Econometrics. 2022-3-25.

Assumptions of linear regression model.

A1. Linearity: $y = X\beta + \varepsilon$

A2. Full rank of X: $\text{rank}(X) = k \quad (n > k)$

A3. Exogeneity: $E[\varepsilon_i | X] = 0$ for $i = 1, \cdots, n$

$\Rightarrow E[\varepsilon | X] = 0, \quad E(\varepsilon) = 0, \quad Cov(\varepsilon_i, X) = 0.$

$\Rightarrow E[y|X] = X\beta$ (with A1).

A4. Homoskedasticity : $Var[\varepsilon_i | X] = \sigma^2$ for $i = 1, \cdots, n$.

Non-autocorrelation : $Cov[\varepsilon_i, \varepsilon_j | X] = 0$ for $i \neq j$.

$\varepsilon_t, \varepsilon_s$

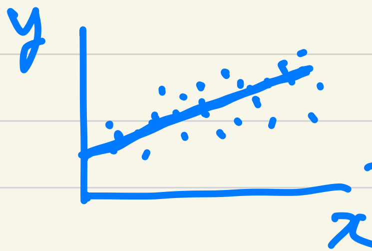
$$E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' | X] = \begin{bmatrix} \underbrace{E[\varepsilon_1\varepsilon_1 | X]}_{=\sigma^2} & \underbrace{E[\varepsilon_1\varepsilon_2 | X]}_{=0} & \cdots & \underbrace{E[\varepsilon_1\varepsilon_n | X]}_{=0} \\ \vdots & \vdots & \ddots & \vdots \\ \underbrace{E[\varepsilon_n\varepsilon_1 | X]}_{=0} & \underbrace{E[\varepsilon_n\varepsilon_2 | X]}_{=0} & \cdots & \underbrace{E[\varepsilon_n\varepsilon_n | X]}_{=\sigma^2} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 I$$

$\Rightarrow$ A4 can be written as $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' | X] = \sigma^2 I$.

Theorem B.4.

$$\mathrm{Var}[y] = \mathrm{Var}_x\big[E[y|x]\big] + E_x\big[\mathrm{Var}[y|x]\big]$$



$\mathrm{Var}[y] =$ 右面第1项

$+$ 右面第2项.

$$\mathrm{Var}[\varepsilon] = \mathrm{Var}_x\big[E[\varepsilon|X]\big] + E_x\big[\mathrm{Var}[\varepsilon|X]\big]$$

$$= \mathrm{Var}_x\big[\underbrace{E[\varepsilon|X]}_{=0}\big] + E_x\big[E[\varepsilon\varepsilon'|X] - \underbrace{E[\varepsilon|X]}_{=0}\underbrace{E[\varepsilon|X]}_{=0}\big]$$

$$= 0 + E_x\big[E[\varepsilon\varepsilon'|X]\big]$$

$$= E[\sigma^2 I] = \sigma^2 I$$

$$\mathrm{Var}[\varepsilon] = E[\varepsilon\varepsilon'] - \mu\mu'$$

A5. Data generation: X may be fixed or random.

