Ch. 9 Heteroscedasticity

(June 13, 2016)

1 Introduction

Regression disturbances whose variance are not constant across observations are **heteroscedastic**. There are several reasons why the disturbance of ε_i may be variable, some of which are as follows.[1]

(a). Following the error-learning models, as people learn, their errors of behavior become smaller over time.

(b). As income grows, people have more discretionary income and hence more scope for choice about the disposition of their income. Hence, σ_i^2 is likely to increase with income.

(c). As data collecting techniques improve, σ_i^2 is likely to decrease. etc. □

In the heteroscedastic model, the variances of the disturbances are

$$Var(\varepsilon_i|\mathbf{x}) = \sigma_i^2,\ i = 1, 2, ..., N.$$

We continue to assume that the disturbances are pairwise uncorrelated. Thus,

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{x}) = \begin{bmatrix} \sigma_1^2 & 0 & . & . & . & 0 \\ 0 & \sigma_2^2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \sigma_N^2 \end{bmatrix}.$$

[1] See Gujarati (2003), *Basic Econometrics*. p. 389.

It will sometimes prove useful to write $\sigma_i^2 = \sigma^2\omega_i$. Hence

$$\begin{aligned} E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{x}) &= \sigma^2 \begin{bmatrix} \omega_1 & 0 & . & . & . & 0 \\ 0 & \omega_2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \omega_N \end{bmatrix}, \\ &= \sigma^2\boldsymbol{\Sigma}, \end{aligned}$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \omega_1 & 0 & . & . & . & 0 \\ 0 & \omega_2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \omega_N \end{bmatrix}. \tag{9-1}$$

This form is an arbitrary scaling which allows us to use a normalization,

$$tr(\boldsymbol{\Sigma}) = \sum_{i=1}^{N} \omega_i = N.$$

(For example, $\sigma^2 = \frac{\sum_{i=1}^{N}\sigma_i^2}{N}$.) This makes the classical regression with homoscedastic disturbances a simple special case with $\omega_i = 1, i = 1, 2, ..., N$.

Example.

See Figure 11.1 (p.216) of Greene 5th edition. ■

2 Testing for Heteroscedasticity

One can rarely be certain that the disturbances are heteroscedastic however, and unfortunately, what form the heteroscedasticity takes if they are. As such, it is useful to be able to test for homoscedasticity and if necessary, modify our estimation procedure accordingly.

Most of the test for heteroscedasticity are based on the following strategy. OLS estimator is a consistent estimator of $\boldsymbol{\beta}$ even in the presence of heteroscedasticity. As such, the OLS residuals will mimic, albeit imperfectly because of sampling variability, the heteroscedasticity of the true disturbance. Therefore, tests designed to detect heteroscedasticity will, in most cases, be applied to the OLS residuals.

2.1 Nonspecific Tests for Heteroscedasticity

2.1.1 White's General Test

White (1980) addressed the case where nothing is known about the structure of the heteroscedasticity other than the heteroscedastic variance σ_i^2 are uniformly bounded. It would be desirable to be able to test a general hypothesis of the form:

$$\begin{aligned} H_0 &: \quad \sigma_i^2 = \sigma^2 \quad for\ all\ i, \\ H_1 &: \quad Not\ H_0. \end{aligned}$$

If there is no heteroscedasticity (under H_0), then $s^2(\mathbf{X}'\mathbf{X})$ will give a consistent estimator of variance $\hat{\boldsymbol{\beta}}$, where if there is, then it will not (see Ch. 8 sec.1). The correct covariance matrix for the OLS is estimated by

$$\begin{aligned} &\widehat{Var(\hat{\boldsymbol{\beta}})}_{HAC} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{t=1}^{T} e_t^2 \mathbf{x}_t \mathbf{x}_t')\right).(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

White derives a test for heteroscedasticity which consists of comparing the elements of $N\mathbf{S}_0(=\sum_{i=1}^{N} e_t^2 \mathbf{x}_t \mathbf{x}_t')$ and $s^2(\mathbf{X}'\mathbf{X})(= s^2 \sum_{t=1}^{N} \mathbf{x}_t \mathbf{x}_t')$, thus indicating whether or not the usual OLS formula $s^2(\mathbf{X}'\mathbf{X})$ is a consistent covariance estimator. Large discrepancies between $N\mathbf{S}_0$ and $s^2(\mathbf{X}'\mathbf{X})$ support the contention of heteroscedasticity while small discrepancies support homoscedasticity.

