

OLS estimation

OLS operates on a simple but powerful principle – minimizing the sum of the squared differences between the observed values and the predicted values by our model. This [iterative process](#) involves finding the 'best fit' line that minimally deviates from the observed data points in our scatterplot. The art and science of finding this line lie at the heart of OLS estimation.

Mathematically, we seek to minimize the expression $\sum \epsilon^2 = \sum (Y - \beta_0 - \beta_1 X)^2$. This is called the Sum of Squared Residuals (SSR). This is achieved via a simple optimization under constraints procedure. Basically a couple of derivatives

The first step is to take the [First Order Conditions](#) (FOCs), which means computing the partial derivative of the objective function (the SSR in this case) with respect to each of the parameters of interest (β_0, β_1 in this case but there could be many more) and solve the equations by setting them to 0 (thus you are computing the value of the parameter for which the change in squared residuals is 0, other things constant). The main necessary conditions we need in order to apply FOC, in case you are curious, is that the function we are maximizing is continuous, concave and that the set of feasible solutions is convex. Anyways, here is the mathematical derivation of OLS estimators.

Let $\sum \epsilon^2 = \sum (Y_i - \beta_0 - \beta_1 X_i)^2$ be denoted by S . To find β_0^{ols} and β_1^{ols} we solve

$$\partial S / \partial \beta_0 = \sum_{i=1}^n 2 * (Y_i - \beta_0 - \beta_1 X_i)(-1) = 0 \text{ (chain rule)}$$

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) = 0 \rightarrow \sum_{i=1}^n Y_i - \sum_{i=1}^n \beta_0 - \sum_{i=1}^n \beta_1 X_i = 0; \text{ use trick } 1/n * \sum_{i=1}^n Z_i = \bar{Z}$$

$$n\bar{Y} - n\beta_0 - n\beta_1 \bar{X} = 0 \rightarrow n\bar{Y} - n\beta_1 \bar{X} = n\beta_0 \rightarrow \bar{Y} - \beta_1 \bar{X} = \beta_0^{ols}$$

Now we must find the estimate for β_1 .

$$\partial S / \partial \beta_1 = \sum_{i=1}^n 2 * (Y_i - \beta_0 - \beta_1 X_i)(-X_i) = 0$$

$$\sum_{i=1}^n Y_i X_i - \sum_{i=1}^n \beta_0 X_i - \sum_{i=1}^n \beta_1 X_i^2 = 0 \rightarrow \sum_{i=1}^n Y_i X_i - n\beta_0 \bar{X} - \beta_1 \sum_{i=1}^n X_i^2 = 0 \text{ (substitute } \beta_0^{ols} \text{)}$$

$$\sum_{i=1}^n Y_i X_i - n(\bar{Y} - \beta_1 \bar{X}) \bar{X} - \beta_1 \sum_{i=1}^n X_i^2 = 0 \rightarrow \sum_{i=1}^n Y_i X_i - n\bar{Y} \bar{X} + n\beta_1 \bar{X}^2 = \beta_1 \sum_{i=1}^n X_i^2$$

$$\sum_{i=1}^n Y_i X_i - n\bar{Y} \bar{X} = \beta_1 (\sum_{i=1}^n X_i^2 - n\bar{X}^2) \rightarrow \frac{\sum_{i=1}^n Y_i X_i - n\bar{Y} \bar{X}}{(\sum_{i=1}^n X_i^2 - n\bar{X}^2)} = \beta_1^{ols} = \frac{COV(X,Y)}{VAR(X)}$$

So the OLS estimators are $\beta_0^{ols} = \bar{Y} - \frac{COV(X,Y)}{VAR(X)} * \bar{X}$ and $\beta_1^{ols} = \frac{COV(X,Y)}{VAR(X)}$

OLS is a bit like a balancing act. It systematically adjusts the slope and intercept of our line (remember, these are our β_1 and β_0 from the equation) to find the sweet spot where the sum of the squares of these residuals is the smallest. Why square the residuals? Well, squaring helps in two ways: it ensures that we treat positive and negative deviations equally (since a square is always positive), and it disproportionately penalizes larger errors, which can be crucial in fitting a model accurately. This is exactly what the **lm()** function in R does in the background when we try to fit a linear model to some data.

In econometrics, we impose a set of (strong) assumptions on our linear regression model. These are called the Gauss-Markov Assumptions (GMA):

- **Linearity in Parameters:** The model is linear in parameters, meaning it can be expressed as $Y = \beta_0 + \beta_1 X + U$.
- **Random Sampling:** The data (observations) are drawn from a random sample. This ensures that the sample provides a good representation of the population.
- **No Perfect Multicollinearity:** There is no exact linear relationship between the independent variables in the model. This ensures that each variable contributes unique information.
- **Zero Conditional Mean:** The expectation of the error term U conditional on the independent variable X is zero, i.e., $E(U|X) = 0$. This implies that the error term does not contain systematic information that is omitted from the model.
- **Homoscedasticity:** The variance of the error term is constant across observations, i.e., $Var(U|X) = \sigma^2$. This means that the uncertainty or 'noise' in the relationship between X and Y is consistent across all levels of X .

Under the GMA assumptions, it can be proven that the OLS estimator is BLUE (best linear unbiased estimator). That is, if these assumptions hold then we are guaranteed to obtain the best possible estimation of the parameters of our linear regression model.

In econometrics, linear regression and OLS estimation are not just mathematical tools; they are the lenses through which we interpret the economic world. These methodologies enable us to quantify relationships between variables, test economic theories, and make informed predictions and policy decisions. The beauty of OLS, under the guidance of Gauss-Markov Assumptions, lies in its simplicity and robustness, making it an indispensable tool in the econometrician's toolkit.

However, it's important to remember that while OLS provides a powerful method for estimating linear relationships, the validity of its results hinges on the adherence to the underlying assumptions. Violations of these assumptions can lead to biased or inefficient estimators, underscoring the importance of a thorough understanding of these principles.