

Ch. 3 : Multiple Regression

Motivation

y

more likely that u & $x_1, x_2, \dots$ are not correl. indep. vars.

$x_1, x_2, \dots$

u more likely to be exogenous in case of multiple regression

wage

educ., exper., IQ, demogr., family background

followers

hrs. per day, age, # posts, ...

$k \rightarrow$ no. of indep. vars.

$$k = 2$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

$$\beta_1 \rightarrow \Delta y \text{ for } \Delta x_1 = 1, \Delta x_2 = 0, \Delta u = 0$$

Estimation

$$E(u) = 0$$

$$E(x_1 u) = 0$$

$$E(x_2 u) = 0$$

OLS estimates of $\beta_0, \beta_1, \beta_2$: $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}) = 0$$

$$\frac{1}{n} \sum_{i=1}^n x_{i1} \square = 0$$

$$\frac{1}{n} \sum_{i=1}^n x_{i2} \square = 0$$

Fitted value

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}$$

$$\hat{u}_i = y_i - \hat{y}_i \quad (\text{residual})$$

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \quad \begin{array}{l} \text{non } \downarrow \text{ing} \\ \text{in } k (\# \text{ } x\text{'s}) \end{array}$$

(goodness of fit)

$$\text{Adjusted } R^2 \rightarrow \bar{R}^2 = 1 - \frac{\frac{SSR}{(n-k-1)}}{\frac{SST}{(n-1)}}$$

$$\text{As } k \uparrow \quad SSR \downarrow \quad (n-k-1) \downarrow$$

Expected Value

unbiasedness of
OLS estimators

$$E(\hat{\beta}) = \beta$$

under certain assumptions

linear in
parameter

random
sample

variation
in each x

No linear relⁿ
among x 's

$$\text{wage} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{age} + u$$

$$\text{age} = 6 + \text{educ.} + \text{exper}$$

$$n \geq k+1$$