9

THE GENERALIZED REGRESSION MODEL AND HETEROSCEDASTICITY

9.1 INTRODUCTION

In this and the next several chapters, we will extend the multiple regression model to disturbances that violate Assumption A.4 of the classical regression model. The **generalized linear regression model** is

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \\ E[\boldsymbol{\varepsilon}\,|\,\mathbf{X}] &= \mathbf{0}, \\ E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\,|\,\mathbf{X}] &= \sigma^2\boldsymbol{\Omega} = \boldsymbol{\Sigma}, \end{aligned} \qquad \textbf{(9-1)}$$

where $\boldsymbol{\Omega}$ is a positive definite matrix. The covariance matrix is written in the form $\sigma^2\boldsymbol{\Omega}$ at several points so that we can obtain the classical model, $\sigma^2\mathbf{I}$ as a convenient special case.

The two leading cases are **heteroscedasticity** and **autocorrelation**. Disturbances are heteroscedastic when they have different variances. Heteroscedasticity arises in numerous applications, in both cross-section and time-series data. Volatile high-frequency time-series data, such as daily observations in financial markets, are heteroscedastic. Heteroscedasticity appears in cross-section data where the scale of the dependent variable and the explanatory power of the model tend to vary across observations. Microeconomic data, such as expenditure surveys, are typical. Even after accounting for firm size, 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.

The disturbances are still assumed to be uncorrelated across observations, so $\sigma^2\boldsymbol{\Omega}$ would be

$$\sigma^2\boldsymbol{\Omega} = \sigma^2\begin{bmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \omega_n \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}.$$

Autocorrelation is usually found in time-series data. Economic time series often display a *memory* in that variation around the regression function is not independent from one period to the next. The seasonally adjusted price and quantity series published by government agencies are examples. Time-series data are usually homoscedastic, so $\sigma^2\mathbf{\Omega}$ might be

$$\sigma^2\mathbf{\Omega} = \sigma^2\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ & & \vdots & \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix}.$$

The values that appear off the diagonal depend on the model used for the disturbance. In most cases, consistent with the notion of a fading memory, the values decline as we move away from the diagonal.

A number of other cases considered later will fit in this framework. **Panel data**, consisting of cross sections observed at several points in time, may exhibit both heteroscedasticity and autocorrelation. In the *random effects model*, $y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + u_i + \varepsilon_{it}$, with $E[\varepsilon_{it}|\mathbf{x}_{it}] = E[u_i|\mathbf{x}_{it}] = 0$, the implication is that

$$\sigma^2\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Gamma} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma} & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Gamma} \end{bmatrix} \text{ where } \mathbf{\Gamma} = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_\varepsilon^2 + \sigma_u^2 & \cdots & \sigma_u^2 \\ & & \vdots & \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_\varepsilon^2 + \sigma_u^2 \end{bmatrix}.$$

The specification exhibits autocorrelation. We shall consider it in Chapter 11. Models of spatial autocorrelation, examined in Chapter 11, and multiple equation regression models, considered in Chapter 10, are also forms of the generalized regression model.

This chapter presents some general results for this extended model. We will focus on the model of heteroscedasticity in this chapter and in Chapter 14. A general model of autocorrelation appears in Chapter 20. Chapters 10 and 11 examine in detail other specific types of generalized regression models. We first consider the consequences for the least squares estimator of the more general form of the regression model. This will include devising an appropriate estimation strategy, still based on least squares. We will then examine alternative estimation approaches that can make better use of the characteristics of the model.

9.2 ROBUST LEAST SQUARES ESTIMATION AND INFERENCE

The generalized regression model in (9-1) drops assumption A.4. If $\mathbf{\Omega} \neq \mathbf{I}$, then the disturbances may be heteroscedastic or autocorrelated or both. The least squares estimator is

$$\mathbf{b} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\textstyle\sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i. \quad \textbf{(9-2)}$$

The covariance matrix of the estimator based on (9-1) and (9-2) would be

$$Var[\mathbf{b}|\mathbf{X}] = \frac{1}{n}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\sigma^2\Sigma_{i=1}^{n}\Sigma_{j=1}^{n}\omega_{ij}\mathbf{x}_i\mathbf{x}_j'}{n}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}$$
$$= \frac{1}{n}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'(\sigma^2\mathbf{\Omega})\mathbf{X}}{n}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}. \quad \textbf{(9-3)}$$

Based on (9-3), we see that $s^2(\mathbf{X}'\mathbf{X})^{-1}$ would not be the appropriate estimator for the asymptotic covariance matrix for the least squares estimator, $\mathbf{b}$. In Section 4.5, we considered a strategy for estimation of the appropriate covariance matrix, without making explicit assumptions about the form of $\boldsymbol{\Omega}$, for two cases, heteroscedasticity and *clustering* (which resembles the random effects model suggested in the Introduction). We will add some detail to that discussion for the heteroscedasticity case. Clustering is revisited in Chapter 11.

The matrix $(\mathbf{X}'\mathbf{X}/n)$ is readily computable using the sample data. The complication is the center matrix that involves the unknown $\sigma^2\boldsymbol{\Omega}$. For estimation purposes, σ^2 is not a separate unknown parameter. We can arbitrarily scale the unknown $\boldsymbol{\Omega}$, say, by κ, and σ^2 by $1/\kappa$ and obtain the same product. We will remove the indeterminacy by assuming that trace $(\boldsymbol{\Omega}) = n$, as it is when $\boldsymbol{\Omega} = \mathbf{I}$. Let $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega}$. It might seem that to estimate $(1/n)\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}$, an estimator of $\boldsymbol{\Sigma}$, which contains $n(n+1)/2$ unknown parameters, is required. But fortunately (because with only n observations, this would be hopeless), this observation is not quite right. What is required is an estimator of the $K(K+1)/2$ unknown elements in the center matrix $\mathbf{Q}_* = \text{plim}\,\dfrac{\mathbf{X}'(\sigma^2\boldsymbol{\Omega})\mathbf{X}}{n} = \text{plim}\,\dfrac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\mathbf{x}_i\mathbf{x}_j'$. The point is that $\mathbf{Q}_*$ is a matrix of sums of squares and cross products that involves σ_{ij} *and* the rows of $\mathbf{X}$. The least squares estimator $\mathbf{b}$ is a consistent estimator of $\boldsymbol{\beta}$, which implies that the least squares residuals e_i are "pointwise" consistent estimators of their population counterparts ε_i. The general approach, then, will be to use $\mathbf{X}$ and $\mathbf{e}$ to devise an estimator of $\mathbf{Q}_*$ for the heteroscedasticity case, $\sigma_{ij} = 0$ when $i \neq j$.

We seek an estimator of $\mathbf{Q}_* = \text{plim}(1/\text{n})\sum_{i=1}^{n}\sigma_i^2\mathbf{x}_i\mathbf{x}_i'$. White (1980, 2001) shows that, under very general conditions, the estimator

$$\mathbf{S}_0 = \frac{1}{n}\sum_{i=1}^{n}e_i^2\mathbf{x}_i\mathbf{x}_i' \tag{9-4}$$

has plim $\mathbf{S}_0 = \mathbf{Q}_*$.[1] The end result is that the **White heteroscedasticity consistent estimator**

$$\begin{aligned}\text{Est.Asy.Var}[\mathbf{b}] &= \frac{1}{n}\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}e_i^2\mathbf{x}_i\mathbf{x}_i'\right)\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\\ &= n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{S}_0(\mathbf{X}'\mathbf{X})^{-1}\end{aligned} \tag{9-5}$$

can be used to estimate the asymptotic covariance matrix of $\mathbf{b}$. This result implies that without actually specifying the type of heteroscedasticity, we can still make appropriate inferences based on the least squares estimator. This implication is especially useful if we are unsure of the precise nature of the heteroscedasticity (which is probably most of the time).

A number of studies have sought to improve on the White estimator for least squares.[2] 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 residuals tend to underestimate the squares of the true disturbances.

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

[2] See, for example, MacKinnon and White (1985) and Messer and White (1984).

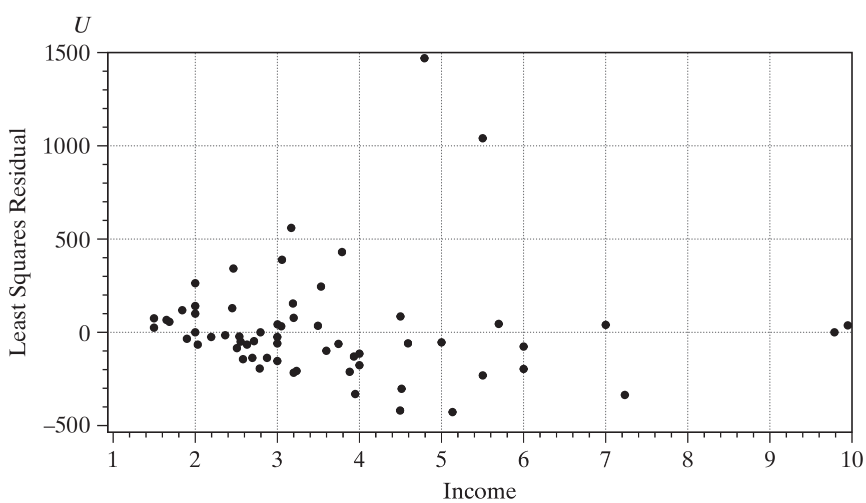
[That is why we use $1/(n - K)$ rather than $1/n$ in computing s^2.] The end result is that in small samples, at least as suggested by some Monte Carlo studies,[3] the White estimator is a bit too optimistic; the matrix is a bit too small, so asymptotic *t* ratios are a little too large. Davidson and MacKinnon (1993) 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 $m_{ii} = 1 - \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$.[4] (See Exercise 9.6.b.) 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 9.1 show the recomputed standard errors with these two modifications.

