[Academic Script]

Generalised Least Squares

Subject:	Business Economics
Course:	B. A. (Hons.), 5th Semester, Undergraduate
Paper No. & Title:	Paper – 531 Elective Paper Q1 – Advanced Econometrics
Unit No. & Title:	Unit – 1 Relaxing the Assumptions of The Classical Linear Model
Lecture No. & Title:	Lecture – 4 Generalised Least Squares

Academic Script

1. Introduction

Hello friends nice meeting you. We will be discussing today about generalized least square. You know that in the GLM when the disturbances are such that there variances are uniform then it is known as homoscadasticity. But if the variances are different then it is known as heteroscadasticity. In a way we come across spherical disturbances. To tackle this problem we should know what are the tests associated with this one and what are the remedial measures which can deal with this problems all this things we will discuss in detail in this lecture.

We want to study now a particular case of k Variate GLM when the assumption about homoscadasticity is no longer valid.

In classical GLM, we assumed that $E(\underline{U}) = \underline{0}$ and $V(\underline{U}) = E(\underline{U}\,\underline{U'})$ $= \sigma^2 I_n$ which meant uniform variance for disturbance term and also that no pair of disturbances are correlated.

Now we wish to examine a situation when this assumption is violated.

In real life situation, the disturbance terms U_i are hardly found to be having uniform variance and also pair wise mutually uncorrelated. This refers to a very broad presentation of the nature of disturbances. Such disturbances are called **non-spherical disturbances**.

We examine this situation which is called heteroscadasticity.

Let us consider our k variate GLM in the form of a single matrix equation

$\underline{Y} = X\underline{\beta} + \underline{U}$ $\qquad$ $\underline{Y} : n \times 1$, $X: n \times K$, $\underline{\beta} : k \times 1$, $\underline{U}: n \times 1$

We assume now that

$E(\underline{U}) = \underline{0}$ and$V(\underline{U}) = E(\underline{U}\,\underline{U'}) = \sigma^2 V = \Omega$

Where $V = (v_{ij}): n \times n$ is a matrix which is symmetric and positive definite.

[This assumption shows that $V(U_i) = \sigma_i^{\,2} = \sigma^2\, v_{ij}$ (i = 1,2,...,n) and $COV(U_i, U_j) = \sigma^2 v_{ij} \neq 0$ i, j $= 1,2,\ldots,n)$ $i \neq j$

Here we have to estimate the following parameters

(i) Unknown σ^2

(ii) (ii)Unknown k beta coefficients

(iii)Unknown terms v_{ij}

(iv)Unknown nc_2 terms v_{ij}

Thus we have totally (1+k+n+ nc_2) unknown parameters to be estimated, which comes to a total of $(1 + k + \frac{n(n+1)}{2})$ unknowns.

We have only n equations stated by the model. Thus it is not possible to obtain unique solution, if we apply our usual ordinary least squares method.

The problem now is how to estimate them?

This is done by <u>Generalised Least Squares</u> method and hence the linear model in this situation can also be called <u>Generalised Least Squares Model</u> (GLS model).

Estimation Procedure

[Suppose that we use ordinary Least Square method to estimate beta coefficients in such case also.

Then minimizing $S^2 = \underline{e}'\,\underline{e}$ gives us $\underline{\hat{\beta}} = (X'X)^{-1}X'\,\underline{U}$ so that $E(\underline{\hat{\beta}}\,)$ $= \underline{\beta}$ but

$$V(\underline{\hat{\beta}}\,) = E[\left(\underline{\hat{\beta}} - \underline{\beta}\,\right)\left(\underline{\hat{\beta}} - \underline{\beta}\right)']$$

$$= E[(X'X)^{-1}X'(\,\underline{U}\,\underline{U'})\text{ X }(X'X)^{-1}]$$

$$= [(X'X)^{-1}X'VX\,((X'X)^{-1}]\,\sigma^2 \neq \sigma^2(X'X)^{-1}$$

Hence ordinary Least Squares give unbiased estimator but will not have minimum variance property.]

We use the same Least Squares principle for estimation here also, but quite differently.

