CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

AACSB ACCREDITED · EFMD EQUIS ACCREDITED · EFMD EPAS ACCREDITED

MSF Chula

Financial Econometrics

Lecture II

x & y are in form of linear relationship

Introduction to Linear Regression & Bivariate OLS Estimator

Narapong Srivisal, Ph.D.

Outline

Bivariate Linear Regression Model

in the population
that we want to sample

- Three Interpretations
- Non-linear random variables in linear regression
- Solving for β_0 and β_1

Estimator

- Ordinary Least Squares Estimator (OLS)
- Properties of the OLS estimator
  - Biasedness
  - Consistency

know if moves relative to mean

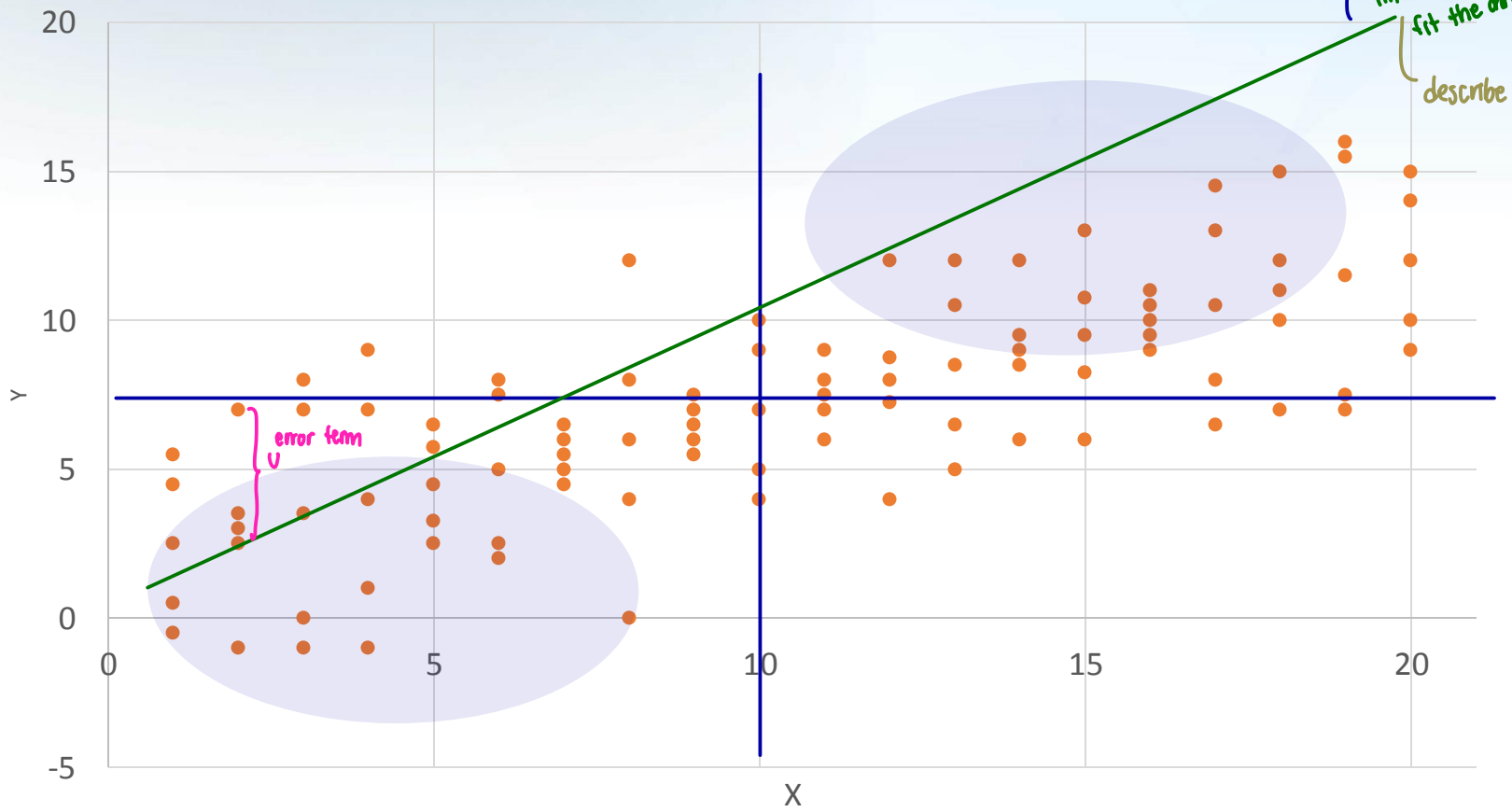
Scatter plot : relationship btw x, y | covariance = ⊕

when x is below mean, y is also below mean

above above

want to estimate this linear equation fit the data

describe relationship



Bivariate Linear Regression Model

- Model: let (Y, X, U) be a random vector and β_0, β_1 be parameters such that

y-intercept rotate line → error term

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$

parameters → we know these then we know the line model completely

- Y is referred to as the **regressand** or dependent variable or explained variable ↳ value of y depends on value of x
- X is referred to as the **regressor** or independent variable or explanatory variable
- U is referred to as the error term
- β_0 is sometimes referred to as the intercept parameter
- β_1 is sometimes referred to as the slope parameter

to be estimated

linear line

- We can interpret this relationship in three ways:
  - linear conditional expectation
  - linear approximation to conditional expectation
  - causal relationship.

