CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

MSF Chula

Financial Econometrics

Lecture V

Skedasticity, GLS and FGLS Estimators

2 new estimators
use in heteroskedasticity case

Narapong Srivisal, Ph.D.

Review:

- Model construction:
  - Multicollinearity — one x can be linear combination of another x
- Estimation
  - * OLS estimator – unbiased : $E[u|x] = 0$ hard to get
    - consistent : $cov(u,x) = 0$ consistency is enough
  - Endogeneity → $cov(u,x) \neq 0$ even we have controls
    - OLS estimator is not consistent
    - ↳ causation may face this
- Inference
  - Heteroskedasticity → affect standard error of $\hat{\beta}$ not estimation
    - ↳ $se(\hat{\beta})$

$y = \beta x + u$

3 Interpretations:

1) $E[y|x]$
2) $BLP[y|x]$

} correlation — $cov(u,x) = 0$ always — don't have to care about endogeneity problem

3) Ceteris Paribus effect of x on y
   "Pure effect"

Causation *** → have to think about endogeneity

Note:

unbiased if $E[u|x] = 0$ → only 1st interpretation

consistent if $cov(x,u) = 0$ → only 1st & 2nd interpretation

$E[ux] = E[u]E[x] + cov(u,x)$ ↳ $cov(x,u) = E[ux] - \cancel{E[u]E[x]}$ → 0 for all interpretation

⇓

$E[ux] = 0$ ถึงจะ consistent

MSF Chula

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

AACSB ACCREDITED

EFMD EQUIS ACCREDITED

EFMD EPAS ACCREDITED

Outline

- **Linear regression and OLS in matrix notation**
- **Skedasticity**
    - Homoskedasticity
    - Heteroskedasticity
- **Tests for Heteroskedasticity** whether we have homo / heteroskedasticity in data
- **GLS and FGLS Estimators**

OLS in Matrix Notation

- Recall that we have a linear regression

vector of all regressors

$$\underbrace{Y_i}_{1\times1} = \underbrace{X_i'}_{1\times(k+1)} \underbrace{\beta}_{(k+1)\times1} + \underbrace{U_i}_{1\times1}$$

- Suppose that all the four assumptions stated before are true. Then, we know the value of the parameters

$$\underbrace{\beta}_{(k+1)\times1} = \underbrace{E\left[\underbrace{X}_{(k+1)\times1} X'\right]^{-1}}_{(k+1)\times(k+1)} \underbrace{E\left[\underbrace{X}_{(k+1)\times1}\underbrace{Y}_{1\times1}\right]}_{(k+1)\times1}$$

- The OLS estimator for β is

sample counterpart of

$$\underbrace{\hat{\beta}}_{(k+1)\times1} = \left(\sum_{i=1}^{n} \underbrace{X_i}_{(k+1)\times1} {X_i}'\right)^{-1} \left(\sum_{i=1}^{n} \underbrace{X_i}_{(k+1)\times1} \underbrace{Y_i}_{1\times1}\right)$$

sum of them across observation

- Now, if we stack all the n observations into the matrix form:

bold letter

vector

$$\boldsymbol{X} = \begin{pmatrix} \underbrace{X_1'}_{1\times(k+1)} \\ \vdots \\ \underbrace{X_n'}_{1\times(k+1)} \end{pmatrix}_{n\times(k+1)} ; \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}_{n\times 1}$$

- Then, we can write our OLS estimator as

$$\underbrace{\hat{\beta}}_{(k+1)\times 1} = \left[\begin{pmatrix} \underbrace{X_1}_{(k+1)\times 1} & \cdots & \underbrace{X_n}_{(k+1)\times 1} \end{pmatrix} \begin{pmatrix} \underbrace{X_1'}_{1\times(k+1)} \\ \vdots \\ \underbrace{X_n'}_{1\times(k+1)} \end{pmatrix} \right]^{-1} \left[\begin{pmatrix} \underbrace{X_1}_{(k+1)\times 1} & \cdots & \underbrace{X_n}_{(k+1)\times 1} \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \right]$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