Thus determine estimators of $\underline{\beta}$ by minimizing residual sum of squares

M $= (\underline{Y} - X\underline{\beta})'.[\mathrm{E}(\underline{U}\,\underline{U}')]^{-1}\,(\underline{Y} - X\underline{\beta})$

$$= \frac{(\underline{Y}-X\underline{\beta})' V^{-1} (\underline{Y}-X\underline{\beta})}{\sigma^2}$$

Since V is a symmetric positive definite matrix, there exists a non-singular matrix P such that

$P'P = V^{-1}$

We have $\underline{Y} = X\underline{\beta} + \underline{U}$ .... (1)

Hence $P\,\underline{Y} = PX\underline{\beta} + P\,\underline{U}$

Write $P\,\underline{Y} = \underline{Y}^*$, $PX = X^*$, $P\underline{U} = \underline{U}^*$

Then $\underline{Y}^* = X^*\,\underline{\beta} + \underline{U}^*$ .... (2)

Here $E(\underline{U}^*) = E(P\underline{U}) = 0$

$E(\underline{U}^*\,\underline{U}'_*) = E(P\underline{U}\,\underline{U}'\,P')$

$= \sigma^2(PVP')$ $\qquad\qquad PVP' = \mathrm{I}_n$

$= \sigma^2\,\mathrm{I}_n$

And Rank of $X^* = \rho(X^*) = \rho(X)$

Thus the transformed model (2) has homoscadasticity even though original has heteroscadasticity so we get estimator of $\underline{\beta}$ denoted by $\hat{\underline{\beta}}_g$ by the formula

$\hat{\underline{\beta}}_g = (X'VX)^{-1} \quad X'V^{-1}\underline{Y}$ ....

(3)

Which is called Generalized Least Squares estimator (GLSE) of $\underline{\beta}$.

$V\left(\hat{\underline{\beta}}_g\right) = (X'V^{-1}X)^{-1}\sigma^2$ ……………………………………………………………………

(4)

Note that GLSE $\hat{\underline{\beta}}_g$ is also BLUE for $\underline{\beta}$.

Writing $\sigma^2 V = \Omega$,

$V\left(\hat{\underline{\beta}}_g\right) = (X'\Omega^{-1}X)^{-1}$ ……………………………………………………………………

(5)

(Note that GLSE of $\underline{\beta}$ is also called Aitken's estimators)

Estimation of $\underline{\sigma^2}$ if V is Known

From the transformed model given in (2) above, we can find estimate of σ^2 when V is known.

As usual, $\hat{\sigma}^2 = \frac{(\underline{e}^{*\prime}\ \underline{e}^{*})}{(n-K)}$ and hence unbiased estimator of σ^2 is given by

$$S^2 = \hat{\sigma}^2 = \frac{1}{(n-K)}\left[\left(\underline{Y} - X\hat{\underline{\beta}}_g\right)' V^{-1}\left(\underline{Y} - X\hat{\underline{\beta}}_g\right)\right]$$

Where $\hat{\underline{\beta}}_g$ is GLSE of $\underline{\beta}$ as shown above.

Some Testing Problems

If we assume normality for disturbances then $\underline{U} \sim \mathrm{N}\left[\underline{0}, \sigma^2 V\right]$

Hence $\hat{\underline{\beta}}_g \sim N[\underline{\beta}, (X'V^{-1}X)^{-1}\sigma^2]$

And any linear combination Z = $\underline{C}'\ \ \hat{\underline{\beta}}_g$ has also normal distribution given by

$$Z = \ \hat{\underline{\beta}}_g \sim N[\underline{C}'\underline{\beta}, \{\underline{C}'(X'V^{-1}X)^{-1}\underline{C}\}\ \sigma^2]$$

This gives some testing problems.

(1) To test $H_0 : \underline{\beta} = \underline{\beta}_0$ Versus $H_1 : \underline{\beta} \neq \underline{\beta}_0$

Here $\underline{\beta}_0$ is a specified value of $\underline{\beta}$

We apply t test by $t = \frac{(\hat{\underline{\beta}}_g - \underline{\beta}_0)}{(S.E.of\ \hat{\underline{\beta}}_g)} = \frac{(\hat{\underline{\beta}}_g - \underline{\beta}_0)}{S}$

Where $s^2 = \frac{1}{(n-k)}\left[\left(\underline{Y} - X\underline{\hat{\beta}_g}\right)' V^{-1}\left(\underline{Y} - X\underline{\hat{\beta}_g}\right)\right]$

Here t has (n-k) degrees of freedom

Hence apply the test as usual.

