

Chapter 9

Heteroskedasticity

In the multiple regression model

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

it is assumed that

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

i.e.,

$$Var(\varepsilon_i) = \sigma^2,$$

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

In this case, the diagonal elements of the covariance matrix of ε are the same, indicating that the variance of each ε_i is the same, and the 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 homoskedastic disturbances.

In many situations, this assumption may not hold, and the variances may change. Disturbances whose variances are not constant across observations are called heteroskedastic, and this property is termed heteroskedasticity. In this case

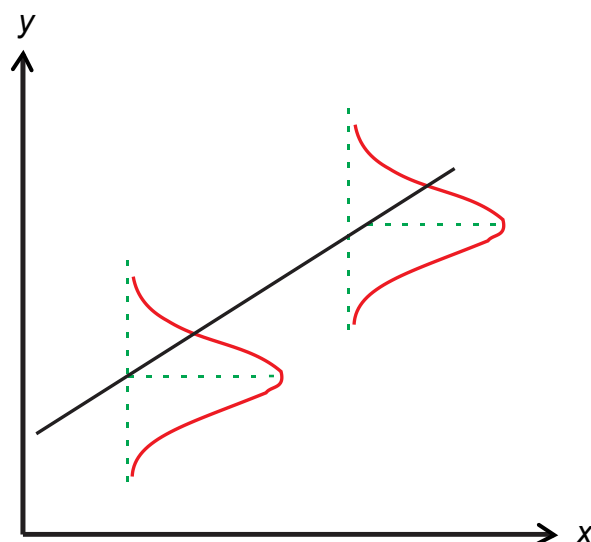
$$Var(\varepsilon_i) = \sigma_i^2, i = 1, 2, \dots, n$$

and disturbances are pairwise uncorrelated.

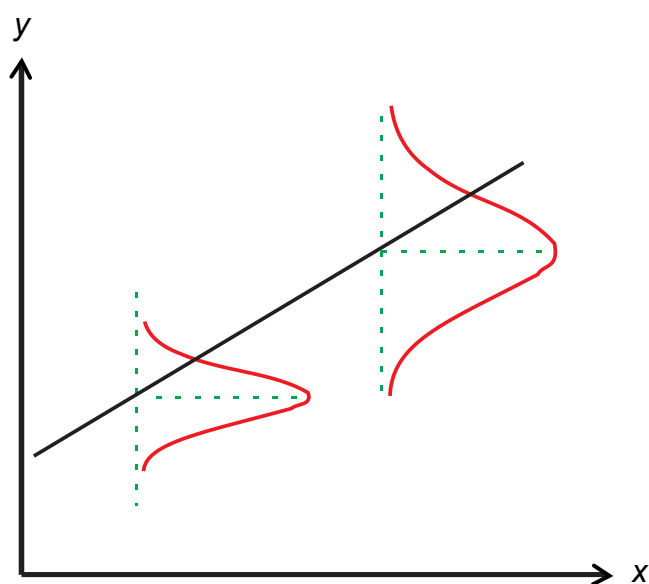
The covariance matrix of disturbances is

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

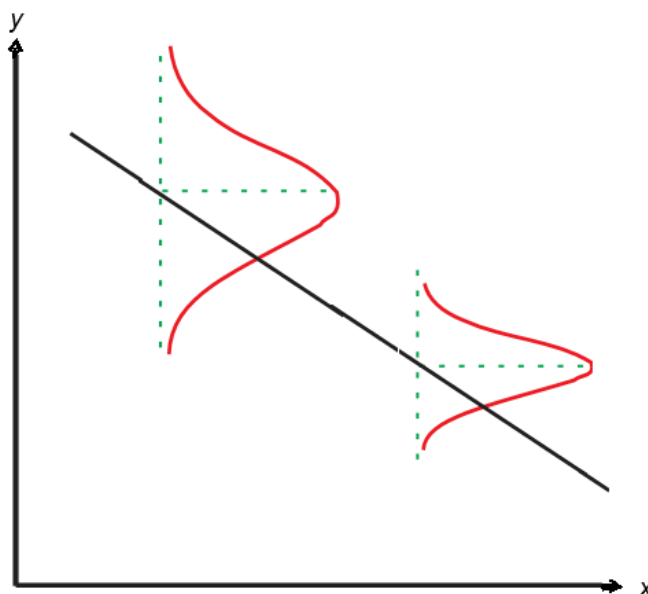
Graphically, the following pictures depict homoskedasticity and heteroskedasticity.



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 the choice and variety of 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 person's income. 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 increases as income rises.

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 homoskedasticity assumption implies that the number of errors per page will remain constant regardless of the number of hours of typing practice, which may not hold in practice.

Possible reasons for heteroskedasticity

There are various reasons why heteroskedasticity may be introduced into 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 food consumption patterns increases as income rises. Similarly, the number of typing mistakes decreases as the number of hours of typing practice increases.
2. Sometimes the observations are in the form of averages, and this introduces heteroskedasticity in the model. For example, it is easier to collect data on family expenditure on clothes than on a particular family member. Suppose in a simple linear regression model

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

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 the data. The variance will remain constant only when all m_j 's are same.

3. Sometimes the theoretical considerations introduce heteroskedasticity in the data.

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 testing of hypotheses. Heteroskedasticity can enter the data for 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 Bartlett's test.

Bartlett's test

It is a test for testing the null hypothesis

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

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

Since in the model

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

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 models 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. Applying 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

$$s_i^2 = \frac{1}{m_i - k} e_i^{*'} e_i^* \quad \text{and} \quad s^2 = \frac{\sum_{i=1}^n (m_i - k) s_i^2}{\sum_{i=1}^n (m_i - k)}.$$

Now apply Bartlett's test as

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

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

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

In experimental sciences, it is easier to obtain replicated data, and this test can be applied readily. In real-world applications, obtaining replicated data is challenging, so this test may not be applicable. This difficulty is overcome in the Breusch-Pagan test.

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 in the presence of 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 a conventional expression. So OLSE is less efficient under heteroskedasticity than under homoskedasticity.

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

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

Thus

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

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 the following two cases arise:

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

We consider both cases as follows:

Case 1: σ_i^2 's are prespecified:

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

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

Now, deflate the model by σ_i , i.e.,

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

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 the 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 addressed in the software output.

Case 2: Ω may not be completely specified

Let $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ are partially known and are supposed

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

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

but σ^2 is not available.

Consider the model

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

and deflate it by X_{ij}^λ as

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

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

One caution to keep in mind while doing so. This is illustrated in the following example with a one-explanatory-variable model.

Consider the model

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

Deflating 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 the 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 needs to be careful in identifying the intercept term and slope parameter, particularly in the software output.