

# The OLS method for estimating the parameters of the regression model

Econometrics for minor Finance, Lecture 3

Tunga Kantarcı, Fall 2025

# OLS as an estimation method: Population parameter to estimate

Consider the population model:

$$y_i = \beta x_i + u_i$$

$\beta$  is an unknown population parameter. It is a fixed number.

We want to estimate them given sample data. What method can we use for estimation?

# OLS as an estimation method: The estimator

$$y_i = \beta x_i + u_i$$

Among several possible estimation methods, we focus on the most widely used: **Ordinary Least Squares**.

The idea is that the best estimate of  $\beta$ , is the one that makes

$$\beta x_i$$

as close as possible to

$$y_i$$

so that the error we make,  $u_i$ , is minimized.

Denote this estimate with  $\hat{\beta}$ .

# OLS as an estimation method: The estimator

Consider the model with an intercept:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

The error term for observation  $i$  is

$$u_i = y_i - \beta_0 - \beta_1 x_i$$

For a sample of  $n$  observations, we want to minimize the sum of squared errors:

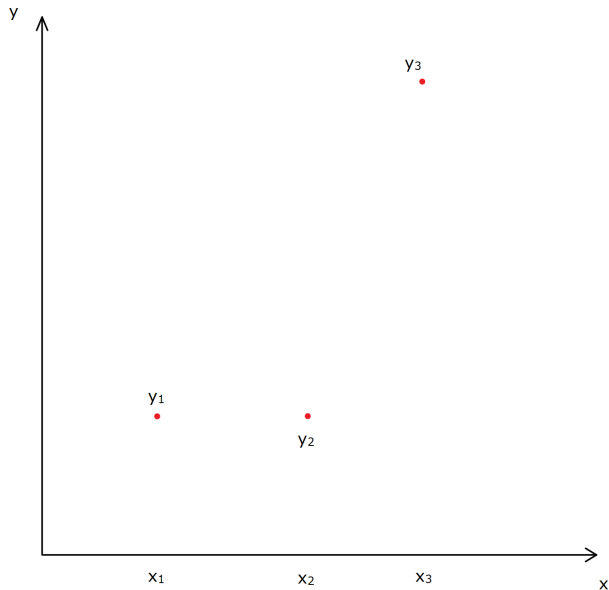
$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

The values of  $\beta_0$  and  $\beta_1$  that minimize this criterion are denoted as

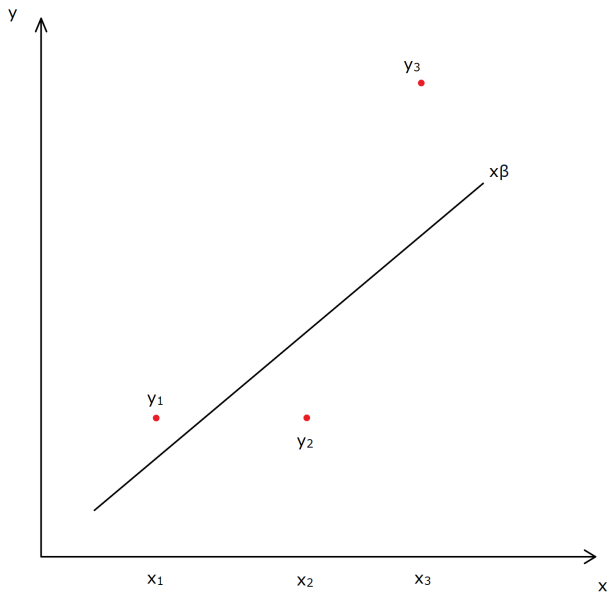
$$\hat{\beta}_0, \hat{\beta}_1$$

and called the OLS estimates.

