Definition

objective : estimate the effect of X on Y

	Y	X
	dependent var.	independent var.
	explained 〃	explanatory 〃
	regressand	regressor
e.g.	pollution	trade
	wage	educ.

Simple linear regression (SLR) model

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

intercept

slope

unobserved or error term

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

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

Objective : estimate β_0 and β_1

2 assumptions :

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

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

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

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

$$\boxed{E(x.u) = 0}$$

Deriving OLS estimates

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

n : sample size

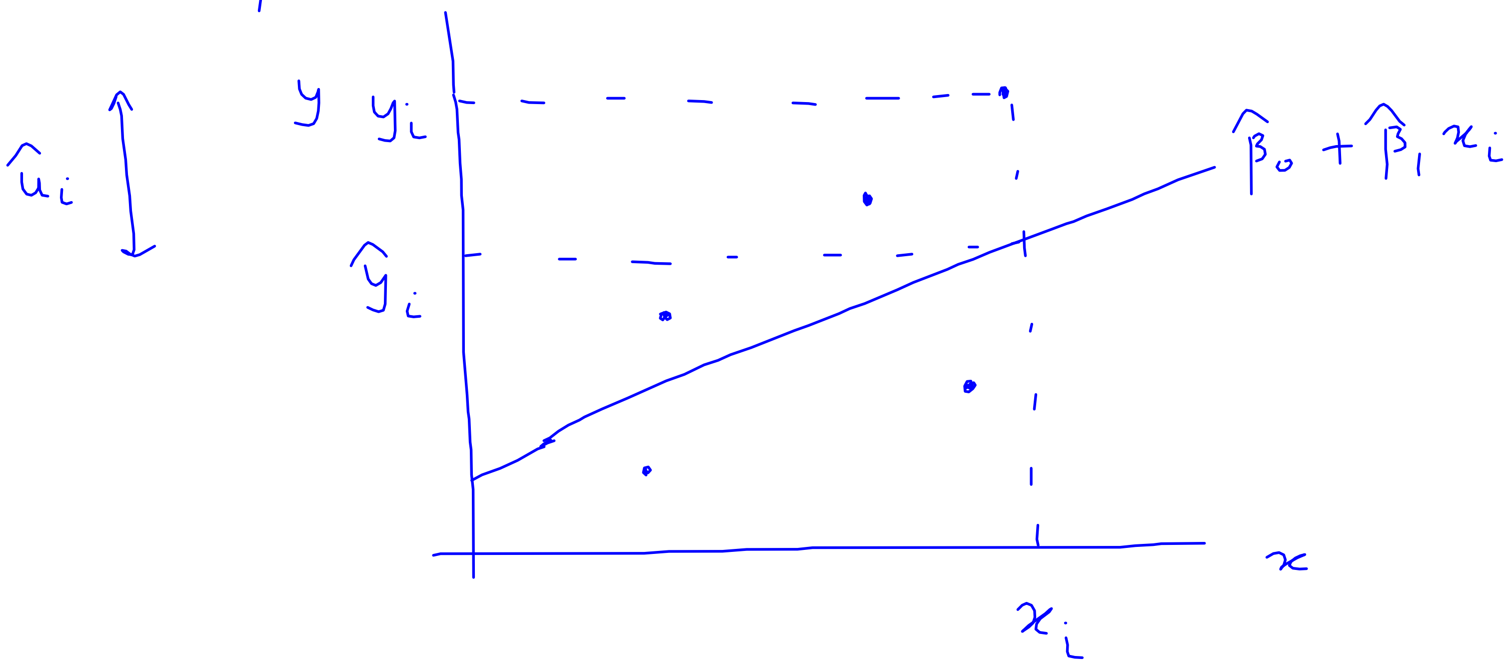
i : obs i 1, 2, ..., n

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

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

Estimated regression line:

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



For ith obs. y_i : dep. var. value

$\hat{y}_i$: fitted value

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

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

Method $\longrightarrow$ also called ordinary least Squares (OLS)

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

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

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

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

For each obs.