LINEAR CONDITIONAL EXPECTATION

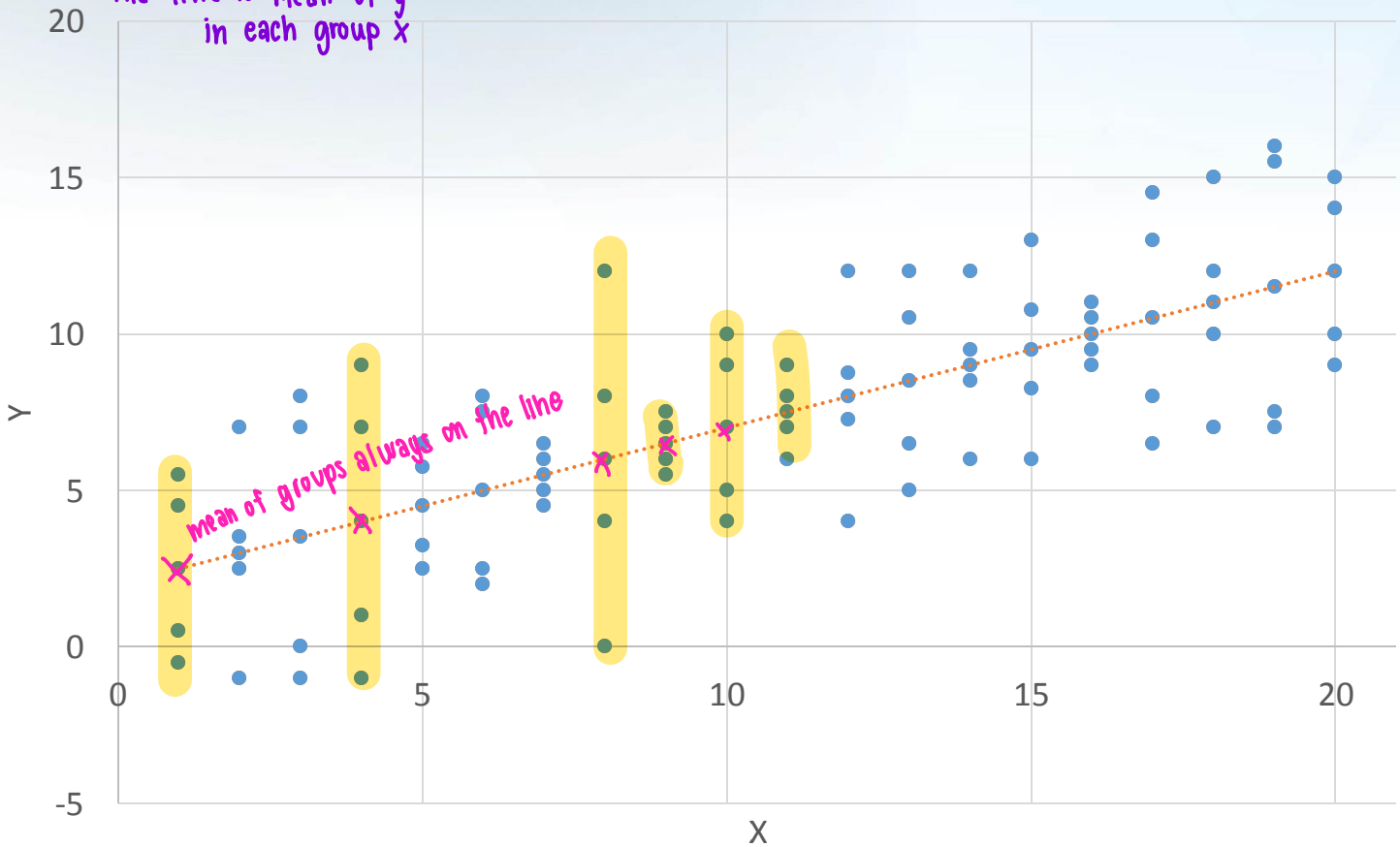
1 Expectation of y

1st interpretation

the line is mean of y in each group x

mean of y for each group X

$$E[Y|X] = 0.5X + 2$$



$Y_i = \beta_0 + \beta_1 X_i + U$

If $E[Y|X] = \beta_0 + \beta_1 X \Rightarrow U = Y - E[Y|X]$

Linear Conditional Expectation

line captures mean of Y given X

- Assume that $E[Y|X] = \beta_0 + \beta_1 X$ and define $U = Y - E[Y|X]$
- This interpretation implies that

$E[E[U|X]] = E[U]$

$E[0] = 0$

$$E[U|X] = E[Y - E[Y|X]|X] = E[Y|X] - E[Y|X] = 0$$

$E[U] = 0$

$E[U|X] = E[U] \longrightarrow cov(U, X) = 0$

mean independent

Property

$E[XY] = E[X] \cdot E[Y] + cov(X, Y)$

$E[XY] = 0$, $E[Y] \to 0$, $cov(X,Y) \to 0$

- Which implies $E[U] = 0, Corr[X, U] = 0, E[XU] = 0$

$\frac{cov(X,U)}{\sigma_U \sigma_X} \to 0$

- In this case, β_0 and β_1 provide a convenient way of summarizing the conditional expectation.
- X does NOT cause Y in this interpretation.

constant because x is constant from conditional on X

$$E[y|X] = E[\overbrace{\beta_0 + \beta_1 X}^{\text{constant because } x \text{ is constant from conditional on } X} + u \mid X]$$

$$\cancel{E[y|X]} = \cancel{\beta_0 + \beta_1 X} + E[u|X] \quad ; \quad E[y|x] = \beta_0 + \beta_1 X$$

If we believe linear line intepret mean of y in each group of x

$$E[u|X] = 0 \Rightarrow E[u] = 0$$

$$\Rightarrow \underbrace{\mathrm{cov}(u, x)}_{E[xu] - E[x]\cancel{E[u]}^{0}} = 0$$

this should also = 0

law of iterated expectation

$$E[E(u|x)] = E[0]$$

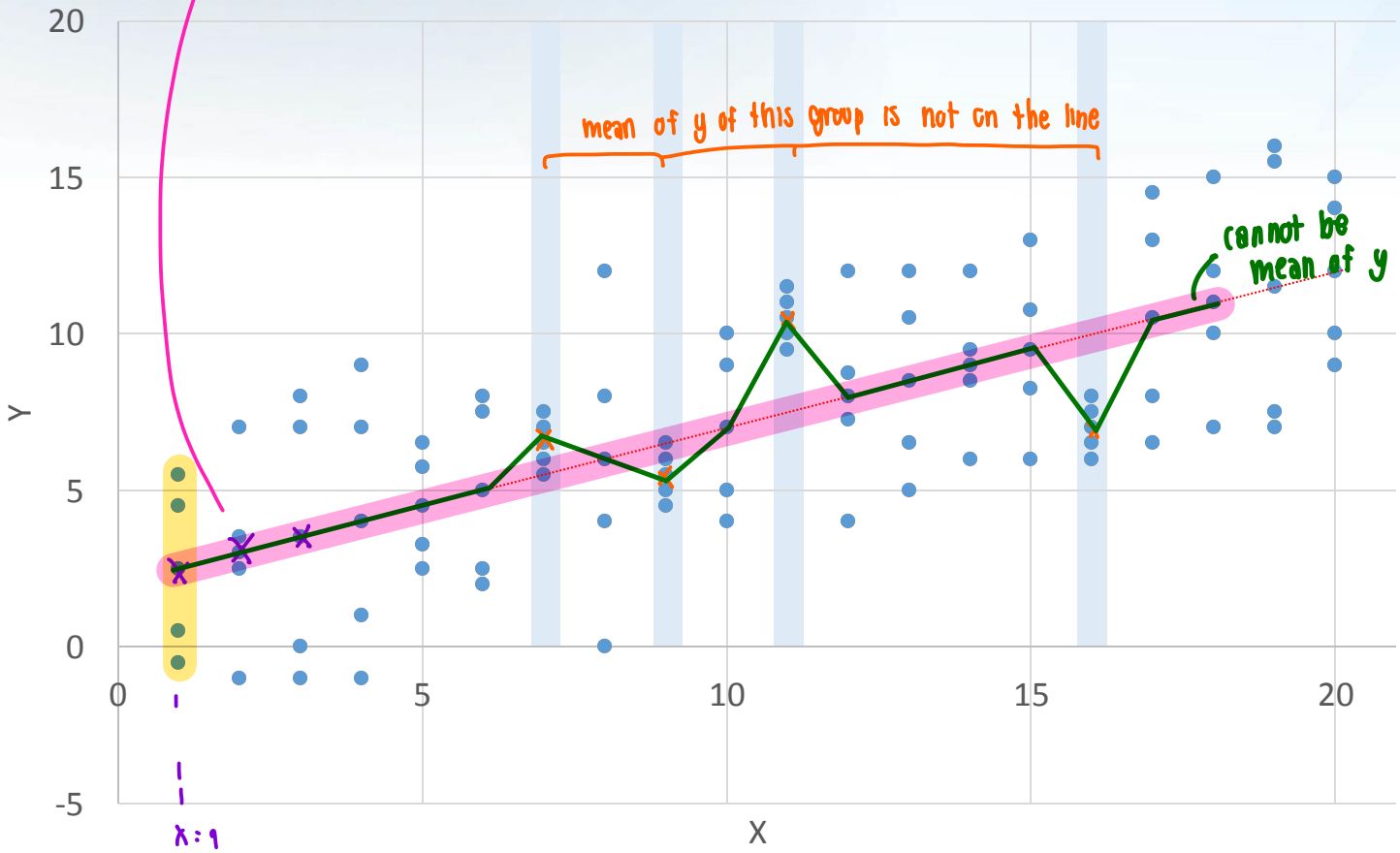
$$E[u] = 0$$

II Intepretation or assumption

Best linear approximation, prediction, projection

The line is the "best" or "closest" to the mean

BLP[Y|X] = 0.5X + 2



Best Linear Prediction of $E[Y|X]$ or of Y

- In many cases, it's likely that the conditional expectation $E[Y|X]$ is a non-linear function of X, but one may ask what is the "best" linear prediction or linear approximation of $E[Y|X]$
- So, we **pick β_0 and β_1 to minimize**

mean that we want line to be closer to

$$E[(E[Y|X] - \beta_0 - \beta_1 X)^2]$$

use β_0 & β_1 that minimize this

average square of distance btw mean and the line ⇒ want to minimize this

- In this case, β_0 and β_1 provide a convenient way of summarizing the "best" *linear* function of X to predict Y, although $E[Y|X]$ may not be linear.
- Again, X does NOT cause Y in this interpretation.
- If we interpret the relationship this way, it can be showed that

$$E[U] = E[XU] = 0$$

$\cancel{E[XU]}^{0} = E[X] \cdot \cancel{E[U]}^{0} + \cancel{Cov(X,U)}^{0}$

min β_0, β_1

Objective : to minimize

- Proof: we can write $E\{[E[Y|X] - (\beta_0 + \beta_1 X)]^2\}$

$$= E\{[(E[Y|X] - Y) + (Y - \beta_0 - \beta_1 X)]^2\}$$

$u^2 + 2ua + a^2$

$$= E\{(E[Y|X] - Y)^2\} + 2E\{(E[Y|X] - Y)(Y - \beta_0 - \beta_1 X)\} + E\{(Y - \beta_0 - \beta_1 X)^2\}$$

