Econ 103: Introduction to Econometrics

Lecture 7 — Properties of OLS & the Gauss–Markov Theorem

Ryan Longmuir

UCLA

Summer Session C, 2026

Reading: Hill, Griffiths & Lim (5th ed.), §2.4–2.6; Stock & Watson (4th ed.), §4.5, 5.5.

Where we are

Last lecture we fit the line: $b_2 = \widehat{\text{Cov}}(x, y)/\widehat{\text{Var}}(x)$, $b_1 = \bar{y} - b_2\bar{x}$, giving $\widehat{\text{FOOD\_EXP}} = 83.42 + 10.21\,\text{INCOME}$.

The natural question: how good is this estimate? HGL's answer is blunt: that question is *unanswerable*. We will never know β_2, so we can't say how close 10.21 is to it.

So we change the question. Instead of judging one estimate, we judge the **estimation procedure** — the random variables b_1, b_2. Today:

- OLS is unbiased: on average across samples, it nails β,
- its variance (precision) and what drives it,
- and Gauss–Markov: OLS is **BLUE**.

Today's plan

1. The estimator as a random variable
2. Unbiasedness
3. Variance and precision
4. The Gauss–Markov theorem

Part 1

The estimator as a random variable

Sampling variation is real

Draw another 40 households (same incomes, new families) and you get *different* estimates — because each y_i is random. HGL report 10 hypothetical samples from the *same* population:

sample	b_1	b_2
1	93.64	8.24
2	91.62	8.90
3	126.76	6.59
4	55.98	11.23
⋮	⋮	⋮
10	128.55	6.99
avg	96.11	8.70

- b_2 ranges from 6.59 to 11.23 — the same procedure, different numbers each time.
- This **sampling variation** is unavoidable: b_1, b_2 are random variables with a distribution.
- Their average (8.70) sits near the truth — a hint of unbiasedness.

We can't see this spread with one sample. So we study it *theoretically*.

Step 1: OLS is a linear estimator

To study b_2's distribution, rewrite it (HGL Appendix 2C). Starting from the deviation form and using $\sum(x_i - \bar{x})\bar{y} = 0$,

$$b_2 = \sum_{i=1}^{N} w_i\, y_i, \qquad w_i = \frac{x_i - \bar{x}}{\sum_j (x_j - \bar{x})^2}.$$

- The weights w_i depend only on x. Conditioning on x, they are constants.
- Any estimator that is a weighted average of the y_i is a **linear estimator**. OLS is one. (Remember this for Gauss–Markov.)

Useful facts about the weights: $\sum w_i = 0$ and $\sum w_i x_i = 1$. They do the bookkeeping in every proof below.

Step 2: the key decomposition

Substitute the model $y_i = \beta_1 + \beta_2 x_i + e_i$ into $b_2 = \sum w_i y_i$ and use $\sum w_i = 0$, $\sum w_i x_i = 1$:

$$\boxed{b_2 = \beta_2 + \sum_{i=1}^{N} w_i \, e_i}$$

This is the workhorse equation of the lecture. It splits the estimator into

$$\underbrace{\beta_2}_{\text{what we want (fixed)}} \quad + \quad \underbrace{\sum w_i e_i}_{\text{estimation error (random)}} \quad .$$

Everything random about b_2 lives in $\sum w_i e_i$. Its mean controls bias; its variance controls precision. We take them in turn.

Part 2

Unbiasedness

OLS is unbiased

Take the conditional expectation of $b_2 = \beta_2 + \sum w_i e_i$ given x:

$$\begin{aligned}\mathbb{E}(b_2 \mid x) &= \beta_2 + \sum w_i \, \mathbb{E}(e_i \mid x) \\ &= \beta_2 + \sum w_i \cdot 0 = \beta_2.\end{aligned}$$

Two ingredients: w_i is constant given x (so it pulls out), and SR2: $\mathbb{E}(e_i \mid x) = 0$.

