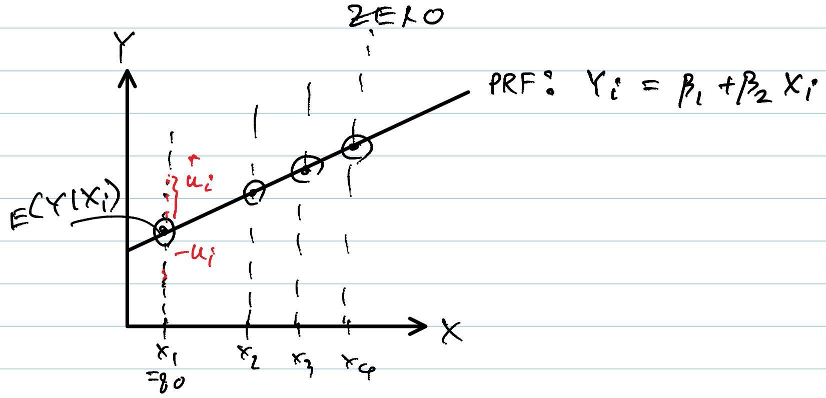
Friday, September 14, 2012
12:40 PM

ASSUMPTIONS UNDERLYING THE OLS (10 ASSUMPTIONS)

① LINEAR MODEL : $Y_i = \beta_1 + \beta_2 X_i + u_i$.

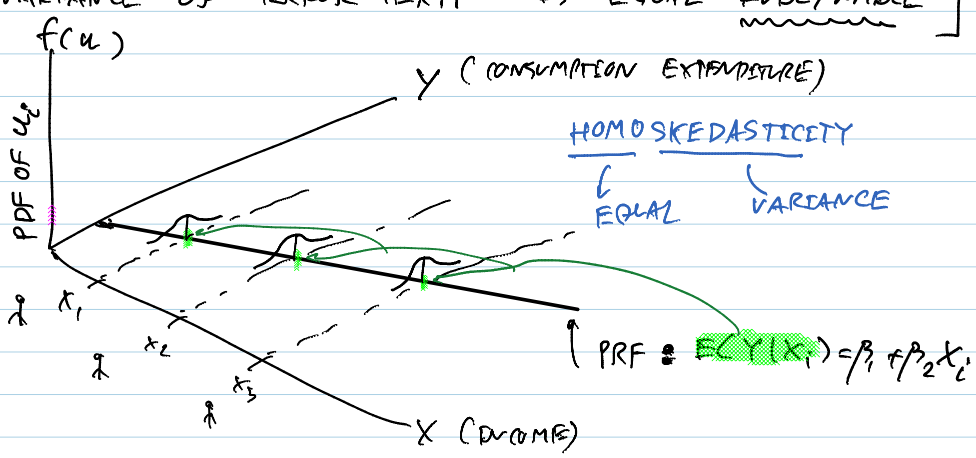
② X IS NONSTOCHASTIC, i.e., X is FIXED IN REPEATED SAMPLING.

③ $E(u_i) = 0$, i.e., MEAN OR AVERAGE VALUES OF DISTURBANCE TERM IS EQUAL TO ZERO



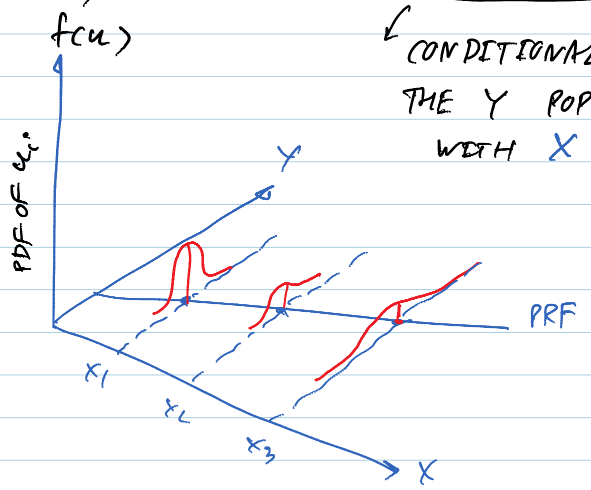
④ $var(u_i) = E(u_i - E(u_i))^2 = E(u_i^2) = \sigma_u^2$

