# Lecture 5: Gauss-Markov Theorem

Economics 326 — Introduction to Econometrics II

Vadim Marmer, UBC

April 5, 2026

## What we have so far

- **Simple linear regression model**:

$$\begin{aligned} &Y_i = \alpha + \beta X_i + U_i, \\ &\mathrm{E}\left[U_i \mid \mathbf{X}\right] = 0, \\ &\mathrm{Var}\left(U_i \mid \mathbf{X}\right) = \sigma^2, \\ &\mathrm{Cov}\left(U_i, U_j \mid \mathbf{X}\right) = 0 \text{ for all } i \neq j. \end{aligned}$$

- **OLS estimator**:

$$\hat{\beta} = \frac{\sum_{i=1}^n (X_i - \bar{X})Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

- **Unbiasedness**: $\mathrm{E}\left[\hat{\beta} \mid \mathbf{X}\right] = \beta$.
- **Variance**:

$$\mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

- **Question**: Is OLS the best we can do?

## There are many alternative estimators

- The OLS estimator is not the only estimator we can construct.
- There are alternative estimators with some desirable properties.

## Example of an alternative estimator

- Using only the first two observations, suppose that $X_2 \neq X_1$.

$$\tilde{\beta} = \frac{Y_2 - Y_1}{X_2 - X_1}.$$

- $\tilde{\beta}$ is **linear**:

$$\tilde{\beta} = c_1 Y_1 + c_2 Y_2,$$

where

$$c_1 = -\frac{1}{X_2 - X_1} \text{ and } c_2 = \frac{1}{X_2 - X_1}.$$

- Signal/noise decomposition of $\tilde{\beta}$:

$$\begin{aligned}\tilde{\beta} &= \frac{Y_2 - Y_1}{X_2 - X_1} \\ &= \frac{(\alpha + \beta X_2 + U_2) - (\alpha + \beta X_1 + U_1)}{X_2 - X_1} \\ &= \frac{\beta(X_2 - X_1)}{X_2 - X_1} + \frac{U_2 - U_1}{X_2 - X_1} \\ &= \beta + \frac{U_2 - U_1}{X_2 - X_1}.\end{aligned}$$

- Unbiasedness of $\tilde{\beta}$:

$$\begin{aligned}\mathrm{E}\left[\tilde{\beta} \mid \mathbf{X}\right] &= \beta + \mathrm{E}\left[\frac{U_2 - U_1}{X_2 - X_1} \mid \mathbf{X}\right] \\ &= \beta + \frac{\mathrm{E}\left[U_2 \mid \mathbf{X}\right] - \mathrm{E}\left[U_1 \mid \mathbf{X}\right]}{X_2 - X_1} \\ &= \beta.\end{aligned}$$

## Optimality of an estimator: BLUE

- Among all **linear** and **unbiased** estimators, an estimator with the **smallest variance** is called the **Best Linear Unbiased Estimator** (**BLUE**).

## Gauss-Markov Theorem

Suppose that

- $Y_i = \alpha + \beta X_i + U_i$.
- $\mathrm{E}\left[U_i \mid \mathbf{X}\right] = 0$.
- $\mathrm{E}\left[U_i^2 \mid \mathbf{X}\right] = \sigma^2$ for all $i = 1, \ldots, n$ (homoskedasticity).
- For all $i \neq j$, $\mathrm{E}\left[U_i U_j \mid \mathbf{X}\right] = 0$.
- Then the OLS estimator is **BLUE**.

## Gauss-Markov Theorem (setup)

- We already know that the OLS estimator $\hat{\beta}$ is linear and unbiased.
- Let $\tilde{\beta}$ be any other estimator of $\beta$ such that
  - $\tilde{\beta}$ is linear:

    $$\tilde{\beta} = \sum_{i=1}^{n} c_i Y_i,$$

    where $c$'s depend only on $\mathbf{X}$.
  - $\tilde{\beta}$ is unbiased: $\mathrm{E}\left[\tilde{\beta} \mid \mathbf{X}\right] = \beta$.
- We need to show that for **any** such $\tilde{\beta} \neq \hat{\beta}$,

$$\mathrm{Var}\left(\tilde{\beta} \mid \mathbf{X}\right) > \mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right).$$

## An outline of the proof

1. Show that the $c$'s in $\tilde{\beta} = \sum_{i=1}^{n} c_i Y_i$ satisfy $\sum_{i=1}^{n} c_i = 0$ and $\sum_{i=1}^{n} c_i X_i = 1$.
2. Using the results of Step 1, show that $\mathrm{Cov}\left(\tilde{\beta}, \hat{\beta} \mid \mathbf{X}\right) = \mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right)$.