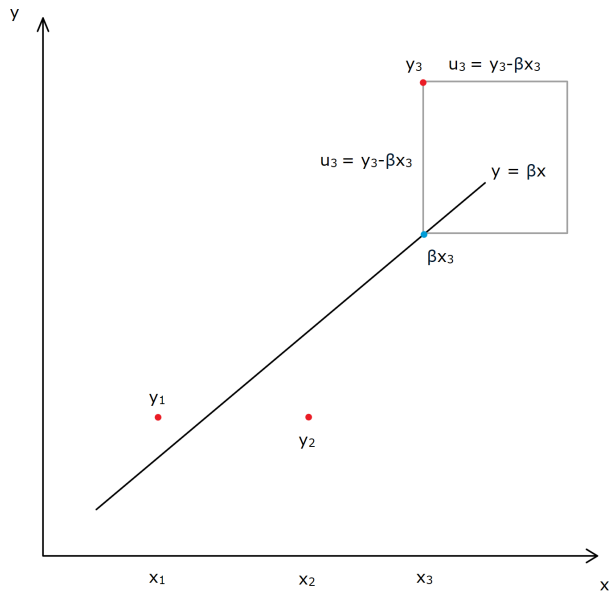
# OLS as an estimation method: The estimator: Intuition



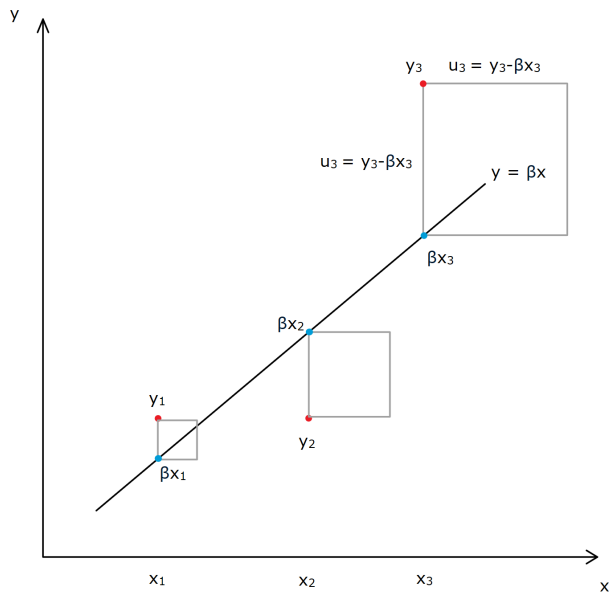
# OLS as an estimation method: The estimator: Intuition



# OLS as an estimation method: The estimator: Intuition



# OLS as an estimation method: The estimator: Intuition





# OLS as an estimation method: The estimator

By minimizing the **sum of squared errors**, OLS finds the

$$\hat{\beta}$$

that makes the regression line **fit** to the data points as good as possible.

# OLS as an estimation method: The estimator

More formally, OLS chooses the parameters to solve

$$\hat{\beta}_0, \hat{\beta}_1 = \arg \min_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

Using some algebra that we skip, the OLS estimator of  $\beta_0$  is

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

and the OLS estimator of  $\beta_1$  is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

## OLS as an estimation method: Fitted value

The estimated population model becomes the fitted model:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

It says: given the OLS estimates of population parameters  $\beta_0, \beta_1$

$$\hat{\beta}_0, \hat{\beta}_1,$$

when

$$x = x_i,$$

our model predicts  $y_i$  as

$$\hat{y}_i$$

We call this the **fitted value**. It represents the part of  $y_i$  that our model is able to explain.

