22.01.2024 **Econometrics-II** Lecture 4

⇒ Simple regression model

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

$i = 1, 2, \ldots, N$

Limitations

1. Linearity between x_i and y_i
2. Causality is not ensured (most important)
   ⇝ we need a key assumption to ensure causal relationships.

⇒ Assumptions on the Simple linear regression model

A1. $E(u_i) = 0$

Expectation operator

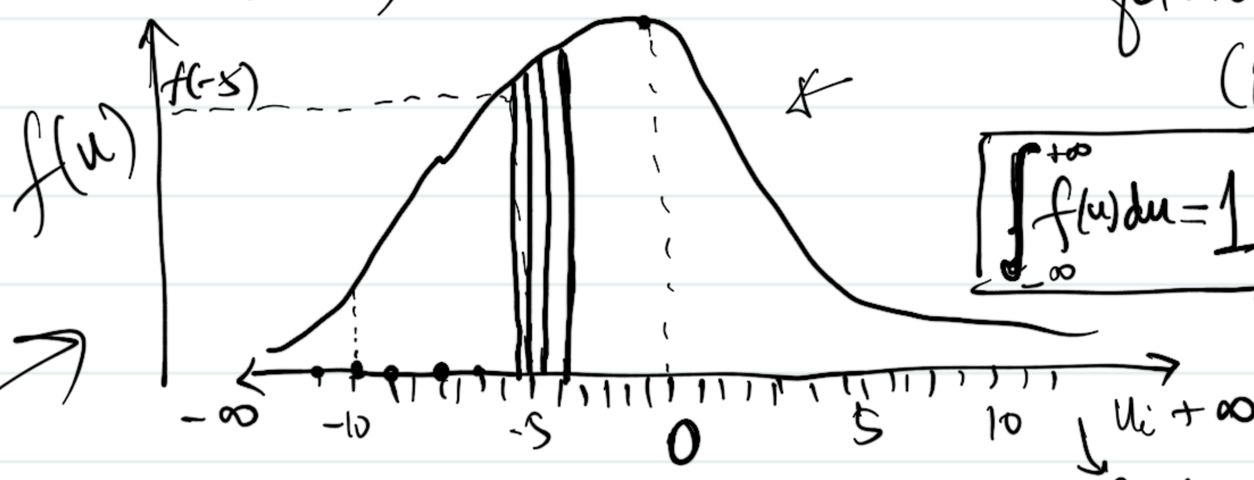
p.d.f.

* u_i is a random error. $\Rightarrow u_i \sim f(u)$

A1. $E(U_i) = 0$

* $U_i \sim f(u)$ where $f(\cdot)$ is a probability density function. (pdf)

U_i ↑ random variable (error)

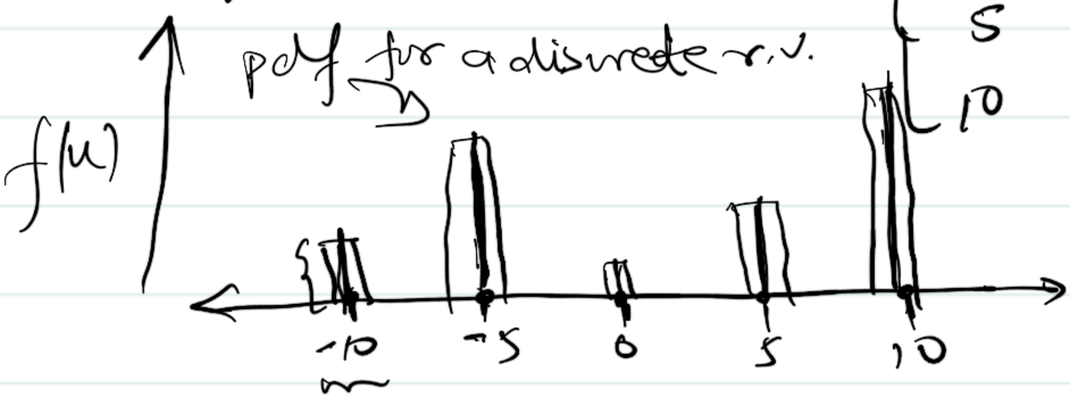


$$\boxed{\int_{-\infty}^{+\infty} f(u)\,du = 1}$$ → Area under the f(u) graph

U_i is a continous random variable

→ What if U_i is a discrete random variable? pdf?

$$U_i \notin \mathbb{R} \quad ; \quad U_i = \begin{cases} -10 & \text{with } f(-10) = p_1 = 0.1 \\ -5 & \text{with } p_2 = 0.3 \\ 0 & \text{with } p_3 = 0.05 \\ 5 & \text{w/ } p_4 = 0.15 \\ 10 & \text{w/ } p_5 = 0.4 \end{cases}$$



$$\boxed{\sum_i p_i = 1}$$

A1. $E(u_i) = 0$ (w.l.o.g)

without loss of generality.