[VARIANCE OF ERROR TERM IS EQUAL EVERYWHERE]

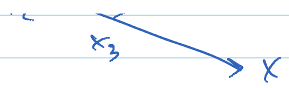


IN CONTRAST, LET'S SEE HETERO SKEDASTICITY

CONDITIONAL VARIANCE OF THE Y POPULATION VARIES WITH X.



VARIANCE OF THE Y POPULATION IS NO LONGER CONSTANT



⑤ $\text{cov}(u_i, u_j) = E[(u_i - E(u_i))(u_j - E(u_j))]$
$= E(u_i u_j) = 0$

[IT SAYS THAT NO AUTOCORRELATION BETWEEN THE DISTURBANCE TERMS]

EX: CONSIDER $Y_t = \beta_1 + \beta_2 X_t + u_t \Rightarrow$ OUR PRF

SUPPOSE THAT u_t IS POSITIVELY LINKED W/ u_{t-1}.

THEN, Y_t DEPENDS NOT ONLY ON X_t BUT ALSO u_{t-1}!

BECAUSE u_t AND u_{t-1} ARE POSITIVELY CORRELATED.

- THIS CASE IS NOT ALLOWED TO HAPPEN IF ASSUMPTION 4 IS VALID.
- WE WILL SEE LATER (A/F MIDTERM) ON WHAT WOULD BE THE CONSEQUENCES IF u_t AND u_{t-1} ARE CORRELATED. (THIS PROBLEM CALLED "AUTOCORRELATION")

⑥ $\text{cov}(X_i, u_i) = E[(X_i - E(X_i))(u_i - E(u_i))]$
$= E(X_i u_i) = 0$

[IT SAYS THAT EXPLANATORY VARIABLE X_i AND DISTURBANCE TERMS ARE NOT CORRELATED]

INTUITION? $\Rightarrow$ $Y_i = \beta_1 + \beta_2 X_i + u_i$ PRF

$\hat{\beta}_2 = 0.8$

AS WE WANT TO SEPARATE THE EFFECT OF X ON Y (INDIVIDUAL EFFECT) FROM THE EFFECT OF ANY OTHER VARIABLES ON Y,

WE NEED THIS ASSUMPTION OF $\text{cov}(X_i, u_i) = 0$.

⑦ NUMBER OF OBSERVATIONS > NUMBER OF PARAMETERS

(OR $n > k$.)

k → # OF EXPLANATORY VARIABLES.

EX: IF $n = k = 2$ $\Rightarrow$ CANNOT ESTIMATE THIS MODEL.

2

(8) X HAS VARIABILITY, i.e., $var(X_i) = \sigma_x^2 \neq 0$

(9) THE MODEL HAS A CORRECT SPECIFICATION.

MEANS THAT
1. CORRECT FUNCTIONAL FORM
2. NO OMITTED VARIABLES
3. NO REDUNDANT VARIABLES

CH.13

A/F MIDTERM

(10) NO PERFECT MULTICOLLINEARITY

IT MEANS THAT THERE ARE NO PERFECT "LINEAR" RELATIONSHIP BETWEEN EXPLANATORY VARIABLES.

FOR EXAMPLE:

Y_i	X_{2i}	X_{3i} ($= 3X_{2i} + 2$)
⋮	1	5
⋮	3	11
⋮	5	17
⋮	7	23
⋮	9	29

PERFECT LINEAR RELATIONSHIP

IF THIS IS SO,

THE MODEL CANNOT BE SOLVED!

