Chap.7 Nonspherical Disturbances (Heterocedasticity)

（5^{th} , Greene ， Chap.10 & Chap.11）

7.1 Generalized Linear Model

7.1.1 Introduction

Assumption 3 of Classical Model states that the $n \times n$ matrix of conditional second moment $E\left(\varepsilon\varepsilon' \middle| X\right) = \sigma^2 I_n$ is spherical. Without this assumption, we extend the multiple regression model to the generalized linear regression model.

$$Y = X\beta + \varepsilon, \quad E\left(\varepsilon \middle| X\right) = 0, \quad E\left(\varepsilon\varepsilon' \middle| X\right) = \sigma^2\Omega \tag{7.1}$$

Where Ω is a positive definite matrix and $\Omega \neq I$, the disturbances are nonspherical disturbances.Two cases will consider for GR model:

$$E\left(\varepsilon\varepsilon' \middle| X\right) = \sigma^2\Omega$$

$$= \begin{cases} E\left(\varepsilon_i\varepsilon_j \middle| X\right) \neq 0, \quad i \neq j \qquad autocorrelation \\ E\left(\varepsilon_i^2 \middle| X\right) = \sigma_i^2 \quad i = j \qquad heteroscedasticity \end{cases}$$

7.1.2 Consequence of Relaxing Assumption 3.

1) OLS estimator $\hat{\beta}$ is unbiased and consistent.

(i) linearity $\hat{\beta} = \left(X'X\right)^{-1} X'Y$

(ii) unbiased $E\left(\hat{\beta} \middle| X\right) = \beta + \left(X'X\right)^{-1} X'E\left(\varepsilon \middle| X\right) = \beta$

(iii) $\operatorname{var}\left(\hat{\beta} \middle| X\right) = E\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)' \middle| X\right]$

$$= E\left[\left(X'X\right)^{-1} X'\varepsilon\varepsilon'X\left(X'X\right)^{-1} \middle| X\right]$$

$$= \left(X'X\right)^{-1} X'E\left[\varepsilon\varepsilon' \middle| X\right] X\left(X'X\right)^{-1} = \sigma^2\left(X'X\right)^{-1} X'\Omega X\left(X'X\right)^{-1} \tag{7.2}$$

Conventional OLS coefficient standard errors are incorrect. The correct variance matrix for the OLS coefficient is $\sigma^2\left(X'X\right)^{-1} X'\Omega X\left(X'X\right)^{-1}$. The variance matrix may also be expressed as

$$\operatorname{var}\left(\hat{\beta}\middle|X\right)=\frac{\sigma^2}{n}\left(\frac{X'X}{n}\right)^{-1}\frac{X'\Omega X}{n}\left(\frac{X'X}{n}\right)^{-1} \tag{7.3}$$

Consistency requires $p\lim\frac{X'X}{n}$ and $p\lim\frac{X'\Omega X}{n}$ are both finite positive definite matrices, which in general will be true if the regressors are well behaved and the elements of Ω are finite. Mean square consistency follows since $\operatorname{var}\left(\hat{\beta}\middle|X\right)$ has a zero probability limit.

2) The OLS estimator $\hat{\sigma}^2$ is biased

$$\hat{\sigma}^2=\frac{e'e}{n-k}$$

$$E\left(\frac{e'e}{n-k}\middle|X\right)=E\left(\frac{\varepsilon'M\varepsilon}{n-k}\middle|X\right)$$

$$=\frac{1}{n-k}Tr\left[\left(E\left(\varepsilon'\varepsilon\middle|X\right)-E\left(\left(X(X'X)^{-1}X'\varepsilon'\varepsilon\middle|\right)X\right)\right)\right]$$

$$=\frac{\sigma^2}{n-k}E\left(Tr\left(\Omega\right)-Tr\left((X'X)^{-1}X'\Omega X\right)\right)\neq\sigma^2$$

3) OLS estimator $\hat{\beta}$ is inefficient

The Gauss-Markov Theorem no longer holds for the OLS estimator $\hat{\beta}$, the BLUE is some other estimator. Thus the t-ratio is not distributed as the t distribution, the t-test is no longer valid. The same comments apply to the F-test.

4) Asymptotic Distribution

In finite sample , $\varepsilon\sim N\left(0,\sigma^2\Omega\right)$, $\hat{\beta}\middle|X\sim N(\beta,\sigma^2\left(X'X\right)^{-1}X'\Omega X\left(X'X\right)^{-1})$

For large sample ,the asymptotic distribution of OLS estimator in the GR model will discuss in the heteroscedastic case and in the autocorrelated case respectively. under specified conditions, both cases have Asymptotic Normality.

7.2 Efficient Estimation by Generalized Least Squares

7.2.1 Ω is known

1) GLS Estimator

Since Ω is positive definite, its inverse is positive definite. Thus it is possible to find a nonsingular matrix P such that

$$\Omega^{-1} = P'P \tag{7.4}$$

Pre-multiply the linear model $Y = X\beta + \varepsilon$ by a nonsingular matrix P, satisfying (7.4), to obtain

$$Y^* = X^*\beta + \varepsilon^* \tag{7.5}$$

Where $Y^* = PY$, $X^* = PX$, and $\varepsilon^* = P\varepsilon$. It follows from (7.4) that $\Omega = P^{-1}\left(P^{-1}\right)'$. Then

$$\operatorname{var}\left(\varepsilon^* \middle| X\right) = E\left(P\varepsilon\varepsilon'P' \middle| X\right) = PE\left(\varepsilon\varepsilon' \middle| X\right)P' = \sigma^2 P\Omega P' = \sigma^2 PP^{-1}\left(P^{-1}\right)' P' = \sigma^2 I$$