Find $\hat{\beta}$

$$\mathbb{Y} = \mathbb{X}\hat{\beta} + \hat{U}$$

$$\mathbb{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad \hat{U} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix}$$

constant term

x_2 of person 1

data for person 1

$$\begin{pmatrix} 1 & x_{1_1} & x_{2_1} & x_{3_1} & \cdots & x_{k_1} \\ 1 & x_{1_2} & x_{2_2} & x_{3_2} & \cdots & x_{k_2} \\ \vdots & \vdots & & & & \\ 1 & x_{1_n} & x_{2_n} & x_{3n} & \cdots & x_{k_n} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_3 \end{pmatrix}$$

times $\mathbb{X}'$

to make it square

$$\mathbb{X}'\mathbb{Y} = \mathbb{X}'\mathbb{X}\hat{\beta} + \underbrace{\mathbb{X}'\hat{U}}_{\sum_{i=1}^{n} x_i \hat{u}_i = 0}$$

$\sum_{i=1}^{n} x_i \hat{u}_i = 0$ by property of OLS estimator

$$(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{X}\hat{\beta}$$

Variance of the OLS estimator

- When stacking all the observations in matrix, we can roughly think of $\hat{\beta}$ as the solving the model

$$\underbrace{\boldsymbol{Y}}_{n\times 1} = \underbrace{\boldsymbol{X}}_{n\times(k+1)}\hat{\beta} + \underbrace{\widehat{\boldsymbol{U}}}_{n\times 1}$$

- Multiply $\boldsymbol{X}'$ to both side of the equation, and note that $\boldsymbol{X}'\widehat{\boldsymbol{U}} = 0$ by the first-ordered condition of OLS derivation; then, we get

$$\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{X}\hat{\beta} + \boldsymbol{X}'\widehat{\boldsymbol{U}} = \boldsymbol{X}'\boldsymbol{X}\hat{\beta} \Rightarrow \hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

- Then, to find the variance of $\hat{\beta}$ write

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}'\beta + \boldsymbol{U}) = \beta + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{U}$$

$$Var\left[\hat{\beta}\right] = Var\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{U}\right]$$

vector of all observations

$$Var\left[\underbrace{\hat{\beta}}_{\text{vector of all } \beta_i \text{ in model}}|\boldsymbol{X}\right] = \underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{Var}[\boldsymbol{U}|\boldsymbol{X}])\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}}$$

VCE matrix:
in STATA

$$\begin{pmatrix} Var(\hat{\beta}_0|x) & & Cov(\ \ ) \\ & Var[\hat{\beta}_1|x] & \\ & \ddots & \\ Cov(\ \ ) & & Var[\hat{\beta}_k|x] \end{pmatrix}$$

cross sectional analysis
iid except time series
SERIAL CORRELATION

$$\begin{pmatrix} Var(u_1|x) & & & Cov \to 0 \\ & Var(u_2|X) & & \\ & & Var(u_3|x) & \\ Cov \to 0 & & \ddots & Var(u_n|x) \end{pmatrix}$$

SKEDASTICITY : $var(u_i | x_i)$

random variable

$$Var[aU] \Rightarrow a^2 \; Var[U]$$

$$Var[U] \Rightarrow \begin{pmatrix} var[U_1] & cov(U_1, U_2) \\ cov(U_1, U_2) & var[U_2] \end{pmatrix}$$

$$U \downarrow \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$$

---

variance of vector

$$Var\left[\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right] = \begin{pmatrix} var(x) & cov(x,y) & cov(x,z) \\ cov(x,y) & var(y) & cov(y,z) \\ cov(x,z) & cov(y,z) & var(z) \end{pmatrix}$$

E.g. $Var[A \cdot y] \Rightarrow A \, Var[y] \, A'$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \longrightarrow \begin{pmatrix} ay_1 + by_2 \\ cy_1 + dy_2 \end{pmatrix}$$

constant matrix

random vector

vector of all observations

$$Var[\hat{\beta}|X] = \underbrace{(X'X)^{-1}X'(Var[U|X])X(X'X)^{-1}}$$

vector of all β_i in model

Vce matrix: in STATA

$$\begin{pmatrix} Var(\hat{\beta}_0|x) \; \sigma_{11}^2 & & Cov(\;) \\ & Var[\hat{\beta}_1|x] \; \sigma_{22}^2 & \\ & \ddots & \\ Cov(\;) & & Var[\hat{\beta}_k|x] \; \sigma_{nn}^2 \end{pmatrix}$$

$$\begin{pmatrix} Var(u_1|x) & & & Cov \to 0 \\ & Var(u_2|x) & & \\ & & Var(u_3|x) & \\ Cov \to 0 & & \ddots & Var(u_n|x) \end{pmatrix}$$

cross sectional analysis # iid except time series

SERIAL CORRELATION

SKEDASTICITY : $var(u_i|x_i)$

(scedasticity)
SKEDASTICITY

Homo
= 1

constant

identity matrix

$$Var[U|X] = \begin{pmatrix} \sigma^2 & & 0 \\ & \sigma^2 \ddots & \\ 0 & & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & & 0 \\ & 1 \ddots & \\ 0 & & 1 \end{pmatrix}$$