(2) To test H_0: $\underline{C}'\underline{\beta} = C_0$ against H_1: $\underline{C}'\underline{\beta} \neq C_0$

Here C_0 is some specified value of $\underline{C}'\underline{\beta}$.

We have $t = \frac{\underline{C}'\,\underline{\hat{\beta}_g} - C_0}{\sqrt{V(\underline{C}'\,\underline{\hat{\beta}_g})}}$ which has student's t distribution with (n$-k$) degrees of freedom and s is S.E. of $(\underline{C}'\ \underline{\hat{\beta}_g})$

Thus $t = \frac{\underline{C}'\,\underline{\hat{\beta}_g} - C_0}{s\sqrt{\underline{C}'\,(X'V^{-1}X)^{-1}\,\underline{C}}}$

And now t test can be applied as usual.

[Note These results can be extended for the case of linear restrictions upon Beta coefficients. e.g. $\underline{C}'\ \underline{\hat{\beta}_g} = d$ is the linear restrictions imposed upon the parameters. Here $\underline{C}'$and $\underline{d}$ are known, and further test procedure can be carried out. As an example, test for constant returns to scale in cobb Douglass production function means to test whether $\alpha + \beta = 1$ or not, which refers to the parametric restrictions.]

2. A specific Heteroscadasticity Problem

Let us consider a specific situation for heteroscadasticity when the matrix V takes a particular form as diagonal matrix.

Thus for the model $\underline{Y} = X\underline{\beta} + \underline{U}$

We have $E(\underline{U}) = \underline{0}$ and $V(\underline{U}) = E(\underline{U}\,\underline{U}') = \sigma^2$ V

Where V$= diag(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_n})$

(Here diagonal elements are arbitrary constants $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_n}$ and non-diagonal elements are zero).

This shows that $V(U_i) = \frac{\sigma^2}{\lambda_i}$ (i$= 1,2,\ldots,n$)

and COV $(U_i, U_j) = 0$ i $= 1,2,..,$n j $= 1,2,\ldots,n$ (i$\neq j$)

If each $\lambda_i = 1$ (i$= 1,2,\ldots n$) there is homoscadasticity.

$V^{-1} = diag(\lambda_1, \lambda_2, \ldots., \lambda_n)$.

Thus $\underline{\hat{\beta}_g} = \{X' diag(\lambda_1, \lambda_2, \ldots., \lambda_n) X\}^{-1} X' diag(\lambda_1, \lambda_2, \ldots., \lambda_n) \underline{Y}$

And $V\left(\underline{\hat{\beta}_g}\right) = (X'V^{-1}X)^{-1}\ \sigma^2 = [X' diag(\lambda_1, \lambda_2, \ldots., \lambda_n) X]^{-1} \sigma^2$.

Choice of λ coefficient is arbitrary, hence it can be shown that $\underline{\hat{\beta}_g}$

can be made more efficient than OLSE of $\underline{\hat{\beta}}$ by choosing λ's.

We illustrate this by two variables model.

Illustration

Let $Y_i = \beta_1 + \beta_2 X_i + U_i$ (i$= 1,2,..,n$)

With heteroscadastic situation.

$$\underline{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} : n \times 1, \text{ X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} : n \times 2\ ,\ \underline{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} : 2 \times 1$$

Then $X'V^{-1}X = \begin{bmatrix} \sum \lambda_i & \sum \lambda_i X_i \\ \sum \lambda_i X_i & \sum \lambda_i X_i^2 \end{bmatrix} : 2 \times 2$

And $X'V^{-1}\underline{Y} = \begin{pmatrix} \sum \lambda_i Y_i \\ \sum \lambda_i X_i Y_i \end{pmatrix} : 2 \times 1$

Hence $\begin{bmatrix} \sum \lambda_i & \sum \lambda_i X_i \\ \sum \lambda_i X_i & \sum \lambda_i X_i^2 \end{bmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \sum \lambda_i Y_i \\ \sum \lambda_i X_i Y_i \end{pmatrix}$

Which determines GLSE of β_1 andβ_2 .

$$V\left(\underline{\hat{\beta}}\right) = \sigma^2 (X'V^{-1}X)^{-1} = \sigma^2 \begin{bmatrix} \sum \lambda_i & \sum \lambda_i X_i \\ \sum \lambda_i X_i & \sum \lambda_i X_i^2 \end{bmatrix}^{-1}$$

Thus $V(\hat{\beta}_2) = \frac{\sigma^2 \sum \lambda_i}{(\sum \lambda_i)(\sum \lambda_i X_i^2) - (\sum \lambda_i X_i)^2}$

Which gives $V(\hat{\beta}_2)$ under heteroscadasticity with

$V(U_i) = \sigma^2/_{\lambda_i}$ (i$= 1,2,..n$)

Here choice of λ coefficients is arbitrary.