Thus the transformed variables in (7.5) satisfy the conditions under which OLS is BLUE. The coefficient vector from the OLS regression of Y^* on X^* is the Generalized Least Squares (GLS) estimator.

$$\hat{\beta}_{GLS} = \left(X^{*\prime}X^*\right)^{-1} X^{*\prime}Y^* =$$

$$= \left(X'\left(P'P\right)X\right)^{-1} X'\left(P'P\right)Y$$

$$= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y \tag{7.6}$$

$$..\operatorname{var}\left(\hat{\beta}_{GLS} \middle| X^*\right) = \sigma^2\left(X^{*\prime}X^*\right)^{-1} = \sigma^2\left(X'\Omega^{-1}X\right)^{-1} \tag{7.7}$$

GLS estimator is BLUE, it is more efficient than OLS estimator in the GR model:

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

$$= \sigma^2\left[\left(X'X\right)^{-1} X' - \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}\right]\Omega\left[\left(X'X\right)^{-1} X' - \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}\right]'$$

$$= A\Omega A'$$

Where $A = \left[\left(X'X\right)^{-1} X' - \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}\right]$. Since Ω is positive definite, we have

$$A\Omega A' \ge 0$$

i.e. $\operatorname{var}\left(\hat{\beta}_{GLS} \middle| X\right) - \operatorname{var}\left(\hat{\beta}_{OLS} \middle| X\right) \le 0$

2) MLE

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma^2\Omega\right)$$

The multivariate normal density for ε is

$$f(\varepsilon) = (2\pi)^{-\frac{n}{2}}\left|\sigma^2\Omega\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\varepsilon'\left(\sigma^2\Omega\right)^{-1}\varepsilon\right\}$$

Noting that $\left|\sigma^2\Omega\right| = \sigma^{2n}\left|\Omega\right|$, we may rewrite the density as

$$f(\varepsilon)=(2\pi)^{-\frac{n}{2}}\left(\sigma^2\right)^{-\frac{n}{2}}|\Omega|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}\varepsilon'\Omega^{-1}\varepsilon\right\}$$

The log-likelihood is

$$\ln L=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma^2-\frac{1}{2}\ln|\Omega|-\frac{1}{2\sigma^2}(Y-X\beta)'\Omega^{-1}(Y-X\beta) \qquad (7.8)$$

$$\frac{\partial \ln L}{\partial\beta}=\frac{1}{\sigma^2}\left(X'\Omega^{-1}Y-X'\Omega^{-1}X\beta\right)=0$$

$$\frac{\partial \ln L}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}(Y-X\beta)'\Omega^{-1}(Y-X\beta)=0$$

The ML estimator

$$\hat{\beta}_{ML}=\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}Y \qquad (7.9)$$

$$\hat{\sigma}^2_{ML}=\frac{1}{n}\left(Y-X\hat{\beta}_{ML}\right)'\Omega^{-1}\left(Y-X\hat{\beta}_{ML}\right) \qquad (7.10)$$

It is quite evident that $\hat{\beta}_{ML}$ is equivalent to $\hat{\beta}_{GLS}$ when Ω is known. But $\hat{\sigma}^2_{ML}$ differs from unbiased GLS estimator $\hat{\sigma}^2_{GLS}$ by the factor $\frac{n-K}{n}$, where

$$\hat{\sigma}^2_{GLS}=\frac{\left(Y^*-X^*\hat{\beta}_{GLS}\right)'\left(Y^*-X^*\hat{\beta}_{GLS}\right)}{n-K}=\frac{\left[P\left(Y-X\hat{\beta}_{GLS}\right)\right]'\left[P\left(Y-X\hat{\beta}_{GLS}\right)\right]}{n-K}$$

$$=\frac{\left(Y-X\hat{\beta}_{GLS}\right)'\Omega^{-1}\left(Y-X\hat{\beta}_{GLS}\right)}{n-K} \qquad (7.11)$$

7.2.2 Hypothesis Testing

1) linear restriction

Since (7.5) satisfies the conditions for the application of OLS, an exact,finite sample test of the linear restriction

$$H_0: R\beta=q$$

can be based on

$$F=\left(R\hat{\beta}_{GLS}-q\right)'\left[R\hat{\sigma}^2_{GLS}\left(X^{*\prime}X^*\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{GLS}-q\right)\Big/J$$

$$=\frac{\left(R\hat{\beta}_{GLS}-q\right)'\left[R\left(X'\Omega^{-1}X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{GLS}-q\right)\Big/J}{\hat{\sigma}^2_{GLS}}$$

$$=\frac{\left(\hat{\varepsilon}_c'\hat{\varepsilon}_c-\hat{\varepsilon}'\hat{\varepsilon}\right)/J}{\hat{\varepsilon}'\hat{\varepsilon}/(n-K)} \qquad (7.12)$$

Having the $F(J,n-K)$ distribution under H_0.

The constrained GLS residuals $\hat{\varepsilon}_c=Y^*-X^*\hat{\beta}_{c,GLS}$ are based on

$$\hat{\beta}_{c,GLS} = \hat{\beta}_{GLS} - \left(X'\Omega^{-1}X\right)^{-1} R'\left[R\left(X'\Omega^{-1}X\right)^{-1} R'\right]^{-1}\left(R\hat{\beta}_{GLS} - q\right)$$

It is just the constrained OLS estimator using the transformed data, and the $\hat{\sigma}^2_{GLS}$ is defined in (7.11), where the residual vector

$$\hat{\varepsilon} = Y^* - X^*\hat{\beta}_{GLS}$$

The residuals from the original model $Y - X\hat{\beta}_{GLS}$ are GLS residuals.

2) R^2 for GR Model

There is no precise counterpart to R^2 in the GR model. The R^2 -like measures in this setting are purely descriptive.

