Properties of OLS

Gov 2001: Quantitative Social Science Methods I

Week 9, Lecture 18

Spring 2026

For Today

Required Reading

- Blackwell, Chapter 7 (pp. 139–157)
- Aronow & Miller, §4.1.1–4.1.3 (pp. 143–151)

Today: When is OLS unbiased? What is its sampling variance?

Roadmap

1. Writing $\hat{\beta}$ in terms of errors
2. Unbiasedness of OLS
3. Sampling variance of $\hat{\beta}$
4. Gauss-Markov theorem
5. Standard errors

Part I: $\hat{\beta}$ as a Function of Errors

The Setup

Assume the population model:

$$Y_i = \alpha^* + \beta^* X_i + \varepsilon_i$$

where:

- α^*, β^* are the population BLP parameters
- $\varepsilon_i = Y_i - \alpha^* - \beta^* X_i$ is the population error
- By construction of BLP: $\mathbb{E}[\varepsilon_i | X_i] = 0$

Question: What are the properties of $\hat{\beta}$ as an estimator of β^*?

Rewriting $\hat{\beta}$

Recall:

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

Using $Y_i = \alpha^* + \beta^* X_i + \varepsilon_i$:

$$Y_i - \bar{Y} = \beta^*(X_i - \bar{X}) + (\varepsilon_i - \bar{\varepsilon})$$

Substitute:

$$\hat{\beta} = \frac{\sum(X_i - \bar{X})[\beta^*(X_i - \bar{X}) + (\varepsilon_i - \bar{\varepsilon})]}{\sum(X_i - \bar{X})^2}$$

Simplifying

$$\hat{\beta} = \frac{\beta^* \sum(X_i - \bar{X})^2 + \sum(X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon})}{\sum(X_i - \bar{X})^2}$$

$$\hat{\beta} = \beta^* + \frac{\sum(X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon})}{\sum(X_i - \bar{X})^2}$$

Since $\sum(X_i - \bar{X})\bar{\varepsilon} = \bar{\varepsilon} \sum(X_i - \bar{X}) = 0$:

$$\boxed{\hat{\beta} = \beta^* + \frac{\sum_{i=1}^{n}(X_i - \bar{X})\varepsilon_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

Key Representation

$$\hat{\beta} = \beta^* + \sum_{i=1}^{n} w_i \varepsilon_i$$

where the **weights** are:

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

Key insight: $\hat{\beta}$ is a **linear combination of the errors**.

Properties of weights:

- $\sum_i w_i = 0$
- $\sum_i w_i X_i = 1$

Part II: Unbiasedness

Unbiasedness (Conditional on X)

Take the expectation conditional on $X_1, \ldots, X_n$:

$$\mathbb{E}[\hat{\beta}|X] = \mathbb{E}\left[\beta^* + \sum_i w_i \varepsilon_i \,\middle|\, X\right]$$

Since β^* and w_i are constants given X:

$$\mathbb{E}[\hat{\beta}|X] = \beta^* + \sum_i w_i \, \mathbb{E}[\varepsilon_i|X]$$

If $\mathbb{E}[\varepsilon_i|X] = 0$ for all i:

$$\boxed{\mathbb{E}[\hat{\beta}|X] = \beta^*}$$

The Key Assumption

Conditional Mean Independence:

$$\mathbb{E}[\varepsilon_i | X_1, \ldots, X_n] = 0$$

This is automatically satisfied when ε_i is the BLP error because:

$$\mathbb{E}[\varepsilon_i | X_i] = \mathbb{E}[Y_i - \alpha^* - \beta^* X_i | X_i] = 0$$

by construction of the BLP.

Conclusion: OLS is unbiased for the BLP coefficients.

Unconditional Unbiasedness

By the Law of Iterated Expectations:

$$\mathbb{E}[\hat{\beta}] = \mathbb{E}\Big[\mathbb{E}[\hat{\beta}|X]\Big] = \mathbb{E}[\beta^*] = \beta^*$$

Result:

OLS is Unbiased

Under the assumption that $\mathbb{E}[\varepsilon_i|X] = 0$:

$$\mathbb{E}[\hat{\beta}] = \beta^*$$

On average, across repeated samples, OLS gets the right answer.

Important Clarification

What are we estimating?
OLS is unbiased for $\beta^* = \mathrm{Cov}(X, Y)/\mathrm{Var}(X)$, the BLP coefficient.

This is NOT necessarily a causal effect!

