HETEROSCEDASTICITY

CONSIDER THE REGRESSION MODEL

$$Y_i = \beta_1 + \sum_{j=2}^{k} \beta_j X_{ij} + u_i$$

ASSUMPTIONS:

$$E(u_i) = 0$$
$$E(u_i^2) = \sigma^2; \quad E(u_i u_j) = 0 \quad \text{FOR } i \neq j$$

THIS IMPLIES THAT THE ERROR-COVARIANCE MATRIX

$$\Sigma = E(\underset{\sim}{u}\underset{\sim}{u}') = \sigma^2 I$$

BECAUSE EACH u_i IS ASSUMED TO HAVE THE SAME VARIANCE, THIS IS KNOWN AS THE HOMOSCEDASTICITY (EQUAL VARIANCE) ASSUMPTION.

EXAMPLES OF HETEROSCEDASTICITY

① CRIME IN CITIES OF DIFFERENT SIZES:

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

Y = # OF CRIMES

X = POPULATION

IN LARGE CITIES

$$E(Y \mid X = X_i) = \beta_1 + \beta_2 X_i$$

WILL BE LARGER THAN IN SMALL CITIES. BUT VARIATION IN THE NUMBER OF CRIMES ACROSS NYC, ATLANTA, MIAMI etc WILL BE MUCH GREATER THAN IN CITIES LIKE WILLIMANTIC, COLCHESTER, & KILLINGLY.

HERE $E(u_i^2) = \sigma_i^2$ IS NOT A CONSTANT.

(2) LEARNING:

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

Y = # OF MISTAKES IN TYPING

X = YEARS OF EXPERIENCE

AS X INCREASES $E(Y|X=X_i)$ DECLINES AND $E(u_i^2) = \sigma_i^2$ ALSO FALLS.

(3) INCOME & SAVING:

$$S_i = \alpha + \beta Y_i + u_i$$

S = SAVING

Y = INCOME

AT LOWER INCOME, SAVING IS LOW & VARIATION ACROSS INDIVIDUALS IS LOW

AT HIGHER INCOME, SAVING IS HIGHER ALSO MORE CHOICE BETWEEN SAVING & LUXURY CONSUMPTION MEANS VARIANCE IN U_i IS ALSO BIGGER (e.g. BUY A H-D TV VS PUT MONEY INTO BANK).

THE ALGEBRAIC PROBLEM:

SUPPOSE THAT $E(uu') = \sigma^2\Omega$
WHERE Ω IS ANY ARBITRARY
SYMMETRIC & POSITIVE DEFINITE MATRIX.

RECALL THAT $\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u$

$E(\hat{\beta}) = \beta$ &

$$E\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\right] = E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right]$$

$$= (X'X)^{-1}X'E(uu')X(X'X)^{-1}$$

$$= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

$$\neq \sigma^2(X'X)^{-1}.$$

THUS, UNDER THESE CONDITIONS THE USUAL MEASURE OF THE ERROR VARIANCE-COVARIANCE MATRIX IS INCORRECT.

A GENERALIZED LEAST SQUARES (GLS) PROCEDURE

CONSIDER Ω & A TRANSFORMATION MATRIX T
SUCH THAT $T\Omega T' = I$

THEN $$T^{-1}(T\Omega T')(T')^{-1} = T^{-1}(T')^{-1}$$

$$\Omega = T^{-1}(T')^{-1} = (T'T)^{-1}.$$

THAT IS $\boxed{T'T = \Omega^{-1}}$.