(WE WILL SEE WHEN WE DO STATA)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

(MULTIPLE REGRESSION MODEL)

H - A - M PROBLEM

ASS. 4 ASS. 5 ASS. 10

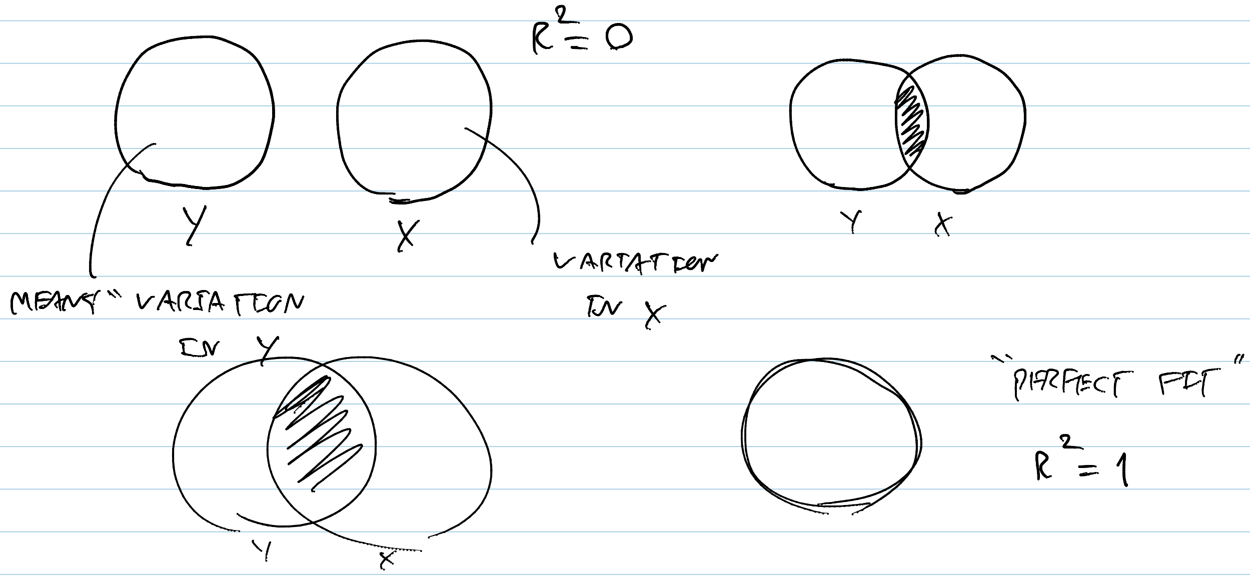
21.09.2012

THE COEFFICIENT OF DETERMINATION R^2:

A MEASURE OF "GOODNESS OF FIT"