Example 9.1 Heteroscedastic Regression and the White Estimator

The data in Appendix Table F7.3 give monthly credit card expenditure, for 13,444 individuals. A subsample of 100 observations used here is given in Appendix Table F9.1. The estimates are based on the 72 of these 100 observations for which expenditure is positive. Linear regression of monthly expenditure on a constant, age, income and its square, and a dummy variable for home ownership produces the residuals plotted in Figure 9.1. The pattern of the residuals is characteristic of a regression with heteroscedasticity.

Using White's estimator for the regression produces the results in the row labeled "White S. E." in Table 9.1. The adjustment of the least squares results is fairly large, but the Davidson and MacKinnon corrections to White are, even in this sample of only 72 observations, quite modest. The two income coefficients are individually and jointly statistically significant based on the

FIGURE 9.1 Plot of Residuals against Income.



[3] For example, MacKinnon and White (1985).

[4] This is the standardized residual in (4-69). The authors 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. Note that both $n/(n - K)$ and m_{ii} converge to 1 (quickly). The Davidson and MacKinnon results are strictly small sample considerations.

TABLE 9.1 Least Squares Regression Results

	Constant	*Age*	*OwnRent*	*Income*	*Income²*
Sample mean		31.28	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	−1.19	−0.5590	0.337	2.916	−2.0080
White S.E.	212.99	3.3017	92.188	88.866	6.9446
D. and M. (1)	220.79	3.4227	95.566	92.122	7.1991
D. and M. (2)	221.09	3.4477	95.672	92.084	7.1995

$R^2 = 0.243578$, $s = 284.7508$, R^2 without Income and Income^2 $= 0.06393$.

Mean expenditure = \$262.53, Income is × \$10,000
Tests for heteroscedasticity: White = 14.239, Breusch–Pagan = 49.061, Koenker–Bassett = 7.241.

individual *t* ratios and $F(2, 67) = [(0.244 - 0.064)/2]/[0.756/(72 - 5)] = 7.976$. The 1% critical value is 4.94. (Using the internal digits, the value is 7.956.)

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

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

and the estimated asymptotic covariance matrix is the White estimator. The *F* statistic based on least squares is 7.976. The Wald statistic based on the White estimator is 20.604; the 95% critical value for the chi-squared distribution with two degrees of freedom is 5.99, so the conclusion is unchanged.

9.3 PROPERTIES OF LEAST SQUARES AND INSTRUMENTAL VARIABLES

The essential results for the classical model with $E[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$ and $E[\boldsymbol{\varepsilon\varepsilon}'|\mathbf{X}] = \sigma^2\mathbf{I}$ are developed in Chapters 2 through 6. The least squares estimator

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \tag{9-6}$$

is best linear unbiased (BLU), consistent and asymptotically normally distributed, and if the disturbances are normally distributed, asymptotically efficient. We now consider which of these properties continue to hold in the model of (9-1). To summarize, the least squares estimator retains only some of its desirable properties in this model. It remains unbiased, consistent, and asymptotically normally distributed. It will, however, no longer be efficient and the usual inference procedures based on the *t* and *F* distributions are no longer appropriate.

9.3.1 FINITE-SAMPLE PROPERTIES OF LEAST SQUARES

By taking expectations on both sides of (9-6), we find that if $E[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$, then

$$E[\mathbf{b}] = E_{\mathbf{X}}[E[\mathbf{b}|\mathbf{X}]] = \boldsymbol{\beta} \tag{9-7}$$

and

$$\begin{aligned}\text{Var}[\mathbf{b}|\mathbf{X}] &= E[(\mathbf{b}-\boldsymbol{\beta})(\mathbf{b}-\boldsymbol{\beta})'|\mathbf{X}]\\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}|\mathbf{X}]\\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\boldsymbol{\Omega})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\\ &= \frac{\sigma^2}{n}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{n}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}.\end{aligned} \tag{9-8}$$

Because the variance of the least squares estimator is not $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, statistical inference based on $s^2(\mathbf{X}'\mathbf{X})^{-1}$ may be misleading. There is usually no way to know whether $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ is larger or smaller than the true variance of $\mathbf{b}$ in (9-8). Without Assumption A.4, the familiar inference procedures based on the F and t distributions will no longer be appropriate even if A.6 (normality of $\boldsymbol{\varepsilon}$) is maintained.

THEOREM 9.1 Finite-Sample Properties of b in the Generalized Regression Model

If the regressors and disturbances are uncorrelated, then the least squares estimator is unbiased in the generalized regression model. With nonstochastic regressors, or conditional on $\mathbf{X}$*, the sampling variance of the least squares estimator is given by (9-8). If the regressors are stochastic, then the unconditional variance is* $E_{\mathbf{X}}[\text{Var}[\mathbf{b}|\mathbf{X}]]$. *From (9-6),* $\mathbf{b}$ *is a linear function of* $\boldsymbol{\varepsilon}$. *Therefore, if* $\boldsymbol{\varepsilon}$ *is normally distributed, then* $\mathbf{b}|\mathbf{X} \sim N[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}]$.

9.3.2 ASYMPTOTIC PROPERTIES OF LEAST SQUARES

If $\text{Var}[\mathbf{b}|\mathbf{X}]$ converges to zero, then $\mathbf{b}$ is mean square consistent.[5] With well-behaved regressors, $(\mathbf{X}'\mathbf{X}/n)^{-1}$ will converge to a constant matrix, and σ^2/n will converge to zero. But $(\sigma^2/n)(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}/n)$ need not converge to zero. By writing this product as

$$\frac{\sigma^2}{n}\left(\frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{n}\right) = \left(\frac{\sigma^2}{n}\right)\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}\mathbf{x}_i\mathbf{x}_j'}{n}\right) \tag{9-9}$$

we see that the matrix is a sum of n^2 terms, divided by n. Thus, the product is a scalar that is $O(1/n)$ times a matrix that is $O(n)$ (at least at this juncture) which is $O(1)$. So, it does appear that if the product in (9-9) does converge, it might converge to a matrix of nonzero constants. In this case, the covariance matrix of the least squares estimator would not converge to zero, and consistency would be difficult to establish. We will examine in some detail the conditions under which the matrix in (9-9) converges to a constant matrix. If it does, then because σ^2/n does vanish, least squares is consistent as well as unbiased.

[5] The argument based on the linear projection in Section 4.4.5 cannot be applied here because, unless $\boldsymbol{\Omega} = \mathbf{I}$, $(\mathbf{X}, \mathbf{y})$ cannot be treated as a random sample from a joint distribution.

Consistency will depend on both $\mathbf{X}$ and $\boldsymbol{\Omega}$. A formula that separates the two components is as follows:[6]

1. The smallest characteristic root of $\mathbf{X}'\mathbf{X}$ increases without bound as $n \to \infty$, which implies that $\text{plim}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}$. If the regressors satisfy the Grenander conditions in Table 4.2, then they will meet this requirement.
2. The largest characteristic root of $\boldsymbol{\Omega}$ is finite for all n. For the heteroscedastic model, the variances are the characteristic roots, which requires them to be finite. For models with autocorrelation, the requirements are that the elements of $\boldsymbol{\Omega}$ be finite and that the off-diagonal elements not be too large relative to the diagonal elements. We will examine this condition in Chapter 20.

The least squares estimator is asymptotically normally distributed if the limiting distribution of

$$\sqrt{n}(\mathbf{b} - \boldsymbol{\beta}) = \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} \frac{1}{\sqrt{n}}\mathbf{X}'\boldsymbol{\varepsilon} \tag{9-10}$$

is normal. If $\text{plim}(\mathbf{X}'\mathbf{X}/n) = \mathbf{Q}$, then the limiting distribution of the right-hand side is the same as that of

$$\mathbf{v}_{n,LS} = \mathbf{Q}^{-1}\frac{1}{\sqrt{n}}\mathbf{X}'\boldsymbol{\varepsilon} = \mathbf{Q}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i, \tag{9-11}$$

where $\mathbf{x}_i'$ is a row of $\mathbf{X}$. The question now is whether a central limit theorem can be applied directly to $\mathbf{v}$. If the disturbances are merely heteroscedastic and still uncorrelated, then the answer is generally yes. In fact, we already showed this result in Section 4.4.2 when we invoked the Lindeberg–Feller central limit theorem (D.19) or the Lyapounov theorem (D.20). The theorems allow unequal variances in the sum. *The proof of asymptotic normality* in *Section 4.4.2 is general enough to include this model without modification*. As long as $\mathbf{X}$ is well behaved and the diagonal elements of $\boldsymbol{\Omega}$ are finite and well behaved, the least squares estimator is asymptotically normally distributed, with the covariance matrix given in (9-8). *In the heteroscedastic case, if the variances of ε_i are finite and are not dominated by any single term, so that the conditions of the Lindeberg–Feller central limit theorem apply to $\mathbf{v}_{n,LS}$ in (9-11), then the least squares estimator is asymptotically normally distributed with covariance matrix*

$$\text{Asy.Var}[\mathbf{b}] = \frac{\sigma^2}{n}\mathbf{Q}^{-1}\,\text{plim}\left(\frac{1}{n}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}\right)\mathbf{Q}^{-1}. \tag{9-12}$$

For the most general case, asymptotic normality is much more difficult to establish because the sums in (9-11) are not necessarily sums of independent or even uncorrelated random variables. Nonetheless, Amemiya (1985) and Anderson (1971) have established the asymptotic normality of $\mathbf{b}$ in a model of autocorrelated disturbances general enough to include most of the settings we are likely to meet in practice. We will revisit this issue in Chapter 20 when we examine time-series modeling. We can conclude that, except in particularly unfavorable cases, we have the following theorem.