A simple operational version of this test is carried out by obtaining NR^2 in the regression of e_i^2 on a constant and all unique variables in $\mathbf{x} \otimes \mathbf{x}$. This statistics is asymptotically distributed as χ_p^2, where p is the number of regressors in the regression, including the constant.

2.2 Specific Tests for Heteroscedasticity

If nothing is known a priori other than the heteroscedastic variance are uniformly bounded, White (1980) general test is applicable.

There may be instance when the form of the heteroscedasticity is not known, but nevertheless, it is known that the disturbance variance in monotonically related to the size of a known exogenous variable Z by which observations on the dependent variable Y can be ordered. One frequently used test in this instance is the Goldfeld-Quandt test.

When it is believed that the broader class of heteroscedasticity is $\sigma_i^2 = h(\mathbf{z}_i'\boldsymbol{\alpha})$, where $h(\cdot)$ is a general function independent of i, is applicable (such as $\sigma_i^2 = \mathbf{z}_i'\boldsymbol{\alpha}$, $\sigma_i^2 = (\mathbf{z}_i'\boldsymbol{\alpha})^2$ and $\sigma_i^2 = \exp(\mathbf{z}_i'\boldsymbol{\alpha})$). If so, the Breush-Pagan test is appropriate.

2.2.1 The Goldfeld-Quandt Test

A very popular test for determining the presence of heteroscedasticity which is monotonically related to an exogenous variable by which observations on the dependent variable can be ordered is the Goldfeld-Quandt (1965) test.

For the GoldfeldQuandt test, we assume that the observations can be divided into two groups in such a way that under the hypothesis of homoscedasticity, the disturbance variances would be the same in the two groups, whereas under the alternative, the disturbance variances would differ systematically. The steps of this test are as follow:

(a). Order the observations (from "supposed" large to small variance) by the values of the variables Z.

(b). Choose p central observations and omit them, provides $(N-p)/2 > k$.

(c). Fit separate regression by OLS to the two groups, with N_1 and N_2 observations, respectively.

(d). Let SSE_1 and SSE_2 denote the sum of squared residuals based on the large variance (which you suppose they do) and the small variance group, respectively. □

Recall that $\frac{\mathbf{e}_1'\mathbf{e}_1}{\sigma_1^2} \sim \chi^2_{[N_1-k]}$ and $\frac{\mathbf{e}_2'\mathbf{e}_2}{\sigma_2^2} \sim \chi^2_{[N_2-k]}$, then the statistics

$$F = \frac{\mathbf{e}_1'\mathbf{e}_1/N_1 - k}{\mathbf{e}_2'\mathbf{e}_2/N_2 - k} = \frac{SSE_1/N_1 - k}{SSE_2/N_2 - k} \sim F_{[N_1-k,N_2-k]},$$

under the null hypothesis of homoscedasticity $\sigma_1^2 = \sigma_2^2 = \sigma^2$,

2.2.2 The Breush-Pagan Test

The Goldfeld-Quandt test has been found to be reasonably powerful when we are able to identify correctly the variable to use in the sample separation. This requirement does limit its generality, however. Breush-Pagan (1979) assume a border class of heteroscedasticity defined by

$$\sigma_i^2 = \sigma^2 h(\alpha_0 + \mathbf{z}_i'\boldsymbol{\alpha}),$$

where $\mathbf{z}_i$ is a $(p \times 1)$vector of exogenous variables. This model is homoscedastic if $\boldsymbol{\alpha} = 0$.

Breush and Pagen (1979) consider the general estimation equation

$$\frac{\hat{e}_i^2}{\bar{\sigma}^2} = \alpha_0 + \mathbf{z}_i'\boldsymbol{\alpha} + v_i,$$

where $\hat{e}_i$ represents the $i-th$ OLS residual and $\bar{\sigma}^2 = \sum_{i=1}^N \hat{e}_i^2/N$. The null hypothesis $\boldsymbol{\alpha_1} = 0$ can be tested if the ε_i are normally distributed. Let SSR denote the sum of squares obtained in an OLS estimation of

$$\frac{\hat{e}_i^2}{\bar{\sigma}^2} = \hat{\alpha}_0 + \mathbf{z}_i'\hat{\boldsymbol{\alpha}} + \hat{v}_i.$$

Denote $W_i = \frac{\hat{e}_i^2}{\bar{\sigma}^2}, \bar{W} = \sum_{i=1}^N W_i/N$, and $\hat{W}_i = \hat{\alpha}_0 + \mathbf{z}_i'\hat{\boldsymbol{\alpha}}_\mathbf{1}$. Then $SSR = \sum_{i=1}^N(\hat{W}_i - \bar{W})^2$. Breush and Pagan show, if $\boldsymbol{\alpha} = 0$, then

$$\frac{1}{2}SSR \sim \chi_p^2.$$

Exercise 1.