- Consider the second term $E\{(E[Y|X] - Y)(Y - \beta_0 - \beta_1 X)\}$ is

$$= E\{(E[Y|X] - Y)Y\} - \beta_0 E\{(E[Y|X] - Y)\} - \beta_1 E\{(E[Y|X] - Y)X\}$$

$$= E\{(E[Y|X] - Y)Y\} - \beta_0 \left(\underbrace{E[E[Y|X]]}_{=E[Y]} - E[Y]\right) - \beta_1 \left(\underbrace{E[XE[Y|X]]}_{=E[XY]} - E[XY]\right)$$

$$= E\{(E[Y|X] - Y)Y\}$$

- So, $E\{[E[Y|X] - (\beta_0 - \beta_1 X)]^2\}$ can be written as

diff with respect to β_0 or β_1 → this become zero

can focus on this only

$$E\{(E[Y|X] - Y)^2\} + 2E\{(E[Y|X] - Y)Y\} + \boldsymbol{E\{(Y - \beta_0 - \beta_1 X)^2}$$

min β_0, β_1

- Note that only the last term depends on the parameters β_0 and β_1.

maximization problem

Best linear prediction / projection / approximation

$$\min_{\beta_0, \beta_1} E[(E(Y|X) - (\beta_0 - \beta_1 X))^2]$$

Best linear approximation of mean of y

⇒ different btw mean to line

$$\min_{\beta_0, \beta_1} E[(y - (\beta_0 + \beta_1 X))^2]$$

Best linear approximation of y

exactly the same

⇒ different btw Y overall to line

↳ called BLP (y|x)

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

try to minimize error

$$\min_{\beta_0, \beta_1} E[(Y - \beta_0 - \beta_1 X)^2] = E[u^2]$$

$u = Y - \beta_0 - \beta_1 X$

differentiale with respect to

$[\beta_0]$: $E[\underbrace{2u}_{\text{antiderivative}} \cdot \underbrace{(-1)}_{\text{diff inside}}] = 0$

$[\beta_1]$: $E[2u \cdot (-x)] = 0$

$E[u] = 0$

$E[ux] = 0$

we have these two moment conditions automatically

↳ use to prove whether the estimator is make sense or not

$\Downarrow$

$E[ux] - E[u]E[x]$

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

→ ingredient for the prove for unbiase & consistency

Moment conditions

make OLS unbiased, satisfy in 3st interpretation only

$\boxed{E[u|x]=0}$

$E[ux]=0$

$E[u]=0$

OLS consistant

$E[ux]=0$

$E[u]=0$

x & y move together

not causal relationship

3 Interpretations

1) Mean of y|x ⟹

2) Approximation (Best) of y|x ⟹

3) Causal relationship

$E[u]=0$

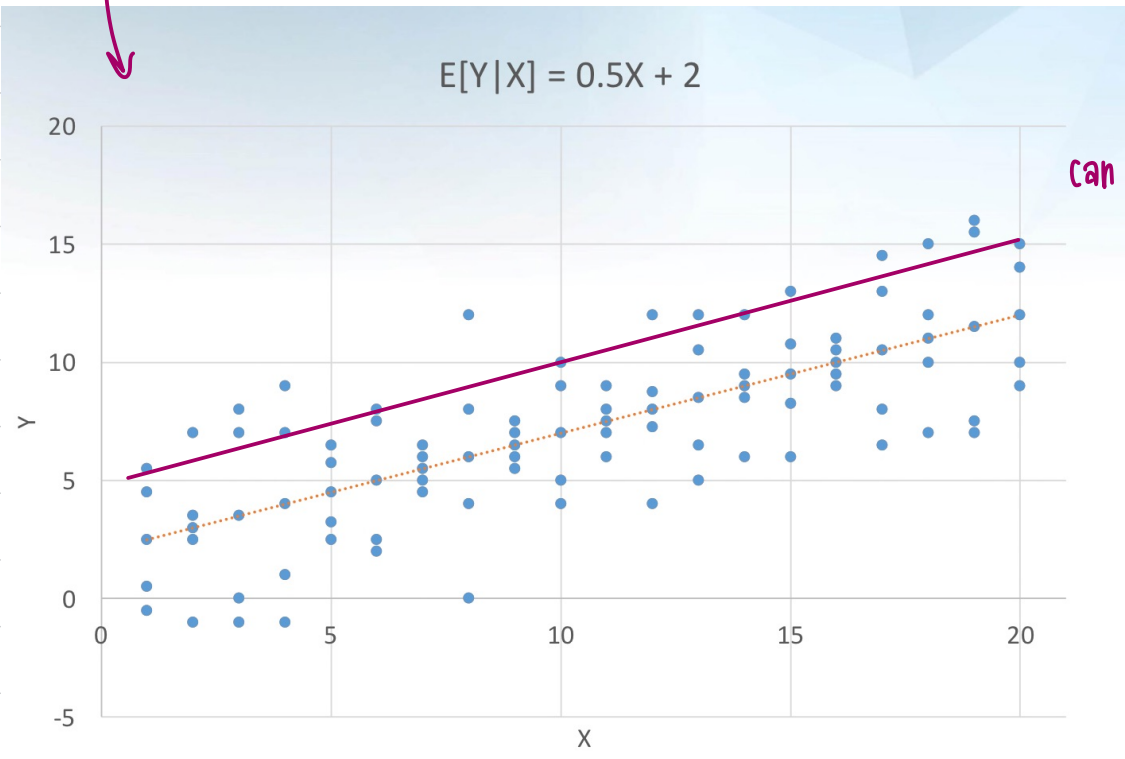
$y = \beta_0 + \beta_1 x$

pure effect of x

→ if you want to use OLS estimator don't have any statistical property

have to justify $\boxed{Cov(u,x)=0}$

have only $E[u]=0$



can always change intercept

- Therefore, **pick β_0 and β_1 to minimize $E[(E[Y|X] - \beta_0 - \beta_1 X)^2]$ is the same problem as**

$$\min_{\beta_0,\beta_1} E[(\overbrace{Y - \beta_0 - \beta_1 X}^{U})^2]$$

- In words, we can interpret the linear line in this case as the best linear prediction/approximation/projection either of the conditional mean $E[Y|X]$ or simply of the variable Y.