[6] Amemiya (1985, p. 184).

THEOREM 9.2 Asymptotic Properties of b in the Generalized Regression Model

If $\mathbf{Q} = \text{plim}(\mathbf{X}'\mathbf{X}/n)$ *and* $\text{plim}(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}/n)$ *are both finite positive definite matrices, then* $\mathbf{b}$ *is consistent for* $\boldsymbol{\beta}$. *Under the assumed conditions,* $\text{plim}\ \mathbf{b} = \boldsymbol{\beta}$. *If the regressors are sufficiently well behaved and the off-diagonal terms in* $\boldsymbol{\Omega}$ *diminish sufficiently rapidly, then the least squares estimator is asymptotically normally distributed with mean* $\boldsymbol{\beta}$ *and asymptotic covariance matrix given in (9-12).*

9.3.3 HETEROSCEDASTICITY AND Var[b|X]

In the presence of heteroscedasticity, the least squares estimator $\mathbf{b}$ is still unbiased, consistent, and asymptotically normally distributed. The asymptotic covariance matrix is given in (9-12). For this case, with well-behaved regressors,

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

The mean square consistency of $\mathbf{b}$ depends on the limiting behavior of the matrix

$$\mathbf{Q}_n^* = \frac{1}{n}\sum_{i=1}^{n}\omega_i\mathbf{x}_i\mathbf{x}_i'.$$

If $\mathbf{Q}_n^*$ converges to a positive definite matrix, then as $n \to \infty$, $\mathbf{b}$ 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 $\mathbf{x}$ 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 $\mathbf{Q}$, so it would be surprising if $\mathbf{Q}_n^*$ did not converge as well. In general, then, we would expect that

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

The conventionally estimated covariance matrix for the least squares estimator $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ is inappropriate; the appropriate matrix is $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$. 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. It is easy to show that if $\mathbf{b}$ is consistent for $\boldsymbol{\beta}$, then $\text{plim}\ s^2 = \text{plim}\ \mathbf{e}'\mathbf{e}/(n - K) = \sigma^2$, assuming $\text{tr}(\boldsymbol{\Omega}) = n$. The normalization $\text{tr}(\boldsymbol{\Omega}) = n$ implies that $\sigma^2 = \overline{\sigma}^2 = (1/n)\Sigma_i\sigma_i^2$ and $\omega_i = \sigma_i^2/\overline{\sigma}^2$. Therefore, the least squares estimator, s^2, converges to $\text{plim}\ \overline{\sigma}^2$, that is, the probability limit of the average variance of the disturbances.

The difference between the conventional estimator and the appropriate (true) covariance matrix for $\mathbf{b}$ is

$$\text{Est.Var}[\mathbf{b}|\mathbf{X}] - \text{Var}[\mathbf{b}|\mathbf{X}] = s^2(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}. \quad \textbf{(9-13)}$$

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

$$\mathbf{D} = \frac{\sigma^2}{n}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left[\frac{\mathbf{X}'\mathbf{X}}{n} - \frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{n}\right]\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}. \quad \textbf{(9-14)}$$

The difference between the two matrices hinges on the bracketed matrix,

$$\boldsymbol{\Delta} = \Sigma_{i=1}^{n}(1/n)\mathbf{x}_i\mathbf{x}_i' - \Sigma_{i=1}^{n}(\omega_i/n)\mathbf{x}_i\mathbf{x}_i' = (1/n)\Sigma_{i=1}^{n}(1 - \omega_i)\mathbf{x}_i\mathbf{x}_i', \qquad \textbf{(9-15)}$$

where $\mathbf{x}_i'$ is the ith row of $\mathbf{X}$. These are two weighted averages of the matrices $\mathbf{x}_i\mathbf{x}_i'$ using weights 1 for the first term and ω_i for the second. The scaling $\text{tr}(\boldsymbol{\Omega}) = n$ implies that $\Sigma_i(\omega_i/n) = 1$. 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.

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*.

9.3.4 INSTRUMENTAL VARIABLE ESTIMATION

Chapter 8 considered cases in which the regressors, $\mathbf{X}$, are correlated with the disturbances, $\boldsymbol{\varepsilon}$. The instrumental variables (IV) estimator developed there enjoys a kind of robustness that least squares lacks in that it achieves consistency whether or not $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ are correlated, while $\mathbf{b}$ is neither unbiased nor consistent. However, efficiency was not a consideration in constructing the IV estimator. We will reconsider the IV estimator here, but because it is inefficient to begin with, there is little to say about the implications of (9-1) for the efficiency of the estimator. As such, the relevant question for us to consider here would be, essentially, does IV still work in the generalized regression model. Consistency and asymptotic normality will be the useful properties.

The IV/2SLS estimator is

$$\begin{aligned}\mathbf{b}_{\text{IV}} &= [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}\\ &= [\hat{\mathbf{X}}'\mathbf{X}]^{-1}\hat{\mathbf{X}}'\mathbf{y}\\ &= \boldsymbol{\beta} + [\hat{\mathbf{X}}'\mathbf{X}]^{-1}\hat{\mathbf{X}}'\boldsymbol{\varepsilon}, \qquad \textbf{(9-16)}\end{aligned}$$

where $\mathbf{X}$ is the set of K regressors and $\mathbf{Z}$ is a set of $L \geq K$ instrumental variables. We now consider the extension of Theorem 9.2 to the IV estimator when $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}] = \sigma^2\boldsymbol{\Omega}$. Suppose that $\mathbf{X}$ and $\mathbf{Z}$ are well behaved as assumed in Section 8.2. That is, $(1/n)\mathbf{Z}'\mathbf{Z}$, $(1/n)\mathbf{X}'\mathbf{X}$, and $(1/n)\mathbf{Z}'\mathbf{X}$ all converge to finite nonzero matrices. For convenience let

$$\begin{aligned}\mathbf{Q}_{\mathbf{XX.Z}} &= \text{plim}\left[\left(\frac{1}{n}\mathbf{X}'\mathbf{Z}\right)\left(\frac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\frac{1}{n}\mathbf{Z}'\mathbf{X}\right)\right]^{-1}\left(\frac{1}{n}\mathbf{X}'\mathbf{Z}\right)\left(\frac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1}\\ &= [\mathbf{Q}_{\mathbf{XZ}}\mathbf{Q}_{\mathbf{ZZ}}^{-1}\mathbf{Q}_{\mathbf{ZX}}]^{-1}\mathbf{Q}_{\mathbf{XZ}}\mathbf{Q}_{\mathbf{ZZ}}^{-1}.\end{aligned}$$

If $\mathbf{Z}$ is a valid set of instrumental variables, that is, if the second term in (9-16) vanishes asymptotically, then

$$\text{plim}\,\mathbf{b}_{\text{IV}} = \boldsymbol{\beta} + \mathbf{Q}_{\mathbf{XX.Z}}\,\text{plim}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{\varepsilon}\right) = \boldsymbol{\beta}.$$

The large sample behavior of $\mathbf{b}_{\text{IV}}$ depends on the behavior of

$$\mathbf{v}_{n,\,\mathrm{IV}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\varepsilon_i.$$

This result is exactly the one we analyzed in Section 4.4.2. If the sampling distribution of $\mathbf{v}_n$ converges to a normal distribution, then we will be able to construct the asymptotic distribution for $\mathbf{b}_{\mathrm{IV}}$. This set of conditions is the same that was necessary for $\mathbf{X}$ when we considered $\mathbf{b}$ above, with $\mathbf{Z}$ in place of $\mathbf{X}$. We will rely on the results of Anderson (1971) or Amemiya (1985) that, under very general conditions,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\varepsilon_i \xrightarrow{d} N\left[\mathbf{0}, \sigma^2 \operatorname{plim}\left(\frac{1}{n}\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)\right].$$

With the other results already in hand, we now have the following.