7.2.3 Estimation when Ω is unknown

1) Feasible Generalized Least Squares

If we do not know the matrix Ω, we must estimate its functional form $\Omega = \Omega(\theta)$ from the sample. This method is called the Feasible Generalized Least Squares (FGLS). Let the FGLS estimator be denoted

$$\hat{\hat{\beta}}_{FGLS} = \left(X'\hat{\Omega}^{-1}X\right)^{-1} X'\hat{\Omega}^{-1}Y \qquad (7.13)$$

Here we use $\hat{\Omega} = \Omega\left(\hat{\theta}\right)$ instead of the true Ω. It would seem that if $p\lim\hat{\theta} = \theta$, then using $\hat{\Omega}$ is asymptotically equivalent to using the true Ω. The conditions that imply that $\hat{\hat{\beta}}_{FGLS}$ is asymptotically equivalent to $\hat{\beta}_{GLS}$ are

$$p\lim\left[\left(\frac{1}{n}X'\hat{\Omega}^{-1}X\right) - \left(\frac{1}{n}X'\Omega^{-1}X\right)\right] = 0$$

$$p\lim\left[\left(\frac{1}{n}X'\hat{\Omega}^{-1}\varepsilon\right) - \left(\frac{1}{n}X'\Omega^{-1}\varepsilon\right)\right] = 0$$

These conditions, in principle, must be verified on a case-by-case basis. Fortunately, in most familiar settings, they are met.

The following theorem is extremely useful.

Theorem 10.8: Efficiency of the FGLS Estimator

An asymptotically efficient FGLS estimator does not require that we have an efficient estimator of θ; only a consistent one is required to achieve full efficiency for the FGLS estimator.

2) Maximum Likelihood Estimation

The iterative two-step method proposed by Oberhofer and Kmenka (1974) .

Step 1: For a given value of θ the estimator of β would be FGLS and the estimator of σ^2 would be the estimator as follow:

$$\hat{\sigma}^2 = \frac{1}{n}\left(Y - X\hat{\hat{\beta}}_{FGLS}\right)'\hat{\Omega}^{-1}\left(Y - X\hat{\hat{\beta}}_{FGLS}\right)$$

Step 2: For given values of β and σ^2, calculate the estimate values of θ straightforward.

Oberhofer and Kmenka showed that under some fairly weak requirements, iterating back and forth between step 1 and step 2 until convergence would produce the maximum likelihood estimator. The most important requirement is that θ not involve σ^2 or any of the parameters in β; If θ and β have no parameter in common, the information matrix for the ML estimator of β, σ^2 and θ will be block diagonal of the form:

$$-E\left[\frac{\partial^2 \ln L}{\partial\begin{pmatrix}\beta\\ \sigma^2\\ \theta\end{pmatrix}\partial\left(\beta' \quad \sigma^2 \quad \theta\right)}\right] = \begin{pmatrix} \frac{1}{\sigma^2}\left(X'\Omega^{-1}X\right) & 0 & 0 \\ 0 & \frac{n}{2\sigma^4} & \delta \\ 0 & \delta & c \end{pmatrix}$$

7.3 Reasons for Heteroscedasticity (leave out)

7.4 Tests for Heteroscedasticity

7.4.1 White's General Test

$$H_0 : \sigma_i^2 = \sigma^2 \quad \left(i = 1, \cdots, n\right)$$
$$H_1 : otherwise$$

Suppose $S_0 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 X_i X_i'$,where $e_i = y - X_i'\hat{\beta}$, and $\hat{\beta}$ is the OLS estimator. Under H_0, White verified that

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2 X_i X_i' - s^2\frac{1}{n}\sum_{i=1}^{n} X_i X_i' = \frac{1}{n}\sum_{i=1}^{n}\left(e_i^2 - s^2\right)X_i X_i' \xrightarrow{P} 0 \qquad (7.14)$$

Where $s^2 \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \xrightarrow{p} \sigma^2 \frac{1}{n}\sum_{i=1}^{n} E(X_i X_i') \xrightarrow{p} \sigma^2 Q_{XX}$

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2 X_i X_i' \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n} E(e_i^2 X_i X_i') = \frac{1}{n}\sum_{i=1}^{n} E\left[E\left(e_i^2 \middle| X\right) X_i X_i'\right]$$
$$= \sigma^2 \frac{1}{n}\sum_{i=1}^{n} E(X_i X_i') \xrightarrow{p} \sigma^2 Q_{XX}$$

Let Ψ_i be a vector collecting unique and nonconstant elements of the $K \times K$ symmetric matrix $X_i X_i'$. Then (7.14) implies

$$Cn = \frac{1}{n}\sum_{i=1}^{n}\left(e_i^2 - s^2\right)\Psi_i \xrightarrow{p} 0$$

This Cn is a sample mean converging to zero. Under some conditions appropriate for CLT to be applicable, we would expect $\sqrt{n}Cn$ to converge to a normal distribution with mean zero and some asymptotic variance Γ. so for any consistent estimator $\hat{\Gamma}$ of Γ, we have

$$n \cdot Cn'\hat{\Gamma}^{-1}Cn \xrightarrow{d} \chi^2(p-1)$$

Where (p-1) is the dimension of Ψ_i.

$$\Gamma = asy.E\left[\left(\sqrt{n}Cn\right)\left(\sqrt{n}Cn\right)'\right] = asy.E\left\{\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(e_i^2 - s^2\right)\Psi_i\right]\left[\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(e_j^2 - s^2\right)\Psi_j'\right]\right\}$$
$$= asy.E\left[\frac{1}{n}\sum_{i=1}^{n}\left(e_i^2 - s^2\right)^2 \Psi_i \Psi_i'\right] = asy.E\left\{E\left[\left(e_i^2 - s^2\right)^2 \middle| \Psi\right] \cdot \left[\frac{1}{n}\sum_{i=1}^{n}\Psi_i \Psi_i'\right]\right\}$$
$$= asy.E\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\left(e_i^2 - s^2\right)^2\right] \cdot \left[\frac{1}{n}\sum_{i=1}^{n}\Psi_i \Psi_i'\right]\right\}$$