$Var[\hat{\beta}|x] \Rightarrow \sigma^2(x'x)^{-1}$ & $Var[U|X] = \sigma^2 I_n$

$se(\hat{\beta})$ come from short formula

Hetero
= more than 1

$Var(\hat{\beta}|x) \neq \sigma^2(x'x)^{-1}$

wrong $se(\hat{\beta})$

$$Var[\hat{\beta}|X] = (X'X)^{-1}X'(\underbrace{Var[U|X]}_{\sigma^2 I_n})X(X'X)^{-1}$$

$$\sigma^2 (x'x)^{-1} \underbrace{x' I_n x}_{(x'x)} (x'x)^{-1}$$

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

can use this simplified formula

*STATA will assume homoskedasticity

Skedasticity

- If $Var[U|X]$ does not depend on X, then we call that the error term U is **homoskedastic.** ← constant
- If $Var[U|X]$ depends on X then, we call that the error term U is **heteroskedastic.** ← not constant
- So, if we have homoscedasticity, $Var[U|X] = Var[U] = \sigma^2$, then the conditional variance of the OLS estimator can be simplified.
- Stacking all data into a matrix:

$$\underbrace{Var[\boldsymbol{U}|\boldsymbol{X}]}_{n\times 1} = Var(\boldsymbol{U}) = \sigma^2 I_n$$

$$Var\left[\hat{\beta}|\boldsymbol{X}\right] = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\left(\sigma^2 I_n\right)\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- Similarly, the asymptotic variance simplifies to $\Sigma = E[XX']^{-1}Var[U]$
- In the case of near perfectly collinear, the term $(XX')^{-1}$ and $E[XX']^{-1}$ exist but are very large; so, the variance and $se(\hat{\beta})$ will be very large as well.

t-stat ↓ → hard to reject or detect effect of x on y

e.g. cov(x_1, x_2) = 0.9 ใกล้ -1,1
-0.88

Homoskedastic Standard Error

- Since we don't observe U_i, we don't know the parameter Σ but will have to find an estimator for this.
- Using the standard error of regression (SER) as the estimator of standard deviation of U, we have for the homoskedastic case,

$$\hat{\Sigma} = SER^2 \cdot (\boldsymbol{X}\boldsymbol{X}')^{-1} = \left(\frac{1}{n-k-1}\sum_{i=1}^{n}\hat{U}_i^2\right)(\boldsymbol{X}\boldsymbol{X}')^{-1}$$

- Estimator for standard deviation or **standard error of the OLS estimator** $\hat{\beta}$ is $se(\hat{\beta}_j) = \sqrt{\hat{\sigma}_{jj}/n}$, for $\hat{\sigma}_{jj}$ is the jj^{th} entry of $\hat{\Sigma}$

$Var[\hat{\beta} | x] = \sigma^2 (x^2 x)^{-1}$; we have to estimate $\underline{\sigma^2}$

Heteroskedasticity

- Now for the heteroskedastic case, $\boldsymbol{Var[U_i|X] = \sigma_i^2}$ **which varies depending on the observation $\boldsymbol{i}$**.
- So, we get the non-identity diagonal matrix of the variance

