11
HETEROSCEDASTICITY

11.1 INTRODUCTION

Regression disturbances whose variances are not constant across observations are **heteroscedastic.** Heteroscedasticity arises in numerous applications, in both cross-section and time-series data. For example, even after accounting for firm sizes, we expect to observe greater variation in the profits of large firms than in those of small ones. The variance of profits might also depend on product diversification, research and development expenditure, and industry characteristics and therefore might also vary across firms of similar sizes. When analyzing family spending patterns, we find that there is greater variation in expenditure on certain commodity groups among high-income families than low ones due to the greater discretion allowed by higher incomes.[1]

In the heteroscedastic regression model,

$$\text{Var}[\varepsilon_i \mid \mathbf{x}_i] = \sigma_i^2, \quad i = 1, \ldots, n.$$

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

$$E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}] = \sigma^2 \boldsymbol{\Omega} = \sigma^2 \begin{bmatrix} \omega_1 & 0 & 0 & \cdots & 0 \\ 0 & \omega_2 & 0 & \cdots & \\ & & & \vdots & \\ 0 & 0 & 0 & \cdots & \omega_n \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & \\ & & & \vdots & \\ 0 & 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}.$$

It will sometimes prove useful to write $\sigma_i^2 = \sigma^2 \omega_i$. This form is an arbitrary scaling which allows us to use a normalization,

$$\text{tr}(\boldsymbol{\Omega}) = \sum_{i=1}^{n} \omega_i = n.$$

This makes the classical regression with homoscedastic disturbances a simple special case with $\omega_i = 1, i = 1, \ldots, n$. Intuitively, one might then think of the ωs as weights that are scaled in such a way as to reflect only the variety in the disturbance variances. The scale factor σ^2 then provides the overall scaling of the disturbance process.

Example 11.1 Heteroscedastic Regression

The data in Appendix Table F9.1 give monthly credit card expenditure for 100 individuals, sampled from a larger sample of 13,444 people. Linear regression of monthly expenditure on a constant, age, income and its square, and a dummy variable for home ownership using the 72 observations for which expenditure was nonzero produces the residuals plotted in Figure 11.1. The pattern of the residuals is characteristic of a regression with heteroscedasticity.

[1] Prais and Houthakker (1955).

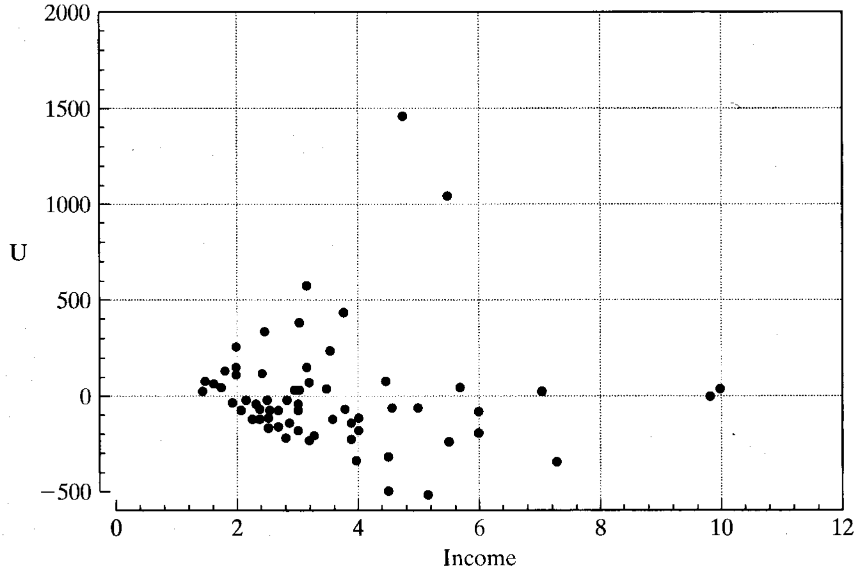

FIGURE 11.1 Plot of Residuals Against Income.

This chapter will present the heteroscedastic regression model, first in general terms, then with some specific forms of the disturbance covariance matrix. We begin by examining the consequences of heteroscedasticity for least squares estimation. We then consider **robust estimation,** in two frameworks. Section 11.2 presents appropriate estimators of the asymptotic covariance matrix of the least squares estimator. Section 11.3 discusses GMM estimation. Sections 11.4 to 11.7 present more specific formulations of the model. Sections 11.4 and 11.5 consider **generalized (weighted) least squares,** which requires knowledge at least of the form of $\mathbf{\Omega}$. Section 11.7 presents maximum likelihood estimators for two specific widely used models of heteroscedasticity. Recent analyses of financial data, such as exchange rates, the volatility of market returns, and inflation, have found abundant evidence of clustering of large and small disturbances,[2] which suggests a form of heteroscedasticity in which the variance of the disturbance depends on the size of the preceding disturbance. Engle (1982) suggested the **A**uto**R**egressive, **C**onditionally **H**eteroscedastic, or **ARCH,** model as an alternative to the standard time-series treatments. We will examine the ARCH model in Section 11.8.

11.2 ORDINARY LEAST SQUARES ESTIMATION

We showed in Section 10.2 that in the presence of heteroscedasticity, the least squares estimator **b** is still unbiased, consistent, and asymptotically normally distributed. The

[2]Pioneering studies in the analysis of macroeconomic data include Engle (1982, 1983) and Cragg (1982).

asymptotic covariance matrix is

$$\text{Asy. Var}[\mathbf{b}] = \frac{\sigma^2}{n}\left(\text{plim } \frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\left(\text{plim } \frac{1}{n}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}\right)\left(\text{plim } \frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}.$$

Estimation of the asymptotic covariance matrix would be based on

$$\text{Var}[\mathbf{b} \mid \mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\left(\sigma^2\sum_{i=1}^{n}\omega_i\mathbf{x}_i\mathbf{x}_i'\right)(\mathbf{X}'\mathbf{X})^{-1}.$$

[See (10-5).] Assuming, as usual, that the regressors are well behaved, so that $(\mathbf{X}'\mathbf{X}/n)^{-1}$ converges to a positive definite matrix, we find that the mean square consistency of **b** depends on the limiting behavior of the matrix:

$$\mathbf{Q}_n^* = \frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{n} = \frac{1}{n}\sum_{i=1}^{n}\omega_i\mathbf{x}_i\mathbf{x}_i'. \qquad \textbf{(11-1)}$$

If $\mathbf{Q}_n^*$ converges to a positive definite matrix $\mathbf{Q}^*$, then as $n \to \infty$, **b** will converge to $\boldsymbol{\beta}$ in mean square. Under most circumstances, if ω_i is finite for all i, then we would expect this result to be true. Note that $\mathbf{Q}_n^*$ is a weighted sum of the squares and cross products of **x** with weights ω_i/n, which sum to 1. We have already assumed that another weighted sum $\mathbf{X}'\mathbf{X}/n$, in which the weights are $1/n$, converges to a positive definite matrix **Q**, so it would be surprising if $\mathbf{Q}_n^*$ did not converge as well. In general, then, we would expect that

$$\mathbf{b} \overset{a}{\sim} N\left[\boldsymbol{\beta}, \frac{\sigma^2}{n}\mathbf{Q}^{-1}\mathbf{Q}^*\mathbf{Q}^{-1}\right], \quad \text{with } \mathbf{Q}^* = \text{plim } \mathbf{Q}_n^*.$$

A formal proof is based on Section 5.2 with $\mathbf{Q}_i = \omega_i\mathbf{x}_i\mathbf{x}_i'$.

11.2.1 INEFFICIENCY OF LEAST SQUARES

It follows from our earlier results that **b** is inefficient relative to the GLS estimator. By how much will depend on the setting, but there is some generality to the pattern. As might be expected, the greater is the dispersion in ω_i across observations, the greater the efficiency of GLS over OLS. The impact of this on the efficiency of estimation will depend crucially on the nature of the disturbance variances. In the usual cases, in which ω_i depends on variables that appear elsewhere in the model, the greater is the dispersion in these variables, the greater will be the gain to using GLS. It is important to note, however, that both these comparisons are based on knowledge of $\boldsymbol{\Omega}$. In practice, one of two cases is likely to be true. If we do have detailed knowledge of $\boldsymbol{\Omega}$, the performance of the inefficient estimator is a moot point. We will use GLS or feasible GLS anyway. In the more common case, we will not have detailed knowledge of $\boldsymbol{\Omega}$, so the comparison is not possible.

11.2.2 THE ESTIMATED COVARIANCE MATRIX OF b

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

The conventionally estimated covariance matrix for the least squares estimator $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ is inappropriate; the appropriate matrix is $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$. It is unlikely that these two would coincide, so the usual estimators of the standard errors are likely to be erroneous. In this section, we consider how erroneous the conventional estimator is likely to be.

As usual,

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{n-K} = \frac{\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}}{n-K}, \tag{11-2}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Expanding this equation, we obtain

$$s^2 = \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n-K} - \frac{\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}}{n-K}. \tag{11-3}$$

Taking the two parts separately yields

$$E\left[\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n-K}\,\middle|\,\mathbf{X}\right] = \frac{\operatorname{tr} E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}]}{n-K} = \frac{n\sigma^2}{n-K}. \tag{11-4}$$

[We have used the scaling $\operatorname{tr}(\mathbf{\Omega}) = n$.] In addition,

$$E\left[\frac{\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}}{n-K}\,\middle|\,\mathbf{X}\right] = \frac{\operatorname{tr}\{E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X} \mid \mathbf{X}]\}}{n-K}$$

$$= \frac{\operatorname{tr}\left[\sigma^2\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{\Omega}\mathbf{X}}{n}\right)\right]}{n-K} = \frac{\sigma^2}{n-K}\operatorname{tr}\left[\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\mathbf{Q}_n^*\right], \tag{11-5}$$

where $\mathbf{Q}_n^*$ is defined in (11-1). As $n \to \infty$, the term in (11-4) will converge to σ^2. The term in (11-5) will converge to zero if $\mathbf{b}$ is consistent because both matrices in the product are finite. Therefore:

If $\mathbf{b}$ is consistent, then $\lim_{n\to\infty} E[s^2] = \sigma^2$.

It can also be shown—we leave it as an exercise—that if the fourth moment of every disturbance is finite and all our other assumptions are met, then

$$\lim_{n\to\infty} \operatorname{Var}\left[\frac{\mathbf{e}'\mathbf{e}}{n-K}\right] = \lim_{n\to\infty} \operatorname{Var}\left[\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n-K}\right] = 0.$$

This result implies, therefore, that:

If $\operatorname{plim} \mathbf{b} = \boldsymbol{\beta}$, then $\operatorname{plim} s^2 = \sigma^2$.

Before proceeding, it is useful to pursue this result. The normalization $\operatorname{tr}(\mathbf{\Omega}) = n$ implies that

$$\sigma^2 = \bar{\sigma}^2 = \frac{1}{n}\sum_i \sigma_i^2 \quad \text{and} \quad \omega_i = \frac{\sigma_i^2}{\bar{\sigma}^2}.$$

Therefore, our previous convergence result implies that the least squares estimator s^2 converges to $\operatorname{plim} \bar{\sigma}^2$, that is, the probability limit of the average variance of the disturbances, *assuming that this probability limit exists*. Thus, some further assumption about these variances is necessary to obtain the result. (For an application, see Exercise 5 in Chapter 13.)

The difference between the conventional estimator and the appropriate (true) covariance matrix for **b** is

$$\text{Est. Var}[\mathbf{b}|\mathbf{X}] - \text{Var}[\mathbf{b}|\mathbf{X}] = s^2(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}. \tag{11-6}$$

In a large sample (so that $s^2 \approx \sigma^2$), this difference is approximately equal to

$$\mathbf{D} = \frac{\sigma^2}{n}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left[\frac{\mathbf{X}'\mathbf{X}}{n} - \frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{n}\right]\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}. \tag{11-7}$$

The difference between the two matrices hinges on

$$\boldsymbol{\Delta} = \frac{\mathbf{X}'\mathbf{X}}{n} - \frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{n} = \sum_{i=1}^{n}\left(\frac{1}{n}\right)\mathbf{x}_i\mathbf{x}_i' - \sum_{i=1}^{n}\left(\frac{\omega_i}{n}\right)\mathbf{x}_i\mathbf{x}_i' = \frac{1}{n}\sum_{i=1}^{n}(1-\omega_i)\mathbf{x}_i\mathbf{x}_i', \tag{11-8}$$

where $\mathbf{x}_i'$ is the ith row of $\mathbf{X}$. These are two weighted averages of the matrices $\mathbf{Q}_i = \mathbf{x}_i\mathbf{x}_i'$, using weights 1 for the first term and ω_i for the second. The scaling $\text{tr}(\boldsymbol{\Omega}) = n$ implies that $\sum_i(\omega_i/n) = 1$. Whether the weighted average based on ω_i/n differs much from the one using $1/n$ depends on the weights. If the weights are related to the values in $\mathbf{x}_i$, then the difference can be considerable. If the weights are uncorrelated with $\mathbf{x}_i\mathbf{x}_i'$, however, then the weighted average will tend to equal the unweighted average.[3]

Therefore, the comparison rests on whether the heteroscedasticity is related to any of x_k or $x_j \times x_k$. The conclusion is that, in general: *If the heteroscedasticity is not correlated with the variables in the model, then at least in large samples, the ordinary least squares computations, although not the optimal way to use the data, will not be misleading.* For example, in the groupwise heteroscedasticity model of Section 11.7.2, if the observations are grouped in the subsamples in a way that is unrelated to the variables in **X**, then the usual OLS estimator of Var[**b**] will, at least in large samples, provide a reliable estimate of the appropriate covariance matrix. It is worth remembering, however, that the least squares estimator will be inefficient, the more so the larger are the differences among the variances of the groups.[4]

The preceding is a useful result, but one should not be overly optimistic. First, it remains true that ordinary least squares is demonstrably inefficient. Second, if the primary assumption of the analysis—that the heteroscedasticity is unrelated to the variables in the model—is incorrect, then the conventional standard errors may be quite far from the appropriate values.

11.2.3 ESTIMATING THE APPROPRIATE COVARIANCE MATRIX FOR ORDINARY LEAST SQUARES

It is clear from the preceding that heteroscedasticity has some potentially serious implications for inferences based on the results of least squares. The application of more

[3]Suppose, for example, that **X** contains a single column and that both $\mathbf{x}_i$ and ω_i are independent and identically distributed random variables. Then $\mathbf{x}'\mathbf{x}/n$ converges to $E[x_i^2]$, whereas $\mathbf{x}'\boldsymbol{\Omega}\mathbf{x}/n$ converges to $\text{Cov}[\omega_i, x_i^2] + E[\omega_i]E[x_i^2]$. $E[\omega_i] = 1$, so if ω and x^2 are uncorrelated, then the sums have the same probability limit.

