Chapter 10
Heteroskedasticity

In the multiple regression model

$$y = X\beta + \varepsilon,$$

it is assumed that

$$V(\varepsilon) = \sigma^2 I,$$

i.e.,

$$Var(\varepsilon_i^2) = \sigma^2,$$
$$Cov(\varepsilon_i \varepsilon_j) = 0,\ i \neq j = 1, 2, ..., n.$$

In this case, the diagonal elements of the covariance matrix of ε are the same indicating that the variance of each ε_i is same and off-diagonal elements of the covariance matrix of ε are zero indicating that all disturbances are pairwise uncorrelated. This property of constancy of variance is termed as **homoskedasticity** and disturbances are called as **homoskedastic disturbances.**

In many situations, this assumption may not be plausible, and the variances may not remain the same. The disturbances whose variances are not constant across the observations are called **heteroskedastic disturbance,** and this property is termed as **heteroskedasticity.** In this case

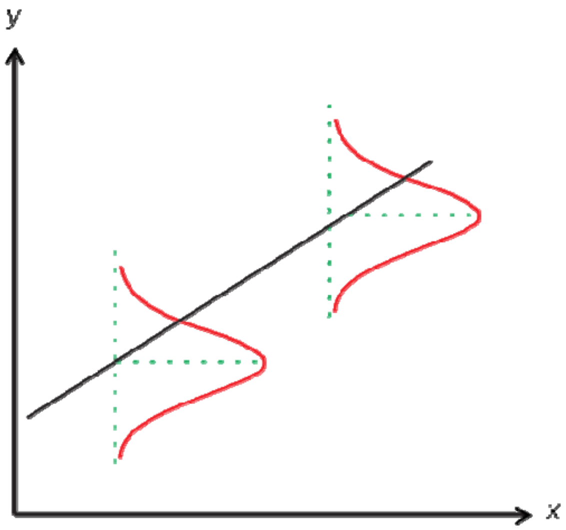
$$Var(\varepsilon_i) = \sigma_i^2,\ i = 1, 2, ..., n$$

and disturbances are pairwise uncorrelated.

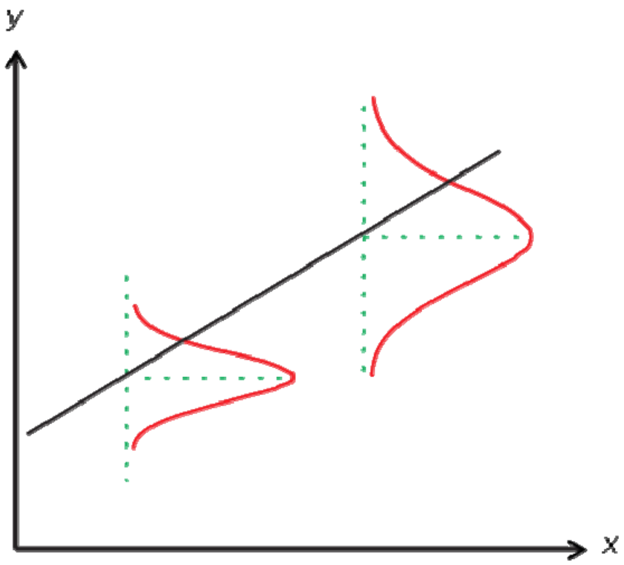
The covariance matrix of disturbances is

$$V(\varepsilon) = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}.$$

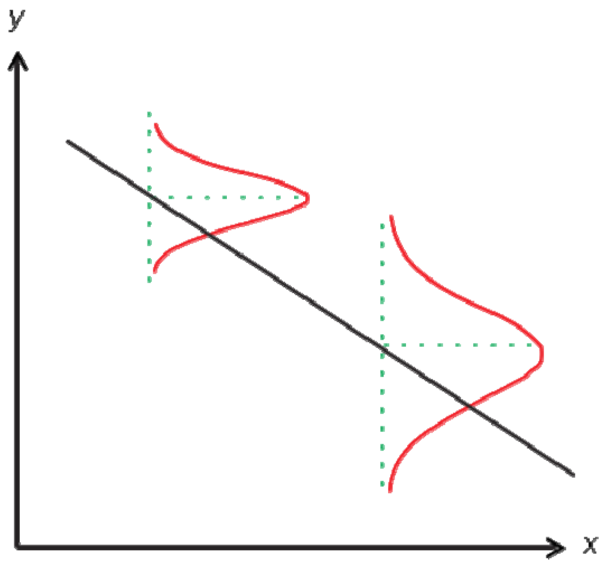
Graphically, the following pictures depict homoskedasticity and heteroskedasticity.



