11

HETEROSCEDASTICITY: WHAT HAPPENS IF THE ERROR VARIANCE IS NONCONSTANT?

An important assumption of the classical linear regression model (Assumption 4) is that the disturbances u_i appearing in the population regression function are homoscedastic; that is, they all have the same variance. In this chapter we examine the validity of this assumption and find out what happens if this assumption is not fulfilled. As in Chapter 10, we seek answers to the following questions:

1. What is the nature of heteroscedasticity?
2. What are its consequences?
3. How does one detect it?
4. What are the remedial measures?

11.1 THE NATURE OF HETEROSCEDASTICITY

As noted in Chapter 3, one of the important assumptions of the classical linear regression model is that the variance of each disturbance term u_i, conditional on the chosen values of the explanatory variables, is some constant number equal to σ^2. This is the assumption of **homoscedasticity,** or *equal* (homo) *spread* (scedasticity), that is, *equal variance*. Symbolically,

$$E(u_i^2) = \sigma^2 \qquad i = 1, 2, \ldots, n \tag{11.1.1}$$

Diagrammatically, in the two-variable regression model homoscedasticity can be shown as in Figure 3.4, which, for convenience, is reproduced as

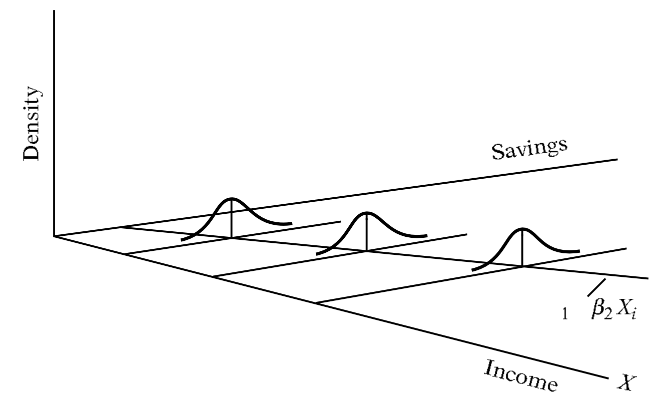

FIGURE 11.1 Homoscedastic disturbances.

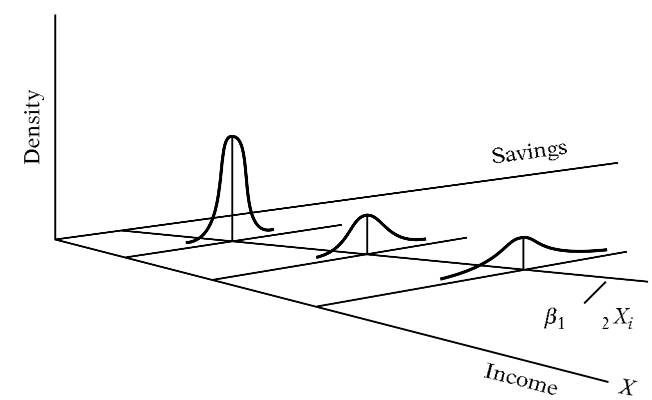

FIGURE 11.2 Heteroscedastic disturbances.

Figure 11.1. As Figure 11.1 shows, the conditional variance of Y_i (which is equal to that of u_i), conditional upon the given X_i, remains the same regardless of the values taken by the variable X.

In contrast, consider Figure 11.2, which shows that the conditional variance of Y_i increases as X increases. Here, the variances of Y_i are not the same. Hence, there is heteroscedasticity. Symbolically,

$$E\left(u_i^2\right) = \sigma_i^2 \tag{11.1.2}$$

Notice the subscript of σ^2, which reminds us that the conditional variances of u_i (= conditional variances of Y_i) are no longer constant.

To make the difference between homoscedasticity and heteroscedasticity clear, assume that in the two-variable model $Y_i = \beta_1 + \beta_2 X_i + u_i$, Y represents savings and X represents income. Figures 11.1 and 11.2 show that as income increases, savings on the average also increase. But in Figure 11.1

the variance of savings remains the same at all levels of income, whereas in Figure 11.2 it increases with income. It seems that in Figure 11.2 the higher-income families on the average save more than the lower-income families, but there is also more variability in their savings.

There are several reasons why the variances of u_i may be variable, some of which are as follows.[1]

1. Following the *error-learning models*, as people learn, their errors of behavior become smaller over time. In this case, σ_i^2 is expected to decrease. As an example, consider Figure 11.3, which relates the number of typing errors made in a given time period on a test to the hours put in typing practice. As Figure 11.3 shows, as the number of hours of typing practice increases, the average number of typing errors as well as their variances decreases.

2. As incomes grow, people have more *discretionary income*[2] and hence more scope for choice about the disposition of their income. Hence, σ_i^2 is likely to increase with income. Thus in the regression of savings on income one is likely to find σ_i^2 increasing with income (as in Figure 11.2) because people have more choices about their savings behavior. Similarly, companies with larger profits are generally expected to show greater variability in their dividend policies than companies with lower profits. Also, *growth-oriented* companies are likely to show more variability in their dividend payout ratio than established companies.

3. As data collecting techniques improve, σ_i^2 is likely to decrease. Thus, banks that have sophisticated data processing equipment are likely to

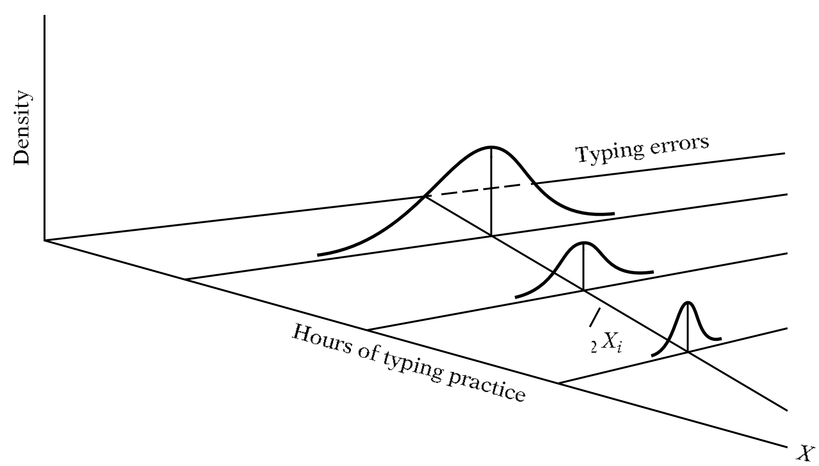

FIGURE 11.3 Illustration of heteroscedasticity.

[1]See Stefan Valavanis, *Econometrics*, McGraw-Hill, New York, 1959, p. 48.

[2]As Valavanis puts it, "Income grows, and people now barely discern dollars whereas previously they discerned dimes," ibid., p. 48.

commit fewer errors in the monthly or quarterly statements of their customers than banks without such facilities.

4. Heteroscedasticity can also arise as a result of the presence of **outliers.** An outlying observation, or outlier, is an observation that is much different (either very small or very large) in relation to the observations in the sample. More precisely, an outlier is an observation from a different population to that generating the remaining sample observations.[3] The inclusion or exclusion of such an observation, especially if the sample size is small, can substantially alter the results of regression analysis.

As an example, consider the scattergram given in Figure 11.4. Based on the data given in exercise 11.22, this figure plots percent rate of change of stock prices (Y) and consumer prices (X) for the post–World War II period through 1969 for 20 countries. In this figure the observation on Y and X for Chile can be regarded as an outlier because the given Y and X values are much larger than for the rest of the countries. In situations such as this, it would be hard to maintain the assumption of homoscedasticity. In exercise 11.22, you are asked to find out what happens to the regression results if the observations for Chile are dropped from the analysis.

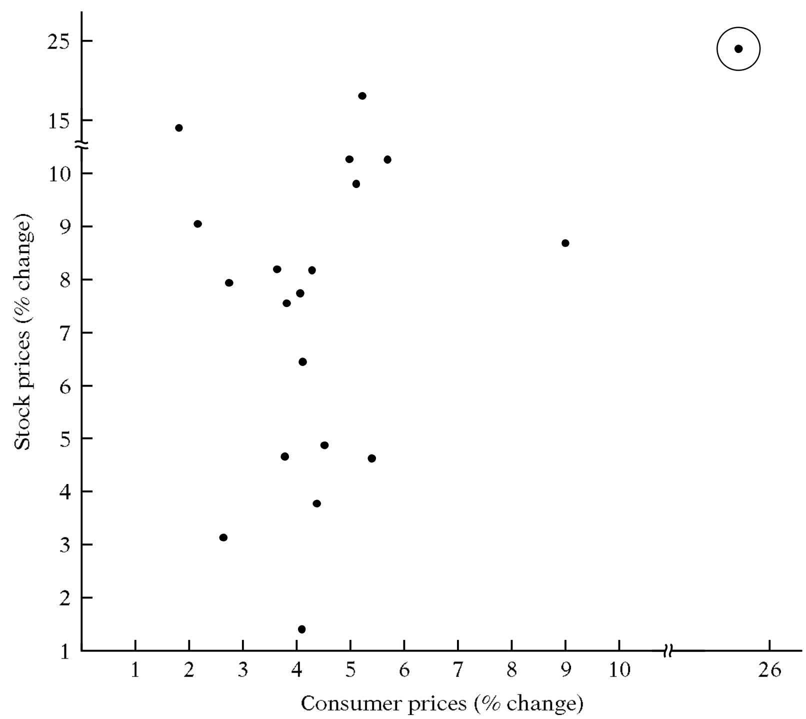

FIGURE 11.4 The relationship between stock prices and consumer prices.

[3]I am indebted to Michael McAleer for pointing this out to me.

5. Another source of heteroscedasticity arises from violating Assumption 9 of CLRM, namely, that the regression model is correctly specified. Although we will discuss the topic of specification errors more fully in Chapter 13, very often what looks like heteroscedasticity may be due to the fact that some important variables are omitted from the model. Thus, in the demand function for a commodity, if we do not include the prices of commodities complementary to or competing with the commodity in question (the omitted variable bias), the residuals obtained from the regression may give the distinct impression that the error variance may not be constant. But if the omitted variables are included in the model, that impression may disappear.

As a concrete example, recall our study of advertising impressions retained (Y) in relation to advertising expenditure (X). (See exercise 8.32.) If you regress Y on X only and observe the residuals from this regression, you will see one pattern, but if you regress Y on X and X^2, you will see another pattern, which can be seen clearly from Figure 11.5. We have already seen that X^2 belongs in the model. (See exercise 8.32.)

6. Another source of heteroscedasticity is **skewness** in the distribution of one or more regressors included in the model. Examples are economic variables such as income, wealth, and education. It is well known that the distribution of income and wealth in most societies is uneven, with the bulk of the income and wealth being owned by a few at the top.

7. Other sources of heteroscedasticity: As David Hendry notes, heteroscedasticity can also arise because of (1) incorrect data transformation (e.g., ratio or first difference transformations) and (2) incorrect functional form (e.g., linear versus log–linear models).[4]

Note that the problem of heteroscedasticity is likely to be more common in cross-sectional than in time series data. In cross-sectional data, one

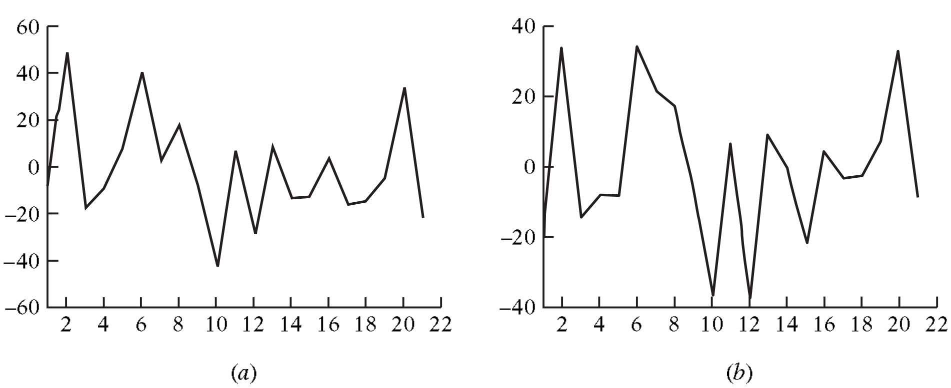

FIGURE 11.5 Residuals from the regression of (*a*) impressions of advertising expenditure and (*b*) impression on Adexp and Adexp^2.

[4]David F. Hendry, *Dynamic Econometrics,* Oxford University Press, 1995, p. 45.

usually deals with members of a population at a given point in time, such as individual consumers or their families, firms, industries, or geographical subdivisions such as state, country, city, etc. Moreover, these members may be of different sizes, such as small, medium, or large firms or low, medium, or high income. In time series data, on the other hand, the variables tend to be of similar orders of magnitude because one generally collects the data for the same entity over a period of time. Examples are GNP, consumption expenditure, savings, or employment in the United States, say, for the period 1950 to 2000.

As an illustration of heteroscedasticity likely to be encountered in cross-sectional analysis, consider Table 11.1. This table gives data on compensation per employee in 10 nondurable goods manufacturing industries, classified by the employment size of the firm or the establishment for the year 1958. Also given in the table are average productivity figures for nine employment classes.

Although the industries differ in their output composition, Table 11.1 shows clearly that on the average large firms pay more than the small firms.