# OLS as an estimation method: Residual

Our population model for observation  $i$  is

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

and hence the error term is

$$u_i = y_i - (\beta_0 + \beta_1 x_i)$$

The estimate of the second term is the fitted value

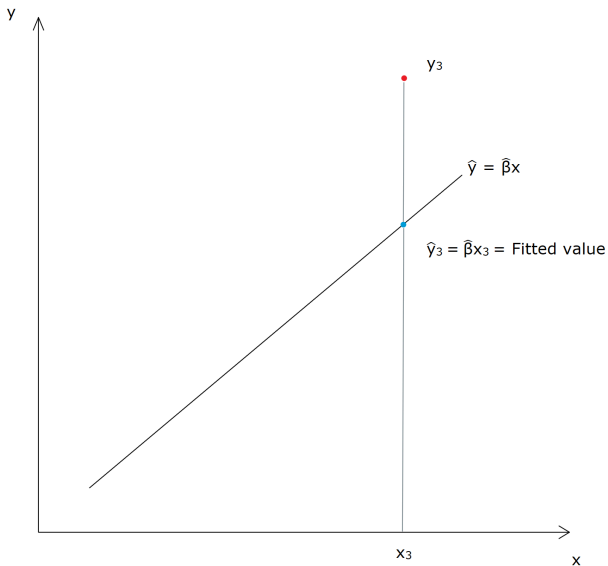
$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Then, the difference between the observed and fitted value is

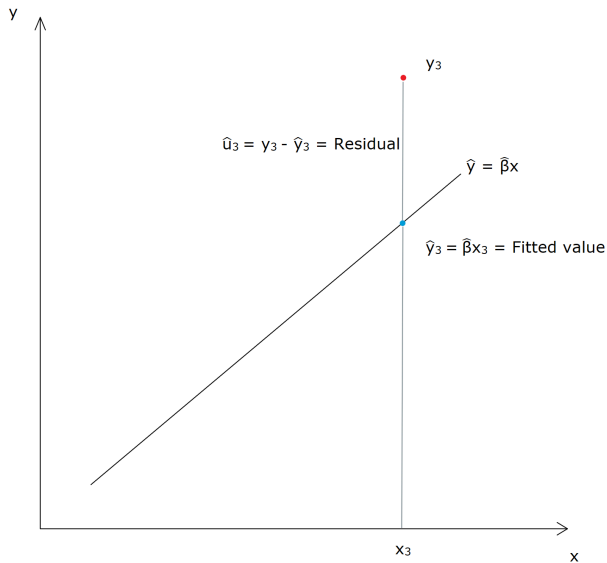
$$\hat{u}_i = y_i - \hat{y}_i$$

We call this the **residual**. It represents the part of  $y_i$  that the model does not explain.

# OLS as an estimation method: Fitted value: Intuition



# OLS as an estimation method: Residual: Intuition



# OLS as an estimation method: Fitted value and residual: Intuition

So the solution to the least squares problem gives

$$y = \hat{y} + \hat{u}$$

It says that the observed outcome variable is fully explained by what our model can and cannot explain. It has to be. How else could it be?

# OLS as an estimation method: The model and the method

OLS is an approximation, or estimation, **method**. It is not a model. There is no such thing as “OLS model”, a term that appears even in good journals.

The **model** is the liner regression model.

We estimate the population, or slope, parameters of the regression **model** using the OLS **method**.



# OLS as an estimation method: Algebraic properties of the OLS method

The solution of the OLS problem leads to three results. First, if the regression includes a constant, the residuals, or deviations from the regression line, sum to zero:

$$\sum_i^n \hat{u}_i = 0$$

Intuition: OLS chooses the line so that residuals cancel out in aggregate: there is no systematic tendency for them to be positive or negative.

# OLS as an estimation method: Algebraic properties of the OLS method

Second,

$$y_i = \hat{\beta}x_i + \hat{u}_i$$

taking the average over  $i$ , we have

$$\bar{y} = \hat{\beta}\bar{x} + \bar{\hat{u}}$$

Since

$$\bar{\hat{u}} = 0$$

by the first result, we have

$$\bar{y} = \hat{\beta}\bar{x}$$

Intuition: OLS aligns the fitted line so that it goes exactly through the average values, ensuring the model captures the central tendency of the sample.