Homoskedasticity



Heteroskedasticity (*Var*(*y*) increases with *x*)



Heteroskedasticity (*Var*(*y*) decreases with *x*)

Examples: Suppose in a simple linear regression model, x denote the income and y denotes the expenditure on food. It is observed that as the income increases, the expenditure on food increases because of the choice and varieties in food increase, in general, up to a certain extent. So the variance of observations on y will not remain constant as income changes. The assumption of homoscedasticity implies that the consumption pattern of food will remain the same irrespective of the income of the person. This may not generally be a correct assumption in real situations. Instead, the consumption pattern changes and hence the variance of y and so the variances of disturbances will not remain constant. In general, it and will be increasing as income increases.

In another example, suppose in a simple linear regression model, x denotes the number of hours of practice for typing and y denotes the number of typing errors per page. It is expected that the number of typing mistakes per page decreases as the person practices more. The homoskedastic disturbances assumption implies that the number of errors per page will remain the same irrespective of the number of hours of typing practice which may not be true in practice.

Possible reasons for heteroskedasticity:

There are various reasons due to which the heteroskedasticity is introduced in the data. Some of them are as follows:

1. The nature of the phenomenon under study may have an increasing or decreasing trend. For example, the variation in consumption pattern on food increases as income increases. Similarly, the number of typing mistakes decreases as the number of hours of typing practise increases.

2. Sometimes the observations are in the form of averages, and this introduces the heteroskedasticity in the model. For example, it is easier to collect data on the expenditure on clothes for the whole family rather than on a particular family member. Suppose in a simple linear regression model

   $$y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}, i = 1, 2, ..., n, j = 1, 2, ..., m_i$$

   y_{ij} denotes the expenditure on cloth for the j^{th} family having m_j members and x_{ij} denotes the age of the i^{th} person in the j^{th} family. It is difficult to record data for an individual family member, but it is easier to get data for the whole family. So y_{ij}'s are known collectively.

   Then instead of per member expenditure, we find the data on average spending for each family member as

   $$\bar{y}_i = \frac{1}{m_j} \sum_{j=1}^{m_j} y_{ij}$$

   and the model becomes

   $$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + \bar{\varepsilon}_i.$$

   If we assume $E(\varepsilon_{ij}) = 0, Var(\varepsilon_{ij}) = \sigma^2,$ then

   $$E(\bar{\varepsilon}_i) = 0$$

   $$Var(\bar{\varepsilon}_i) = \frac{\sigma^2}{m_j}$$

which indicates that the resultant variance of disturbances does not remain constant but depends on the number of members in a family m_j. So heteroskedasticity enters in the data. The variance will remain constant only when all $m_j's$ are same.

3. Sometimes the theoretical considerations introduce the heteroskedasticity in the data. For example, suppose in the simple linear model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ i = 1,2,...,n,$$

y_i denotes the yield of rice and x_i denotes the quantity of fertilizer in an agricultural experiment. It is observed that when the quantity of fertilizer increases, then yield increases. In fact, initially, the yield increases when the quantity of fertilizer increases. Gradually, the rate of increase slows down, and if fertilizer is increased further, the crop burns. So notice that β_1 changes with different levels of fertilizer. In such cases, when β_1 changes, a possible way is to express it as a random variable with constant mean $\bar{\beta}_1$ and constant variance θ^2 like