[4]Some general results, including analysis of the properties of the estimator based on estimated variances, are given in Taylor (1977).

appropriate estimation techniques requires a detailed formulation of $\boldsymbol{\Omega}$, however. It may well be that the form of the heteroscedasticity is unknown. White (1980) has shown that it is still possible to obtain an appropriate estimator for the variance of the least squares estimator, even if the heteroscedasticity is related to the variables in $\mathbf{X}$. The **White estimator** [see (10-14) in Section 10.3[5]]

$$\text{Est. Asy. Var}[\mathbf{b}] = \frac{1}{n}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2\mathbf{x}_i\mathbf{x}_i'\right)\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}, \tag{11-9}$$

where e_i is the ith least squares residual, can be used as an estimate of the asymptotic variance of the least squares estimator.

A number of studies have sought to improve on the White estimator for OLS.[6] The asymptotic properties of the estimator are unambiguous, but its usefulness in small samples is open to question. The possible problems stem from the general result that the squared OLS residuals tend to underestimate the squares of the true disturbances. [That is why we use $1/(n-K)$ rather than $1/n$ in computing s^2.] The end result is that in small samples, at least as suggested by some Monte Carlo studies [e.g., MacKinnon and White (1985)], the White estimator is a bit too optimistic; the matrix is a bit too small, so asymptotic t ratios are a little too large. Davidson and MacKinnon (1993, p. 554) suggest a number of fixes, which include (1) scaling up the end result by a factor $n/(n-K)$ and (2) using the squared residual scaled by its true variance, e_i^2/m_{ii}, instead of e_i^2, where $m_{ii} = 1 - \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$.[7] [See (4-20).] On the basis of their study, Davidson and MacKinnon strongly advocate one or the other correction. Their admonition "One should *never* use [the White estimator] because [(2)] *always* performs better" seems a bit strong, but the point is well taken. The use of sharp asymptotic results in small samples can be problematic. The last two rows of Table 11.1 show the recomputed standard errors with these two modifications.

Example 11.2 The White Estimator

Using White's estimator for the regression in Example 11.1 produces the results in the row labeled "White S. E." in Table 11.1. The two income coefficients are individually and jointly statistically significant based on the individual t ratios and $F(2, 67) = [(0.244-0.064)/2]/[0.776/(72-5)] = 7.771$. The 1 percent critical value is 4.94.

The differences in the estimated standard errors seem fairly minor given the extreme heteroscedasticity. One surprise is the decline in the standard error of the age coefficient. The F test is no longer available for testing the joint significance of the two income coefficients because it relies on homoscedasticity. A Wald test, however, may be used in any event. The chi-squared test is based on

$$W = (\mathbf{Rb})'\left[\mathbf{R}\left(\text{Est. Asy. Var}[\mathbf{b}]\right)\mathbf{R}'\right]^{-1}(\mathbf{Rb}) \quad \text{where } \mathbf{R} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

and the estimated asymptotic covariance matrix is the White estimator. The F statistic based on least squares is 7.771. The Wald statistic based on the White estimator is 20.604; the 95 percent critical value for the chi-squared distribution with two degrees of freedom is 5.99, so the conclusion is unchanged.

[5]See also Eicker (1967), Horn, Horn, and Duncan (1975), and MacKinnon and White (1985).

[6]See, e.g., MacKinnon and White (1985) and Messer and White (1984).

[7]They also suggest a third correction, e_i^2/m_{ii}^2, as an approximation to an estimator based on the "jackknife" technique, but their advocacy of this estimator is much weaker than that of the other two.

TABLE 11.1 Least Squares Regression Results

	Constant	*Age*	*OwnRent*	*Income*	*Income²*
Sample Mean		32.08	0.36	3.369	
Coefficient	−237.15	−3.0818	27.941	234.35	−14.997
Standard Error	199.35	5.5147	82.922	80.366	7.4693
t ratio	−1.10	−0.5590	0.337	2.916	−2.008
White S.E.	212.99	3.3017	92.188	88.866	6.9446
D. and M. (1)	270.79	3.4227	95.566	92.122	7.1991
D. and M. (2)	221.09	3.4477	95.632	92.083	7.1995

$R^2 = 0.243578,\ s = 284.75080$

Mean Expenditure = \$189.02. Income is × \$10,000
Tests for Heteroscedasticity: White = 14.329, Goldfeld–Quandt = 15.001,
Breusch–Pagan = 41.920, Koenker–Bassett = 6.187.
(Two degrees of freedom. $\chi^2_* = 5.99$.)

11.3 GMM ESTIMATION OF THE HETEROSCEDASTIC REGRESSION MODEL

The **GMM estimator** in the heteroscedastic regression model is produced by the empirical moment equations

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\left(y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_{GMM}\right) = \frac{1}{n}\mathbf{X}'\hat{\boldsymbol{\varepsilon}}\left(\hat{\boldsymbol{\beta}}_{GMM}\right) = \bar{\mathbf{m}}\left(\hat{\boldsymbol{\beta}}_{GMM}\right) = \mathbf{0}. \tag{11-10}$$

The estimator is obtained by minimizing

$$q = \bar{\mathbf{m}}'\left(\hat{\boldsymbol{\beta}}_{GMM}\right)\mathbf{W}\bar{\mathbf{m}}\left(\hat{\boldsymbol{\beta}}_{GMM}\right)$$

where $\mathbf{W}$ is a positive definite weighting matrix. The optimal weighting matrix would be

$$\mathbf{W} = \left\{\text{Asy. Var}[\sqrt{n}\,\bar{\mathbf{m}}(\boldsymbol{\beta})]\right\}^{-1}$$

which is the inverse of

$$\text{Asy. Var}[\sqrt{n}\bar{\mathbf{m}}(\boldsymbol{\beta})] = \text{Asy. Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i\right] = \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sigma^2\omega_i\mathbf{x}_i\mathbf{x}_i' = \sigma^2\mathbf{Q}^*$$

[see (11-1)]. The optimal weighting matrix would be $[\sigma^2\mathbf{Q}^*]^{-1}$. But, recall that this minimization problem is an exactly identified case, so, the weighting matrix is irrelevant to the solution. You can see that in the moment equation—that equation is simply the normal equations for least squares. We can solve the moment equations exactly, so there is no need for the weighting matrix. *Regardless of the covariance matrix of the moments, the GMM estimator for the heteroscedastic regression model is ordinary least squares.* (This is Case 2 analyzed in Section 10.4.) We can use the results we have already obtained to find its asymptotic covariance matrix. The result appears in Section 11.2. The implied estimator is the White estimator in (11-9). [Once again, see Theorem 10.6.] The conclusion to be drawn at this point is that until we make some specific assumptions about the variances, we do not have a more efficient estimator than least squares, but we do have to modify the estimated asymptotic covariance matrix.

11.4 TESTING FOR HETEROSCEDASTICITY

Heteroscedasticity poses potentially severe problems for inferences based on least squares. 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 procedures accordingly.[8] Several types of tests have been suggested. They can be roughly grouped in descending order in terms of their generality and, as might be expected, in ascending order in terms of their power.[9]

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

11.4.1 WHITE'S GENERAL TEST

To formulate most of the available tests, it is necessary to specify, at least in rough terms, the nature of the heteroscedasticity. It would be desirable to be able to test a general hypothesis of the form

$$H_0 : \sigma_i^2 = \sigma^2 \quad \text{for all } i,$$
$$H_1 : \text{Not } H_0.$$

In view of our earlier findings on the difficulty of estimation in a model with n unknown parameters, this is rather ambitious. Nonetheless, such a test has been devised by White (1980b). The correct covariance matrix for the least squares estimator is

$$\text{Var}[\mathbf{b}|\mathbf{X}] = \sigma^2[\mathbf{X}'\mathbf{X}]^{-1}[\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}][\mathbf{X}'\mathbf{X}]^{-1}, \tag{11-11}$$

which, as we have seen, can be estimated using (11-9). The conventional estimator is $\mathbf{V} = s^2[\mathbf{X}'\mathbf{X}]^{-1}$. If there is no heteroscedasticity, then $\mathbf{V}$ will give a consistent estimator of $\text{Var}[\mathbf{b}|\mathbf{X}]$, whereas if there is, then it will not. White has devised a statistical test based on this observation. A simple operational version of his test is carried out by obtaining nR^2 in the regression of e_i^2 on a constant and all unique variables contained in $\mathbf{x}$ and all the squares and cross products of the variables in $\mathbf{x}$. The statistic is asymptotically distributed as chi-squared with $P - 1$ degrees of freedom, where P is the number of regressors in the equation, including the constant.

The **White test** is extremely general. To carry it out, we need not make any specific assumptions about the nature of the heteroscedasticity. Although this characteristic is a virtue, it is, at the same time, a potentially serious shortcoming. The test may reveal

[8]There is the possibility that a preliminary test for heteroscedasticity will incorrectly lead us to use weighted least squares or fail to alert us to heteroscedasticity and lead us improperly to use ordinary least squares. Some limited results on the properties of the resulting estimator are given by Ohtani and Toyoda (1980). Their results suggest that it is best to test first for heteroscedasticity rather than merely to assume that it is present.

[9]A study that examines the power of several tests for heteroscedasticity is Ali and Giaccotto (1984).

heteroscedasticity, but it may instead simply identify some other specification error (such as the omission of x^2 from a simple regression).[10] Except in the context of a specific problem, little can be said about the power of White's test; it may be very low against some alternatives. In addition, unlike some of the other tests we shall discuss, the White test is **nonconstructive.** If we reject the null hypothesis, then the result of the test gives no indication of what to do next.

11.4.2 THE GOLDFELD–QUANDT TEST

By narrowing our focus somewhat, we can obtain a more powerful test. Two tests that are relatively general are the **Goldfeld–Quandt (1965) test** and the Breusch–Pagan (1979) **Lagrange multiplier test.**

For the Goldfeld–Quandt 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 most favorable case for this would be the **groupwise heteroscedastic** model of Section 11.7.2 and Example 11.7 or a model such as $\sigma_i^2 = \sigma^2 x_i^2$ for some variable x. By ranking the observations based on this x, we can separate the observations into those with high and low variances. The test is applied by dividing the sample into two groups with n_1 and n_2 observations. To obtain statistically independent variance estimators, the regression is then estimated separately with the two sets of observations. The test statistic is

$$F[n_1 - K, n_2 - K] = \frac{\mathbf{e}_1'\mathbf{e}_1/(n_1 - K)}{\mathbf{e}_2'\mathbf{e}_2/(n_2 - K)}, \tag{11-12}$$

where we assume that the disturbance variance is larger in the first sample. (If not, then reverse the subscripts.) Under the null hypothesis of homoscedasticity, this statistic has an F distribution with $n_1 - K$ and $n_2 - K$ degrees of freedom. The sample value can be referred to the standard F table to carry out the test, with a large value leading to rejection of the null hypothesis.

To increase the power of the test, Goldfeld and Quandt suggest that a number of observations in the middle of the sample be omitted. The more observations that are dropped, however, the smaller the degrees of freedom for estimation in each group will be, which will tend to diminish the power of the test. As a consequence, the choice of how many central observations to drop is largely subjective. Evidence by Harvey and Phillips (1974) suggests that no more than a third of the observations should be dropped. If the disturbances are normally distributed, then the Goldfeld–Quandt statistic is exactly distributed as F under the null hypothesis and the nominal size of the test is correct. If not, then the F distribution is only approximate and some alternative method with known large-sample properties, such as White's test, might be preferable.

11.4.3 THE BREUSCH–PAGAN/GODFREY LM 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. For example, several of the models we will consider allow

[10] Thursby (1982) considers this issue in detail.

the disturbance variance to vary with a set of regressors. Breusch and Pagan[11] have devised a **Lagrange multiplier test** of the hypothesis that $\sigma_i^2 = \sigma^2 f(\alpha_0 + \boldsymbol{\alpha}'\mathbf{z}_i)$, where $\mathbf{z}_i$ is a vector of independent variables.[12] The model is homoscedastic if $\boldsymbol{\alpha} = \mathbf{0}$. The test can be carried out with a simple regression:

$$\text{LM} = \tfrac{1}{2} \text{ explained sum of squares in the regression of } e_i^2/(\mathbf{e}'\mathbf{e}/n) \text{ on } \mathbf{z}_i.$$

For computational purposes, let $\mathbf{Z}$ be the $n \times P$ matrix of observations on $(1, \mathbf{z}_i)$, and let $\mathbf{g}$ be the vector of observations of $g_i = e_i^2/(\mathbf{e}'\mathbf{e}/n) - 1$. Then

$$\text{LM} = \tfrac{1}{2}[\mathbf{g}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{g}].$$

Under the null hypothesis of homoscedasticity, LM has a limiting chi-squared distribution with degrees of freedom equal to the number of variables in $\mathbf{z}_i$. This test can be applied to a variety of models, including, for example, those examined in Example 11.3 (3) and in Section 11.7.[13]

It has been argued that the **Breusch–Pagan Lagrange multiplier test** is sensitive to the assumption of normality. Koenker (1981) and Koenker and Bassett (1982) suggest that the computation of LM be based on a more robust estimator of the variance of ε_i^2,

$$V = \frac{1}{n}\sum_{i=1}^{n}\left[e_i^2 - \frac{\mathbf{e}'\mathbf{e}}{n}\right]^2.$$

The variance of ε_i^2 is not necessarily equal to $2\sigma^4$ if ε_i is not normally distributed. Let $\mathbf{u}$ equal $(e_1^2, e_2^2, \ldots, e_n^2)$ and $\mathbf{i}$ be an $n \times 1$ column of 1s. Then $\bar{u} = \mathbf{e}'\mathbf{e}/n$. With this change, the computation becomes

$$\text{LM} = \left[\frac{1}{V}\right](\mathbf{u} - \bar{u}\,\mathbf{i})'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{u} - \bar{u}\,\mathbf{i}).$$

Under normality, this modified statistic will have the same asymptotic distribution as the Breusch–Pagan statistic, but absent normality, there is some evidence that it provides a more powerful test. Waldman (1983) has shown that if the variables in $\mathbf{z}_i$ are the same as those used for the White test described earlier, then the two tests are algebraically the same.

Example 11.3 Testing for Heteroscedasticity

1. White's Test: For the data used in Example 11.1, there are 15 variables in $\mathbf{x} \otimes \mathbf{x}$ including the constant term. But since Ownrent^2 = OwnRent and Income × Income = Income^2, only 13 are unique. Regression of the squared least squares residuals on these 13 variables produces $R^2 = 0.199013$. The chi-squared statistic is therefore 72(0.199013) = 14.329. The 95 percent critical value of chi-squared with 12 degrees of freedom is 21.03, so despite what might seem to be obvious in Figure 11.1, the hypothesis of homoscedasticity is not rejected by this test.