- The FOCs are

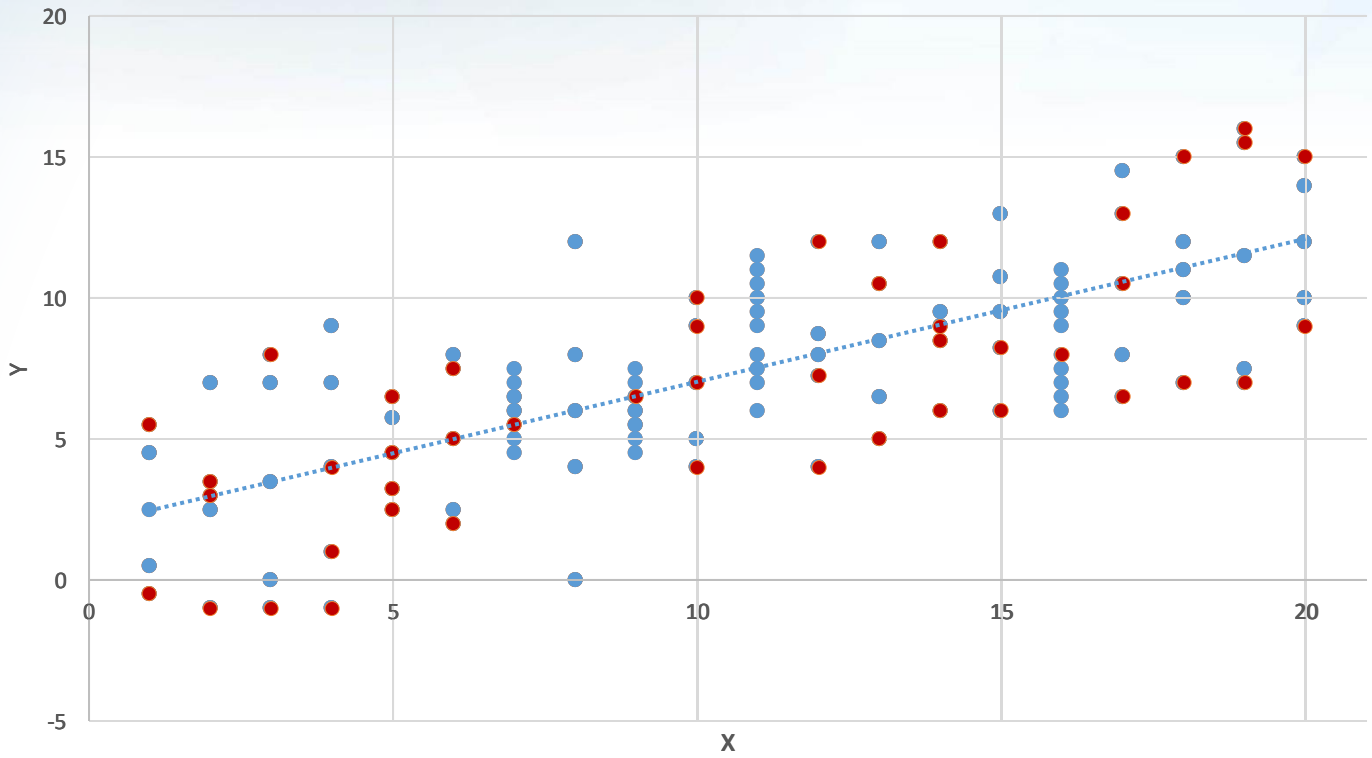
$$[\beta_0]: E[-2(Y - \beta_0 - \beta_1 X)] = 0$$

$$[\beta_1]: E[-2X(Y - \beta_0 - \beta_1 X)] = 0$$

- Since $U = Y - \beta_0 - \beta_1 X$, this interpretation then implies

$$E[U] = E[XU] = 0$$

Remark: We have to "interpret" or assume the relationship because we only observe a sample, not the whole population.



Causal Model

- In many cases, we assume that Y is a linear function of observed variable X and unobserved determinant U

$$Y = g(X, U) = \beta_0 + \beta_1 X + U$$

- Generally, making such assumption has to be based on some sort of "models" or "theories"
- X causes Y in this interpretation, and β_1 describes the change in Y when X changes by one unit, **given other factors constant.**

ทำให้บอกได้เลยว่า x เปลี่ยนทำให้ y เปลี่ยน

- We can always normalize β_0 to have $E[U] = 0$, because if $E[U] = c$, we can define another model:

so, we can change intercept

$$Y = \underbrace{\beta_0 + c}_{\tilde{\beta}_0} + \beta_1 X + \underbrace{U - c}_{\tilde{U}}$$

- In this interpretation, **it's not necessary that $E[XU] = 0$ or $E[U|X] = 0$**

Example

- Let R_i be return of stock i, and X_i be net buy volume of stock i.

$$R_i = \beta_0 + \beta_1 X_i + U_i$$

correlation

- **Interpretation 1**: If we classify all the stocks into groups by their net buy volumes, the group that has net buy volume equal to X has mean returns equal to $\beta_0 + \beta_1 X$. In other words, the expected value of returns for stocks that have net buy volumes equal to X is $\beta_0 + \beta_1 X$.
- **Interpretation 2**: $\beta_0 + \beta_1 X$ is the "best" linear prediction of returns or $E[R|X]$, which may be non linear. In other words, if we see a stock with net buy volume equal to X, we could predict that its return or expected return would be $\beta_0 + \beta_1 X$, although the actual expected value of returns may not be $\beta_0 + \beta_1 X$.

→ doesn't expect x to have effect on y

predictor, forecast

- In this case, it is like we use X as a signal or indicator for returns, but X does not necessarily have any effect or impact on returns.

or determinant

, e.g. effectiveness of policy

Causation

Causation → if we want to make return (y) higher → have to adjust volume (x) ← want the effect of it

- **Interpretation 3**: If we make net buy volume increases by one unit, the stock return will rise by β_1 unit. In this case, X has an impact on returns.

Return = β_0 + 0.1 Volume + U

THB mm

1% 2%

10 million 20

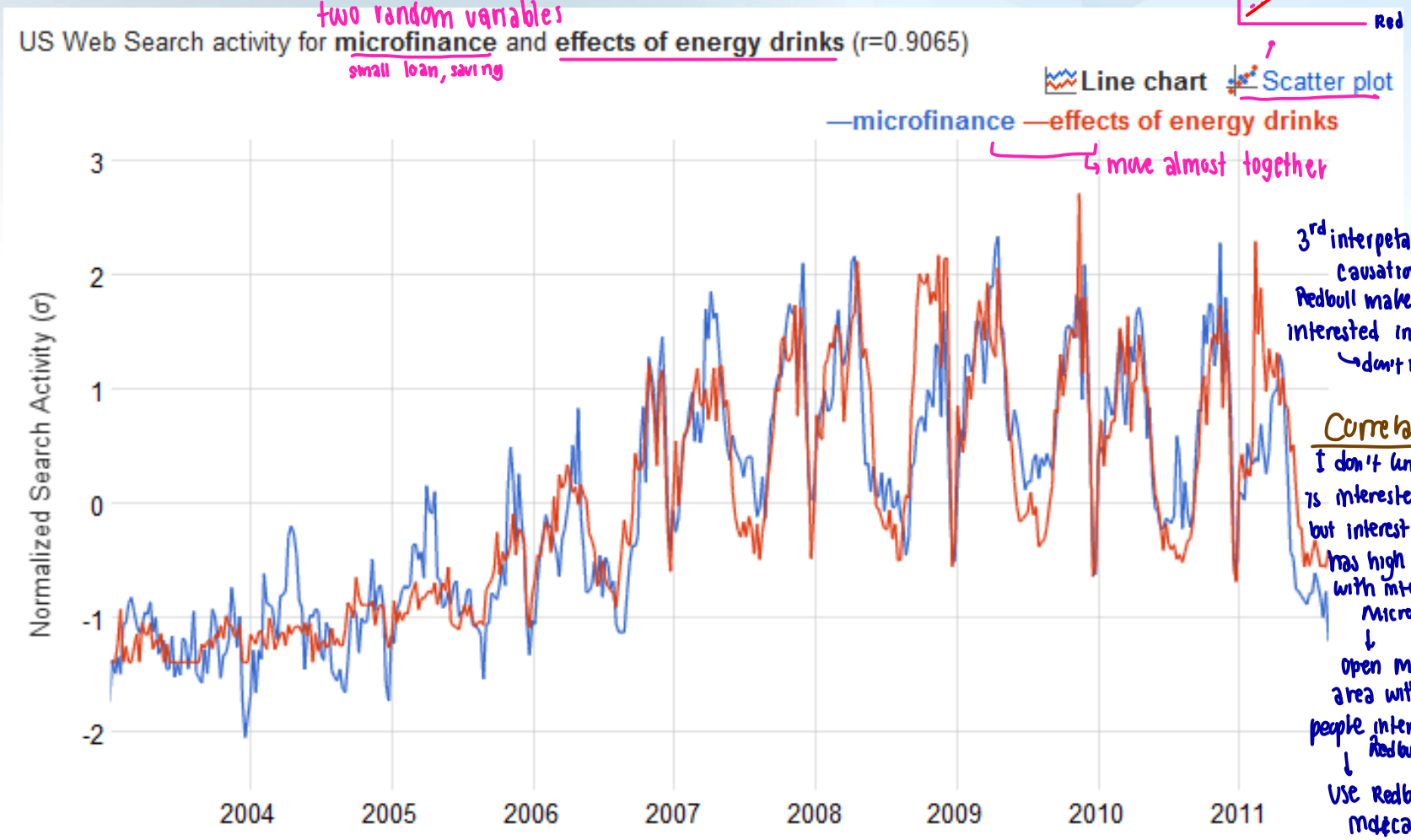
Return = 2 + 0.1 volume + U

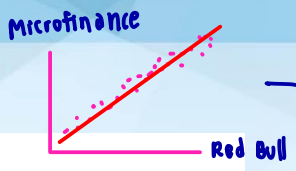
Pick a stock with approximate expected return