Reproduce the results of Example 11.3 at p.224 of Greene 5th edition. ■

3 OLS Estimation

We showed in Section 8.2 that in the presence of heteroscedasticity, the OLS estimator $\hat{\boldsymbol{\beta}}$ is unbiased and consistent. However it is inefficient relative to the GLS estimator.

3.1 Estimating the Appropriate Covariance Matrix for OLS Estimators

If the type of heteroscedasticity is known with certainty, then the OLS estimator is undesirable; we should use the GLS instead. The precise form of the heteroscedasticity is usually unknown, however. In that case, GLS is not usable, and we may need to salvage what we can from the results of OLS estimators.

The conventional estimated covariance matrix for the OLS estimator $\sigma^2(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}$ is inappropriate; the appropriate matrix is $\sigma^2(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}$. White (1980) has shown that it is still possible to obtain an appropriate covariance estimator of the OLS estimators even the form of heteroscedasticity is unknown. What is actually required is an estimate of

$$\boldsymbol{\Omega} = \frac{1}{N}\sigma^2\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X} = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2\mathbf{x}_i\mathbf{x}_i'.$$

White (1980) shows that under very general conditions, the matrix

$$\mathbf{S}_0 = \frac{1}{N}\sum_{i=1}^{N}e_i^2\mathbf{x}_i\mathbf{x}_i',$$

where e_i is the $i-th\ OLS$ residual, is a consistent estimator of $\boldsymbol{\Omega}$. Therefore, the **White estimator**,

$$\widehat{Var(\hat{\boldsymbol{\beta}})} = N(\mathbf{X}'\mathbf{X})^{-1}\mathbf{S}_0(\mathbf{X}'\mathbf{X})^{-1}, \tag{9-2}$$

can be used as an estimator of the true variance of the OLS estimator. Inference concerning $\boldsymbol{\beta}$ is still possible by means of OLS estimator even when the specific structure of $\boldsymbol{\Sigma}$ is not specified as $\hat{\boldsymbol{\beta}}$ is normally distributed asymptotically.

More generally, White shows that tests of the general linear hypothesis $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, under the null hypothesis, the statistics

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}N\mathbf{S}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}) \xrightarrow{L} \chi_m^2,$$

where m denote the number of restrictions imposed.

Exercise 2.

Reproduce the results at Table 11.1 on p. 221 of Greene 5th edition. ■

4 GLS (Weighted Least Squares)

Having tested for and found evidence of heteroscedasticity, the logical next is to revise the estimation technique to account for it if $\mathbf{\Sigma}$ is known. The GLS estimator is

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y}. \tag{9-3}$$

Consider the most general case, $\sigma_i^2 = \sigma^2\omega_i$. Then $\mathbf{\Sigma}^{-1}$ is a diagonal matrix whose $i-th$ diagonal element is $1/\omega_i$ (See Eq. (9-1)), that is

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} \frac{1}{\omega_1} & 0 & . & . & . & 0 \\ 0 & \frac{1}{\omega_2} & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \frac{1}{\omega_N} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{\omega_1}} & 0 & . & . & . & 0 \\ 0 & \frac{1}{\sqrt{\omega_2}} & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \frac{1}{\sqrt{\omega_N}} \end{bmatrix} \times \begin{bmatrix} \frac{1}{\sqrt{\omega_1}} & 0 & . & . & . & 0 \\ 0 & \frac{1}{\sqrt{\omega_2}} & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \frac{1}{\sqrt{\omega_N}} \end{bmatrix}$$

$$= \mathbf{P}'\mathbf{P}.$$

The GLS is obtained by regressing (See Eq (8-18))

$$\mathbf{Py} = \begin{bmatrix} Y_1/\sqrt{\omega_1} \\ Y_2/\sqrt{\omega_2} \\ . \\ . \\ . \\ Y_N/\sqrt{\omega_N} \end{bmatrix} \; on \quad \mathbf{PX} = \begin{bmatrix} \mathbf{x}'_1/\sqrt{\omega_1} \\ \mathbf{x}'_2/\sqrt{\omega_2} \\ . \\ . \\ . \\ \mathbf{x}'_N/\sqrt{\omega_N} \end{bmatrix}.$$