$$\beta_{1i} = \bar{\beta}_1 + v_i,\ i = 1,2,...,n$$

with

$$E(v_i) = 0, Var(v_i) = \theta^2, E(\varepsilon_i v_i) = 0.$$

So the complete model becomes

$$\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \varepsilon_i \\
\beta_i &= \bar{\beta}_1 + v_i \\
\Rightarrow y_i &= \beta_0 + \bar{\beta} x_i + (\varepsilon_i + x_i v_i) \\
&= \beta_0 + \bar{\beta} x_i + w_i
\end{aligned}$$

where $w_i = \varepsilon_i + x_i v_i$ is like a new random error component. So

$$\begin{aligned}
E(w_i) &= 0 \\
Var(w_i) &= E(w_i^2) \\
&= E(\varepsilon_i^2) + x_i^2 E(v_i^2) + 2x_i E(\varepsilon_i v_i) \\
&= \sigma^2 + x_i^2\theta^2 + 0 \\
&= \sigma^2 + x_i^2\theta^2.
\end{aligned}$$

So variance depends on i, and thus heteroskedasticity is introduced in the model. Note that we assume homoskedastic disturbances for the model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ \beta_{1i} = \bar{\beta}_1 + v_i$$

but finally ends up with heteroskedastic disturbances. This is due to theoretical considerations.

4. The skewness in the distribution of one or more explanatory variables in the model also causes heteroskedasticity in the model.
5. The incorrect data transformations and wrong functional form of the model can also give rise to the heteroskedasticity problem.

Tests for heteroskedasticity

The presence of heteroskedasticity affects the estimation and test of hypothesis. The heteroskedasticity can enter into the data due to various reasons. The tests for heteroskedasticity assume a specific nature of heteroskedasticity. Various tests are available in the literature, e.g.,

1. Bartlett test
2. Breusch Pagan test
3. Goldfeld Quandt test
4. Glesjer test
5. Test based on Spearman's rank correlation coefficient
6. White test
7. Ramsey test
8. Harvey Phillips test
9. Szroeter test
10. Peak test (nonparametric) test

We discuss the first five tests.

1. Bartlett's test

It is a test for testing the null hypothesis

$$H_0: \sigma_1^2 = \sigma_2^2 = ... = \sigma_i^2 = ... = \sigma_n^2$$

This hypothesis is termed as **the hypothesis of homoskedasticity**. This test can be used only when replicated data is available.

Since in the model

$$y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i,\ E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma_i^2,\ i = 1,2,...,n,$$

only one observation y_i is available to find σ_i^2, so the usual tests can not be applied. This problem can be overcome if replicated data is available. So consider the model of the form

$$y_i^* = X_i\beta + \varepsilon_i^*$$

where y_i^* is a $m_i \times 1$ vector, X_i is $m_i \times k$ matrix, β is $k \times 1$ vector and ε_i^* is $m_i \times 1$ vector. So replicated data is now available for every y_i^* in the following way:

$y_1^* = X_1\beta + \varepsilon_1^*$ consists of m_1 observations

$y_2^* = X_2\beta + \varepsilon_2^*$ consists of m_2 observations

$\vdots$

$y_n^* = X_n\beta + \varepsilon_n^*$ consists of m_n observations.

All the individual model can be written as

$$\begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1^* \\ \varepsilon_2^* \\ \vdots \\ \varepsilon_n^* \end{pmatrix}$$

or $y^* = X\beta + \varepsilon^*$

where y^* is a vector of order $\left(\sum_{i=1}^n m_i\right) \times 1$, X is $\left(\sum_{i=1}^n m_i\right) \times k$ matrix, β is $k \times 1$ vector and ε^* is $\left(\sum_{i=1}^n m_i\right) \times 1$ vector. Apply OLS to this model yields

$$\hat{\beta} = (X'X)^{-1}X'y^*$$

and obtain the residual vector

$$e_i^* = y_i^* - X_i\hat{\beta}.$$

Based on this, obtain

$$s_i^2 = \frac{1}{m_i - k} e_i^{*\prime} e_i^*$$

$$s^2 = \frac{\sum_{i=1}^n (m_i - k)s_i^2}{\sum_{i=1}^n (m_i - k)}.$$

Now apply Bartlett's test as

$$\chi^2 = \frac{1}{C}\sum_{i=1}^n (m_i - k)\log\frac{s^2}{s_i^2}$$

which has asymptotic $\chi^2 -$distribution with $(n-1)$ degrees of freedom where

$$C = 1 + \frac{1}{3(n-1)}\left[\sum_{i=1}^n \left(\frac{1}{m_i - k}\right) - \frac{1}{\sum_{i=1}^n (m_i - k)}\right].$$