1 mm → 2.1

2 mm → 2.2

3 mm → 2.3

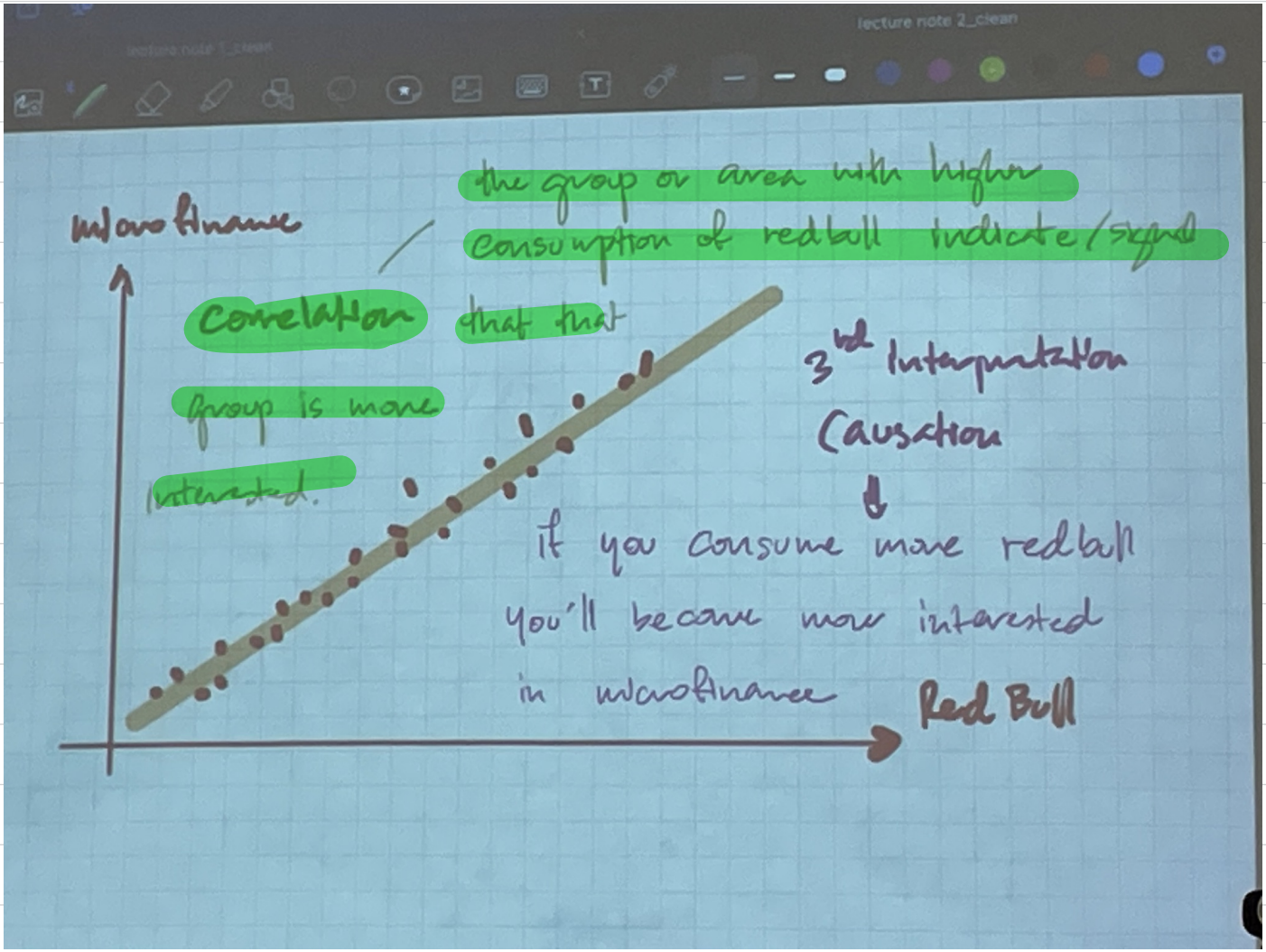




3rd interpetation
Causation
Redbull make you interested in microfin
↳ don't make sense

Correlation
I don't know who is interested in microfin but interest in redbull has high correlation with interest in microfin
↓
open microfin in area with lot of people interested in Redbull
↓
use Redbull as indicator

* Source: "Promise and Pitfalls of Digital Financial Services" by Shawn Cole, Aug 17, 2015


lecture note 2_clean
Macrofinance
the group or area with higher consumption of redbull indicate/signal that that group is more interested.
Correlation
3rd Interpretation
Causation
if you consume more redbull you'll become more interested in macrofinance
Red Bull

Nonlinearity in Linear Regression

- Certain nonlinear models can be modified to fit in the linear regression framework by defining the regressor and/or regressand as non-linear function of random variables.
- E.g. Let $X = \tilde{X}^2$. Hence, the model becomes

$$Y = \beta_0 + \beta_1 \tilde{X}^2 + U$$

- E.g. Let $Y = \ln \tilde{Y}$. Then, the model becomes

$$\ln \tilde{Y} = \beta_0 + \beta_1 X + U$$

If we use the third interpretation (causal relationship), then

$$\beta_1 = \frac{\partial \ln \tilde{Y}}{\partial X} = \frac{1}{\tilde{Y}} \frac{\partial \tilde{Y}}{\partial X} \approx \frac{\left(\frac{\Delta \tilde{Y}}{\tilde{Y}}\right)}{\Delta X}$$

- This means that, when X increases by one unit or $\Delta X = 1$, $100\beta_1$ is the percentage change in $\tilde{Y}$.

Linear model

Can insert r.v. function because function of r.v. also is r.v.

$$\underbrace{\boxed{y}}_{\text{regressand r.v.}} = \beta_0 + \beta_1 \underbrace{x^2}_{\text{regressor r.v.}} + u$$

we have to interpret in x not x^2

y is linear in x^2 but not linear in x

$$dy = \boxed{2\beta_1 x}\, dx$$

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

What is the effect of x on y?

$$\frac{dy}{dx} = \beta_1$$

$$dy = \beta_1 \cancel{dx}^{\,1}$$

$x\uparrow$ 1 unit $\rightarrow$ $y\uparrow$ β_1 unit

E.g. $\ln y = \beta_0 + \beta_1 x + u$

impact of x on y

$$100\left(\frac{1}{y}\, dy\right) = 100\left(\beta_1 \cancel{dx}^{\,1}\right)$$

this change by β_1

$$100\left(\frac{y^{now} - y^{old}}{y^{old}}\right)$$ rate of change in y %

$\therefore$ % change in y is $100\,\beta_1$ %

$x\uparrow$ 1 unit then $y\uparrow 100\,\beta_1$ %

ln → interpret it in % change

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

$$\cancel{100}\, dy = \frac{\beta_1}{100} \left(\frac{1}{x} dx \cancel{100} \right)$$