2. Goldfeld–Quandt Test: The 72 observations are sorted by Income, and then the regression is computed with the first 36 observations and the second. The two sums of squares are 326,427 and 4,894,130, so the test statistic is $F[31, 31] = 4{,}894{,}130/326{,}427 = 15.001$. The critical value from this table is 1.79, so this test reaches the opposite conclusion.

[11]Breusch and Pagan (1979).

[12]Lagrange multiplier tests are discussed in Section 17.5.3.

[13]The model $\sigma_i^2 = \sigma^2 \exp(\boldsymbol{\alpha}'\mathbf{z}_i)$ is one of these cases. In analyzing this model specifically, Harvey (1976) derived the same test statistic.

3. Breusch–Pagan Test: This test requires a specific alternative hypothesis. For this purpose, we specify the test based on $\mathbf{z} = [1, \text{Income}, \text{IncomeSq}]$. Using the least squares residuals, we compute $g_i = e_i^2/(\mathbf{e}'\mathbf{e}/72) - 1$; then $\text{LM} = \frac{1}{2}\mathbf{g}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{g}$. The sum of squares is 5,432,562.033. The computation produces LM = 41.920. The critical value for the chi-squared distribution with two degrees of freedom is 5.99, so the hypothesis of homoscedasticity is rejected. The Koenker and Bassett variant of this statistic is only 6.187, which is still significant but much smaller than the LM statistic. The wide difference between these two statistics suggests that the assumption of normality is erroneous. Absent any knowledge of the heteroscedasticity, we might use the Bera and Jarque (1981, 1982) and Kiefer and Salmon (1983) test for normality,

$$\chi^2[2] = n[(m_3/s^3)^2 + ((m_4 - 3)/s^4)^2]$$

where $m_j = (1/n)\sum_i e_i^j$. Under the null hypothesis of *homoscedastic* and normally distributed disturbances, this statistic has a limiting chi-squared distribution with two degrees of freedom. Based on the least squares residuals, the value is 482.12, which certainly does lead to rejection of the hypothesis. Some caution is warranted here, however. It is unclear what part of the hypothesis should be rejected. We have convincing evidence in Figure 11.1 that the disturbances are heteroscedastic, so the assumption of homoscedasticity underlying this test is questionable. This does suggest the need to examine the data before applying a specification test such as this one.

11.5 WEIGHTED LEAST SQUARES WHEN Ω IS KNOWN

Having tested for and found evidence of heteroscedasticity, the logical next step is to revise the estimation technique to account for it. The GLS estimator is

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

Consider the most general case, $\text{Var}[\varepsilon_i \mid \mathbf{x}_i] = \sigma_i^2 = \sigma^2\omega_i$. Then $\boldsymbol{\Omega}^{-1}$ is a diagonal matrix whose ith diagonal element is $1/\omega_i$. The GLS estimator is obtained by regressing

$$\mathbf{Py} = \begin{bmatrix} y_1/\sqrt{\omega_1} \\ y_2/\sqrt{\omega_2} \\ \vdots \\ y_n/\sqrt{\omega_n} \end{bmatrix} \quad \text{on} \quad \mathbf{PX} = \begin{bmatrix} \mathbf{x}_1/\sqrt{\omega_1} \\ \mathbf{x}_2/\sqrt{\omega_2} \\ \vdots \\ \mathbf{x}_n/\sqrt{\omega_n} \end{bmatrix}.$$

Applying ordinary least squares to the transformed model, we obtain the **weighted least squares (WLS)** estimator.

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n} w_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} \left[\sum_{i=1}^{n} w_i \mathbf{x}_i y_i\right], \qquad \textbf{(11-13)}$$

where $w_i = 1/\omega_i$.[14] The logic of the computation is that observations with smaller variances receive a larger weight in the computations of the sums and therefore have greater influence in the estimates obtained.

[14]The weights are often denoted $w_i = 1/\sigma_i^2$. This expression is consistent with the equivalent $\hat{\boldsymbol{\beta}} = [\mathbf{X}'(\sigma^2\boldsymbol{\Omega})^{-1}\mathbf{X}]^{-1}\mathbf{X}'(\sigma^2\boldsymbol{\Omega})^{-1}\mathbf{y}$. The σ^2's cancel, leaving the expression given previously.

A common specification is that the variance is proportional to one of the regressors or its square. Our earlier example of family expenditures is one in which the relevant variable is usually income. Similarly, in studies of firm profits, the dominant variable is typically assumed to be firm size. If

$$\sigma_i^2 = \sigma^2 x_{ik}^2,$$

then the transformed regression model for GLS is

$$\frac{y}{x_k} = \beta_k + \beta_1\left(\frac{x_1}{x_k}\right) + \beta_2\left(\frac{x_2}{x_k}\right) + \cdots + \frac{\varepsilon}{x_k}. \tag{11-14}$$

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

In (11-14), the coefficient on x_k becomes the constant term. But if the variance is proportional to any power of x_k other than two, then the transformed model will no longer contain a constant, and we encounter the problem of interpreting R^2 mentioned earlier. For example, no conclusion should be drawn if the R^2 in the regression of y/z on $1/z$ and x/z is higher than in the regression of y on a constant and x for any z, including x. The good fit of the weighted regression might be due to the presence of $1/z$ on both sides of the equality.

It is rarely possible to be certain about the nature of the heteroscedasticity in a regression model. In one respect, this problem is only minor. The weighted least squares estimator

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n} w_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{n} w_i \mathbf{x}_i y_i\right]$$

is consistent regardless of the weights used, as long as the weights are uncorrelated with the disturbances.

But using the wrong set of weights has two other consequences that may be less benign. First, the improperly weighted least squares estimator is inefficient. This point might be moot if the correct weights are unknown, but the GLS standard errors will also be incorrect. The asymptotic covariance matrix of the estimator

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{11-15}$$

is

$$\text{Asy. Var}[\hat{\boldsymbol{\beta}}] = \sigma^2[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{\Omega}\mathbf{V}^{-1}\mathbf{X}[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}. \tag{11-16}$$

This result may or may not resemble the usual estimator, which would be the matrix in brackets, and underscores the usefulness of the White estimator in (11-9).

The standard approach in the literature is to use OLS with the White estimator or some variant for the asymptotic covariance matrix. One could argue both flaws and virtues in this approach. In its favor, **robustness to unknown heteroscedasticity** is a compelling virtue. In the clear presence of heteroscedasticity, however, least squares can be extremely inefficient. The question becomes whether using the wrong weights is better than using no weights at all. There are several layers to the question. If we use one of the models discussed earlier—Harvey's, for example, is a versatile and flexible candidate—then we may use the wrong set of weights and, in addition, estimation of the variance parameters introduces a new source of variation into the slope estimators for the model. A heteroscedasticity robust estimator for weighted least squares can be formed by combining (11-16) with the White estimator. The weighted least squares estimator in (11-15) is consistent with any set of weights $\mathbf{V} = \text{diag}[v_1, v_2, \ldots, v_n]$. Its asymptotic covariance matrix can be estimated with

$$\text{Est.Asy. Var}[\hat{\boldsymbol{\beta}}] = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\left[\sum_{i=1}^{n}\left(\frac{e_i^2}{v_i^2}\right)\mathbf{x}_i\mathbf{x}_i'\right](\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{11-17}$$

Any consistent estimator can be used to form the residuals. The weighted least squares estimator is a natural candidate.

11.6 ESTIMATION WHEN Ω CONTAINS UNKNOWN PARAMETERS

The general form of the heteroscedastic regression model has too many parameters to estimate by ordinary methods. Typically, the model is restricted by formulating $\sigma^2\boldsymbol{\Omega}$ as a function of a few parameters, as in $\sigma_i^2 = \sigma^2 x_i^\alpha$ or $\sigma_i^2 = \sigma^2[\mathbf{x}_i'\boldsymbol{\alpha}]^2$. Write this as $\boldsymbol{\Omega}(\boldsymbol{\alpha})$. FGLS based on a consistent estimator of $\boldsymbol{\Omega}(\boldsymbol{\alpha})$ (meaning a consistent estimator of $\boldsymbol{\alpha}$) is asymptotically equivalent to full GLS, and FGLS based on a maximum likelihood estimator of $\boldsymbol{\Omega}(\boldsymbol{\alpha})$ will produce a maximum likelihood estimator of $\boldsymbol{\beta}$ if $\boldsymbol{\Omega}(\boldsymbol{\alpha})$ does not contain any elements of $\boldsymbol{\beta}$. The new problem is that we must first find consistent estimators of the unknown parameters in $\boldsymbol{\Omega}(\boldsymbol{\alpha})$. Two methods are typically used, two-step GLS and maximum likelihood.

11.6.1 TWO-STEP ESTIMATION

For the heteroscedastic model, the GLS estimator is

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n}\left(\frac{1}{\sigma_i^2}\right)\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{n}\left(\frac{1}{\sigma_i^2}\right)\mathbf{x}_i y_i\right]. \tag{11-18}$$

The **two-step estimators** are computed by first obtaining estimates $\hat{\sigma}_i^2$, usually using some function of the ordinary least squares residuals. Then, $\hat{\hat{\boldsymbol{\beta}}}$ uses (11-18) and $\hat{\sigma}_i^2$. The ordinary least squares estimator of $\boldsymbol{\beta}$, although inefficient, is still consistent. As such, statistics computed using the ordinary least squares residuals, $e_i = (y_i - \mathbf{x}_i'\mathbf{b})$, will have the same asymptotic properties as those computed using the true disturbances, $\varepsilon_i = (y_i - \mathbf{x}_i'\boldsymbol{\beta})$. This result suggests a regression approach for the true disturbances and variables $\mathbf{z}_i$ that may or may not coincide with $\mathbf{x}_i$. Now $E[\varepsilon_i^2 \mid \mathbf{z}_i] = \sigma_i^2$, so

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

where v_i is just the difference between ε_i^2 and its conditional expectation. Since ε_i is unobservable, we would use the least squares residual, for which $e_i = \varepsilon_i - \mathbf{x}_i'(\mathbf{b} - \boldsymbol{\beta}) = \varepsilon_i + u_i$. Then, $e_i^2 = \varepsilon_i^2 + u_i^2 + 2\varepsilon_i u_i$. But, in large samples, as $\mathbf{b} \xrightarrow{p} \boldsymbol{\beta}$, terms in u_i will become negligible, so that at least approximately,[15]

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

The procedure suggested is to treat the variance function as a regression and use the squares or some other functions of the least squares residuals as the dependent variable.[16] For example, if $\sigma_i^2 = \mathbf{z}_i'\alpha$, then a consistent estimator of $\boldsymbol{\alpha}$ will be the least squares slopes, **a**, in the "model,"

$$e_i^2 = \mathbf{z}_i'\boldsymbol{\alpha} + v_i^*.$$

In this model, v_i^* is both heteroscedastic and autocorrelated, so **a** is consistent but inefficient. But, consistency is all that is required for asymptotically efficient estimation of $\boldsymbol{\beta}$ using $\boldsymbol{\Omega}(\hat{\boldsymbol{\alpha}})$. It remains to be settled whether improving the estimator of $\boldsymbol{\alpha}$ in this and the other models we will consider would improve the small sample properties of the two-step estimator of $\boldsymbol{\beta}$.[17]

The two-step estimator may be iterated by recomputing the residuals after computing the FGLS estimates and then reentering the computation. The asymptotic properties of the iterated estimator are the same as those of the two-step estimator, however. In some cases, this sort of iteration will produce the maximum likelihood estimator at convergence. Yet none of the estimators based on regression of squared residuals on other variables satisfy the requirement. Thus, iteration in this context provides little additional benefit, if any.

11.6.2 MAXIMUM LIKELIHOOD ESTIMATION[18]

The log-likelihood function for a sample of normally distributed observations 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 **(11-19)**

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

where $\boldsymbol{\alpha}$ is the vector of unknown parameters in $\boldsymbol{\Omega}(\boldsymbol{\alpha})$ and $f_i(\boldsymbol{\alpha})$ is indexed by i to indicate that it is a function of $\mathbf{z}_i$—note that $\boldsymbol{\Omega}(\boldsymbol{\alpha}) = \text{diag}[f_i(\boldsymbol{\alpha})]$ so it is also. Assume as well that no elements of $\boldsymbol{\beta}$ appear in $\boldsymbol{\alpha}$. The log-likelihood function is

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

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

[15] See Amemiya (1985) for formal analysis.

[16] See, for example, Jobson and Fuller (1980).

[17] Fomby, Hill, and Johnson (1984, pp. 177–186) and Amemiya (1985, pp. 203–207; 1977a) examine this model.

[18] The method of maximum likelihood estimation is developed in Chapter 17.

log-likelihood function are

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \mathbf{x}_i \frac{\varepsilon_i}{\sigma^2 f_i}$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\frac{\varepsilon_i^2}{f_i} = \sum_{i=1}^{n}\left(\frac{1}{2\sigma^2}\right)\left(\frac{\varepsilon_i^2}{\sigma^2 f_i} - 1\right) \tag{11-20}$$

$$\frac{\partial \ln L}{\partial \boldsymbol{\alpha}} = \sum_{i=1}^{n}\left(\frac{1}{2}\right)\left(\frac{\varepsilon_i^2}{\sigma^2 f_i} - 1\right)\left(\frac{1}{f_i}\right)\mathbf{g}_i.$$

Since $E[\varepsilon_i \,|\, \mathbf{x}_i, \mathbf{z}_i] = 0$ and $E[\varepsilon_i^2 \,|\, \mathbf{x}_i, \mathbf{z}_i] = \sigma^2 f_i$, it is clear that all derivatives have expectation zero as required. The **maximum likelihood estimators** are those values of $\boldsymbol{\beta}, \sigma^2$, and $\boldsymbol{\alpha}$ that simultaneously equate these derivatives to zero. The likelihood equations are generally highly nonlinear and will usually require an iterative solution.

Let $\mathbf{G}$ be the $n \times M$ matrix with ith row equal to $\partial f_i/\partial \boldsymbol{\alpha}' = \mathbf{g}_i'$ and let $\mathbf{i}$ denote an $n \times 1$ column vector of 1s. The asymptotic covariance matrix for the maximum likelihood estimator in this model is

$$\left(-E\left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma}\,\partial \boldsymbol{\gamma}'}\right]\right)^{-1} = \begin{bmatrix} (1/\sigma^2)\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & n/(2\sigma^4) & (1/(2\sigma^2))\mathbf{i}'\boldsymbol{\Omega}^{-1}\mathbf{G} \\ \mathbf{0}' & (1/(2\sigma^2))\mathbf{G}'\boldsymbol{\Omega}^{-1}\mathbf{i} & (1/2)\mathbf{G}'\boldsymbol{\Omega}^{-2}\mathbf{G} \end{bmatrix}^{-1}, \tag{11-21}$$

where $\boldsymbol{\gamma}' = [\boldsymbol{\beta}', \sigma^2, \boldsymbol{\alpha}']$. (One convenience is that terms involving $\partial^2 f_i/\partial \boldsymbol{\alpha}\,\partial \boldsymbol{\alpha}'$ fall out of the expectations. The proof is considered in the exercises.)