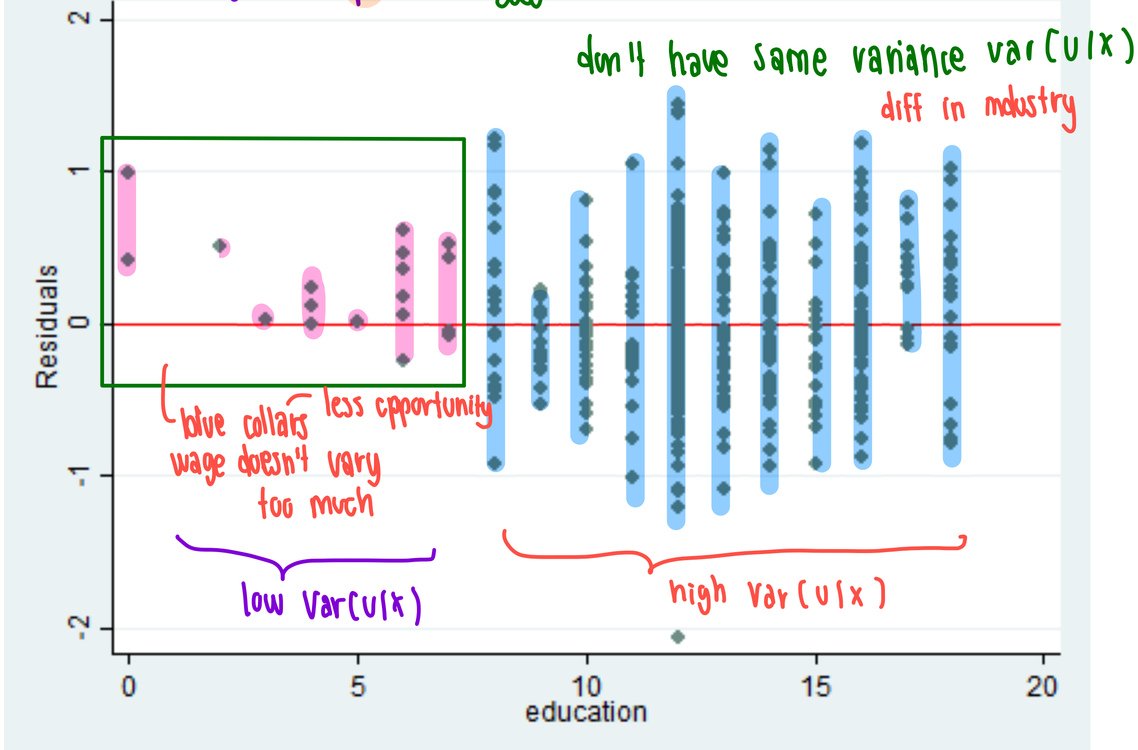
$$Var[\boldsymbol{U}|\boldsymbol{X}] = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix} \equiv \Omega$$

- Therefore, **the covariance matrix of the OLS estimator cannot be simplified**; we need to use this complicated form of variance for inference

$$Var\left[\hat{\beta}|\boldsymbol{X}\right] = (\boldsymbol{XX}')^{-1}\boldsymbol{X}\Omega\boldsymbol{X}'(\boldsymbol{XX}')^{-1}$$

An example of heteroscedasticity: the high-education samples have more spread-out errors than those at the lower end.

$wage = \hat{\beta}_0 + \hat{\beta}_1 edu + controls + \hat{u}$



- Notice that we need to estimate Ω which contains n parameters $(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$ instead of just one value σ^2 as in the homoscedastic case
- In general, we cannot estimate each of the σ_i^2 because it is for each observation i, which means we have one data point to estimate a variance.
- Two ways to proceed:
  - **Construct and use another standard error which is a consistent estimator of standard deviation of $\hat{\beta}$ for the case of heteroskedasticity**
  - Using another estimator than the OLS GLS estimator → transform model from hetero to homoskedastic then use OLS on new model (OLS applied to the transformed)
- Notice that → only variance of error term, impact on standard error **heteroskedasticity** does not affect $E[U|X]$ or $E[XU]$; so, it **has no impact on unbiasedness or consistency of OLS**.

Heteroskedasticity-Robust SE$(\widehat{\beta})$

- Huber White (1980) proved that, instead of estimating each of the σ_i^2, we only need a consistent estimator of the matrix

$$\boldsymbol{X}\Omega\boldsymbol{X}' = \sum_{i=1}^{n} \sigma_i^2 X_i X_i'$$

- and a consistent estimator of such matrix is simply replace σ_i^2 with the residual of the OLS

$(x'x)^{-1} x' \, Var(u|x) \, x(x'x)^{-1}$

$$\sum_{i=1}^{n} \widehat{U}_i^2 X_i X_i'$$

- Thus, he get a heteroskedasticity-consistent covariance matrix is

$$\hat{\Sigma} = (\boldsymbol{XX}')^{-1}\left(\sum_{i=1}^{n} \widehat{U}_i^2 X_i X_i'\right)(\boldsymbol{XX}')^{-1}$$

- We can get the standard error of $\hat{\beta}_j$ as the square root of the (j, j) entry of $\hat{\Sigma}$. This is also called the **heteroskedasticity-robust standard error** or **White's robust standard error**.

