ECON 351* -- NOTE 4

Statistical Properties of the OLS Coefficient Estimators

1. Introduction

We derived in *Note 2* the OLS (Ordinary Least Squares) estimators $\hat{\beta}_j$ (j = 0, 1) of the regression coefficients β_j (j = 0, 1) in the simple linear regression model given by the **population regression equation**, or PRE

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad (i = 1, \ldots, N) \qquad (1)$$

where u_i is an iid random error term. The **OLS sample regression equation** (SRE) corresponding to PRE (1) is

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i \qquad (i = 1, \ldots, N) \qquad (2)$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the **OLS coefficient estimators** given by the formulas

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \qquad (3)$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \qquad (4)$$

$x_i \equiv X_i - \overline{X}$, $y_i \equiv Y_i - \overline{Y}$, $\overline{X} = \sum_i X_i / N$, and $\overline{Y} = \sum_i Y_i / N$.

Why Use the OLS Coefficient Estimators?

The reason we use these OLS coefficient estimators is that, under assumptions A1-A8 of the classical linear regression model, they have **several desirable statistical properties**. This note examines these desirable statistical properties of the OLS coefficient estimators primarily in terms of the OLS slope coefficient estimator $\hat{\beta}_1$; the same properties apply to the intercept coefficient estimator $\hat{\beta}_0$.

2. Statistical Properties of the OLS Slope Coefficient Estimator

➢ **PROPERTY 1: Linearity of $\hat{\beta}_1$**

The OLS coefficient estimator $\hat{\beta}_1$ can be written as a **linear function of the sample values of Y, the Y_i** (i = 1, ..., N).

Proof: Starts with formula (3) for $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

$$= \frac{\sum_i x_i (Y_i - \bar{Y})}{\sum_i x_i^2}$$

$$= \frac{\sum_i x_i Y_i}{\sum_i x_i^2} - \frac{\bar{Y} \sum_i x_i}{\sum_i x_i^2}$$

$$= \frac{\sum_i x_i Y_i}{\sum_i x_i^2} \qquad \text{because } \sum_i x_i = 0.$$

- Defining the observation weights $\mathbf{k_i = x_i / \sum_i x_i^2}$ for i = 1, …, N, we can re-write the last expression above for $\hat{\beta}_1$ as:

$$\hat{\beta}_1 = \sum_i k_i Y_i \qquad \text{where } k_i \equiv \frac{x_i}{\sum_i x_i^2} \qquad (i = 1, ..., N) \qquad \textbf{... (P1)}$$

- Note that the formula (3) and the definition of the weights k_i imply that $\hat{\beta}_1$ is also a linear function of the y_i's such that

$$\hat{\beta}_1 = \sum_i k_i y_i .$$

❑ ***Result:*** The OLS slope coefficient estimator $\hat{\beta}_1$ is a ***linear*** **function of the sample values Y_i or y_i** (i = 1,…,N), where the coefficient of Y_i or y_i is k_i.

❑ Properties of the Weights k_i

In order to establish the remaining properties of $\hat{\beta}_1$, it is necessary to know the arithmetic properties of the weights k_i.

[K1] $\sum_i \mathbf{k_i} = \mathbf{0}$, i.e., the weights k_i sum to zero.

$$\sum_i k_i = \sum_i \frac{x_i}{\sum_i x_i^2} = \frac{1}{\sum_i x_i^2}\sum_i x_i = 0, \qquad \text{because } \sum_i x_i = 0.$$

[K2] $\sum_i \mathbf{k_i^2} = \dfrac{\mathbf{1}}{\sum_i \mathbf{x_i^2}}$.

$$\sum_i k_i^2 = \sum_i \left(\frac{x_i}{\sum_i x_i^2}\right)^2 = \sum_i \frac{x_i^2}{\left(\sum_i x_i^2\right)^2} = \frac{\left(\sum_i x_i^2\right)}{\left(\sum_i x_i^2\right)^2} = \frac{1}{\sum_i x_i^2}.$$

