - 1 = 1 - 1 = 0.$$

For the conditional mean: given $X = x$, the value $u = x^2 - 1$ is known, so

$$E[u \mid X = x] = x^2 - 1.$$

For example,

$$E[u \mid X = 0] = 0^2 - 1 = -1 < 0,$$
$$E[u \mid X = 2] = 2^2 - 1 = 3 > 0,$$

and $E[u \mid X = x] = 0$ only at $x = \pm 1$, so Assumption 1 fails at every other value of X. Economically, this is the case of a curved relationship: if the population relationship is $Y = \beta_0 + \beta_1 X + (X^2 - 1)$, a straight-line model puts the curvature $X^2 - 1$ into u, which is then systematically positive for large $|X|$ and negative for small $|X|$.

$$\boxed{\begin{array}{c}\text{Counterexample: continuous } u = X^2 - 1,\ X \sim N(0,1): \\ E[u] = 0 \text{ but } E[u|X] \neq 0.\end{array}}$$

Unofficial Solutions

Method 3: Covariance test via the Law of Iterated Expectations

This method applies to part (c). It uses the Law of Iterated Expectations, $E[Z] = E\big[E[Z \mid X]\big]$, which Lecture 3 states without requiring a proof.

First, Assumption 1 forces $\mathrm{Cov}(X, u) = 0$. Suppose $E[u \mid X] = 0$. Then

$$E[u] = E\big[E[u \mid X]\big] = E[0] = 0.$$

Given X, the factor X is a known constant, so it comes out of the inner expectation:

$$E[Xu \mid X] = X\, E[u \mid X] = X \cdot 0 = 0,$$

and therefore

$$E[Xu] = E\big[E[Xu \mid X]\big] = E[0] = 0.$$

Hence

$$\mathrm{Cov}(X, u) = E[Xu] - E[X]E[u] = 0 - E[X] \cdot 0 = 0.$$

Taking the contrapositive: if $E[u] = 0$ and $\mathrm{Cov}(X, u) \neq 0$, then $E[u \mid X] \neq 0$ for some value of X.

Apply this to the binary design $\Pr(X = 0) = \Pr(X = 1) = \frac{1}{2}$, $E[u \mid X = 0] = -1$, $E[u \mid X = 1] = 1$. By the Law of Iterated Expectations,

$$\begin{aligned} E[u] &= E\big[E[u \mid X]\big] = \tfrac{1}{2}(-1) + \tfrac{1}{2}(1) = 0, \\ E[Xu] &= E\big[X\, E[u \mid X]\big] = \tfrac{1}{2} \cdot 0 \cdot (-1) + \tfrac{1}{2} \cdot 1 \cdot 1 = 0 + \tfrac{1}{2} = \tfrac{1}{2}, \\ E[X] &= \tfrac{1}{2} \cdot 0 + \tfrac{1}{2} \cdot 1 = \tfrac{1}{2}. \end{aligned}$$

Therefore

$$\mathrm{Cov}(X, u) = E[Xu] - E[X]E[u] = \tfrac{1}{2} - \tfrac{1}{2} \cdot 0 = \tfrac{1}{2} \neq 0,$$

and the contrapositive gives $E[u \mid X] \neq 0$ while $E[u] = 0$. The test works in one direction only: for the continuous design of Method 2, $\mathrm{Cov}(X, u) = E[X(X^2 - 1)] - E[X]E[u] = E[X^3] - E[X] - 0 = 0 - 0 = 0$ (odd moments of $N(0, 1)$ vanish), so the covariance is silent there even though $E[u \mid X] = X^2 - 1 \neq 0$. Zero covariance does not imply Assumption 1.

Counterexample: binary X with $E[u\|X{=}0]{=}-1$, $E[u\|X{=}1]{=}1$: $E[u] = 0$ but $E[u\|X] \neq 0$.