THEOREM 9.3 Asymptotic Properties of the IV Estimator in the Generalized Regression Model

If the regressors and the instrumental variables are well behaved in the fashions discussed above, then $\mathbf{b}_{\mathrm{IV}}$ is consistent and asymptotically normally distributed with

$$\mathbf{b}_{\mathrm{IV}} \overset{a}{\sim} N[\boldsymbol{\beta}, \mathbf{V}_{\mathrm{IV}}],$$

where

$$\mathbf{V}_{\mathrm{IV}} = \frac{\sigma^2}{n}(\mathbf{Q}_{\mathbf{XX.Z}}) \operatorname{plim}\left(\frac{1}{n}\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)(\mathbf{Q}'_{\mathbf{XX.Z}}).$$

9.4 EFFICIENT ESTIMATION BY GENERALIZED LEAST SQUARES

Efficient estimation of $\boldsymbol{\beta}$ in the generalized regression model requires knowledge of $\boldsymbol{\Omega}$. To begin, it is useful to consider cases in which $\boldsymbol{\Omega}$ is a known, symmetric, positive definite matrix. This assumption will occasionally be true, though in most models $\boldsymbol{\Omega}$ will contain unknown parameters that must also be estimated. We shall examine this case in Section 9.4.2.

9.4.1 GENERALIZED LEAST SQUARES (GLS)

Because $\boldsymbol{\Omega}$ is a positive definite symmetric matrix, it can be factored into

$$\boldsymbol{\Omega} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}',$$

where the columns of $\mathbf{C}$ are the characteristic vectors of $\boldsymbol{\Omega}$ and the characteristic roots of $\boldsymbol{\Omega}$ are arrayed in the diagonal matrix, $\boldsymbol{\Lambda}$. Let $\boldsymbol{\Lambda}^{1/2}$ be the diagonal matrix with ith diagonal element, $\sqrt{\lambda_i}$, and let $\mathbf{T} = \mathbf{C}\boldsymbol{\Lambda}^{1/2}$. Then, $\boldsymbol{\Omega} = \mathbf{T}\mathbf{T}'$. Also, let $\mathbf{P}' = \mathbf{C}\boldsymbol{\Lambda}^{-1/2}$, so $\boldsymbol{\Omega}^{-1} = \mathbf{P}'\mathbf{P}$. Premultiply the model in (9-1) by $\mathbf{P}$ to obtain

$$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon}$$

or

$$\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*. \tag{9-17}$$

The conditional variance of $\boldsymbol{\varepsilon}_*$ is

$$E[\boldsymbol{\varepsilon}_*\boldsymbol{\varepsilon}_*' \mid \mathbf{X}_*] = \mathbf{P}\sigma^2\boldsymbol{\Omega}\mathbf{P}' = \sigma^2\mathbf{I},$$

so the classical regression model applies to this transformed model. Because $\boldsymbol{\Omega}$ is assumed to be known, $\mathbf{y}_*$ and $\mathbf{X}_*$ are observed data. In the classical model, ordinary least squares is efficient; hence,

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{X}_*'\mathbf{y}_*\\ &= (\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{y}\\ &= (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}\end{aligned}$$

is the **efficient estimator** of $\boldsymbol{\beta}$. This estimator is the **generalized least squares (GLS)** or Aitken (1935) estimator of $\boldsymbol{\beta}$. This estimator is in contrast to the **ordinary least squares (OLS)** estimator, which uses a *weighting matrix*, $\mathbf{I}$, instead of $\boldsymbol{\Omega}^{-1}$. By appealing to the classical regression model in (9-17), we have the following theorem, which includes the generalized regression model analogs to our results of Chapter 4:

THEOREM 9.4 Properties of the Generalized Least Squares Estimator

If $E[\boldsymbol{\varepsilon}_* \mid \mathbf{X}_*] = \mathbf{0}$, *then*

$$E[\hat{\boldsymbol{\beta}} \mid \mathbf{X}_*] = E[(\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{X}_*'\mathbf{y}_* \mid \mathbf{X}_*] = \boldsymbol{\beta} + E[(\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{X}_*'\boldsymbol{\varepsilon}_* \mid \mathbf{X}_*] = \boldsymbol{\beta}.$$

The GLS estimator $\hat{\boldsymbol{\beta}}$ *is unbiased. This result is equivalent to* $E[\mathbf{P}\boldsymbol{\varepsilon} \mid \mathbf{P}\mathbf{X}] = \mathbf{0}$, *but because* $\mathbf{P}$ *is a matrix of known constants, we return to the familiar requirement* $E[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}$. *The requirement that the regressors and disturbances be uncorrelated is unchanged.*

The GLS estimator is consistent if $\text{plim}(1/n)\mathbf{X}_*'\mathbf{X}_* = \mathbf{Q}_*$, *where* $\mathbf{Q}_*$ *is a finite positive definite matrix. Making the substitution, we see that this implies*

$$\text{plim}[(1/n)\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1} = \mathbf{Q}_*^{-1}. \tag{9-18}$$

We require the transformed data $\mathbf{X}_* = \mathbf{P}\mathbf{X}$, *not the original data* $\mathbf{X}$, *to be well behaved.*[7] *Under the assumption in (9-1), the following hold:*

The GLS estimator is asymptotically normally distributed, with mean $\boldsymbol{\beta}$ *and sampling variance*

$$\text{Var}[\hat{\boldsymbol{\beta}} \mid \mathbf{X}_*] = \sigma^2(\mathbf{X}_*'\mathbf{X}_*)^{-1} = \sigma^2(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}. \tag{9-19}$$

The GLS estimator $\hat{\boldsymbol{\beta}}$ *is the minimum variance linear unbiased estimator in the generalized regression model. This statement follows by applying the Gauss–Markov theorem to the model in (9-17). The result in Theorem 9.5 is* ***Aitken's theorem*** *(1935), and* $\hat{\boldsymbol{\beta}}$ *is sometimes called the Aitken estimator. This broad result includes the Gauss–Markov theorem as a special case when* $\boldsymbol{\Omega} = \mathbf{I}$.

[7] Once again, to allow a time trend, we could weaken this assumption a bit.

For testing hypotheses, we can apply the full set of results in Chapter 5 to the transformed model in (9-17). For testing the J linear restrictions, $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, the appropriate statistic is

$$F[J, n - K] = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}\hat{\sigma}^2(\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})}{J} = \frac{(\hat{\boldsymbol{\varepsilon}}_c'\hat{\boldsymbol{\varepsilon}}_c - \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}})/J}{\hat{\sigma}^2},$$

where the residual vector is

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}}$$

and

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n - K} = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - K}. \quad \textbf{(9-20)}$$

The constrained GLS residuals, $\hat{\boldsymbol{\varepsilon}}_c = \mathbf{y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}}_c$, are based on

$$\hat{\boldsymbol{\beta}}_c = \hat{\boldsymbol{\beta}} - [\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}).$$ [8]

To summarize, all the results for the classical model, including the usual inference procedures, apply to the transformed model in (9-17).

There is no precise counterpart to R^2 in the generalized regression model. Alternatives have been proposed, but care must be taken when using them. For example, one choice is the R^2 in the transformed regression, (9-17). But this regression need not have a constant term, so the R^2 is not bounded by zero and one. Even if there is a constant term, the transformed regression is a computational device, not the model of interest. That a good (or bad) fit is obtained in the model in (9-17) may be of no interest; the dependent variable in that model, y_*, is different from the one in the model as originally specified. The usual R^2 often suggests that the fit of the model is improved by a correction for heteroscedasticity and degraded by a correction for autocorrelation, but both changes can often be attributed to the computation of y_*. A more appealing fit measure might be based on the residuals from the original model once the GLS estimator is in hand, such as

$$R_G^2 = 1 - \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Like the earlier contender, however, this measure is not bounded in the unit interval. In addition, this measure cannot be reliably used to compare models. The generalized least squares estimator minimizes the **generalized sum of squares**

$$\boldsymbol{\varepsilon}_*'\boldsymbol{\varepsilon}_* = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

not $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$. As such, there is no assurance, for example, that dropping a variable from the model will result in a decrease in R_G^2, as it will in R^2. Other goodness-of-fit measures, designed primarily to be a function of the sum of squared residuals (raw or weighted by $\boldsymbol{\Omega}^{-1}$) and to be bounded by zero and one, have been proposed.[9] Unfortunately, they all suffer from at least one of the previously noted shortcomings. The R^2-like measures in

[8] Note that this estimator is the constrained OLS estimator using the transformed data. [See (5-23).]

[9] See Judge et al. (1985, p. 32) and Buse (1973).

this setting are purely descriptive. That being the case, the squared sample correlation between the actual and predicted values, $r^2_{y,\hat{y}} = corr^2(y, \hat{y}) = corr^2(y, \mathbf{x}'\hat{\beta})$, would likely be a useful descriptor. Note, though, that this is not a proportion of variation explained, as is R^2; it is a measure of the agreement of the model predictions with the actual data.

9.4.2 FEASIBLE GENERALIZED LEAST SQUARES (FGLS)

To use the results of Section 9.4.1, $\mathbf{\Omega}$ must be known. If $\mathbf{\Omega}$ contains unknown parameters that must be estimated, then generalized least squares is not feasible. But with an unrestricted $\mathbf{\Omega}$, there are $n(n + 1)/2$ additional parameters in $\sigma^2\mathbf{\Omega}$. This number is far too many to estimate with n observations. Obviously, some structure must be imposed on the model if we are to proceed.