$u_i \sim f(u)$

what does expectation of random error imply?

for the discrete case: $u_i = \begin{cases} -10 \text{ w/ } p_1 \\ -5 \text{ w/ } p_2 \\ 0 \text{ w/ } p_3 \\ 5 \text{ w/ } p_4 \\ 10 \text{ w/ } p_5 \end{cases}$

$$E(u_i) = \sum_i u_i p_i$$

for continuous r.v.:

$$\rightarrow E(u_i) = \int_{-\infty}^{\infty} u_i f(u_i)\, du$$

yields, fertiliz, soil i

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

w.l.o.g assumption : $E(u_i) = 0 \implies$

when β_0 is included in the SLR model then $E(u_i) = 0$

$$yield_i = \beta_0 + \beta_1 ferti_i + u_i$$

$$E(yield_i) = E[\beta_0 + \beta_1 ferti_i + u_i]$$

$$E(yield_i) = E(\beta_0) + E(\beta_1 ferti_i) + E(u_i)$$

$$E(yield_i) = \underline{\beta_0} + \beta_1 E(ferti_i) + \underbrace{E(u_i)}_{=\boxed{a}\ (say)}$$

No difference b/w
$E(u_i) = a$ and the const. β_0
so far as our SLRM is concerned

$\therefore$ We can assume that $E(u_i) = 0$

$$\beta_0' = \underline{\beta_0 + a}$$

$$E(u_i) = 0 \checkmark$$

A2. The crucial assumption for causality

$$E(\underbrace{u_i}\mid x_i) = 0$$

Expected value of u_i given x_i is zero (constant)

$\Rightarrow$ Expected value of u_i does not depend on x_i

$$E(u_i \mid x_i) = E(u_i) = 0$$

$$\int_{-\infty}^{+\infty} u_i \underbrace{f(u_i \mid x_i)} du = \int_{-\infty}^{+\infty} u_i f(u_i) du.$$

$$= \frac{f(u_i, x_i)}{f(x_i)}$$

independence of u_i & x_i

$$= \frac{f(u_i) \cdot \cancel{f(x_i)}}{\cancel{f(x_i)}}$$

$\Rightarrow$ Implication of A_2 is u and x are independent.

$\Rightarrow$ x_i and u_i are uncorrelated

$$Corr(x_i, u_i) = 0$$

$$Corr(x_i, u_i) = \frac{Cov(x_i, u_i)}{sd(x_i)\, sd(u_i)} = 0$$

$$Cov(x_i, u_i) = 0$$

$$Cov(x_i, u_i) = E\Big((x_i - E(x_i))(u_i - E(u_i))\Big)$$

Aside

$$V(u_i) = E(u_i - E(u_i))^2 = \underline{E(u_i^2) - [E u_i]^2}$$

$$\parallel$$

$$Cov(u_i, u_i)$$

$$= E\Big[u_i^2 + (\underbrace{E(u_i)}_{\text{const.}})^2 - 2\, u_i \underbrace{E(u_i)}_{\text{const.}}\Big]$$

$$= E[u_i^2] + [E[u_i]]^2 - 2E(u_i)\, E(u_i)$$

$$\underline{E[u_i^2] - [E(u_i)]^2} \Leftarrow \; = E[u_i^2] + (E[u_i])^2 - 2\,(E(u_i))^2$$

$$\mathrm{Cor}(u_i, x_i) = E\left(\left[u_i - E(u_i)\right]\left[x_i - E(x_i)\right]\right)$$

$$= E\left[u_i x_i + E(u_i)E(x_i) - x_i E(u_i) - u_i E(x_i)\right]$$

$$= E[u_i x_i] + E(u_i)E(x_i) - E(u_i)E(x_i) - E(x_i)E(u_i)$$

$$= E[u_i x_i] - \underbrace{E(u_i)}_{=0\ (\because A1)} \cdot E(x_i)$$

$$= E[u_i x_i]$$

$$A2 \Rightarrow \boxed{E[u_i x_i] = 0}$$

$$A2: \quad E(u|x) = 0$$ → allows us to simulate the ceteris paribus experiment (when true)

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

$\Rightarrow$ x and u are independent / orthogonal to each other

$$\Rightarrow E(ux) = 0$$

$$\Rightarrow \text{Corr}(u, x) = \text{Cov}(u, x) = 0$$

$$yield_i = \beta_0 + \beta_1 \, fert_i + \underline{u_i}$$

$$wage_i = \gamma_0 + \gamma_1 \, educ_i + u_i$$

u_i → ability of farmer/employee.