TABLE 11.1
COMPENSATION PER EMPLOYEE ($) IN NONDURABLE MANUFACTURING INDUSTRIES ACCORDING TO EMPLOYMENT SIZE OF ESTABLISHMENT, 1958

	Employment size (average number of employees)								
Industry	1–4	5–9	10–19	20–49	50–99	100–249	250–499	500–999	1000–2499
Food and kindred products	2994	3295	3565	3907	4189	4486	4676	4968	5342
Tobacco products	1721	2057	3336	3320	2980	2848	3072	2969	3822
Textile mill products	3600	3657	3674	3437	3340	3334	3225	3163	3168
Apparel and related products	3494	3787	3533	3215	3030	2834	2750	2967	3453
Paper and allied products	3498	3847	3913	4135	4445	4885	5132	5342	5326
Printing and publishing	3611	4206	4695	5083	5301	5269	5182	5395	5552
Chemicals and allied products	3875	4660	4930	5005	5114	5248	5630	5870	5876
Petroleum and coal products	4616	5181	5317	5337	5421	5710	6316	6455	6347
Rubber and plastic products	3538	3984	4014	4287	4221	4539	4721	4905	5481
Leather and leather products	3016	3196	3149	3317	3414	3254	3177	3346	4067
Average compensation	3396	3787	4013	4104	4146	4241	4388	4538	4843
Standard deviation	742.2	851.4	727.8	805.06	929.9	1080.6	1241.2	1307.7	1110.5
Average productivity	9355	8584	7962	8275	8389	9418	9795	10,281	11,750

Source: The Census of Manufacturers, U.S. Department of Commerce, 1958 (computed by author).

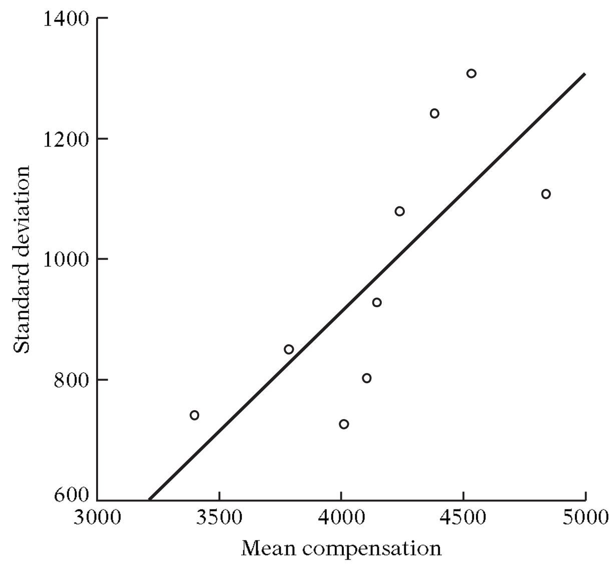

FIGURE 11.6 Standard deviation of compensation and mean compensation.

As an example, firms employing one to four employees paid on the average about \$3396, whereas those employing 1000 to 2499 employees on the average paid about \$4843. But notice that there is considerable variability in earning among various employment classes as indicated by the estimated standard deviations of earnings. This can be seen also from Figure 11.6, which plots the standard deviation of compensation and average compensation in each employment class. As can be seen clearly, on average, the standard deviation of compensation increases with the average value of compensation.

11.2 OLS ESTIMATION IN THE PRESENCE OF HETEROSCEDASTICITY

What happens to OLS estimators and their variances if we introduce heteroscedasticity by letting $E(u_i^2) = \sigma_i^2$ but retain all other assumptions of the classical model? To answer this question, let us revert to the two-variable model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Applying the usual formula, the OLS estimator of β_2 is

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum x_i y_i}{\sum x_i^2}\\ &= \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}\end{aligned} \tag{11.2.1}$$

but its variance is now given by the following expression (see Appendix 11A, Section 11A.1):

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2} \tag{11.2.2}$$

which is obviously different from the usual variance formula obtained under the assumption of homoscedasticity, namely,

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{11.2.3}$$

Of course, if $\sigma_i^2 = \sigma^2$ for each i, the two formulas will be identical. (Why?)

Recall that $\hat{\beta}_2$ is best linear unbiased estimator (BLUE) if the assumptions of the classical model, including homoscedasticity, hold. Is it still BLUE when we drop only the homoscedasticity assumption and replace it with the assumption of heteroscedasticity? It is easy to prove that $\hat{\beta}_2$ is still linear and unbiased. As a matter of fact, as shown in Appendix 3A, Section 3A.2, to establish the unbiasedness of $\hat{\beta}_2$ it is not necessary that the disturbances (u_i) be homoscedastic. In fact, the variance of u_i, homoscedastic or heteroscedastic, plays no part in the determination of the unbiasedness property. Recall that in Appendix 3A, Section 3A.7, we showed that $\hat{\beta}_2$ is a consistent estimator under the assumptions of the classical linear regression model. Although we will not prove it, it can be shown that $\hat{\beta}_2$ is a consistent estimator despite heteroscedasticity; that is, as the sample size increases indefinitely, the estimated β_2 converges to its true value. Furthermore, it can also be shown that under certain conditions (called regularity conditions), $\hat{\beta}_2$ is *asymptotically normally distributed*. Of course, what we have said about $\hat{\beta}_2$ also holds true of other parameters of a multiple regression model.

Granted that $\hat{\beta}_2$ is still linear unbiased and consistent, is it "efficient" or "best"; that is, does it have minimum variance in the class of unbiased estimators? And is that minimum variance given by Eq. (11.2.2)? The answer is *no* to both the questions: $\hat{\beta}_2$ is no longer best and the minimum variance is not given by (11.2.2). Then what is BLUE in the presence of heteroscedasticity? The answer is given in the following section.

11.3 THE METHOD OF GENERALIZED LEAST SQUARES (GLS)

Why is the usual OLS estimator of β_2 given in (11.2.1) not best, although it is still unbiased? Intuitively, we can see the reason from Table 11.1. As the table shows, there is considerable variability in the earnings between employment classes. If we were to regress per-employee compensation on the size of employment, we would like to make use of the knowledge that there is considerable interclass variability in earnings. Ideally, we would like to devise the estimating scheme in such a manner that observations coming from populations with greater variability are given less weight than those coming from populations with smaller variability. Examining Table 11.1, we would like to weight observations coming from employment classes 10–19 and 20–49 more heavily than those coming from employment classes like 5–9 and 250–499, for the former are more closely clustered around their mean values than the latter, thereby enabling us to estimate the PRF more accurately.

Unfortunately, the usual OLS method does not follow this strategy and therefore does not make use of the "information" contained in the unequal variability of the dependent variable Y, say, employee compensation of Table 11.1: It assigns equal weight or importance to each observation. But a method of estimation, known as **generalized least squares (GLS),** takes such information into account explicitly and is therefore capable of producing estimators that are BLUE. To see how this is accomplished, let us continue with the now-familiar two-variable model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{11.3.1}$$

which for ease of algebraic manipulation we write as

$$Y_i = \beta_1 X_{0i} + \beta_2 X_i + u_i \tag{11.3.2}$$

where $X_{0i} = 1$ for each i. The reader can see that these two formulations are identical.

Now assume that the heteroscedastic variances σ_i^2 are *known*. Divide (11.3.2) through by σ_i to obtain

$$\frac{Y_i}{\sigma_i} = \beta_1 \left(\frac{X_{0i}}{\sigma_i}\right) + \beta_2 \left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right) \tag{11.3.3}$$

which for ease of exposition we write as

$$Y_i^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^* \tag{11.3.4}$$

where the starred, or transformed, variables are the original variables divided by (the known) σ_i. We use the notation β_1^* and β_2^*, the parameters of the transformed model, to distinguish them from the usual OLS parameters β_1 and β_2.

What is the purpose of transforming the original model? To see this, notice the following feature of the transformed error term u_i^*:

$$\begin{aligned}
\operatorname{var}(u_i^*) &= E(u_i^*)^2 = E\left(\frac{u_i}{\sigma_i}\right)^2 \\
&= \frac{1}{\sigma_i^2} E(u_i^2) \qquad \text{since } \sigma_i^2 \text{ is known} \\
&= \frac{1}{\sigma_i^2}(\sigma_i^2) \qquad \text{since } E(u_i^2) = \sigma_i^2 \\
&= 1
\end{aligned} \tag{11.3.5}$$

which is a constant. That is, the variance of the transformed disturbance term u_i^* is now homoscedastic. Since we are still retaining the other assumptions of the classical model, the finding that it is u^* that is homoscedastic suggests that if we apply OLS to the transformed model (11.3.3) it will produce estimators that are BLUE. In short, the estimated β_1^* and β_2^* are now BLUE and not the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$.

This procedure of transforming the original variables in such a way that the transformed variables satisfy the assumptions of the classical model and then applying OLS to them is known as the method of generalized least squares (GLS). *In short, GLS is OLS on the transformed variables that satisfy the standard least-squares assumptions*. The estimators thus obtained are known as **GLS estimators,** and it is these estimators that are BLUE.

The actual mechanics of estimating β_1^* and β_2^* are as follows. First, we write down the SRF of (11.3.3)

$$\frac{Y_i}{\sigma_i} = \hat{\beta}_1^*\left(\frac{X_{0i}}{\sigma_i}\right) + \hat{\beta}_2^*\left(\frac{X_i}{\sigma_i}\right) + \left(\frac{\hat{u}_i}{\sigma_i}\right)$$

or

$$Y_i^* = \hat{\beta}_1^* X_{0i}^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^* \tag{11.3.6}$$

Now, to obtain the GLS estimators, we minimize

$$\sum \hat{u}_i^{2*} = \sum (Y_i^* - \hat{\beta}_1^* X_{0i}^* - \hat{\beta}_2^* X_i^*)^2$$

that is,

$$\sum \left(\frac{\hat{u}_i}{\sigma_i}\right)^2 = \sum \left[\left(\frac{Y_i}{\sigma_i}\right) - \hat{\beta}_1^*\left(\frac{X_{0i}}{\sigma_i}\right) - \hat{\beta}_2^*\left(\frac{X_i}{\sigma_i}\right)\right]^2 \tag{11.3.7}$$

The actual mechanics of minimizing (11.3.7) follow the standard calculus techniques and are given in Appendix 11A, Section 11A.2. As shown there, the GLS estimator of β_2^* is

$$\hat{\beta}_2^* = \frac{\left(\sum w_i\right)\left(\sum w_i X_i Y_i\right) - \left(\sum w_i X_i\right)\left(\sum w_i Y_i\right)}{\left(\sum w_i\right)\left(\sum w_i X_i^2\right) - \left(\sum w_i X_i\right)^2} \tag{11.3.8}$$

and its variance is given by

$$\operatorname{var}(\hat{\beta}_2^*) = \frac{\sum w_i}{\left(\sum w_i\right)\left(\sum w_i X_i^2\right) - \left(\sum w_i X_i\right)^2} \tag{11.3.9}$$

where $w_i = 1/\sigma_i^2$.

Difference between OLS and GLS

Recall from Chapter 3 that in OLS we minimize

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \quad \textbf{(11.3.10)}$$

but in GLS we minimize the expression (11.3.7), which can also be written as

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* X_{0i} - \hat{\beta}_2^* X_i)^2 \quad \textbf{(11.3.11)}$$

where $w_i = 1/\sigma_i^2$ [verify that (11.3.11) and (11.3.7) are identical].

Thus, in GLS we minimize a *weighted sum of residual squares* with $w_i = 1/\sigma_i^2$ acting as the weights, but in OLS we minimize an unweighted or (what amounts to the same thing) equally weighted RSS. As (11.3.7) shows, in GLS the weight assigned to each observation is inversely proportional to its σ_i, that is, observations coming from a population with larger σ_i will get relatively smaller weight and those from a population with smaller σ_i will get proportionately larger weight in minimizing the RSS (11.3.11). To see the difference between OLS and GLS clearly, consider the hypothetical scattergram given in Figure 11.7.

In the (unweighted) OLS, each $\hat{u}_i^2$ associated with points A, B, and C will receive the same weight in minimizing the RSS. Obviously, in this case the $\hat{u}_i^2$ associated with point C will dominate the RSS. But in GLS the extreme observation C will get relatively smaller weight than the other two observations. As noted earlier, this is the right strategy, for in estimating the

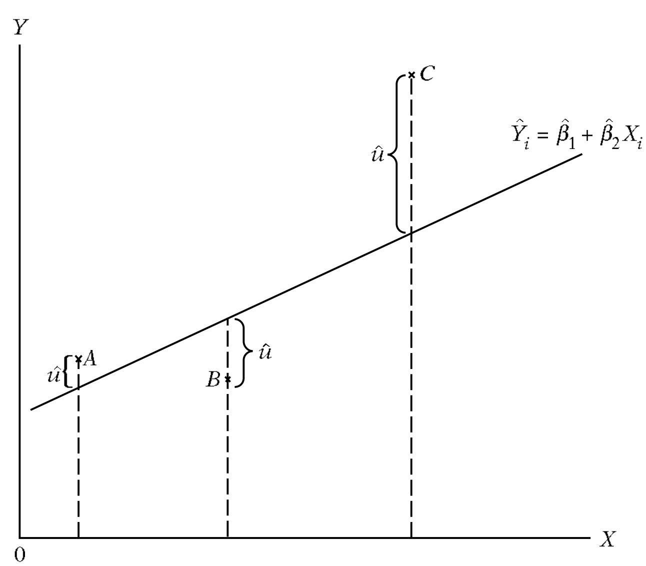

FIGURE 11.7 Hypothetical scattergram.

population regression function (PRF) more reliably we would like to give more weight to observations that are closely clustered around their (population) mean than to those that are widely scattered about.

Since (11.3.11) minimizes a weighted RSS, it is appropriately known as **weighted least squares (WLS),** and the estimators thus obtained and given in (11.3.8) and (11.3.9) are known as **WLS estimators.** But WLS is just a special case of the more general estimating technique, GLS. In the context of heteroscedasticity, one can treat the two terms WLS and GLS interchangeably. In later chapters we will come across other special cases of GLS.

In passing, note that if $w_i = w$, a constant for all i, $\hat{\beta}_2^*$ is identical with $\hat{\beta}_2$ and $\operatorname{var}(\hat{\beta}_2^*)$ is identical with the usual (i.e., homoscedastic) $\operatorname{var}(\hat{\beta}_2)$ given in (11.2.3), which should not be surprising. (Why?) (See exercise 11.8.)