LET C BE THE MATRIX OF CHARACTERISTIC VECTORS OF Ω. THEN

$$C \Omega C' = \Lambda$$

WHERE $\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_n \end{pmatrix}$ IS THE DIAGONAL MATRIX OF THE CHARACTERISTIC ROOTS OF Ω

DEFINE $\Lambda^{-1/2} = \begin{pmatrix} 1/\sqrt{\lambda_1} & & \\ & 1/\sqrt{\lambda_2} & \\ & & 1/\sqrt{\lambda_n} \end{pmatrix}$ &

$$\boxed{T = \Lambda^{-1/2} C}$$

THEN $T \Omega T' = \Lambda^{-1/2} C \, \Omega \, C' \Lambda^{-1/2}$

$= \Lambda^{-1/2} \Lambda \Lambda^{-1/2} = I$

NOW CONSIDER $y = x\beta + u$

$E(uu') = \sigma^2 \Omega$

USING THE TRANSFORMATION T,

$$Ty = Tx\beta + Tu$$

DEFINE $y^* = Ty$, $X^* = TX$ & $u^* = Tu$

THEN THE TRANSFORMED MODEL IS

$$y^* = x^*\beta + u^*$$

NOW, $E(u^* u^{*\prime}) = E[Tuu'T'] = TE(uu')T'$

$= \sigma^2 T \Omega T' = \sigma^2 I$.

THUS THE TRANSFORMED MODEL IS HOMOSCEDASTIC.

$$y^* = x^*\beta + u^*$$ 12-4A

$$\beta_* = (x^{*\prime}x^*)^{-1} x^{*\prime} y^*$$

$$= \left[(Tx)'(Tx)\right]^{-1} (Tx)'(Ty)$$

$$= (x'T'Tx)^{-1} x'T'Ty$$

$$= (x'\Omega^{-1}x)^{-1} x'\Omega^{-1} y$$

CONSEQUENCE OF HETEROSCEDASTICITY

RECALL THE 2-VARIABLE REGRESSION

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

$$\hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta_2 + \frac{\sum (X_i - \bar{X})(u_i - \bar{u})}{\sum (X_i - \bar{X})^2}$$

$$E(\hat{\beta}_2) = \beta_2 \quad \left[\text{DOES NOT DEPEND ON THE ASSUMPT} \quad E(u_i^2) = \sigma^2 \right]$$

BUT

$$\text{VAR}(\hat{\beta}_2) = E\left[(\hat{\beta}_2 - \beta_2)^2\right]$$

$$= \frac{\sum (X_i - \bar{X})^2 E(u_i^2)}{\left[\sum (X_i - \bar{X})^2\right]^2} = \frac{\sum (X_i - \bar{X})^2 \sigma_i^2}{\left[\sum (X_i - \bar{X})^2\right]^2}$$

THIS REDUCES TO $\dfrac{\sigma^2}{\sum (X_i - \bar{X})^2}$ ONLY IF $\sigma_i^2 = \sigma^2$ FOR ALL i.

THUS, IN THE ABSENCE OF HOMOSCEDASTICITY,

IF WE USE $se(\hat{\beta}_2) = \dfrac{\hat{\sigma}}{\sqrt{\sum (X_i - \bar{X})^2}}$ TO PERFORM 't' TESTS, THE CONCLUSIONS WILL BE WRONG.

TESTING FOR HETEROSCEDASTICITY

GOLDFELD - QUANDT TEST

CONSIDER THE MODEL

$$y = X\beta + u$$

SUPPOSE THAT THE SAMPLE CAN BE SPLIT INTO 2 GROUPS THUS

$[y_1 \vdots X_1]$ IS THE DATA FOR GROUP 1

& $[y_2 \vdots X_2]$ IS THE DATA FOR GROUP 2

WE SUSPECT THAT EVEN THOUGH β IS THE SAME FOR BOTH GROUPS, THE VARIANCE OF u_i IS σ_1^2 FOR GROUP 1 & σ_2^2 FOR GROUP 2. GREENE CALL THIS GROUP-WISE HETEROSCEDASTICITY.