Unbiasedness

An estimator is **unbiased** if its expected value equals the parameter. Under SR1–SR5,

$$\mathbb{E}(b_2 \mid x) = \beta_2 \qquad \text{and} \qquad \mathbb{E}(b_1 \mid x) = \beta_1.$$

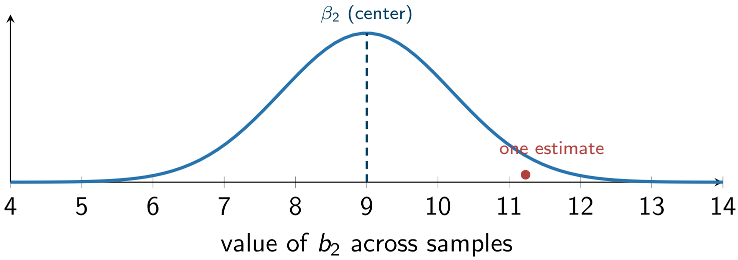
What unbiasedness does and does not say

Does say

Over all possible samples, the estimates average out to the true β_2. The procedure is centered on the target — no systematic over- or under-shooting.

Does *not* say

That *your* estimate 10.21 is close to β_2. One draw can land far from center. Unbiasedness is a property of the estimator, never of a single estimate.



When unbiasedness fails: omitted variables

The proof leaned entirely on SR2: $\mathbb{E}(e_i \mid x) = 0$. If that breaks, so does unbiasedness:

$$\mathbb{E}(e_i \mid x) \neq 0 \implies \mathbb{E}(b_2 \mid x) = \beta_2 + \sum w_i \, \mathbb{E}(e_i \mid x) \neq \beta_2.$$

Example (Omitting ability from a wage equation)

$WAGE = \beta_1 + \beta_2 \, EDUC + e$, with ability buried in e. Ability is correlated with education, so $\mathbb{E}(e \mid EDUC) \neq 0$, b_2 is biased, and it confounds the return to schooling with the payoff to ability.

This is **omitted-variable bias** — the formal face of "correlation $\neq$ causation." We quantify it in Lecture 18.

Part 3

Variance and precision

The variances and covariance of OLS

Unbiased is not enough — we want estimates that are tightly clustered around β. That is the **variance**. Under SR1–SR5 (HGL Appendix 2E):

$$\mathrm{Var}(b_2 \mid x) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}, \qquad \mathrm{Var}(b_1 \mid x) = \sigma^2 \left[\frac{\sum x_i^2}{N \sum (x_i - \bar{x})^2} \right],$$

$$\mathrm{Cov}(b_1, b_2 \mid x) = \sigma^2 \left[\frac{-\bar{x}}{\sum (x_i - \bar{x})^2} \right].$$

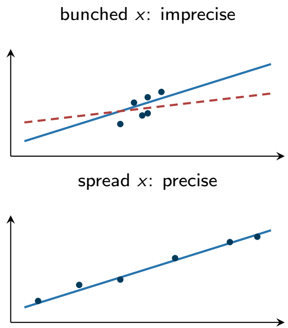
Smaller variance = more precise

Two unbiased estimators with the same center: prefer the one with the smaller variance — it has a higher chance of landing near β_2. We will care a lot about $\mathrm{Var}(b_2 \mid x)$.

What drives the precision of b_2?

Look hard at $\mathrm{Var}(b_2 \mid x) = \dfrac{\sigma^2}{\sum(x_i - \bar{x})^2}$. Three levers:

1. Error variance σ^2 (numerator). Noisier data about the line $\Rightarrow$ less precise. Can't control it.
2. Spread of x, $\sum(x_i - \bar{x})^2$ (denominator). More variation in income $\Rightarrow$ more precise. A wide lever arm pins the slope.
3. Sample size N (more terms in the sum) $\Rightarrow$ the denominator grows $\Rightarrow$ more precise. More data, tighter estimates.



From variance to standard error

The variances above contain the unknown σ^2. Replace it with an estimate $\hat{\sigma}^2$ (Lecture 8) to get an *estimated* variance; its square root is the **standard error**:

$$\mathrm{se}(b_2) = \sqrt{\widehat{\mathrm{Var}}(b_2 \mid x)} = \sqrt{\frac{\hat{\sigma}^2}{\sum(x_i - \bar{x})^2}}.$$

- The standard error is our estimate of the sampling standard deviation of b_2 — "how far, typically, would b_2 land from β_2 across samples?"
- For the food data, software reports $\mathrm{se}(b_1) = 43.41$, $\mathrm{se}(b_2) = 2.09$.
- Standard errors are the raw material of every confidence interval and t-test to come (Lectures 9–10).