If we choose $\lambda_i = \frac{1}{X_i^2}$ then it can be shown that $\hat{\beta}_2$ is more efficient than its OLSE.

Milder and Harder Heteroscadasticity

(I) If $V(U_i) = E(U_i^2) = \sigma^2 X_i^2$ (i=1 ,2,..n), it is called Harder Heteroscadasticity. (Choosing $\lambda_i = \frac{1}{X_i^2}$)

(II) If $V(U_i) = E(U_i^2) = \sigma^2 X_i$ (i= 1, 2,..,n) it is called Milder Heteroscadasticity.(Choosing $\lambda_i = \frac{1}{X_i}$)

We can have similar other forms also for harder and milder Heteroscadasticity.

As for example, $V(U_i) = \sigma^2 \sqrt{X_i}$ is still milder form and $V(U_i) = \sigma^2 X_i^p$ will be harder form of heteroscadasticity.

It may be noted that in all such cases variance of disturbance term changes rapidly with the change in the corresponding explanatory variable. This also shows that Heteroscadasticity is an important issue to be taken care of.

We now want to study the problem of Heteroscadasticity in more details by means of the following

(1) What are the reasons for Heteroscadasticity ?

(2) How Heteroscadasticity is generated?

(3) How to detect Heteroscadasticity?

(4) How to tackle the problem of Heteroscadasticity?

What are the reasons for Heteroscadasticity?

(1) Due to Error Learning models

As time goes on, the behavior of errors becomes well known and thus it tends to decrease the errors. Due to this reason the spread or dispersion decreases.

(2) As Incomes Grow

Due to increase in Income, as per the passage of time, the choice of disposition of income also changes. Thus savings on income grows and also varies with time.

Due to this, the variance of the disturbance term also changes.

(3) As Efficiency increases

Due to the sophisticated technological development the data processing works in banks, government and corporate offices will improve and hence efficiency of the system increases. This can result in reducing the error.

(4) More heteroscadasticity in cross sectional data

In cross sectional data, since the observations are given over a point of time only as compared to the entire time series, it is generally observed that the error may not be the same throughout and hence there is heteroscadasticity problem.

(5) Due to presence of outliers

An outlier is an observation that is much different (very small or very large) in relation to the other observations in the sample. If such an outlier is imposed (or detected) in the sample data it causes variation in error thus resulting into heteroscadasticity phenomena.

How heteroscadasticity is generated?

While dealing with analysis of data due to some of the reasons mentioned below, heteroscadasticity is generated even though originally there is no such phenomena.

(1) By grouping of observations

Many times the cross section or time series data are grouped as per the range of the number of values of the dependent variable. If we are dealing with Large data, then also we need grouping of observations. This has an effect that even though originally there is homoscadasticity, due to grouping of observations we find that there is heteroscadasticity. Hence GLSE should be used instead of OLSE for grouped data.

(2) By grouping of equations

In a similar way, when we deal with Large number of equations, all having similar properties of homoscadasticity, we may group these equations as per some laid characteristics or norms.

What we find that due to grouping of equations there is heteroscadasticity and hence proper equations should be used to deal with the problem.

(3) By imposing stochastic Linear Constraints on the parameters

Due to restrictions done by means of stochastic linear constraints upon the parameters, even though the original model has homoscadasticity, the new model so developed has heteroscadasticity. Here also we should use GLSE instead of OLSE.

3. How heteroscadasticity is detected?

There are different ways to detect the presence of heteroscadasticity. We now want to know them one by one in nutshell.

(A) Graphical Method

We run the regression data as usual assuming no heteroscadasticity. Then we can draw scatter diagrams as shown in the following diagrams. On X axis we can take $\hat{Y}_i$ (or X_i) and on Y axis we can take e_i^2. This shows the variational pattern representing the situation of heteroscadasticity. In case of more $X's$ as explanatory variables, we may take any particular X_i on X axis e_i^2 gives $\hat{u}_i^2$ thus showing what type of heteroscadasticity, we may have.