# OLS as an estimation method: Algebraic properties of the OLS method

Third, we showed that

$$y_i = \hat{y}_i + \hat{u}_i$$

Taking the average over  $i$ , we have

$$\bar{y} = \bar{\hat{y}} + \bar{\hat{u}}$$

Since  $\bar{\hat{u}} = 0$  by the first result, we obtain

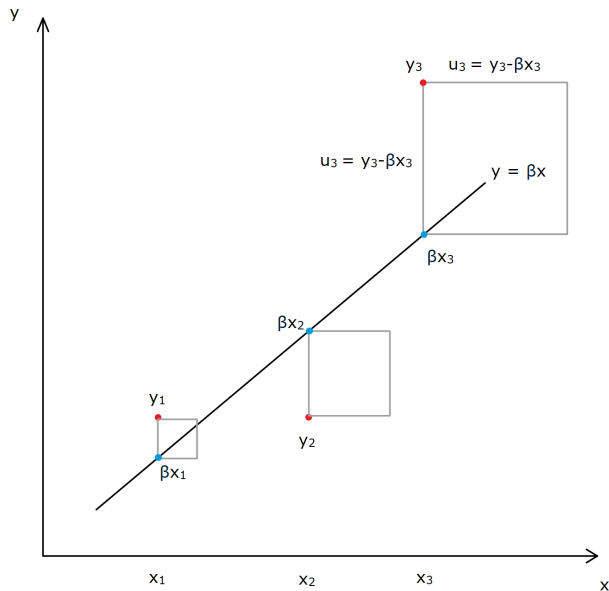
$$\bar{y} = \bar{\hat{y}}$$

Intuition: the average of the fitted values equals the average of the actual values, so OLS ensures the model reproduces the sample mean without bias.

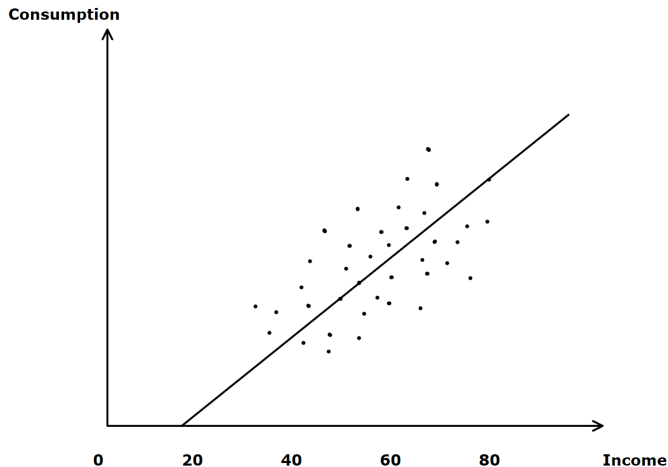
## OLS as an estimation method: Insights: Outlier

The OLS estimator is not robust to outliers.  $y_3$  contributes too much to the sum of squared errors, the minimization problem of OLS. It pulls the regression towards itself. A given observation is not supposed to dictate a contribution much larger than other observations do.

# OLS as an estimation method: Insights: Outlier



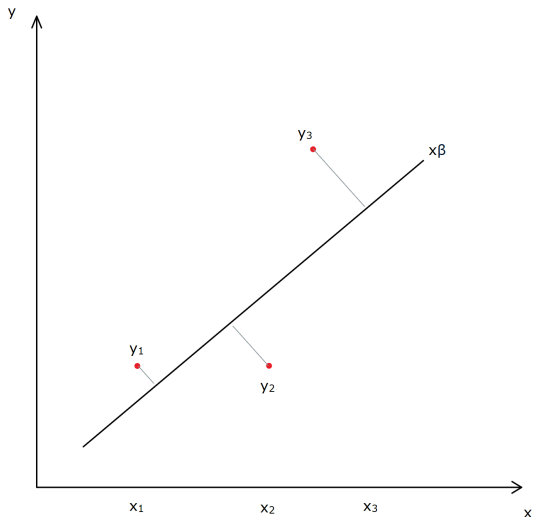
# OLS as an estimation method: Insights: Models are approximations



## OLS as an estimation method: Insights: Models are approximations

For incomes between 40 and 80, the consumption function can be approximated by the model. Does the model describe the consumption-income relationship for all incomes, or only for the those around the center? Only in the center. What is the predicted consumption when income is 10? A negative value. Models are approximations. Approximations do not work well if we move too far away from the point of approximation. OLS is a good approximation around the average value of  $x$ .