1%

% change in x

x ↑ 1% → a 100 times change in y = β_1

x ↑ 1% ⇒ y ↑ by $\frac{\beta_1}{100}$ unit

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

$$\left(100 \frac{dy}{y} \right) = \beta_1 \left(\frac{dx}{x} 100 \right) \Rightarrow \frac{\left(100 \frac{dy}{y} \right)}{\left(\frac{dx}{x} 100 \right)} = \beta_1$$

elasticity

x ↑ 1%

y ↑ β_1

Price elasticity of demand (p, q^d) ⇒ price change how much demand respond

Income elasticity of demand (y, q^d)

price is x

price change first

random variable

$$\ln q^d = \beta_0 + \beta_1 \ln p + u$$

price elasticity of demand

or elasticity of demand with respect to price

monthly salary — work exp in year

① $wage = \beta_0 + 100\, exp + u$ → ↑1 yr of exp → wage ↑ 100 THB > based on population

which one is better

② $\ln wage = \beta_0 + 0.1\, exp + u$ → " → wage ↑ 10%

it based on based salary

$wage = \beta_0 + 10{,}000 \ln(exp) + u$ → stay 1% with company → get 100 THB X doesn't make sense

longer

$\ln wage = \beta_0 + 1 \ln(exp) + u$ → " → 1% increase in wage X regressive

increase in wage

① $d\,wage = 100\, d\,exp$

② $100\left(\frac{d\,wage}{wage}\right)\% = 100(0.1)\%\, d\,exp$

③ $d\,wage = \frac{\cancel{10{,}000}^{100}}{\cancel{100}^{100}}\left(\frac{d\,exp}{exp}\right)\%$

④ $100\left(\frac{d\,wage}{wage}\right)\% = 1\, 100\left(\frac{d\,exp}{exp}\right)\%$

- E.g. Let $Y = \ln \tilde{Y}$ and $X = \ln \tilde{X}$. Then, the model becomes

$$\ln \tilde{Y} = \beta_0 + \beta_1 \ln \tilde{X} + U$$

If we use the third interpretation (causal relationship), then

$$\beta_1 = \frac{\partial \ln \tilde{Y}}{\partial \ln \tilde{X}} = \frac{\left(\frac{1}{\tilde{Y}} \partial \tilde{Y}\right)}{\left(\frac{1}{\tilde{X}} \partial \tilde{X}\right)} \approx \frac{\left(\frac{\Delta \tilde{Y}}{\tilde{Y}}\right)}{\left(\frac{\Delta \tilde{X}}{\tilde{X}}\right)}$$

- This means that β_1 is the percentage change in $\tilde{Y}$ when $\tilde{X}$ increases by one percent. In this case, β_1 can also be thought of as the "**elasticity** of $\tilde{Y}$ with respect to $\tilde{X}$" or "$\tilde{Y}$ elasticity of $\tilde{X}$".

Solving for the Parameters β_0, β_1

- For now, we assume the following conditions
  - $Y = \beta_0 + \beta_1 X + U$
  - $E[U] = 0$
  - $E[XU] = 0$
  - $0 < Var[X] < \infty$

- Solving for β_0 and β_1:

$$E[U] = 0 \quad \Rightarrow E[Y - \beta_0 - \beta_1 X] = 0$$

$$\Rightarrow \boldsymbol{\beta_0 = E[Y] - \beta_1 E[X]}$$

$E[Y] - \beta_0 - \beta_1 E[X] = 0$

$\beta_0 = E[Y] - \beta_1 E[X]$

$$E[XU] = 0 \Rightarrow 0 = E[X(Y - \beta_0 - \beta_1 X)]$$

$$\Rightarrow 0 = E[X(Y - E[Y] + \beta_1 E[X] - \beta_1 X)]$$

$$\Rightarrow 0 = E[X(Y - E[Y])] - \beta_1 E[X(X - E[X])]$$

$E[X(Y - \beta_0 - \beta_1 X)] = 0$

$E[X(Y - E(Y) + \beta_1 E(X) - \beta_1 X)] = 0$

$E[X(Y - E(Y))] - \beta_1 E[X(X - E(X))]$

$- E[X(-\beta_1 E[X] + \beta_1 X)]$

$- \beta_1 E[X(X - E[X])]$

$$\frac{E[(x-\mu)(y-\mu)]}{E((x-\mu)^2)}$$

- Note the following two facts:

$$E[X(Y-E[Y])] = E[XY] - E[X]E[Y] = Cov[X,Y]$$

$$E[X(X-E[X])] = E[X^2] - E[X]E[X] = Var[X]$$

sample counterpart of population

- So we have $0 = Cov[X,Y] - \beta_1 Var[X]$, and thus

OLS : $\frac{\widehat{cov}(x,y)}{\widehat{var}(x)}$

$$\boldsymbol{\beta_1 = \frac{Cov[X,Y]}{Var[X]}}$$

sign of slope depends on cov(x,y)

scale by variance regressor in the model

$$\boldsymbol{\beta_0 = E[Y] - \left(\frac{Cov[X,Y]}{Var[X]}\right)E[X]}$$

- Notice that the slope coefficient reflects how X and Y are correlated: positive (negative) correlation would result in upward (downward) sloping.

Special Case: Binary Regressor

- Suppose the regressor $X \sim$ Bernoulli(p). We can also call it a **binary regressor** or **dummy variable**
- Recall: $E[X] = p, Var[X] = p(1-p)$
- Note:

$$\begin{aligned} E[XY] &= E\big[E[XY|X]\big] \\ &= E[XY|X=1]P\{X=1\} + E[XY|X=0]P\{X=0\} \\ &= E[Y|X=1]p \end{aligned}$$

- Note:

$$E[Y] = E[E(Y|X)] = E[Y|X=1]p + E[Y|X=0](1-p)$$

- Then, we have

$$\begin{aligned} Cov[X,Y] &= E[Y|X=1]p - p(E[Y|X=1]p + E[Y|X=0](1-p)) \\ &= p(1-p) \cdot \{E[Y|X=1] - E[Y|X=0]\} \end{aligned}$$

- Therefore,

$$\beta_1 = \frac{Cov[X,Y]}{Var[X]} = E[Y|X=1] - E[Y|X=0]$$

- Similarly,

$$\begin{aligned}\beta_0 &= E[Y] - \beta_1 E[X] \\ &= E[Y|X=1]p + E[Y|X=0](1-p) - \beta_1 p \\ &= E[Y|X=1]p + E[Y|X=0](1-p) - (E[Y|X=1] - E[Y|X=0])p \\ &= E[Y|X=0]\end{aligned}$$

- In words, **$\boldsymbol{\beta_0}$ is the mean of $\boldsymbol{Y}$ when $\boldsymbol{X = 0}$**.
- and **$\boldsymbol{\beta_0 + \beta_1}$ is the mean of $\boldsymbol{Y}$ when $\boldsymbol{X = 1}$**.
- **$\boldsymbol{\beta_1}$ thus captures the difference in means of the two groups** of the population identified by X.

Special cases

no regressor (x)

$E[y] = E(\beta_0 + U)$

$\boxed{E[y] = \beta_0} + \cancel{E(U)}$

any interpretation → $E[U] = 0$

β_0 is mean of y

$\hat{\beta}_0^{OLS} = \bar{y}_n = \frac{1}{n}\sum y_i$

sample counterpart of population mean

---

dummy

$y = \beta_0 + \beta_1 D + U$

random Bernoulli variable

Bernoulli $\{^1_0$

Binary

Dummy

two cases:

Set $\underline{D=0}$ ⇒ $y = \beta_0 + U$

sample average for group that $D=0$

$E[y \mid D=0]$

Set $D=1$ ⇒ $y = \beta_0 + \beta_1 + U$

$E[y \mid D=1]$

capture mean of different group

$\beta_1 = E(y \mid D=1) - E(y \mid D=0)$

How much average of y for the group $D=1$ is higher than average of y for $D=0$

Now, let's find

an estimator for β_0 and β_1

The OLS Estimator

- Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be an iid sample from the distribution of (X, Y). Suppose that all the assumptions stated before are true:
  - $Y = \beta_0 + \beta_1 X + U$
  - $E[U] = 0$
  - $E[XU] = 0$
  - $0 < Var[X] < \infty$
- Then, recall that the values of the *parameters*

$$\beta_1 = \frac{Cov[X, Y]}{Var[X]}$$

$$\beta_0 = E[Y] - \left(\frac{Cov[X, Y]}{Var[X]}\right) E[X]$$

- We want to find estimators $\hat{\beta}_0, \hat{\beta}_1$ for these two parameters.