Question 4 (Binary regressor: SSS vs. SOH wages)

Problem

Suppose we want to find out if there exists a difference in average wages between graduates from the School of Social Sciences (SSS) and graduates from the School of Humanities (SOH), using the linear regression model with an indicator variable ($X = 1$ or 0).

(a) Write down the linear regression model for this problem. Define the dependent variable, explanatory variable, and error term in your model.

(b) Write down the corresponding OLS Assumption 1. Explain in what case the OLS Assumption 1 is likely to hold/unlikely to hold.

(c) Now suppose that the OLS Assumption 1 does hold; explain the specific meaning of the parameters β_0 and β_1 using conditional expectation. *[Hint: derive the conditional expectation $E[Y_i \mid X_i]$ for $X_i = 1$ and $X_i = 0$, respectively.]*

(d) Illustrate the meaning of β_0 and β_1 by using a diagram.

(Tutorial Question Set No. 3, Q4.)

Solution

Method 1: Conditional expectation and the OLS formula

(a) The model is

$$\text{wage}_i = \beta_0 + \beta_1 X_i + u_i, \qquad i = 1, \ldots, n.$$

- The dependent variable wage_i is the wage of graduate i, measured for example in SGD per month.
- The explanatory variable is the indicator $X_i = 1$ if graduate i is from SSS and $X_i = 0$ if graduate i is from SOH. It has no units and takes only these two values.
- The error term u_i collects every other factor that affects the wage of graduate i: ability, family background, work experience, effort, choice of industry. It is a population object and is not observed.

$$\boxed{\text{wage}_i = \beta_0 + \beta_1 X_i + u_i, \quad X_i = 1 \text{ (SSS)},\ 0 \text{ (SOH)}}$$

(b) With $X_i \in \{0, 1\}$, conditioning on X_i means conditioning on one of two groups, so

$$E[u_i \mid X_i] = 0 \iff E[u_i \mid X_i = 1] = 0 \text{ and } E[u_i \mid X_i = 0] = 0.$$

To see what a failure does, take the conditional expectation of the model within each group, using linearity of conditional expectation:

$$E[\text{wage}_i \mid X_i = 1] = \beta_0 + \beta_1 + E[u_i \mid X_i = 1],$$
$$E[\text{wage}_i \mid X_i = 0] = \beta_0 + E[u_i \mid X_i = 0].$$

Subtract the second line from the first:

$$E[\text{wage}_i \mid X_i = 1] - E[\text{wage}_i \mid X_i = 0] = \beta_1 + \big(E[u_i \mid X_i = 1] - E[u_i \mid X_i = 0]\big).$$

- Likely to hold: if graduates were assigned to SSS or SOH at random (for example by a lottery among admitted students), ability, background and effort would have the same distribution in both schools, so the other factors would average to the same value, zero, in each group.
- Unlikely to hold: in practice students choose their school. If students with stronger quantitative ability or better-connected families are more likely to choose SSS, then $E[u_i \mid X_i = 1] > 0 > E[u_i \mid X_i = 0]$ and Assumption 1 fails.
- Hypothetical illustration: with 40% of graduates in SSS, $E[u_i \mid X_i = 1] = 150$ and $E[u_i \mid X_i = 0] = -100$ SGD per month give $E[u_i] = 0.4(150) + 0.6(-100) = 60 - 60 = 0$, yet the gap in mean wages is $\beta_1 + (150 - (-100)) = \beta_1 + 250$: the school difference would be overstated by 250 SGD per month.