# OLS as an estimation method: Insights: Why not minimize orthogonal deviations?





# OLS as an estimation method: Insights: Why not minimize orthogonal deviations?

Why not minimize the sum of orthogonal deviations but squared deviations? Such a method leads to an estimator that does not have statistical properties as desirable.

# Population regression function

Our population model for observation  $i$  is

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

We do not want to work with the entire distribution of the random variable  $y_i$ , but a summary measure of it, the expected value. So take the expected value conditional on  $x_i$ , to obtain

$$E(y_i | x_i) = E(\beta_0 + \beta_1 x_i | x_i) + E(u_i | x_i)$$

Under the exogeneity assumption, that is,  $E(u_i | x_i) = 0$ , we obtain

$$E(y_i | x_i) = E(\beta_0 + \beta_1 x_i | x_i)$$

Using the con. exp. property 1, we obtain

$$E(y_i | x_i) = \beta_0 + \beta_1 x_i$$

This is the **population regression function**.

# Population regression function: Intuition

$$E(y_i | x_i) = \beta_0 + \beta_1 x_i$$

On average, what outcome do we expect at a specific value of the explanatory variable?

# Sample regression function

The population regression function

$$E(y_i | x_i) = \beta_0 + \beta_1 x_i$$

is not known to us. Using the OLS estimates of the population parameters, we can estimate it as

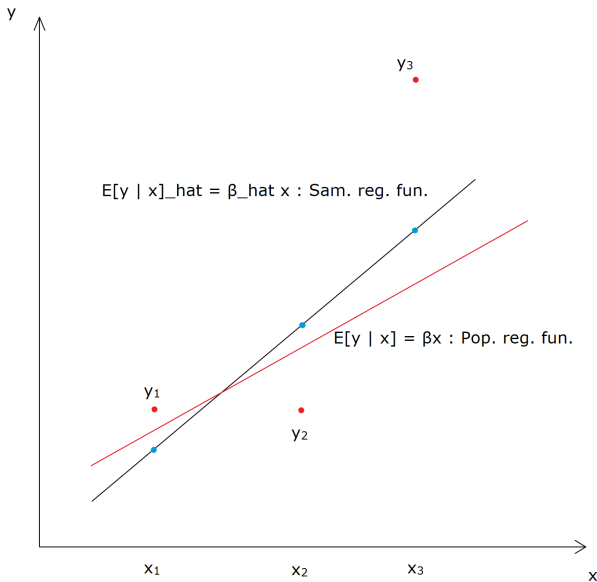
$$\begin{aligned} E(\widehat{y_i | x_i}) &= \hat{\beta}_0 + \hat{\beta}_1 x_i \\ &= \hat{y}_i \end{aligned}$$

which is just the prediction we have learned.

$$E(\widehat{y_i | x_i}) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

is the **sample regression function**. It is the estimate of the population regression function.