- OLS estimates the best linear prediction relationship
- Whether this is causal depends on additional assumptions
- Selection bias, confounding, etc. can make $\beta^* \neq$ causal effect

Unbiasedness for BLP ≠ Unbiasedness for causal effect

Part III: Sampling Variance

Sampling Variance of $\hat{\beta}$

We've established:

$$\hat{\beta} = \beta^* + \sum_{i=1}^{n} w_i \varepsilon_i$$

To find the variance, compute:

$$\text{Var}(\hat{\beta}|X) = \text{Var}\left(\sum_{i=1}^{n} w_i \varepsilon_i \middle| X\right)$$

Since w_i are constants given X:

$$\text{Var}(\hat{\beta}|X) = \sum_{i=1}^{n} w_i^2 \,\text{Var}(\varepsilon_i|X) + 2\sum_{i<j} w_i w_j \,\text{Cov}(\varepsilon_i, \varepsilon_j|X)$$

Homoskedasticity Assumption

Assume:

1. $\text{Var}(\varepsilon_i|X) = \sigma^2$ for all i (homoskedasticity)
2. $\text{Cov}(\varepsilon_i, \varepsilon_j|X) = 0$ for $i \neq j$ (no autocorrelation)

Under these assumptions:

$$\text{Var}(\hat{\beta}|X) = \sigma^2 \sum_{i=1}^{n} w_i^2 = \sigma^2 \cdot \frac{\sum_i (X_i - \bar{X})^2}{[\sum_i (X_i - \bar{X})^2]^2}$$

$$\boxed{\text{Var}(\hat{\beta}|X) = \frac{\sigma^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}}$$

Intuition for the Variance Formula

$$\mathrm{Var}(\hat{\beta}|X) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sigma^2}{n \cdot \widehat{\mathrm{Var}}(X)}$$

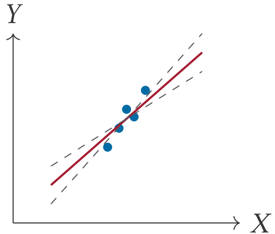
Variance decreases when:

1. σ^2 **is smaller**: Less noise in Y given X
2. n **is larger**: More data points
3. $\widehat{\mathrm{Var}}(X)$ **is larger**: More spread in X

Spread in X gives you "leverage" to estimate the slope.

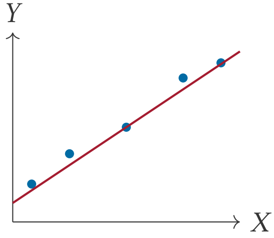
Visual Intuition

Low variance in X:



Many lines fit similarly well
$\Rightarrow$ High variance in $\hat{\beta}$

High variance in X:



Only one line fits well
$\Rightarrow$ Low variance in $\hat{\beta}$

Part IV: Gauss-Markov Theorem

The Gauss-Markov Theorem

Gauss-Markov

Under the assumptions:

1. $\mathbb{E}[\varepsilon_i|X] = 0$ (conditional mean zero)
2. $\mathrm{Var}(\varepsilon_i|X) = \sigma^2$ (homoskedasticity)
3. $\mathrm{Cov}(\varepsilon_i, \varepsilon_j|X) = 0$ for $i \neq j$ (no autocorrelation)

OLS is the **Best Linear Unbiased Estimator** (BLUE).

"Best" = lowest variance among all linear unbiased estimators.

What BLUE Means

Linear: $\hat{\beta}$ is a linear function of $Y_1, \ldots, Y_n$:

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

for some weights c_i that depend only on X.

Unbiased: $\mathbb{E}[\hat{\beta}] = \beta^*$

Best: Among all linear unbiased estimators, OLS has the smallest variance.

Interpreting Gauss-Markov

What Gauss-Markov says:

- OLS is efficient within the class of linear unbiased estimators
- You can't do better (lower variance) without either:
  - Introducing bias
  - Using nonlinear estimators
  - Making stronger assumptions

What Gauss-Markov does NOT say:

- OLS is always best (might not be if assumptions fail)
- OLS gives causal estimates
- Linearity is the right functional form

Political Science Example: Campaign Ads

Estimating returns to negative advertising:

$$\text{Vote share}_i = \alpha + \beta \times \text{Negative ads}_i + \varepsilon_i$$

Gauss-Markov: OLS gives lowest-variance linear unbiased estimate of β.