[K3] $\sum_i \mathbf{k_i x_i} = \sum_i \mathbf{k_i X_i}$.

$$\begin{aligned} \sum_i k_i x_i &= \sum_i k_i (X_i - \bar{X}) \\ &= \sum_i k_i X_i - \bar{X}\sum_i k_i \\ &= \sum_i k_i X_i \qquad \text{since } \sum_i k_i = 0 \text{ by [K1] above.} \end{aligned}$$

[K4] $\sum_i \mathbf{k_i x_i} = \mathbf{1}$.

$$\sum_i k_i x_i = \sum_i \left(\frac{x_i}{\sum_i x_i^2}\right) x_i = \sum_i \frac{x_i^2}{\left(\sum_i x_i^2\right)} = \frac{\left(\sum_i x_i^2\right)}{\left(\sum_i x_i^2\right)} = 1.$$

Implication: $\sum_i \mathbf{k_i X_i} = \mathbf{1}$.

➢ **PROPERTY 2: Unbiasedness of $\hat{\beta}_1$ and $\hat{\beta}_0$.**

The **OLS coefficient estimator $\hat{\beta}_1$ is unbiased**, meaning that $\mathbf{E(\hat{\beta}_1) = \beta_1}$.
The **OLS coefficient estimator $\hat{\beta}_0$ is unbiased**, meaning that $\mathbf{E(\hat{\beta}_0) = \beta_0}$.

- **Definition of unbiasedness:** The coefficient estimator $\hat{\beta}_1$ is unbiased if and only if $E(\hat{\beta}_1) = \beta_1$; i.e., its mean or expectation is equal to the true coefficient β_1.

▪ **Proof of unbiasedness of $\hat{\beta}_1$:** Start with the formula $\hat{\beta}_1 = \sum_i k_i Y_i$.

1. Since assumption A1 states that the PRE is $Y_i = \beta_0 + \beta_1 X_i + u_i$,

$$\begin{aligned}
\hat{\beta}_1 &= \textstyle\sum_i k_i Y_i \\
&= \textstyle\sum_i k_i(\beta_0 + \beta_1 X_i + u_i) && \text{since } Y_i = \beta_0 + \beta_1 X_i + u_i \text{ by A1} \\
&= \beta_0 \textstyle\sum_i k_i + \beta_1 \sum_i k_i X_i + \sum_i k_i u_i \\
&= \beta_1 + \textstyle\sum_i k_i u_i, && \text{since } \textstyle\sum_i k_i = 0 \text{ and } \sum_i k_i X_i = 1.
\end{aligned}$$

2. Now take expectations of the above expression for $\hat{\beta}_1$, conditional on the sample values $\{X_i: i = 1, \ldots, N\}$ of the regressor X. Conditioning on the sample values of the regressor X means that the k_i are treated as nonrandom, since the k_i are functions only of the X_i.

$$\begin{aligned}
E(\hat{\beta}_1) &= E(\beta_1) + E[\textstyle\sum_i k_i u_i] \\
&= \beta_1 + \textstyle\sum_i k_i E(u_i \mid X_i) && \text{since } \beta_1 \text{ is a constant and the } k_i \text{ are nonrandom} \\
&= \beta_1 + \textstyle\sum_i k_i 0 && \text{since } E(u_i \mid X_i) = 0 \text{ by assumption A2} \\
&= \beta_1.
\end{aligned}$$

❑ ***Result:*** The **OLS slope coefficient estimator $\hat{\beta}_1$ is an *unbiased* estimator** of the slope coefficient β_1: that is,

$$\mathbf{E(\hat{\beta}_1) = \beta_1}. \qquad \textbf{... (P2)}$$

- **Proof of unbiasedness of $\hat{\beta}_0$:** Start with the formula $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$.