Applying OLS to the transformed model, we obtain the GLS estimator, which is also called **weighted least squares (WLS)** estimator,

$$\tilde{\boldsymbol{\beta}} = \left[\sum_{i=1}^{N} w_i\mathbf{x}_i\mathbf{x}'_i\right]^{-1}\left[\sum_{i=1}^{N} w_i\mathbf{x}_iY_i\right],$$

where $w_i = 1/\omega_i$.

The logic of the computation is that observations with smaller variances receive a large weight in the computations of the sums and therefore have greater influence in

the estimate obtained.

Example.

A common specification in linear regression model with heteroscedasticity is that the variance of the disturbances is proportional to one of the regressors or its square. For example, if the model is

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i,\ i = 1, 2, ..., N,$$

where

$$\sigma_i^2 = \sigma^2 X_{il}^2,$$

then

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} \frac{1}{X_{1l}^2} & 0 & . & . & . & 0 \\ 0 & \frac{1}{X_{2l}^2} & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \frac{1}{X_{Nl}^2} \end{bmatrix},$$

and

$$\mathbf{P} = \begin{bmatrix} \frac{1}{X_{1l}} & 0 & . & . & . & 0 \\ 0 & \frac{1}{X_{2l}} & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \frac{1}{X_{Nl}} \end{bmatrix}.$$

Hence, the transformed regression model for the GLS is

$$\begin{aligned} \frac{Y_i}{X_{il}} &= \beta_1 \left(\frac{X_{i1}}{X_{il}}\right) + \beta_2 \left(\frac{X_{i2}}{X_{il}}\right) + ... + \beta_l \left(\frac{X_{il}}{X_{il}}\right) + ... + \beta_k \left(\frac{X_{ik}}{X_{il}}\right) + \frac{\varepsilon_i}{X_{il}} \\ &= \beta_l + \beta_1 \left(\frac{X_{i1}}{X_{il}}\right) + \beta_2 \left(\frac{X_{i2}}{X_{il}}\right) + ... + \beta_k \left(\frac{X_{ik}}{X_{il}}\right) + \frac{\varepsilon_i}{X_{il}}, \end{aligned}$$

where $E\left(\frac{\varepsilon_i}{X_{il}}\right)^2 = \frac{\sigma^2 X_{il}^2}{X_{il}^2} = \sigma^2$, $\forall i$.

If the variance σ_i^2 is proportional to X_{il} instead of X_{il}^2, then the weight applied to each observation is $1/\sqrt{X_{il}}$ instead of $1/X_{il}$. ■

5 Estimation When Σ is Unknown

The general form of the heteroscedastic regression model has too many parameters to estimate by ordinary method as shown in Section 8.4. Typically, the model is restricted by formulating $\sigma^2\boldsymbol{\Sigma}$ as a function of a few parameters, such as $\sigma_i^2 = \sigma^2 X_{il}^{\alpha}$ or $\sigma_i^2 = \sigma^2[\mathbf{z}_i'\boldsymbol{\alpha}]^2$. Write this as $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$, FGLS based on a consistent estimator of $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$ is asymptotically equivalent to GLS. The new problem is that we must first find consistent estimators of the unknown parameters in $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$. Two methods are typically used, two step GLS and maximum likelihood.

5.1 FGLS, Two-Step Estimation

For the heteroscedastic model, the GLS estimator is

$$\tilde{\boldsymbol{\beta}} = \left[\sum_{i=1}^{N}\left(\frac{1}{\sigma_i^2}\right)\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{N}\left(\frac{1}{\sigma_i^2}\right)\mathbf{x}_i Y_i\right].$$

The **two step estimators** are computed by first obtaining estimators $\hat{\sigma}_i^2$, usually using some function of the OLS residuals, then the FGLS will be

$$\check{\boldsymbol{\beta}} = \left[\sum_{i=1}^{N}\left(\frac{1}{\hat{\sigma}_i^2}\right)\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{N}\left(\frac{1}{\hat{\sigma}_i^2}\right)\mathbf{x}_i Y_i\right]. \tag{9-4}$$