From the likelihood equations, it is apparent that for a given value of $\boldsymbol{\alpha}$, the solution for $\boldsymbol{\beta}$ is the GLS estimator. The scale parameter, σ^2, is ultimately irrelevant to this solution. The second likelihood equation shows that for given values of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, σ^2 will be estimated as the mean of the squared generalized residuals, $\hat{\sigma}^2 = (1/n)\sum_{i=1}^{n}[(y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}})/\hat{f}_i]^2$. This term is the **generalized sum of squares.** Finally, there is no general solution to be found for the estimator of $\boldsymbol{\alpha}$; it depends on the model. We will examine two examples. If $\boldsymbol{\alpha}$ is only a single parameter, then it may be simplest just to scan a range of values of α to locate the one that, with the associated FGLS estimator of $\boldsymbol{\beta}$, maximizes the log-likelihood. The fact that the Hessian is block diagonal does provide an additional convenience. The parameter vector $\boldsymbol{\beta}$ may always be estimated conditionally on $[\sigma^2, \boldsymbol{\alpha}]$ and, likewise, if $\boldsymbol{\beta}$ is given, then the solutions for σ^2 and $\boldsymbol{\alpha}$ can be found conditionally, although this may be a complicated optimization problem. But, by going back and forth in this fashion, as suggested by Oberhofer and Kmenta (1974), we may be able to obtain the full solution more easily than by approaching the full set of equations simultaneously.

11.6.3 MODEL BASED TESTS FOR HETEROSCEDASTICITY

The tests for heteroscedasticity described in Section 11.4 are based on the behavior of the least squares residuals. The general approach is based on the idea that if heteroscedasticity of any form is present in the disturbances, it will be discernible in the behavior of the residuals. Those **residual based tests** are robust in the sense that they will detect heteroscedasticity of a variety of forms. On the other hand, their power is a function of the specific alternative. The model considered here is fairly narrow. The tradeoff is that within the context of the specified model, a test of heteroscedasticity will have greater power than the residual based tests. (To come full circle, of course, that means that if the model specification is incorrect, the tests are likely to have limited or no power at all to reveal an incorrect hypothesis of homoscedasticity.)

Testing the hypothesis of homoscedasticity using any of the three standard methods is particularly simple in the model outlined in this section. The trio of tests for parametric models is available. The model would generally be formulated so that the heteroscedasticity is induced by a nonzero $\boldsymbol{\alpha}$. Thus, we take the test of $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ to be a test against homoscedasticity.

Wald Test The Wald statistic is computed by extracting from the full parameter vector and its estimated asymptotic covariance matrix the subvector $\hat{\boldsymbol{\alpha}}$ and its asymptotic covariance matrix. Then,

$$W = \hat{\boldsymbol{\alpha}}'\left\{\text{Est.Asy. Var}[\hat{\boldsymbol{\alpha}}]\right\}^{-1}\hat{\boldsymbol{\alpha}}.$$

Likelihood Ratio Test The results of the homoscedastic least squares regression are generally used to obtain the initial values for the iterations. The restricted log-likelihood value is a by-product of the initial setup; log-$L_R = -(n/2)[1 + \ln 2\pi + \ln(\mathbf{e}'\mathbf{e}/n)]$. The unrestricted log-likelihood, log-L_U, is obtained as the objective function for the estimation. Then, the statistic for the test is

$$\text{LR} = -2(\ln\text{-}L_R - \ln\text{-}L_U].$$

Lagrange Multiplier Test To set up the LM test, we refer back to the model in (11-19)–(11-21). At the restricted estimates $\boldsymbol{\alpha} = \mathbf{0}$, $\boldsymbol{\beta} = \mathbf{b}$, $\sigma^2 = \mathbf{e}'\mathbf{e}/n$ (not $n - K$), $f_i = 1$ and $\boldsymbol{\Omega}(\mathbf{0}) = \mathbf{I}$. Thus, the first derivatives vector evaluated at the least squares estimates is

$$\left.\frac{\partial \ln L}{\partial \boldsymbol{\beta}}\right| (\boldsymbol{\beta} = \mathbf{b}, \sigma^2 = \mathbf{e}'\mathbf{e}/n, \hat{\boldsymbol{\alpha}} = \mathbf{0}) = \mathbf{0}$$

$$\left.\frac{\partial \ln L}{\partial \sigma^2}\right| (\boldsymbol{\beta} = \mathbf{b}, \sigma^2 = \mathbf{e}'\mathbf{e}/n, \hat{\boldsymbol{\alpha}} = \mathbf{0}) = 0$$

$$\left.\frac{\partial \ln L}{\partial \boldsymbol{\alpha}}\right| (\boldsymbol{\beta} = \mathbf{b}, \sigma^2 = \mathbf{e}'\mathbf{e}/n, \hat{\boldsymbol{\alpha}} = \mathbf{0}) = \sum_{i=1}^{n} \frac{1}{2}\left(\frac{e_i^2}{\mathbf{e}'\mathbf{e}/n} - 1\right)\mathbf{g}_i = \sum_{i=1}^{n} \frac{1}{2} v_i \mathbf{g}_i.$$

The negative expected inverse of the Hessian, from (11-21) is

$$\left(-E\left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma}\,\partial \boldsymbol{\gamma}'}\right]_{\boldsymbol{\alpha}=\mathbf{0}}\right)^{-1} = \begin{bmatrix} (1/\sigma^2)\mathbf{X}'\mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & n/(2\sigma^4) & [1/(2\sigma^2)]\mathbf{g} \\ \mathbf{0}' & [1/(2\sigma^2)]\mathbf{g}' & (1/2)\mathbf{G}'\mathbf{G} \end{bmatrix}^{-1} = \left\{-E[\mathbf{H}]\right\}^{-1}$$

where $\mathbf{g} = \sum_{i=1}^{n} \mathbf{g}_i$ and $\mathbf{G}'\mathbf{G} = \sum_{i=1}^{n} \mathbf{g}_i\mathbf{g}_i'$. The LM statistic will be

$$\text{LM} = \left[\left.\frac{\partial \ln L}{\partial \boldsymbol{\gamma}}\right| (\boldsymbol{\gamma} = \mathbf{b}, \mathbf{e}'\mathbf{e}/n, \mathbf{0})\right]' \left\{-E[\mathbf{H}]\right\}^{-1} \left[\left.\frac{\partial \ln L}{\partial \boldsymbol{\gamma}}\right| (\boldsymbol{\gamma} = \mathbf{b}, \mathbf{e}'\mathbf{e}/n, \mathbf{0})\right].$$

With a bit of algebra and using (B-66) for the partitioned inverse, you can show that this reduces to

$$\text{LM} = \frac{1}{2}\left\{\sum_{i=1}^{n} v_i \mathbf{g}_i\right\}\left[\sum_{i=1}^{n}(\mathbf{g}_i - \bar{\mathbf{g}})(\mathbf{g}_i - \bar{\mathbf{g}})'\right]^{-1}\left\{\sum_{i=1}^{n} v_i \mathbf{g}_i\right\}.$$

This result, as given by Breusch and Pagan (1980), is simply one half times the regression sum of squares in the regression of v_i on a constant and $\mathbf{g}_i$. This actually simplifies even further if, as in the cases studied by Bruesch and Pagan, the variance function is $f_i = f(\mathbf{z}_i'\boldsymbol{\alpha})$ where $f(\mathbf{z}_i'\mathbf{0}) = 1$. Then, the derivative will be of the form $\mathbf{g}_i = r(\mathbf{z}_i'\boldsymbol{\alpha})\mathbf{z}_i$ and it will follow that $r_i(\mathbf{z}_i'\mathbf{0}) =$ a constant. In this instance, the same statistic will result from the regression of v_i on a constant and $\mathbf{z}_i$ which is the result reported in Section 11.4.3. The remarkable aspect of the result is that the same statistic results regardless of the choice of variance function, so long as it satisfies $f_i = f(\mathbf{z}_i'\boldsymbol{\alpha})$ where $f(\mathbf{z}_i'\mathbf{0}) = 1$. The model studied by Harvey, for example has $f_i = \exp(\mathbf{z}_i'\boldsymbol{\alpha})$, so $\mathbf{g}_i = \mathbf{z}_i$ when $\boldsymbol{\alpha} = \mathbf{0}$.

Example 11.4 Two-Step Estimation of a Heteroscedastic Regression

Table 11.2 lists weighted least squares and two-step FGLS estimates of the parameters of the regression model in Example 11.1 using various formulations of the scedastic function. The method used to compute the weights for weighted least squares is given below each model formulation. The procedure was iterated to convergence for the model $\sigma_i^2 = \sigma^2 z_i^\alpha$ — convergence required 13 iterations. (The two-step estimates are those computed by the first iteration.) ML estimates for this model are also shown. As often happens, the iteration produces fairly large changes in the estimates. There is also a considerable amount of variation produced by the different formulations.

For the model $f_i = z_i^\alpha$, the concentrated log-likelihood is simple to compute. We can find the maximum likelihood estimate for this model just by scanning over a range of values for α. For any α, the maximum likelihood estimator of $\boldsymbol{\beta}$ is weighted least squares, with weights $w_i = 1/z_i^\alpha$. For our expenditure model, we use income for z_i. Figure 11.2 shows a plot of the log-likelihood function. The maximum occurs at $\alpha = 3.65$. This value, with the FGLS estimates of $\boldsymbol{\beta}$, is shown in Table 11.2.

TABLE 11.2 Two-Step and Weighted Least Squares Estimates

		Constant	*Age*	*OwnRent*	*Income*	*Income²*
$\sigma_i^2 = \sigma^2$ (OLS)	est.	**−237.15**	−**3.0818**	**27.941**	**234.35**	−**14.997**
	s.e.	199.35	5.5147	82.922	80.366	7.4693
$\sigma_i^2 = \sigma^2 I_i$ (WLS)	est.	**−181.87**	−**2.9350**	**50.494**	**202.17**	−**12.114**
	s.e.	165.52	4.6033	69.879	76.781	8.2731
$\sigma_i^2 = \sigma^2 I_i^2$ (WLS)	est.	−**114.11**	−**2.6942**	**60.449**	**158.43**	−**7.2492**
	s.e.	139.69	3.8074	58.551	76.392	9.7243
$\sigma_i^2 = \sigma^2 \exp(\mathbf{z}_i'\boldsymbol{\alpha})$	est.	**−117.88**	−**1.2337**	**50.950**	**145.30**	−**7.9383**
($\ln e_i^2$ on $\mathbf{z}_i = (1, \ln I_i)$)	s.e.	101.39	2.5512	52.814	46.363	3.7367
$\sigma_i^2 = \sigma^2 z_i^\alpha$ (2 Step)	est.	−**193.33**	−**2.9579**	**47.357**	**208.86**	−**12.769**
($\ln e_i^2$ on $(1, \ln z_i)$)	s.e.	171.08	4.7627	72.139	77.198	8.0838
(iterated)	est.	−**130.38**	−**2.7754**	**59.126**	**169.74**	−**8.5995**
($\alpha = 1.7623$)	s.e.	145.03	3.9817	61.0434	76.180	9.3133
(ML)	est.	−**19.929**	−**1.7058**	**58.102**	**75.970**	**4.3915**
($\alpha = 3.6513$)	s.e.	113.06	2.7581	43.5084	81.040	13.433

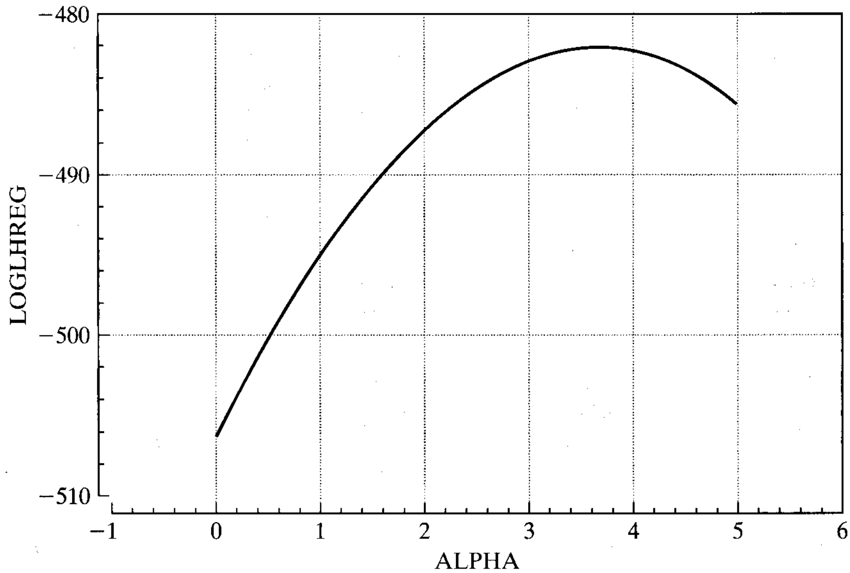

FIGURE 11.2 Plot of Log-Likelihood Function.

Note that this value of α is very different from the value we obtained by iterative regression of the logs of the squared residuals on log income. In this model, $g_i = f_i \ln z_i$. If we insert this into the expression for $\partial \ln L/\partial\alpha$ and manipulate it a bit, we obtain the implicit solution

$$\sum_{i=1}^{n}\left(\frac{\varepsilon_i^2}{\sigma^2 z_i^\alpha} - 1\right)\ln z_i = 0.$$

(The $\frac{1}{2}$ disappears from the solution.) For given values of σ^2 and β, this result provides only an implicit solution for α. In the next section, we examine a method for finding a solution. At this point, we note that the solution to this equation is clearly not obtained by regression of the logs of the squared residuals on log z_i. Hence, the strategy we used for the two-step estimator does not seek the maximum likelihood estimator.

11.7 APPLICATIONS

This section will present two common applications of the heteroscedastic regression model, Harvey's model of **multiplicative heteroscedasticity** and a model of **groupwise heteroscedasticity** that extends to the disturbance variance some concepts that are usually associated with variation in the regression function.

11.7.1 MULTIPLICATIVE HETEROSCEDASTICITY

Harvey's (1976) model of multiplicative heteroscedasticity is a very flexible, general model that includes most of the useful formulations as special cases. The general formulation is

$$\sigma_i^2 = \sigma^2 \exp(\mathbf{z}_i'\boldsymbol{\alpha}).$$

The model examined in Example 11.4 has $z_i = \ln \text{income}_i$. More generally, a model with heteroscedasticity of the form

$$\sigma_i^2 = \sigma^2 \prod_{m=1}^{M} z_{im}^{\alpha_m}$$

results if the logs of the variables are placed in z_i. The groupwise heteroscedasticity model described below is produced by making $\mathbf{z}_i$ a set of group dummy variables (one must be omitted). In this case, σ^2 is the disturbance variance for the base group whereas for the other groups, $\sigma_g^2 = \sigma^2 \exp(\alpha_g)$.