  1. Average the PRE $Y_i = \beta_0 + \beta_1 X_i + u_i$ across i:

  $$\sum_{i=1}^{N} Y_i = N\beta_0 + \beta_1 \sum_{i=1}^{N} X_i + \sum_{i=1}^{N} u_i \qquad \text{(sum the PRE over the N observations)}$$

  $$\frac{\sum_{i=1}^{N} Y_i}{N} = \frac{N\beta_0}{N} + \beta_1 \frac{\sum_{i=1}^{N} X_i}{N} + \frac{\sum_{i=1}^{N} u_i}{N} \qquad \text{(divide by N)}$$

  $$\bar{Y} = \beta_0 + \beta_1\bar{X} + \bar{u} \qquad \text{where } \bar{Y} = \textstyle\sum_i Y_i / N,\ \bar{X} = \sum_i X_i / N, \text{ and } \bar{u} = \sum_i u_i / N.$$

  2. Substitute the above expression for $\bar{Y}$ into the formula $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$:

  $$\begin{aligned} \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1\bar{X} \\ &= \beta_0 + \beta_1\bar{X} + \bar{u} - \hat{\beta}_1\bar{X} \qquad \text{since } \bar{Y} = \beta_0 + \beta_1\bar{X} + \bar{u} \\ &= \beta_0 + (\beta_1 - \hat{\beta}_1)\bar{X} + \bar{u}. \end{aligned}$$

  3. Now take the expectation of $\hat{\beta}_0$ conditional on the sample values $\{X_i: i = 1, \ldots, N\}$ of the regressor X. Conditioning on the X_i means that $\bar{X}$ is treated as nonrandom in taking expectations, since $\bar{X}$ is a function only of the X_i.

  $$\begin{aligned} E(\hat{\beta}_0) &= E(\beta_0) + E\left[(\beta_1 - \hat{\beta}_1)\bar{X}\right] + E(\bar{u}) \\ &= \beta_0 + \bar{X}\,E(\beta_1 - \hat{\beta}_1) + E(\bar{u}) \qquad \text{since } \beta_0 \text{ is a constant} \\ &= \beta_0 + \bar{X}\,E(\beta_1 - \hat{\beta}_1) \qquad \text{since } E(\bar{u}) = 0 \text{ by assumptions A2 and A5} \\ &= \beta_0 + \bar{X}\left[E(\beta_1) - E(\hat{\beta}_1)\right] \\ &= \beta_0 + \bar{X}(\beta_1 - \beta_1) \qquad \text{since } E(\beta_1) = \beta_1 \text{ and } E(\hat{\beta}_1) = \beta_1 \\ &= \beta_0 \end{aligned}$$

- ***Result:*** The **OLS intercept coefficient estimator $\hat{\beta}_0$ is an *unbiased* estimator** of the intercept coefficient β_0: that is,

  $$\mathbf{E(\hat{\beta}_0) = \beta_0}. \qquad \textbf{... (P2)}$$

➢ **PROPERTY 3: Variance of $\hat{\beta}_1$.**

- **Definition:** The variance of the OLS slope coefficient estimator $\hat{\beta}_1$ is defined as

$$\text{Var}\left(\hat{\beta}_1\right) \equiv \text{E}\left\{\left[\hat{\beta}_1 - \text{E}(\hat{\beta}_1)\right]^2\right\}.$$

- **Derivation of Expression for Var($\hat{\beta}_1$):**

1. Since $\hat{\beta}_1$ is an unbiased estimator of β_1, $\text{E}(\hat{\beta}_1) = \beta_1$. The variance of $\hat{\beta}_1$ can therefore be written as

$$\text{Var}\left(\hat{\beta}_1\right) = \text{E}\left\{\left[\hat{\beta}_1 - \beta_1\right]^2\right\}.$$