The OLS estimator $\hat{\boldsymbol{\beta}}$, although inefficient, is still consistent. As such, statistics computed using the OLS residual, $e_i = (Y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}})$, will have the same asymptotic properties as those computed using the true disturbance, $\varepsilon_i = (Y_i - \mathbf{x}_i'\boldsymbol{\beta})$

Let

$$\varepsilon_i^2 = \sigma_i^2 + v_i,$$

where v_i is just the difference between the random variable ε_i^2 and its expectation. Since ε_i is unobservable, we would use the OLS residual, for which

$$e_i = \varepsilon_i - \mathbf{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \varepsilon_i + u_i.$$

But in large sample, as $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$, terms in u_i will become negligible, so that at least approximately,

$$e_i^2 = \sigma_i^2 + v_i^*.$$

The procedure suggested is to treat the variance function as a regression and use the squares of the OLS residual as the dependent variable. For example, if $\sigma_i^2 = \mathbf{z}_i'\boldsymbol{\alpha}$, then a consistent estimator of $\boldsymbol{\alpha}$ will be the OLS in the model[2]

$$e_i^2 = \mathbf{z}_i'\boldsymbol{\alpha} + v_i^*, \quad i = 1, 2, ..., N. \tag{9-5}$$

Having obtained the estimated $\hat{\boldsymbol{\alpha}}$ in the first step from (9-6), then we substitute $\hat{\sigma}_i^2 = \mathbf{z}_i'\hat{\boldsymbol{\alpha}}$ into Eq. (9-5), we finish the second step and the FGLS estimator is thus obtained.

The two-step estimator may be iterated by recomputing the residuals after computing the FGLS estimate and then reentering the computation

$$OLS \to \hat{\boldsymbol{\beta}} \to \mathbf{e} \overset{Eq.(9-6)}{\longrightarrow} \hat{\boldsymbol{\alpha}}^{(1)} \overset{Eq.(9-5)}{\longrightarrow} \check{\boldsymbol{\beta}}^{(1)} \to \check{\mathbf{e}}^{(1)} \overset{Eq.(9-6)}{\longrightarrow} \hat{\boldsymbol{\alpha}}^{(2)} \overset{Eq.(9-5)}{\longrightarrow} \check{\boldsymbol{\beta}}^{(2)} \to$$
$$\to \check{\mathbf{e}}^{(2)} \to .....,$$

where

$$\check{\mathbf{e}}^{(1)} = \mathbf{y} - \mathbf{X}\check{\boldsymbol{\beta}}^{(1)}.$$

5.2 Maximum Likelihood Estimation

The log-likelihood function for a sample of normally distributed observations with heteroscedastic variance is

$$\ln L = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\left[\ln\sigma_i^2 + \frac{1}{\sigma_i^2}(Y_i - \mathbf{x}_i'\boldsymbol{\beta})^2\right].$$

For simplicity, let

$$\sigma_i^2 = \sigma^2 f_i(\boldsymbol{\alpha}),$$

where $\boldsymbol{\alpha}$ is the vector of unknown parameters in $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$ and $f_i(\boldsymbol{\alpha})$ is indexed by i to indicate that is a function of $\mathbf{z}_i$. Assume as well that no elements of $\boldsymbol{\beta}$ appear in $\boldsymbol{\alpha}$. The log-likelihood function is

$$\ln L = -\frac{N}{2}[\ln(2\pi) + \ln\sigma^2] - \frac{1}{2}\sum_{i=1}^{N}\left[\ln f_i(\boldsymbol{\alpha}) + \frac{1}{\sigma^2}\left(\frac{1}{f_i(\boldsymbol{\alpha})}\right)(Y_i - \mathbf{x}_i'\boldsymbol{\beta})^2\right].$$

[2]In this model, v_i^* may be both heteroscedastic and autocorrelated, so $\hat{\boldsymbol{\alpha}}$ is consistent but inefficient. But, consistency is all that is required for asymptotically efficient estimation of $\boldsymbol{\beta}$ using $\boldsymbol{\Sigma}(\hat{\boldsymbol{\alpha}})$.