11.4 CONSEQUENCES OF USING OLS IN THE PRESENCE OF HETEROSCEDASTICITY

As we have seen, both $\hat{\beta}_2^*$ and $\hat{\beta}_2$ are (linear) unbiased estimators: In repeated sampling, on the average, $\hat{\beta}_2^*$ and $\hat{\beta}_2$ will equal the true β_2; that is, they are both unbiased estimators. But we know that it is $\hat{\beta}_2^*$ that is efficient, that is, has the smallest variance. What happens to our confidence interval, hypotheses testing, and other procedures if we continue to use the OLS estimator $\hat{\beta}_2$? We distinguish two cases.

OLS Estimation Allowing for Heteroscedasticity

Suppose we use $\hat{\beta}_2$ and use the variance formula given in (11.2.2), which takes into account heteroscedasticity explicitly. Using this variance, and assuming σ_i^2 are known, can we establish confidence intervals and test hypotheses with the usual t and F tests? The answer generally is no because it can be shown that $\operatorname{var}(\hat{\beta}_2^*) \le \operatorname{var}(\hat{\beta}_2)$,[5] which means that confidence intervals based on the latter will be unnecessarily larger. As a result, the t and F tests are likely to give us inaccurate results in that $\operatorname{var}(\hat{\beta}_2)$ is overly large and what appears to be a statistically insignificant coefficient (because the t value is smaller than what is appropriate) may in fact be significant if the correct confidence intervals were established on the basis of the GLS procedure.

OLS Estimation Disregarding Heteroscedasticity

The situation can become serious if we not only use $\hat{\beta}_2$ but also continue to use the usual (homoscedastic) variance formula given in (11.2.3) even if heteroscedasticity is present or suspected: Note that this is the more likely

[5] A formal proof can be found in Phoebus J. Dhrymes, *Introductory Econometrics*, Springer-Verlag, New York, 1978, pp. 110–111. In passing, note that the loss of efficiency of $\hat{\beta}_2$ [i.e., by how much $\operatorname{var}(\hat{\beta}_2)$ exceeds $\operatorname{var}(\hat{\beta}_2^*)$] depends on the sample values of the X variables and the value of σ_i^2.

case of the two we discuss here, because running a standard OLS regression package and ignoring (or being ignorant of) heteroscedasticity will yield variance of $\hat{\beta}_2$ as given in (11.2.3). First of all, var $(\hat{\beta}_2)$ given in (11.2.3) is a *biased* estimator of var $(\hat{\beta}_2)$ given in (11.2.2), that is, on the average it overestimates or underestimates the latter, and *in general* we cannot tell whether the bias is positive (overestimation) or negative (underestimation) because it depends on the nature of the relationship between σ_i^2 and the values taken by the explanatory variable X, as can be seen clearly from (11.2.2) (see exercise 11.9). The bias arises from the fact that $\hat{\sigma}^2$, the conventional estimator of σ^2, namely, $\sum \hat{u}_i^2/(n-2)$ is no longer an unbiased estimator of the latter when heteroscedasticity is present (see Appendix 11A.3). As a result, we can no longer rely on the conventionally computed confidence intervals and the conventionally employed t and F tests.[6] **In short, if we persist in using the usual testing procedures despite heteroscedasticity, whatever conclusions we draw or inferences we make may be very misleading.**

To throw more light on this topic, we refer to a **Monte Carlo** study conducted by Davidson and MacKinnon.[7] They consider the following simple model, which in our notation is

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \textbf{(11.4.1)}$$

They assume that $\beta_1 = 1$, $\beta_2 = 1$, and $u_i \sim N(0, X_i^{\alpha})$. As the last expression shows, they assume that the error variance is heteroscedastic and is related to the value of the regressor X with power α. If, for example, $\alpha = 1$, the error variance is proportional to the value of X; if $\alpha = 2$, the error variance is proportional to the square of the value of X, and so on. In Section 11.6 we will consider the logic behind such a procedure. Based on 20,000 replications and allowing for various values for α, they obtain the standard errors of the two regression coefficients using OLS [see Eq. (11.2.3)], OLS allowing for heteroscedasticity [see Eq. (11.2.2)], and GLS [see Eq. (11.3.9)]. We quote their results for selected values of α:

	Standard error of $\hat{\beta}_1$			Standard error of $\hat{\beta}_2$		
Value of α	OLS	OLS_{het}	GLS	OLS	OLS_{het}	GLS
0.5	0.164	0.134	0.110	0.285	0.277	0.243
1.0	0.142	0.101	0.048	0.246	0.247	0.173
2.0	0.116	0.074	0.0073	0.200	0.220	0.109
3.0	0.100	0.064	0.0013	0.173	0.206	0.056
4.0	0.089	0.059	0.0003	0.154	0.195	0.017

Note: OLS_{het} means OLS allowing for heteroscedasticity.

[6]From (5.3.6) we know that the $100(1-\alpha)\%$ confidence interval for β_2 is $[\hat{\beta}_2 \pm t_{\alpha/2}\,\text{se}\,(\hat{\beta}_2)]$. But if se $(\hat{\beta}_2)$ cannot be estimated unbiasedly, what trust can we put in the conventionally computed confidence interval?

[7]Russell Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993, pp. 549–550.

The most striking feature of these results is that OLS, with or without correction for heteroscedasticity, consistently overestimates the true standard error obtained by the (correct) GLS procedure, especially for large values of α, thus establishing the superiority of GLS. These results also show that if we do not use GLS and rely on OLS—allowing for or not allowing for heteroscedasticity—the picture is mixed. The usual OLS standard errors are either too large (for the intercept) or generally too small (for the slope coefficient) in relation to those obtained by OLS allowing for heteroscedasticity. The message is clear: In the presence of heteroscedasticity, use GLS. However, for reasons explained later in the chapter, in practice it is not always easy to apply GLS. Also, as we discuss later, unless heteroscedasticity is very severe, one may not abandon OLS in favor of GLS or WLS.

From the preceding discussion it is clear that heteroscedasticity is potentially a serious problem and the researcher needs to know whether it is present in a given situation. If its presence is detected, then one can take corrective action, such as using the weighted least-squares regression or some other technique. Before we turn to examining the various corrective procedures, however, we must first find out whether heteroscedasticity is present or likely to be present in a given case. This topic is discussed in the following section.

A Technical Note

Although we have stated that, in cases of heteroscedasticity, it is the GLS, not the OLS, that is BLUE, there are examples where OLS can be BLUE, despite heteroscedasticity.[8] But such examples are infrequent in practice.

11.5 DETECTION OF HETEROSCEDASTICITY

As with multicollinearity, the important practical question is: How does one know that heteroscedasticity is present in a specific situation? Again, as in the case of multicollinearity, there are no hard-and-fast rules for detecting heteroscedasticity, only a few rules of thumb. But this situation is inevitable because σ_i^2 can be known only if we have the entire Y population corresponding to the chosen X's, such as the population shown in Table 2.1 or Table 11.1. But such data are an exception rather than the rule in most

[8]The reason for this is that the Gauss–Markov theorem provides the sufficient (but not necessary) condition for OLS to be efficient. The necessary and sufficient condition for OLS to be BLUE is given by **Kruskal's Theorem.** But this topic is beyond the scope of this book. I am indebted to Michael McAleer for bringing this to my attention. For further details, see Denzil G. Fiebig, Michael McAleer, and Robert Bartels, "Properties of Ordinary Least Squares Estimators in Regression Models with Nonspherical Disturbances," *Journal of Econometrics,* vol. 54, No. 1–3, Oct.–Dec., 1992, pp. 321–334. For the mathematically inclined student, I discuss this topic further in **App. C**, using matrix algebra.

economic investigations. In this respect the econometrician differs from scientists in fields such as agriculture and biology, where researchers have a good deal of control over their subjects. More often than not, in economic studies there is only one sample Y value corresponding to a particular value of X. And there is no way one can know σ_i^2 from just one Y observation. Therefore, in most cases involving econometric investigations, heteroscedasticity may be a matter of intuition, educated guesswork, prior empirical experience, or sheer speculation.

With the preceding caveat in mind, let us examine some of the informal and formal methods of detecting heteroscedasticity. As the following discussion will reveal, most of these methods are based on the examination of the OLS residuals $\hat{u}_i$ since they are the ones we observe, and not the disturbances u_i. One hopes that they are good estimates of u_i, a hope that may be fulfilled if the sample size is fairly large.

Informal Methods

Nature of the Problem Very often the nature of the problem under consideration suggests whether heteroscedasticity is likely to be encountered. For example, following the pioneering work of Prais and Houthakker on family budget studies, where they found that the residual variance around the regression of consumption on income increased with income, one now generally assumes that in similar surveys one can expect unequal variances among the disturbances.[9] As a matter of fact, in cross-sectional data involving heterogeneous units, heteroscedasticity may be the rule rather than the exception. Thus, in a cross-sectional analysis involving the investment expenditure in relation to sales, rate of interest, etc., heteroscedasticity is generally expected if small-, medium-, and large-size firms are sampled together.

As a matter of fact, we have already come across examples of this. In Chapter 2 we discussed the relationship between mean, or average, hourly wages in relation to years of schooling in the United States. In that chapter we also discussed the relationship between expenditure on food and total expenditure for 55 families in India (see exercise 11.16).

Graphical Method If there is no a priori or empirical information about the nature of heteroscedasticity, in practice one can do the regression analysis on the assumption that there is no heteroscedasticity and then do a postmortem examination of the residual squared $\hat{u}_i^2$ to see if they exhibit any systematic pattern. Although $\hat{u}_i^2$ are not the same thing as u_i^2, they can be

[9]S. J. Prais and H. S. Houthakker, *The Analysis of Family Budgets*, Cambridge University Press, New York, 1955.

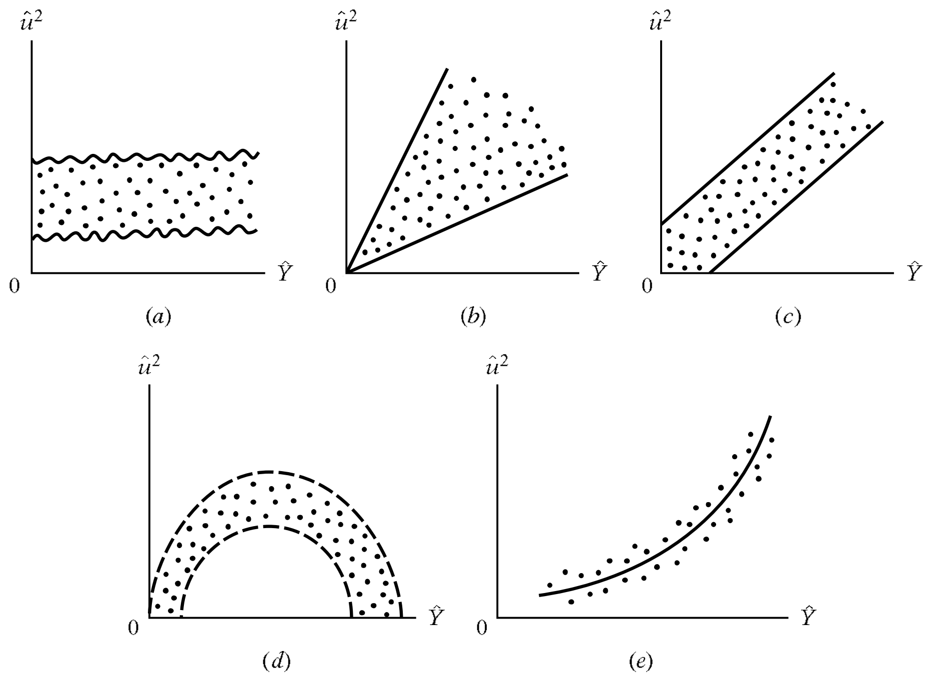

FIGURE 11.8 Hypothetical patterns of estimated squared residuals.

used as proxies especially if the sample size is sufficiently large.[10] An examination of the $\hat{u}_i^2$ may reveal patterns such as those shown in Figure 11.8.

In Figure 11.8, $\hat{u}_i^2$ are plotted against $\hat{Y}_i$, the estimated Y_i from the regression line, the idea being to find out whether the estimated mean value of Y is systematically related to the squared residual. In Figure 11.8*a* we see that there is no systematic pattern between the two variables, suggesting that perhaps no heteroscedasticity is present in the data. Figure 11.8*b* to *e*, however, exhibits definite patterns. For instance, Figure 11.8*c* suggests a linear relationship, whereas Figure 11.8*d* and *e* indicates a quadratic relationship between $\hat{u}_i^2$ and $\hat{Y}_i$. Using such knowledge, albeit informal, one may transform the data in such a manner that the transformed data do not exhibit heteroscedasticity. In Section 11.6 we shall examine several such transformations.

Instead of plotting $\hat{u}_i^2$ against $\hat{Y}_i$, one may plot them against one of the explanatory variables, especially if plotting $\hat{u}_i^2$ against $\hat{Y}_i$ results in the pattern shown in Figure 11.8*a*. Such a plot, which is shown in Figure 11.9, may reveal patterns similar to those given in Figure 11.8. (In the case of the two-variable model, plotting $\hat{u}_i^2$ against $\hat{Y}_i$ is equivalent to plotting it against

[10] For the relationship between $\hat{u}_i$ and u_i, see E. Malinvaud, *Statistical Methods of Econometrics*, North Holland Publishing Company, Amsterdam, 1970, pp. 88–89.