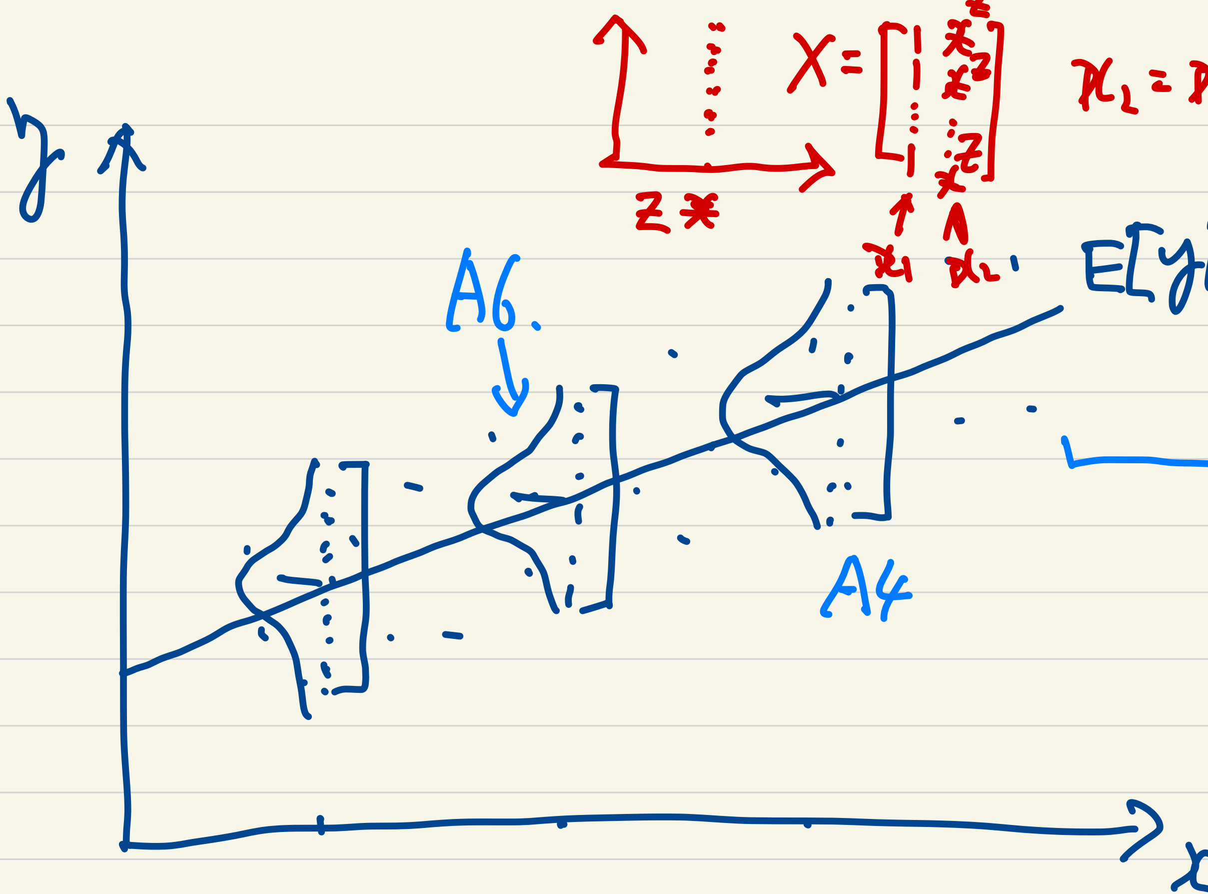
Fixed: from experiments. $E[\varepsilon|X] = E[\varepsilon] = 0$

$Var[\varepsilon] = \sigma^2 I$.

Random: observational data.

A6. Normality: $\varepsilon | X \sim N[0, \sigma^2 I]$

A3 A4.



$x_2 = x_1 .* z$

$E[y|x] = \underbrace{\alpha + \beta x}_{A1}$

A3

About independence.

(I) · Statistical independence : $f(x,y) = f_x(x) \cdot f_y(y)$.

(II) · Uncorrelatedness : $Cov(x, y) = 0$.

(III) · Mean independence : $E[y|x]$ does not depend on x.

· Linear independence : $X = [x_1, \cdots, x_K]$

$x_1, \cdots, x_K$ are not linearly dependent

$x_1 \neq \alpha_2 x_2 + \cdots + \alpha_K x_K$.

$$(I) \underset{\not\Leftarrow}{\Rightarrow} (III) \underset{\not\Leftarrow}{\Rightarrow} (II)$$

$$y = X\beta + \varepsilon$$

known. calculat β

Model fitting.

$$y_i = \hat{y}_i + e_i$$
$$= x_i' b + e_i$$

· Population regression function. unknown

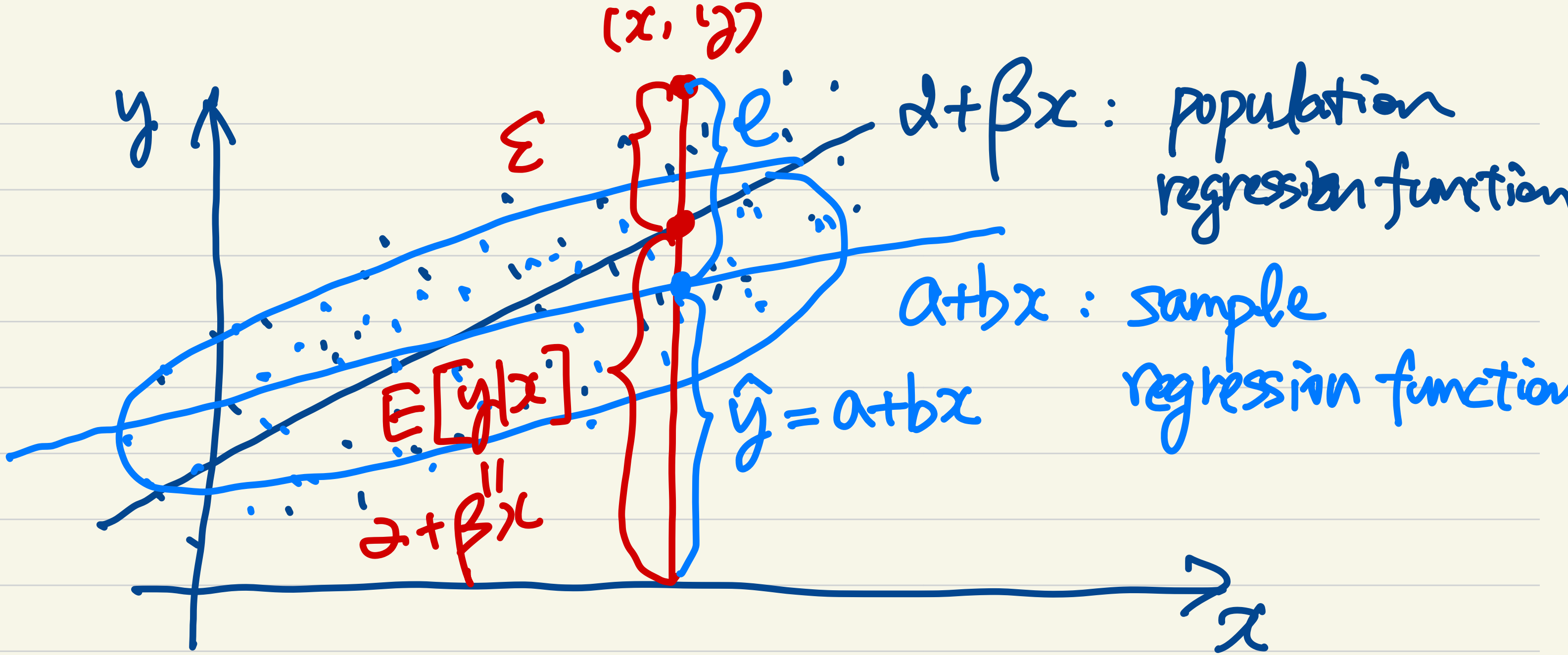
$$E[y_i \mid x_i] = x_i'\beta. \quad \varepsilon_i = y_i - x_i'\beta.$$