But is $\hat{\beta}$ causal? Only if $\mathbb{E}[\varepsilon|\text{Ads}] = 0$:

- Competitive races use more ads *and* have closer outcomes
- Candidate quality affects both ad strategy and votes
- Without randomization: $\hat{\beta}$ describes correlation, not causation

When Assumptions Fail

Heteroskedasticity: $\text{Var}(\varepsilon_i|X) \neq \sigma^2$

- OLS is still unbiased
- OLS is still consistent
- OLS is **no longer efficient** (not BLUE)
- Standard errors are **wrong**

Solution: Use heteroskedasticity-robust standard errors.
(We'll cover this in Week 12.)

Part V: Standard Errors

Estimating σ^2

We need $\sigma^2 = \text{Var}(\varepsilon_i | X)$, but we don't observe ε_i.

Estimator:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{e}_i^2 = \frac{SSR}{n-2}$$

Why $n-2$?

- We estimated 2 parameters ($\hat{\alpha}$, $\hat{\beta}$)
- "Degrees of freedom" correction
- Makes $\hat{\sigma}^2$ unbiased: $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$

Standard Error of $\hat{\beta}$

True variance:

$$\text{Var}(\hat{\beta}|X) = \frac{\sigma^2}{\sum_i (X_i - \bar{X})^2}$$

Estimated variance:

$$\widehat{\text{Var}}(\hat{\beta}) = \frac{\hat{\sigma}^2}{\sum_i (X_i - \bar{X})^2}$$

Standard error:

$$SE(\hat{\beta}) = \sqrt{\widehat{\text{Var}}(\hat{\beta})} = \frac{\hat{\sigma}}{\sqrt{\sum_i (X_i - \bar{X})^2}}$$

The t-Statistic

To test $H_0 : \beta^* = \beta_0$:

t-statistic:

$$t = \frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})}$$

Under the null hypothesis (and normality of ε):

$$t \sim t_{n-2}$$

For large n: $t \approx N(0, 1)$ (by CLT).

Confidence Interval for β^*

A $(1 - \alpha)$ confidence interval:

$$\hat{\beta} \pm t_{n-2,1-\alpha/2} \times SE(\hat{\beta})$$

For large n and 95% confidence:

$$\hat{\beta} \pm 1.96 \times SE(\hat{\beta})$$

Interpretation: In repeated sampling, 95% of intervals constructed this way contain β^*.

Part VI: Putting It Together

Example: Education and Wages

Regression output:

$$\widehat{\log(\text{wage})} = 1.80 + 0.087 \times \text{Education}$$

	Estimate	SE	t	p-value
Intercept	1.80	0.15	12.0	< 0.001
Education	0.087	0.011	7.9	< 0.001

$n = 1000$, $R^2 = 0.20$, $\hat{\sigma} = 0.48$

Interpreting the Output

Point estimate: $\hat{\beta} = 0.087$
Each additional year of education is associated with 8.7% higher wages.

Standard error: $SE(\hat{\beta}) = 0.011$
Measures uncertainty about $\hat{\beta}$ due to sampling variability.

***t*-statistic:** $t = 0.087/0.011 = 7.9$
Highly significant—strong evidence against $H_0 : \beta^* = 0$.

95% CI: $0.087 \pm 1.96 \times 0.011 = [0.065, 0.109]$

Summary: Properties of OLS

Property	**Condition**	**Meaning**
Unbiased	$\mathbb{E}[\varepsilon \mid X] = 0$	$\mathbb{E}[\hat{\beta}] = \beta^*$
Consistent	I.I.D. data	$\hat{\beta} \xrightarrow{p} \beta^*$
BLUE	Gauss-Markov	Lowest variance (linear, unbiased)
Asymp. Normal	CLT	$\sqrt{n}(\hat{\beta} - \beta^*) \xrightarrow{d} N(0, V)$

Summary

Key results:

1. $\hat{\beta} = \beta^* + \sum_i w_i \varepsilon_i$ (linear in errors)
2. OLS is **unbiased**: $\mathbb{E}[\hat{\beta}] = \beta^*$
3. Sampling variance: $\text{Var}(\hat{\beta}) = \sigma^2 / \sum (X_i - \bar{X})^2$
4. **Gauss-Markov**: OLS is BLUE under homoskedasticity
5. $SE(\hat{\beta}) = \hat{\sigma} / \sqrt{\sum (X_i - \bar{X})^2}$

Looking Ahead

Next week: Multiple Regression

- Adding more X variables
- Frisch-Waugh-Lovell theorem
- Omitted variable bias
- "Controlling for" covariates

The properties we derived extend naturally to multiple regression.

OLS is unbiased for the BLP coefficients.

Under Gauss-Markov conditions,
OLS is the **best** linear unbiased estimator.

But “best” estimator of a predictive relationship
is not the same as “causal effect.”