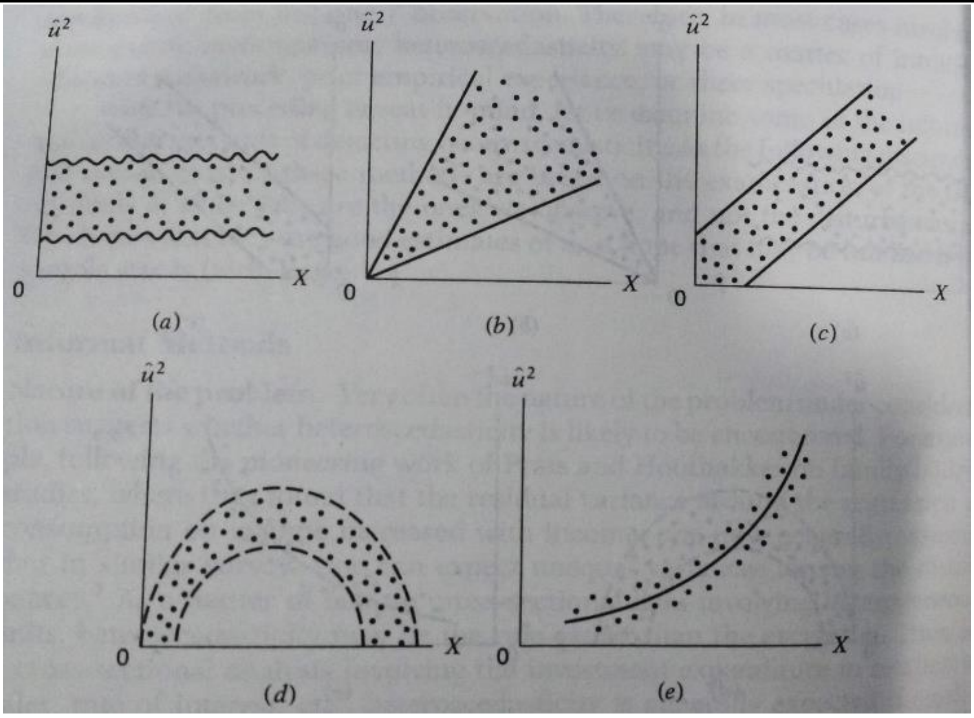


Fig (a) → No systematic pattern, showing no heteroscadasticity

Fig (b) → exhibits some systematic pattern
Fig (c) → exhibits some systematic pattern (May be linear)

Fig (d) → exhibits some systematic pattern
Fig (e) → exhibits some systematic pattern (May be quadratic or curvilinear)

This is an informal way for detecting heteroscadasticity.

(B) Formal Methods

We have some tests for detecting heteroscadasticity. Some of them are (1) spearman's Rank correlation test (2) Park test (3) Glejser test (4) Goldfield and Quandt test etc.

(1) Spearman's Rank Correlation test

We consider model $Y_i = \beta_1 + \beta_2 X_i + U_i$

Step1 Run this regression and find $e_i = Y_i - \hat{Y}_i$

Step2 Find $|e_i|$ and find rank correlation coefficient between $|e_i|$ and X_i.

If $d_i =$ Difference between Rank of $|e_i|$ and X_i, find rank correlation coefficient by the formula

$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$ $\quad r_s =$ Spearman's rank correlation coefficient

Step3 Test $H_0: \rho_s = 0$ against $H_1: \rho_s \neq 0$.

This is done by t test $t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}}$ with (n-2) degrees of freedom

$\rho_s =$ Population correlation coefficient

If $t(calculated) > t(tabulated) \to$ we reject H_0

(i.e. there is heteroscadasticity.)

If there are more X variables in the model, we can carry out this test for each respective X_i against $|e_i|$ and decide about heteroscadasticity.

(2) Park test

Park assumed that σ_i^2 is some function of X_i.

He gave the functional form $\sigma_i^2 = \sigma^2 X_i^{\beta}\, e^{\vartheta_i}$

Where ϑ_i is the stochastic disturbance term.

Then $\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + \vartheta_i$

(ln is natural logarithm).