THUS, THE MODEL IS

$$y_1 = X_1\beta + u_1 \quad ; \quad E(u_1)=0 ; \; E(u_1 u_1') = \sigma_1^2 I$$

$$y_2 = X_2\beta + u_2 \quad ; \quad E(u_2)=0 ; \; E(u_2 u_2') = \sigma_2^2 I$$

WE WANT TO PERFORM A TEST OF

$$H_0 : \quad \sigma_1^2 = \sigma_2^2$$

AGAINST $H_1 : \underbrace{\sigma_1^2 < \sigma_2^2}$.

GOLDFELD & QUANDT PROPOSED THE FOLLOWING F TEST.

STEP 1: RUN 2 SEPARATE REGRESSIONS TO GET

$$\hat{\beta}_{(1)} = (x_1'x_1)^{-1}x_1'y_1 \quad \& \quad \hat{\sigma}_1^2 = \frac{e_1'e_1}{n_1-k}$$

$$\hat{\beta}_{(2)} = (x_2'x_2)^{-1}x_2'y_2 \quad \& \quad \hat{\sigma}_2^2 = \frac{e_2'e_2}{n_2-k}$$

n_1 = # OBSERVATIONS IN GROUP 1

n_2 = # OBSERVATIONS IN GROUP 2

NOTE THAT BOTH $\hat{\beta}_{(1)}$ & $\hat{\beta}_{(2)}$ ARE UNBIASED ESTIMATORS OF THE (COMMON) PARAMETER VECTOR β THAT APPLIES TO BOTH GROUPS. THUS $\hat{\sigma}_1^2$ & $\hat{\sigma}_2^2$ ARE UNBIASED ESTIMATORS OF σ_1^2 & σ_2^2.

STEP 2 COMPUTE THE GOLDFELD-QUANDT STATISTIC

$$GQ = \frac{\hat{\sigma}_2^2/\cancel{n_2-k}}{\hat{\sigma}_1^2/\cancel{n_1-k}} \qquad \frac{e_2'e_2/n_2-k}{e_1'e_1/n_1-k}$$

UNDER THE NULL HYPOTHESIS

$$\frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \sim F(n_2-k,\ n_1-k)$$

STEP 3: REJECT H_0 IN FAVOR OF H_1 IS

GQ EXCEEDS THE CRITICAL VALUE FOR $(n_2-k,\ n_1-k)$ DEGREES OF FREEDOM.

A DIFFERENT USE OF THE G-Q STATE

IN MANY CASES, THERE DOES NOT EXIST A CLEAR CUT SAMPLE SEPARATION BUT IT IS BELIEVED THAT THE VARIANCE OF u_i INCREASES WITH SOME VARIABLE THAT MAY EITHER BE AN EXPLANATORY VARIABLE OR SOME OTHER VARIABLE (NOT INCLUDED IN THE MODEL).

SUPPOSE $\sigma_i^2 = E(u_i^2) = f(x_i)$, $f' > 0$

THEN ONE MAY USE THE G-Q TEST IN THE FOLLOWING WAY:

(1) REARRANGE THE DATA IN ASCENDING ORDER OF X (THE VARIABLE CAUSING THE VARIANCE OF u_i TO INCREASE).

NOTE: THIS DOES NOT MEAN THAT AS A RESULT Y_i ALSO GETS ARRANGED IN ASCENDING ORDER.

(2) DISCARD c CENTRAL OBSERVATIONS. THIS LEAVES $\frac{n-c}{2}$ OBSERVATIONS IN 2 REMAINING GROUPS.

(3) RUN 2 SEPARATE REGRESSIONS & OBTAIN $\hat{\sigma}_1^2$ & $\hat{\sigma}_2^2$ FROM THE SUB-SAMPLES AT THE LOWER END (SMALLER X) & AT THE UPPER END (LARGER X).

(4) COMPUTE

$$GQ = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \sim F_{\frac{n-c}{2}-k,\ \frac{n-c}{2}-k}$$

& PERFORM AN F TEST FOR

$$H_0: \quad \sigma_1^2 = \sigma_2^2$$

AGAINST $H_1: \quad \sigma_1^2 < \sigma_2^2$.