We begin with a useful simplification. Let $\mathbf{z}_i$ include a constant term so that $\mathbf{z}_i' = [1, \mathbf{q}_i']$, where $\mathbf{q}_i$ is the original set of variables, and let $\boldsymbol{\gamma}' = [\ln \sigma^2, \boldsymbol{\alpha}']$. Then, the model is simply $\sigma_i^2 = \exp(\boldsymbol{\gamma}'\mathbf{z}_i)$. Once the full parameter vector is estimated, $\exp(\gamma_1)$ provides the estimator of σ^2. (This estimator uses the invariance result for maximum likelihood estimation. See Section 17.4.5.d.)

The log-likelihood is

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

$$= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n} \mathbf{z}_i'\boldsymbol{\gamma} - \frac{1}{2}\sum_{i=1}^{n} \frac{\varepsilon_i^2}{\exp(\mathbf{z}_i'\boldsymbol{\gamma})}.$$

The likelihood equations are

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \mathbf{x}_i \frac{\varepsilon_i}{\exp(\mathbf{z}_i'\boldsymbol{\gamma})} = \mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} = \mathbf{0},$$

$$\frac{\partial \ln L}{\partial \boldsymbol{\gamma}} = \frac{1}{2}\sum_{i=1}^{n} \mathbf{z}_i \left(\frac{\varepsilon_i^2}{\exp(\mathbf{z}_i'\boldsymbol{\gamma})} - 1\right) = \mathbf{0}.$$

For this model, the method of scoring turns out to be a particularly convenient way to maximize the log-likelihood function. The terms in the Hessian are

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}'} = -\sum_{i=1}^{n} \frac{1}{\exp(\mathbf{z}_i'\boldsymbol{\gamma})} \mathbf{x}_i\mathbf{x}_i' = -\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X},$$

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\gamma}'} = -\sum_{i=1}^{n} \frac{\varepsilon_i}{\exp(\mathbf{z}_i'\boldsymbol{\gamma})} \mathbf{x}_i\mathbf{z}_i',$$

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma}\, \partial \boldsymbol{\gamma}'} = -\frac{1}{2}\sum_{i=1}^{n} \frac{\varepsilon_i^2}{\exp(\mathbf{z}_i'\boldsymbol{\gamma})} \mathbf{z}_i\mathbf{z}_i'.$$

The expected value of $\partial^2 \ln L/\partial \boldsymbol{\beta}\, \partial \boldsymbol{\gamma}'$ is $\mathbf{0}$ since $E[\varepsilon_i | \mathbf{x}_i, \mathbf{z}_i] = 0$. The expected value of the fraction in $\partial^2 \ln L/\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'$ is $E[\varepsilon_i^2/\sigma_i^2 | \mathbf{x}_i, \mathbf{z}_i] = 1$. Let $\boldsymbol{\delta} = [\boldsymbol{\beta}, \boldsymbol{\gamma}]$. Then

$$-E\left(\frac{\partial^2 \ln L}{\partial \boldsymbol{\delta}\, \partial \boldsymbol{\delta}'}\right) = \begin{bmatrix} \mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X} & \mathbf{0} \\ \mathbf{0}' & \frac{1}{2}\mathbf{Z}'\mathbf{Z} \end{bmatrix} = -\mathbf{H}.$$

The scoring method is

$$\boldsymbol{\delta}_{t+1} = \boldsymbol{\delta}_t - \mathbf{H}_t^{-1}\mathbf{g}_t,$$

where $\boldsymbol{\delta}_t$ (i.e., $\boldsymbol{\beta}_t$, $\boldsymbol{\gamma}_t$, and $\boldsymbol{\Omega}_t$) is the estimate at iteration t, $\mathbf{g}_t$ is the two-part vector of first derivatives $[\partial \ln L/\partial \boldsymbol{\beta}_t', \partial \ln L/\partial \boldsymbol{\gamma}_t']'$ and $\mathbf{H}_t$ is partitioned likewise. Since $\mathbf{H}_t$ is block diagonal, the iteration can be written as separate equations:

$$\begin{aligned}\boldsymbol{\beta}_{t+1} &= \boldsymbol{\beta}_t + (\mathbf{X}'\boldsymbol{\Omega}_t^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}_t^{-1}\boldsymbol{\varepsilon}_t)\\ &= \boldsymbol{\beta}_t + (\mathbf{X}'\boldsymbol{\Omega}_t^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}_t^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_t)\\ &= (\mathbf{X}'\boldsymbol{\Omega}_t^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}_t^{-1}\mathbf{y} \text{ (of course)}.\end{aligned}$$

Therefore, the updated coefficient vector $\boldsymbol{\beta}_{t+1}$ is computed by FGLS using the previously computed estimate of $\boldsymbol{\gamma}$ to compute $\boldsymbol{\Omega}$. We use the same approach for $\boldsymbol{\gamma}$:

$$\boldsymbol{\gamma}_{t+1} = \boldsymbol{\gamma}_t + [2(\mathbf{Z}'\mathbf{Z})^{-1}]\left[\frac{1}{2}\sum_{i=1}^{n}\mathbf{z}_i\left(\frac{\varepsilon_i^2}{\exp(\mathbf{z}_i'\boldsymbol{\gamma})} - 1\right)\right].$$

The 2 and $\frac{1}{2}$ cancel. The updated value of $\boldsymbol{\gamma}$ is computed by adding the vector of slopes in the least squares regression of $[\varepsilon_i^2/\exp(\mathbf{z}_i'\boldsymbol{\gamma}) - 1]$ on $\mathbf{z}_i$ to the old one. Note that the correction is $2(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\partial \ln L/\partial \boldsymbol{\gamma})$, so convergence occurs when the derivative is zero.

The remaining detail is to determine the starting value for the iteration. Since any consistent estimator will do, the simplest procedure is to use OLS for $\boldsymbol{\beta}$ and the slopes in a regression of the logs of the squares of the least squares residuals on $\mathbf{z}_i$ for $\boldsymbol{\gamma}$. Harvey (1976) shows that this method will produce an inconsistent estimator of $\gamma_1 = \ln \sigma^2$, but the inconsistency can be corrected just by adding 1.2704 to the value obtained.[19] Thereafter, the iteration is simply:

1. Estimate the disturbance variance σ_i^2 with $\exp(\boldsymbol{\gamma}_t'\mathbf{z}_i)$.
2. Compute $\boldsymbol{\beta}_{t+1}$ by FGLS.[20]
3. Update $\boldsymbol{\gamma}_t$ using the regression described in the preceding paragraph.
4. Compute $\mathbf{d}_{t+1} = [\boldsymbol{\beta}_{t+1}, \boldsymbol{\gamma}_{t+1}] - [\boldsymbol{\beta}_t, \boldsymbol{\gamma}_t]$. If $\mathbf{d}_{t+1}$ is large, then return to step 1.

If $\mathbf{d}_{t+1}$ at step 4 is sufficiently small, then exit the iteration. The asymptotic covariance matrix is simply $-\mathbf{H}^{-1}$, which is block diagonal with blocks

$$\begin{aligned}\text{Asy. Var}[\hat{\boldsymbol{\beta}}_{\text{ML}}] &= (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1},\\ \text{Asy. Var}[\hat{\boldsymbol{\gamma}}_{\text{ML}}] &= 2(\mathbf{Z}'\mathbf{Z})^{-1}.\end{aligned}$$

If desired, then $\hat{\sigma}^2 = \exp(\hat{\gamma}_1)$ can be computed. The asymptotic variance would be $[\exp(\gamma_1)]^2(\text{Asy. Var}[\hat{\gamma}_{1,\text{ML}}])$.

[19] He also presents a correction for the asymptotic covariance matrix for this first step estimator of $\boldsymbol{\gamma}$.

[20] The two-step estimator obtained by stopping here would be fully efficient if the starting value for $\boldsymbol{\gamma}$ were consistent, but it would not be the maximum likelihood estimator.

TABLE 11.3 Multiplicative Heteroscedasticity Model

	Constant	*Age*	*OwnRent*	*Income*	*Income²*
Ordinary Least Squares Estimates					
Coefficient	−237.15	−3.0818	27.941	234.35	−14.997
Standard error	199.35	5.5147	82.922	80.366	7.469
t ratio	−1.1	−0.559	0.337	2.916	−2.008
$R^2 = 0.243578$, $s = 284.75080$, Ln-L $= -506.488$					
Maximum Likelihood Estimates (standard errors for estimates of γ in parentheses)					
Coefficient	−58.437	−0.37607	33.358	96.823	−3.3008
Standard error	62.098	0.55000	37.135	31.798	2.6248
t ratio	−0.941	−0.684	0.898	3.045	−1.448
$[\exp(c_1)]^{1/2} = 0.9792(0.79115)$, $c_2 = 5.355(0.37504)$, $c_3 = -0.56315(0.036122)$					
Ln-L $= -465.9817$, Wald $= 251.423$, LR $= 81.0142$, LM $= 115.899$					

Example 11.5 Multiplicative Heteroscedasticity

Estimates of the regression model of Example 11.1 based on Harvey's model are shown in Table 11.3 with the ordinary least squares results. The scedastic function is

$$\sigma_i^2 = \exp\left(\gamma_1 + \gamma_2 \text{income}_i + \gamma_3 \text{income}_i^2\right).$$

The estimates are consistent with the earlier results in suggesting that Income and its square significantly explain variation in the disturbance variances across observations. The 95 percent critical value for a chi-squared test with two degrees of freedom is 5.99, so all three test statistics lead to rejection of the hypothesis of homoscedasticity.

11.7.2 GROUPWISE HETEROSCEDASTICITY

A groupwise heteroscedastic regression has structural equations

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n,$$
$$E[\varepsilon_i \mid \mathbf{x}_i] = 0, \quad i = 1, \ldots, n.$$

The n observations are grouped into G groups, each with n_g observations. The slope vector is the same in all groups, but within group g:

$$\text{Var}[\varepsilon_{ig} \mid \mathbf{x}_{ig}] = \sigma_g^2, \quad i = 1, \ldots, n_g.$$

If the variances are known, then the GLS estimator is

$$\hat{\boldsymbol{\beta}} = \left[\sum_{g=1}^{G}\left(\frac{1}{\sigma_g^2}\right)\mathbf{X}_g'\mathbf{X}_g\right]^{-1}\left[\sum_{g=1}^{G}\left(\frac{1}{\sigma_g^2}\right)\mathbf{X}_g'\mathbf{y}_g\right]. \qquad \textbf{(11-22)}$$

Since $\mathbf{X}_g'\mathbf{y}_g = \mathbf{X}_g'\mathbf{X}_g\mathbf{b}_g$, where $\mathbf{b}_g$ is the OLS estimator in the gth subset of observations,

$$\hat{\boldsymbol{\beta}} = \left[\sum_{g=1}^{G}\left(\frac{1}{\sigma_g^2}\right)\mathbf{X}_g'\mathbf{X}_g\right]^{-1}\left[\sum_{g=1}^{G}\left(\frac{1}{\sigma_g^2}\right)\mathbf{X}_g'\mathbf{X}_g\mathbf{b}_g\right] = \left[\sum_{g=1}^{G}\mathbf{V}_g\right]^{-1}\left[\sum_{g=1}^{G}\mathbf{V}_g\mathbf{b}_g\right] = \sum_{g=1}^{G}\mathbf{W}_g\mathbf{b}_g.$$

This result is a matrix weighted average of the G least squares estimators. The weighting matrices are $\mathbf{W}_g = \left[\sum_{g=1}^{G}\left(\text{Var}[\mathbf{b}_g]\right)^{-1}\right]^{-1}\left(\text{Var}[\mathbf{b}_g]\right)^{-1}$. The estimator with the smaller covariance matrix therefore receives the larger weight. (If $\mathbf{X}_g$ is the same in every group, then the matrix $\mathbf{W}_g$ reduces to the simple scalar, $w_g = h_g / \sum_g h_g$ where $h_g = 1/\sigma_g^2$.)

The preceding is a useful construction of the estimator, but it relies on an algebraic result that might be unusable. If the number of observations in any group is smaller than the number of regressors, then the group specific OLS estimator cannot be computed. But, as can be seen in (11-22), that is not what is needed to proceed; what is needed are the weights. As always, pooled least squares is a consistent estimator, which means that using the group specific subvectors of the OLS residuals,

$$\hat{\sigma}_g^2 = \frac{\mathbf{e}_g'\mathbf{e}_g}{n_g} \tag{11-23}$$

provides the needed estimator for the group specific disturbance variance. Thereafter, (11-22) is the estimator and the inverse matrix in that expression gives the estimator of the asymptotic covariance matrix.

Continuing this line of reasoning, one might consider iterating the estimator by returning to (11-23) with the two-step FGLS estimator, recomputing the weights, then returning to (11-22) to recompute the slope vector. This can be continued until convergence. It can be shown [see Oberhofer and Kmenta (1974)] that so long as (11-23) is used without a degrees of freedom correction, then if this does converge, it will do so at the maximum likelihood estimator (with normally distributed disturbances). Another method of estimating this model is to treat it as a form of Harvey's model of multiplicative heteroscedasticity where $\mathbf{z}_i$ is a set (minus one) of group dummy variables.

For testing the homoscedasticity assumption in this model, one can use a likelihood ratio test. The log-likelihood function, assuming homoscedasticity, is

$$\ln L_0 = -(n/2)[1 + \ln 2\pi + \ln(\mathbf{e}'\mathbf{e}/n)]$$

where $n = \sum_g n_g$ is the total number of observations. Under the alternative hypothesis of heteroscedasticity across G groups, the log-likelihood function is

$$\ln L_1 = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{g=1}^{G} n_g \ln \sigma_g^2 - \frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{n_g}\left(\varepsilon_{ig}^2/\sigma_g^2\right). \tag{11-24}$$

The maximum likelihood estimators of σ^2 and σ_g^2 are $\mathbf{e}'\mathbf{e}/n$ and $\hat{\sigma}_g^2$ from (11-23), respectively. The OLS and maximum likelihood estimators of $\boldsymbol{\beta}$ are used for the slope vector under the null and alternative hypothesis, respectively. If we evaluate $\ln L_0$ and $\ln L_1$ at these estimates, then the likelihood ratio test statistic for homoscedasticity is

$$-2(\ln L_0 - \ln L_1) = n \ln s^2 - \sum_{g=1}^{G} n_g \ln s_g^2.$$

Under the null hypothesis, the statistic has a limiting chi-squared distribution with $G-1$ degrees of freedom.