3. Using the results of Step 2, show that $\mathrm{Var}\left(\tilde{\beta} \mid \mathbf{X}\right) \geq \mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right)$.
4. Show that $\mathrm{Var}\left(\tilde{\beta} \mid \mathbf{X}\right) = \mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right)$ if and only if $\tilde{\beta} = \hat{\beta}$.

## Proof: Step 1

- Since $\tilde{\beta} = \sum_{i=1}^n c_i Y_i$ (linearity),

$$\begin{aligned}\tilde{\beta} &= \sum_{i=1}^n c_i(\alpha + \beta X_i + U_i) \\ &= \alpha \sum_{i=1}^n c_i + \beta \sum_{i=1}^n c_i X_i + \sum_{i=1}^n c_i U_i.\end{aligned}$$

- Therefore,

$$\begin{aligned}\mathrm{E}\left[\tilde{\beta} \mid \mathbf{X}\right] &= \mathrm{E}\left[\alpha \sum_{i=1}^n c_i + \beta \sum_{i=1}^n c_i X_i + \sum_{i=1}^n c_i U_i \mid \mathbf{X}\right] \\ &= \alpha \sum_{i=1}^n c_i + \beta \sum_{i=1}^n c_i X_i + \sum_{i=1}^n c_i \mathrm{E}\left[U_i \mid \mathbf{X}\right] \\ &= \alpha \sum_{i=1}^n c_i + \beta \sum_{i=1}^n c_i X_i.\end{aligned}$$

- From the **linearity** we have that

$$\mathrm{E}\left[\tilde{\beta} \mid \mathbf{X}\right] = \alpha \sum_{i=1}^n c_i + \beta \sum_{i=1}^n c_i X_i.$$

- From the **unbiasedness** we have that

$$\beta = \mathrm{E}\left[\tilde{\beta} \mid \mathbf{X}\right] = \alpha \sum_{i=1}^n c_i + \beta \sum_{i=1}^n c_i X_i.$$

- Since this has to be true for **any** $\alpha$, $\beta$, and the $X$'s, it follows now that

$$\sum_{i=1}^n c_i = 0, \quad \sum_{i=1}^n c_i X_i = 1.$$

## Proof: Step 2

- We have

$$\tilde{\beta} = \beta + \sum_{i=1}^n c_i U_i, \text{ with } \sum_{i=1}^n c_i = 0, \sum_{i=1}^n c_i X_i = 1.$$

$$\hat{\beta} = \beta + \sum_{i=1}^n w_i U_i, \text{ with } w_i = \frac{X_i - \bar{X}}{\sum_{j=1}^n (X_j - \bar{X})^2}.$$

- Then,

$$\begin{aligned}\mathrm{Cov}\left(\tilde{\beta}, \hat{\beta} \mid \mathbf{X}\right) &= \mathrm{E}\left[(\tilde{\beta} - \beta)(\hat{\beta} - \beta) \mid \mathbf{X}\right] \\ &= \mathrm{E}\left[\left(\sum_{i=1}^n c_i U_i\right)\left(\sum_{i=1}^n w_i U_i\right) \mid \mathbf{X}\right] \\ &= \sum_{i=1}^n c_i w_i \mathrm{E}\left[U_i^2 \mid \mathbf{X}\right] + \sum_{i=1}^n \sum_{j \neq i} c_i w_j \mathrm{E}\left[U_i U_j \mid \mathbf{X}\right].\end{aligned}$$

- Since $\mathrm{E}\left[U_i^2 \mid \mathbf{X}\right] = \sigma^2$ for all $i$'s:

$$\sum_{i=1}^{n} c_i w_i \mathrm{E}\left[U_i^2 \mid \mathbf{X}\right] = \sigma^2 \sum_{i=1}^{n} c_i w_i.$$

- Since $\mathrm{E}\left[U_i U_j \mid \mathbf{X}\right] = 0$ for all $i \neq j$,

$$\sum_{i=1}^{n} \sum_{j \neq i} c_i w_j \mathrm{E}\left[U_i U_j \mid \mathbf{X}\right] = 0.$$