# Sample regression function



# Sample regression function

Consider the sample regression function

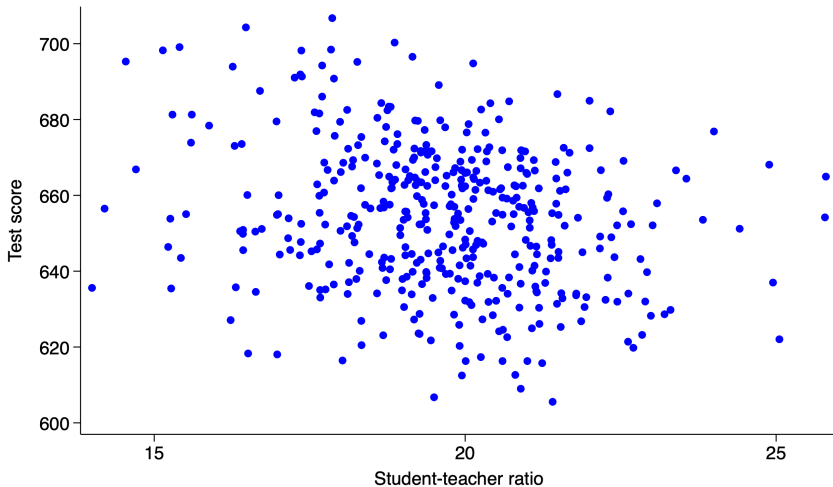
$$\widehat{E[y | x]} = \hat{\beta}_0 + \hat{\beta}_1 x$$

The discrete change in the predicted dependent variable with respect to a discrete change in the independent variable is

$$\frac{\Delta \widehat{E[y | x]}}{\Delta x} = \hat{\beta}_1$$

This shows that when the independent variable changes by some unit, the dependent variable changes **on average** by our OLS estimate.

# Sample regression function: Example



## Sample regression function: Example

Assume that the population regression model is

$$E(\text{testscr} \mid \text{str}) = \beta_0 + \beta_1 \text{str}$$

We want to estimate this model. Given sample data, we can use the OLS method in Stata.



# Sample regression function: Example

```
. regress testscr str
```

Source	SS	df	MS	Number of obs	=	420
Model	7794.11012	1	7794.11012	F(1, 418)	=	22.58
Residual	144315.484	418	345.252353	Prob > F	=	0.0000
				R-squared	=	0.0512
				Adj R-squared	=	0.0490
Total	152109.594	419	363.030056	Root MSE	=	18.581

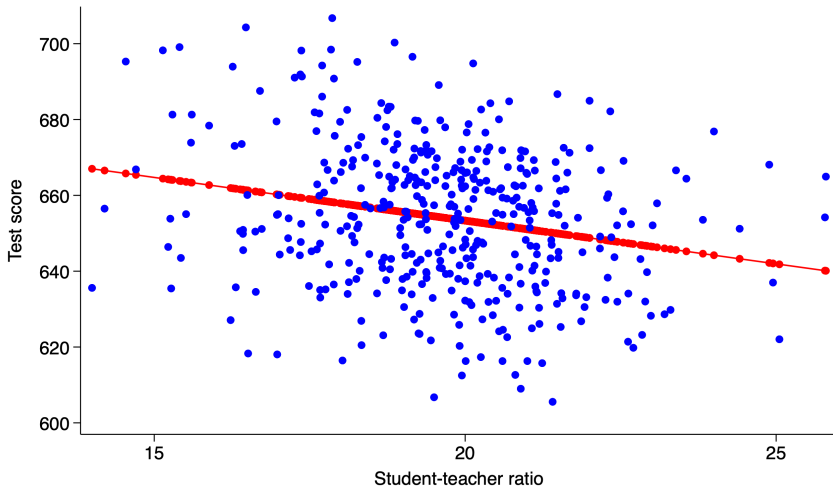
testscr	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
str	-2.279808	.4798256	-4.75	0.000	-3.22298	-1.336637
_cons	698.933	9.467491	73.82	0.000	680.3231	717.5428

## Sample regression function: Example

Our sample regression model, or the fitted model, becomes

$$E(\widehat{testscr} \mid str) = 698.93 - 2.28str$$

# Sample regression function: Example



## Sample regression function: Example

Interpretation of the estimated coefficient of the independent variable:

$$\frac{\Delta E(\widehat{testscr} \mid str)}{\Delta str} = -2.28$$

A unit increase in student-teacher ratio is associated with a  $-2.28$  points decrease in test scores, on average.