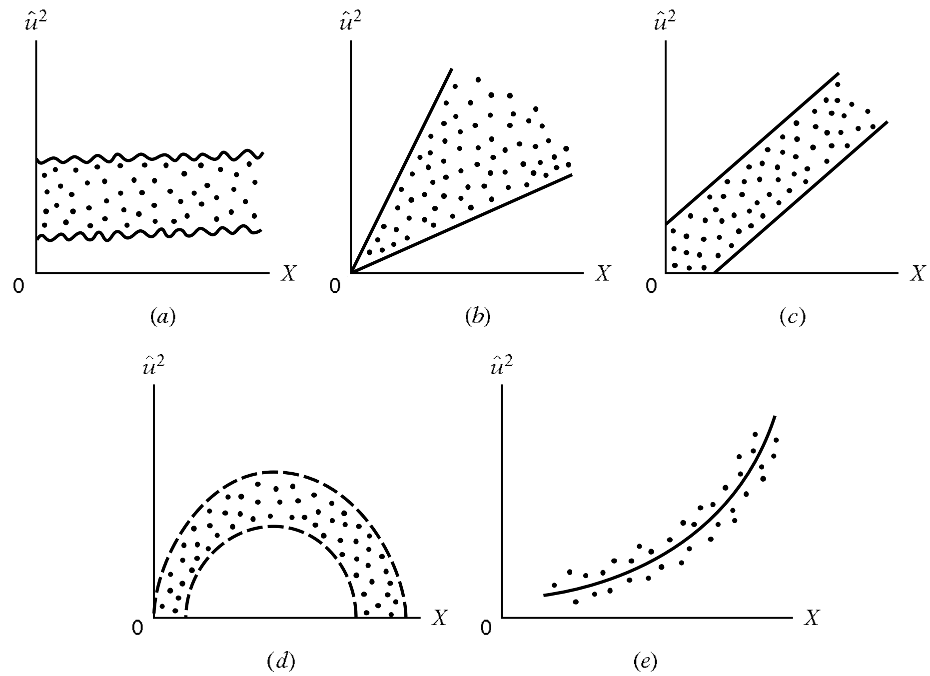

FIGURE 11.9 Scattergram of estimated squared residuals against *X*.

X_i, and therefore Figure 11.9 is similar to Figure 11.8. But this is not the situation when we consider a model involving two or more X variables; in this instance, $\hat{u}_i^2$ may be plotted against any X variable included in the model.)

A pattern such as that shown in Figure 11.9*c*, for instance, suggests that the variance of the disturbance term is linearly related to the X variable. Thus, if in the regression of savings on income one finds a pattern such as that shown in Figure 11.9*c*, it suggests that the heteroscedastic variance may be *proportional* to the value of the income variable. This knowledge may help us in transforming our data in such a manner that in the regression on the transformed data the variance of the disturbance is homoscedastic. We shall return to this topic in the next section.

Formal Methods

Park Test[11] Park formalizes the graphical method by suggesting that σ_i^2 is some function of the explanatory variable X_i. The functional form he

[11] R. E. Park, "Estimation with Heteroscedastic Error Terms," *Econometrica*, vol. 34, no. 4, October 1966, p. 888. The Park test is a special case of the general test proposed by A. C. Harvey in "Estimating Regression Models with Multiplicative Heteroscedasticity," *Econometrica*, vol. 44, no. 3, 1976, pp. 461–465.

suggested was

$$\sigma_i^2 = \sigma^2 X_i^{\beta} e^{v i}$$

or

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i \tag{11.5.1}$$

where v_i is the stochastic disturbance term.

Since σ_i^2 is generally not known, Park suggests using $\hat{u}_i^2$ as a proxy and running the following regression:

$$\begin{aligned} \ln \hat{u}_i^2 &= \ln \sigma^2 + \beta \ln X_i + v_i \\ &= \alpha + \beta \ln X_i + v_i \end{aligned} \tag{11.5.2}$$

If β turns out to be statistically significant, it would suggest that heteroscedasticity is present in the data. If it turns out to be insignificant, we may accept the assumption of homoscedasticity. The Park test is thus a two-stage procedure. In the first stage we run the OLS regression disregarding the heteroscedasticity question. We obtain $\hat{u}_i$ from this regression, and then in the second stage we run the regression (11.5.2).

Although empirically appealing, the Park test has some problems. Goldfeld and Quandt have argued that the error term v_i entering into (11.5.2) may not satisfy the OLS assumptions and may itself be heteroscedastic.[12] Nonetheless, as a strictly exploratory method, one may use the Park test.

EXAMPLE 11.1

RELATIONSHIP BETWEEN COMPENSATION AND PRODUCTIVITY

To illustrate the Park approach, we use the data given in Table 11.1 to run the following regression:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

where Y = average compensation in thousands of dollars, X = average productivity in thousands of dollars, and i = ith employment size of the establishment. The results of the regression were as follows:

$$\begin{aligned} \hat{Y}_i &= 1992.3452 + 0.2329 X_i \\ \text{se} &= (936.4791) \quad (0.0998) \\ t &= (2.1275) \quad (2.333) \qquad R^2 = 0.4375 \end{aligned} \tag{11.5.3}$$

The results reveal that the estimated slope coefficient is significant at the 5 percent level on the basis of a one-tail t test. The equation shows that as labor productivity increases by, say, a dollar, labor compensation on the average increases by about 23 cents.

The residuals obtained from regression (11.5.3) were regressed on X_i as suggested in Eq. (11.5.2), giving the following results:

$$\begin{aligned} \widehat{\ln \hat{u}_i^2} &= 35.817 - 2.8099 \ln X_i \\ \text{se} &= (38.319) \quad (4.216) \\ t &= (0.934) \quad (-0.667) \qquad R^2 = 0.0595 \end{aligned} \tag{11.5.4}$$

Obviously, there is no statistically significant relationship between the two variables. Following the Park test, one may conclude that there is no heteroscedasticity in the error variance.[13]

[12] Stephen M. Goldfeld and Richard E. Quandt, *Nonlinear Methods in Econometrics*, North Holland Publishing Company, Amsterdam, 1972, pp. 93–94.

[13] The particular functional form chosen by Park is only suggestive. A different functional form may reveal significant relationships. For example, one may use $\hat{u}_i^2$ instead of $\ln \hat{u}_i^2$ as the dependent variable.

Glejser Test[14] The Glejser test is similar in spirit to the Park test. After obtaining the residuals $\hat{u}_i$ from the OLS regression, Glejser suggests regressing the absolute values of $\hat{u}_i$ on the X variable that is thought to be closely associated with σ_i^2. In his experiments, Glejser used the following functional forms:

$$|\hat{u}_i| = \beta_1 + \beta_2 X_i + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \sqrt{X_i} + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \frac{1}{X_i} + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

where v_i is the error term.

Again as an empirical or practical matter, one may use the Glejser approach. But Goldfeld and Quandt point out that the error term v_i has some problems in that its expected value is nonzero, it is serially correlated (see Chapter 12), and ironically it is heteroscedastic.[15] An additional difficulty with the Glejser method is that models such as

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

and

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

are nonlinear in the parameters and therefore cannot be estimated with the usual OLS procedure.

Glejser has found that for large samples the first four of the preceding models give generally satisfactory results in detecting heteroscedasticity. As a practical matter, therefore, the Glejser technique may be used for large samples and may be used in the small samples strictly as a qualitative device to learn something about heteroscedasticity.

[14]H. Glejser, "A New Test for Heteroscedasticity," *Journal of the American Statistical Association*, vol. 64, 1969, pp. 316–323.

[15]For details, see Goldfeld and Quandt, op. cit., Chap. 3.

EXAMPLE 11.2

RELATIONSHIP BETWEEN COMPENSATION AND PRODUCTIVITY:
THE GLEJSER TEST

Continuing with Example 11.1, the absolute value of the residuals obtained from regression (11.5.3) were regressed on average productivity (X), giving the following results:

$$\widehat{|\hat{u}_i|} = 407.2783 - 0.0203X_i$$
$$\text{se} = (633.1621) \quad (0.0675) \qquad r^2 = 0.0127 \tag{11.5.5}$$
$$t = (0.6432) \quad (-0.3012)$$

As you can see from this regression, there is no relationship between the absolute value of the residuals and the regressor, average productivity. This reinforces the conclusion based on the Park test.

Spearman's Rank Correlation Test. In exercise 3.8 we defined the Spearman's rank correlation coefficient as

$$r_s = 1 - 6\left[\frac{\sum d_i^2}{n(n^2 - 1)}\right] \tag{11.5.6}$$

where d_i = difference in the ranks assigned to two different characteristics of the ith individual or phenomenon and n = number of individuals or phenomena ranked. The preceding rank correlation coefficient can be used to detect heteroscedasticity as follows: Assume $Y_i = \beta_0 + \beta_1 X_i + u_i$.

Step 1. Fit the regression to the data on Y and X and obtain the residuals $\hat{u}_i$.

Step 2. Ignoring the sign of $\hat{u}_i$, that is, taking their absolute value $|\hat{u}_i|$, rank both $|\hat{u}_i|$ and X_i (or $\hat{Y}_i$) according to an ascending or descending order and compute the Spearman's rank correlation coefficient given previously.

Step 3. Assuming that the population rank correlation coefficient ρ_s is zero and $n > 8$, the significance of the sample r_s can be tested by the t test as follows[16]:

$$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} \tag{11.5.7}$$

with df $= n - 2$.

[16]See G. Udny Yule and M. G. Kendall, *An Introduction to the Theory of Statistics*, Charles Griffin & Company, London, 1953, p. 455.

If the computed t value exceeds the critical t value, we may accept the hypothesis of heteroscedasticity; otherwise we may reject it. If the regression model involves more than one X variable, r_s can be computed between $|\hat{u}_i|$ and each of the X variables separately and can be tested for statistical significance by the t test given in Eq. (11.5.7).

EXAMPLE 11.3

ILLUSTRATION OF THE RANK CORRELATION TEST

To illustrate the rank correlation test, consider the data given in Table 11.2. The data pertain to the average annual return (E, %) and the standard deviation of annual return (σ_i, %) of 10 mutual funds.

The capital market line (CML) of portfolio theory postulates a linear relationship between expected return (E_i) and risk (as measured by the standard deviation, σ) of a portfolio as follows:

$$E_i = \beta_i + \beta_2 \sigma_i$$

Using the data in Table 11.2, the preceding model was estimated and the residuals from this model were computed. Since the data relate to 10 mutual funds of differing sizes and investment goals, a priori one might expect heteroscedasticity. To test this hypothesis, we apply the rank correlation test. The necessary calculations are given in Table 11.2.

Applying formula (11.5.6), we obtain

$$r_s = 1 - 6\frac{110}{10(100-1)} \qquad (11.5.8)$$

$$= 0.3333$$

Applying the t test given in (11.5.7), we obtain

$$t = \frac{(0.3333)(\sqrt{8})}{\sqrt{1-0.1110}} \qquad (11.5.9)$$

$$= 0.9998$$

For 8 df this t value is not significant even at the 10% level of significance; the p value is 0.17. Thus, there is no evidence of a systematic relationship between the explanatory variable and the absolute values of the residuals, which might suggest that there is no heteroscedasticity.

TABLE 11.2
RANK CORRELATION TEST OF HETEROSCEDASTICITY

Name of mutual fund	E_i, average annual return, %	σ_i, standard deviation of annual return, %	$\hat{E}_i^*$	$\lvert\hat{u}_i\rvert^\dagger$ residuals, $\lvert(E_i - \hat{E}_i)\rvert$	Rank of $\lvert\hat{u}_i\rvert$	Rank of σ_i	d, difference between two rankings	d^2
Boston Fund	12.4	12.1	11.37	1.03	9	4	5	25
Delaware Fund	14.4	21.4	15.64	1.24	10	9	1	1
Equity Fund	14.6	18.7	14.40	0.20	4	7	–3	9
Fundamental Investors	16.0	21.7	15.78	0.22	5	10	–5	25
Investors Mutual	11.3	12.5	11.56	0.26	6	5	1	1
Loomis-Sales Mutual Fund	10.0	10.4	10.59	0.59	7	2	5	25
Massachusetts Investors Trust	16.2	20.8	15.37	0.83	8	8	0	0
New England Fund	10.4	10.2	10.50	0.10	3	1	2	4
Putnam Fund of Boston	13.1	16.0	13.16	0.06	2	6	–4	16
Wellington Fund	11.3	12.0	11.33	0.03	1	3	–2	4
Total							0	110

*Obtained from the regression: $\hat{E}_i = 5.8194 + 0.4590\sigma_i$.
†Absolute value of the residuals.
Note: The ranking is in ascending order of values.

Goldfeld-Quandt Test.[17] This popular method is applicable if one assumes that the heteroscedastic variance, σ_i^2, is positively related to *one* of the explanatory variables in the regression model. For simplicity, consider the usual two-variable model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Suppose σ_i^2 is positively related to X_i as

$$\sigma_i^2 = \sigma^2 X_i^2 \tag{11.5.10}$$

where σ^2 is a constant.[18]

Assumption (11.5.10) postulates that σ_i^2 is proportional to the square of the X variable. Such an assumption has been found quite useful by Prais and Houthakker in their study of family budgets. (See Section 11.6.)

If (11.5.10) is appropriate, it would mean σ_i^2 would be larger, the larger the values of X_i. If that turns out to be the case, heteroscedasticity is most likely to be present in the model. To test this explicitly, Goldfeld and Quandt suggest the following steps:

Step 1. Order or rank the observations according to the values of X_i, beginning with the lowest X value.

Step 2. Omit c central observations, where c is specified a priori, and divide the remaining $(n - c)$ observations into two groups each of $(n - c)/2$ observations.

Step 3. Fit separate OLS regressions to the first $(n - c)/2$ observations and the last $(n - c)/2$ observations, and obtain the respective residual sums of squares RSS_1 and RSS_2, RSS_1 representing the RSS from the regression corresponding to the smaller X_i values (the small variance group) and RSS_2 that from the larger X_i values (the large variance group). These RSS each have

$$\frac{(n-c)}{2} - k \qquad \text{or} \qquad \left(\frac{n-c-2k}{2}\right) \text{df}$$

where k is the number of parameters to be estimated, including the intercept. (Why?) For the two-variable case k is of course 2.