Since σ_i^2 is unknown, it is estimated by $\hat{u}_i^2$ as a proxy, so that above relation becomes

$$\ln \hat{u}_i^2 = \ln \sigma^2 + \beta \ln X_i + \vartheta_i$$

$$= \alpha + \beta \ln X_i + \vartheta_i \quad \text{where } \alpha = \ln \sigma^2$$

We carry out this test in two stages.

First stage

(i) Run the regression of Y on X_i as usual without bothering about heteroscadasticity.

(ii) Then find $\hat{u}_i^2 = e_i^2$ from this regression

Second stage

(i)Now run the regression

$$\ln e_i^2 = \alpha + \beta \ln X_i + \vartheta_i$$

(ii) Test for β . If it is significant, there is heteroscadasticity.

Comments This test suits graphical approach. Since ϑ_i is stochastic error term, it does not satisfy the assumptions of OLS. It may itself be heteroscadastic. (Goldfield and Quandt)

(3) Glejser Test

Similar to park test, Glejser assumed certain relationship between| $\hat{u}_i^2$| and X_i.

(Note that $\widehat{U}_i^2 = \mathrm{e}_i^2$)

First stage Run the regression as usual and find $\hat{u}_i^2 = \mathrm{e}_i$ and hence $|\mathrm{e}_i|$

(1) $|\mathrm{e}_i| = \beta_1 + \beta_2 X_i + \vartheta_i$

(2) $|\mathrm{e}_i| = \beta_1 + \beta_2 \sqrt{X_i} + \vartheta_i$

(3) $|\mathrm{e}_i| = \beta_1 + {}^{\beta_2}/_{X_i} + \vartheta_i$

(4) $|\mathrm{e}_i| = \beta_1 + {}^{\beta_2}\Big/_{\sqrt{X_i}} + \vartheta_i$

(5) $|\mathrm{e}_i| = \sqrt{\beta_1 + \beta_2 X_i} + \vartheta_i$

(6) $|\mathrm{e}_i| = \sqrt{\beta_1 + \beta_2 {X_i}^2} + \vartheta_i$

Where ϑ_i is the error term

Run regression and if β_2 is significant , there is heteroscadasticity.

Comments $E(\vartheta_i)$ can be non-zero and it is serially correlated and hence heteroscadastic. Last 2 models are non-linear and OLS cannot be obtained with usual procedure.

Glejser test is good for large samples.

(4) Goldfield and Quandt test

Here assumption is that one of the explanatory variables is responsible for heteroscadasticity and hence σ_i^2 is positively related to one of the explanatory variables.

Consider two variables model

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

And $\sigma_i^2 = \sigma^2 X_i^2$ (Harder heteroscadasticity)

Here σ^2 is a constant.

Here σ_i^2 will be larger for larger value of X_i.

Step: 1 Order or rank the observations according to the values of X_i, starting with the lowest X value.

Step: 2 Omit C central observations as per values of X_i, starting with the lowest X value.

Step: 3 Fit separate regressions for first $\frac{(n-c)}{2}$ and last $\frac{(n-c)}{2}$ observations, and find respective residual sum of squares $(RSS)_1$ and $(RSS)_2$ for the two groups. Choice of c is arbitrary.

$(RSS)_1$ = RSS for the smaller X_i values

(Small variance group)

$(RSS)_2$ = RSS for the Larger X_i values

(Large variance group)

Each RSS has $\left(\frac{n-c}{2}\right)$ – k $= \left(\frac{n-c-2K}{2}\right)$ d. f.

Where k is the number of parameters to be estimated (including the intercept).

Step: 4 compute the ratio $\lambda = \frac{(RSS)_2 / d.f}{(RSS)_1 / d.f}$

Under normality assumptions for u_i, λ has F distribution with d. f. each $\left(\frac{n-c-2K}{2}\right)$.

If F is significant, there is heteroscadasticity, otherwise there is homoscadasticity.

Note

(1) If there are more than two X variables, ranking of observations can be done for any one of the $X's$ and then others follow accordingly.

This way we can conduct the test for each of the X variables.

(2) Choose $c = 4$ for $n = 30$ and $c = 10$ for $n = 60$.

