

1 OLS estimator is unbiased

Let us first substitute the expression for the population regression itself¹ into our solution for the slope coefficient

$$\begin{aligned}b &= (X'X)^{-1}X'y \\b &= (X'X)^{-1}X'(X\beta + \epsilon)\end{aligned}$$

A bit more of rearranging (remembering that $(X'X)^{-1}(X'X) = I$) and we get:

$$\begin{aligned}b &= (X'X)^{-1}X'(X\beta + \epsilon) \\b &= (X'X)^{-1}(X'X)\beta + (X'X)^{-1}X'\epsilon \\b &= \beta + (X'X)^{-1}X'\epsilon\end{aligned}$$

Now, one of the assumptions underlying the OLS estimator is that

$$E(X'\epsilon) = 0$$

and another one says that

$$E(\epsilon) = 0$$

, so, in any case:

$$\begin{aligned}\text{If} \\b &= \beta + (X'X)^{-1}X'\epsilon \\ \text{then} \\E(b) &= \beta + 0 = \beta\end{aligned}$$

as long as those assumptions hold.

¹This only works if the functional form is correct. Otherwise this substitution would not be valid.

2 Variances-covariance matrix of OLS estimator

Since

$$\begin{aligned} b &= \beta + (X'X)^{-1}X'\epsilon \\ \text{and} \\ E(b) &= \beta \end{aligned}$$

we have that

$$\begin{aligned} b - E(b) &= (X'X)^{-1}X'\epsilon \\ \text{so} \\ \text{var}(b) &= E(b - E(b))^2 \\ &= E[((X'X)^{-1}X'\epsilon)'((X'X)^{-1}X'\epsilon)] = \\ &= E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] = \\ &= (X'X)^{-1}X'E(\epsilon\epsilon')X(X'X)^{-1} = \\ &= (X'X)^{-1}X'\sigma_\epsilon^2 I X(X'X)^{-1} = \\ &= \sigma_\epsilon^2 (X'X)^{-1}X X(X'X)^{-1} = \\ &= \sigma_\epsilon^2 (X'X)^{-1} \end{aligned}$$

Where I stands for the identity matrix, and σ_ϵ^2 is a scalar, since OLS assumes homoskedasticity, that is the error terms are all normally distributed with mean zero, and all the same variance of σ_ϵ^2 . Check that $\epsilon\epsilon'$ is an $n \times n$ matrix, whose off-diagonal elements are all zero. This is because OLS assume that errors between observations are uncorrelated with each other. That is, the variance covariance matrix of the OLS estimator is:

$$\begin{bmatrix} \sigma_\epsilon^2 & 0 & \dots & 0 \\ 0 & \sigma_\epsilon^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \sigma_\epsilon^2 \end{bmatrix} = \sigma_\epsilon^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 1 \end{bmatrix}$$

Therefore, to recap, we have to estimate the OLS slope coefficients:

$$b = \beta + (X'X)^{-1}X'\epsilon$$

and their variances:

$$\text{var}(b) = \sigma_\epsilon^2 (X'X)^{-1}$$

The off diagonal elements of the $(X'X)$ matrix are usually nonzero (although we do not want them to be very close to one either!). Essentially, the explanatory variables or covariates are very likely somewhat correlated (but they should not be perfectly correlated!).

We typically do not know the value of σ_ϵ^2 , so we would estimate is as follows:

$$\hat{\sigma}_\epsilon^2 = \frac{ee'}{n - (k + 1)}$$

3 Gauss Markov Theorem: OLS estimator is BLUE

This theorem states that the OLS estimator (which yields the estimates in vector b) is, under the conditions imposed, the best (the one with the smallest variance) among the linear unbiased estimators of the parameters in vector β . In order to prove this theorem, let us conceive an alternative linear estimator such as

$$\tilde{\beta} = A'y$$

where A is an $n(k+1)$ matrix. This matrix can contain only nonrandom numbers and functions of X , for $\tilde{\beta}$ to be unbiased conditional on X . It cannot, for example, contain functions of y .

For $\tilde{\beta}$ to be a linear unbiased estimator of β , we need further restrictions. In order to show you these, let us rewrite $\tilde{\beta}$ as:

$$\begin{aligned}\tilde{\beta} &= A'(X\beta + \epsilon) = \\ &= (A'X)\beta + A'\epsilon\end{aligned}\tag{1}$$

Then the expectation of $\tilde{\beta}$ is:

$$\begin{aligned}E(\tilde{\beta}) &= A'X\beta + E(A'\epsilon|X) \\ &\text{which, since } A \text{ is a function of } X \\ &= A'X\beta + A'E(\epsilon|X) \\ &\text{which, since we assumed that } E(\epsilon|X) \text{ is null} \\ &= A'X\beta\end{aligned}$$

So, for $\tilde{\beta}$ to be an unbiased estimator of β it must be true that $A'X\beta = \beta$ for all $(k+1) \times 1$ vectors β . Since $A'X$ is a $(k+1) \times (k+1)$ matrix, $A'X\beta = \beta$ only happens if $A'X = I_{k+1}$

Now, since we have Expression 1 telling us that

$$\tilde{\beta} = (A'X)\beta + A'\epsilon$$

we can work out that

$$\begin{aligned}var(\tilde{\beta}|X) &= A'[var(\epsilon|X)]A \\ &= \sigma^2 A'A\end{aligned}$$

since we assumed homoskedasticity of the errors for the OLS estimator. Therefore

$$\begin{aligned}var(\tilde{\beta}|X) - var(\hat{\beta}|X) &= \sigma^2[A'A - (X'X)^{-1}] \\ \text{premultiply and postmultiply by } A'X &= I_{k+1} \\ &= \sigma^2[A'A - A'X(X'X)^{-1}X'A] \\ &= \sigma^2 A'[I_n - X(X'X)^{-1}X']A \\ &= \sigma^2 A'MA\end{aligned}$$

where $M = I_n - X(X'X)^{-1}X'$.

Since M is symmetric and idempotent (that is its trace equals its rank $M \times M = M$), $A'MA$ is positive semidefinite² for any $n \times (k+1)$ matrix A . This means that the OLS estimator is BLUE. Any other linear unbiased estimator has a larger variance, the difference between variances given by $\sigma^2 A'MA$, which as just proved boils down to something that is not null nor negative.

²A matrix F is positive-semidefinite (or nonnegative-definite) if $x'Fx \geq 0$ for any x .