HETEROSCEDASTICITY: WHAT HAPPENS IF THE ERROR VARIANCE IS NOT CONSTANT?

- NATURE OF THE PROBLEM
- CONSEQUENCES OF THE PRESENCE OF HETEROSCEDASTICITY
- HOW TO DETECT THE PROBLEM (WHETHER IT EXISTS)
- REMEDY MEASURES

HOMOSCEDASTICITY: $E(u_i^2) = \sigma^2$ FOR $i = 1, 2, \ldots, n$.

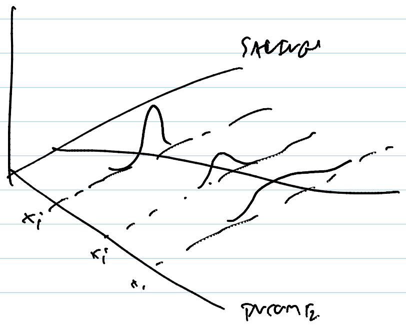
HETEROSCEDASTICITY: $E(u_i^2) = \sigma_i^2$

NOTICE THAT THE SUBSCRIPT OF σ^2, REMINDS US THAT THE CONDITIONAL VARIANCES OF u_i (= CONDITIONAL VARIANCE OF Y_i) ARE NO LONGER CONSTANT.

- WHY HETEROSCEDASTICITY ARISES?

(1) ASSUME THAT $\underbrace{Y_i}_{\text{SAVING}} = \beta_1 + \beta_2 \underbrace{X_i}_{\text{INCOME}} + u_i$

FROM THE FIGURES, HIGHER INCOME FAMILIES ON AVERAGE SAVE MORE THAN THE LOW INCOME FAMILIES BUT FOR HIGH INCOME FAMILIES, THERE ARE MORE VARIABILITY IN THEIR SAVING



(2) ERROR-LEARNING MODEL: AS PEOPLE LEARN, THEIR ERRORS OF BEHAVIOR BECOME SMALLER OVER TIME SO IN THIS CASE, σ_i^2 IS EXPECTED TO DECREASE.

(3) AS INCOMES GROW, PEOPLE HAVE MORE OPTIONS HOW THEY ARE GOING TO SPEND THEIR INCOME.

SO IN THIS CASE, σ_i^2 IS LIKELY TO INCREASE W/ INCOME.

(4) AS DATA COLLECTING TECHNIQUES IMPROVE, σ_i^2 IS LIKELY TO DECREASE

EX: $\boxed{\text{BANK W/ SOPHISTICATED DATA PROCESSING DEVICES}}$

VS.

$\boxed{\text{BANK W/O THAT FACILITIES}}$

SO, BANKS W/ FACILITIES ARE LIKELY TO COMMIT FEWER ERRORS W/ THEIR MONTHLY OR

QUARTERLY STATEMENT OF THEIR CUSTOMERS THAN THOSE W/O SUCH FACILITIES

(5) HETEROSCEDASTICITY MIGHT OCCUR DUE TO THE PRESENCE OF OUTLIERS.

EX: CHILI IS AN OUTLIER IN THE DIAGRAM.

INCLUDING CHILE IN THE DATA SET MIGHT CHANGE THE RESULTS OF REGRESSION ANALYSIS

(6) MODEL MISSPECIFICATION:

- PUT IRRELEVANT VARIABLE(S) INTO THE MODEL

AND/OR • IGNORE IMPORTANT VARIABLE(S) IN THE MODEL

EX:
$$Q_x^d = f(P_x, \text{INCOME } P_y) \quad \text{— (1)}$$
$$Q_x^d = f(P_x, \text{INCOME}) \quad \text{— (2)}$$

SUPPOSE (1) IS THE CORRECT MODEL

(7) HETEROSCEDASTICITY MAY ARISE BECAUSE

→ (1) INCORRECT DATA TRANSFORMATION

$$Y = \alpha_1 + \alpha_2 L + \alpha_2 K + u_i$$
$$\frac{Y}{L} = \frac{\alpha_1}{L} + \alpha_2 \frac{L}{L} + \alpha \frac{K}{L} + \frac{u_i}{L}.$$

EX: RATIO OR FIRST-DIFFERENCE TRANSFORMATIONS

→ (2) INCORRECT FUNCTIONAL FORM

EX: LINEAR VERSUS LOG-LINEAR MODELS.

- CONSEQUENCES OF HETEROSCEDASTICITY

Q: WHAT HAPPENS TO OLS ESTIMATORS AND THEIR VARIANCES IF WE INTRODUCE HETEROSCEDASTICITY BY LETTING $E(u_i^2) = \sigma_i^2$. BUT RETAIN ALL OTHER ASSUMPTIONS OF CLRM?