For convenience in what follows, substitute ε_i for $(Y_i - \mathbf{x}_i'\boldsymbol{\beta})$, denote $f_i(\boldsymbol{\alpha})$ as simply f_i, and denote the vector of derivatives $\partial f_i(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}$ as $\mathbf{g}_i$. Then the derivatives of the log-likelihood functions are

$$\begin{aligned}
\frac{\partial \ln L}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{N} \mathbf{x}_i \frac{\varepsilon_i}{\sigma^2 f_i}, \\
\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}\frac{\varepsilon_i^2}{f_i} = \sum_{i=1}^{N}\left(\frac{1}{2\sigma^2}\right)\left(\frac{\varepsilon_i^2}{\sigma^2 f_i} - 1\right), \\
\frac{\partial \ln L}{\partial \boldsymbol{\alpha}} &= \sum_{i=1}^{N}\left(\frac{1}{2}\right)\left(\frac{\varepsilon_i^2}{\sigma^2 f_i} - 1\right)\left(\frac{1}{f_i}\right)\mathbf{g}_i.
\end{aligned}$$

The maximum likelihood estimators are those values of $\boldsymbol{\beta}$, σ^2, and $\boldsymbol{\alpha}$ that simultaneously equate these derivatives to zero. The likelihood equations are generally **highly nonlinear** and will usually require an iteration solution.

Exercise 3

Reproduce the results at Table 11.2 on p.231 of Greene 5th edition. ■

6 $ARCH$ Model

Heteroscedasticity is often associated with cross-sectional data, whereas time series are usually studied in the context of homoscedastic processes. In analyses of macroeconomic data, Engle (1982, 1983) and Cragg (1982) found evidence that for some kinds of data, the disturbance variances in time-series models were less stable than usually assumed.

With time-series data, it is not uncommon to see that the OLS residuals to be quite small in absolute value for a number of successive periods of time, then much larger for a while, then smaller again, and so on. This phenomenon of time-varying volatility (or disturbances occur in clusters) is often encountered in models for stock returns, foreign exchange rates, and other series that are determined in financial markets. Numerous models for dealing with this phenomenon have been proposed. One very popular approach is based on the concept of autoregressive, conditionally heteroscedastic, or ARCH, that was introduced by Engle (1982). The basic idea of ARCH models is that the variance of the disturbance at time t depends on the realized values of squared disturbances in previous time periods.

A model which allows the conditional variance to depend on the past realization of the series is considered in the following. Suppose that

$$u_t = \sqrt{h_t}\varepsilon_t \tag{9-6}$$

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2, \tag{9-7}$$

with $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = 1$, then this is an example of what will be called an **autoregressive conditional heteroscedasticity** ($ARCH(1)$) model.

Example.

See Figure 11.3 on p.239 of Greene 5th edition. ■

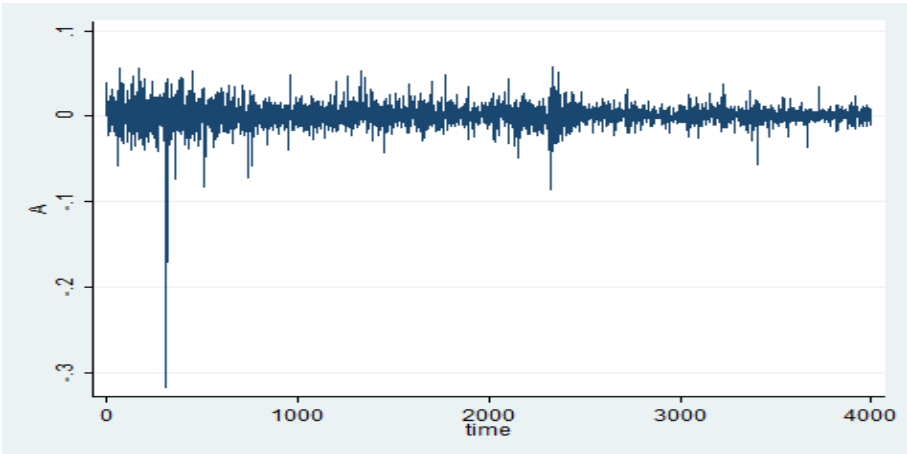

Figure (9-1). An Example of Volatility Clustering.