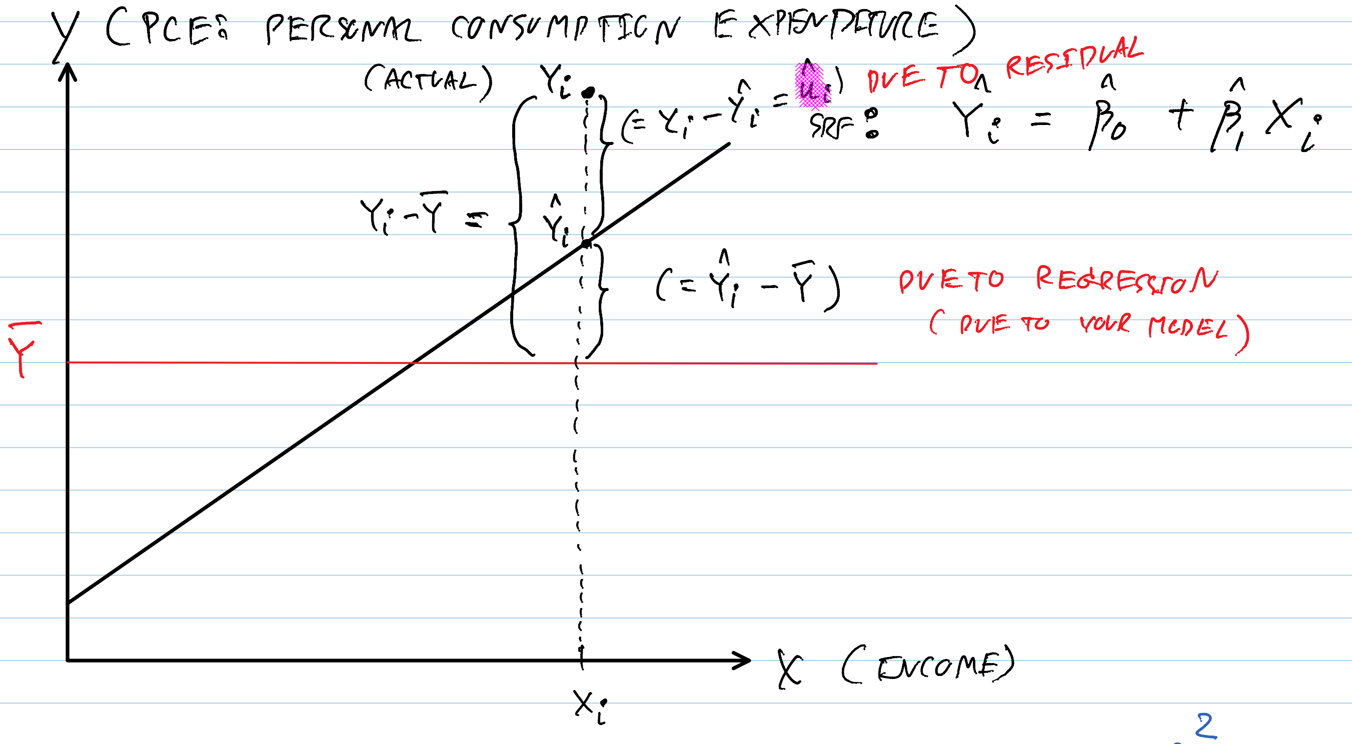
Q: HOW "WELL" THE SAMPLE REGRESSION LINE FITS THE DATA?

A: WE USE R^2 TO MEASURE THE DEGREE OF EXPLANATORY POWER



- THE HIGHER THE OVERLAPPING AREA, THE HIGHER PROPORTION THAT VARIATION IN Y IS EXPLAINED BY VARIATION IN X.

- $0 \leq R^2 \leq 1$



LET'S LOOK AT HOW TO COMPUTE OR TO DERIVE R^2

CONSIDER $Y_i = \hat{Y}_i + \hat{u}_i$ — ① (TRUE EVERY WHERE)

(ACTUAL) (ESTIMATED) (RESIDUAL)

OR IN DEVIATION FORM: $y_i = \hat{y}_i + \hat{u}_i$. — ②

TAKING SQUARE FOR BOTH SIDES AND SUMMING OVER ALL OBSERVATION

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} (\hat{y}_i + \hat{u}_i)^2$$

$$= \sum_{i=1}^{n} [\hat{y}_i^2 + 2\hat{y}_i\hat{u}_i + \hat{u}_i^2]$$

$$= \sum_{i=1}^{n} \hat{y}_i^2 + 2\underbrace{\sum_{i=1}^{n} \hat{y}_i\hat{u}_i}_{=0 \text{ BY ASSUMPTION}} + \sum_{i=1}^{n} \hat{u}_i^2$$

$$= \sum (\hat{\beta}_2 x_i)^2 + \sum_{i=1}^{n} \hat{u}_i^2$$

$$\sum_{i=1}^{n} y_i^2 = \hat{\beta}_2 \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} \hat{u}_i^2$$

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \qquad [\because \bar{\hat{Y}} = \bar{Y}]$$

TSS

TOTAL SUM OF SQUARE = TOTAL VARIATION OF ACTUAL Y FROM THEIR SAMPLE MEAN