can be applied to all cases <both homo & hetero>

MSF Chula

CHULALONGKORN BUSINESS SCHOOL FLAGSHIP FOR LIFE

```
. reg lwage educ exper exper2, nohead
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0903658    .007468    12.10   0.000     .0756948    .1050368
       exper |   .0410089   .0051965     7.89   0.000     .0308002    .0512175
      exper2 |  -.0007136   .0001158    -6.16   0.000     -.000941   -.0004861
       _cons |   .1279975   .1059323     1.21   0.227    -.0801085    .3361035
------------------------------------------------------------------------------
```

same OLS estimator → same β

```
. reg lwage educ exper exper2, nohead robust
```

bcs we use diff formula

```
------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0903658   .0077827    11.61   0.000     .0750766     .105655
       exper |   .0410089   .0050237     8.16   0.000     .0311398     .050878
      exper2 |  -.0007136   .0001098    -6.50   0.000    -.0009292   -.0004979
       _cons |   .1279975   .1071261     1.19   0.233    -.0824537    .3384487
------------------------------------------------------------------------------
```

SKEDASTICITY

Homo

Var (U | X) constant

⇓

Homo se($\hat{\beta}$) → more efficient

White Robust se($\hat{\beta}$)

Hetero

Var (U | X) = f(X)

⇓

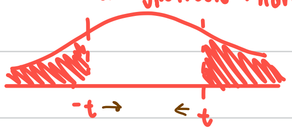
~~Homo se($\hat{\beta}$)~~

White Robust se($\hat{\beta}$)

Often se($\hat{\beta}$) < robust se($\hat{\beta}$)

we use robust se($\hat{\beta}$) in t-stat = $\dfrac{\hat{\beta} - \beta}{se(\hat{\beta}) \uparrow}$ ↓

closer to null hypothesis → harder to reject H_0



, robust → nothing significant (fail to reject H_0) → test: Homo

o.o

Tests for Heteroskedasticity

- If the model is homoscedastic, the simple homoscedastic OLS standard error is more efficient than the robust standard error.
- This implies that usually (but not always!), the robust standard error is larger than the simple homoscedastic OLS standard error.
- So, if the model is homoscedastic, but we use the robust standard error, we may be less likely to reject null hypotheses.
- Hence, if we are not sure if there is heteroskedasticity, we should conduct a hypothesis test for heteroskedasticity based on available dataset first.
- Commonly used tests for heteroskedasticity are **Breusch-Pagan Test** and **White Test**

MSF Chula

$E[U]=0$ for all interpretations

$E[(U-E(U))^2]$

$E[U^2]$

Breusch-Pagan Test

$Var(U|X) = \delta_0 + \delta_1 X_1 + \dots \delta_k X_k$

depends on x → capture by linear function of x

- Since variance is related to the second moment, the idea here is to test whether the U^2 is related to any of the regressors by looking at the linear approximation:

not depend on x

If $\delta_1, \delta_2, \dots, \delta_k = 0$, then it is homo

proxy of variance

$$U^2 = \delta_0 + \delta_1 X_1 + \cdots + \delta_k X_k + \nu$$ → not homo

should be a linear function of regressor

- Then, if variance does not depens on X, all the coefficient except for the intercept should be zero.
- So, this is like a test for multiple linear restriction of the OLS model:

as proxy of U^2 and test this instead to get consistent

$$\widehat{U}^2 = \delta_0 + \delta_1 X_1 + \cdots + \delta_k X_k + \nu$$

$$\boldsymbol{H_0: \delta_1 = \cdots = \delta_k = 0} \text{ vs } \boldsymbol{H_1: \delta_1 \neq 0 \text{ or } \dots \text{ or } \delta_k \neq 0}$$

- Note that we don't know U_i; so, we need to use $\widehat{U}_i$ in the regression instead.

- Then we can conduct an *F*-test using the test statistics:

$$F = \frac{\left(\frac{R^2_{\hat{U}^2}}{k}\right)}{\left(\frac{1-R^2_{\hat{U}^2}}{n-k-1}\right)} \sim F_{k,n-k-1}$$

- Note that here, our restricted model is $\hat{U}^2 = \delta_0 + \nu$; so, the OLS estimator for this restricted model is simply the sample mean of $\hat{U}^2$. Then, ESS and R^2 of the restricted model is zero.

