Simple Regression Model

Definition

Objective : estimate the effect of X on Y

Y	X
dep. var.	indep. var.
explained	explanatory
regressand	regressor
e.g. wage	educ.
pollution	trade

The simple linear regression (SLR) model

$$y = \underbrace{\beta_0}_{\text{intercept}} + \underbrace{\beta_1}_{\text{slope}} x + \underbrace{u}_{\text{unobserved or error term}}$$

$$\Delta y = \beta_1 \Delta x + \Delta u$$

β_1 : Δy for $\Delta x = 1$ (all else const.)

objective : estimate β_0 and β_1

2 assumptions : $\boxed{E(u) = 0}$

$$y = \beta_0 + \beta_1 x + u$$

$\beta_0 + 150$

$u - 150$

if $E(u) = 150$

$$E(u|x) = E(u)$$

$$\Rightarrow E(u|x) = 0$$

$$\text{corr.}(x, u) = 0$$

$$\boxed{E(xu) = 0}$$

Deriving OLS Estimates

Sample analogs : $\hat{\beta}_0$ and $\hat{\beta}_1$ such that

$$\frac{1}{n} \sum_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0$$

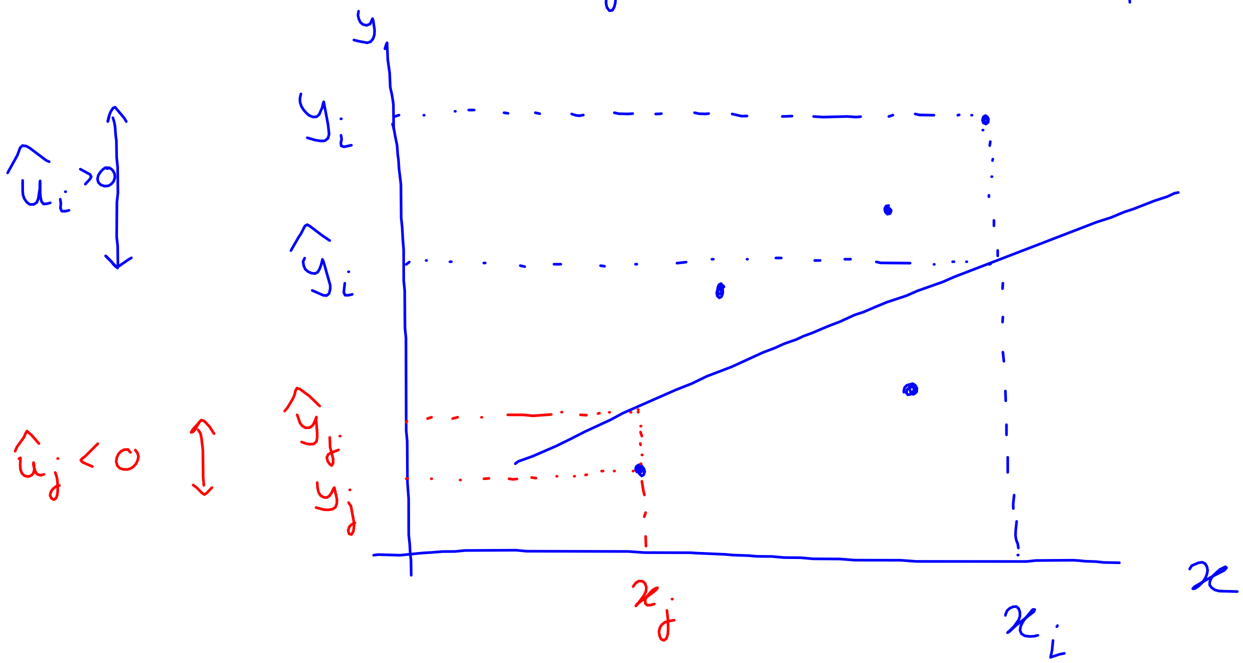
n: sample size

i: obs. i

$1, 2, \ldots, n$

$$\frac{1}{n} \sum_i x_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0$$

Estimated regression line : $\hat{\beta}_0 + \hat{\beta}_1 x_i = \hat{y}_i$



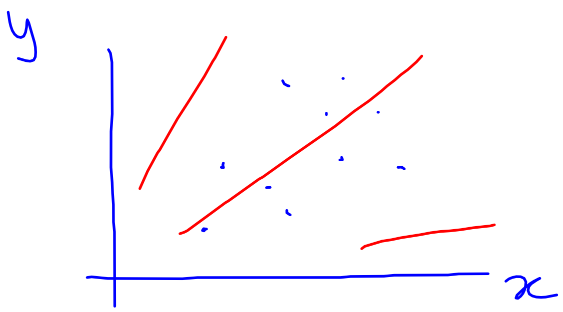
For ith obs. : y_i : value of dep. var.

$\hat{y}_i$: fitted value

$\hat{u}_i$: residual

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

$\hat{\beta}_0$ and $\hat{\beta}_1$ also minimize the sum of squared residuals (SSR) $\longrightarrow$ $\sum_{i=1}^{n} \hat{u}_i^2$



Method $\rightarrow$ ordinary least squares (OLS)

$$\hat{\beta}_1 = \frac{\text{how } x \text{ and } y \text{ covary}}{\text{how } x \text{ varies}}$$

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

$\bar{x}, \bar{y}$: sample means

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

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

Properties of OLS

Sum / avg. of OLS residuals

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

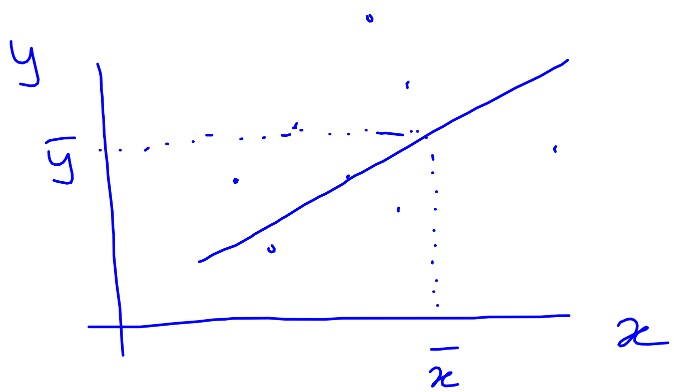
Sample correlⁿ b/w x and $\hat{u}$

$$\sum_i x_i \hat{u}_i = 0$$

Sample correlⁿ b/w $\hat{y}$ and $\hat{u}$

$$\sum_i \hat{y}_i \hat{u}_i = 0$$

$(\bar{x}, \bar{y})$: on the OLS reg. line



For each obs. i

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

Total variation

explained variation

residual variation