Taking $\hat{\Gamma} = \left[\frac{1}{n}\sum_{i=1}^{n}\left(e_i^2 - s^2\right)^2\right] \cdot \left[\frac{1}{n}\sum_{i=1}^{n}\Psi_i \Psi_i'\right]$, thus

$$n \cdot Cn'\hat{\Gamma}^{-1}Cn = n \frac{\left[\sum_{i=1}^{n}\left(e_i^2 - s^2\right)\Psi_i'\right]\left[\sum_{i=1}^{n}\Psi_i \Psi_i'\right]^{-1}\left[\sum_{i=1}^{n}\left(e_i^2 - s^2\right)\Psi_i\right]}{\sum_{i=1}^{n}\left(e_i^2 - s^2\right)^2} = nR^2 \qquad (7.15)$$

This statistic can be computed as nR^2 from the regression e_i^2 on a constant and Ψ_i Then we have

$$nR^2 \xrightarrow{d} \chi^2(p-1) \qquad (7.16)$$

The construction of Ψ_i from $X_i X_i'$, for example, is illustrated as follow:

$$X_i = \begin{pmatrix} 1 & x_{i2} & x_{i3} \end{pmatrix}'$$
$$\Psi_i = \begin{pmatrix} x_{i2} & x_{i3} & x_{i2}^2 & x_{i3}^2 & x_{i2}x_{i3} \end{pmatrix}$$

The power of this test approach unity as $n \to \infty$, against most heteroscedasticity alternatives but may require a fairly large sample to have power close to unity.

7.4.2 Goldfeld-Quandt Test

Assume that the observations can be divided into two groups, σ_I^2 and σ_{II}^2 are variances of disturbances of the two settings separately.

$$H_0 : \sigma_I^2 = \sigma_{II}^2$$
$$H_1 : \sigma_I^2 \neq \sigma_{II}^2$$

Suppose $\sigma_i^2 = \sigma^2 x_{ik}^2$ for some variable x_{ik} $(i = 1, \cdots, n; k = 1, \cdots, K)$. The test procedure is:

i) Reorder the observations by the value of x_{ik}.

ii) Omit c central observations.

iii) Fit separate regressions by OLS to the two sets of observations.

Then, we have the test statistic

$$F = \frac{e_2'e_2 / n_2 - K}{e_1'e_1 / n_1 - K} \sim F\left(n_2 - K, n_1 - K\right) \qquad (7.17)$$

When $n_1 = n_2 = \frac{n-c}{2}$.

$$F = \frac{e_2'e_2}{e_1'e_1} \sim F\left(\frac{n-c}{2} - K, \frac{n-c}{2} - K\right) \qquad (7.18)$$

This test requires

i) ε follows normal distribution.

ii) That $\frac{n-c}{2}$ exceeds the number of parameter.

The power of the test will depend, among other things, on the number of central observations excluded. The power will be low if c is too small or if c is too large. A rough guide is the set $c = \frac{n}{4} \sim \frac{n}{3}$.

7.4.3 The Breusch-Pagan / Godfrey LM Test

Consider the model:

$$y_i = X_i'\beta + \varepsilon_i$$

$$X_i = (1, x_{i2}, \cdots, x_{iK})' \quad \sigma_i^2 = \sigma^2 f_i(\alpha) \quad f_i(\alpha) = e^{\alpha' Z_i}$$

$$\alpha = (\alpha_1, \cdots, \alpha_p)' \qquad Z_i = (1, z_{i2}, \cdots, z_{ip})'$$

$$H_0 : \alpha_2 = \cdots = \alpha_p = 0$$

$f_i(\alpha) = f_i(\alpha_1) = constant$, setting $f_i(\alpha_1) = 1$.

By assuming normally distributed disturbances:

$$f(\varepsilon_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{\varepsilon_i^2}{2\sigma_i^2}\right)$$

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\ln\sigma_i^2 - \frac{1}{2}\sum_{i=1}^{n}\frac{\varepsilon_i^2}{\sigma_i^2}$$

The information matrix $I(\beta,\alpha)$ is block diagonal. Thus we need only concentrate on $\frac{\partial \ln L}{\partial \alpha}$ and the submatrix $I_{\alpha\alpha} = -E\left(\frac{\partial^2 \ln L}{\partial\alpha\partial\alpha'}\right)$.

$$\because \frac{\partial \sigma_i^2}{\partial \alpha} = \sigma_i^2 Z_i$$

$$\frac{\partial \ln \sigma_i^2}{\partial \alpha} = Z_i$$

$$\frac{\partial \left(\sigma_i^2\right)^{-1}}{\partial \alpha} = -\frac{Z_i}{\sigma_i^2}$$

$$\therefore \frac{\partial \ln L}{\partial \alpha} = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\varepsilon_i^2}{\sigma_i^2} - 1\right) Z_i$$

$$\frac{\partial^2 \ln L}{\partial\alpha\partial\alpha'} = -\frac{1}{2}\sum_{i=1}^{n}\frac{\varepsilon_i^2}{\sigma_i^2} Z_i Z_i'$$

Taking expectations, we have

$$I_{\alpha\alpha} = -E\left(\frac{\partial^2 \ln L}{\partial\alpha\partial\alpha'}\right) = \frac{1}{2}\sum_{i=1}^{n} Z_i Z_i' \quad \left(E\left(\varepsilon_i^2 \middle| Z_i\right) = \sigma_i^2 \text{ for all i}\right)$$