Step 4. Compute the ratio

$$\lambda = \frac{\text{RSS}_2/\text{df}}{\text{RSS}_1/\text{df}} \tag{11.5.11}$$

If u_i are assumed to be normally distributed (which we usually do), and *if the assumption of homoscedasticity is valid,* then it can be shown that λ of (11.5.10) follows the F distribution with numerator and denominator df each of $(n - c - 2k)/2$.

[17] Goldfeld and Quandt, op. cit., Chap. 3.

[18] This is only one plausible assumption. Actually, what is required is that σ_i^2 be monotonically related to X_i.

If in an application the computed $\lambda (= F)$ is greater than the critical F at the chosen level of significance, we can reject the hypothesis of homoscedasticity, that is, we can say that heteroscedasticity is very likely.

Before illustrating the test, a word about omitting the c central observations is in order. These observations are omitted to sharpen or accentuate the difference between the small variance group (i.e., RSS_1) and the large variance group (i.e., RSS_2). But the ability of the Goldfeld–Quandt test to do this successfully depends on how c is chosen.[19] For the two-variable model the Monte Carlo experiments done by Goldfeld and Quandt suggest that c is about 8 if the sample size is about 30, and it is about 16 if the sample size is about 60. But Judge et al. note that $c = 4$ if $n = 30$ and $c = 10$ if n is about 60 have been found satisfactory in practice.[20]

Before moving on, it may be noted that in case there is more than one X variable in the model, the ranking of observations, the first step in the test, can be done according to any one of them. Thus in the model: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$, we can rank-order the data according to any one of these X's. If a priori we are not sure which X variable is appropriate, we can conduct the test on each of the X variables, or via a Park test, in turn, on each X.

EXAMPLE 11.4

THE GOLDFELD–QUANDT TEST

To illustrate the Goldfeld–Quandt test, we present in Table 11.3 data on consumption expenditure in relation to income for a cross section of 30 families. Suppose we postulate that consumption expenditure is linearly related to income but that heteroscedasticity is present in the data. We further postulate that the nature of heteroscedasticity is as given in (11.5.10). The necessary reordering of the data for the application of the test is also presented in Table 11.3.

Dropping the middle 4 observations, the OLS regressions based on the first 13 and the last 13 observations and their associated residual sums of squares are as shown next (standard errors in the parentheses).

Regression based on the first 13 observations:

$$\hat{Y}_i = 3.4094 + 0.6968X_i$$
$$(8.7049) \quad (0.0744) \qquad r^2 = 0.8887 \qquad \text{RSS}_1 = 377.17 \qquad \text{df} = 11$$

Regression based on the last 13 observations:

$$\hat{Y}_i = -28.0272 + 0.7941X_i$$
$$(30.6421) \quad (0.1319) \qquad r^2 = 0.7681 \qquad \text{RSS}_2 = 1536.8 \qquad \text{df} = 11$$

(*Continued*)

[19]Technically, the **power** of the test depends on how c is chosen. In statistics, the *power of a test* is measured by the probability of rejecting the null hypothesis when it is false [i.e., by 1 − Prob (type II error)]. Here the null hypothesis is that the variances of the two groups are the same, i.e., homoscedasticity. For further discussion, see M. M. Ali and C. Giaccotto, "A Study of Several New and Existing Tests for Heteroscedasticity in the General Linear Model," *Journal of Econometrics*, vol. 26, 1984, pp. 355–373.

[20]George G. Judge, R. Carter Hill, William E. Griffiths, Helmut Lütkepohl, and Tsoung-Chao Lee, *Introduction to the Theory and Practice of Econometrics*, John Wiley & Sons, New York, 1982, p. 422.

EXAMPLE 11.4 (*Continued*)

From these results we obtain

$$\lambda = \frac{\text{RSS}_2/\text{df}}{\text{RSS}_1/\text{df}} = \frac{1536.8/11}{377.17/11}$$

$$\lambda = 4.07$$

The critical F value for 11 numerator and 11 denominator df at the 5 percent level is 2.82. Since the estimated $F(=\lambda)$ value exceeds the critical value, we may conclude that there is heteroscedasticity in the error variance. However, if the level of significance is fixed at 1 percent, we may not reject the assumption of homoscedasticity. (Why?) Note that the p value of the observed λ is 0.014.

TABLE 11.3
HYPOTHETICAL DATA ON CONSUMPTION EXPENDITURE *Y*($) AND INCOME *X*($) TO ILLUSTRATE THE GOLDFELD–QUANDT TEST

		Data ranked by *X* values		
Y	*X*	*Y*	*X*	
55	80	55	80	
65	100	70	85	
70	85	75	90	
80	110	65	100	
79	120	74	105	
84	115	80	110	
98	130	84	115	
95	140	79	120	
90	125	90	125	
75	90	98	130	
74	105	95	140	
110	160	108	145	
113	150	113	150	
125	165	110	160	Middle 4 observations
108	145	125	165	
115	180	115	180	
140	225	130	185	
120	200	135	190	
145	240	120	200	
130	185	140	205	
152	220	144	210	
144	210	152	220	
175	245	140	225	
180	260	137	230	
135	190	145	240	
140	205	175	245	
178	265	189	250	
191	270	180	260	
137	230	178	265	
189	250	191	270	

Breusch–Pagan–Godfrey Test.[21] The success of the Goldfeld–Quandt test depends not only on the value of c (the number of central observations to be omitted) but also on identifying the correct X variable with which to order the observations. This limitation of this test can be avoided if we consider the **Breusch–Pagan–Godfrey (BPG) test.**

To illustrate this test, consider the k-variable linear regression model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \quad \textbf{(11.5.12)}$$

Assume that the error variance σ_i^2 is described as

$$\sigma_i^2 = f(\alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi}) \quad \textbf{(11.5.13)}$$

that is, σ_i^2 is some function of the nonstochastic variables Z's; some or all of the X's can serve as Z's. Specifically, assume that

$$\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi} \quad \textbf{(11.5.14)}$$

that is, σ_i^2 is a linear function of the Z's. If $\alpha_2 = \alpha_3 = \cdots = \alpha_m = 0$, $\sigma_i^2 = \alpha_1$, which is a constant. Therefore, to test whether σ_i^2 is homoscedastic, one can test the hypothesis that $\alpha_2 = \alpha_3 = \cdots = \alpha_m = 0$. This is the basic idea behind the Breusch–Pagan test. The actual test procedure is as follows.

Step 1. Estimate (11.5.12) by OLS and obtain the residuals $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n$.

Step 2. Obtain $\tilde{\sigma}^2 = \sum \hat{u}_i^2/n$. Recall from Chapter 4 that this is the maximum likelihood (ML) estimator of σ^2. [*Note:* The OLS estimator is $\sum \hat{u}_i^2/(n-k)$.]

Step 3. Construct variables p_i defined as

$$p_i = \hat{u}_i^2 / \tilde{\sigma}^2$$

which is simply each residual squared divided by $\tilde{\sigma}^2$.

Step 4. Regress p_i thus constructed on the Z's as

$$p_i = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi} + v_i \quad \textbf{(11.5.15)}$$

where v_i is the residual term of this regression.

Step 5. Obtain the ESS (explained sum of squares) from (11.5.15) and define

$$\Theta = \frac{1}{2}(\text{ESS}) \quad \textbf{(11.5.16)}$$

Assuming u_i are normally distributed, one can show that if there is homoscedasticity and if the sample size n increases indefinitely, then

$$\Theta \underset{\text{asy}}{\sim} \chi^2_{m-1} \quad \textbf{(11.5.17)}$$

[21]T. Breusch and A. Pagan, "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, vol. 47, 1979, pp. 1287–1294. See also L. Godfrey, "Testing for Multiplicative Heteroscedasticity," *Journal of Econometrics*, vol. 8, 1978, pp. 227–236. Because of similarity, these tests are known as Breusch–Pagan–Godfrey tests of heteroscedasticity.

that is, Θ follows the chi-square distribution with $(m - 1)$ degrees of freedom. (*Note: asy* means asymptotically.)

Therefore, if in an application the computed $\Theta\ (= \chi^2)$ exceeds the critical χ^2 value at the chosen level of significance, one can reject the hypothesis of homoscedasticity; otherwise one does not reject it.

The reader may wonder why BPG chose $\frac{1}{2}$ESS as the test statistic. The reasoning is slightly involved and is left for the references.[22]

EXAMPLE 11.5

THE BREUSCH–PAGAN–GODFREY (BPG) TEST

As an example, let us revisit the data (Table 11.3) that were used to illustrate the Goldfeld–Quandt heteroscedasticity test. Regressing Y on X, we obtain the following:

Step 1.

$$\hat{Y}_i = 9.2903 + 0.6378X_i$$
$$\text{se} = (5.2314) \quad (0.0286) \qquad \text{RSS} = 2361.153 \qquad R^2 = 0.9466 \qquad \textbf{(11.5.18)}$$

Step 2.

$$\tilde{\sigma}^2 = \sum \hat{u}_i^2/30 = 2361.153/30 = 78.7051$$

Step 3. Divide the squared residuals $\hat{u}_i$ obtained from regression (11.5.18) by 78.7051 to construct the variable p_i.

Step 4. Assuming that p_i are linearly related to $X_i(= Z_i)$ as per (11.5.14), we obtain the regression

$$\hat{p}_i = -0.7426 + 0.0101X_i$$
$$\text{se} = (0.7529) \quad (0.0041) \qquad \text{ESS} = 10.4280 \qquad R^2 = 0.18 \qquad \textbf{(11.5.19)}$$

Step 5.

$$\Theta = \frac{1}{2}(\text{ESS}) = 5.2140 \qquad \textbf{(11.5.20)}$$

Under the assumptions of the BPG test Θ in (11.5.20) asymptotically follows the chi-square distribution with 1 df. [*Note:* There is only one regressor in (11.5.19).] Now from the chi-square table we find that for 1 df the 5 percent critical chi-square value is 3.8414 and the 1 percent critical χ^2 value is 6.6349. Thus, the observed chi-square value of 5.2140 is significant at the 5 percent but not the 1 percent level of significance. Therefore, we reach the same conclusion as the Goldfeld–Quandt test. But keep in mind that, strictly speaking, the BPG test is an asymptotic, or large-sample, test and in the present example 30 observations may not constitute a large sample. It should also be pointed out that in small samples the test is sensitive to the assumption that the disturbances u_i are normally distributed. Of course, we can test the normality assumption by the tests discussed in Chapter 5.[23]

---

[22]See Adrian C. Darnell, *A Dictionary of Econometrics,* Edward Elgar, Cheltenham, U.K., 1994, pp. 178–179.

[23]On this, see R. Koenker, "A Note on Studentizing a Test for Heteroscedasticity," *Journal of Econometrics*, vol. 17, 1981, pp. 1180–1200.

White's General Heteroscedasticity Test. Unlike the Goldfeld–Quandt test, which requires reordering the observations with respect to the X variable that supposedly caused heteroscedasticity, or the BPG test, which is sensitive to the normality assumption, the general test of heteroscedasticity proposed by White does not rely on the normality assumption and is easy to implement.[24] As an illustration of the basic idea, consider the following three-variable regression model (the generalization to the k-variable model is straightforward):

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{11.5.21}$$

The White test proceeds as follows:

Step 1. Given the data, we estimate (11.5.21) and obtain the residuals, $\hat{u}_i$.

Step 2. We then run the following (*auxiliary*) regression:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i \qquad \textbf{(11.5.22)}^{25}$$

That is, the squared residuals from the original regression are regressed on the original X variables or regressors, their squared values, and the cross product(s) of the regressors. Higher powers of regressors can also be introduced. Note that there is a constant term in this equation even though the original regression may or may not contain it. Obtain the R^2 from this (auxiliary) regression.

Step 3. Under the null hypothesis that there is no heteroscedasticity, it can be shown that sample size (n) times the R^2 obtained from the auxiliary regression *asymptotically* follows the chi-square distribution with df equal to the number of regressors (excluding the constant term) in the auxiliary regression. That is,

$$n \cdot R^2 \underset{\text{asy}}{\sim} \chi^2_{\text{df}} \tag{11.5.23}$$

where df is as defined previously. In our example, there are 5 df since there are 5 regressors in the auxiliary regression.

Step 4. If the chi-square value obtained in (11.5.23) exceeds the critical chi-square value at the chosen level of significance, the conclusion is that there is heteroscedasticity. If it does not exceed the critical chi-square value, there is no heteroscedasticity, which is to say that in the auxiliary regression (11.5.21), $\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$ (see footnote 25).

---

[24] H. White, "A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity," *Econometrica*, vol. 48, 1980, pp. 817–818.

[25] Implied in this procedure is the assumption that the error variance of u_i, σ_i^2, is functionally related to the regressors, their squares, and their cross products. If all the partial slope coefficients in this regression are simultaneously equal to zero, then the error variance is the homoscedastic constant equal to α_1.

EXAMPLE 11.6

WHITE'S HETEROSCEDASTICITY TEST

From cross-sectional data on 41 countries, Stephen Lewis estimated the following regression model[26]:

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \qquad \textbf{(11.5.24)}$$

where Y = ratio of trade taxes (import and export taxes) to total government revenue, X_2 = ratio of the sum of exports plus imports to GNP, and X_3 = GNP per capita; and ln stands for natural log. His hypotheses were that Y and X_2 would be positively related (the higher the trade volume, the higher the trade tax revenue) and that Y and X_3 would be negatively related (as income increases, government finds it is easier to collect direct taxes—e.g., income tax—than rely on trade taxes).

