

Economics 326
Methods of Empirical Research in Economics

Lecture 4: Properties of OLS

Vadim Marmer
University of British Columbia

February 20, 2009

The OLS estimators are random variables

- ▶ The model

$$Y_i = \alpha + \beta X_i + U_i,$$
$$E(U_i | X_1, \dots, X_n) = 0.$$

Conditioning on X in $E(U_i | X_1, \dots, X_n) = 0$ allows us to treat all X 's as fixed, but Y is still random.

- ▶ The estimators

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

are random because they are functions of random data.

The estimators are linear

- Since $\hat{\beta} = \frac{\sum_{i=1}^n (X_i - \bar{X}) Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2}$, we can write $\hat{\beta} = \sum_{i=1}^n w_i Y_i$, where

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

After conditioning on X 's, w_i 's are not random.

- For $\hat{\alpha}$,

$$\begin{aligned}\hat{\alpha} &= \bar{Y} - \hat{\beta} \bar{X} \\ &= \frac{1}{n} \sum_{i=1}^n Y_i - \left(\sum_{i=1}^n w_i Y_i \right) \bar{X} \\ &= \sum_{i=1}^n \left(\frac{1}{n} - \bar{X} w_i \right) Y_i \\ &= \sum_{i=1}^n \left(\frac{1}{n} - \bar{X} \frac{X_i - \bar{X}}{\sum_{l=1}^n (X_l - \bar{X})^2} \right) Y_i.\end{aligned}$$

Unbiasedness

- ▶ $\hat{\beta}$ is called an unbiased estimator if $E\hat{\beta} = \beta$.
- ▶ Suppose that $Y_i = \alpha + \beta X_i + U_i$, $E(U_i | X_1, \dots, X_n) = 0$.
Then $E\hat{\beta} = \beta$.

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{i=1}^n (X_i - \bar{X}) Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X}) (\alpha + \beta X_i + U_i)}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ &= \alpha \frac{\sum_{i=1}^n (X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} + \beta \frac{\sum_{i=1}^n (X_i - \bar{X}) X_i}{\sum_{i=1}^n (X_i - \bar{X})^2} + \frac{\sum_{i=1}^n (X_i - \bar{X}) U_i}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ &= \alpha \frac{0}{\sum_{i=1}^n (X_i - \bar{X})^2} + \beta \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} + \frac{\sum_{i=1}^n (X_i - \bar{X}) U_i}{\sum_{i=1}^n (X_i - \bar{X})^2},\end{aligned}$$

- ▶ or

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

Unbiasedness

- ▶ Once we condition on X_1, \dots, X_n , all X 's in

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

can be treated as fixed.

- ▶ Thus,

$$\begin{aligned} E(\hat{\beta} | X_1, \dots, X_n) &= E\left(\beta + \frac{\sum_{i=1}^n (X_i - \bar{X}) U_i}{\sum_{i=1}^n (X_i - \bar{X})^2} \mid X_1, \dots, X_n\right) \\ &= \beta + E\left(\frac{\sum_{i=1}^n (X_i - \bar{X}) U_i}{\sum_{i=1}^n (X_i - \bar{X})^2} \mid X_1, \dots, X_n\right) \\ &= \beta + \frac{\sum_{i=1}^n (X_i - \bar{X}) E(U_i | X_1, \dots, X_n)}{\sum_{i=1}^n (X_i - \bar{X})^2}. \end{aligned}$$

Unbiasedness

- ▶ Thus, with $E(U_i|X_1, \dots, X_n) = 0$, we have

$$\begin{aligned} E(\hat{\beta}|X_1, \dots, X_n) &= \beta + \frac{\sum_{i=1}^n (X_i - \bar{X}) E(U_i|X_1, \dots, X_n)}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ &= \beta + \frac{\sum_{i=1}^n (X_i - \bar{X}) \cdot 0}{\sum_{i=1}^n (X_i - \bar{X})^2} = \beta. \end{aligned}$$

- ▶ By the LIE, $E\hat{\beta} = E[E(\hat{\beta}|X_1, \dots, X_n)] = E[\beta] = \beta$.

Strong exogeneity of regressors

- ▶ The regressor X is strongly exogenous if $E(U_i|X_1, \dots, X_n) = 0$.
- ▶ Alternatively, we can assume that $E(U_i|X_i) = 0$ and all observations are independent:

$$\begin{aligned} E(U_1|X_1, \dots, X_n) &= E(U_1|X_1), \\ E(U_2|X_1, \dots, X_n) &= E(U_2|X_2) \text{ and etc.} \end{aligned}$$

- ▶ The OLS estimator is in general biased if the strong exogeneity assumption is violated.