Rewrite $\frac{\partial \ln L}{\partial \alpha}$ at the restricted estimates:

$$\left.\frac{\partial \ln L}{\partial \alpha}\right|_{\alpha=0} = \frac{1}{2}\sum_{i=1}^{n} g_i Z_i \text{ , where } g_i = \frac{e_i^2}{\hat{\sigma}^2} - 1 .$$

$$\therefore LM = \left(\frac{1}{2}\sum_{i=1}^{n} g_i Z_i\right)' \left(\frac{1}{2}\sum_{i=1}^{n} Z_i Z_i'\right)^{-1} \left(\frac{1}{2}\sum_{i=1}^{n} g_i Z_i\right) = \frac{1}{2} SSR \sim \chi^2 (p-1) \qquad (7.19)$$

This statistic measure one-half the regression sum of squares from regression of g_i on Z_i, where e_i is the residual from the OLS regression of y_i on X_i and $\hat{\sigma}^2 = e'e/n$.

This statistic is the LM test for multiplicative heteroscedasticity. It also can be written as:

$$LM = \frac{1}{2} SSR = nR^2 \qquad (7.20)$$

where R^2 is the coefficient of determination of the regression of g_i on Z_i.

Return to the regression of g_i on Z_i, the g_i variable has zero mean thus $SSR = R^2 \sum_{i=1}^{n} g_i^2$, and

$$\sum_{i=1}^{n} g_i^2 = \sum_{i=1}^{n} \left(\frac{e_i^2}{\hat{\sigma}^2} - 1\right)^2 = \frac{1}{\hat{\sigma}^4} \sum_{i=1}^{n} e_i^4 - \frac{2}{\hat{\sigma}^2} \sum_{i=1}^{n} e_i^2 + n$$

$$\frac{1}{n} \sum_{i=1}^{n} g_i^2 = \frac{m_4}{m_2^2} - 2 + 1$$

Where m_2 and m_4 denote the second and fourth sample moments about the mean of the OLS residuals. For a normally distributed variable the corresponding population moments obey the relation $\mu_4 = 3\mu_2^2$. Replacing the sample moments by the population equivalents gives $\sum_{i=1}^{n} g_i^2 \simeq 2n$. Thus (7.20) holds.

7.4.4 Modified LM Statistic

The Breusch-Pagan / Godfrey LM test is sensitive to the assumption of normality. Koenker (1981) and Koenker and Bassett (1984) proposed a modified version for this test:

$$V = \frac{1}{n} \sum_{i=1}^{n} \left(e_i^2 - \frac{e'e}{n}\right)^2 \qquad (7.21)$$

The variance of ε_i^2 is not necessary equal to $2\sigma^4$ if ε_i is not normally distributed.

$$Modified\ \ LM = \frac{1}{V}\left(u - \bar{u}i\right)' Z\left(Z'Z\right)^{-1} Z'\left(u - \bar{u}i\right) \qquad (7.22)$$

Where $u = \left(e_i^2, \cdots, e_n^2\right)'$ $i = (1, \cdots, 1)'$, $\bar{u} = \frac{e'e}{n}$. If ε_i is not normal, there is some evidence that it provides a more powerful test.

7.4.5 Glesjer's Tests

Glesjer suggested some specific formulations of the disturbance variance. In particular:

i) $\sigma_i^2 = \mathrm{var}\left(\varepsilon_i \middle| X\right) = \sigma^2\left(\alpha' Z_i\right)$

ii) $\sigma_i^2 = \mathrm{var}\left(\varepsilon_i \middle| X\right) = \sigma^2\left(\alpha' Z_i\right)^2$

iii) $\sigma_i^2 = \mathrm{var}\left(\varepsilon_i \middle| X\right) = \sigma^2 \exp\left(\alpha' Z_i\right)$

For the three cases, testing the hypothesis that all the coefficients except the constant term are zero constitutes a test of the homoscedasticity assumption in the context of the specific formulation. The tests are carried out by the following regressions:

$$e_i^2 = \alpha' Z_i + v_i^* \qquad (7.23)$$

$$e_i = \alpha' Z_i + v_i^* \qquad (7.24)$$

$$\log\left(e_i^2\right) = \alpha' Z_i + v_i^* \qquad (7.25)$$

$$H_0 : \alpha_2 = \cdots = \alpha_p = 0$$

Then, the Wald statistic is computed by

$$W = \hat{\alpha}^{*\prime}\left[Est.A\mathrm{var}\left(\hat{\alpha}^*\right)\right]^{-1}\hat{\alpha}^* \overset{a}{\sim} \chi^2\left(p-1\right) \qquad (7.26)$$

Where $\hat{\alpha}^* = \left(0, I_{p-1}\right)\alpha$.

However, each of these regressions is heteroscedasticity, we would discuss the problem in detail in next section. On the other hand, their power is a function of the specified alternative. If the heteroscedasticity form is incorrect, the tests are like to have limited or no power at all to reveal an incorrect hypothesis of homoscedasticity.