CONSIDER
$$Y_i = \beta_1 + \beta_2 X_i + u_i.$$

BY USING OLS, WE OBTAIN
$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$
$$= \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}.$$

NOW,
$$var(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} \Rightarrow var(\hat{\beta}_2) \text{ UNDER HETEROSCEDASTICITY}$$

PROOF: $var(\hat{\beta}_2) = E(k_1^2 u_1^2 + k_2^2 u_2^2 + \dots k_n^2 u_n^2 + 2 \text{ CROSS-PRODUCT TERMS})$

$$= E(k_1^2 u_1^2 + k_2^2 u_2^2 + \dots k_n^2 u_n^2)$$

(SINCE EXPECTATIONS OF CROSS-PRODUCT TERMS ARE ZERO)
(WHY?)

$$var(\hat{\beta}_2) = k_1^2 E(u_1^2) + k_2^2 E(u_2^2) + \dots + k_n^2 E(u_n^2)$$

$$= k_1^2 \sigma_1^2 + k_2^2 \sigma_2^2 + \dots + k_n^2 \sigma_n^2$$

B/C OF NO AUTOCORRELATION ASSUMPTION (LATER STUDIED)

(SINCE $E(u_i^2) = \sigma_i^2$)

$$var(\hat{\beta}_2) = \sum_{i=1}^{n} k_i^2 \sigma_i^2$$

$$= \sum \left[\left(\frac{x_i}{\sum x_i^2} \right)^2 \sigma_i^2 \right]$$

SINCE $k_i = \dfrac{x_i}{\sum x_i^2}$.

$$\boxed{var(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2}} \rightarrow \text{HETEROSKEDASTICITY CASE}$$

COMPARE WITH

$$\boxed{var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}} \rightarrow \text{HOMOSCEDASTICITY CASE}$$

* NOTICE ALSO THAT IF $\sigma_i^2 = \sigma^2$ FOR EACH i, THE TWO FORMULA WILL BE IDENTICAL.

Q: W/ HETEROSCEDASTICITY, IS $\hat{\beta}_2$ STILL BLUE?

BEST LINEAR UNBIASED ESTIMATOR

A:

UNBIASEDNESS ✓

CONSISTENT ✓

BEST X (NO)

$\boxed{E(\hat{\beta}_2) = \beta_2}$

- $\hat{\beta}_2$ IS STILL ~~UNBIASED~~ DESPITE OF HETERO
- $\hat{\beta}_2$ IS STILL ~~CONSISTENT~~ " ______ "
- $\hat{\beta}_2$ WILL NO LONGER PRODUCE MINIMUM VARIANCE!

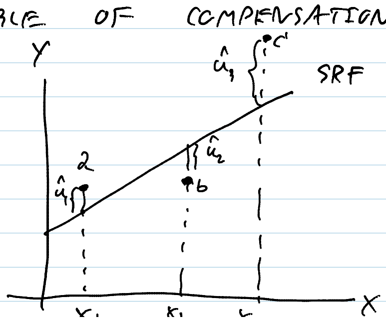
16.11.12

THE METHOD OF GENERALIZED LEAST SQUARES (GLS)

OBJECTIVE: FIND A METHOD SUCH THAT OBSERVATION COMING FROM POPULATIONS WITH GREATER VARIABILITY ARE GIVEN LESS WEIGHT THAN THOSE COMING FROM POPULATIONS WITH SMALLER VARIABILITY

TAKE A LOOK AT TABLE OF COMPENSATION & SIZES OF FIRMS:

ANOTHER EXAMPLE:



$\text{MIN} \sum \hat{u}_i^2$

RSS

OK, ... TAKE A LOOK AT HOW TO CONSTRUCT GLS

CONSIDER (1) $Y_i = \beta_1 + \beta_2 X_i + u_i$ (TWO-VARIABLE REGRESSION MODEL)

OR (2) $Y_i = \beta_1 X_{0i} + \beta_2 X_i + u_i$

WHERE $X_0 = 1$

NOW ASSUME THAT σ_i^2 ARE KNOWN

HETEROSCEDASTIC VARIANCE

(3) $$\frac{Y_i}{\sigma_i} = \beta_1 \left(\frac{X_{0i}}{\sigma_i}\right) + \beta_2 \left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right)$$

(4) $$Y_i^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^*$$

WHERE THE STARRED DENOTES TRANSFORMED PARAMETERS.

Q: WHAT IS THE PURPOSE OF DOING THIS TRANSFORMATION OF THE ORIGINAL MODEL?