Another variant of Bartlett's test

Another variant of Bartlett's test is based on the likelihood ratio test statistic

$$u = \sum_{i=1}^{m} \left(\frac{s_i^2}{s^2} \right)^{n_i/2}$$

where

$$s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y}_i \right)^2, \; i = 1, 2, ..., m; \; j = 1, 2, ..., n_i$$

$$s^2 = \frac{1}{n} \sum_{i=1}^{m} n_i s_i^2$$

$$n = \sum_{i=1}^{m} n_i \; .$$

To obtain an unbiased test and modification of $-2\ln u$ which is a closer approximation to χ^2_{m-1} under H_0, Bartlett test replaces n_i by $(n_i - 1)$ and divide by a scalar constant. This leads to the statistic

$$M = \frac{(n-m)\log \hat{\sigma}^2 - \sum_{i=1}^{m} (n_i - 1) \log \hat{\sigma}_i^2}{1 + \frac{1}{3(m-1)} \left[\sum_{i=1}^{m} \left(\frac{1}{n_i - 1} \right) - \frac{1}{n-m} \right]}$$

which has a χ^2 distribution with $(m-1)$ degrees of freedom under H_0 and

$$\hat{\sigma}_i^2 = \frac{1}{n-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

$$\hat{\sigma}^2 = \frac{1}{n-m} \sum_{i=1}^{m} (n_i - 1) \hat{\sigma}_i^2 .$$

In experimental sciences, it is easier to get replicated data, and this test can be easily applied. In real-life applications, it is challenging to get replicated data, and this test may not be applied. This difficulty is overcome in Breusch Pagan test.

2. Breusch Pagan test

This test can be applied when the replicated data is not available, but only single observations are available. When it is suspected that the variance is some function (but not necessarily multiplicative) of more than one explanatory variable, then Breusch Pagan test can be used.

Assuming that under the alternative hypothesis σ_i^2 is expressible as

$$\sigma_i^2 = h(Z_i'\hat{\gamma}) = h(\gamma_1 + Z_i^* \gamma^*)$$

where h is some unspecified function and is independent of i,

$$Z_i' = (1, Z_i^{*'})(1, Z_{i2}, Z_{i3}, ..., Z_{ip})$$

is the vector of observable explanatory variables with first element unity and $\gamma' = (\gamma_1, \gamma_i^*) = (\gamma_1, \gamma_2, ..., \gamma_p)$ is a vector of unknown coefficients related to β with the first element being the intercept term. The heterogeneity is defined by these p variables. These Z_i's may also include some X's also.

Specifically, assume that

$$\sigma_i^2 = \gamma_1 + \gamma_2 Z_{i2} + ... + \gamma_p Z_{ip}.$$

The null hypothesis

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2$$

can be expressed as

$$H_0 : \gamma_2 = \gamma_3 = ... = \gamma_p = 0 .$$

If H_0 is accepted , it implies that $\gamma_2 Z_{i2}, \gamma_3 Z_{i3}, ..., \gamma_p Z_{ip}$ do not have any effect on σ_i^2 and we get $\sigma_i^2 = \gamma_1$.

The test procedure is as follows:

1. Ignoring heterogeneity, apply OLS to

$$y_i = \beta_1 + \beta_2 X_{i1} + ... + \beta_k X_{ik} + \varepsilon_i$$

and obtain residual

$$e = y - Xb$$
$$b = (X'X)^{-1} X'Y.$$

2. Construct the variables

$$g_i = \frac{e_i^2}{\left(\sum_{i=1}^n e_i^2 \middle/ n \right)} = \frac{ne_i^2}{SS_{res}}$$

where SS_{res} is the residual sum of squares based on e_i's.