(3) Besides the above mentioned tests for heteroscadasticity, we also have some other tests which are mentioned here as-

Breusch-Pagan-Godfrey test (BPG test),

White's general heteroscadasticity test etc. as some of the advanced methods

4. How to tackle the problem of Heteroscadasticity?

We study in brief about some remedial measures to solve the problem of Heteroscadasticity.

Since OLS estimators under Heteroscadasticity are not efficient, some remedial measures are needed.

We study them in two parts:

I. When σ_i^2 is known. II. When σ_i^2 is unknown.

I. When σ_i^2 is known

Here we use the **method of weighted least squares (WLS).**

We write $V(U_i) = E\left(U_i^2\right) = \sigma_i^2, (i = 1, 2, \ldots\ldots, n)$ and since σ_i^2 are known, we use them as weights to resolve the problem. Let us write $W_i = \frac{1}{\sigma_i^2} (i = 1, 2, \ldots\ldots, n)$. This ensures to give less weightage for higher variation and more weightage for lower variation. Thus weights are inversely proportional to the variance σ_i^2.

Let us consider two variables model

PRF $\quad Y_i = \beta_1 + \beta_2 X_i + U_i$

SRF $\quad Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i, \quad (i = 1, 2, \ldots, n)$

We obtain estimators of β_1 and β_2 under the case of heteroscadasticity by writing β_1^* and β_2^*, then these estimators are obtained by minimizing the error sum of squares

$S^2 = \sum_1^n W_i e_i^2 = \sum_1^n W_i (Y_i - \beta_1^* - \beta_2^* X_i^*)^2$

Which gives $\beta_2^* = \frac{\sum w_i x_i^* y_i^*}{\sum w_i x_i^{*2}}$ and $\beta_1^* = \bar{Y}^* - \beta_2^* \bar{X}^*$

Here $x_i^* = X_i - \bar{X}^*$ $\quad\quad$ $\bar{X}^* = \frac{\sum w_i x_i}{\sum w_i} = weighted\ mean$

$y_i^* = Y_i - \bar{Y}^*$ $\quad\quad$ $\bar{Y}^* = \frac{\sum w_i Y_i}{\sum w_i} = weighted\ mean$

(Check that if $W_i = \frac{1}{\sigma^2}$ (No Heteroscadasticity) we get the results for OLSE under homoscadasticity)

Also if all weights are same, that is $W_i = w$ for all i=1, 2, ….,n. We conclude that here WLSE and OLSE are the same.

For WLSE, formula for $V(\beta_2^*)$ is given by,

$$V(\beta_2^*) = \frac{\sum W_i}{(\sum W_i)(\sum W_i X_i^2) - (\sum W_i X_i)^2}$$

Note that under Heteroscadasticity the OLS estimators are unbiased and consistent, however they are not efficient in small as well as large samples.

II. When σ_i^2 is Unknown

In this case we have a number of assumptions to solve the problem of Heteroscadasticity.

Assumption (1) When $V(U_i) = E(U_i^2) = \sigma^2 X_i^2\ for\ all\ i$

(This means that variance of the disturbance term is proportional to square of the explanatory variable which is Harder Heteroscadasticity)

Consider the model $Y_i = \beta_1 + \beta_2 X_i + U_i$………………………… (1)

We divide both the side by X_i

Hence $\frac{Y_i}{X_i} = \frac{\beta_1}{X_i} + \beta_2 + \vartheta_i$…………………………………….… (2)

Where $\vartheta_i = \frac{U_i}{X_i}$

$E(\vartheta_i) = 0, E(\vartheta_i^2) = \sigma^2$

This shows that model (1) has heteroscadasticity but transformed model in (2) has homoscadasticity.

We can run this regression of Y_i against $\frac{1}{X_i}$ and find OLS estimators, Note that slope and intercept terms are changed in new model compared with original.