The empirical results supported the hypotheses. For our purpose, the important point is whether there is heteroscedasticity in the data. Since the data are cross-sectional involving a heterogeneity of countries, a priori one would expect heteroscedasticity in the error variance. By applying **White's heteroscedasticity test** to the residuals obtained from regression (11.5.24), the following results were obtained[27]:

$$\begin{aligned}\widehat{u_i^2} = &-5.8417 + 2.5629 \ln \text{Trade}_i + 0.6918 \ln \text{GNP}_i \\ &-0.4081(\ln \text{Trade}_i)^2 - 0.0491(\ln \text{GNP}_i)^2 \qquad \textbf{(11.5.25)} \\ &+0.0015(\ln \text{Trade}_i)(\ln \text{GNP}_i) \qquad R^2 = 0.1148\end{aligned}$$

Note: The standard errors are not given, as they are not pertinent for our purpose here.

Now $n \cdot R^2 = 41(0.1148) = 4.7068$, which has, asymptotically, a chi-square distribution with 5 df (why?). The 5 percent critical chi-square value for 5 df is 11.0705, the 10 percent critical value is 9.2363, and the 25 percent critical value is 6.62568. For all practical purposes, one can conclude, on the basis of the White test, that there is no heteroscedasticity.

A comment is in order regarding the White test. If a model has several regressors, then introducing all the regressors, their squared (or higher-powered) terms, and their cross products can quickly consume degrees of freedom. Therefore, one must use caution in using the test.[28]

In cases where the White test statistic given in (11.5.25) is statistically significant, heteroscedasticity may not necessarily be the cause, but specification errors, about which more will be said in Chapter 13 (recall point 5 of Section 11.1). In other words, **the White test can be a test of (pure) heteroscedasticity or specification error or both.** It has been argued that if no cross-product terms are present in the White test procedure, then it is a test of pure heteroscedasticity. If cross-product terms are present, then it is a test of both heteroscedasticity and specification bias.[29]

[26]Stephen R. Lewis, "Government Revenue from Foreign Trade," *Manchester School of Economics and Social Studies*, vol. 31, 1963, pp. 39–47.

[27]These results, with change in notation, are reproduced from William F. Lott and Subhash C. Ray, *Applied Econometrics: Problems with Data Sets*, Instructor's Manual, Chap. 22, pp. 137–140.

[28]Sometimes the test can be modified to conserve degrees of freedom. See exercise 11.18.

[29]See Richard Harris, *Using Cointegration Analysis in Econometrics Modelling,* Prentice Hall & Harvester Wheatsheaf, U.K., 1995, p. 68.

Other Tests of Heteroscedasticity. There are several other tests of heteroscedasticity, each based on certain assumptions. The interested reader may want to consult the references.[30] We mention but one of these tests because of its simplicity. This is the **Koenker–Bassett (KB) test.** Like the Park, Breusch–Pagan–Godfrey, and White's tests of heteroscedasticity, the KB test is based on the squared residuals, $\hat{u}_i^2$, but instead of being regressed on one or more regressors, the squared residuals are regressed on the squared estimated values of the regressand. Specifically, if the original model is:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \qquad \textbf{(11.5.26)}$$

you estimate this model, obtain $\hat{u}_i$ from this model, and then estimate

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 (\hat{Y}_i)^2 + v_i \qquad \textbf{(11.5.27)}$$

where $\hat{Y}_i$ are the estimated values from the model (11.5.26). The null hypothesis is that $\alpha_2 = 0$. If this is not rejected, then one could conclude that there is no heteroscedasticity. The null hypothesis can be tested by the usual t test or the F test. (Note that $F_{1,k} = t_k{}^2$.) If the model (11.5.26) is double log, then the squared residuals are regressed on $(\log \hat{Y}_i)^2$. One other advantage of the KB test is that it is applicable even if the error term in the original model (11.5.26) is not normally distributed. If you apply the KB test to Example 11.1, you will find that the slope coefficient in the regression of the squared residuals obtained from (11.5.3) on the estimated $\hat{Y}_i^2$ from (11.5.3) is statistically not different from zero, thus reinforcing the Park test. This result should not be surprising since in the present instance we only have a single regressor. But the KB test is applicable if there is one regressor or many.

11.6 REMEDIAL MEASURES

As we have seen, heteroscedasticity does not destroy the unbiasedness and consistency properties of the OLS estimators, but they are no longer efficient, not even asymptotically (i.e., large sample size). This lack of efficiency makes the usual hypothesis-testing procedure of dubious value. Therefore, remedial measures may be called for. There are two approaches to remediation: when σ_i^2 is known and when σ_i^2 is not known.

When σ_i^2 Is Known: The Method of Weighted Least Squares

As we have seen in Section 11.3, if σ_i^2 is known, the most straightforward method of correcting heteroscedasticity is by means of weighted least squares, for the estimators thus obtained are BLUE.

[30]See M. J. Harrison and B. P. McCabe, "A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals," *Journal of the American Statistical Association*, vol. 74, 1979, pp. 494–499; J. Szroeter, "A Class of Parametric Tests for Heteroscedasticity in Linear Econometric Models," *Econometrica*, vol. 46, 1978, pp. 1311–1327; M. A. Evans and M. L. King, "A Further Class of Tests for Heteroscedasticity," *Journal of Econometrics*, vol. 37, 1988, pp. 265–276; R. Koenker and G. Bassett, "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica*, vol. 50, 1982, pp. 43–61.

EXAMPLE 11.7

ILLUSTRATION OF THE METHOD OF WEIGHTED LEAST SQUARES

To illustrate the method, suppose we want to study the relationship between compensation and employment size for the data presented in Table 11.1. For simplicity, we measure employment size by 1 (1–4 employees), 2 (5–9 employees), . . . , 9 (1000–2499 employees), although we could also measure it by the midpoint of the various employment classes given in the table.

Now letting Y represent average compensation per employee ($) and X the employment size, we run the following regression [see Eq. (11.3.6)]:

$$Y_i/\sigma_i = \hat{\beta}_1^*(1/\sigma_i) + \hat{\beta}_2^*(X_i/\sigma_i) + (\hat{u}_i/\sigma_i) \qquad \textbf{(11.6.1)}$$

where σ_i are the standard deviations of wages as reported in Table 11.1. The necessary raw data to run this regression are given in Table 11.4.

Before going on to the regression results, note that (11.6.1) has no intercept term. (Why?) Therefore, one will have to use the regression-through-the-origin model to estimate β_1^* and β_2^*, a topic discussed in Chapter 6. But most computer packages these days have an option to suppress the intercept term (see Minitab or Eviews, for example). Also note another interesting feature of (11.6.1): It has two explanatory variables, $(1/\sigma_i)$ and (X_i/σ_i), whereas if we were to use OLS, regressing compensation on employment size, that regression would have a single explanatory variable, X_i. (Why?)

The regression results of WLS are as follows:

$$\widehat{(Y_i/\sigma_i)} = 3406.639(1/\sigma_i) + 154.153(X_i/\sigma_i)$$
$$(80.983) \qquad (16.959) \qquad \textbf{(11.6.2)}$$
$$t = (42.066) \qquad (9.090)$$
$$R^2 = 0.9993^{31}$$

For comparison, we give the usual or unweighted OLS regression results:

$$\hat{Y}_i = 3417.833 + 148.767\,X_i$$
$$(81.136) \quad (14.418) \qquad \textbf{(11.6.3)}$$
$$t = (42.125) \quad (10.318) \qquad R^2 = 0.9383$$

In exercise 11.7 you are asked to compare these two regressions.

TABLE 11.4
ILLUSTRATION OF WEIGHTED LEAST-SQUARES REGRESSION

Compensation, Y	Employment size, X	σ_i	Y_i/σ_i	X_i/σ_i
3396	1	743.7	4.5664	0.0013
3787	2	851.4	4.4480	0.0023
4013	3	727.8	5.5139	0.0041
4104	4	805.06	5.0978	0.0050
4146	5	929.9	4.4585	0.0054
4241	6	1080.6	3.9247	0.0055
4387	7	1243.2	3.5288	0.0056
4538	8	1307.7	3.4702	0.0061
4843	9	1112.5	4.3532	0.0081

Note: In regression (11.6.2), the dependent variable is (Y_i/σ_i) and the independent variables are $(1/\sigma_i)$ and (X_i/σ_i).

Source: Data on Y and σ_i (standard deviation of compensation) are from Table 11.1. Employment size: 1 = 1–4 employees, 2 = 5–9 employees, etc. The latter data are also from Table 11.1.

[31] As noted in footnote 3 of Chap. 6, the R^2 of the regression through the origin is not directly comparable with the R^2 of the intercept-present model. The reported R^2 of 0.9993 takes this difference into account. (See the SAS package for further details about how the R^2 is corrected to take into account the absence of the intercept term. See also App. 6A, Sec. 6A1.)

When σ_i^2 Is Not Known

As noted earlier, if true σ_i^2 are known, we can use the WLS method to obtain BLUE estimators. Since the true σ_i^2 are rarely known, is there a way of obtaining *consistent* (in the statistical sense) estimates of the variances and covariances of OLS estimators even if there is heteroscedasticity? The answer is yes.

White's Heteroscedasticity-Consistent Variances and Standard Errors. White has shown that this estimate can be performed so that *asymptotically* valid (i.e., large-sample) statistical inferences can be made about the true parameter values.[32] We will not present the mathematical details, for they are beyond the scope of this book. However, Appendix 11A.4 outlines White's procedure. Nowadays, several computer packages present White's heteroscedasticity-corrected variances and standard errors along with the usual OLS variances and standard errors.[33] Incidentally, White's heteroscedasticity-corrected standard errors are also known as **robust standard errors.**

EXAMPLE 11.8

ILLUSTRATION OF WHITE'S PROCEDURE

As an example, we quote the following results due to Greene[34]:

$$\begin{aligned}
\hat{Y}_i &= 832.91 \;-\; 1834.2\,(\text{Income}) + 1587.04\,(\text{Income})^2 \\
\text{OLS se} &= (327.3) \qquad (829.0) \qquad\qquad (519.1) \\
t &= (2.54) \qquad\; (2.21) \qquad\qquad (3.06) \\
\text{White se} &= (460.9) \qquad (1243.0) \qquad\quad (830.0) \\
t &= (1.81) \qquad (-1.48) \qquad\qquad (1.91)
\end{aligned} \tag{11.6.4}$$

where Y = per capita expenditure on public schools by state in 1979 and Income = per capita income by state in 1979. The sample consisted of 50 states plus Washington, D.C.

As the preceding results show, (White's) heteroscedasticity-corrected standard errors are considerably larger than the OLS standard errors and therefore the estimated t values are much smaller than those obtained by OLS. On the basis of the latter, both the regressors are statistically significant at the 5 percent level, whereas on the basis of White estimators they are not. However, it should be pointed out that White's heteroscedasticity-corrected

[32]See H. White, op. cit.

[33]More technically, they are known as **heteroscedasticity-consistent covariance matrix estimators.**

[34]William H. Greene, *Econometric Analysis,* 2d ed., Macmillan, New York, 1993, p. 385.

standard errors can be larger or smaller than the uncorrected standard errors.

Since White's heteroscedasticity-consistent estimators of the variances are now available in established regression packages, it is recommended that the reader report them. As Wallace and Silver note:

> Generally speaking, it is probably a good idea to use the WHITE option [available in regression programs] routinely, perhaps comparing the output with regular OLS output as a check to see whether heteroscedasticity is a serious problem in a particular set of data.[35]

Plausible Assumptions about Heteroscedasticity Pattern. Apart from being a large-sample procedure, one drawback of the White procedure is that the estimators thus obtained may not be so efficient as those obtained by methods that transform data to reflect specific types of heteroscedasticity. To illustrate this, let us revert to the two-variable regression model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

We now consider several assumptions about the pattern of heteroscedasticity.

Assumption 1: The error variance is proportional to X_i^2:

$$E(u_i^2) = \sigma^2 X_i^2 \tag{11.6.5}$$[36]

If, as a matter of "speculation," graphical methods, or Park and Glejser approaches, it is believed that the variance of u_i is proportional to the square of the explanatory variable X (see Figure 11.10), one may transform the original model as follows. Divide the original model through by X_i:

$$\begin{aligned} \frac{Y_i}{X_i} &= \frac{\beta_1}{X_i} + \beta_2 + \frac{u_i}{X_i} \\ &= \beta_1 \frac{1}{X_i} + \beta_2 + v_i \end{aligned} \tag{11.6.6}$$

where v_i is the transformed disturbance term, equal to u_i/X_i. Now it is easy to verify that

$$\begin{aligned} E(v_i^2) &= E\left(\frac{u_i}{X_i}\right)^2 = \frac{1}{X_i^2} E(u_i^2) \\ &= \sigma^2 \qquad \text{using (11.6.5)} \end{aligned}$$

[35] T. Dudley Wallace and J. Lew Silver, *Econometrics: An Introduction,* Addison-Wesley, Reading, Mass., 1988, p. 265.

[36] Recall that we have already encountered this assumption in our discussion of the Goldfeld–Quandt test.

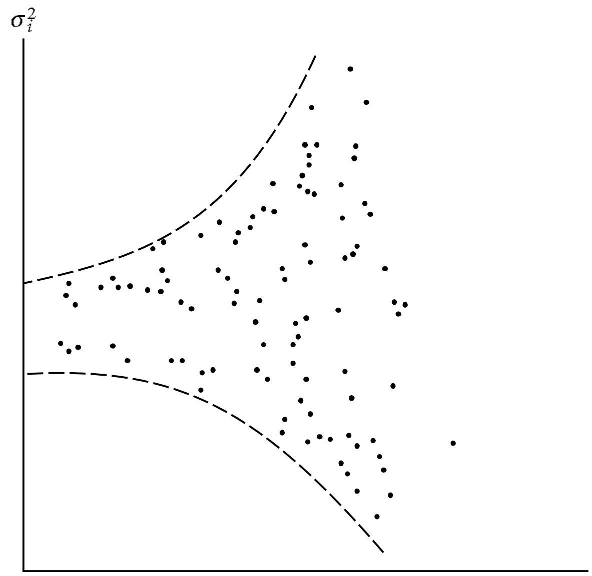

FIGURE 11.10 Error variance proportional to X^2.