- Another easier test statistics proposed by Breusch and Pagan is the **Lagrange Multiplier statistics**:

follows Chi-square

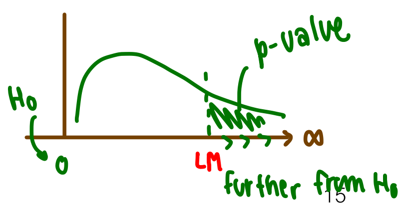
$$\boldsymbol{LM = n \cdot R^2_{\hat{U}^2} \sim \chi^2_k}$$

# of regressors in model
or # of restrictions in H_0

when regress $\hat{U}^2$ for all x

# of observations that you use to run regression

the better x can explain y
the better R^2



1) regress y on all x   No Robust

2) keep $\hat{u}$ $\Rightarrow$ gen $\hat{u}^2$

3) regress $\hat{u}^2$ on all x $\rightarrow$ F-test

$\rightarrow$ LM test

White Test

- The idea of this test is based on the variance matrix that involves the term XX', which contains squares and interactions of the regressors.
- White (1980) then proposes another test which is more general than the Breusch-Pagan test by also restricting that the error terms cannot related to the squares and interactions of the regressors.
- More precisely, White test is like a test for multiple linear restrictions of the OLS model:

proxy for variance

$x'x$

add interaction term

as explanatory variable that can explain variance

$$\hat{U}^2 = \delta_0 + \Sigma_{j=1}^{k} \delta_j X_j + \Sigma_{j=1}^{k} \Sigma_{r=1}^{j} \gamma_{jk} X_j X_k + \nu$$

$$H_0: \text{all the coefficients} = 0 \text{ except } \delta_0$$

- Again we do not know U_i; so, we need to use $\hat{U}_i$ in the regression instead.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

⇒ more control → more terms in model :(but it able to detect HETERO :)

x : x_1 x_2 x_3

x^2 : x_1^2 x_2^2 x_3^2

xx : $x_1 x_2$, $x_1 x_3$, $x_2 x_3$

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_1^2 + \delta_4 x_2^2 + \delta_5 x_1 x_2 + \text{error}$$

- For example, if we have three regressors X_1, X_2, X_3, then running the following regression:

$$\widehat{U}^2 = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3 + \boldsymbol{\gamma_{11} X_1^2 + \gamma_{22} X_2^2 + \gamma_{33} X_3^2}$$
$$+ \boldsymbol{\gamma_{12} X_1 X_2 + \gamma_{13} X_1 X_3 + \gamma_{23} X_2 X_3} + \nu$$

- Then we can conduct an *F*-test in the usual way.

- Or easier, the test statistics proposed by White is the **Lagrange Multiplier statistics**:

$$\boldsymbol{LM = n \cdot R^2_{\widehat{U}^2} \sim \chi^2_M}$$

- where M is the number of regressors (excluding the constant term) in the residual regression model. E.g. if we have X_1, X_2, X_3 like above, then the degree of freedom M is 9

Breusch – Pagan Test by STATA

```
. reg lwage educ exper exper2, nohead
```

suppose to have 9 df in white test ↓ reduce to 8 as $x_3 = x_2^2$

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
x_1 educ	.0903658	.007468	12.10	0.000	.0756948	.1050368
x_2 exper	.0410089	.0051965	7.89	0.000	.0308002	.0512175
x_3 exper2 is x_2^2	-.0007136	.0001158	-6.16	0.000	-.000941	-.0004861
_cons	.1279975	.1059323	1.21	0.227	-.0801085	.3361035

```
.
. * Breusch-pagan test for heteroskedasticity
. estat hettest, rhs
```

tell STATA that you want to use all the right-hand side X as explanatory variable

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: educ exper exper2

         chi2(3)      =    25.48
         Prob > chi2  =   0.0000
```

homo

3 regressor df

→ LM stat

p-value < 1%, 5%, 10%

∴ we reject H_0 ⇒ hetero

You should
1) understand hypothesis
2) the rule to reject H_0
- p-value < significant level
- test stat is further from critical value (need to know more that it is > or <)

White Test by STATA

```
. * White test for heteroskedasticity
. imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(8)      =     23.39
         Prob > chi2  =    0.0029

Cameron & Trivedi's decomposition of IM-test
---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |      23.39      8    0.0029
            Skewness |       0.84      3    0.8406
            Kurtosis |       1.13      1    0.2874
---------------------+-----------------------------
               Total |      25.36     12    0.0132
---------------------------------------------------
```

LM · test

p-value

```
. *ssc install whitetst
. whitetst
```

need to install it first

reject → Hetero

0.29%