Example 11.6 Heteroscedastic Cost Function for Airline Production

To illustrate the computations for the groupwise heteroscedastic model, we will reexamine the cost model for the total cost of production in the airline industry that was fit in Example 7.2.

TABLE 11.4 Least Squares and Maximum Likelihood Estimates of a Groupwise Heteroscedasticity Model

	Least Squares: Homoscedastic			*Maximum Likelihood*		
	Estimate	*Std. Error*	*t Ratio*	*Estimate*	*Std. Error*	*t Ratio*
β_1	9.706	0.193	50.25	10.057	0.134	74.86
β_2	0.418	−.0152	27.47	0.400	0.0108	37.12
β_3	−1.070	0.202	−5.30	−1.129	0.164	−7.87
β_4	0.919	0.0299	30.76	0.928	0.0228	40.86
δ_2	−0.0412	0.0252	−1.64	−0.0487	0.0237	−2.06
δ_3	−0.209	0.0428	−4.88	−0.200	0.0308	−6.49
δ_4	0.185	0.0608	3.04	0.192	0.0499	3.852
δ_5	0.0241	0.0799	0.30	0.0419	0.0594	0.71
δ_6	0.0871	0.0842	1.03	0.0963	0.0631	1.572
γ_1				−7.088	0.365	−19.41
γ_2				2.007	0.516	3.89
γ_3				0.758	0.516	1.47
γ_4				2.239	0.516	4.62
γ_5				0.530	0.516	1.03
γ_6				1.053	0.516	2.04
σ_1^2		0.001479			0.0008349	
σ_2^2		0.004935			0.006212	
σ_3^2		0.001888			0.001781	
σ_4^2		0.005834			0.009071	
σ_5^2		0.002338			0.001419	
σ_6^2		0.003032			0.002393	
	$R^2 = 0.997, s^2 = 0.003613, \ln L = 130.0862$				$\ln L = 140.7591$	

(A description of the data appears in the earlier example.) For a sample of six airlines observed annually for 15 years, we fit the cost function

$$\ln \text{cost}_{it} = \beta_1 + \beta_2 \ln \text{output}_{it} + \beta_3 \text{ load factor}_{it} + \beta_4 \ln \text{ fuel price}_{it}$$

$$\delta_2 \text{ Firm}_2 + \delta_3 \text{ Firm}_3 + \delta_4 \text{ Firm}_4 + \delta_5 \text{ Firm}_5 + \delta_6 \text{ Firm}_6 + \varepsilon_{it}.$$

Output is measured in "revenue passenger miles." The load factor is a rate of capacity utilization; it is the average rate at which seats on the airline's planes are filled. More complete models of costs include other factor prices (materials, capital) and, perhaps, a quadratic term in log output to allow for variable economies of scale. The "firm$_j$" terms are firm specific dummy variables.

Ordinary least squares regression produces the set of results at the left side of Table 11.4. The variance estimates shown at the bottom of the table are the firm specific variance estimates in (11-23). The results so far are what one might expect. There are substantial economies of scale; e.s.$_{it}$ = (1/0.919) − 1 = 0.088. The fuel price and load factors affect costs in the predictable fashions as well. (Fuel prices differ because of different mixes of types and regional differences in supply characteristics.) The second set of results shows the model of groupwise heteroscedasticity. From the least squares variance estimates in the first set of results, which are quite different, one might guess that a test of homoscedasticity would lead to rejection of the hypothesis. The easiest computation is the likelihood ratio test. Based on the log likelihood functions in the last row of the table, the test statistic, which has a limiting chi-squared distribution with 5 degrees of freedom, equals 21.3458. The critical value from the table is 11.07, so the hypothesis of homoscedasticity is rejected.

11.8 AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY

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. Engle's results suggested that in models of inflation, large and small forecast errors appeared to occur in clusters, suggesting a form of heteroscedasticity in which the variance of the forecast error depends on the size of the previous disturbance. He suggested the autoregressive, conditionally heteroscedastic, or ARCH, model as an alternative to the usual time-series process. More recent studies of financial markets suggest that the phenomenon is quite common. The ARCH model has proven to be useful in studying the volatility of inflation [Coulson and Robins (1985)], the term structure of interest rates [Engle, Hendry, and Trumbull (1985)], the volatility of stock market returns [Engle, Lilien, and Robins (1987)], and the behavior of foreign exchange markets [Domowitz and Hakkio (1985) and Bollerslev and Ghysels (1996)], to name but a few. This section will describe specification, estimation, and testing, in the basic formulations of the ARCH model and some extensions.[21]

Example 11.7 Stochastic Volatility
Figure 11.3 shows Bollerslev and Ghysel's 1974 data on the daily percentage nominal return for the Deutschmark/Pound exchange rate. (These data are given in Appendix Table F11.1.) The variation in the series appears to be fluctuating, with several clusters of large and small movements.

11.8.1 THE ARCH(1) MODEL

The simplest form of this model is the ARCH(1) model,

$$\begin{aligned} y_t &= \boldsymbol{\beta}'\mathbf{x}_t + \varepsilon_t \\ \varepsilon_t &= u_t\sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2}, \end{aligned} \tag{11-25}$$

where u_t is distributed as standard normal.[22] It follows that $E[\varepsilon_t \mid \mathbf{x}_t, \varepsilon_{t-1}] = 0$, so that $E[\varepsilon_t \mid \mathbf{x}_t] = 0$ and $E[y_t \mid \mathbf{x}_t] = \boldsymbol{\beta}'\mathbf{x}_t$. Therefore, this model is a classical regression model. But

$$\text{Var}[\varepsilon_t \mid \varepsilon_{t-1}] = E\left[\varepsilon_t^2 \mid \varepsilon_{t-1}\right] = E\left[u_t^2\right]\left[\alpha_0 + \alpha_1\varepsilon_{t-1}^2\right] = \alpha_0 + \alpha_1\varepsilon_{t-1}^2,$$

so ε_t is *conditionally heteroscedastic,* not with respect to $\mathbf{x}_t$ as we considered in the preceding sections, but with respect to ε_{t-1}. The unconditional variance of ε_t is

$$\text{Var}[\varepsilon_t] = \text{Var}\{E[\varepsilon_t \mid \varepsilon_{t-1}]\} + E\{\text{Var}[\varepsilon_t \mid \varepsilon_{t-1}]\} = \alpha_0 + \alpha_1 E\left[\varepsilon_{t-1}^2\right] = \alpha_0 + \alpha_1 \text{Var}[\varepsilon_{t-1}].$$

[21]Engle and Rothschild (1992) give a recent survey of this literature which describes many extensions. Mills (1993) also presents several applications. See, as well, Bollerslev (1986) and Li, Ling, and McAleer (2001). See McCullough and Renfro (1999) for discussion of estimation of this model.

[22]The assumption that u_t has unit variance is not a restriction. The scaling implied by any other variance would be absorbed by the other parameters.

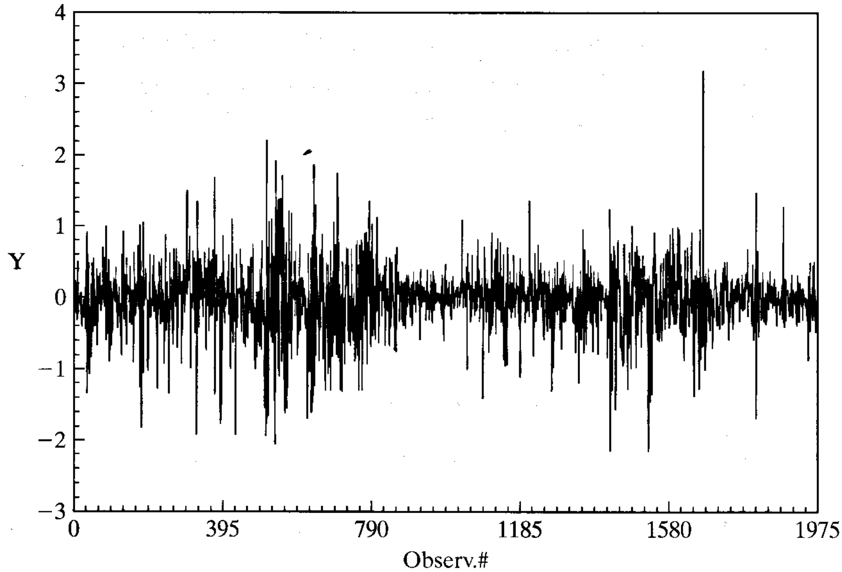

FIGURE 11.3 Nominal Exchange Rate Returns.

If the process generating the disturbances is weakly (covariance) stationary (see Definition 12.2),[23] then the unconditional variance is not changing over time so

$$\text{Var}[\varepsilon_t] = \text{Var}[\varepsilon_{t-1}] = \alpha_0 + \alpha_1 \text{Var}[\varepsilon_{t-1}] = \frac{\alpha_0}{1-\alpha_1}.$$

For this ratio to be finite and positive, $|\alpha_1|$ must be less than 1. Then, unconditionally, ε_t is distributed with mean zero and variance $\sigma^2 = \alpha_0/(1-\alpha_1)$. Therefore, the model obeys the classical assumptions, and ordinary least squares is the most efficient *linear* unbiased estimator of $\boldsymbol{\beta}$.

But there is a more efficient *nonlinear* estimator. The log-likelihood function for this model is given by Engle (1982). Conditioned on starting values y_0 and $\mathbf{x}_0$ (and ε_0), the conditional log-likelihood for observations $t = 1, \ldots, T$ is the one we examined in Section 11.6.2 for the general heteroscedastic regression model [see (11-19)],

$$\ln L = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln\left(\alpha_0 + \alpha_1\varepsilon_{t-1}^2\right) - \frac{1}{2}\sum_{t=1}^{T}\frac{\varepsilon_t^2}{\alpha_0 + \alpha_1\varepsilon_{t-1}^2}, \quad \varepsilon_t = y_t - \boldsymbol{\beta}'\mathbf{x}_t. \tag{11-26}$$

Maximization of log L can be done with the conventional methods, as discussed in Appendix E.[24]

[23]This discussion will draw on the results and terminology of time series analysis in Section 12.3 and Chapter 20. The reader may wish to peruse this material at this point.

[24]Engle (1982) and Judge et al. (1985, pp. 441–444) suggest a four-step procedure based on the method of scoring that resembles the two-step method we used for the multiplicative heteroscedasticity model in Section 11.6. However, the full MLE is now incorporated in most modern software, so the simple regression based methods, which are difficult to generalize, are less attractive in the current literature. But, see McCullough and Renfro (1999) and Fiorentini, Calzolari and Panattoni (1996) for commentary and some cautions related to maximum likelihood estimation.

11.8.2 ARCH(q), ARCH-IN-MEAN AND GENERALIZED ARCH MODELS

The natural extension of the ARCH(1) model presented before is a more general model with longer lags. The ARCH(q) process,

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2,$$

is a qth order **moving average** [MA(q)] process. (Much of the analysis of the model parallels the results in Chapter 20 for more general time series models.) [Once again, see Engle (1982).] This section will generalize the ARCH(q) model, as suggested by Bollerslev (1986), in the direction of the autoregressive-moving average (ARMA) models of Section 20.2.1. The discussion will parallel his development, although many details are omitted for brevity. The reader is referred to that paper for background and for some of the less critical details.

The capital asset pricing model (CAPM) is discussed briefly in Chapter 14. Among the many variants of this model is an intertemporal formulation by Merton (1980) that suggests an approximate linear relationship between the return and variance of the market portfolio. One of the possible flaws in this model is its assumption of a constant variance of the market portfolio. In this connection, then, the **ARCH-in-Mean,** or ARCH-M, model suggested by Engle, Lilien, and Robins (1987) is a natural extension. The model states that

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \delta\sigma_t^2 + \varepsilon_t,$$
$$\text{Var}[\varepsilon_t \mid \Psi_t] = \text{ARCH}(q).$$

Among the interesting implications of this modification of the standard model is that under certain assumptions, δ is the coefficient of relative risk aversion. The ARCH-M model has been applied in a wide variety of studies of volatility in asset returns, including the daily Standard and Poor's Index [French, Schwert, and Stambaugh (1987)] and weekly New York Stock Exchange returns [Chou (1988)]. A lengthy list of applications is given in Bollerslev, Chou, and Kroner (1992).

The ARCH-M model has several noteworthy statistical characteristics. Unlike the standard regression model, misspecification of the variance function does impact on the consistency of estimators of the parameters of the mean. [See Pagan and Ullah (1988) for formal analysis of this point.] Recall that in the classical regression setting, weighted least squares is consistent even if the weights are misspecified as long as the weights are uncorrelated with the disturbances. That is not true here. If the ARCH part of the model is misspecified, then conventional estimators of $\boldsymbol{\beta}$ and δ will not be consistent. Bollerslev, Chou, and Kroner (1992) list a large number of studies that called into question the specification of the ARCH-M model, and they subsequently obtained quite different results after respecifying the model. A closely related practical problem is that the mean and variance parameters in this model are no longer uncorrelated. In analysis up to this point, we made quite profitable use of the block diagonality of the Hessian of the log-likelihood function for the model of heteroscedasticity. But the Hessian for the ARCH-M model is not block diagonal. In practical terms, the estimation problem cannot be segmented as we have done previously with the heteroscedastic regression model. All the parameters must be estimated simultaneously.

The model of generalized autoregressive conditional heteroscedasticity (GARCH) is defined as follows.[25] The underlying regression is the usual one in (11-25). *Conditioned on an information set at time t*, denoted Ψ_t, the distribution of the disturbance is assumed to be

$$\varepsilon_t \mid \Psi_t \sim N[0, \sigma_t^2],$$

where the conditional variance is

$$\sigma_t^2 = \alpha_0 + \delta_1\sigma_{t-1}^2 + \delta_2\sigma_{t-2}^2 + \cdots + \delta_p\sigma_{t-p}^2 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2. \quad \textbf{(11-27)}$$

Define

$$\mathbf{z}_t = \left[1, \sigma_{t-1}^2, \sigma_{t-2}^2, \ldots, \sigma_{t-p}^2, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots, \varepsilon_{t-q}^2\right]'$$

and

$$\boldsymbol{\gamma} = [\alpha_0, \delta_1, \delta_2, \ldots, \delta_p, \alpha_1, \ldots, \alpha_q]' = [\alpha_0, \boldsymbol{\delta}', \boldsymbol{\alpha}']'.$$

Then

$$\sigma_t^2 = \boldsymbol{\gamma}'\mathbf{z}_t.$$

Notice that the conditional variance is defined by an autoregressive-moving average [ARMA (p, q)] process in the innovations ε_t^2, exactly as in Section 20.2.1. The difference here is that the *mean* of the random variable of interest y_t is described completely by a heteroscedastic, but otherwise ordinary, regression model. The *conditional variance,* however, evolves over time in what might be a very complicated manner, depending on the parameter values and on p and q. The model in (11-27) is a GARCH(p, q) model, where p refers, as before, to the order of the autoregressive part.[26] As Bollerslev (1986) demonstrates with an example, the virtue of this approach is that a GARCH model with a small number of terms appears to perform as well as or better than an ARCH model with many.