Assumption 1: $E[u_i \mid X_i{=}1] = E[u_i \mid X_i{=}0] = 0$; holds under random assignment, fails under self-selection by ability/background

(c) Take the conditional expectation of the model given X_i and use linearity:

$$\begin{aligned} E[\text{wage}_i \mid X_i] &= \beta_0 + \beta_1 X_i + E[u_i \mid X_i] \\ &= \beta_0 + \beta_1 X_i + 0 \qquad \text{(Assumption 1)} \\ &= \beta_0 + \beta_1 X_i. \end{aligned}$$

Evaluate at the two values of X_i:

$$\begin{aligned} X_i = 0: &\quad E[\text{wage}_i \mid X_i = 0] = \beta_0 + \beta_1(0) = \beta_0, \\ X_i = 1: &\quad E[\text{wage}_i \mid X_i = 1] = \beta_0 + \beta_1(1) = \beta_0 + \beta_1. \end{aligned}$$

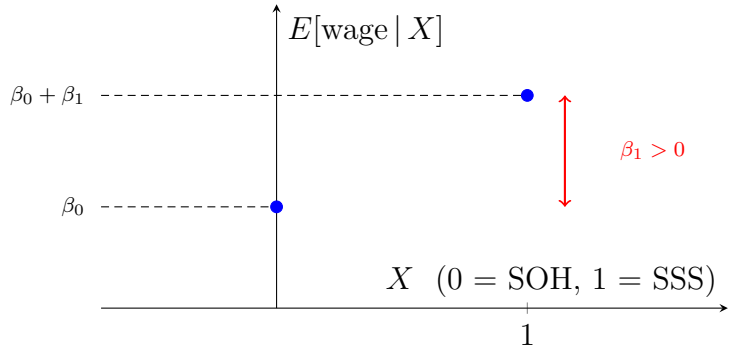
Subtract the first from the second:

$$E[\text{wage}_i \mid X_i = 1] - E[\text{wage}_i \mid X_i = 0] = (\beta_0 + \beta_1) - \beta_0 = \beta_1.$$

- β_0 is the population mean wage of SOH graduates, in SGD per month.
- β_1 is the population difference in mean wages, SSS minus SOH, in SGD per month. Because X jumps from 0 to 1 and takes no values in between, β_1 is a level shift between two groups, not a slope along a continuum.
- Both are population objects; the sample counterparts below estimate them.
- Hypothetical illustration: if $E[\text{wage} \mid \text{SOH}] = 3{,}900$ and $E[\text{wage} \mid \text{SSS}] = 4{,}300$ SGD per month, then $\beta_0 = 3{,}900$ and $\beta_1 = 4{,}300 - 3{,}900 = 400$ SGD per month.

$\beta_0 = E[\text{wage}\mid\text{SOH}], \quad \beta_1 = E[\text{wage}\mid\text{SSS}] - E[\text{wage}\mid\text{SOH}]$ (not a slope)

(d) Plot the two conditional means against X: β_0 at $X = 0$ and $\beta_0 + \beta_1$ at $X = 1$, drawn for the case $\beta_1 > 0$. The vertical distance between the two dots is β_1; in the illustration of (c) the dots would sit at 3,900 and 4,300 and the arrow would measure 400 SGD per month.



$$\boxed{\beta_0 \text{ at } X = 0, \qquad \beta_0 + \beta_1 \text{ at } X = 1}$$

Sample counterpart. Let n_1 and n_0 be the numbers of SSS ($X_i = 1$) and SOH ($X_i = 0$) graduates in the sample, $n = n_0 + n_1$ with $n_0, n_1 > 0$, and let $\bar{Y}_1, \bar{Y}_0$ be the sample mean wages of the two groups, so that

$$\sum_{X_i=1} Y_i = n_1 \bar{Y}_1, \qquad \sum_{X_i=0} Y_i = n_0 \bar{Y}_0, \qquad \bar{Y} = \frac{n_1 \bar{Y}_1 + n_0 \bar{Y}_0}{n}.$$

Because $X_i \in \{0, 1\}$, we have $X_i^2 = X_i$, $\sum_{i=1}^n X_i = n_1$ and $\bar{X} = n_1/n$. Expand the denominator of $\hat{\beta}_1$, using $\sum_{i=1}^n X_i = n\bar{X}$:

$$\begin{aligned}
\sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 \\
&= \sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 \\
&= n_1 - n\left(\frac{n_1}{n}\right)^2 = n_1 - \frac{n_1^2}{n} = \frac{n_1(n - n_1)}{n} = \frac{n_1 n_0}{n}.
\end{aligned}$$