- Thus,

$$\mathrm{Cov}\left(\tilde{\beta}, \hat{\beta} \mid \mathbf{X}\right) = \sigma^2 \sum_{i=1}^{n} c_i w_i.$$

and

$$w_i = \frac{X_i - \bar{X}}{\sum_{j=1}^{n} (X_j - \bar{X})^2}.$$

- We have:

$$\begin{aligned}
\mathrm{Cov}\left(\tilde{\beta}, \hat{\beta} \mid \mathbf{X}\right) &= \sigma^2 \sum_{i=1}^{n} c_i \frac{X_i - \bar{X}}{\sum_{j=1}^{n} (X_j - \bar{X})^2} \\
&= \frac{\sigma^2}{\sum_{j=1}^{n} (X_j - \bar{X})^2} \sum_{i=1}^{n} c_i (X_i - \bar{X}) \\
&= \frac{\sigma^2}{\sum_{j=1}^{n} (X_j - \bar{X})^2} \left( \sum_{i=1}^{n} c_i X_i - \bar{X} \sum_{i=1}^{n} c_i \right) \\
&= \frac{\sigma^2}{\sum_{j=1}^{n} (X_j - \bar{X})^2} (1 - \bar{X} \cdot 0) \\
&= \mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right).
\end{aligned}$$

## Proof: Step 3

- We know now that for any **linear** and **unbiased** $\tilde{\beta}$,

$$\mathrm{Cov}\left(\tilde{\beta}, \hat{\beta} \mid \mathbf{X}\right) = \mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right).$$

- Let's consider $\mathrm{Var}\left(\tilde{\beta} - \hat{\beta} \mid \mathbf{X}\right)$:

$$\begin{aligned}
\mathrm{Var}\left(\tilde{\beta} - \hat{\beta} \mid \mathbf{X}\right) &= \mathrm{Var}\left(\tilde{\beta} \mid \mathbf{X}\right) + \mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right) - 2\mathrm{Cov}\left(\tilde{\beta}, \hat{\beta} \mid \mathbf{X}\right) \\
&= \mathrm{Var}\left(\tilde{\beta} \mid \mathbf{X}\right) + \mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right) - 2\mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right) \\
&= \mathrm{Var}\left(\tilde{\beta} \mid \mathbf{X}\right) - \mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right).
\end{aligned}$$

- But since $\mathrm{Var}\left(\tilde{\beta} - \hat{\beta} \mid \mathbf{X}\right) \geq 0$,

$$\mathrm{Var}\left(\tilde{\beta} \mid \mathbf{X}\right) - \mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right) \geq 0$$

or

$$\mathrm{Var}\left(\tilde{\beta} \mid \mathbf{X}\right) \geq \mathrm{Var}\left(\hat{\beta} \mid \mathbf{X}\right).$$

## Proof: Step 4 (Uniqueness)

Suppose that $\text{Var}\left(\tilde{\beta} \mid \mathbf{X}\right) = \text{Var}\left(\hat{\beta} \mid \mathbf{X}\right)$.

- Then,

$$\text{Var}\left(\tilde{\beta} - \hat{\beta} \mid \mathbf{X}\right) = \text{Var}\left(\tilde{\beta} \mid \mathbf{X}\right) - \text{Var}\left(\hat{\beta} \mid \mathbf{X}\right) = 0.$$

- Thus, $\tilde{\beta} - \hat{\beta}$ is not random or

$$\tilde{\beta} - \hat{\beta} = \text{constant}.$$

- This constant also has to be zero because

$$\begin{aligned} \text{E}\left[\tilde{\beta} \mid \mathbf{X}\right] &= \text{E}\left[\hat{\beta} \mid \mathbf{X}\right] + \text{constant} \\ &= \beta + \text{constant}, \end{aligned}$$

  and in order for $\tilde{\beta}$ to be **unbiased**

$$\text{constant} = 0 \text{ or } \tilde{\beta} = \hat{\beta}.$$

## Why does unbiasedness matter?

- The Gauss-Markov Theorem says that OLS is the best among **linear unbiased** estimators. Can we do better if we drop unbiasedness?
- Consider $\tilde{\beta} = 0$ (always "estimate" the slope as zero, regardless of the data).
- $\tilde{\beta} = 0$ is **linear**: $\tilde{\beta} = \sum_{i=1}^{n} 0 \cdot Y_i = 0$.
- $\tilde{\beta} = 0$ is **biased**: $\text{E}\left[\tilde{\beta} \mid \mathbf{X}\right] = 0 \neq \beta$ (unless $\beta = 0$).
- $\text{Var}\left(\tilde{\beta} \mid \mathbf{X}\right) = 0 < \text{Var}\left(\hat{\beta} \mid \mathbf{X}\right)$.
- So $\tilde{\beta} = 0$ has a **smaller variance** than OLS, but it is useless as an estimator because it ignores the data entirely.
- This illustrates why the **unbiasedness** requirement in the Gauss-Markov Theorem is essential: without it, one can trivially achieve zero variance by using a constant.