Assumption (2)When$E(U_i^2) = \sigma^2 X_i$ (Milder heteroscadasticity)

$E(U_i^2) = \sigma^2 X_i = \sigma^2 (\sqrt{X_i})^2$

Hence we transform the model dividing by $\sqrt{X_i}$

So that $\frac{Y_i}{\sqrt{X_i}} = \beta_1(\frac{1}{\sqrt{X_i}}) + \beta_2(\sqrt{X_i}) + \vartheta_i$.. (3)

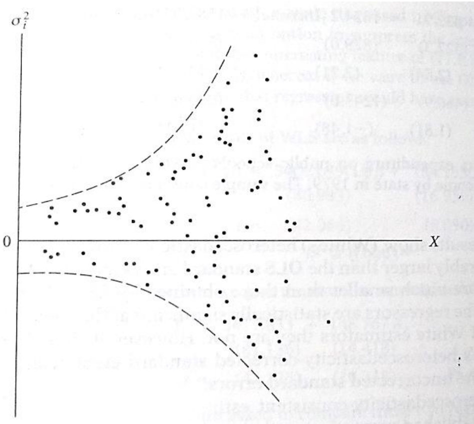
Where $\vartheta_i = \frac{U_i}{\sqrt{X_i}}, E(\vartheta_i) = 0, V(\vartheta_i) = \sigma^2$

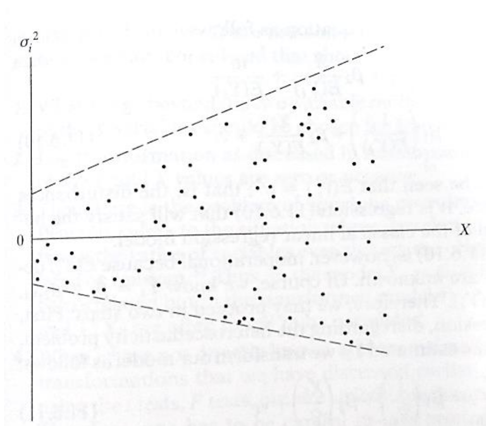
Thus transformed model has homoscadasticity even though original model has heteroscadasticity.

We run the regression of $\frac{Y_i}{\sqrt{X_i}}$ on the variable $\frac{1}{\sqrt{X_i}}$ and $\sqrt{X_i}$ and find OLS estimators.

This has an important feature. There is no intercept term. Hence regression is to be done through the original model to estimate β_1 and β_2.

Following diagrams represent these two situations for assumption (1) and assumption (2)





Assumption (1) $\sigma_i^2 \propto X_i^2$ **Assumption (2)** $\sigma_i^2 \propto X_i$

Harder Heteroscadasticity Milder Heteroscadasticity

Assumption (3) When $E(U_i^2) = \sigma^2[E(Y_i)]^2$ *for all i*

We have the model $Y_i = \beta_1 + \beta_2 X_i + U_i$.................................. (1)

Then $E(Y_i) = \beta_1 + \beta_2 X_i$... (2)

We transform the model by writing

$$\frac{Y_i}{E(Y_i)} = \frac{\beta_1}{E(Y_i)} + \beta_2 \frac{X_i}{E(Y_i)} + \frac{U_i}{E(Y_i)}$$

$$= \beta_1 \left(\frac{1}{E(Y_i)}\right) + \beta_2 \left(\frac{X_i}{E(Y_i)}\right) + \vartheta_i \text{.................................} \quad (3)$$

Here $\vartheta_i = \frac{U_i}{E(Y_i)}$ $\quad E(\vartheta_i) = 0 \; and \; V(\vartheta_i) = \sigma^2$

Thus transformed model has homoscadasticity.

We apply two stage Least squares method as under.

First stage (i) We regress as usual model (1) and find $\widehat{Y_i} = \widehat{\beta_1} + \widehat{\beta_2} X_i$

For sufficiently large samples, $\widehat{Y_i}$ and $E(Y_i)$ can be considered to be near. We go to the second stage

Second Stage $\frac{Y_i}{\widehat{Y_i}} = \beta_1 \left(\frac{1}{\widehat{Y_i}}\right) + \beta_2 \left(\frac{Y_i}{\widehat{Y_i}}\right) + \vartheta_i$................................ (4)

Run the regression in (4) and find $\widehat{\beta_1}$ and $\widehat{\beta_2}$ which are now two stage least squares estimates of β_1 and β_2. Here also there is no intercept term.

Assumption (4) Logarithmic transformation of the variables

Due to logarithmic values, original observations are reduced which controls heteroscadasticity.

$LnY_i = \beta_1 + LnX_i + U_i$

Where Ln denotes log values at natural base.

Run this regression instead of original and find estimators. Here advantage is that $\widehat{\beta_2}$ give estimate of elasticity and serves a basis of measuring sensitivity. Due to this, the method is popular in empirical econometrics.

(The method does not work if some observation is zero or negative)