Expand the numerator in the same way, using $\sum_{i=1}^n Y_i = n\bar{Y}$; only the SSS graduates contribute to $\sum_{i=1}^n X_i Y_i$:

$$\begin{aligned}
\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) &= \sum_{i=1}^n X_i Y_i - \bar{X} \sum_{i=1}^n Y_i - \bar{Y} \sum_{i=1}^n X_i + n\bar{X}\bar{Y} \\
&= \sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y} \\
&= n_1 \bar{Y}_1 - n_1 \bar{Y} \\
&= n_1 \left(\bar{Y}_1 - \frac{n_1 \bar{Y}_1 + n_0 \bar{Y}_0}{n}\right) = n_1 \cdot \frac{(n - n_1)\bar{Y}_1 - n_0 \bar{Y}_0}{n} \\
&= \frac{n_1 n_0}{n} (\bar{Y}_1 - \bar{Y}_0).
\end{aligned}$$

Divide:

$$\hat{\beta}_1 = \frac{\frac{n_1 n_0}{n}(\bar{Y}_1 - \bar{Y}_0)}{\frac{n_1 n_0}{n}} = \bar{Y}_1 - \bar{Y}_0.$$

Then

$$\begin{aligned}\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\ &= \frac{n_1 \bar{Y}_1 + n_0 \bar{Y}_0}{n} - (\bar{Y}_1 - \bar{Y}_0)\frac{n_1}{n} \\ &= \frac{n_1 \bar{Y}_1 + n_0 \bar{Y}_0 - n_1 \bar{Y}_1 + n_1 \bar{Y}_0}{n} \\ &= \frac{(n_0 + n_1)\bar{Y}_0}{n} = \bar{Y}_0.\end{aligned}$$

For a numerical illustration, take a hypothetical sample of three SOH graduates earning 3,600, 3,900, 4,200 and two SSS graduates earning 4,100, 4,500 SGD per month. Then $n_0 = 3$, $n_1 = 2$, $n = 5$ and

$$\bar{Y}_0 = \frac{3600 + 3900 + 4200}{3} = \frac{11700}{3} = 3900, \qquad \bar{Y}_1 = \frac{4100 + 4500}{2} = \frac{8600}{2} = 4300,$$
$$\bar{Y} = \frac{11700 + 8600}{5} = \frac{20300}{5} = 4060, \qquad \bar{X} = \frac{2}{5} = 0.4.$$

The formula pieces are

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = 2 - 5(0.4)^2 = 2 - 0.8 = 1.2 = \frac{(2)(3)}{5},$$
$$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y} = 8600 - 5(0.4)(4060) = 8600 - 8120 = 480,$$

so

$$\hat{\beta}_1 = \frac{480}{1.2} = 400 = 4300 - 3900, \qquad \hat{\beta}_0 = 4060 - 400(0.4) = 4060 - 160 = 3900 = \bar{Y}_0.$$

The estimated SSS–SOH gap in this sample is 400 SGD per month.

$$\boxed{\text{Sample: } \hat{\beta}_0 = \bar{Y}_0, \quad \hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0}$$

Unofficial QRS@NTU material

Method 2: Moment conditions and residual orthogonality

This method covers parts (b), (c) and the sample counterpart; parts (a) and (d) are as in Method 1. Let $p = \Pr(X_i = 1)$ be the population share of SSS graduates, with $0 < p < 1$.