3. Run regression of g on $Z_1, Z_2, ..., Z_p$ and get residual sum of squares SS^*_{res}.

4. For testing, calculate the test statistic

$$Q = \frac{1}{2}\left(\sum_{i=1}^{n} g_i^2 - SS^*_{res}\right)$$

which is asymptotically distributed as χ^2 distribution with $(p-1)$ degrees of freedom.

5. The decision rule is to reject H_0 if $Q > \chi^2_{1-\alpha}(m-1)$.

- This test is very simple to perform.
- A fairly general form is assumed for heterogeneity, so it is a very general test.
- This is an asymptotic test.
- This test is quite powerful in the presence of heteroskedasticity.

3. Goldfeld Quandt test

This test is based on the assumption that σ_i^2 is positively related to X_{ij}, i.e., one of the explanatory variables explains the heteroskedasticity in the model. Let j^{th} explanatory variable explains the heteroskedasticity, so

$$\sigma_i^2 \propto X_{ij}$$

$$\text{or } \sigma_i^2 = \sigma^2 X_{ij}.$$

The test procedure is as follows:

1. Rank the observations according to the decreasing order of X_j.
2. Split the observations into two equal parts leaving c observations in the middle.

   So each part contains $\frac{n-c}{2}$ observations provided $\frac{n-c}{2} > k.$
3. Run two separate regression in the two parts using OLS and obtain the residual sum of squares SS_{res1} and SS_{res2}.
4. The test statistic is

$$F_0 = \frac{SS_{res2}}{SS_{res1}}$$

   which follows $F-$distribution, i.e., $F\left(\frac{n-c}{2}-k, \frac{n-c}{2}-k\right)$ when H_0 true.

5. The decision rule is to reject H_0 whenever $F_0 > F_{1-\alpha}\left(\frac{n-c}{2}-k, \frac{n-c}{2}-k\right)$.

- This test is a simple test, but it is based on the assumption that one of the explanatory variables helps in determining the heteroskedasticity.
- Then the test is an exact finite sample test.
- The only difficulty in this test is that the choice of c is not obvious. If a large value of c is chosen, then it reduces the degrees of freedom $\frac{n-c}{2}-k$, and the condition $\frac{n-c}{2}>k$ may be violated.

On the other hand, if a smaller value of c is chosen, then the test may fail to reveal the heteroskedasticity. The basic objective of the ordering of observations and deletion of observations in the middle part may not reveal the heteroskedasticity effect. Since the first and last values of σ_i^2 gives the maximum discretion, so removal of smaller value may not give the proper idea of heteroskedasticity. Considering these two points, the working choice of c is suggested as $c=\frac{n}{3}$.

Moreover, the choice of X_{ij} is also difficult. Since $\sigma_i^2 \propto X_{ij}$, so if all important variables are included in the model, then it may be difficult to decide that which of the variable is influencing the heteroskedasticity.

4. Glesjer test:

This test is based on the assumption that σ_i^2 is influenced by one variable Z, i.e., there is only one variable which is influencing the heteroskedasticity. This variable could be either one of the explanatory variable or it can be chosen from some extraneous sources also.

The test procedure is as follows:

1. Use OLS and obtain the residual vector e on the basis of available study and explanatory variables.
2. Choose Z and apply OLS to

   $$|e_i| = \delta_0 + \delta_1 Z_i^h + v_i$$

   where v_i is the associated disturbance term.
3. Test $H_0: \delta_1 = 0$ using t-ratio test statistic.
4. Conduct the test for $h = \pm 1, \pm \frac{1}{2}$. So the test procedure is repeated four times.

In practice, one can choose any value of h. For simplicity, we choose $h=1$.

- The test has only asymptotic justification and the four choices of h give generally satisfactory results.
- This test sheds light on the nature of heteroskedasticity.