## What if homoskedasticity fails?

- Suppose that

$$\text{E}\left[U_i^2 \mid \mathbf{X}\right] = \sigma^2(X_i) = \sigma_i^2,$$

  where $\sigma_i^2$ may differ across observations (**heteroskedasticity**).
- **Unbiasedness** of the OLS estimator $\hat{\beta}$ **still holds**. The proof of unbiasedness only uses:
  1. $Y_i = \alpha + \beta X_i + U_i$.
  2. $\text{E}\left[U_i \mid \mathbf{X}\right] = 0$.
  - It does **not** require homoskedasticity.
- However, the OLS estimator is **no longer BLUE**.
- The Gauss-Markov proof relied on $\text{E}\left[U_i^2 \mid \mathbf{X}\right] = \sigma^2$ being the **same** for all $i$'s.
- There exists another linear unbiased estimator with a **smaller variance** than OLS.

## Transforming the model

- Suppose we know $\sigma_i^2 = \mathrm{Var}(U_i \mid \mathbf{X})$ for each $i$. Divide both sides of

$$Y_i = \alpha + \beta X_i + U_i$$

by $\sigma_i$:

- We have:

$$\frac{Y_i}{\sigma_i} = \alpha \cdot \frac{1}{\sigma_i} + \beta \cdot \frac{X_i}{\sigma_i} + \frac{U_i}{\sigma_i}.$$

- Define the **transformed variables**:

$$Y_i^* = \frac{Y_i}{\sigma_i}, \quad X_{0i}^* = \frac{1}{\sigma_i}, \quad X_{1i}^* = \frac{X_i}{\sigma_i}, \quad U_i^* = \frac{U_i}{\sigma_i}.$$

- The transformed model is:

$$Y_i^* = \alpha\, X_{0i}^* + \beta\, X_{1i}^* + U_i^*.$$

- The transformed errors $U_i^*$ satisfy:
  - $\mathrm{E}[U_i^* \mid \mathbf{X}] = \frac{1}{\sigma_i}\mathrm{E}[U_i \mid \mathbf{X}] = 0$.
  - $\mathrm{Var}(U_i^* \mid \mathbf{X}) = \frac{1}{\sigma_i^2}\mathrm{Var}(U_i \mid \mathbf{X}) = \frac{\sigma_i^2}{\sigma_i^2} = 1$.
- The transformed errors are **homoskedastic**: $\mathrm{Var}(U_i^* \mid \mathbf{X}) = 1$ for all $i$'s.
- The Gauss-Markov assumptions hold for the transformed model, so OLS applied to the **transformed** data is **BLUE**.

## Weighted Least Squares (WLS)

- OLS applied to the transformed model minimizes (w.r.t. $a$ and $b$):

$$\sum_{i=1}^{n} \frac{(Y_i - a - bX_i)^2}{\sigma_i^2} = \sum_{i=1}^{n} w_i (Y_i - a - bX_i)^2,$$

where $w_i = 1/\sigma_i^2$.

- This is called **Weighted Least Squares** (WLS).
- **Interpretation**: WLS gives **more weight** to observations with smaller variance (more precise observations) and **less weight** to observations with larger variance (noisier observations).
- The WLS estimator is BLUE for $\beta$ in the original model.
- The WLS estimator of $\beta$ is:

$$\hat{\beta}_{WLS} = \frac{\sum_{i=1}^{n} w_i (X_i - \bar{X}_w)(Y_i - \bar{Y}_w)}{\sum_{i=1}^{n} w_i (X_i - \bar{X}_w)^2},$$

where $w_i = 1/\sigma_i^2$, $\bar{X}_w = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$, and $\bar{Y}_w = \frac{\sum_{i=1}^{n} w_i Y_i}{\sum_{i=1}^{n} w_i}$.

- **Practical limitation**: WLS requires knowledge of $\sigma_i^2$, which is typically unknown.