(b) By the Law of Total Expectation over the two groups,

$$\begin{aligned} E[u_i] &= p\,E[u_i \mid X_i = 1] + (1-p)\,E[u_i \mid X_i = 0], \\ E[X_i u_i] &= p\,E[1 \cdot u_i \mid X_i = 1] + (1-p)\,E[0 \cdot u_i \mid X_i = 0] = p\,E[u_i \mid X_i = 1]. \end{aligned}$$

Since $p > 0$, the second line gives

$$E[X_i u_i] = 0 \iff E[u_i \mid X_i = 1] = 0.$$

When $E[u_i \mid X_i = 1] = 0$, the first line reduces to $E[u_i] = (1-p)\,E[u_i \mid X_i = 0]$, and since $1 - p > 0$,

$$E[u_i] = 0 \iff E[u_i \mid X_i = 0] = 0.$$

Hence, for a binary regressor,

$$E[u_i \mid X_i = 1] = E[u_i \mid X_i = 0] = 0 \iff E[u_i] = 0 \text{ and } E[X_i u_i] = 0.$$

With $E[u_i] = 0$, $E[X_i u_i] = \mathrm{Cov}(X_i, u_i) + E[X_i]E[u_i] = \mathrm{Cov}(X_i, u_i)$, so here Assumption 1 says that school choice is uncorrelated with the other wage factors.

- Random assignment makes X_i independent of u_i, so $E[X_i u_i] = E[X_i]E[u_i] = p \cdot 0 = 0$ and Assumption 1 holds.
- Self-selection by ability or background makes X_i correlated with the ability and background terms inside u_i, so $E[X_i u_i] = \mathrm{Cov}(X_i, u_i) \neq 0$ and Assumption 1 fails.

Assumption 1: $E[u_i \mid X_i{=}1] = E[u_i \mid X_i{=}0] = 0$; holds under random assignment, fails under self-selection by ability/background

(c) Under Assumption 1, part (b) gives the two population moment conditions

$$E[u_i] = 0, \qquad E[X_i u_i] = 0.$$

Write $\mu_1 = E[\text{wage}_i \mid X_i = 1]$ and $\mu_0 = E[\text{wage}_i \mid X_i = 0]$. By the Law of Total Expectation,

$$\begin{aligned} E[\text{wage}_i] &= p\mu_1 + (1-p)\mu_0, \\ E[X_i\,\text{wage}_i] &= p\mu_1 + (1-p)\cdot 0 = p\mu_1, \\ E[X_i] &= E[X_i^2] = p. \end{aligned}$$

Substitute $u_i = \text{wage}_i - \beta_0 - \beta_1 X_i$ into the first condition:

$$0 = E[\text{wage}_i - \beta_0 - \beta_1 X_i] = p\mu_1 + (1-p)\mu_0 - \beta_0 - \beta_1 p. \tag{1}$$

Substitute into the second condition, using $X_i^2 = X_i$:

$$\begin{aligned} 0 = E\big[X_i(\text{wage}_i - \beta_0 - \beta_1 X_i)\big] &= E[X_i\,\text{wage}_i] - \beta_0 E[X_i] - \beta_1 E[X_i^2] \\ &= p\mu_1 - \beta_0 p - \beta_1 p = p\big(\mu_1 - \beta_0 - \beta_1\big). \end{aligned} \tag{2}$$

Since $p > 0$, (2) gives

$$\mu_1 = \beta_0 + \beta_1.$$

Substitute this into (1):

$$\begin{aligned} 0 &= p(\beta_0 + \beta_1) + (1-p)\mu_0 - \beta_0 - \beta_1 p \\ &= p\beta_0 + (1-p)\mu_0 - \beta_0 \\ &= (1-p)\mu_0 - (1-p)\beta_0 = (1-p)(\mu_0 - \beta_0). \end{aligned}$$

Since $1-p > 0$, $\beta_0 = \mu_0$, and then $\beta_1 = \mu_1 - \beta_0 = \mu_1 - \mu_0$. In words, β_0 is the mean wage of SOH graduates and β_1 is the SSS–SOH difference in mean wages, a difference between two groups rather than a slope.

$$\boxed{\beta_0 = E[\text{wage}|\text{SOH}], \quad \beta_1 = E[\text{wage}|\text{SSS}] - E[\text{wage}|\text{SOH}] \text{ (not a slope)}}$$