2. From part (1) of the unbiasedness proofs above, the term $[\hat{\beta}_1 - \beta_1]$, which is called the **sampling error of $\hat{\beta}_1$**, is given by

$$\left[\hat{\beta}_1 - \beta_1\right] = \sum_i k_i u_i .$$

3. The square of the sampling error is therefore

$$\left[\hat{\beta}_1 - \beta_1\right]^2 = \left(\sum_i k_i u_i\right)^2$$

4. Since the square of a sum is equal to the sum of the squares plus twice the sum of the cross products,

$$\begin{aligned}\left[\hat{\beta}_1 - \beta_1\right]^2 &= \left(\sum_i k_i u_i\right)^2 \\ &= \sum_{i=1}^{N} k_i^2 u_i^2 + 2\sum_{i<s}\sum_{s=2}^{N} k_i k_s u_i u_s .\end{aligned}$$

For example, if the summation involved only three terms, the square of the sum would be

$$\left(\sum_{i=1}^{3} k_i u_i\right)^2 = \left(k_1 u_1 + k_2 u_2 + k_3 u_3\right)^2$$
$$= k_1^2 u_1^2 + k_2^2 u_2^2 + k_3^2 u_3^2 + 2k_1 k_2 u_1 u_2 + 2k_1 k_3 u_1 u_3 + 2k_2 k_3 u_2 u_3.$$

5. Now use assumptions A3 and A4 of the classical linear regression model (CLRM):

(A3) $\mathrm{Var}(u_i \mid X_i) = E(u_i^2 \mid X_i) = \sigma^2 > 0$ for all i = 1, ..., N;

(A4) $\mathrm{Cov}(u_i, u_s \mid X_i, X_s) = E(u_i u_s \mid X_i, X_s) = 0$ for all $i \neq s$.

6. We take expectations conditional on the sample values of the regressor X:

$$E\left[\left(\hat{\beta}_1 - \beta_1\right)^2\right] = \sum_{i=1}^{N} k_i^2 E(u_i^2 \mid X_i) + 2\sum_{i<s}\sum_{s=2}^{N} k_i k_s E(u_i u_s \mid X_i, X_s)$$
$$= \sum_{i=1}^{N} k_i^2 E(u_i^2 \mid X_i) \quad \text{since } E(u_i u_s \mid X_i, X_s) = 0 \text{ for } i \neq s \text{ by (A4)}$$
$$= \sum_{i=1}^{N} k_i^2 \sigma^2 \quad \text{since } E(u_i^2 \mid X_i) = \sigma^2 \ \forall i \text{ by (A3)}$$
$$= \frac{\sigma^2}{\sum_i x_i^2} \quad \text{since } \sum_i k_i^2 = \frac{1}{\sum_i x_i^2} \text{ by (K2).}$$

- ***Result:*** The ***variance*** of the OLS slope coefficient estimator $\hat{\beta}_1$ is

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i x_i^2} = \frac{\sigma^2}{\sum_i (X_i - \bar{X})^2} = \frac{\sigma^2}{TSS_X} \quad \text{where } TSS_X = \sum_i x_i^2 . \quad \textbf{... (P3)}$$

The ***standard error*** of $\hat{\beta}_1$ is the square root of the variance: i.e.,

$$se(\hat{\beta}_1) = \sqrt{\mathrm{Var}(\hat{\beta}_1)} = \left(\frac{\sigma^2}{\sum_i x_i^2}\right)^{\frac{1}{2}} = \frac{\sigma}{\sqrt{\sum_i x_i^2}} = \frac{\sigma}{\sqrt{TSS_X}} .$$

➢ **<u>PROPERTY 4</u>: Variance of $\hat{\beta}_0$ (given without proof).**

❑ ***<u>Result</u>:*** The ***variance*** of the OLS intercept coefficient estimator $\hat{\beta}_0$ is

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_i X_i^2}{N \sum_i x_i^2} = \frac{\sigma^2 \sum_i X_i^2}{N \sum_i (X_i - \bar{X})^2}. \qquad \textbf{... (P4)}$$

The ***standard error*** of $\hat{\beta}_0$ is the square root of the variance: i.e.,

$$\text{se}(\hat{\beta}_0) = \sqrt{\text{Var}(\hat{\beta}_0)} = \left(\frac{\sigma^2 \sum_i X_i^2}{N \sum_i x_i^2} \right)^{\frac{1}{2}}.$$