5. Spearman's rank correlation test

It d_i denotes the difference in the ranks assigned to two different characteristics of the i^{th} object or phenomenon and n is the number of objects or phenomenon ranked, then the Spearman's rank correlation coefficient is defined as

$$r = 1 - 6\left(\frac{\sum_{i=1}^{n} d_i^2}{n(n^2-1)}\right); \quad -1 \le r \le 1.$$

This can be used for testing the hypothesis about the heteroskedasticity.

Consider the model

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i .$$

1. Run the regression of y on X and obtain the residuals e.
2. Consider $|e_i|$.
3. Rank both $|e_i|$ and X_i (or $\hat{y}_i$) in an ascending (or descending) order.
4. Compute rank correlation coefficient r based on $|e_i|$ and X_i (or $\hat{y}_i$).
5. Assuming that the population rank correlation coefficient is zero and $n > 8$, use the test statistic

   $$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

   which follows a t-distribution with $(n-2)$ degrees of freedom.
6. The decision rule is to reject the null hypothesis of heteroskedasticity whenever $t_0 \ge t_{1-\alpha}(n-2)$.

   If there are more than one explanatory variables, then rank correlation coefficient can be computed between $|e_i|$ and each of the explanatory variables separately and can be tested using t_0.

Estimation under heteroskedasticity

Consider the model

$$y = X\beta + \varepsilon$$

with k explanatory variables and assume that

$$E(\varepsilon) = 0$$

$$V(\varepsilon) = \Omega = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}.$$

The OLSE is

$$b = (X'X)^{-1}X'y.$$

Its estimation error is

$$b - \beta = (X'X)^{-1}X'\varepsilon$$

and

$$E(b-\beta) = (X'X)^{-1}X'E(\varepsilon) = 0.$$

Thus OLSE remains unbiased even under heteroskedasticity.

The covariance matrix of b is

$$\begin{aligned} V(b) &= E(b-\beta)(b-\beta)' \\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} \\ &= (X'X)^{-1}X'\Omega X(X'X)^{-1} \end{aligned}$$

which is not the same as conventional expression. So OLSE is not efficient under heteroskedasticity as compared under homoskedasticity.

Now we check if $E(e_i^2) = \sigma_i^2$ or not where e_i is the i^{th} residual.

The residual vector is

$$e = y - Xb = \bar{H}\varepsilon$$

$$\begin{aligned} e_i &= [0,0,...,0,1,0,...0]\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \\ &= \ell_i' e = \ell_i' \bar{H}\varepsilon \end{aligned}$$

where ℓ_i is a $n \times 1$ vector with all elements zero except the i^{th} element which is unity and $\bar{H} = I - X(X'X)^{-1}X'$. Then

$$\begin{aligned} e_i^2 &= \ell_i' e.e'\ell_i = \ell_i'\bar{H}\varepsilon\varepsilon'\bar{H}\ell_i \\ E(e_i^2) &= \ell_i'\bar{H}E(\varepsilon\varepsilon')\bar{H}\ell_i = \ell_i'\bar{H}\Omega\bar{H}\ell_i \end{aligned}$$

$$\bar{H}\ell_i = \begin{bmatrix} \bar{h}_{11} & \cdots & \bar{h}_{1n} \\ \vdots & \ddots & \vdots \\ \bar{h}_{n1} & \cdots & \bar{h}_{nn} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \bar{h}_{1i}\, i \\ \bar{h}_{2i} i \\ \vdots \\ \bar{h}_{ni}\, i \end{bmatrix}$$

$$E(e_i^2) = \left[\bar{h}_{1i} i,\ \bar{h}_{2i} i, \ldots,\ \bar{h}_{ni}\, i\right] \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} \begin{bmatrix} \bar{h}_{1i}\, i \\ \bar{h}_{2i}\, i \\ \vdots \\ \bar{h}_{n1}\, i \end{bmatrix}.$$

Thus $E(e_i^2) \neq \sigma_i^2$ and so e_i^2 becomes a biased estimator of σ_i^2 in the presence of heteroskedasticity.