Sample counterpart. OLS imposes the sample analogues of the two moment conditions, which are the residual identities of Question 2:

$$\sum_{i=1}^{n} \hat{u}_i = 0, \qquad \sum_{i=1}^{n} \hat{u}_i X_i = 0.$$

For an SSS graduate ($X_i = 1$) the residual is $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1$; for an SOH graduate ($X_i = 0$) it is $\hat{u}_i = Y_i - \hat{\beta}_0$. In $\sum_{i=1}^{n} \hat{u}_i X_i$ the SOH terms are multiplied by 0, so

$$\begin{aligned} 0 = \sum_{i=1}^{n} \hat{u}_i X_i &= \sum_{X_i=1} \hat{u}_i \\ &= \sum_{X_i=1} (Y_i - \hat{\beta}_0 - \hat{\beta}_1) \\ &= n_1 \bar{Y}_1 - n_1 \hat{\beta}_0 - n_1 \hat{\beta}_1 = n_1 \big(\bar{Y}_1 - \hat{\beta}_0 - \hat{\beta}_1\big). \end{aligned}$$

Since $n_1 > 0$,

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

Subtract the two identities to isolate the SOH group:

$$\begin{aligned} 0 = \sum_{i=1}^{n} \hat{u}_i - \sum_{i=1}^{n} \hat{u}_i X_i &= \sum_{X_i=0} \hat{u}_i \\ &= \sum_{X_i=0} (Y_i - \hat{\beta}_0) \\ &= n_0 \bar{Y}_0 - n_0 \hat{\beta}_0 = n_0 \big(\bar{Y}_0 - \hat{\beta}_0\big). \end{aligned}$$

Since $n_0 > 0$, $\hat{\beta}_0 = \bar{Y}_0$, and then $\hat{\beta}_1 = \bar{Y}_1 - \hat{\beta}_0 = \bar{Y}_1 - \bar{Y}_0$. In the five-graduate sample of Method 1, $\hat{\beta}_0 = 3900$ and $\hat{\beta}_1 = 400$ give SOH residuals $3600 - 3900 = -300$, $3900 - 3900 = 0$, $4200 - 3900 = 300$ (sum 0) and SSS residuals $4100 - 4300 = -200$, $4500 - 4300 = 200$ (sum 0), so both identities hold.

$$\boxed{\text{Sample: } \hat{\beta}_0 = \bar{Y}_0, \quad \hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0}$$

Method 3: Splitting the sum of squared residuals by group

This method covers the sample counterpart; parts (a)–(d) are as in Method 1. For a candidate line $b_0 + b_1 X$, the fitted value is b_0 for an SOH graduate ($X_i = 0$) and $b_0 + b_1$ for an SSS graduate ($X_i = 1$). Split the sum of squares by group:

$$S(b_0, b_1) = \sum_{X_i=0} (Y_i - b_0)^2 + \sum_{X_i=1} (Y_i - b_0 - b_1)^2.$$

Let $c_0 = b_0$ and $c_1 = b_0 + b_1$. The map $(b_0, b_1) \mapsto (c_0, c_1)$ is one-to-one, with inverse $b_0 = c_0$, $b_1 = c_1 - c_0$, so minimising over (b_0, b_1) is the same as minimising over (c_0, c_1):

$$S = Q_0(c_0) + Q_1(c_1), \qquad Q_g(c) = \sum_{X_i=g} (Y_i - c)^2, \quad g = 0, 1.$$