- **<u>Interpretation of the Coefficient Estimator Variances</u>**

  - $\text{Var}(\hat{\beta}_0)$ and $\text{Var}(\hat{\beta}_1)$ measure the **statistical precision** of the OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i x_i^2}; \qquad \text{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_i X_i^2}{N \sum_i x_i^2}.$$

  - **Determinants of $\text{Var}(\hat{\beta}_0)$ and $\text{Var}(\hat{\beta}_1)$**

$\text{Var}(\hat{\beta}_0)$ and $\text{Var}(\hat{\beta}_1)$ are ***smaller***:

(1) the ***smaller*** is the **error variance σ^2**, i.e., the smaller the variance of the unobserved and unknown random influences on Y_i;

(2) the ***larger*** is the **sample variation of the X_i** about their sample mean, i.e., the larger the values of $x_i^2 = (X_i - \bar{X})^2$, i = 1, …, N;

(3) the ***larger*** is the **size of the sample**, i.e., the ***larger* is N**.

➢ **<u>PROPERTY 5</u>: Covariance of $\hat{\beta}_0$ and $\hat{\beta}_1$.**

- **Definition:** The covariance of the OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ is defined as

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \equiv E\{[\hat{\beta}_0 - E(\hat{\beta}_0)][\hat{\beta}_1 - E(\hat{\beta}_1)]\}.$$

- **Derivation of Expression for Cov($\hat{\beta}_0, \hat{\beta}_1$):**

1. Since $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}$, the expectation of $\hat{\beta}_0$ can be written as

$$\begin{aligned} E(\hat{\beta}_0) &= \overline{Y} - E(\hat{\beta}_1)\overline{X} \\ &= \overline{Y} - \beta_1\overline{X} \qquad \text{since } E(\hat{\beta}_1) = \beta_1. \end{aligned}$$

Therefore, the term $\hat{\beta}_0 - E(\hat{\beta}_0)$ can be written as

$$\begin{aligned} \hat{\beta}_0 - E(\hat{\beta}_0) &= [\overline{Y} - \hat{\beta}_1\overline{X}] - [\overline{Y} - \beta_1\overline{X}] \\ &= \overline{Y} - \hat{\beta}_1\overline{X} - \overline{Y} + \beta_1\overline{X} \\ &= -\hat{\beta}_1\overline{X} + \beta_1\overline{X} \\ &= -\overline{X}(\hat{\beta}_1 - \beta_1). \end{aligned}$$

2. Since $E(\hat{\beta}_1) = \beta_1$, the term $\hat{\beta}_1 - E(\hat{\beta}_1)$ takes the form

$$\hat{\beta}_1 - E(\hat{\beta}_1) = \hat{\beta}_1 - \beta_1.$$

3. The product $[\hat{\beta}_0 - E(\hat{\beta}_0)][\hat{\beta}_1 - E(\hat{\beta}_1)]$ thus takes the form

$$\begin{aligned} [\hat{\beta}_0 - E(\hat{\beta}_0)][\hat{\beta}_1 - E(\hat{\beta}_1)] &= -\overline{X}(\hat{\beta}_1 - \beta_1)(\hat{\beta}_1 - \beta_1) \\ &= -\overline{X}(\hat{\beta}_1 - \beta_1)^2. \end{aligned}$$

4. The expectation of the product $[\hat{\beta}_0 - E(\hat{\beta}_0)][\hat{\beta}_1 - E(\hat{\beta}_1)]$ is therefore

$$
\begin{aligned}
E\left\{[\hat{\beta}_0 - E(\hat{\beta}_0)][\hat{\beta}_1 - E(\hat{\beta}_1)]\right\} &= E\left[-\overline{X}(\hat{\beta}_1 - \beta_1)^2\right] \\
&= -\overline{X}\,E(\hat{\beta}_1 - \beta_1)^2 \qquad \text{b/c } \overline{X} \text{ is a constant} \\
&= -\overline{X}\,\text{Var}(\hat{\beta}_1) \qquad \text{b/c } E(\hat{\beta}_1 - \beta_1)^2 = \text{Var}(\hat{\beta}_1) \\
&= -\overline{X}\left(\frac{\sigma^2}{\sum_i x_i^2}\right) \qquad \text{b/c } \text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i x_i^2}.
\end{aligned}
$$