7.5 Estimation under Heteroscedasticity

7.5.1 WLS with Known Ω

If Ω is known, i.e.

$$\sigma^2\Omega = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

Taking $P = \begin{pmatrix} 1/\sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_n \end{pmatrix}$ for the transformed model $y^* = X^*\beta + \varepsilon^*$,we have:

$$y^* = Py = \begin{pmatrix} y_1/\sigma_1 \\ \vdots \\ y_n/\sigma_n \end{pmatrix}, \quad X^* = PX = \begin{pmatrix} 1/\sigma_1 & x_{12}/\sigma_1 & \dots & x_{1K}/\sigma_1 \\ \vdots & \vdots & \ddots & \vdots \\ 1/\sigma_n & x_{n2}/\sigma_n & \dots & x_{nK}/\sigma_n \end{pmatrix}.$$

It is equivalent to dividing both sides of the original equation $y_i = X_i'\beta + \varepsilon_i$ by the square root of σ_i^2, to obtain

$$\frac{y_i}{\sigma_i} = \frac{X_i'}{\sigma_i}\beta + \frac{\varepsilon_i}{\sigma_i} \tag{7.26}$$

and then apply OLS. The GLS reduces to a simple application of weighted least squares.

$$\underset{\hat\beta}{Min}\left[\left(Y^* - X^*\hat\beta\right)'\left(Y^* - X^*\hat\beta\right)\right] = \underset{\hat\beta}{Min}\left[\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\left(y_i - \beta_1 - \beta_2 x_{i2} - \cdots - \beta_K x_{iK}\right)^2\right]$$

For example, suppose $\sigma_i^2 = \sigma^2 x_{iK}$, we minimized the weighted sum of squares

$$\sum_{i=1}^{n}\frac{\varepsilon_i^2}{\sigma^2 x_{iK}} = \sum_{i=1}^{n}\frac{1}{\sigma^2 x_{iK}}\left(y_i - \beta_1 - \beta_2 x_{i2} - \cdots - \beta_K x_{iK}\right)^2$$

It then follows that

$$\sigma^2\Omega = \sigma^2\begin{pmatrix} x_{1K} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{nK} \end{pmatrix}$$

$$\hat\beta_{GLS} = \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}Y = \left[\sum_{i=1}^{n}\frac{1}{x_{iK}}X_i X_i'\right]^{-1}\left[\sum_{i=1}^{n}\frac{1}{x_{iK}}X_i' y_i\right] \tag{7.27}$$

Another example is the heteroscedasticity caused by different grouping frequencies. The model is

$$\bar{y}_g = \bar{X}_g \beta + \bar{\varepsilon}_g \quad g = 1, \cdots, G$$

$$\operatorname{var}\left(\bar{\varepsilon}_g \middle| \bar{X}_g\right) = \frac{\sigma^2}{n_g}$$

$$\operatorname{var}\left(\bar{\varepsilon} \middle| \bar{X}\right) = \sigma^2 diag\left(\frac{1}{n_1}, \cdots, \frac{1}{n_G}\right)$$

So we have $\sigma^2 \Omega = \sigma^2 diag\left(\frac{1}{n_1}, \cdots, \frac{1}{n_G}\right)$.

$$\Omega^{-1} = diag\left(n_1, \cdots, n_G\right), \quad P = diag\left(\sqrt{n_1}, \cdots, \sqrt{n_G}\right).$$

$$\hat{\beta}_{GLS} = \left(\sum_{g=1}^{G} n_g \bar{X}_g \bar{X}_g'\right)^{-1} \left(\sum_{g=1}^{G} n_g \bar{X}_g' \bar{y}_g\right)$$

7.5.2 WLS with Estimated Ω: Two-Step Estimation

When Ω is unknown, we must first find consistent estimator of the unknown parameters in Ω, that is to shed light on the function form of the heteroscedasticity.

Suppose the heteroscedasticity pattern $\sigma_i^2 = \alpha' Z_i$, where Z_i may or may not include the variable in X

$$E\left(\varepsilon_i^2 \middle| Z_i\right) = \sigma_i^2$$

$\varepsilon_i^2 = \sigma_i^2 + v_i$, where v_i is the difference between ε_i^2 and its conditional expectation.

Since ε_i is unobservable, we would use the OLS residuals

$$e_i = y_i - X_i' \hat{\beta}_{OLS} = y_i - X_i' \beta - X_i'\left(\hat{\beta}_{OLS} - \beta\right) = \varepsilon_i + u_i$$

$$e_i^2 = \varepsilon_i^2 + u_i^2 + 2\varepsilon_i u_i$$

As $\hat{\beta}_{OLS} \xrightarrow{P} \beta$, the terms in u_i will become negligible, we have the "model" about variance function

$$e_i^2 = \alpha' Z_i + v_i^* \qquad (7.28)$$

The operational implication for the WLS estimation of the model $y_i = X_i' \beta + \varepsilon_i$ is called two-step estimation:

Step 1: Estimate the equation $y_i = X_i'\beta + \varepsilon_i$ by OLS and compute the OLS residuals e_i; Regress e_i^2 on Z_i to obtain the OLS coefficient estimate $\hat{\alpha}$.

Step 2: Re-estimate the equation $y_i = X_i'\beta + \varepsilon_i$ by WLS, using $1\big/\sqrt{\hat{\alpha}' Z_i}$ as the weight for observation i.

In model (7.28), v_i^* is both heteroscedastic and autocorrelated. So $\hat{\alpha}$ is consistent but inefficient. But, consistency is all that is required for asymptotic efficient estimation of β using $\Omega(\hat{\alpha})$. And we may use White's estimator for the covariance matrix in the regression of e_i^2 on Z_i.

For the both case of WLS estimation: Ω is known and Ω is unknown, the standard t- and F-ratios can be used to do statistical inference.