Each Q_g depends on its own variable only, so each can be minimised separately. Complete the square by writing $Y_i - c = (Y_i - \bar{Y}_g) + (\bar{Y}_g - c)$; the cross term vanishes because $\sum_{X_i=g} Y_i = n_g \bar{Y}_g$:

$$\begin{aligned} Q_g(c) &= \sum_{X_i=g} \big[(Y_i - \bar{Y}_g) + (\bar{Y}_g - c)\big]^2 \\ &= \sum_{X_i=g} (Y_i - \bar{Y}_g)^2 + 2(\bar{Y}_g - c) \sum_{X_i=g} (Y_i - \bar{Y}_g) + n_g(\bar{Y}_g - c)^2 \\ &= \sum_{X_i=g} (Y_i - \bar{Y}_g)^2 + 2(\bar{Y}_g - c)\big(n_g \bar{Y}_g - n_g \bar{Y}_g\big) + n_g(\bar{Y}_g - c)^2 \\ &= \sum_{X_i=g} (Y_i - \bar{Y}_g)^2 + n_g(\bar{Y}_g - c)^2. \end{aligned}$$

The first term does not depend on c, and the second is nonnegative and equals 0 only at $c = \bar{Y}_g$. Hence Q_g is minimised at $c = \bar{Y}_g$:

$$\hat{c}_0 = \bar{Y}_0, \qquad \hat{c}_1 = \bar{Y}_1.$$

Translate back:

$$\hat{\beta}_0 = \hat{c}_0 = \bar{Y}_0, \qquad \hat{\beta}_1 = \hat{c}_1 - \hat{c}_0 = \bar{Y}_1 - \bar{Y}_0.$$

The OLS line therefore passes through the two group means $(0, \bar{Y}_0)$ and $(1, \bar{Y}_1)$, and its coefficient on X is the gap between them over the unit distance from $X = 0$ to $X = 1$. In the five-graduate sample of Method 1,

$$\begin{aligned} Q_0(c) &= \big[(-300)^2 + 0^2 + 300^2\big] + 3(3900 - c)^2 = 180000 + 3(3900 - c)^2, \\ Q_1(c) &= \big[(-200)^2 + 200^2\big] + 2(4300 - c)^2 = 80000 + 2(4300 - c)^2, \end{aligned}$$

which are minimised at $c_0 = 3900$ and $c_1 = 4300$, giving $\hat{\beta}_0 = 3900$, $\hat{\beta}_1 = 4300 - 3900 = 400$ SGD per month and a minimum sum of squared residuals of $180000 + 80000 = 260000$.

$$\boxed{\text{Sample: } \hat{\beta}_0 = \bar{Y}_0, \quad \hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0}$$

Final Answers

1. Population line $\beta_0 + \beta_1 X$; error $u_i = Y_i - (\beta_0 + \beta_1 X_i)$ (both unobserved); OLS line $\hat{\beta}_0 + \hat{\beta}_1 X$; predicted value $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$; residual $\hat{u}_i = Y_i - \hat{Y}_i$ (all computed).

2. $\sum_{i=1}^{n} \hat{u}_i = 0$; $\sum_{i=1}^{n} \hat{u}_i X_i = 0$; $\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = 0$.

3. (a) $E[u_i|X_i] = 0$ at every X (Figure 1). (b) $E[Y_i|X_i] = \beta_0 + \beta_1 X_i$ (Figure 2). (c) Binary X: $E[u|X{=}0] = -1$, $E[u|X{=}1] = 1$ (Figure 3); continuous: $u = X^2 - 1$, $X \sim N(0, 1)$; both have $E[u] = 0$, $E[u|X] \neq 0$.

4. (a) $\text{wage}_i = \beta_0 + \beta_1 X_i + u_i$, $X_i = 1$ (SSS) or 0 (SOH). (b) $E[u_i|X_i{=}1] = E[u_i|X_i{=}0] = 0$; holds under random assignment, fails under self-selection. (c) $\beta_0 = E[\text{wage}|\text{SOH}]$, $\beta_1 = E[\text{wage}|\text{SSS}] - E[\text{wage}|\text{SOH}]$. (d) β_0 at $X = 0$, $\beta_0 + \beta_1$ at $X = 1$. Sample: $\hat{\beta}_0 = \bar{Y}_0$, $\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$.

Unofficial QRS@NTU Tutorials