The typical problem involves a small set of parameters $\boldsymbol{\alpha}$ such that $\mathbf{\Omega} = \mathbf{\Omega}(\boldsymbol{\alpha})$. For example, a commonly used formula in time-series settings is

$$\mathbf{\Omega}(\rho) = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{n-2} \\ & & & & \vdots & \\ \rho^{n-1} & \rho^{n-2} & & \cdots & & 1 \end{bmatrix},$$

which involves only one additional unknown parameter. A model of heteroscedasticity that also has only one new parameter is

$$\sigma_i^2 = \sigma^2 z_i^{\theta} \qquad \textbf{(9-21)}$$

for some exogenous variable z. Suppose, then, that $\hat{\boldsymbol{\alpha}}$ is a consistent estimator of $\boldsymbol{\alpha}$. (We consider later how such an estimator might be obtained.) To make GLS estimation feasible, we shall use $\hat{\mathbf{\Omega}} = \mathbf{\Omega}(\hat{\boldsymbol{\alpha}})$ instead of the true $\mathbf{\Omega}$. The issue we consider here is whether using $\mathbf{\Omega}(\hat{\boldsymbol{\alpha}})$ requires us to change any of the results of Section 9.4.1.

It would seem that if plim $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$, then using $\hat{\mathbf{\Omega}}$ is asymptotically equivalent to using the true $\mathbf{\Omega}$.[10] Let the **feasible generalized least squares** estimator be denoted

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

Conditions that imply that $\hat{\hat{\boldsymbol{\beta}}}$ is asymptotically equivalent to $\hat{\boldsymbol{\beta}}$ are

$$\text{plim}\left[\left(\frac{1}{n}\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X}\right) - \left(\frac{1}{n}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}\right)\right] = \mathbf{0} \qquad \textbf{(9-22)}$$

and

$$\text{plim}\left[\left(\frac{1}{\sqrt{n}}\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\boldsymbol{\varepsilon}\right) - \left(\frac{1}{\sqrt{n}}\mathbf{X}'\mathbf{\Omega}^{-1}\boldsymbol{\varepsilon}\right)\right] = \mathbf{0}. \qquad \textbf{(9-23)}$$

The first of these equations states that if the weighted sum of squares matrix based on the true $\mathbf{\Omega}$ converges to a positive definite matrix, then the one based on $\hat{\mathbf{\Omega}}$ converges to the same matrix. We are assuming that this is true. In the second condition, if the

[10] This equation is sometimes denoted plim $\hat{\mathbf{\Omega}} - \mathbf{\Omega}$. Because $\mathbf{\Omega}$ is $n \times n$, it cannot have a probability limit. We use this term to indicate convergence element by element.

transformed regressors are well behaved, then the right-hand-side sum will have a limiting normal distribution. This condition is exactly the one we used in Chapter 4 to obtain the asymptotic distribution of the least squares estimator; here we are using the same results for $\mathbf{X}_*$ and $\boldsymbol{\varepsilon}_*$. Therefore, (9-23) requires the same condition to hold when $\boldsymbol{\Omega}$ is replaced with $\hat{\boldsymbol{\Omega}}$.[11]

These conditions, in principle, must be verified on a case-by-case basis. Fortunately, in most familiar settings, they are met. If we assume that they are, then the FGLS estimator based on $\hat{\boldsymbol{\alpha}}$ has the same **asymptotic properties** as the GLS estimator. This result is extremely useful. Note, especially, the following theorem.

THEOREM 9.5 Efficiency of the FGLS Estimator
An asymptotically efficient FGLS estimator does not require that we have an efficient estimator of $\boldsymbol{\alpha}$; *only a consistent one is required to achieve full efficiency for the FGLS estimator.*

Except for the simplest cases, the **finite-sample properties** and exact distributions of FGLS estimators are unknown. The asymptotic efficiency of FGLS estimators may not carry over to small samples because of the variability introduced by the estimated $\boldsymbol{\Omega}$. Some analyses for the case of heteroscedasticity are given by Taylor (1977). A model of autocorrelation is analyzed by Griliches and Rao (1969). In both studies, the authors find that, over a broad range of parameters, FGLS is more efficient than least squares. But if the departure from the classical assumptions is not too severe, then least squares may be more efficient than FGLS in a small sample.

9.5 HETEROSCEDASTICITY AND WEIGHTED LEAST SQUARES

In the heteroscedastic regression model,

$$\text{Var}[\varepsilon_i | \mathbf{X}] = \sigma_i^2 = \sigma^2 \omega_i, \quad i = 1, \ldots, n.$$

This form is an arbitrary scaling which allows us to use a normalization, $\text{trace}(\boldsymbol{\Omega}) = \Sigma_i \omega_i = n$. 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.

We will examine the heteroscedastic regression model, first in general terms, then with some specific forms of the disturbance covariance matrix. Specification tests for heteroscedasticity are considered in Section 9.6. Section 9.6 considers generalized (weighted) least squares, which requires knowledge at least of the form of $\boldsymbol{\Omega}$. Finally, two common applications are examined in Section 9.7.

[11] The condition actually requires only that if the right-hand-side sum has *any* limiting distribution, then the left-hand one has the same one. Conceivably, this distribution might not be the normal distribution, but that seems unlikely except in a specially constructed, theoretical case.

9.5.1 WEIGHTED LEAST SQUARES

The GLS estimator is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}. \tag{9-24}$$

In the most general case, $\text{Var}[\varepsilon_i|\mathbf{X}] = \sigma_i^2 = \sigma^2\omega_i$, $\boldsymbol{\Omega}^{-1}$ 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** estimator.

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n} w_i\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{n} w_i\mathbf{x}_i y_i\right], \tag{9-25}$$

where $w_i = 1/\omega_i$.[12] 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.

9.5.2 WEIGHTED LEAST SQUARES WITH KNOWN Ω

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

$$\frac{y}{x_k} = \beta_k + \beta_1\left(\frac{x_1}{x_k}\right) + \beta_2\left(\frac{x_2}{x_k}\right) + \cdots + \frac{\varepsilon}{x_k}. \tag{9-26}$$

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 (9-26), 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

[12] 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^2s cancel, leaving the expression given previously.

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

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

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{9-27}$$

is

$$\text{Asy.Var}[\hat{\boldsymbol{\beta}}] = \sigma^2[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{\Omega}\mathbf{V}^{-1}\mathbf{X}[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}. \tag{9-28}$$

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 (9-5).

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 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 mentioned 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. However, the weights we use might well be better than none. A heteroscedasticity robust estimator for weighted least squares can be formed by combining (9-27) with the White estimator. The weighted least squares estimator in (9-27) 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

$$\text{Est.Asy.Var}[\hat{\boldsymbol{\beta}}] = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\left[\sum_{i=1}^{n}\left(\frac{e_i^2}{v_i^2}\right)\mathbf{x}_i\mathbf{x}_i'\right](\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{9-29}$$

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

9.5.3 ESTIMATION WHEN $\boldsymbol{\Omega}$ 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. 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. We consider the two-step estimator here and the maximum likelihood estimator in Chapter 14.

For the heteroscedastic model, the GLS estimator is

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n}\left(\frac{1}{\sigma_i^2}\right)\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{n}\left(\frac{1}{\sigma_i^2}\right)\mathbf{x}_i y_i\right]. \tag{9-30}$$

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 (9-30) 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 $E[\varepsilon_i^2|\mathbf{z}_i] = \sigma_i^2$, so $\varepsilon_i^2 = \sigma_i^2 + v_i$, where v_i is just the difference between ε_i^2 and its conditional expectation. Because 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,[13]

$$e_i^2 = \sigma_i^2 + v_i^*. \tag{9-31}$$

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.[14] For example, if $\sigma_i^2 = \mathbf{z}_i'\boldsymbol{\alpha}$, then a consistent estimator of $\boldsymbol{\alpha}$ will be the least squares slopes, $\mathbf{a}$, in the "model,"

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

In this model, v_i^* is both heteroscedastic and autocorrelated, so $\mathbf{a}$ 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 $\boldsymbol{\beta}$.[15]

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.

9.6 TESTING FOR HETEROSCEDASTICITY

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 general, be applied to the ordinary least squares residuals.

[13] See Amemiya (1985) and Harvey (1976) for formal analyses.

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

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

9.6.1 WHITE'S GENERAL TEST

To formulate the available tests, it is necessary to specify, at least in rough terms, the nature of the heteroscedasticity. White's (1980) test proposes a general hypothesis of the form

$$H_0 : \sigma_i^2 = E[\varepsilon_i^2 | \mathbf{x}_i] = \sigma^2 \text{ for all } i,$$
$$H_1 : Not\ H_0.$$

A simple operational version of his test is carried out by obtaining nR^2 in the regression of the squared OLS residuals, e_i^2, on a constant and all unique variables contained in $\mathbf{x}$ and $\mathbf{x} \otimes \mathbf{x}$. The statistic has a limiting chi-squared distribution with $P - 1$ degrees of freedom, where P is the number of regressors in the equation, including the constant. An equivalent approach is to use an F test to test the hypothesis that $\boldsymbol{\gamma}_1 = \mathbf{0}$ and $\boldsymbol{\gamma}_2 = \mathbf{0}$ in the regression

$$e_i^2 = \gamma_0 + \mathbf{x}_i'\boldsymbol{\gamma}_1 + (\mathbf{x}_i \otimes \mathbf{x}_i)'\boldsymbol{\gamma}_2 + v_i^*.$$

[As before, $(\mathbf{x}_i \otimes \mathbf{x}_i)$ contains only the unique components.] The **White test** is extremely general. To carry it out, we need not make any specific assumptions about the nature of the heteroscedasticity.