```
White's general test statistic :  23.39184  Chi-sq(8)  P-value =  .0029
```

$$y = \beta_0 + \beta_1 \text{female} + \beta_2 \text{edu} + \beta_3 \text{edu} \times \text{female} + u$$

What is the remaining unique term:

x : female , edu , edu · female

x^2 : $\cancel{\text{female}^2}$ (repeat f) , edu^2 , $(\text{edu} \cdot \text{female})^2$ } same

xx : $\cancel{\text{female} \cdot \text{edu}}$, $\cancel{f^2 \cdot \text{edu}}$ (repeat edu · f) , $f \cdot \text{edu}^2$

∴ degree of freedom = 5

เอาตัวซ้ำออก → จะได้ไม่เกิด multicollinearity problem

$$\hat{u}^2 = \delta_0 + \delta_1 f + \delta_2 \text{edu} + \delta_3 \text{edu} \cdot f + \delta_4 \text{edu}^2 + \delta_5 f \cdot \text{edu}^2$$

GLS & FGLS Estimators

MSF Chula

→ transform model until it becomes homo

Generalized Least Squares (GLS)

- Now suppose that we **must know (or willing to assume) a functional form of the heteroskedasticity**, then we can transform the model so that it becomes homoscedastic.

- For example, consider a model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + U_i$$

$$Var[U_i | X_{1i}, X_{2i}] = \sigma^2 X_{1i}$$

- We can instead run the OLS regression of the transformed model

$$\left(\frac{Y_i}{\sqrt{X_{1i}}}\right) = \beta_0 \left(\frac{1}{\sqrt{X_{1i}}}\right) + \beta_1 \sqrt{X_{1i}} + \beta_2 \left(\frac{X_{2i}}{\sqrt{X_{1i}}}\right) + \left(\frac{U_i}{\sqrt{X_{1i}}}\right)$$

- Then, we have homoscedasticity because

$$Var\left[\left(\frac{U_i}{\sqrt{X_{1i}}}\right) \middle| X_{1i}, X_{2i}\right] = \left(\frac{1}{\sqrt{X_{1i}}}\right)^2 Var[U_i | X_{1i}, X_{2i}] = \sigma^2$$

$$y = \beta_0 + \beta_1 x_{1_i} + \beta_2 x_{2_i} + \beta_3 x_{3_i} + u_i$$

need to justify this!!! hard to justify → Umi to robust sd

→ if not → cannot use GLS

treat it as constant

$$\frac{1}{x_{1_i}} \operatorname{Var}(u_i \mid x_i) = \frac{\sigma^2 x_{1_i}}{x_{1_i}} = \sigma^2$$

become constant

$$\operatorname{Var}\left[\frac{u_i}{\sqrt{x_{1_i}}} \,\middle|\, x_i\right] = \sigma^2$$

transform to this

$$\frac{y}{\sqrt{x_1}} = \beta_0\left(\frac{1}{\sqrt{x_1}}\right) + \beta_1\sqrt{x_1} + \beta_2\frac{x_2}{\sqrt{x_1}} + \beta_3\frac{x_3}{\sqrt{x_1}} + \frac{u}{\sqrt{x_1}}$$

} It is not recommended to do this

GENERALIZED LEAST SQUARE

or WEIGHTED LEAST SQUARE (WLS)

* GLS

1st step : $y = \beta x + u$

Assume $Var(u_i \mid x_i) = h^2(x)$ → need to justify whether this is value

$var\left(\frac{u_i}{h(x)} \mid x\right) = 1$

2nd step : $\frac{y}{h} = \beta_0\left(\frac{1}{h}\right) + \beta_1\left(\frac{x_1}{h}\right) + \dots + \beta_k\left(\frac{x_k}{h}\right) + \frac{u}{h}$

* FGLS (FEASIBLE)

estimate h rather than make assumption about it

e.g. $h = \underbrace{\exp}_{\text{always returns positive value}}(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k)$

$\ln h = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k$

proxy as $\hat{u}^2$

$\ln \hat{u}^2 = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k + \varepsilon$

$\hat{\alpha}_0$ $\hat{\alpha}_1$ $\hat{\alpha}_k$

substitute in this

$\hat{h} = \exp(\hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \dots + \hat{\alpha}_k x_k)$ ⇒ then, use it

แนะนำให้ใช้ Robust

- Since we run the OLS regression of the transformed model, then the estimator is

$$\hat{\beta}_{GLS} = \left(\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i{}'\right)^{-1} \left(\sum_{i=1}^{n} \tilde{X}_i \tilde{Y}_i\right)$$

where

$$\tilde{X}_i' = \left(\frac{1}{\sqrt{X_{1i}}} \quad \sqrt{X_{1i}} \quad \frac{X_{2i}}{\sqrt{X_{1i}}}\right); \quad \tilde{Y}_i = \frac{Y_i}{\sqrt{X_{1i}}}$$

- Notice there is **no constant term in this transformed regression.**