We need $\hat{\sigma}^2$ to actually compute these — that is the first order of business next lecture.

The shape of the distribution

Center (β_2) and spread ($\mathrm{Var}(b_2 \mid x)$) describe b_2. What about its *shape*?

If SR6 holds (normal errors)

$b_2 = \sum w_i y_i$ is a weighted sum of normals, hence exactly normal:

$$b_2 \mid x \sim N\left(\beta_2,\ \frac{\sigma^2}{\sum (x_i - \bar{x})^2}\right).$$

If SR6 fails (a CLT)

Even non-normal errors give an approximately normal b_2 in *large* samples — a central limit theorem, since b_2 is (a kind of) average.

Either way, b_2 ends up normal (exactly, or in large samples). That normal shape is what makes the t-based inference of Lectures 9–10 possible.

Part 4

The Gauss–Markov theorem

OLS is BLUE

We now know OLS is linear ($b_2 = \sum w_i y_i$) and unbiased ($\mathbb{E}(b_2 \mid x) = \beta_2$). Is it the *best* such estimator?

Gauss–Markov Theorem

Given x and under assumptions SR1–SR5, the OLS estimators b_1, b_2 have the smallest variance among all **linear and unbiased** estimators of β_1, β_2. They are the **Best Linear Unbiased Estimators** (BLUE).

In words: within the class of estimators that are (i) weighted averages of the y_i and (ii) correct on average, *nothing beats OLS on precision*. There is no point searching for a cleverer linear unbiased rule — you already have the winner.

Reading the fine print

Gauss–Markov is precise about what it promises:

1. "Best" is only within a class — linear *and* unbiased. A nonlinear or biased estimator could, in principle, have smaller variance.
2. "Best" means minimum variance among that class.
3. It requires SR1–SR5. Break one (e.g. heteroskedasticity breaks SR3) and OLS need no longer be best.
4. It does not need normality (SR6). Pure first- and second-moment result.
5. It applies to the estimators (the procedure), not to the estimates from one sample.

S&W call the same idea the Gauss–Markov theorem too (§5.5); their efficiency statement is the textbook's reason OLS is the default estimator everywhere.

What each assumption buys you

A scorecard of which property leans on which assumption:

Property	**Needs**	**Why**
b_1, b_2 exist / are computable	SR1, SR5	line defined; $\sum(x_i - \bar{x})^2 \neq 0$
Unbiased: $\mathbb{E}(b \mid x) = \beta$	SR1, SR2	$\mathbb{E}(e \mid x) = 0$ kills the bias term
Variance formulas as stated	SR1–SR4	SR3 (homosk.), SR4 (uncorrelated)
BLUE (Gauss–Markov)	SR1–SR5	minimum variance in class
Exact normal b, exact inference	$+$ SR6	normal errors $\Rightarrow$ normal b

Much of the rest of econometrics is a tour of which assumption is failing and what to do about it — robust standard errors (SR3), clustering (SR4), instrumental variables (SR2), and so on.

Recap

The estimator is random

- $b_2 = \sum w_i y_i$ (linear), $w_i = \dfrac{x_i - \bar{x}}{\sum(x_j - \bar{x})^2}$
- $b_2 = \beta_2 + \sum w_i e_i$

Unbiased

- $\mathbb{E}(b_2 \mid x) = \beta_2$ (needs SR2)
- about the procedure, not one estimate
- fails under omitted variables

Precision

- $\mathrm{Var}(b_2 \mid x) = \dfrac{\sigma^2}{\sum(x_i - \bar{x})^2}$
- smaller with: low σ^2, spread-out x, big N
- $\mathrm{se}(b_2) = \sqrt{\hat{\sigma}^2 / \sum(x_i - \bar{x})^2}$

Gauss–Markov

- under SR1–SR5, OLS is BLUE
- normality optional (SR6) for *exact* normal b

Next time (Lecture 8): the last unknown, σ^2

Every variance and standard error above hides σ^2. Next: estimate it with $\hat{\sigma}^2 = \mathrm{SSE}/(N-2)$, get the standard error of the regression, and make point predictions.

Questions?