AN INHERENT PROBLEM WITH THIS PROCEDURE IS THAT THE POWER OF THE TEST INCREASES WITH THE NUMBER OF OBSERVATIONS DISCARDED WHILE AT THE SAME TIME DEGREES OF FREEDOM DECLINES AS c BECOMES LOWER. larger

WHITE'S TEST FOR HETEROSCEDASTICITY

WHEN $E(uu') = \sigma^2\Omega$

THE COVARIANCE MATRIX OF THE OLS ESTIMATOR $\hat{\beta}$ IS

$$E\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\right] = \underline{\sigma^2 (X'X)^{-1} X'[\Omega] X (X'X)^{-1}}.$$

THE ESTIMATED $V(\hat{\beta}) = \left\{ (X'X)^{-1} \left[\sum_{i=1}^{n} e_i^2 x_i x_i'\right] (X'X)^{-1} \right\}$

THE CONVENTIONAL ESTIMATOR IS

$$V = \hat{\sigma}^2 (X'X)^{-1}.$$

WHITE PROPOSED THE FOLLOWING TEST:

① USE OLS TO GET THE RESIDUALS $\tilde{e}$

② COMPUTE $e_i^2 \quad i=1,2,\ldots,n$

③ RUN AN AUXILIARY REGRESSION WITH e_i^2 AS THE DEPENDENT VARIABLE & ALL x_j's, x_j^2's & $x_j x_k$'s ALONG WITH A CONSTANT AS RIGHT HAND VARIABLES

④ OBTAIN nR_e^2 WHERE R_e^2 IS THE COEFFICIENT OF DETERMINATION OF THE AUXILIARY REGRESSION.

⑤ UNDER THE NULL HYPOTHESIS OF HOMOSCEDASTICITY nR_e^2 HAS A χ^2 DISTRIBUTION WITH $\boxed{p-1}$ DEGREES OF FREEDOM WHERE p = # OF NON-INTERCEPT COEFFICIENTS IN THE AUXILIARY REGRESSION.

⑥ REJECT H_0 IF $nR_e^2 > \chi_*^2$ WHERE χ_*^2 IS THE CRITICAL VALUE OF χ^2 WITH $p-1$ DEGREES OF FREEDOM.

BREUSCH - PAGAN / GODFREY TEST

THE BP TEST ALLOWS A GENERAL FORM OF HETEROSCEDASTICITY OF THE FORM

$$E(u_i^2) = \sigma_i^2 = \sigma^2 f(\alpha_0 + \alpha' z_i)$$

ALTERNATIVELY $$\frac{\sigma_i^2}{\sigma^2} = f(\alpha_0 + \alpha' z_i)$$

NOTE THAT $E(u_i^2) = \sigma_i^2$.

ONE CAN USE $\hat{u}_i^2$ FOR σ_i^2.

WE THEN RUN THE AUXILIARY REGRESSION

$$\frac{\hat{u}_i^2}{\hat{\sigma}^2} = \alpha_0 + \alpha' z_i + v_i$$

HERE v_i IS A RANDOM ERROR THAT CORRESPONDS TO THE DIFFERENCE BEWEEN $\frac{\hat{u}_i^2}{\hat{\sigma}^2}$ & $\frac{\sigma_i^2}{\sigma^2}$.

STEPS: (1) RUN OLS TO GET RESIDUALS e_i $\boxed{\hat{u}_i \text{ is } e_i}$ They are the same.

(2) COMPUTE $\hat{\sigma}_*^2 = \frac{e'e}{n}$ (THIS IS THE MAXIMUM LIKELIHOOD ESTIMATOR OF σ^2)

(3) REGRESS $\frac{e_i^2}{\sigma_*^2} = e_i^{*2}$ ON THE Z A. Zs

R-square

(4) GET nR^2 FROM THIS REGRESSION

(5) UNDER H_0 OF NO HETEROSCEDASTICITY

$BP = nR^2 \sim \chi^2_p$ WHERE

p = # OF Z VARIABLES INCLUDED IN (3).