6.1 Population's Properties of $ARCH$ Models

6.1.1 The Conditional Mean and Variance

Let F_{t-1} denote the information set available at time $t-1$. The conditional mean of u_t is

$$E(u_t|F_{t-1}) = \sqrt{h_t} \cdot E(\varepsilon_t|F_{t-1}) = 0. \tag{9-8}$$

From (9-9) it implies that the conditional variance of u_t is

$$\begin{aligned}
\sigma_t^2 &= Var(u_t|F_{t-1}) \\
&= E\{[u_t - E(u_t|F_{t-1})]^2|F_{t-1}\} \\
&= E(u_t^2|F_{t-1}) \qquad (since\ E(u_t|F_{t-1}) = 0) \\
&= E(h_t\varepsilon_t^2|F_{t-1}) \\
&= E(\varepsilon_t^2)E(\alpha_0 + \alpha_1 u_{t-1}^2|F_{t-1}) \\
&= \alpha_0 + \alpha_1 u_{t-1}^2 \\
&= h_t,
\end{aligned}$$

so u_t is conditionally heteroscedastic. From the structure of the model, it is seen that large past squared shocked shocks $u_{t-i}^2, i = 1, .., m$ imply a large conditional variance $\sigma_t^2 (= Var(u_t|F_{t-1}))$ for this variable u_t. Consequently, u_t tends to assume a large value. This means that, under the $ARCH$ framework, large shocks tend to be followed by another large chock. This feature is similar to the volatility clustering observed in asset

returns.

6.1.2 The Conditional Density

By assuming that ε_t is a Gaussian variate, the condition density of u_t given all the information update to $t-1$ is

$$f(u_t|F_{t-1}) = \sqrt{h_t} f(\varepsilon_t|F_{t-1}) = \sqrt{h_t} \cdot N(0,1) \sim N(0, h_t).$$

6.1.3 The Unconditional Mean and Variance

The unconditional mean of u_t is

$$E(u_t) = E[E(u_t|F_{t-1})] = E(0) = 0. \tag{9-9}$$

While u_t is conditional heteroscedastic, the unconditional variance of u_t is

$$\begin{aligned} Var(u_t) &= Var[E(u_t|F_{t-1})] + E[Var(u_t|F_{t-1})] \\ &= 0 + \alpha_0 + \alpha_1 E(u_{t-1}^2) \\ &= \alpha_0 + \alpha_1 Var(u_{t-1}). \end{aligned}$$

If the process generating the disturbance u_t is weakly stationary, then

$$\begin{aligned} Var(u_t) &= Var(u_{t-1}) \\ &= \alpha_0 \alpha_1 Var(u_{t-1}) \\ &= \frac{\alpha_0}{1-\alpha_1}. \end{aligned} \tag{9-10}$$

For this ratio to be finite and positive, we require that $\alpha_0 > 0$ and $|\alpha_1| < 1$.

Moreover, since $E(u_t|F_{t-j}) = 0$, so

$$E(u_t u_{t-j}|F_{t-j}) = u_{t-j} E(u_t|F_{t-j}) = 0.$$

Hence

$$E(u_t u_{t-j}) = E[E(u_t u_{t-j}|F_{t-j})] = 0. \tag{9-11}$$

Based on (9-10)–(9-12), u_t follows ideal conditions.

6.2 Linear Regression Model With $ARCH(1)$ Disturbance

Suppose that we are interested in estimating the parameter of a regression model with $ARCH$ disturbances. Let the regression equation be

$$Y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t.$$

Here $\mathbf{x}_t$ denote a vector of predetermined explanatory variables, which could include lagged value of Y. The disturbance term u_t is assumed to satisfy (9-7) and (9-8).

Because as is shown in last section that u_t is satisfied the classical assumptions, the OLS estimator of $\boldsymbol{\beta}$ is most efficient *linear* estimator according to Gauss-Markov theorem.

But there is a more efficient *nonlinear* estimator. If $\varepsilon_t \overset{i.i.d.}{\sim} N(0,1)$, then conditioned on starting value, the sample conditional log likelihood function is then

$$l(\boldsymbol{\theta}) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{T}\frac{(Y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{\sigma_t^2}, \tag{9-12}$$

where

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 \\ &= \alpha_0 + \alpha_1(Y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta})^2,\end{aligned} \tag{9-13}$$

and the vector of parameters to be estimated $\boldsymbol{\theta} = (\boldsymbol{\beta}', \alpha_0, \alpha_1)'$. For a given numerical value for the parameter vector $\boldsymbol{\theta}$, the sequence of conditional variances can be calculated from (9-14) and used to evaluate the log likelihood function (9-13). This can then be maximized numerically using the methods described in Chapter 3.

NSYSU from a Sea Looking.

End of this Chapter