Hence the variance of v_i is now homoscedastic, and one may proceed to apply OLS to the transformed equation (11.6.6), regressing Y_i/X_i on $1/X_i$.

Notice that in the transformed regression the intercept term β_2 is the slope coefficient in the original equation and the slope coefficient β_1 is the intercept term in the original model. Therefore, to get back to the original model we shall have to multiply the estimated (11.6.6) by X_i. An application of this transformation is given in exercise 11.20.

Assumption 2: The error variance is proportional to X_i. The **square root transformation:**

$$E\left(u_i^2\right) = \sigma^2 X_i \tag{11.6.7}$$

If it is believed that the variance of u_i, instead of being proportional to the squared X_i, is proportional to X_i itself, then the original model can be transformed as follows (see Figure 11.11):

$$\begin{aligned}\frac{Y_i}{\sqrt{X_i}} &= \frac{\beta_1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + \frac{u_i}{\sqrt{X_i}} \\ &= \beta_1\frac{1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + v_i\end{aligned} \tag{11.6.8}$$

where $v_i = u_i/\sqrt{X_i}$ and where $X_i > 0$.

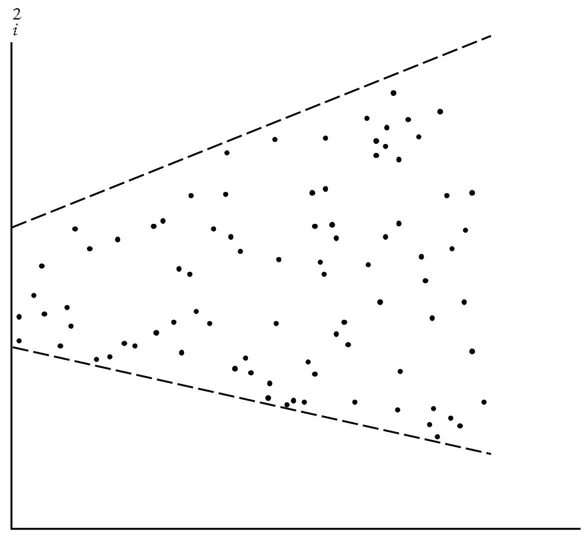

FIGURE 11.11 Error variance proportional to *X*.

Given assumption 2, one can readily verify that $E(v_i^2) = \sigma^2$, a homoscedastic situation. Therefore, one may proceed to apply OLS to (11.6.8), regressing $Y_i/\sqrt{X_i}$ on $1/\sqrt{X_i}$ and $\sqrt{X_i}$.

Note an important feature of the transformed model: It has no intercept term. Therefore, one will have to use the regression-through-the-origin model to estimate β_1 and β_2. Having run (11.6.8), one can get back to the original model simply by multiplying (11.6.8) by $\sqrt{X_i}$.

Assumption 3: The error variance is proportional to the square of the mean value of *Y*.

$$E\left(u_i^2\right) = \sigma^2[E(Y_i)]^2 \qquad \textbf{(11.6.9)}$$

Equation (11.6.9) postulates that the variance of u_i is proportional to the square of the expected value of Y (see Figure 11.8*e*). Now

$$E(Y_i) = \beta_1 + \beta_2 X_i$$

Therefore, if we transform the original equation as follows,

$$\begin{aligned} \frac{Y_i}{E(Y_i)} &= \frac{\beta_1}{E(Y_i)} + \beta_2 \frac{X_i}{E(Y_i)} + \frac{u_i}{E(Y_i)} \\ &= \beta_1 \left(\frac{1}{E(Y_i)}\right) + \beta_2 \frac{X_i}{E(Y_i)} + v_i \end{aligned} \qquad \textbf{(11.6.10)}$$

where $v_i = u_i/E(Y_i)$, it can be seen that $E(v_i^2) = \sigma^2$; that is, the disturbances v_i are homoscedastic. Hence, it is regression (11.6.10) that will satisfy the homoscedasticity assumption of the classical linear regression model.

The transformation (11.6.10) is, however, inoperational because $E(Y_i)$ depends on β_1 and β_2, which are unknown. Of course, we know $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$, which is an estimator of $E(Y_i)$. Therefore, we may proceed in two steps: First, we run the usual OLS regression, disregarding the heteroscedasticity problem, and obtain $\hat{Y}_i$. Then, using the estimated $\hat{Y}_i$, we transform our model as follows:

$$\frac{Y_i}{\hat{Y}_i} = \beta_1\left(\frac{1}{\hat{Y}_i}\right) + \beta_2\left(\frac{X_i}{\hat{Y}_i}\right) + v_i \qquad \textbf{(11.6.11)}$$

where $v_i = (u_i/\hat{Y}_i)$. In Step 2, we run the regression (11.6.11). Although $\hat{Y}_i$ are not exactly $E(Y_i)$, *they are consistent estimators;* that is, as the sample size increases indefinitely, they converge to true $E(Y_i)$. Hence, the transformation (11.6.11) will perform satisfactorily in practice if the sample size is reasonably large.

Assumption 4: A log transformation such as

$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i \qquad \textbf{(11.6.12)}$$

very often reduces heteroscedasticity when compared with the regression $Y_i = \beta_1 + \beta_2 X_i + u_i$.

This result arises because log transformation compresses the scales in which the variables are measured, thereby reducing a tenfold difference between two values to a twofold difference. Thus, the number 80 is 10 times the number 8, but $\ln 80\ (= 4.3280)$ is about twice as large as $\ln 8\ (= 2.0794)$.

An additional advantage of the log transformation is that the slope coefficient β_2 measures the elasticity of Y with respect to X, that is, the percentage change in Y for a percentage change in X. For example, if Y is consumption and X is income, β_2 in (11.6.12) will measure income elasticity, whereas in the original model β_2 measures only the rate of change of mean consumption for a unit change in income. It is one reason why the log models are quite popular in empirical econometrics. (For some of the problems associated with log transformation, see exercise 11.4.)

To conclude our discussion of the remedial measures, we reemphasize that all the transformations discussed previously are ad hoc; we are essentially speculating about the nature of σ_i^2. Which of the transformations discussed previously will work will depend on the nature of the problem and the severity of heteroscedasticity. There are some additional problems with the transformations we have considered that should be borne in mind:

1. When we go beyond the two-variable model, we may not know a priori which of the X variables should be chosen for transforming the data.[37]

2. Log transformation as discussed in Assumption 4 is not applicable if some of the Y and X values are zero or negative.[38]

3. Then there is the problem of **spurious correlation.** This term, due to Karl Pearson, refers to the situation where correlation is found to be present between the ratios of variables even though the original variables are uncorrelated or random.[39] Thus, in the model $Y_i = \beta_1 + \beta_2 X_i + u_i$, Y and X may not be correlated but in the transformed model $Y_i/X_i = \beta_1(1/X_i) + \beta_2$, Y_i/X_i and $1/X_i$ are often found to be correlated.

4. When σ_i^2 are not directly known and are estimated from one or more of the transformations that we have discussed earlier, all our testing procedures using the t tests, F tests, etc., are *strictly speaking valid only in large samples*. Therefore, one has to be careful in interpreting the results based on the various transformations in small or finite samples.[40]

11.7 CONCLUDING EXAMPLES

In concluding our discussion of heteroscedasticity we present two examples illustrating the main points made in this chapter.

EXAMPLE 11.9

CHILD MORTALITY REVISITED

Let us return to the child mortality example we have considered on several occasions. From data for 64 countries, we obtained the regression results shown in Eq. (8.2.1). Since the data are cross sectional, involving diverse countries with different child mortality experiences, it is likely that we might encounter heteroscedasticity. To find this out, let us first consider the residuals obtained from Eq. (8.2.1). These residuals are plotted in Figure 11.12. From this figure it seems that the residuals do not show any distinct pattern that might suggest heteroscedasticity. Nonetheless, appearances can be deceptive. So, let us apply the Park, Glejser, and White tests to see if there is any evidence of heteroscedasticity.

Park Test

Since there are two regressors, GNP and FLR, we can regress the squared residuals from regression (8.2.1) on either of these variables. Or, we can regress them on the estimated CM values ($= \widehat{\text{CM}}$) from regression (8.2.1). Using the latter, we obtained the following results.

$$\widehat{\hat{u}_i^2} = 854.4006 + 5.7016\,\widehat{\text{CM}}_t$$
$$t = \quad (1.2010) \quad (1.2428) \qquad r^2 = 0.024 \qquad \textbf{(11.7.1)}$$

Note: $\hat{u}_i$ are the residuals obtained from regression (8.2.1) and $\widehat{\text{CM}}$ are the estimated values of CM from regression (8.2.1).

As this regression shows, there is no systematic relation between the squared residuals and the estimated CM values (why?), suggesting that the assumption of

(*Continued*)

[37]However, as a practical matter, one may plot $\hat{u}_i^2$ against each variable and decide which X variable may be used for transforming the data. (See Fig. 11.9.)

[38]Sometimes we can use $\ln(Y_i + k)$ or $\ln(X_i + k)$, where k is a positive number chosen in such a way that all the values of Y and X become positive.

[39]For example, if X_1, X_2, and X_3 are mutually uncorrelated $r_{12} = r_{13} = r_{23} = 0$ and we find that the (values of the) ratios X_1/X_3 and X_2/X_3 are correlated, then there is spurious correlation. "More generally, correlation may be described as spurious if it is induced by the method of handling the data and is not present in the original material." M. G. Kendall and W. R. Buckland, *A Dictionary of Statistical Terms,* Hafner Publishing, New York, 1972, p. 143.

[40]For further details, see George G. Judge et al., op. cit., Sec. 14.4, pp. 415–420.

EXAMPLE 11.9 (*Continued*)

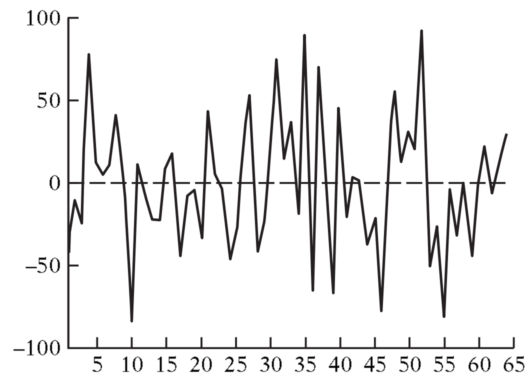

FIGURE 11.12 Residuals from regression (8.2.1).

homoscedasticity may be valid. Incidentally, regressing the log of the squared residual values on the log of $\widehat{\text{CM}}$ did not change the conclusion.

Glejser Test

The absolute values of the residual obtained from (8.2.1), when regressed on the estimated CM value from the same regression, gave the following results:

$$\widehat{|\hat{u}_i|} = 22.3127 + 0.0646\ \widehat{\text{CM}}_i \qquad \textbf{(11.7.2)}$$
$$t = (2.8086) \quad (1.2622) \qquad r^2 = 0.0250$$

Again, there is not much systematic relationship between the absolute values of the residuals and the estimated CM values, as the t value of the slope coefficient is not statistically significant.

White Test

Applying White's heteroscedasticity test with and without cross-product terms, we did not find any evidence of heteroscedasticity. We also reestimated (8.2.1) to obtain White's heteroscedasticity-consistent standard errors and t values, but the results were quite similar to those given in Eq. (8.2.1), which should not be surprising in view of the various heteroscedasticity tests we conducted earlier.

In sum, it seems that our child mortality regression (8.2.1) does not suffer from heteroscedasticity.

EXAMPLE 11.10

R&D EXPENDITURE, SALES, AND PROFITS IN 18 INDUSTRY GROUPINGS IN THE UNITED STATES, 1988

Table 11.5 gives data on research and development (R&D) expenditure, sales, and profits for 18 industry groupings in the United States, all figures in millions of dollars. Since the cross-sectional data presented in this table are quite heterogeneous, in a regression of R&D on sales (or profits), heteroscedasticity is likely. The regression results were as follows:

$$\widehat{\text{R\&D}}_i = 192.9931 + 0.0319\ \text{Sales}_i$$
$$\text{se} = (533.9317) \quad (0.0083) \qquad \textbf{(11.7.3)}$$
$$t = (0.3614) \quad (3.8433) \qquad r^2 = 0.4783$$

Unsurprisingly, there is a significant positive relationship between R&D and sale.

To see if the regression (11.7.3) suffers from heteroscedasticity, we obtained the residuals, $\hat{u}_i$, and the squared residuals, $\hat{u}_i^2$, from the preceding regression and plotted them against sales, as shown in Figure 11.13. It seems from this figure that there is a systematic

(*Continued*)

EXAMPLE 11.10 (*Continued*)

TABLE 11.5
INNOVATION IN AMERICA: RESEARCH AND DEVELOPMENT (R&D) EXPENDITURE IN THE UNITED STATES, 1988 (All Figures in Millions of Dollars)

Industry grouping	Sales	R&D expenses	Profits
1. Containers and packaging	6,375.3	62.5	185.1
2. Nonbank financial	11,626.4	92.9	1,569.5
3. Service industries	14,655.1	178.3	276.8
4. Metals and mining	21,869.2	258.4	2,828.1
5. Housing and construction	26,408.3	494.7	225.9
6. General manufacturing	32,405.6	1,083.0	3,751.9
7. Leisure time industries	35,107.7	1,620.6	2,884.1
8. Paper and forest products	40,295.4	421.7	4,645.7
9. Food	70,761.6	509.2	5,036.4
10. Health care	80,552.8	6,620.1	13,869.9
11. Aerospace	95,294.0	3,918.6	4,487.8
12. Consumer products	101,314.1	1,595.3	10,278.9
13. Electrical and electronics	116,141.3	6,107.5	8,787.3
14. Chemicals	122,315.7	4,454.1	16,438.8
15. Conglomerates	141,649.9	3,163.8	9,761.4
16. Office equipment and computers	175,025.8	13,210.7	19,774.5
17. Fuel	230,614.5	1,703.8	22,626.6
18. Automotive	293,543.0	9,528.2	18,415.4

Source: Business Week, Special 1989 Bonus Issue, R&D Scorecard, pp. 180–224.
Note: The industries are listed in increasing order of sales volume.

pattern between the residuals and squared residuals and sales, perhaps suggesting that there is heteroscedasticity. To test this formally, we used the Park, Glejser, and White tests, which gave the following results:

Park Test

$$\begin{aligned} \widehat{|\hat{u}_i^2|} &= -974{,}469.1 \quad + 86.2321\ \text{Sales}_i \\ \text{se} &= (4{,}802{,}343) \quad (40.3625) \qquad r^2 = 0.2219 \\ t &= (-0.2029) \quad (2.1364) \end{aligned} \qquad \textbf{(11.7.4)}$$

The Park test suggests that there is a statistically significant positive relationship between squared residuals and sales.