How can I correct for heteroscedasticity.

ESTIMATION

SUPPOSE Ω IS KNOWN:

$$VAR(u_i) = \sigma^2 \omega_i$$

$$\Omega = \begin{bmatrix} \omega_1 & & & 0 \\ & \omega_2 & & \\ & & \ddots & \\ 0 & & & \omega_n \end{bmatrix}$$

THEN IF WE DEFINE $T = \begin{bmatrix} 1/\sqrt{\omega_1} & & \\ & 1/\sqrt{\omega_2} & \\ & & 1/\sqrt{\omega_n} \end{bmatrix}$

$$\text{THEN} \quad T'T = \begin{bmatrix} 1/\omega_1 & & \\ & 1/\omega_2 & \\ & & 1/\omega_n \end{bmatrix} = \Omega^{-1}$$

HENCE, FOR GLS,

DEFINE $y^* = Ty = \begin{bmatrix} y_1/\sqrt{\omega_1} \\ y_2/\sqrt{\omega_2} \\ y_n/\sqrt{\omega_n} \end{bmatrix}$

SIMILARLY, SUPPOSE

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix}$$

THEN $\quad TX = X^* = \begin{bmatrix} \frac{1}{\sqrt{\omega_1}} & x_1/\sqrt{\omega_1} \\ \frac{1}{\sqrt{\omega_2}} & x_2/\sqrt{\omega_2} \\ \frac{1}{\sqrt{\omega_n}} & x_n/\sqrt{\omega_n} \end{bmatrix}$

WE TRANSFORM THE MODEL AS:

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

$$\Rightarrow \quad \frac{Y_i}{\sqrt{\omega_i}} = \beta_1 \frac{1}{\sqrt{\omega_i}} + \beta_2 \frac{X_i}{\sqrt{\omega_i}} + u_i/\sqrt{\omega_i}$$

No intercept
w_i is ava

NOW $\quad E\left[\frac{u_i}{\sqrt{\omega_i}}\right] = \frac{E(u_i)}{\sqrt{\omega_i}} = 0$

$$Var\left[\frac{u_i}{\sqrt{\omega_i}}\right] = \frac{E(u_i^2)}{\omega_i} = \frac{\sigma^2 \omega_i}{\omega_i} = \sigma^2.$$

WE NEED TO REGRESS

$y_i^* = \frac{y_i}{\sqrt{\omega_i}}$ ON $x_i^* = \frac{x_i}{\sqrt{\omega_i}}$

& A NEW VARIABLE $\quad z_i = \frac{1}{\sqrt{\omega_i}}$

AS $\quad y_i^* = \beta_1 z_i + \beta_2 x_i^* + u_i^*$

WITHOUT AN INTERCEPT

A FEASIBLE GLS PROCEDURE

ASSUME

$$\mathrm{var}(u \mid x) = \sigma^2 e^{(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k)}$$

is related to any Xs or linear comb. of Xs.

$$\Rightarrow \ln \sigma_i^2 = \ln \sigma^2 + \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k.$$

$$= \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k$$

HERE $\alpha = \ln \sigma^2 + \delta_0$

UNDER THE H_0 OF HOMOSCEDASTICITY,

$$\delta_1 = \delta_2 = \cdots = \delta_k = 0.$$

AGAIN USE e_i^2 FOR σ_i^2.

RUN THE REGRESSION.

$$\ln e_i^2 = \alpha_0 + \delta_1 x_{1i} + \delta_2 x_{2i} + \cdots + \delta_k x_{ki} + v_i$$

GET THE PREDICTED VALUE $\ln \widehat{(e_i)^2}$ & TAKE ITS ANTI LOG TO GET $\hat{\omega}_i$. NOW USE THESE ESTIMATED ω_is TO TRANSFORM THE DATA AS EXPLAINED BEFORE.