The **stationarity conditions** discussed in Section 20.2.2 are important in this context to ensure that the moments of the normal distribution are finite. The reason is that higher moments of the normal distribution are finite powers of the variance. A normal distribution with variance σ_t^2 has fourth moment $3\sigma_t^4$, sixth moment $15\sigma_t^6$, and so on. [The precise relationship of the even moments of the normal distribution to the variance is $\mu_{2k} = (\sigma^2)^k(2k)!/(k!2^k)$.] Simply ensuring that σ_t^2 is stable does not ensure that higher powers are as well.[27] Bollerslev presents a useful figure that shows the conditions needed to ensure stability for moments up to order 12 for a GARCH(1, 1) model and gives some additional discussion. For example, for a GARCH(1, 1) process, for the fourth moment to exist, $3\alpha_1^2 + 2\alpha_1\delta_1 + \delta_1^2$ must be less than 1.

[25]As have most areas in time-series econometrics, the line of literature on GARCH models has progressed rapidly in recent years and will surely continue to do so. We have presented Bollerslev's model in some detail, despite many recent extensions, not only to introduce the topic as a bridge to the literature, but also because it provides a convenient and interesting setting in which to discuss several related topics such as double-length regression and pseudo–maximum likelihood estimation.

[26]We have changed Bollerslev's notation slightly so as not to conflict with our previous presentation. He used $\boldsymbol{\beta}$ instead of our $\boldsymbol{\delta}$ in (18-25) and **b** instead of our $\boldsymbol{\beta}$ in (18-23).

[27]The conditions cannot be imposed a priori. In fact, there is no nonzero set of parameters that guarantees stability of *all* moments, even though the normal distribution has finite moments of all orders. As such, the normality assumption must be viewed as an approximation.

It is convenient to write (11-27) in terms of polynomials in the lag operator:

$$\sigma_t^2 = \alpha_0 + D(L)\sigma_t^2 + A(L)\varepsilon_t^2.$$

As discussed in Section 20.2.2, the stationarity condition for such an equation is that the roots of the characteristic equation, $1 - D(z) = 0$, must lie outside the unit circle. For the present, we will assume that this case is true for the model we are considering and that $A(1) + D(1) < 1$. [This assumption is stronger than that needed to ensure stationarity in a higher-order autoregressive model, which would depend only on $D(L)$.] The implication is that the GARCH process is covariance stationary with $E[\varepsilon_t] = 0$ (unconditionally), $\text{Var}[\varepsilon_t] = \alpha_0/[1 - A(1) - D(1)]$, and $\text{Cov}[\varepsilon_t, \varepsilon_s] = 0$ for all $t \neq s$. Thus, unconditionally the model is the classical regression model that we examined in Chapters 2–8.

The usefulness of the GARCH specification is that it allows the variance to evolve over time in a way that is much more general than the simple specification of the ARCH model. The comparison between simple finite-distributed lag models and the dynamic regression model discussed in Chapter 19 is analogous. For the example discussed in his paper, Bollerslev reports that although Engle and Kraft's (1983) ARCH(8) model for the rate of inflation in the GNP deflator appears to remove all ARCH effects, a closer look reveals GARCH effects at several lags. By fitting a GARCH (1, 1) model to the same data, Bollerslev finds that the ARCH effects out to the same eight-period lag as fit by Engle and Kraft and his observed GARCH effects are all satisfactorily accounted for.

11.8.3 MAXIMUM LIKELIHOOD ESTIMATION OF THE GARCH MODEL

Bollerslev describes a method of estimation based on the BHHH algorithm. As he shows, the method is relatively simple, although with the line search and first derivative method that he suggests, it probably involves more computation and more iterations than necessary. Following the suggestions of Harvey (1976), it turns out that there is a simpler way to estimate the GARCH model that is also very illuminating. This model is actually very similar to the more conventional model of multiplicative heteroscedasticity that we examined in Section 11.7.1.

For normally distributed disturbances, the log-likelihood for a sample of T observations is

$$\ln L = \sum_{t=1}^{T} -\frac{1}{2}\left[\ln(2\pi) + \ln\sigma_t^2 + \frac{\varepsilon_t^2}{\sigma_t^2}\right] = \sum_{t=1}^{T} \ln f_t(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_t(\boldsymbol{\theta}),^{28}$$

where $\varepsilon_t = y_t - \mathbf{x}_t'\boldsymbol{\beta}$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \alpha_0, \boldsymbol{\alpha}', \boldsymbol{\delta}')' = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$. Derivatives of $\ln L$ are obtained by summation. Let l_t denote $\ln f_t(\boldsymbol{\theta})$. The first derivatives with respect to the variance parameters are

$$\frac{\partial l_t}{\partial \boldsymbol{\gamma}} = -\frac{1}{2}\left[\frac{1}{\sigma_t^2} - \frac{\varepsilon_t^2}{(\sigma_t^2)^2}\right]\frac{\partial \sigma_t^2}{\partial \boldsymbol{\gamma}} = \frac{1}{2}\left(\frac{1}{\sigma_t^2}\right)\frac{\partial \sigma_t^2}{\partial \boldsymbol{\gamma}}\left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right) = \frac{1}{2}\left(\frac{1}{\sigma_t^2}\right)\mathbf{g}_t v_t = \mathbf{b}_t v_t. \tag{11-28}$$

[28] There are three minor errors in Bollerslev's derivation that we note here to avoid the apparent inconsistencies. In his (22), $\frac{1}{2}h_t$ should be $\frac{1}{2}h_t^{-1}$. In (23), $-2h_t^{-2}$ should be $-h_t^{-2}$. In (28), $h\,\partial h/\partial\omega$ should, in each case, be $(1/h)\,\partial h/\partial\omega$. [In his (8), $\alpha_0\alpha_1$ should be $\alpha_0 + \alpha_1$, but this has no implications for our derivation.]

Note that $E[v_t] = 0$. Suppose, for now, that there are no regression parameters. Newton's method for estimating the variance parameters would be

$$\hat{\gamma}^{i+1} = \hat{\gamma}^i - \mathbf{H}^{-1}\mathbf{g}, \qquad \textbf{(11-29)}$$

where **H** indicates the Hessian and **g** is the first derivatives vector. Following Harvey's suggestion (see Section 11.7.1), we will use the method of scoring instead. To do this, we make use of $E[v_t]=0$ and $E[\varepsilon_t^2/\sigma_t^2]=1$. After taking expectations in (11-28), the iteration reduces to a linear regression of $v_{*_t} = (1/\sqrt{2})v_t$ on regressors $\mathbf{w}_{*_t} = (1/\sqrt{2})\mathbf{g}_t/\sigma_t^2$. That is,

$$\hat{\gamma}^{i+1} = \hat{\gamma}^i + [\mathbf{W}_*'\mathbf{W}_*]^{-1}\mathbf{W}_*'\mathbf{v}_* = \hat{\gamma}^i + [\mathbf{W}_*'\mathbf{W}_*]^{-1}\left(\frac{\partial \ln L}{\partial \gamma}\right), \qquad \textbf{(11-30)}$$

where row t of $\mathbf{W}_*$ is $\mathbf{w}_{*_t}'$. The iteration has converged when the slope vector is zero, which happens when the first derivative vector is zero. When the iterations are complete, the estimated asymptotic covariance matrix is simply

$$\text{Est.Asy. Var}[\hat{\gamma}] = [\hat{\mathbf{W}}_*'\mathbf{W}_*]^{-1}$$

based on the estimated parameters.

The usefulness of the result just given is that $E[\partial^2 \ln L/\partial\gamma\,\partial\boldsymbol{\beta}']$ is, in fact, zero. Since the expected Hessian is block diagonal, applying the method of scoring to the full parameter vector can proceed in two parts, exactly as it did in Section 11.7.1 for the multiplicative heteroscedasticity model. That is, the updates for the mean and variance parameter vectors can be computed separately. Consider then the slope parameters, $\boldsymbol{\beta}$. The same type of modified scoring method as used earlier produces the iteration

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{i+1} &= \hat{\boldsymbol{\beta}}^i + \left[\sum_{t=1}^T \frac{\mathbf{x}_t\mathbf{x}_t'}{\sigma_t^2} + \frac{1}{2}\left(\frac{\mathbf{d}_t}{\sigma_t^2}\right)\left(\frac{\mathbf{d}_t}{\sigma_t^2}\right)'\right]^{-1}\left[\sum_{t=1}^T \frac{\mathbf{x}_t\varepsilon_t}{\sigma_t^2} + \frac{1}{2}\left(\frac{\mathbf{d}_t}{\sigma_t^2}\right)v_t\right]\\ &= \hat{\boldsymbol{\beta}}^i + \left[\sum_{t=1}^T \frac{\mathbf{x}_t\mathbf{x}_t'}{\sigma_t^2} + \frac{1}{2}\left(\frac{\mathbf{d}_t}{\sigma_t^2}\right)\left(\frac{\mathbf{d}_t}{\sigma_t^2}\right)'\right]^{-1}\left(\frac{\partial \ln L}{\partial \boldsymbol{\beta}}\right) \qquad \textbf{(11-31)}\\ &= \hat{\boldsymbol{\beta}}^i + \mathbf{h}^i,\end{aligned}$$

which has been referred to as a **double-length regression.** [See Orme (1990) and Davidson and MacKinnon (1993, Chapter 14).] The update vector $\mathbf{h}^i$ is the vector of slopes in an augmented or double-length generalized regression,

$$\mathbf{h}^i = [\mathbf{C}'\boldsymbol{\Omega}^{-1}\mathbf{C}]^{-1}[\mathbf{C}'\boldsymbol{\Omega}^{-1}\mathbf{a}], \qquad \textbf{(11-32)}$$

where **C** is a $2T \times K$ matrix whose first T rows are the **X** from the original regression model and whose next T rows are $(1/\sqrt{2})\mathbf{d}_t'/\sigma_t^2$, $t = 1, \ldots, T$; **a** is a $2T \times 1$ vector whose first T elements are ε_t and whose next T elements are $(1/\sqrt{2})v_t/\sigma_t^2$, $t = 1, \ldots, T$; and $\boldsymbol{\Omega}$ is a diagonal matrix with $1/\sigma_t^2$ in positions $1, \ldots, T$ and ones below observation T. At convergence, $[\mathbf{C}'\boldsymbol{\Omega}^{-1}\mathbf{C}]^{-1}$ provides the asymptotic covariance matrix for the MLE. The resemblance to the familiar result for the generalized regression model is striking, but note that this result is based on the double-length regression.

The iteration is done simply by computing the update vectors to the current parameters as defined above.[29] An important consideration is that to apply the scoring method, the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are updated simultaneously. That is, one does not use the updated estimate of $\boldsymbol{\gamma}$ in (11-30) to update the weights for the GLS regression to compute the new $\boldsymbol{\beta}$ in (11-31). The same estimates (the results of the prior iteration) are used on the right-hand sides of both (11-30) and (11-31). The remaining problem is to obtain starting values for the iterations. One obvious choice is **b**, the OLS estimator, for $\boldsymbol{\beta}$, $\mathbf{e}'\mathbf{e}/T = s^2$ for α_0, and zero for all the remaining parameters. The OLS slope vector will be consistent under all specifications. A useful alternative in this context would be to start $\boldsymbol{\alpha}$ at the vector of slopes in the least squares regression of e_t^2, the squared OLS residual, on a constant and q lagged values.[30] As discussed below, an LM test for the presence of GARCH effects is then a by-product of the first iteration. In principle, the updated result of the first iteration is an **efficient two-step estimator** of all the parameters. But having gone to the full effort to set up the iterations, nothing is gained by not iterating to convergence. One virtue of allowing the procedure to iterate to convergence is that the resulting log-likelihood function can be used in likelihood ratio tests.

11.8.4 TESTING FOR GARCH EFFECTS

The preceding development appears fairly complicated. In fact, it is not, since at each step, nothing more than a linear least squares regression is required. The intricate part of the computation is setting up the derivatives. On the other hand, it does take a fair amount of programming to get this far.[31] As Bollerslev suggests, it might be useful to test for GARCH effects first.

The simplest approach is to examine the squares of the least squares residuals. The autocorrelations (correlations with lagged values) of the squares of the residuals provide evidence about ARCH effects. An LM test of ARCH(q) against the hypothesis of no ARCH effects [ARCH(0), the classical model] can be carried out by computing $\chi^2 = TR^2$ in the regression of e_t^2 on a constant and q lagged values. Under the null hypothesis of no ARCH effects, the statistic has a limiting chi-squared distribution with q degrees of freedom. Values larger than the critical table value give evidence of the presence of ARCH (or GARCH) effects.

Bollerslev suggests a Lagrange multiplier statistic that is, in fact, surprisingly simple to compute. The LM test for GARCH(p, 0) against GARCH(p, q) can be carried out by referring T times the R^2 in the linear regression defined in (11-30) to the chi-squared critical value with q degrees of freedom. There is, unfortunately, an indeterminacy in this test procedure. The test for ARCH(q) against GARCH(p, q) is exactly the same as that for ARCH(p) against ARCH($p + q$). For carrying out the test, one can use as

[29]See Fiorentini et al. (1996) on computation of derivatives in GARCH models.

[30]A test for the presence of q ARCH effects against none can be carried out by carrying TR^2 from this regression into a table of critical values for the chi-squared distribution. But in the presence of GARCH effects, this procedure loses its validity.

[31]Since this procedure is available as a preprogrammed procedure in many computer programs, including TSP, E-Views, Stata, RATS, LIMDEP, and Shazam, this warning might itself be overstated.

TABLE 11.5 Maximum Likelihood Estimates of a GARCH(1, 1) Model[32]

	μ	α_0	α_1	δ	$\alpha_0/(1-\alpha_1-\delta)$
Estimate	−0.006190	0.01076	0.1531	0.8060	0.2631
Std. Error	0.00873	0.00312	0.0273	0.0302	0.594
t ratio	−0.709	3.445	5.605	26.731	0.443
	$\ln L = -1106.61, \ln L_{OLS} = -1311.09, \bar{y} = -0.01642, s^2 = 0.221128$				

starting values a set of estimates that includes $\boldsymbol{\delta} = \mathbf{0}$ and any consistent estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. Then TR^2 for the regression at the initial iteration provides the test statistic.[33]

A number of recent papers have questioned the use of test statistics based solely on normality. Wooldridge (1991) is a useful summary with several examples.