- This method is called the **Weighted Least Squares** or **Generalized Least Squares (GLS)** estimation because we are still minimizing sum of squares of the residuals but weighted by $\frac{1}{\sqrt{X_{1i}}}$

- In general, if we know the covariance matrix

$$Var[\boldsymbol{U}|\boldsymbol{X}] = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix} \equiv \Omega$$

- We can find the matrix $\Omega^{-\frac{1}{2}} = \begin{pmatrix} 1/\sigma_1 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_n \end{pmatrix}$

- We can transform the regression $Y_i = X_i'\beta + U_i$ to

$$\Omega^{-\frac{1}{2}}Y_i = \Omega^{-\frac{1}{2}}X_i'\beta + \Omega^{-\frac{1}{2}}U_i$$

- Then, we the **generalized least squares estimator** is

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{GLS}} = \left(\boldsymbol{X}\boldsymbol{\Omega}^{-1}\boldsymbol{X}'\right)^{-1}\boldsymbol{X}\boldsymbol{\Omega}^{-1}\boldsymbol{Y}$$

Feasible GLS

- The problem with GLS is that it requires us to know the true covariance matrix Ω which is unlikely to be the case in real life.
- In practice, we use an estimate of the covariance matrix instead. This method is called **Feasible Generalized Least Squares** or **FGLS**

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{FGLS}} = \left(\boldsymbol{X}\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}'\right)^{-1}\boldsymbol{X}\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{Y}$$

- Typical steps of the FGLS estimation are:
    1. **Run OLS** $Y_i = X_i'\beta + U_\mathrm{i}$ **and get** $\widehat{U}_i$
    2. **Assume a functional form of** Ω **and use** $\widehat{U}_i$ **to get estimate it.**
    3. **Transform the model to** $\widehat{\Omega}^{-\frac{1}{2}}Y_i = \widehat{\Omega}^{-\frac{1}{2}}X_i'\beta + \widehat{\Omega}^{-\frac{1}{2}}U_i$ **and get the OLS estimate of this transformed model.**

- For example, $Var[\boldsymbol{U}|\boldsymbol{X}] = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix} \equiv \Omega$, and assume that the

  variance σ_i^2 takes the exponential form:

$$\sigma_i^2 = \exp\{\alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ki} + \eta_i\}$$

- Then, using $\widehat{U}_i^2$ as an approximation for σ_i^2 we can use OLS to estimate $\alpha_0, \ldots, \alpha_k$ from the regression:

$$\log \widehat{U}_i^2 = \alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_k X_{ki} + \eta_i$$

- The estimated $\hat{\sigma}_i^2 = \exp(\hat{\alpha}_0 + \hat{\alpha}_1 X_{1i} + \cdots + \hat{\alpha}_k X_{ki})$

Review

$$y = \beta x + u$$

$$\hat{\beta}^{OLS} = (x'x)^{-1} x'y$$

vector

$$Var(\hat{\beta}|x) = (x'x)^{-1} \, Var(u|x) \, x \, (x'x)^{-1}$$

variance-covariance matrix

$$\begin{pmatrix} Var(u_1|x_1) & & & & \\ cov(u_2,u_1|x) & Var(u_2|x) & & & \\ cov(u_3,u_1|x) & cov(u_3,u_2|x) & Var(u_3|x) & & \\ & & & \ddots & \\ & & & & var(u_n|x) \end{pmatrix}$$

Serial correlation

↳ SKEDASTICITY

is all about $se(\hat{\beta})$, affect inference step

$\hat{\beta}^{OLS}$ still the same

estimator

transform model until error term become
(homo without serial correlation ⇒ use OLS

Homo

$$Var(u_i|x) = \sigma^2$$

$$Var(\hat{\beta}|x) \text{ simplified}$$

default option in STATA simplified formula

$se(\hat{\beta})$ Homo ← this is more efficient!!!

White's Robust $se(\hat{\beta})$

↳ this is higher

└ may not find significant result ← we want this!

Hetero

FGLS instead of OLS
(then get different $\hat{\beta}$ based on some assumptions

$$Var(u_i|x) = \sigma_i^2$$

~~$se(\hat{\beta})$ homo~~

White's Robust $se(\hat{\beta})$ → can be used in both homo & hetero

recommend to use white

Detect Heteroskedasticity

⇓

* Breusch - Pagan test $\hat{u}^2 = \boxed{SX}$

Lagrange multiplier White test

x^2

xx

when use white test

↳ check whether there's repeat x

LM : $n \cdot R^2 \sim \chi^2$