$$E[abil. \mid educ = 10] = E[abili \mid educ = 12]$$

$$\ldots\ldots = 0$$

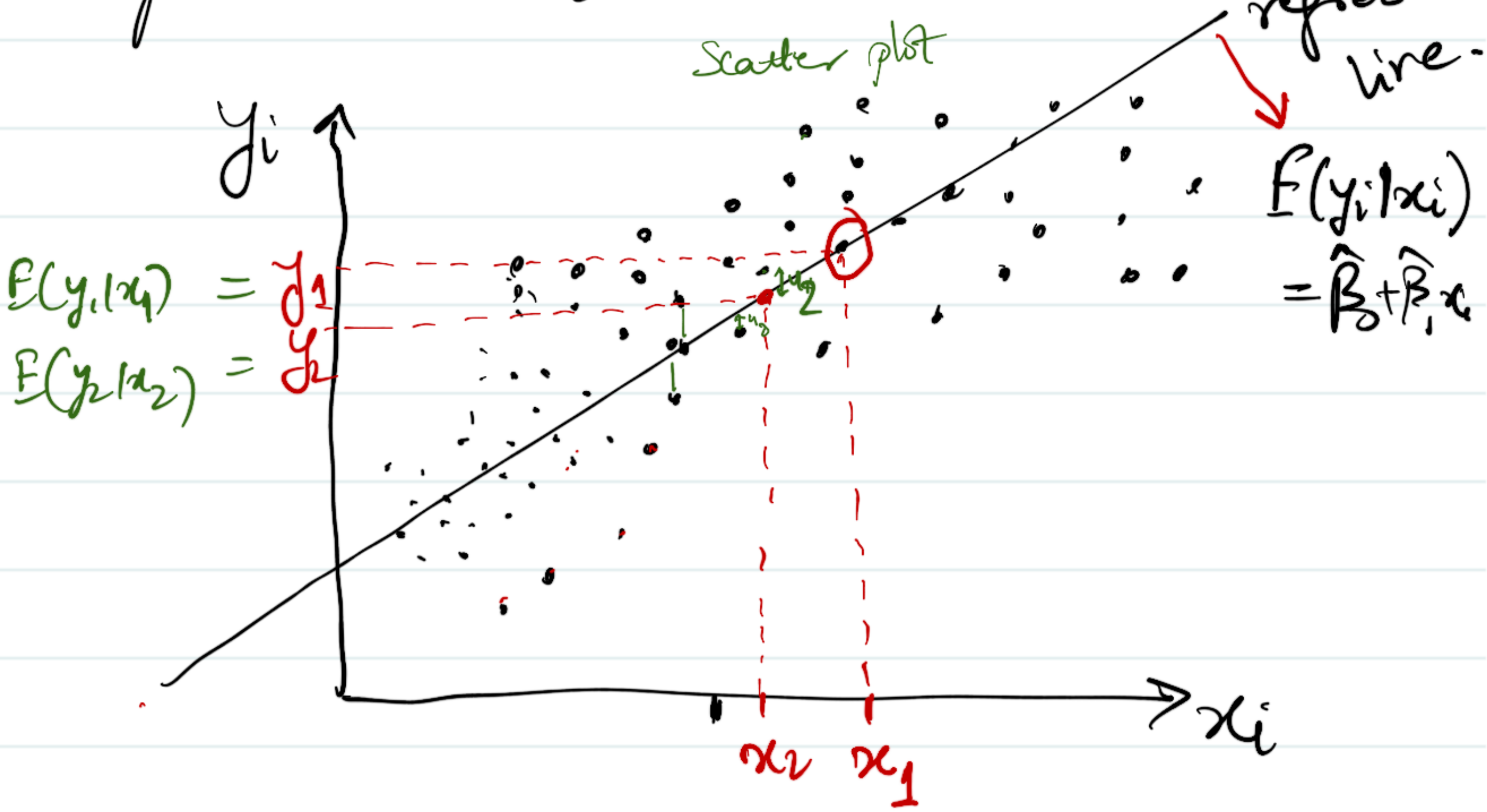
Simulated the ceteris paribus.

→ Estimating β_0 and β_1 from the given data:

$$\boxed{y_i = \beta_0 + \beta_1 x_i} + u_i \quad \text{(Model)}$$

data → $\{y_i, x_i\}_{i=1}^{N=1,000,000,000,000}$

β_0 and β_1 are to be estimated from the given data!!



$$E(y_i) = E(\beta_0 + \beta_1 x_i + u_i) \xrightarrow[E(u_i)=0]{A1} E(y_i|x_i) = \beta_0 + \beta_1 x_i$$