In the presence of heteroskedasticity, use the generalized least squares estimation. The generalized least squares estimator (GLSE) of β is

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} y.$$

Its estimation error is obtained as

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}(X\beta + \varepsilon)$$
$$\hat{\beta} - \beta = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}\varepsilon).$$

Thus

$$E(\hat{\beta} - \beta) = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} E(\varepsilon) = 0$$
$$V(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)$$
$$= (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} E(\varepsilon\varepsilon)\Omega^{-1} X (X'\Omega^{-1}X)^{-1}$$
$$= (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}\Omega\Omega^{-1} X (X'\Omega^{-1}X)^{-1}$$
$$= (X'\Omega^{-1}X)^{-1}.$$

Example: Consider a simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ i = 1, 2, ..., n.$$

The variances of OLSE and GLSE of β are

$$Var(b) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sigma_i^2}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right)^2} \text{ and } Var(\hat{\beta}) = \sum_{i=1}^{n} \frac{\sigma_i^2}{(x_i - \bar{x})^2} \text{ respectively.}$$

Consider

$$\frac{Var(\hat{\beta})}{Var(b)} = \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma_i^2\right]\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 \frac{1}{\sigma_i^2}\right]}} \right]^2$$

= Square of the correlation coefficient betweene $\sigma_i(x_i - \bar{x})$ and $\left(\frac{x_i - \bar{x}}{\sigma_i}\right)$

≤ 1

$\Rightarrow Var(\hat{\beta}) \leq Var(b).$

So efficient of OLSE and GLSE depends upon the correlation coefficient between $(x_i - \bar{x})\sigma_i$ and $\frac{(x_i - \bar{x})}{\sigma_i}$.

The generalized least squares estimation assumes that Ω is known, i.e., the nature of heteroskedasticity is completely specified. Based on this assumption, the possibilities of following two cases arise:

- Ω is completely specified or
- Ω is not completely specified.
-

We consider both the cases as follows:

Case 1: σ_i^2's are prespecified:

Suppose $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$ are completely known in the model

$$y_i = \beta_1 + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i .$$

Now deflate the model by σ_i, i.e.,

$$\frac{y_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{X_{i2}}{\sigma_i} + ... + \beta_k \frac{X_{ik}}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}.$$

Let $\varepsilon_i^* = \frac{\varepsilon_i}{\sigma_i}$, then $E(\varepsilon_i^*) = 0, Var(\varepsilon_i^*) = \frac{\sigma_i^2}{\sigma_i^2} = 1$. Now OLS can be applied to this model and usual tools for drawing statistical inferences can be used.

Note that when the model is deflated, the intercept term is lost as β_1 / σ_i is itself a variable. This point has to be taken care of in software output.

Case 2: Ω may not be completely specified

Let $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$ are partially known and suppose

$$\sigma_i^2 \propto X_{ij}^{2\lambda}$$

or $\sigma_i^2 = \sigma^2 X_{ij}^{2\lambda}$

but σ^2 is not available. Consider the model

$$y_i = \beta_1 + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i$$

and deflate it by X_{ij}^{λ} as

$$\frac{y_i}{X_{ij}^{\lambda}} = \frac{\beta_1}{X_{ij}^{\lambda}} + \beta_2 \frac{X_{i2}}{X_{ij}^{\lambda}} + ... + \beta_k \frac{X_{ik}}{X_{ij}^{\lambda}} + \frac{\varepsilon_i}{X_{ij}^{\lambda}}.$$

Now apply OLS to this transformed model and use the usual statistical tools for drawing inferences.

A caution is to be kept is mind while doing so. This is illustrated in the following example with one explanatory variable model.

Consider the model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$

Deflate it by x_i, so we get

$$\frac{y_i}{x_i} = \frac{\beta_0}{x_i} + \beta_1 + \frac{\varepsilon_i}{x_i}.$$

Note that the roles of β_0 and β_1 in original and deflated models are interchanged. In the original model, β_0 is the intercept term and β_1 is the slope parameter whereas in the deflated model, β_1 becomes the intercept term and β_0 becomes the slope parameter. So essentially, one can use OLS but need to be careful in identifying the intercept term and slope parameter, particularly in the software output.