A: TAKE A CLOSER LOOK AT u_i^*!

$$\text{var}(u_i^*) = E(u_i^*)^2 = E\left(\frac{u_i}{\sigma_i}\right)^2$$

$$= \frac{1}{\sigma_i^2} E(u_i^2)$$

$$= \frac{1}{\cancel{\sigma_i^2}} \cdot \cancel{(\sigma_i^2)}$$ SINCE $E(u_i^2) = \sigma_i^2$

$= 1$!

YOU CAN SEE THAT NOW THE VARIANCE OF THE TRANSFORMED DISTURBANCE TERM u_i^* IS NOW HOMOSCEDASTIC. :)

IN SHORT, GLS IS "ACTUALLY" OLS ON THE TRANSFORMED VARIABLES THAT SATISFY THE OLS ASSUMPTIONS. THE ESTIMATORS ARE NOW BLUE. THE ESTIMATORS OBTAINED ARE KNOWN AS GLS ESTIMATORS.

- THE ACTUAL METHOD OF ESTIMATING β_1^* AND β_2^* IS THE FOLLOWING

STEP 1 WRITE DOWN SRF: $\frac{Y_i}{\sigma_i} = \hat{\beta}_1^* \frac{X_{0i}}{\sigma_i} + \hat{\beta}_2^* \frac{X_i}{\sigma_i} + \frac{\hat{u}_i}{\sigma_i}$

STEP 1 WRITE DOWN SRF :

$$\frac{Y_i}{\sigma_i} = \beta_1^* \frac{X_{0i}}{\sigma_i} + \beta_2^* \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

OR

$$Y_i^* = \hat{\beta}_1^* X_{0i}^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^*$$

STEP 2 TO OBTAIN GLS ESTIMATORS, WE MINIMIZE

$$\sum \hat{u}_i^{*2} = \sum \Big[\underbrace{Y_i^* - \hat{\beta}_1^* X_{0i}^* - \hat{\beta}_2^* X_i^*}_{\hat{u}_i^*} \Big]^2$$

$w_i = \frac{1}{\sigma_i^2}$

$$\sum \left(\frac{\hat{u}_i}{\sigma_i} \right)^2 = \sum \left[\left(\frac{Y_i}{\sigma_i} \right) - \hat{\beta}_1^* \left(\frac{X_{0i}}{\sigma_i} \right) - \hat{\beta}_2^* \left(\frac{X_i}{\sigma_i} \right) \right]^2$$

BY USING LAGRANGE METHOD, WE CAN OBTAIN

$$\boxed{\hat{\beta}_2^* = \frac{(\sum w_i)(\sum w_i X_i Y_i) - (\sum w_i X_i)(\sum w_i Y_i)}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2}}$$

WHERE $w_i = \frac{1}{\sigma_i^2}$.

(SEE APPENDIX 11A, SECTION 11A.2)

$$\boxed{\operatorname{var} \hat{\beta}_2^* = \frac{\sum w_i}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2}}$$

WHERE $w_i = \frac{1}{\sigma_i^2}$.

Q: WHAT ARE THE DIFFERENCE BETWEEN OLS AND GLS?

OLS

. MIN $\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$

IN WORDS, OLS TRYS TO MINIMIZE "SUM OF RESIDUALS SQUARED"

OR

"UNWEIGHTED SUM OF RESIDUAL SQUARED"

GLS

MIN $\underbrace{w_i \hat{u}_i^2}_{\hat{u}_i^{*2}} = \sum w_i \left(Y_i - \hat{\beta}_1^* X_{0i} - \hat{\beta}_2^* X_i \right)^2$

, WHERE $w_i = \frac{1}{\sigma_i^2}$.

IN GLS, IT TRYS TO MINIMIZE "WEIGHTED SUM OF RESIDUAL SQUARED"

KEY IDEA OF GLS : IN GLS, THE WEIGHT ASSIGNED TO EACH OBSERVATION IS INVERSELY PROPORTIONAL TO ITS σ_i, THAT IS

$$\boxed{w_i = \frac{1}{\sigma_i^2}}$$

OBSERVATIONS COMING FROM A POPULATION

$$\boxed{w_i = \frac{1}{\sigma_i^2}}$$

TO FITS σ_i, THAT IS

OBSERVATIONS COMING FROM A POPULATION W/ LARGER σ_i WILL GET RELATIVELY <u>SMALLER</u> WEIGHT AND OBSERVATIONS COMING FROM A POPULATION W/ SMALLER σ_i WILL GET "PROPORTIONATELY" LARGER WEIGHT IN THE JOB OF MINIMIZING RSS.