❑ ***<u>Result</u>:*** The covariance of $\hat{\beta}_0$ and $\hat{\beta}_1$ is

$$\text{Cov}(\hat{\beta}_0, \beta_1) = -\overline{X}\left(\frac{\sigma^2}{\sum_i x_i^2}\right). \qquad \textbf{... (P5)}$$

- **Interpretation of the Covariance Cov($\hat{\beta}_0, \hat{\beta}_1$).**

Since both σ^2 and $\sum_i x_i^2$ are positive, the sign of Cov($\hat{\beta}_0$, $\hat{\beta}_1$) depends on the sign of $\overline{X}$.

(1) If $\overline{X} > 0$, Cov($\hat{\beta}_0$, $\hat{\beta}_1$) < 0: the sampling errors $(\hat{\beta}_0 - \beta_0)$ and $(\hat{\beta}_1 - \beta_1)$ are of ***opposite* sign**.

(2) If $\overline{X} < 0$, Cov($\hat{\beta}_0$, $\hat{\beta}_1$) > 0: the sampling errors $(\hat{\beta}_0 - \beta_0)$ and $(\hat{\beta}_1 - \beta_1)$ are of the ***same* sign**.

➢ THE GAUSS-MARKOV THEOREM

- **Importance of the Gauss-Markov Theorem:**

  1) The Gauss-Markov Theorem summarizes the statistical properties of the OLS coefficient estimators $\hat{\beta}_j$ (j = 0, 1).

  2) More specifically, it establishes that **the OLS coefficient estimators $\hat{\beta}_j$ (j = 0, 1) have several desirable statistical properties**.

- **Statement of the Gauss-Markov Theorem:** Under assumptions A1-A8 of the CLRM, the OLS coefficient estimators $\hat{\beta}_j$ (j = 0, 1) are the ***minimum variance* estimators** of the regression coefficients β_j (j = 0, 1) in the class of ***all linear unbiased estimators*** of β_j.

  That is, under assumptions A1-A8, the **OLS coefficient estimators $\hat{\beta}_j$** are **the BLUE** of β_j (j = 0, 1) in the class of all linear unbiased estimators, where

  1) **BLUE** ≡ **B**est **L**inear **U**nbiased **E**stimator

  2) "Best" means "minimum variance" or "smallest variance".

  So the Gauss-Markov Theorem says that the OLS coefficient estimators $\hat{\beta}_j$ are the best of all linear unbiased estimators of β_j, where "best" means "minimum variance".

- **<u>Interpretation of the G-M Theorem</u>:**

1. Let $\tilde{\beta}_j$ be ***any other linear unbiased estimator*** of β_j.

Let $\hat{\beta}_j$ be the ***OLS estimator*** of β_j; it too is linear and unbiased.

2. Both estimators $\tilde{\beta}_j$ and $\hat{\beta}_j$ are ***unbiased*** **estimators** of β_j:

$$E(\hat{\beta}_j) = \beta_j \quad \text{and} \quad E(\tilde{\beta}_j) = \beta_j .$$

3. But the OLS estimator $\hat{\beta}_j$ has a ***smaller*** **variance** than $\tilde{\beta}_j$:

$$Var(\hat{\beta}_j) \le Var(\tilde{\beta}_j) \quad \Rightarrow \quad \hat{\beta}_j \text{ is } \textit{efficient} \text{ relative to } \tilde{\beta}_j .$$

This means that the OLS estimator $\hat{\beta}_j$ is **statistically *more precise*** than $\tilde{\beta}_j$, ***any other linear unbiased*** **estimator of β_j.**