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

Total variation (sum of squares)

$$\sum_i (y_i - \bar{y})^2$$

SST : total sum of squares

explained variation

$$\sum_i (\hat{y}_i - \bar{y})^2$$

SSE : explained sum of squares

residual variation

SSR

$$\sum_i \hat{u}_i^2$$

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

$0 \leq R^2 \leq 1$

also = square of correl^n b/w y and $\hat{y}$

High R^2 : not the ultimate objective

Example

y (wage)	x (educ)
3.1	11
3.2	12
3	11
6	8
5.3	12
8.8	16
11	18
5	12
3.6	12
18	17

$$\begin{aligned} \bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} &= 0 \\ \overline{xy} - \hat{\beta}_0 \bar{x} - \hat{\beta}_1 \overline{x^2} &= 0 \end{aligned}$$

$$\begin{aligned} 6.742 - \hat{\beta}_0 - 12.9\hat{\beta}_1 &= 0 \\ 97.234 - 12.9\hat{\beta}_0 - 175.1\hat{\beta}_1 &= 0 \end{aligned}$$

$$\begin{aligned} \hat{\beta}_0 &= -8.492 \\ \hat{\beta}_1 &= 1.181 \end{aligned}$$

$\bar{x}$, $\bar{y}$, $\overline{xy}$, and $\overline{x^2}$: sample average of x, y, xy, and x^2

Functional form

dep. var. $(y) \longrightarrow$ price of house (\$1,000)

indep. var. $(x) \rightarrow$ size in sq.ft.

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

$\hat{\beta}_1 = 0.14$ For $\Delta x = 1$, $\Delta y = 0.14$ (i.e. $\hat{\beta}_1$)

$$\log(y) = \beta_0 + \beta_1 x + u$$

$$\hat{\beta}_1 = 0.0004$$

Approx. effect of $\Delta x = 1$: $\% \Delta y = 100 \hat{\beta}_1$

$= 0.04\%$

$$\log(y) = \beta_0 + \beta_1 \log(x) + u$$

$$\hat{\beta}_1 = 0.873$$

For $\Delta x = 1\%$

$\Delta y = 0.873\%$ (i.e. $\hat{\beta}_1$) approx. effect

Expected value of OLS estimators

Under certain assumptions $\rightarrow$ OLS unbiased

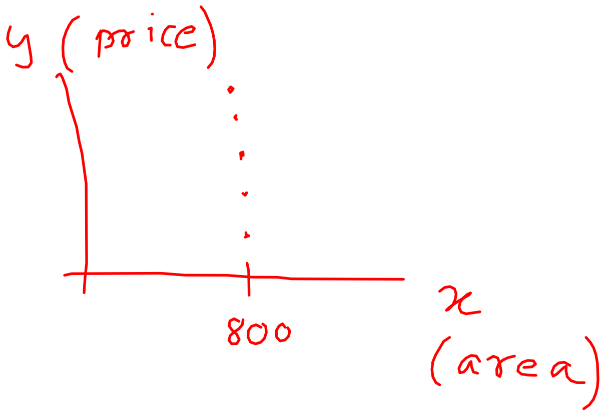
$$E(\hat{\beta}) = \beta$$

Assumptions :

linear model (in parameters) i.e. β's

random sample

variation in x



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

$= 0 \rightarrow x$ exogenous

$E(u|x) \neq 0 \rightarrow x$ endogenous

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

$$E(y|x) = \beta_0 + \beta_1 x$$

Variance of OLS estimators

Variance → precision

Assumption of homoskedasticity

$$Var(u|x) = Var(u) = \sigma^2$$

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

$$Var(y|x) = \sigma^2$$

$$\mathrm{var}(\hat{\beta}_1) = \frac{\sigma^2}{\text{variation in } x} = \frac{\sigma^2}{SST_x} = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\mathrm{sd}(\hat{\beta}_1) = \frac{\sigma}{\sqrt{SST_x}}$$

σ : unknown

unbiased estimator of σ^2 : $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{SSR}{(n-2)} = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{(\underbrace{n-2}_{\text{2 }\beta\text{'s estimated}})}$$

std. error of $\hat{\beta}_1$: $se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{SST_x}}$

$\hat{\sigma}$: standard error of regression