· Sample regression function

$$\hat{y}_i = x_i' b. \quad e_i = y_i - \hat{y}_i$$

residual.

known.



(y_i, x_i) is known. for $i = 1, \cdots, n$.

$\boldsymbol{\beta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ is unknown.

$\mathbf{b} = \begin{bmatrix} a \\ b \end{bmatrix}$ is obtained from $[\mathbf{y} . X]$.

Least Squares.

- The least squares fitting criterion.

$$\min_{\mathbb{b}_0} S(\mathbb{b}_0) = \underline{\underline{\mathbb{e}_0'\mathbb{e}_0}} = (\mathbb{y} - X\mathbb{b}_0)'(\mathbb{y} - X\mathbb{b}_0).$$

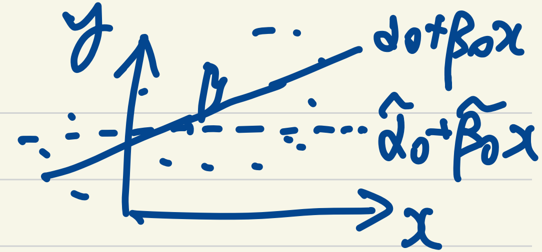
where $\mathbb{e}_0 = \mathbb{y} - X\mathbb{b}_0$. $\mathbb{e}_0'\mathbb{e}_0 = [e_{0,1} \cdots e_{0,n}]\begin{bmatrix} e_{0,1} \\ \vdots \\ e_{0,n} \end{bmatrix}$

$$= \sum_{i=1}^{n} e_{0,i}^2$$

sum of squared residuals.

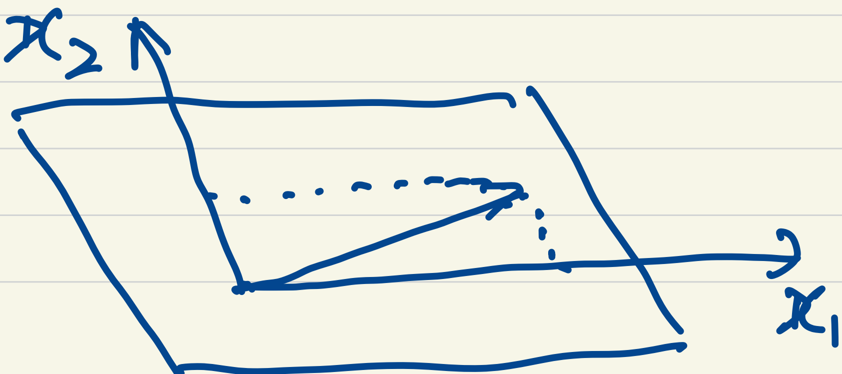
$$\mathbb{b} = \underset{\mathbb{b}_0}{\operatorname{argmin}}\, S(\mathbb{b}_0)$$

$$\mathbf{y} = X\mathbf{b}_0 + \mathbf{e}_0$$



$$X = [x_1, \cdots, x_k]$$

$$X\mathbf{b}_0 = [x_1, \cdots, x_k]\begin{bmatrix} b_{0,1} \\ \vdots \\ b_{0,k} \end{bmatrix} = \sum_{m=1}^{k} x_m \cdot b_{0,m}$$



$X = [x_1, x_2]$