7.5.3 Testing and Estimating for Groupwise Heteroscedasticity

Susuppose the model

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_G \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_G \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_G \end{pmatrix} \qquad (7.29)$$

With n_g observations in the g^th group, and $\sum_{g=1}^{G} n_g = n$. Within the g^th group:

$$E\left(\varepsilon_{ig} \middle| X_{ig}\right) = 0, \quad \operatorname{var}\left(\varepsilon_{ig} \middle| X_{ig}\right) = \sigma_g^2, \quad i = 1, \cdots, n_g; g = 1, \cdots, G$$

The null hypothesis of homoscedasticity is

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_G^2 = \sigma^2 \qquad (7.30)$$

The log-likelihood for restricted model is

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\sigma^2 I_n\right| - \frac{1}{2}\varepsilon'\left(\sigma^2 I_n\right)^{-1}\varepsilon$$

$$. \qquad = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\varepsilon'\varepsilon$$

Concentrated log-likelihood is:

$$\ln L_{0c} = -\frac{n}{2}\left[1 + \ln(2\pi)\right] - \frac{n}{2}\ln\hat{\sigma}^2 \qquad (7.31)$$

Under the alternative hypothesis of heteroscadasticity across G groups, the covariance matrix of $\varepsilon = \left(\varepsilon_1, \cdots \varepsilon_G\right)'$ is

$$V=\sigma^2\Omega=\begin{pmatrix}\sigma_1^2 I_{n_1} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sigma_G^2 I_{n_G}\end{pmatrix}$$

And the log-likelihood function is

$$\ln L_1=-\frac{n}{2}\ln(2\pi)-\frac{1}{2}\sum_{g=1}^{G}n_g\ln\sigma_g^2-\frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{n_g}\left(\varepsilon_{ig}/\sigma_g\right)^2$$

$$.\qquad =-\frac{n}{2}\ln(2\pi)-\frac{1}{2}\sum_{g=1}^{G}n_g\ln\sigma_g^2-\sum_{g=1}^{G}\frac{1}{2\sigma_g^2}\left(y_g-X_g{}'\beta\right)'\left(y_g-X_g{}'\beta\right)$$

The concentrated log-likelihood function is:

$$\ln L_{1c}=-\frac{n}{2}\left[1+\ln(2\pi)\right]-\frac{1}{2}\sum_{g=1}^{G}n_g\ln\hat{\sigma}_g^2 \qquad (7.32)$$

The LR statistic is

$$LR=-2\left(\ln L_{0c}-\ln L_{1c}\right)=n\ln\hat{\sigma}^2-\sum_{g=1}^{G}n_g\ln\hat{\sigma}_g^2 \qquad (7.33)$$

$$\hat{\sigma}^2=\frac{e'e}{n},\quad \hat{\sigma}_g^2=\frac{e_g'e_g}{n_g},\quad (g=1,\cdots,G) \qquad (7.34)$$

Where $\hat{\sigma}^2=\frac{e'e}{n},\quad \hat{\sigma}_g^2=\frac{e_g'e_g}{n_g},\quad (g=1,\cdots,G)$ are MLE of σ^2 and σ_g^2 respectively, the degree of freedom of this LR statistic is G-1.

If the variances are known, then the GLS estimator is

$$\hat{\beta}=\left(X'V^{-1}X\right)^{-1}X'V^{-1}Y$$

$$=\left(\sum_{g=1}^{G}\frac{1}{\sigma_g^2}X_g'X_g\right)^{-1}\left(\sum_{g=1}^{G}\frac{1}{\sigma_g^2}X_g'y_g\right)$$

$$=\left(\sum_{g=1}^{G}\frac{1}{\sigma_g^2}X_g'X_g\right)^{-1}\left(\sum_{g=1}^{G}\frac{1}{\sigma_g^2}X_g'X_g\hat{\beta}_g\right)\qquad\left(X_g'y_g=X_g'X_g\hat{\beta}_g\right)$$

$$=\left(\sum_{g=1}^{G}V_g\right)^{-1}\left(\sum_{g=1}^{G}V_g\hat{\beta}_g\right)$$

$$=\left(\sum_{g=1}^{G}w_g\hat{\beta}_g\right) \qquad (7.35)$$

Where $\hat{\beta}_g$ is the OLS estimator in the g^th^ group, $V_g=\mathrm{var}\left(\hat{\beta}_g\right)$. Thus the $\hat{\beta}$ is a matrix weighted average of G least estimators, the weighted matrices are:

$$w_g=\left(\sum_{g=1}^{G}V_g\right)^{-1}V_g$$

If X_g is same in every group, $w_g = \left(1/\sigma_g^2\right) \Big/ \sum \left(1/\sigma_g^2\right)$ .If the variances are unknown, we have FGLS estimator:

$$\hat{\beta}_{FGLS} = \left(X'\hat{V}^{-1}X\right)^{-1} X'\hat{V}^{-1}Y$$

$$. = \left(\sum_{g=1}^{G} \frac{1}{\hat{\sigma}_g^2} X_g' X_g\right)^{-1} \left(\sum_{g=1}^{G} \frac{1}{\hat{\sigma}_g^2} X_g' y_g\right) \qquad (7.36)$$

One might consider iterating the estimator with two-step FGLS estimator. Under the assumption of normally distributed disturbances, so long as (7.34) is used without a degrees of freedom correction, and if the iterating does converge, then it will converge to MLE.

7.6 White's Heteroscedasticity Consistent **Covariance** Estimator.

7.6.1 White Heteroscedasticity Consistent Covariance Estimator

In the GRM, If Q=plim$\left(\frac{X'X}{n}\right)$and plim$\left(\frac{X'\Omega X}{n}\right)$ are both finite positive definite matrices, we have $\sqrt{n}\left(\hat{\beta} - \beta\right) = \left(\frac{X'X}{n}\right)^{-1} \frac{1}{\sqrt{n}} X'\varepsilon$

Where $\hat{\beta}$ is the OLS estimator. In the heteroscedastic case, if the variance of ε_i are not dominated by any single term, so that the conditions of the Lindberg-Feller central limit theorem can be applied to

$$v_{n,ls} = Q^{-1} \frac{1}{\sqrt{n}} X'\varepsilon = Q^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \varepsilon_i$$

Then the least squares estimator is asymptotically normally distributed with covariance matrix

$$AVar\left(\hat{\beta}\right) = \frac{\sigma^2}{n} Q^{-1} p\lim\left(\frac{1}{n} X'\Omega X\right) Q^{-1}$$

$$= \frac{1}{n} Q^{-1} p\lim\left(\frac{1}{n} X'\left(\sigma^2 \Omega\right) X\right) Q^{-1} \qquad (7.37)$$

$$= \frac{1}{n} Q^{-1} p\lim\left(\frac{1}{n} \sum_{i=1}^{n} \sigma_i^{\,2} X_i X_i'\right) Q^{-1}$$

White(1980) had shown that under very general conditions, the estimator

$$S_0 = \frac{1}{n} \sum_{i=1}^{n} e_i^2 X_i X_i'$$

has plim$S_0 =$plimQ^*, where $Q^* = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 X_i X_i'$

In order to obtain the consistent estimator of Q^*, we need two assumptions as following:

Assumption 1: (X_i, ε_i) is independent but not identical distribution for i, and its fourth-moment exists, and satisfies $E(X_i \varepsilon_i) = 0$.