- Natural estimators are

$$\beta_1 = \frac{Cov[X,Y]}{Var[X]} \Rightarrow \boldsymbol{\widehat{\beta}_1 = \frac{\widehat{\sigma}_{XY}}{\widehat{\sigma}_X^2}}$$

$$\beta_0 = E[Y] - \left(\frac{Cov[X,Y]}{Var[X]}\right)E[X] \Rightarrow \boldsymbol{\widehat{\beta}_0 = \overline{Y}_n - \frac{\widehat{\sigma}_{XY}}{\widehat{\sigma}_X^2}\overline{X}_n}$$

- Recall that β_0, β_1 solve $E[U] = 0, E[XU] = 0$. Hence, it isn't surprising that $\boldsymbol{\widehat{\beta}_0, \widehat{\beta}_1}$ **solve the two corresponding *sample* moment conditions**:

$$\frac{1}{n}\sum_{i=1}^{n}(\widehat{U}_i) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i\widehat{U}_i) = \frac{1}{n}\sum_{i=1}^{n}X_i(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i) = 0$$

- Theoretically, think of these sample moments as the FOCs of

มันจะไม่สำคัญค่าเท่ากัน

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{U}_i\right)^2$$

actual

different btw Y and estimation line

Try to minimize $(\hat{U}_i)^2$

- For this reason, **the $\hat{\beta}_0, \hat{\beta}_1$ is called the Ordinary Least Squares (OLS) estimators of the parameters β_0, β_1**

- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ is the predicted value of Y_i, referred to as **fitted value**

- $\hat{U}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$ is an estimate of U_i, referred to as the **residual** for the ith observation

- Recall the immediate properties of OLS by construction are

doesn't matter

$$\frac{1}{n}\sum_{i=1}^{n}\hat{U}_i = 0; \quad \frac{1}{n}\sum_{i=1}^{n}X_i\hat{U}_i = 0$$

- For the special case that X is binary, $\hat{\beta}_0$ is the sample mean of Y_i when $X = 0$, and $\hat{\beta}_0 + \hat{\beta}_1$ is the sample mean of Y_i when $X = 1$

OLS:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \frac{1}{n} \sum \left[(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \right]$$

$$\Downarrow$$

$$\frac{1}{2} \sum \hat{u} = 0$$

$$\frac{1}{n} \sum \hat{u} x = 0$$

BLP, best linear approximation

$$\min_{\beta_0, \beta_1} E\left[(y - \beta_0 - \beta_1 x)^2 \right]$$

$$\Downarrow$$

$$E[u] = 0$$

$$E[xu] = 0$$

don't observe / draw from pop

Blue + Red = Population

draw from pop

population model

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$

error

estimate → we see it

$$Y_i = \underbrace{\hat{\beta}_0 + \hat{\beta}_1 X_i}_{\hat{Y}_i} + \hat{U}_i$$

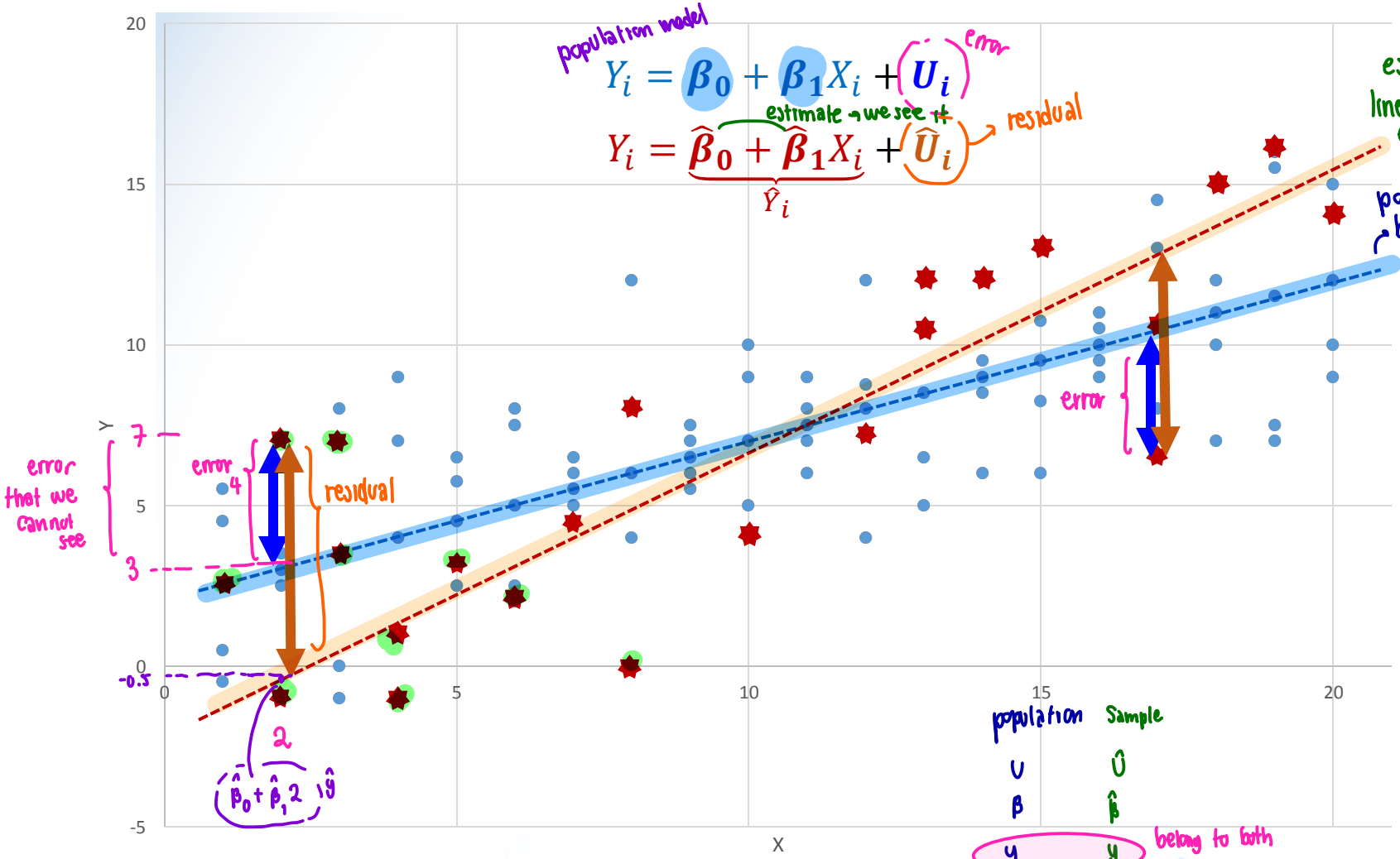
residual

estimator line

line that we estimate from sample

population line

belong to population that we not see it as we don't observe



population	Sample
U	$\hat{U}$
β	$\hat{\beta}$
y	y
	$\hat{y}$

belong to both

Prove of unbiased estimator

Properties of OLS: Bias?

- Recall that the OLS estimator for the slope parameter is

$$\hat{\beta}_1 = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X^2} = \frac{\Sigma_{i=1}^n[(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)]}{\Sigma_{i=1}^n[(X_i - \bar{X}_n)^2]}$$

- Note that

$$\Sigma_{i=1}^n[(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)] = \Sigma_{i=1}^n[(X_i - \bar{X}_n)Y_i\ ]$$

$$\Sigma_{i=1}^n[(X_i - \bar{X}_n)^2] = \Sigma_{i=1}^n[(X_i - \bar{X}_n)X_i]$$

- So, the denominator is

$$\begin{aligned}\Sigma_{i=1}^n[(X_i - \bar{X}_n)Y_i] &= \Sigma_{i=1}^n[(X_i - \bar{X}_n)(\beta_0 + \beta_1 X_i + U_i)] \\ &= \underbrace{\Sigma_{i=1}^n \beta_0 (X_i - \bar{X}_n)}_{0} + \Sigma_{i=1}^n \beta_1 X_i (X_i - \bar{X}_n) + \Sigma_{i=1}^n U_i (X_i - \bar{X}_n) \\ &= \beta_1 \Sigma_{i=1}^n X_i (X_i - \bar{X}_n) + \Sigma_{i=1}^n U_i (X_i - \bar{X}_n)\end{aligned}$$

- Then,

$$\hat{\beta}_1 = \frac{\beta_1 \Sigma_{i=1}^n X_i(X_i - \bar{X}_n) + \Sigma_{i=1}^n U_i(X_i - \bar{X}_n)}{\Sigma_{i=1}^n X_i(X_i - \bar{X}_n)} = \beta_1 + \frac{\Sigma_{i=1}^n U_i(X_i - \bar{X}_n)}{\Sigma_{i=1}^n X_i(X_i - \bar{X}_n)}$$

- **Under the interpretation 1**, we have $E[U_i|X_1, \dots, X_n] = 0$. Then,

$$E\left[\frac{\Sigma_{i=1}^n U_i(X_i - \bar{X}_n)}{\Sigma_{i=1}^n X_i(X_i - \bar{X}_n)} \middle| X_1, \dots, X_n\right] = \frac{E[\Sigma_{i=1}^n (X_i - \bar{X}_n) U_i | X_1, \dots, X_n]}{\Sigma_{i=1}^n X_i(X_i - \bar{X}_n)}$$

If we don't have this, we cannot prove.