9.6.2 THE LAGRANGE MULTIPLIER TEST

Breusch and Pagan (1979) and Godfrey (1988) present 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. The disturbance is homoscedastic if $\boldsymbol{\alpha} = \mathbf{0}$. The test can be carried out with a simple regression:

$$\text{LM} = \tfrac{1}{2} \times \text{explained sum of squared residuals in the regression of } e_i^2/(\mathbf{e}'\mathbf{e}/n) \text{ on } (1, \mathbf{z}_i). \quad \textbf{(9-32)}$$

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 LM $= (1/2)[\mathbf{g}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{g}]$. Under the null hypothesis of homoscedasticity, LM has a limiting chi-squared distribution with $P - 1$ degrees of freedom.

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,

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

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 LM $= (1/V)(\mathbf{u} - \bar{u}\,\mathbf{i})'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{u} - \bar{u}\,\mathbf{i})$. Under normality, this modified statistic will have the same limiting 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 9.2 Testing for Heteroscedasticity

We use the suggested diagnostics to test for heteroscedasticity in the credit card expenditure data in Example 9.2.

1. **White's Test:** There are 15 variables in ($\mathbf{x}, \mathbf{x} \otimes \mathbf{x}$), including the constant term. But because OwnRent^2 = OwnRent and Income × Income = Income^2, which is also in the equation, only 13 of the 15 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% critical value of chi-squared with 12 degrees of freedom is 21.03, so despite what might seem to be obvious in Figure 9.1, the hypothesis of homoscedasticity is not rejected by this test.
2. **Breusch–Pagan Test:** This test requires a specific alternative hypothesis. For this purpose, we specify the test based on $\mathbf{z} = [1, \text{Income}, \text{Income}^2]$. 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 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. If the Breusch and Pagan test is based on (1, x), the chi squared statistic is 49.061 with 4 degrees of freedom, while the Koenker and Bassett version is 7.241. The same conclusions are reached.

9.7 TWO 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.

9.7.1 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

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

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 $\mathbf{z}_i$. The groupwise heteroscedasticity model described in Example 9.4 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, $\sigma_g^2 = \sigma^2 \exp(\alpha_g)$.

Example 9.3 Multiplicative Heteroscedasticity

In Example 6.6, we fit a cost function for the U.S. airline industry of the form

$$\ln C_{it} = \beta_1 + \beta_2 \ln Q_{it} + \beta_3(\ln Q_{it})^2 + \beta_4 \ln P_{fuel,i,t} + \beta_5\, Loadfactor_{i,t} + \varepsilon_{i,t}$$

where $C_{i,t}$ is total cost, $Q_{i,t}$ is output, and $P_{fuel,i,t}$ is the price of fuel, and the 90 observations in the data set are for six firms observed for 15 years. (The model also included dummy variables for firm and year, which we will omit for simplicity.) We now consider a revised model in which the load factor appears in the variance of $\varepsilon_{i,t}$ rather than in the regression function. The model is

$$\sigma_{i,t}^2 = \sigma^2 \exp(\gamma \, Loadfactor_{i,t})$$
$$= \exp(\gamma_1 + \gamma_2 \, Loadfactor_{i,t}).$$

The constant in the implied regression is $\gamma_1 = \ln \sigma^2$. Figure 9.2 shows a plot of the least squares residuals against *Loadfactor* for the 90 observations. The figure does suggest the presence of heteroscedasticity. (The dashed lines are placed to highlight the effect.) We computed the LM statistic using (9-32). The chi-squared statistic is 2.959. This is smaller than the critical value of 3.84 for one degree of freedom, so on this basis, the null hypothesis of homoscedasticity with respect to the load factor is not rejected.

To begin, we use OLS to estimate the parameters of the cost function and the set of residuals, $e_{i,t}$. Regression of $\log(e_{it}^2)$ on a constant and the load factor provides estimates of γ_1 and γ_2, denoted c_1 and c_2. The results are shown in Table 9.2. As Harvey notes, $\exp(c_1)$ does not necessarily estimate σ^2 consistently—for normally distributed disturbances, it is low by a factor of 1.2704. However, as seen in (9-24), the estimate of σ^2 (biased or otherwise) is not needed to compute the FGLS estimator. Weights $w_{i,t} = \exp(-c_1 - c_2 \, Loadfactor_{i,t})$ are computed using these estimates, then weighted least squares using (9-25) is used to obtain the FGLS estimates of $\boldsymbol{\beta}$. The results of the computations are shown in Table 9.2.

We might consider iterating the procedure. Using the results of FGLS at step 2, we can recompute the residuals, then recompute c_1 and c_2 and the weights, and then reenter the iteration. The process converges when the estimate of c_2 stabilizes. This requires seven iterations. The results are shown in Table 9.2. As noted earlier, iteration does not produce any gains here. The second step estimator is already fully efficient. Moreover, this does not produce the MLE, either. That would be obtained by regressing $[e_{i,t}^2/\exp(c_1 + c_2 \, Loadfactor_{i,t}) - 1]$ on the constant and load factor at each iteration to obtain the new estimates. We will revisit this in Chapter 14.

FIGURE 9.2 Plot of Residuals against Load Factor.

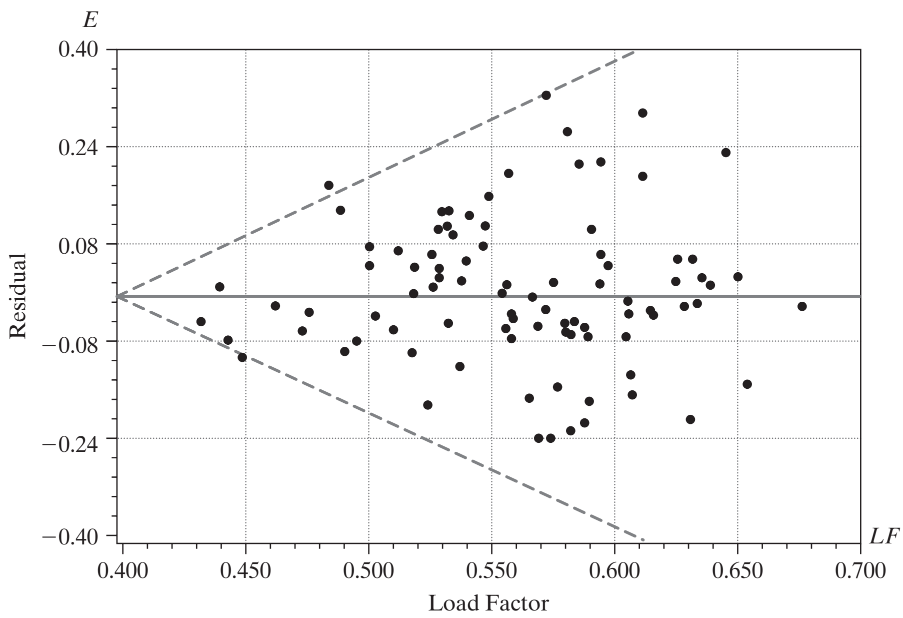


TABLE 9.2 Multiplicative Heteroscedasticity Model

	Constant	**ln** Q	**ln**^{2} Q	**ln** P_f	R^2	*Sum of Squares*
OLS	9.1382	0.92615	0.029145	0.41006		
	0.24507[a]	0.032306	0.012304	0.018807	0.9861674[b]	1.577479[c]
	0.22595[d]	0.030128	0.011346	0.017524		
Two step	9.2463	0.92136	0.024450	0.40352	0.986119	1.612938
	0.21896	0.033028	0.011412	0.016974		
Iterated[e]	9.2774	0.91609	0.021643	0.40174	0.986071	1.645693
	0.20977	0.032993	0.011017	0.016332		

[a]Conventional OLS standard errors
[b]Squared correlation between actual and fitted values
[c]Sum of squared residuals
[d]White robust standard errors
[e]Values of c_2 by iteration: 8.254344, 11.622473, 11.705029, 11.710618, 11.711012, 11.711040, 11.711042

9.7.2 GROUPWISE HETEROSCEDASTICITY

A groupwise heteroscedastic regression has the structural equations

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

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

$$\hat{\boldsymbol{\beta}} = \left[\sum_{g=1}^{G}\left(\frac{1}{\sigma_g^2}\right)\mathbf{X}_g'\mathbf{X}_g\right]^{-1}\left[\sum_{g=1}^{G}\left(\frac{1}{\sigma_g^2}\right)\mathbf{X}_g'\mathbf{y}_g\right]. \qquad \textbf{(9-33)}$$

Because $\mathbf{X}_g'\mathbf{y}_g = \mathbf{X}_g'\mathbf{X}_g\mathbf{b}_g$, 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 $\mathbf{W}_g = \left[\sum_{g=1}^{G}(\text{Var}[\mathbf{b}_g \mid \mathbf{X}_g])^{-1}\right]^{-1}(\text{Var}[\mathbf{b}_g \mid \mathbf{X}_g])^{-1}$. 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, $w_g\mathbf{I} = \left(h_g/\sum_g h_g\right)\mathbf{I}$ 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 (9-33), 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,

$$\hat{\sigma}_g^2 = \frac{\mathbf{e}_g'\mathbf{e}_g}{n_g}, \tag{9-34}$$

provides the needed estimator for the group-specific disturbance variance. Thereafter, (9-33) 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 (9-34) with the two-step FGLS estimator, recomputing the weights, then returning to (9-33) to recompute the slope vector. This can be continued until convergence. It can be shown that so long as (9-34) 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).[16]

