

Lecture 41: Why OLS fails: consistency; a regressor correlated with the disturbance

Econometrics ■ Unit III: Autocorrelation and instrumental variables

Kapil Kumar

Department of Statistics
Chaudhary Charan Singh University, Meerut

B.Sc. (Hons) Statistics, Semester V
Academic Session 2026–2027

Previous lecture

Lecture 40: A complete worked example; summary of key formulas.

Today

- Why another method is needed
 - Consistency
 - OLS is inconsistent when a regressor is correlated with the disturbance

OLS requires that the regressors are uncorrelated with the disturbance. In several important situations this fails:

1. **Errors in variables.** If a regressor is measured with error, the measurement error enters both the regressor and the disturbance, so they are correlated.
2. **Simultaneity.** In a demand equation, price is determined together with quantity and depends on the demand disturbance (Chapters 7 and 8).
3. **Omitted variables.** If an omitted variable affects y and is correlated with a regressor, its effect goes into the disturbance, which is then correlated with that regressor.

In these cases OLS is biased, and the bias does not disappear as the sample size grows. To explain this we need the idea of consistency.

Definition 6.1

An estimator $\hat{\theta}_n$ based on n observations is **consistent** for θ if, for every $\epsilon > 0$,

$$P(|\hat{\theta}_n - \theta| > \epsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

We then write $\text{plim } \hat{\theta}_n = \theta$ (“probability limit”).

Consistency is a large-sample property: as n grows, the estimator gets closer and closer to the true value. Unbiasedness is a fixed-sample property.

We use the following facts:

- a sample average converges in probability to the corresponding population mean (law of large numbers);
- $\text{plim}(A_n B_n) = \text{plim } A_n \cdot \text{plim } B_n$ and $\text{plim}(A_n/B_n) = \text{plim } A_n / \text{plim } B_n$ when the limits exist and $\text{plim } B_n \neq 0$.

OLS is inconsistent when $\text{Cov}(x, \varepsilon) \neq 0$

Take $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. As in Chapter 4,

$$b_1 = \beta_1 + \frac{\sum_i (x_i - \bar{x}) \varepsilon_i}{\sum_i (x_i - \bar{x})^2} = \beta_1 + \frac{\frac{1}{n} \sum_i (x_i - \bar{x}) \varepsilon_i}{\frac{1}{n} \sum_i (x_i - \bar{x})^2}.$$

- The numerator is the sample covariance of x and ε , and the denominator is the sample variance of x .

Taking probability limits,

$$\text{plim } b_1 = \beta_1 + \frac{\text{Cov}(x, \varepsilon)}{\text{Var}(x)}.$$

If $\text{Cov}(x, \varepsilon) \neq 0$, then $\text{plim } b_1 \neq \beta_1$: OLS is **inconsistent**. More data do not help.

Example 6.1 (Omitted variable)

Suppose the true model is

$$y_i = \beta_0 + \beta_1 x_i + \delta a_i + v_i,$$

where a_i (for example, ability in a wage equation) is not observed and v_i is uncorrelated with x_i . Fitting y on x alone, the disturbance is $\varepsilon_i = \delta a_i + v_i$, and

$$\text{Cov}(x_i, \varepsilon_i) = \delta \text{Cov}(x_i, a_i) + \text{Cov}(x_i, v_i) = \delta \text{Cov}(x_i, a_i).$$

If ability affects wages ($\delta \neq 0$) and is correlated with education x , OLS is inconsistent.

Practice

- Exercises 6.1 and 6.6 (Exercises and Solutions booklet).

Next lecture

Lecture 42: Instruments; the IV estimator in the simple regression.