OLS Consistency and Matrix Formalization

Aarón Rey Vázquez

January 31, 2026

1. DERIVATION OF THE OLS ESTIMATOR FOR β_1

Consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, \ldots, n. \tag{1}$$

The OLS method estimates the parameters by minimizing the sum of squared residuals:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \hat{u}_i^2 = \min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{2}$$

Substituting $\hat{u}_i = y_i - (\beta_0 + \beta_1 x_i)$, this problem can be written as

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{3}$$

Taking the first-order condition with respect to β_1 yields

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i = 0. \tag{4}$$

Definition 1.1. Solving the first-order condition with respect to β_0 yields

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \tag{5}$$

where $\bar{y}$ and $\bar{x}$ denote the sample means of y_i and x_i, respectively.

Substituting this expression into the FOC for β_1 and simplifying leads to:

$$\sum_{i=1}^{n} \left[y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i \right] x_i = 0, \tag{6}$$

$$\sum_{i=1}^{n} \left[(y_i - \bar{y}) x_i - \hat{\beta}_1 (x_i - \bar{x}) x_i \right] = 0, \tag{7}$$

$$\sum_{i=1}^{n} (y_i - \bar{y}) x_i - \hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{x}) x_i = 0. \tag{8}$$

Solving for $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{9}$$

2. SAMPLING DISTRIBUTION OF THE OLS ESTIMATOR

2.1. Unbiasedness: Conditional Mean

To derive the sampling distribution of $\hat{\beta}_1$, we begin by computing its expectation. Taking sample averages in the regression model yields

$$\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{u}. \tag{10}$$

Subtracting this expression from the original model gives

$$y_i - \bar{y} = \beta_1(x_i - \bar{x}) + (u_i - \bar{u}). \tag{11}$$

Substituting into the OLS estimator previously calculated:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})[\beta_1(x_i - \bar{x}) + (u_i - \bar{u})]}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{12}$$

$$= \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{13}$$

Taking expectations,

$$\mathbb{E}(\hat{\beta}_1) = \beta_1 + \mathbb{E}\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right). \tag{14}$$

Using the Law of Iterated Expectations, we write

$$\mathbb{E}(\hat{\beta}_1) = \mathbb{E}\Big[\mathbb{E}(\hat{\beta}_1 \mid X)\Big]. \tag{15}$$

From the expression

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

it follows that

$$\mathbb{E}(\hat{\beta}_1 \mid X) = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,\mathbb{E}(u_i - \bar{u} \mid X)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{16}$$

Definition 2.1. Under the assumption of strict exogeneity,

$$\mathbb{E}(u_i \mid X) = 0 \quad \forall i. \tag{17}$$

By linearity of conditional expectation,

$$\mathbb{E}(\bar{u} \mid X) = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(u_j \mid X) = 0. \tag{18}$$

Therefore,

$$\mathbb{E}(u_i - \bar{u} \mid X) = \mathbb{E}(u_i \mid X) - \mathbb{E}(\bar{u} \mid X) = 0. \tag{19}$$

Consequently,

$$\mathbb{E}(\hat{\beta}_1) = \beta_1. \tag{20}$$

This proves that the **OLS estimator $\hat{\beta}_1$ is unbiased under the classical assumptions**.

2.2. Asymptotic Variance of the OLS Estimator

From the previous section, the OLS estimator for the slope coefficient can be written as

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{21}$$

Notice that the numerator can be equivalently rewritten. Expanding the product yields

$$\sum_{i=1}^{n}(x_i - \bar{x})(u_i - \bar{u}) = \sum_{i=1}^{n}(x_i u_i - x_i \bar{u} - \bar{x} u_i + \bar{x}\bar{u}) \tag{22}$$

$$= \sum_{i=1}^{n} x_i u_i - \bar{u}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} u_i + n\bar{x}\bar{u}. \tag{23}$$

Since $\sum_{i=1}^{n} x_i = n\bar{x}$ and $\sum_{i=1}^{n} u_i = n\bar{u}$, the expression simplifies to

$$\sum_{i=1}^{n}(x_i - \bar{x})(u_i - \bar{u}) = \sum_{i=1}^{n}(x_i - \bar{x})u_i. \tag{24}$$

Therefore, the estimator admits the equivalent representation

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{25}$$

To study the variance of $\hat{\beta}_1$, we focus on its asymptotic behavior as $n \to \infty$. By the Law of Large Numbers, under standard regularity conditions,

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \xrightarrow{p} \mathbb{E}\Big[(x_i - \mu_x)^2\Big] \equiv \sigma_x^2, \tag{26}$$

where $\mu_x = \mathbb{E}(x_i)$ (it behaves like a constant whenever $n \xrightarrow{p} \infty$), and $\sigma_x^2 = \mathrm{Var}(x_i)$.