For testing the homoscedasticity assumption, both White's test and the LM test are straightforward. The variables thought to enter the conditional variance are simply a set of $G - 1$ group dummy variables, not including one of them (to avoid the dummy variable trap), which we'll denote $\mathbf{Z}^*$. Because the columns of $\mathbf{Z}^*$ are binary and orthogonal, to carry out White's test, we need only regress the squared least squares residuals on a constant and $\mathbf{Z}^*$ and compute NR^2 where $N = \sum_g n_g$. The LM test is also straightforward. For purposes of this application of the LM test, it will prove convenient to replace the overall constant in $\mathbf{Z}$ in (9-32) with the remaining group dummy variable. Because the column space of the full set of dummy variables is the same as that of a constant and $G - 1$ of them, all results that follow will be identical. In (9-32), the vector $\mathbf{g}$ will now be G subvectors where each subvector is the n_g elements of $[(e_{ig}^2/\hat{\sigma}^2) - 1]$, and $\hat{\sigma}^2 = \mathbf{e}'\mathbf{e}/N$. By multiplying it out, we find that $\mathbf{g}'\mathbf{Z}$ is the G vector with elements $n_g[(\hat{\sigma}_g^2/\hat{\sigma}^2) - 1]$, while $(\mathbf{Z}'\mathbf{Z})^{-1}$ is the $G \times G$ matrix with diagonal elements $1/n_g$. It follows that

$$LM = \frac{1}{2}\mathbf{g}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{g} = \frac{1}{2}\sum_{g=1}^{G} n_g\left(\frac{\hat{\sigma}_g^2}{\hat{\sigma}^2} - 1\right)^2. \tag{9-35}$$

Both statistics have limiting chi-squared distributions with $G - 1$ degrees of freedom under the null hypothesis of homoscedasticity. (There are only $G - 1$ degrees of freedom because the hypothesis imposes $G - 1$ restrictions, that the G variances are all equal to each other. Implicitly, one of the variances is free and the other $G - 1$ equal that one.)

Example 9.4 Groupwise Heteroscedasticity

Baltagi and Griffin (1983) is a study of gasoline usage in 18 of the 30 OECD countries. The model analyzed in the paper is

$$\begin{aligned}\ln(\textit{Gasoline usage/car})_{i,t} = \beta_1 &+ \beta_2 \ln(\textit{Per capita income})_{i,t} + \beta_3 \ln \textit{Price}_{i,t} \\ &+ \beta_4 \ln(\textit{Cars per capita})_{i,t} + \varepsilon_{i,t},\end{aligned}$$

[16] See Oberhofer and Kmenta (1974).

where i = country and t = 1960, . . . , 1978. This is a balanced panel (see Section 11.2) with 19(18) = 342 observations in total. The data are given in Appendix Table F9.2.

Figure 9.3 displays the OLS residuals using the least squares estimates of the model above with the addition of 18 country dummy variables (1 to 18) (and without the overall constant). (The country dummy variables are used so that the country-specific residuals will have mean zero.) The F statistic for testing the null hypothesis that all the constants are equal is

$$F[(G-1),(\Sigma_{g=1}^{G} n_g - K - G)] = \frac{(\mathbf{e}_0'\mathbf{e}_0 - \mathbf{e}_1'\mathbf{e}_1)/(G-1)}{\mathbf{e}_1'\mathbf{e}_1/(\Sigma_{g=1}^{G} n_g - K - G)}$$

$$= \frac{(14.90436 - 2.73649)/17}{2.73649/(342 - 3 - 18)} = 83.960798,$$

where $\mathbf{e}_0$ is the vector of residuals in the regression with a single constant term and $\mathbf{e}_1$ is the regression with country-specific constan t terms. The critical value from the F table with 17 and 321 degrees of freedom is 1.655. The regression results are given in Table 9.3. Figure 9.3 does convincingly suggest the presence of groupwise heteroscedasticity. The White and LM statistics are 342(0.38365) = 131.21 and 279.588, respectively. The critical value from the chi-squared distribution with 17 degrees of freedom is 27.587. So, we reject the hypothesis of homoscedasticity and proceed to fit the model by feasible GLS. The two-step estimates are shown in Table 9.3. The FGLS estimator is computed by using weighted least squares, where the weights are $1/\hat{\sigma}_g^2$ for each observation in country g. Comparing the White standard errors to the two-step estimators, we see that in this instance, there is a substantial gain to using feasible generalized least squares.

FIGURE 9.3 Plot of OLS Residuals by Country.

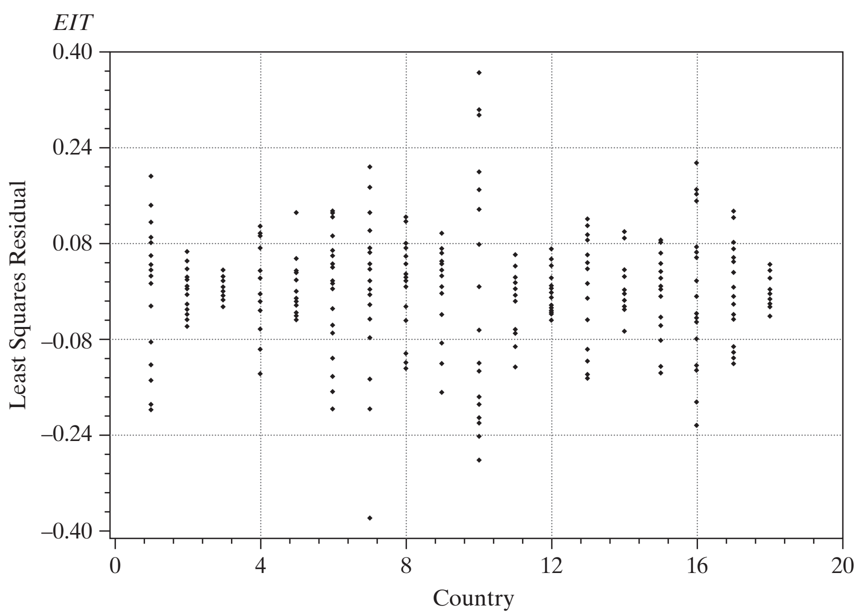


TABLE 9.3 Estimated Gasoline Consumption Equations

	OLS			*FGLS*	
	Coefficient	***Std. Error***	***White Std. Err.***	***Coefficient***	***Std. Error***
ln *Income*	0.66225	0.07339	0.07277	0.57507	0.02927
ln *Price*	−0.32170	0.04410	0.05381	−0.27967	0.03519
ln *Cars/Cap.*	−0.64048	0.02968	0.03876	−0.56540	0.01613
Country 1	2.28586	0.22832	0.22608	2.43707	0.11308
Country 2	2.16555	0.21290	0.20983	2.31699	0.10225
Country 3	3.04184	0.21864	0.22479	3.20652	0.11663
Country 4	2.38946	0.20809	0.20783	2.54707	0.10250
Country 5	2.20477	0.21647	0.21087	2.33862	0.10101
Country 6	2.14987	0.21788	0.21846	2.30066	0.10893
Country 7	2.33711	0.21488	0.21801	2.57209	0.11206
Country 8	2.59233	0.24369	0.23470	2.72376	0.11384
Country 9	2.23255	0.23954	0.22973	2.34805	0.10795
Country 10	2.37593	0.21184	0.22643	2.58988	0.11821
Country 11	2.23479	0.21417	0.21311	2.39619	0.10478
Country 12	2.21670	0.20304	0.20300	2.38486	0.09950
Country 13	1.68178	0.16246	0.17133	1.90306	0.08146
Country 14	3.02634	0.39451	0.39180	3.07825	0.20407
Country 15	2.40250	0.22909	0.23280	2.56490	0.11895
Country 16	2.50999	0.23566	0.26168	2.82345	0.13326
Country 17	2.34545	0.22728	0.22322	2.48214	0.10955
Country 18	3.05525	0.21960	0.22705	3.21519	0.11917

9.8 SUMMARY AND CONCLUSIONS

This chapter has introduced a major extension of the classical linear model. By allowing for heteroscedasticity and autocorrelation in the disturbances, we expand the range of models to a large array of frameworks. We will explore these in the next several chapters. The formal concepts introduced in this chapter include how this extension affects the properties of the least squares estimator, how an appropriate estimator of the asymptotic covariance matrix of the least squares estimator can be computed in this extended modeling framework, and, finally, how to use the information about the variances and covariances of the disturbances to obtain an estimator that is more efficient than ordinary least squares.

We have analyzed in detail one form of the generalized regression model, the model of heteroscedasticity. We first considered least squares estimation. The primary result for 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. After examining two general tests for heteroscedasticity, we then narrowed the model to some specific parametric forms, and considered weighted (generalized) least squares for efficient estimation 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.