Example 11.8 GARCH Model for Exchange Rate Volatility

Bollerslev and Ghysels analyzed the exchange rate data in Example 11.7 using a GARCH(1, 1) model,

$$y_t = \mu + \varepsilon_t,$$
$$E[\varepsilon_t \mid \varepsilon_{t-1}] = 0,$$
$$\text{Var}[\varepsilon_t \mid \varepsilon_{t-1}] = \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \delta\sigma_{t-1}^2.$$

The least squares residuals for this model are simply $e_t = y_t - \bar{y}$. Regression of the squares of these residuals on a constant and 10 lagged squared values using observations 11-1974 produces an $R^2 = 0.025255$. With $T = 1964$, the chi-squared statistic is 49.60, which is larger than the critical value from the table of 18.31. We conclude that there is evidence of GARCH effects in these residuals. The maximum likelihood estimates of the GARCH model are given in Table 11.5. Note the resemblance between the OLS unconditional variance (0.221128) and the estimated equilibrium variance from the GARCH model, 0.2631.

11.8.5 PSEUDO-MAXIMUM LIKELIHOOD ESTIMATION

We now consider an implication of nonnormality of the disturbances. Suppose that the assumption of normality is weakened to only

$$E[\varepsilon_t \mid \Psi_t] = 0, \quad E\left[\frac{\varepsilon_t^2}{\sigma_t^2} \,\middle|\, \Psi_t\right] = 1, \quad E\left[\frac{\varepsilon_t^4}{\sigma_t^4} \,\middle|\, \Psi_t\right] = \kappa < \infty,$$

where σ_t^2 is as defined earlier. Now the normal log-likelihood function is inappropriate. In this case, the nonlinear (ordinary or weighted) least squares estimator would have the properties discussed in Chapter 9. It would be more difficult to compute than the MLE discussed earlier, however. It has been shown [see White (1982a) and Weiss (1982)] that the *pseudo-MLE* obtained by maximizing the same log-likelihood as if it were

[32] These data have become a standard data set for the evaluation of software for estimating GARCH models. The values given are the benchmark estimates. Standard errors differ substantially from one method to the next. Those given are the Bollerslev and Wooldridge (1992) results. See McCullough and Renfro (1999).

[33] Bollerslev argues that in view of the complexity of the computations involved in estimating the GARCH model, it is useful to have a test for GARCH effects. This case is one (as are many other maximum likelihood problems) in which the apparatus for carrying out the test is the same as that for estimating the model, however. Having computed the LM statistic for GARCH effects, one can proceed to estimate the model just by allowing the program to iterate to convergence. There is no additional cost beyond waiting for the answer.

correct produces a consistent estimator despite the misspecification.[34] The asymptotic covariance matrices for the parameter estimators must be adjusted, however.

The general result for cases such as this one [see Gourieroux, Monfort, and Trognon (1984)] is that the appropriate asymptotic covariance matrix for the pseudo-MLE of a parameter vector $\boldsymbol{\theta}$ would be

$$\text{Asy. Var}[\hat{\boldsymbol{\theta}}] = \mathbf{H}^{-1}\mathbf{F}\mathbf{H}^{-1}, \tag{11-33}$$

where

$$\mathbf{H} = -E\left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'}\right]$$

and

$$\mathbf{F} = E\left[\left(\frac{\partial \ln L}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \ln L}{\partial \boldsymbol{\theta}'}\right)\right]$$

(that is, the BHHH estimator), and $\ln L$ is the used but inappropriate log-likelihood function. For current purposes, $\mathbf{H}$ and $\mathbf{F}$ are still block diagonal, so we can treat the mean and variance parameters separately. In addition, $E[v_t]$ is still zero, so the second derivative terms in both blocks are quite simple. (The parts involving $\partial^2\sigma_t^2/\partial\boldsymbol{\gamma}\,\partial\boldsymbol{\gamma}'$ and $\partial^2\sigma_t^2/\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}'$ fall out of the expectation.) Taking expectations and inserting the parts produces the corrected asymptotic covariance matrix for the variance parameters:

$$\text{Asy. Var}[\hat{\boldsymbol{\gamma}}_{\text{PMLE}}] = [\mathbf{W}_*'\mathbf{W}_*]^{-1}\mathbf{B}'\mathbf{B}[\mathbf{W}_*'\mathbf{W}_*]^{-1},$$

where the rows of $\mathbf{W}^*$ are defined in (18-30) and those of $\mathbf{B}$ are in (11-28). For the slope parameters, the adjusted asymptotic covariance matrix would be

$$\text{Asy. Var}[\hat{\boldsymbol{\beta}}_{\text{PMLE}}] = [\mathbf{C}'\boldsymbol{\Omega}^{-1}\mathbf{C}]^{-1}\left[\sum_{t=1}^{T}\mathbf{b}_t\mathbf{b}_t'\right][\mathbf{C}'\boldsymbol{\Omega}^{-1}\mathbf{C}]^{-1},$$

where the outer matrix is defined in (11-31) and, from the first derivatives given in (11-29) and (11-31),

$$\mathbf{b}_t = \frac{\mathbf{x}_t\varepsilon_t}{\sigma_t^2} + \frac{1}{2}\left(\frac{v_t}{\sigma_t^2}\right)\mathbf{d}_t.$$ [35]

11.9 SUMMARY AND CONCLUSIONS

This chapter has analyzed one form of the generalized regression model, the model of heteroscedasticity. We first considered least squares estimation. The primary result for

[34] White (1982a) gives some additional requirements for the true underlying density of ε_t. Gourieroux, Monfort, and Trognon (1984) also consider the issue. Under the assumptions given, the expectations of the matrices in (18-27) and (18-32) remain the same as under normality. The consistency and asymptotic normality of the pseudo-MLE can be argued under the logic of GMM estimators.

[35] McCullough and Renfro (1999) examined several approaches to computing an appropriate asymptotic covariance matrix for the GARCH model, including the conventional Hessian and BHHH estimators and three sandwich style estimators including the one suggested above, and two based on the method of scoring suggested by Bollerslev and Wooldridge (1992). None stand out as obviously better, but the Bollerslev and QMLE estimator based on an actual Hessian appears to perform well in Monte Carlo studies.

least squares estimation is that it retains its consistency and asymptotic normality, but some correction to the estimated asymptotic covariance matrix may be needed for appropriate inference. The White estimator is the standard approach for this computation. These two results also constitute the GMM estimator for this model. After examining some general tests for heteroscedasticity, we then narrowed the model to some specific parametric forms, and considered weighted (generalized) least squares and maximum likelihood estimation. If the form of the heteroscedasticity is known but involves unknown parameters, then it remains uncertain whether FGLS corrections are better than OLS. Asymptotically, the comparison is clear, but in small or moderately sized samples, the additional variation incorporated by the estimated variance parameters may offset the gains to GLS. The final section of this chapter examined a model of stochastic volatility, the GARCH model. This model has proved especially useful for analyzing financial data such as exchange rates, inflation, and market returns.

Key Terms and Concepts

- ARCH model
- ARCH-in-mean
- Breusch–Pagan test
- Double-length regression
- Efficient two-step estimator
- GARCH model
- Generalized least squares
- Generalized sum of squares
- GMM estimator
- Goldfeld–Quandt test
- Groupwise heteroscedasticity
- Lagrange multiplier test
- Heteroscedasticity
- Likelihood ratio test
- Maximum likelihood estimators
- Model based test
- Moving average
- Multiplicative heteroscedasticity
- Nonconstructive test
- Residual based test
- Robust estimator
- Robustness to unknown heteroscedasticity
- Stationarity condition
- Stochastic volatility
- Two-step estimator
- Wald test
- Weighted least squares
- White estimator
- White's test

Exercises

1. Suppose that the regression model is $y_i = \mu + \varepsilon_i$, where $E[\varepsilon_i \mid x_i] = 0$, $\text{Cov}[\varepsilon_i, \varepsilon_j \mid x_i, x_j] = 0$ for $i \neq j$, but $\text{Var}[\varepsilon_i \mid x_i] = \sigma^2 x_i^2$, $x_i > 0$.
   a. Given a sample of observations on y_i and x_i, what is the most efficient estimator of μ? What is its variance?
   b. What is the OLS estimator of μ, and what is the variance of the ordinary least squares estimator?
   c. Prove that the estimator in part a is at least as efficient as the estimator in part b.
2. For the model in the previous exercise, what is the probability limit of $s^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$? Note that s^2 is the least squares estimator of the residual variance. It is also n times the conventional estimator of the variance of the OLS estimator,

$$\text{Est. Var}[\bar{y}] = s^2(\mathbf{X}'\mathbf{X})^{-1} = \frac{s^2}{n}.$$

   How does this equation compare with the true value you found in part b of Exercise 1? Does the conventional estimator produce the correct estimate of the true asymptotic variance of the least squares estimator?

3. Two samples of 50 observations each produce the following moment matrices. (In each case, **X** is a constant and one variable.)

	Sample 1	Sample 2
$\mathbf{X'X}$	$\begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix}$	$\begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix}$
$\mathbf{y'X}$	[300 2000]	[300 2200]
$\mathbf{y'y}$	2100	2800

   a. Compute the least squares regression coefficients and the residual variances s^2 for each data set. Compute the R^2 for each regression.
   b. Compute the OLS estimate of the coefficient vector assuming that the coefficients and disturbance variance are the same in the two regressions. Also compute the estimate of the asymptotic covariance matrix of the estimate.
   c. Test the hypothesis that the variances in the two regressions are the same without assuming that the coefficients are the same in the two regressions.
   d. Compute the two-step FGLS estimator of the coefficients in the regressions, assuming that the constant and slope are the same in both regressions. Compute the estimate of the covariance matrix and compare it with the result of part b.
4. Using the data in Exercise 3, use the Oberhofer–Kmenta method to compute the maximum likelihood estimate of the common coefficient vector.
5. This exercise is based on the following data set.

50 Observations on Y:

−1.42	2.75	2.10	−5.08	1.49	1.00	0.16	−1.11	1.66
−0.26	−4.87	5.94	2.21	−6.87	0.90	1.61	2.11	−3.82
−0.62	7.01	26.14	7.39	0.79	1.93	1.97	−23.17	−2.52
−1.26	−0.15	3.41	−5.45	1.31	1.52	2.04	3.00	6.31
5.51	−15.22	−1.47	−1.48	6.66	1.78	2.62	−5.16	−4.71
−0.35	−0.48	1.24	0.69	1.91				

50 Observations on X_1:

−1.65	1.48	0.77	0.67	0.68	0.23	−0.40	−1.13	0.15
−0.63	0.34	0.35	0.79	0.77	−1.04	0.28	0.58	−0.41
−1.78	1.25	0.22	1.25	−0.12	0.66	1.06	−0.66	−1.18
−0.80	−1.32	0.16	1.06	−0.60	0.79	0.86	2.04	−0.51
0.02	0.33	−1.99	0.70	−0.17	0.33	0.48	1.90	−0.18
−0.18	−1.62	0.39	0.17	1.02				

50 Observations on X_2:

−0.67	0.70	0.32	2.88	−0.19	−1.28	−2.72	−0.70	−1.55
−0.74	−1.87	1.56	0.37	−2.07	1.20	0.26	−1.34	−2.10
0.61	2.32	4.38	2.16	1.51	0.30	−0.17	7.82	−1.15
1.77	2.92	−1.94	2.09	1.50	−0.46	0.19	−0.39	1.54
1.87	−3.45	−0.88	−1.53	1.42	−2.70	1.77	−1.89	−1.85
2.01	1.26	−2.02	1.91	−2.23				

   a. Compute the ordinary least squares regression of Y on a constant, X_1, and X_2. Be sure to compute the conventional estimator of the asymptotic covariance matrix of the OLS estimator as well.

b. Compute the White estimator of the appropriate asymptotic covariance matrix for the OLS estimates.
c. Test for the presence of heteroscedasticity using White's general test. Do your results suggest the nature of the heteroscedasticity?
d. Use the Breusch–Pagan Lagrange multiplier test to test for heteroscedasticity.
e. Sort the data keying on X_1 and use the Goldfeld–Quandt test to test for heteroscedasticity. Repeat the procedure, using X_2. What do you find?

6. Using the data of Exercise 5, reestimate the parameters using a two-step FGLS estimator. Try the estimator used in Example 11.4.
7. For the model in Exercise 1, suppose that ε is normally distributed, with mean zero and variance $\sigma^2[1 + (\gamma x)^2]$. Show that σ^2 and γ^2 can be consistently estimated by a regression of the least squares residuals on a constant and x^2. Is this estimator efficient?
8. Derive the log-likelihood function, first-order conditions for maximization, and information matrix for the model $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$, $\varepsilon_i \sim N[0, \sigma^2(\boldsymbol{\gamma}'\mathbf{z}_i)^2]$.
9. Suppose that y has the pdf $f(y \mid \mathbf{x}) = (1/\boldsymbol{\beta}'\mathbf{x})e^{-y/(\boldsymbol{\beta}'\mathbf{x})}$, $y > 0$. Then $E[y \mid \mathbf{x}] = \boldsymbol{\beta}'\mathbf{x}$ and $\text{Var}[y \mid \mathbf{x}] = (\boldsymbol{\beta}'\mathbf{x})^2$. For this model, prove that GLS and MLE are the same, even though this distribution involves the same parameters in the conditional mean function and the disturbance variance.
10. In the discussion of Harvey's model in Section 11.7, it is noted that the initial estimator of γ_1, the constant term in the regression of $\ln e_i^2$ on a constant, and $\mathbf{z}_i$ is inconsistent by the amount 1.2704. Harvey points out that if the purpose of this initial regression is only to obtain starting values for the iterations, then the correction is not necessary. Explain why this statement would be true.
11. (This exercise requires appropriate computer software. The computations required can be done with *RATS*, *EViews*, *Stata*, *TSP*, *LIMDEP*, and a variety of other software using only preprogrammed procedures.) Quarterly data on the consumer price index for 1950.1 to 2000.4 are given in Appendix Table F5.1. Use these data to fit the model proposed by Engle and Kraft (1983). The model is

$$\pi_t = \beta_0 + \beta_1\pi_{t-1} + \beta_2\pi_{t-2} + \beta_3\pi_{t-3} + \beta_4\pi_{t-4} + \varepsilon_t$$

where $\pi_t = 100 \ln[p_t/p_{t-1}]$ and p_t is the price index.

a. Fit the model by ordinary least squares, then use the tests suggested in the text to see if ARCH effects appear to be present.
b. The authors fit an ARCH(8) model with declining weights,

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{8}\left(\frac{9-i}{36}\right)\varepsilon_{t-i}^2$$

Fit this model. If the software does not allow constraints on the coefficients, you can still do this with a two-step least squares procedure, using the least squares residuals from the first step. What do you find?
c. Bollerslev (1986) recomputed this model as a GARCH(1,1). Use the GARCH(1,1) form and refit your model.