Dividing both numerator and denominator by n, the estimator can be approximated as

$$\hat{\beta}_1 \approx \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \approx \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)u_i}{\sigma_x^2}. \tag{27}$$

Since β_1 is a constant (Var(k)=0), the variance of $\hat{\beta}_1$ is driven by the second term. Hence,

$$\mathrm{Var}(\hat{\beta}_1) \approx \mathrm{Var}\left(\frac{1}{\sigma_x^2} \cdot \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)u_i\right). \tag{28}$$

Using standard properties of the variance operator, we obtain

$$\operatorname{Var}(\hat{\beta}_1) = \frac{1}{\sigma_x^4}\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)u_i\right). \tag{29}$$

Assuming that the observations are independent and identically distributed, the variance of the sum (i.e., the inner parenthesis) satisfies

$$\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)u_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}((x_i - \mu_x)u_i). \tag{30}$$

Therefore, the asymptotic variance of the OLS estimator is given by

$$\operatorname{Var}(\hat{\beta}_1) \approx \frac{1}{n\sigma_x^4}\operatorname{Var}((x_i - \mu_x)u_i). \tag{31}$$

Under homoskedasticity and strict exogeneity, this expression further simplifies to the familiar variance formula.

3. MATRIX APPROACH TO UNBIASEDNESS OF THE OLS ESTIMATOR

To generalize the previous results and facilitate multivariate analysis, we now express the OLS estimator and its properties using matrix notation.

3.1. The Linear Regression Model in Matrix Form

Recall the simple linear regression model written for each observation:

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, \dots, n.$$

Instead of writing one equation per observation, we can stack all observations together. This leads to the matrix representation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}. \tag{32}$$

Each object in this expression has a clear interpretation:

- $\mathbf{y}$ is an $n \times 1$ vector containing all dependent variable observations.
- $\mathbf{X}$ is an $n \times 2$ matrix of regressors. Its first column is a column of ones (corresponding to the intercept), and its second column contains the values of x_i.
- $\boldsymbol{\beta} = \binom{\beta_0}{\beta_1}$ is a 2×1 vector of unknown parameters.
- $\mathbf{u}$ is an $n \times 1$ vector of error terms.

This notation does not change the model; it simply rewrites the same information in a compact form.

3.2. The OLS Estimator in Matrix Notation

As in the scalar case, the Ordinary Least Squares (OLS) method chooses the parameter vector that minimizes the sum of squared residuals. In matrix notation, the vector of residuals is defined as

$$\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}.$$

The sum of squared residuals, $\sum_{i=1}^{n} u_i^2$, can be written compactly as the scalar product

$$\mathbf{u}'\mathbf{u} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \tag{33}$$

since the inner product of a vector with itself equals the sum of the squares of its components.

The OLS problem therefore consists of minimizing this scalar expression with respect to $\boldsymbol{\beta}$.

To find the minimizer, we differentiate the objective function with respect to $\boldsymbol{\beta}$. To do so, we first expand the expression:

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}. \tag{34}$$

We now differentiate term by term:

- The term $\mathbf{y}'\mathbf{y}$ does not depend on $\boldsymbol{\beta}$ and therefore its derivative is zero.
- The term $-2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}$ is linear in $\boldsymbol{\beta}$, and its derivative is $-2\mathbf{X}'\mathbf{y}$.
- The term $\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ is a quadratic form. Since $\mathbf{X}'\mathbf{X}$ is symmetric, its derivative is $2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$.

Adding the derivatives and setting the result equal to zero yields the first-order condition

$$-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{0}. \tag{35}$$

Rearranging terms leads to the so-called *normal equations*:

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}. \tag{36}$$

Provided that the matrix $\mathbf{X}'\mathbf{X}$ is invertible, this system has a unique solution. Solving for $\boldsymbol{\beta}$ gives

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{37}$$

This expression is the OLS estimator in matrix form. It is the direct matrix analogue of the scalar estimator derived earlier and naturally extends to models with multiple regressors.

3.3. Expressing the Estimator in Terms of the Error Term

To study the statistical properties of the estimator, substitute the true model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ into the OLS formula:

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

Using basic properties of matrix multiplication, this expression can be rewritten as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \mathbf{X}'\mathbf{u}) \tag{38}$$

$$= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{u}) \tag{39}$$

$$= \mathbf{I}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}, \quad \text{since } (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}) = \mathbf{I} \tag{40}$$

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}. \tag{41}$$

This result is fundamental. It shows that the OLS estimator equals the true parameter vector plus a term that depends on the error vector.

3.4. Taking Conditional Expectations

To determine whether the estimator is unbiased, we compute its conditional expectation given the matrix of regressors $\mathbf{X}$:

$$\mathbb{E}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}(\mathbf{u} \mid \mathbf{X}). \tag{42}$$

At this point, the key assumption enters the analysis.

3.5. Strict Exogeneity in Matrix Form

In the scalar approach, strict exogeneity was stated as

$$\mathbb{E}(u_i \mid X) = 0 \quad \forall i.$$

In matrix notation, this assumption becomes

$$\mathbb{E}(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}. \tag{43}$$

That is, the entire vector of error terms has zero conditional mean given the regressors. Substituting this into the previous expression yields

$$\mathbb{E}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \boldsymbol{\beta}. \tag{44}$$

3.6. Unconditional Expectation and Unbiasedness

Finally, applying the Law of Iterated Expectations,

$$\mathbb{E}(\hat{\boldsymbol{\beta}}) = \mathbb{E}\Big[\mathbb{E}(\hat{\boldsymbol{\beta}} \mid \mathbf{X})\Big] = \boldsymbol{\beta}. \tag{45}$$

Therefore, the OLS estimator is unbiased.