Assumption 2: For some finite positive scalar δ and Δ, we have $E\left(\left|x_{ij}^2\right|^{1+\delta}\right) < \Delta$ and $E\left(\left|\varepsilon_i^2 x_{ij}^2\right|^{1+\delta}\right) < \Delta$.

To indicated why the fourth-moment assumption is need for the regressors, we provide a sketch of the proof for the special case of k=1. So, both x_i and $x_i\varepsilon_i$ are scalars, we have

$$\begin{aligned} e_i^2 &= \left(y_i - x_i'\hat{\beta}\right)^2 \\ &= \left(\varepsilon_i - \left(\hat{\beta} - \beta\right) x_i\right)^2 \\ &= \varepsilon_i^2 - 2\left(\hat{\beta} - \beta\right)\varepsilon_i x_i + \left(\hat{\beta} - \beta\right)^2 x_i^2 \end{aligned} \tag{7.38}$$

Where $\hat{\beta}$ is the OLS estimator, plim $\hat{\beta} = \beta$, by multiplying both sides of (7.38) by x_i^2 and summing over i, we obtain

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i^2 - \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 x_i^2 &= -2\left(\hat{\beta} - \beta\right)\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i x_i^3 + \left(\hat{\beta} - \beta\right)^2 \frac{1}{n}\sum_{i=1}^{n} x_i^4 \\ &\leq \left|\hat{\beta} - \beta\right| \frac{1}{n}\sum_{i=1}^{n}\left(\varepsilon_i^2 x_i^2 + x_i^4\right) + \left(\hat{\beta} - \beta\right)^2 \frac{1}{n}\sum_{i=1}^{n} x_i^4 \end{aligned} \tag{7.39}$$

If the fourth-moment $E\left(x_i^4\right)$ is finite number, then $\frac{1}{n}\sum x_i^4$ converges in probability to some finite number, and $\left(\hat{\beta} - \beta\right)^2 \frac{1}{n}\sum x_i^4 \xrightarrow{P} 0$. And since $E\left(\varepsilon_i^2 x_i^2\right)$ is finite, $\frac{1}{n}\sum x_i^2 \varepsilon_i^2$ converges in probability to some finite number, then $\left(\hat{\beta} - \beta\right)\frac{1}{n}\sum \varepsilon_i^2 x_i^2 \xrightarrow{P} 0$.

From Assumption 2, use Markov's Strong LLN, we can have

$$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 x_i^2 - \frac{1}{n}\sum_{i=1}^{n}E\left(\varepsilon_i^2 x_i^2\right) \xrightarrow{a.s.} 0$$

So, $P\lim\frac{1}{n}\sum e_i^2 x_i x_i' = P\lim\frac{1}{n}\sum \varepsilon_i^2 x_i x_i' = P\lim\frac{1}{n}\sum E\left(\varepsilon_i^2 x_i x_i'\right)$

Defining so is the n*n diagonal matrix whose i-th element is e_i^2, then the $Est.Av\text{ar}\left(\hat{\beta}\right)$ can be rewritten as

$$Est.Av\text{ar}\left(\hat{\beta}\right) = n\left(X'X\right)^{-1} S_0 \left(X'X\right)^{-1}$$
$$= \left(X'X\right)^{-1}\left[X'\begin{pmatrix} e_1^2 & & 0 \\ & \ddots & \\ 0 & & e_n^2 \end{pmatrix}X\right]\left(X'X\right)^{-1} \qquad (7.40)$$

7.6.2 The improvement on the white estimator:

The white heteros-consistent covariance estimator tends to underestimate the squares of the true disturbances. Davidson and Mackinnon(1993) report that at least for the Monte Carlo simulations they have seen, the robust t-ratio based on $S_0 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 X_i X_i'$ rejects the null too often, and that simply multiplying S_0 by $\frac{n}{n-k}$ mitigates the problem of over-rejected. They also report that the robust t-ration based on the following adjustment on S_0 reform over better

$$S_0 = \frac{1}{n}\sum_{i=1}^{n}\frac{e_i^2}{(1-p_i)^d} X_i X_i' \quad \text{d=1 or 2} \qquad (7.41)$$

Where $P_i = X_i'\left(X'X\right)^{-1} X_i = \frac{1}{n} X_i'\left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} X_i$

The estimated asymptotic variance

$$Est.Av\text{ar}\left(\hat{\beta}\right) = \frac{1}{n}\left(\frac{1}{n}\sum X_i X_i'\right)^{-1}\left(\frac{1}{n}\sum e_i^2 X_i X_i'\right)\left(\frac{1}{n}\sum X_i X_i'\right)^{-1}$$
$$= \left(X'X\right)^{-1}\left(X'\hat{\Omega}X\right)\left(X'X\right)^{-1} \qquad (7.42)$$

Which is consistent of $Av\text{ar}\left(\hat{\beta}\right)$, where $\hat{\Omega} = diag(e_1^2, \ldots\ldots, e_n^2)$.