Key Terms and Concepts

- Asymptotic properties
- Autocorrelation
- Breusch–Pagan Lagrange multiplier test
- Efficient estimator
- Feasible generalized least squares (FGLS)
- Finite-sample properties
- Generalized least squares (GLS)
- Generalized linear regression model
- Generalized sum of squares
- Groupwise heteroscedasticity
- Heteroscedasticity
- Lagrange multiplier test
- Multiplicative heteroscedasticity
- Ordinary least squares (OLS)
- Panel data
- Robust estimator
- Robustness to unknown heteroscedasticity
- Two-step estimator
- Weighted least squares (WLS)
- White heteroscedasticity robust estimator
- White test
- Aitken's theorem

Exercises

1. What is the covariance matrix, $\text{Cov}[\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}} - \mathbf{b}]$, of the GLS estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$ and the difference between it and the OLS estimator, $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$? The result plays a pivotal role in the development of specification tests in Hausman (1978).
2. This and the next two exercises are based on the test statistic usually used to test a set of J linear restrictions in the generalized regression model,

$$F[J, n - K] = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})/J}{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})/(n - K)},$$

where $\hat{\boldsymbol{\beta}}$ is the GLS estimator. Show that if $\boldsymbol{\Omega}$ is known, if the disturbances are normally distributed and if the null hypothesis, $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, is true, then this statistic is exactly distributed as F with J and $n - K$ degrees of freedom. What assumptions about the regressors are needed to reach this conclusion? Need they be nonstochastic?
3. Now suppose that the disturbances are not normally distributed, although $\boldsymbol{\Omega}$ is still known. Show that the limiting distribution of the previous statistic is $(1/J)$ times a chi-squared variable with J degrees of freedom. (*Hint*: The denominator converges to σ^2.) Conclude that, in the generalized regression model, the limiting distribution of the Wald statistic,

$$W = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'\{\mathbf{R}(\text{Est. Var}[\hat{\boldsymbol{\beta}}])\mathbf{R}'\}^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}),$$

is chi-squared with J degrees of freedom, regardless of the distribution of the disturbances, as long as the data are otherwise well behaved. Note that in a finite sample, the true distribution may be approximated with an $F[J, n - K]$ distribution. It is a bit ambiguous, however, to interpret this fact as implying that the statistic is asymptotically distributed as F with J and $n - K$ degrees of freedom, because the limiting distribution used to obtain our result is the chi-squared, not the F. In this instance, the $F[J, n - K]$ is a random variable that tends asymptotically to the chi-squared variate.

4. Finally, suppose that $\mathbf{\Omega}$ must be estimated, but that assumptions (9-22) and (9-23) are met by the estimator. What changes are required in the development of the previous problem?
5. In the generalized regression model, if the K columns of $\mathbf{X}$ are characteristic vectors of $\mathbf{\Omega}$, then ordinary least squares and generalized least squares are identical. (The result is actually a bit broader; $\mathbf{X}$ may be any linear combination of exactly K characteristic vectors. This result is Kruskal's theorem.)
   a. Prove the result directly using matrix algebra.
   b. Prove that if $\mathbf{X}$ contains a constant term and if the remaining columns are in deviation form (so that the column sum is zero), then the model of Exercise 8 is one of these cases. (The seemingly unrelated regressions model with identical regressor matrices, discussed in Chapter 10, is another.)
6. In the generalized regression model, suppose that $\mathbf{\Omega}$ is known.
   a. What is the covariance matrix of the OLS and GLS estimators of $\boldsymbol{\beta}$?
   b. What is the covariance matrix of the OLS residual vector $\mathbf{e} = \mathbf{y} - \mathbf{Xb}$?
   c. What is the covariance matrix of the GLS residual vector $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$?
   d. What is the covariance matrix of the OLS and GLS residual vectors?
7. Suppose that y has the pdf $f(y|\mathbf{x}) = (1/\mathbf{x}'\boldsymbol{\beta})e^{-y/(\mathbf{x}'\boldsymbol{\beta})}$, $y > 0$. Then $E[y|\mathbf{x}] = \mathbf{x}'\boldsymbol{\beta}$ and $\text{Var}[y|\mathbf{x}] = (\mathbf{x}'\boldsymbol{\beta})^2$. 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.
8. Suppose that the regression model is $y = \mu + \varepsilon$, where ε has a zero mean, constant variance, and equal correlation, ρ, across observations. Then $\text{Cov}[\varepsilon_i, \varepsilon_j] = \sigma^2\rho$ if $i \neq j$. Prove that the least squares estimator of μ is inconsistent. Find the characteristic roots of $\mathbf{\Omega}$ and show that Condition 2 before (9-10) is violated.
9. Suppose that the regression model is $y_i = \mu + \varepsilon_i$, where

$$E[\varepsilon_i|x_i] = 0, \text{Cov}[\varepsilon_i, \varepsilon_j|x_i, x_j] = 0 \quad \text{for } i = j, \text{but Var}[\varepsilon_i|x_i] = \sigma^2 x_i^2, x_i > 0.$$

   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 OLS estimator?
   c. Prove that the estimator in part a is at least as efficient as the estimator in part b.
10. For the model in Exercise 9, 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 9? Does the conventional estimator produce the correct estimator of the true asymptotic variance of the least squares estimator?

11. For the model in Exercise 9, 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?
12. Two samples of 50 observations each produce the following moment matrices. (In each case, **X** is a constant and one variable.)

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

   a. Compute the least squares regression coefficients and the residual variances s^2 for each data set. Compute the R^2s 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.
13. Suppose that in the groupwise heteroscedasticity model of Section 9.7.2, $\mathbf{X}_i$ is the same for all i. What is the generalized least squares estimator of $\boldsymbol{\beta}$? How would you compute the estimator if it were necessary to estimate σ_i^2?
14. The model

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \beta + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}$$

satisfies the groupwise heteroscedastic regression model of Section 9.7.2 All variables have zero means. The following sample second-moment matrix is obtained from a sample of 20 observations:

$$\begin{array}{c} \\ y_1 \\ y_2 \\ x_1 \\ x_2 \end{array}\begin{array}{c} \begin{array}{cccc} y_1 & y_2 & x_1 & x_2 \end{array} \\ \begin{bmatrix} 20 & 6 & 4 & 3 \\ 6 & 10 & 3 & 6 \\ 4 & 3 & 5 & 2 \\ 3 & 6 & 2 & 10 \end{bmatrix} \end{array}.$$

   a. Compute the two separate OLS estimates of β, their sampling variances, the estimates of σ_1^2 and σ_2^2, and the R^2s in the two regressions.
   b. Carry out the Lagrange multiplier test of the hypothesis that $\sigma_1^2 = \sigma_2^2$.
   c. Compute the two-step FGLS estimate of β and an estimate of its sampling variance. Test the hypothesis that β equals 1.
   d. Compute the maximum likelihood estimates of β, σ_1^2, and σ_2^2 by iterating the FGLS estimates to convergence.

15. The following table presents a hypothetical panel of data:

	$i = 1$		$i = 2$		$i = 3$	
t	y	x	y	x	y	x
1	30.27	24.31	38.71	28.35	37.03	21.16
2	35.59	28.47	29.74	27.38	43.82	26.76
3	17.90	23.74	11.29	12.74	37.12	22.21
4	44.90	25.44	26.17	21.08	24.34	19.02
5	37.58	20.80	5.85	14.02	26.15	18.64
6	23.15	10.55	29.01	20.43	26.01	18.97
7	30.53	18.40	30.38	28.13	29.64	21.35
8	39.90	25.40	36.03	21.78	30.25	21.34
9	20.44	13.57	37.90	25.65	25.41	15.86
10	36.85	25.60	33.90	11.66	26.04	13.28

a. Estimate the groupwise heteroscedastic model of Section 9.7.2. Include an estimate of the asymptotic variance of the slope estimator. Use a two-step procedure, basing the FGLS estimator at the second step on residuals from the pooled least squares regression.
b. Carry out the Lagrange multiplier tests of the hypothesis that the variances are all equal.

Applications

1. This application 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 OLS 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 (1980) and Godfrey (1988) Lagrange multiplier test to test for heteroscedasticity.
e. Reestimate the parameters using a two-step FGLS estimator. Use Harvey's formulation, $\text{Var}[\varepsilon_i | x_{i1}, x_{i2}] = \sigma^2 \exp(\gamma_1 x_{i1} + \gamma_2 x_{i2})$.

2. (We look ahead to our use of maximum likelihood to estimate the models discussed in this chapter in Chapter 14.) In Example 9.3, we computed an iterated FGLS estimator using the airline data and the model $\text{Var}[\varepsilon_{it} | \mathit{Loadfactor}_{i,t}] = \exp(\gamma_1 + \gamma_2\, \mathit{Loadfactor}_{i,t})$. The weights computed at each iteration were computed by estimating (γ_1, γ_2) by least squares regression of $\ln \hat{\varepsilon}^2_{i,t}$ on a constant and *Loadfactor*. The maximum likelihood estimator would proceed along similar lines, however the weights would be computed by regression of $[\hat{\varepsilon}^2_{i,t}/\hat{\sigma}^2_{i,t} - 1]$ on a constant and $\mathit{Loadfactor}_{i,t}$ instead. Use this alternative procedure to estimate the model. Do you get different results?