Variance of $\hat{\beta}$

- ▶ If $Y_i = \alpha + \beta X_i + U_i$, $E(U_i | X_1, \dots, X_n) = 0$, and

$$E(U_i^2 | X_1, \dots, X_n) = \sigma^2 = \text{constant},$$

and for $i \neq j$

$$E(U_i U_j | X_1, \dots, X_n) = 0,$$

Then

$$\text{Var}(\hat{\beta} | X_1, \dots, X_n) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

- ▶ The assumption $E(U_i^2 | X_1, \dots, X_n) = \sigma^2 = \text{constant}$ is called (conditional) homoskedasticity.
- ▶ The assumption $E(U_i U_j | X_1, \dots, X_n) = 0$ for $i \neq j$ can be replaced by the assumption that the observations are independent.

Variance of $\hat{\beta}$

$$\text{Var}(\hat{\beta} | X_1, \dots, X_n) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

- ▶ The variance of $\hat{\beta}$ is positively related to the variance of the errors $\sigma^2 = \text{Var}(U_i)$.
- ▶ The variance of $\hat{\beta}$ is smaller when X 's are more dispersed.

Variance of $\hat{\beta}$

- ▶ We are going to condition on X 's and will treat them as constants. All expectations below are implicitly conditional on X 's.
- ▶ We have $\hat{\beta} = \beta + \frac{\sum_{i=1}^n (X_i - \bar{X}) U_i}{\sum_{i=1}^n (X_i - \bar{X})^2}$ and $E\hat{\beta} = \beta$,

$$\begin{aligned} \text{Var}(\hat{\beta}) &= E\left[(\hat{\beta} - E\hat{\beta})^2\right] \\ &= E\left[\left(\frac{\sum_{i=1}^n (X_i - \bar{X}) U_i}{\sum_{i=1}^n (X_i - \bar{X})^2}\right)^2\right] \\ &= \left(\frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2}\right)^2 E\left[\left(\sum_{i=1}^n (X_i - \bar{X}) U_i\right)^2\right] \end{aligned}$$

Variance of $\hat{\beta}$



$$\begin{aligned} & \left(\sum_{i=1}^n (X_i - \bar{X}) U_i \right)^2 \\ &= \sum_{i=1}^n \sum_{j=1}^n (X_i - \bar{X}) (X_j - \bar{X}) U_i U_j \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 U_i^2 + \sum_{i=1}^n \sum_{j \neq i} (X_i - \bar{X}) (X_j - \bar{X}) U_i U_j. \end{aligned}$$

► Since $E(U_i U_j) = 0$ for $i \neq j$,

$$\begin{aligned} E \left[\left(\sum_{i=1}^n (X_i - \bar{X}) U_i \right)^2 \right] &= \sum_{i=1}^n (X_i - \bar{X})^2 E U_i^2 + 0 \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 \sigma^2. \end{aligned}$$

Variance of $\hat{\beta}$

We have

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \left(\frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)^2 E \left[\left(\sum_{i=1}^n (X_i - \bar{X}) U_i \right)^2 \right], \\ E \left[\left(\sum_{i=1}^n (X_i - \bar{X}) U_i \right)^2 \right] &= \sigma^2 \sum_{i=1}^n (X_i - \bar{X})^2, \end{aligned}$$

and therefore,

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \left(\frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)^2 \sigma^2 \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \left(\frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2} \right) \sigma^2. \end{aligned}$$

Normality of $\hat{\beta}$

- ▶ Assume that U_i 's are jointly normally distributed conditional on X 's.
- ▶ Then $Y_i = \alpha + \beta X_i + U_i$ are also jointly normally distributed.
- ▶ Since $\hat{\beta} = \sum_{i=1}^n w_i Y_i$, where $w_i = \frac{X_i - \bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2}$ depend only on X 's, $\hat{\beta}$ is also normally distributed conditional on X 's.
- ▶ Conditional on X_1, \dots, X_n

$$\begin{aligned}\hat{\beta} &\sim N(E\hat{\beta}, \text{Var}(\hat{\beta})) \\ &\sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right).\end{aligned}$$