If this equal to 0

$$= \frac{\Sigma_{i=1}^n \{(X_i - \bar{X}_n) \boldsymbol{E[U_i|X_1, \dots, X_n]}\}}{\Sigma_{i=1}^n X_i(X_i - \bar{X}_n)}$$

then

- Then, by LIE, $E[\hat{\beta}_1] = E\left[E[\hat{\beta}_1|X_1, \dots, X_n]\right] = E[\beta_1 + 0] = \beta_1$

- Recall the OLS estimator of β_0 is $\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n$

- Then, under the interpretation 1,

$$
\begin{aligned}
E[\hat{\beta}_0 | X_1, \dots, X_n] &= E[\bar{Y}_n - \hat{\beta}_1 \bar{X}_n | X_1, \dots, X_n] \\
&= E[\bar{Y}_n | X_1, \dots,] - \bar{X}_n E[\hat{\beta}_1 | X_1, \dots, X_n] \\
&= E[\bar{Y}_n | X_1, \dots,] - \bar{X}_n \beta_1
\end{aligned}
$$

$$
\begin{aligned}
E[\hat{\beta}_0] &= E\left[E[\hat{\beta}_0 | X_1, \dots, X_n]\right] \\
&= E[E[\bar{Y}_n | X_1, \dots,]] - \beta_1 E[\bar{X}_n] \\
&= E[\bar{Y}_n] - \beta_1 E[\bar{X}_n] \\
&= E[Y] - \beta_1 E[X] = \beta_0
\end{aligned}
$$

- Note that **we cannot prove this with the linear approximation and the causal model interpretations**, because we do not have $E[U|X] = 0$. So, OLS may not be unbiased in those cases.

Properties of OLS: Consistent?

- *Assume that* $E[U] = E[XU] = 0, Var[X] < \infty, E[X^4] < \infty$ and $E[Y^4] < \infty$, then we know that WLLN implies

$$\hat{\sigma}_X^2 \xrightarrow{P} \sigma_X^2 \text{ and } \hat{\sigma}_{XY} \xrightarrow{P} \sigma_{XY}$$

- Then, by the CMT,

$$\hat{\beta}_1 = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X^2} \xrightarrow{P} \frac{\sigma_{XY}}{\sigma_X^2} = \beta_1$$

- and we know that $\bar{X}_n \xrightarrow{P} E[X]$ and $\bar{Y}_n \xrightarrow{P} E[Y]$ Then, by CMT,

$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n \xrightarrow{P} E[Y] - \beta_1 E[X] = \beta_0$$

- Therefore, the OLS estimators are consistent under the interpretations 1 and 2

- What if $E[XU] \neq 0$?
- In this case, we still have that

weak law of large number
convert to

$$\hat{\beta}_1 = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X^2} \xrightarrow{P} \frac{\sigma_{XY}}{\sigma_X^2}$$

- However, in this case

$$\frac{\sigma_{XY}}{\sigma_X^2} = \frac{Cov[X,Y]}{Var[X]} = \frac{Cov[X, \beta_0 + \beta_1 X + U]}{Var[X]}$$

$$= \frac{Cov[X, \beta_1 X] + Cov[X,U]}{Var[X]} = \boxed{\beta_1 + \frac{Cov[X,U]}{Var[X]}}$$

require
if only $Cov[X,U] = 0$
for OLS
to be consistent

- Therefore, **if $E[XU] \neq 0$, the OLS estimators are not consistent**. This can be the case for the causal model if the unobserved determinant and the regressor are correlated.

must be zero too
always zero for all intepretation

$$Cov(U,X) = E[UX] - \cancel{E[U]}^{0} E[X] = 0$$

MSF Chula

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

AACSB ACCREDITED · EFMD EQUIS ACCREDITED · EFMD EPAS ACCREDITED

Research Question

Correlation vs Causation

more moment conditions

Interpretation 1, 2, or 3?

OLS Estimator

Unbiased or Consistent?

What it is?

How to derive it?

When you can or cannot use it?

(unbiased or consistent

$$y = \underbrace{\beta_0 + \beta_1}_{\text{parameter we want to estimate}} x + u$$

OLS estimator : $\min_{\hat{\beta}_0, \hat{\beta}_1} \sum \hat{u}^2$

$\hat{\beta}_1 = 2$ → can we trust ?

consistent — justify for 3rd: $\boxed{cov(u, x) = 0}$

1st interpretation ✓

2nd interpretation ✓

unbiased: $E[u|x] = 0$

1st interpretation ✓

1st & 2nd :

causal : x increases by 1 , keep other constant , y increases by 2

UNBIASED $\not\Leftrightarrow$ CONSISTENT

only OLS : $E[u|x] = 0 \Rightarrow cov(u, x) = 0$

01/09/2023

Review

Population

we're interested in

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

Parameters → know these

know line

- unknown
- fixed / constant

Sample (Data/observation)

OLS estimator (random)

↳ try to minimize distance

residual square

$\min_{\hat{\beta}_0, \hat{\beta}_1} \frac{1}{n} \sum \hat{u}_i^2$ in sample

↳ sample counterpart of

BLP [y|x] : min $E[u^2]$ in population

Population Conditions

always has this

1st

$E[u|x] = 0 \Rightarrow$ UNBIASED

$E[u] = E[xu] = cov(u, x) = 0 \Rightarrow$ CONSISTENT

2nd, 1st (mean independent)

3 ways to interpret the line $\beta_0 + \beta_1 x$

① $E[y|x]$

② $BLP(y|x)$

Correlation:
signal, indicator, approximator, predict, forecast
say something about y

③ "Pure" or "ceteris paribus" effect of x on y

need to put some controls in it

CAUSATION

$E[u] = 0$

Interpretic β

1) $y = \beta + u$ || $\hat{\beta}^{ols} \Rightarrow \bar{y}_n$

$\beta \Rightarrow$ mean of y

2) $y = \beta_0 + \beta_1 D + u$ || β_0 as a mean of y for $D=0$

dummy 0 1

β_1 as how mean of y for $D=1$ is higher than when $D=0$

3) $\ln y = \beta_0 + \beta_1 x = 0$ || $x\uparrow$ 1 unit $\Rightarrow y\uparrow$ $100\beta_1$%

% change in that r.v.

4) $y = \beta_0 + \beta_1 \ln x + u$ || $x\uparrow$ 1% $\Rightarrow y\uparrow \frac{\beta_1}{100}$ unit

5) $\ln y = \beta_0 + \beta_1 \ln x + u$ || $x\uparrow$ 1% $\Rightarrow y\uparrow$ β_1%

implies elasticity of y

research

Q. interesting in if work exp increase wage :

1) $wage = \beta_0 + \overset{100}{\beta_1} exp + u$ exp↑ 1 year, wage ↑ 100 THB ⇒ not make sense as every employees have different wage base

2) $\ln wage = \beta_0 + \overset{0.1}{\beta_1} exp + u$ exp↑ 1 year, wage ↑ 10%

3) $wage = \beta_0 + \overset{10000}{\beta_1} \ln \overset{\frac{d\,exp}{exp} \times 100}{exp} + u$ exp↑ 1 %, wage ↑ 100 THB

4) $\ln(wage) = \beta_0 + \overset{10}{\beta_1} \ln(exp) + u$