Glejser Test

$$\begin{aligned} \widehat{|\hat{u}_i|} &= 578.5710 + 0.0119\ \text{Sales}_i \\ \text{se} &= (678.6950) \quad (0.0057) \qquad r^2 = 0.214 \\ t &= (0.8524) \quad (2.0877) \end{aligned} \qquad \textbf{(11.7.5)}$$

The Glejser test also suggests that there is a systematic relationship between the absolute values of the residuals and sales, raising the possibility that the regression (11.7.3) suffers from heteroscedasticity.

(*Continued*)

EXAMPLE 11.10 (*Continued*)

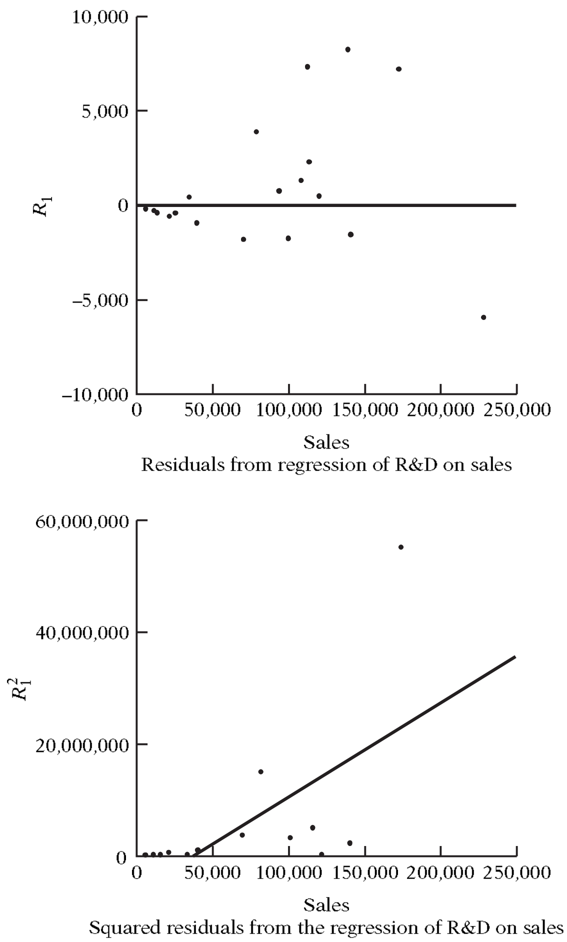

FIGURE 11.13
Residuals R_1 and squared residuals (R_1^2) on sales.

White Test

$$
\begin{aligned}
\widehat{u_i^2} &= -6{,}219{,}665 \quad + 229.3508\,\text{Sales}_i - \quad 0.000537\,\text{Sales}_i^2 \\
\text{se} &= (6{,}459{,}809) \qquad (126.2197) \qquad (0.0004) \\
t &= \quad (0.9628) \qquad (1.8170) \qquad (-1.3425)
\end{aligned}
\qquad \textbf{(11.7.6)}
$$

$$R^2 = 0.2895$$

Using the R^2 value and $n = 18$, we obtain $nR^2 = 5.2124$, which, under the null hypothesis of no heteroscedasticity, has a chi-square distribution with 2 df [because there are two regressors in (11.7.6)]. The p value of obtaining a chi-square value of as much as 5.2124 or greater

(*Continued*)

EXAMPLE 11.10 (*Continued*)

is about 0.074. If this p value is deemed sufficiently low, the White test also suggests that there is heteroscedasticity.

In sum, then, on the basis of the residual graphs and the Park, Glejser, and White tests, it seems that our R&D regression (11.7.3) suffers from heteroscedasticity. Since the true error variance is unknown, we cannot use the method of weighted least squares to obtain heteroscedasticity-corrected standard errors and t values. Therefore, we will have to make some educated guesses about the nature of the error variance.

Looking at the residual graphs given in Figure 11.13, it seems that the error variance is proportional to sales as in Eq. (11.6.7), that is, the *square root transformation.* Effecting this transformation, we obtain the following results.

$$\begin{aligned}
\frac{\widehat{\text{R\&D}}}{\sqrt{\text{Sales}}} &= -246.6769\frac{1}{\sqrt{\text{Sales}_i}} + 0.0367\sqrt{\text{Sales}_i} \\
\text{se} &= \quad (381.1285) \qquad\qquad (0.0071) \qquad\qquad R^2 = 0.3648 \qquad \textbf{(11.7.7)} \\
t &= \quad (-0.6472) \qquad\qquad (5.1690)
\end{aligned}$$

If you want, you can multiply the preceding equation by $\sqrt{\text{Sales}_i}$ to get back to the original model. Comparing (11.7.7) with (11.7.3), you can see that the slope coefficients in the two equations are about the same, but their standard errors are different. In (11.7.3) it was 0.0083, whereas in (11.7.7) it is only 0.0071, a decrease of about 14 percent.

To conclude our example, we present below White's heteroscedasticity-consistent standard errors, as discussed in Section 11.6.

$$\begin{aligned}
\widehat{\text{R\&D}}_i &= \quad 192.9931 + 0.0319\ \text{Sales}_i \\
\text{se} &= (533.9931) \quad (0.0101) \qquad\qquad r^2 = 0.4783 \qquad \textbf{(11.7.8)} \\
t &= \quad (0.3614) \quad (3.1584)
\end{aligned}$$

Comparing with the original (i.e., without correction for heteroscedasticity) regression (11.7.3), we see that although the parameter estimates have not changed (as we would expect), the standard error of the intercept coefficient has decreased and that of the slope coefficient has slightly increased. But remember that the White procedure is strictly a large-sample procedure, whereas we only have 18 observations.

11.8 A CAUTION ABOUT OVERREACTING TO HETEROSCEDASTICITY

Reverting to the R&D example discussed in the previous section, we saw that when we used the square root transformation to correct for heteroscedasticity in the original model (11.7.3), the standard error of the slope coefficient decreased and its t value increased. Is this change so significant that one should worry about it in practice? To put the matter differently, when should we really worry about the heteroscedasticity problem? As one author contends, "heteroscedasticity has never been a reason to throw out an otherwise good model."[41]

[41]N. Gregory Mankiw, "A Quick Refresher Course in Macroeconomics," *Journal of Economic Literature,* vol. XXVIII, December 1990, p. 1648.

Here it may be useful to bear in mind the caution sounded by John Fox:

> . . . unequal error variance is worth correcting only when the problem is severe.
> The impact of nonconstant error variance on the efficiency of ordinary least-squares estimator and on the validity of least-squares inference depends on several factors, including the sample size, the degree of variation in the σ_i^2, the configuration of the X [i.e., regressor] values, and the relationship between the error variance and the X's. It is therefore not possible to develop wholly general conclusions concerning the harm produced by heteroscedasticity.[42]

Returning to the model (11.3.1), we saw earlier that variance of the slope estimator, var $(\hat{\beta}_2)$, is given by the usual formula shown in (11.2.3). Under GLS the variance of the slope estimator, var $(\hat{\beta}_2^*)$, is given by (11.3.9). We know that the latter is more efficient than the former. But how large does the former (i.e., OLS) variance have to be in relation to the GLS variance before one should really worry about it? As a rule of thumb, Fox suggests that we worry about this problem ". . . when the largest error variance is more than about 10 times the smallest."[43] Thus, returning to the Monte Carlo simulations results of Davidson and MacKinnon presented earlier, consider the value of $\alpha = 2$. The variance of the estimated β_2 is 0.04 under OLS and 0.012 under GLS, the ratio of the former to the latter thus being about 3.33.[44] According to the Fox rule, the severity of heteroscedasticity in this case may not be large enough to worry about it.

Also remember that, despite heteroscedasticity, OLS estimators are linear unbiased and are (under general conditions) asymptotically (i.e., in large samples) normally distributed.

As we will see when we discuss other violations of the assumptions of the classical linear regression model, the caution sounded in this section is appropriate as a general rule. Otherwise, one can go overboard.

11.9 SUMMARY AND CONCLUSIONS

1. A critical assumption of the classical linear regression model is that the disturbances u_i have all the same variance, σ^2. If this assumption is not satisfied, there is heteroscedasticity.

2. Heteroscedasticity does not destroy the unbiasedness and consistency properties of OLS estimators.

3. But these estimators are no longer minimum variance or efficient. That is, they are not BLUE.

4. The BLUE estimators are provided by the method of weighted least squares, provided the heteroscedastic error variances, σ_i^2, are known.

5. In the presence of heteroscedasticity, the variances of OLS estimators are not provided by the usual OLS formulas. But if we persist in using the

[42]John Fox, *Applied Regression Analysis, Linear Models, and Related Methods,* Sage Publications, California, 1997, p. 306.

[43]Ibid., p. 307.

[44]Note that we have squared the standard errors to obtain the variances.

usual OLS formulas, the t and F tests based on them can be highly misleading, resulting in erroneous conclusions.

6. Documenting the consequences of heteroscedasticity is easier than detecting it. There are several diagnostic tests available, but one cannot tell for sure which will work in a given situation.

7. Even if heteroscedasticity is suspected and detected, it is not easy to correct the problem. If the sample is large, one can obtain White's heteroscedasticity corrected standard errors of OLS estimators and conduct statistical inference based on these standard errors.

8. Otherwise, on the basis of OLS residuals, one can make educated guesses of the likely pattern of heteroscedasticity and transform the original data in such a way that in the transformed data there is no heteroscedasticity.

EXERCISES

Questions

11.1. State *with brief reason* whether the following statements are true, false, or uncertain:

a. In the presence of heteroscedasticity OLS estimators are biased as well as inefficient.

b. If heteroscedasticity is present, the conventional t and F tests are invalid.

c. In the presence of heteroscedasticity the usual OLS method always overestimates the standard errors of estimators.

d. If residuals estimated from an OLS regression exhibit a systematic pattern, it means heteroscedasticity is present in the data.

e. There is no general test of heteroscedasticity that is free of any assumption about which variable the error term is correlated with.

f. If a regression model is mis-specified (e.g., an important variable is omitted), the OLS residuals will show a distinct pattern.

g. If a regressor that has nonconstant variance is (incorrectly) omitted from a model, the (OLS) residuals will be heteroscedastic.

11.2. In a regression of average wages (W, \$) on the number of employees (N) for a random sample of 30 firms, the following regression results were obtained*:

$$\begin{aligned}\widehat{W} &= 7.5 + 0.009N \\ t &= \text{n.a.} \quad (16.10) \qquad R^2 = 0.90\end{aligned} \tag{1}$$

$$\begin{aligned}\widehat{W}/N &= 0.008 + 7.8(1/N) \\ t &= (14.43) \quad (76.58) \qquad R^2 = 0.99\end{aligned} \tag{2}$$

a. How do you interpret the two regressions?

b. What is the author assuming in going from Eq. (1) to (2)? Was he worried about heteroscedasticity? How do you know?

*See Dominick Salvatore, *Managerial Economics,* McGraw-Hill, New York, 1989, p. 157.

c. Can you relate the slopes and intercepts of the two models?
d. Can you compare the R^2 values of the two models? Why or why not?

11.3. **a.** Can you estimate the parameters of the models

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

by the method of ordinary least squares? Why or why not?
b. If not, can you suggest a method, informal or formal, of estimating the parameters of such models? (See Chapter 14.)

11.4. Although log models as shown in Eq. (11.6.12) often reduce heteroscedasticity, one has to pay careful attention to the properties of the disturbance term of such models. For example, the model

$$Y_i = \beta_1 X_i^{\beta_2} u_i \tag{1}$$

can be written as

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + \ln u_i \tag{2}$$

a. If $\ln u_i$ is to have zero expectation, what must be the distribution of u_i?
b. If $E(u_i) = 1$, will $E(\ln u_i) = 0$? Why or why not?
c. If $E(\ln u_i)$ is not zero, what can be done to make it zero?

11.5. Show that β_2^* of (11.3.8) can also be expressed as

$$\beta_2^* = \frac{\sum w_i y_i^* x_i^*}{\sum w_i x_i^{2*}}$$

and var (β_2^*) given in (11.3.9) can also be expressed as

$$\text{var}\,(\beta_2^*) = \frac{1}{\sum w_i x_i^{2*}}$$

where $y_i^* = Y_i - \bar{Y}^*$ and $x_i^* = X_i - \bar{X}^*$ represent deviations from the weighted means $\bar{Y}^*$ and $\bar{X}^*$ defined as

$$\bar{Y}^* = \sum w_i Y_i \Big/ \sum w_i$$

$$\bar{X}^* = \sum w_i X_i \Big/ \sum w_i$$

11.6. For pedagogic purposes Hanushek and Jackson estimate the following model:

$$C_t = \beta_1 + \beta_2 \text{GNP}_t + \beta_3 D_t + u_i \tag{1}$$

where C_t = aggregate private consumption expenditure in year t, GNP_t = gross national product in year t, and D = national defense expenditures in year t, the objective of the analysis being to study the effect of defense expenditures on other expenditures in the economy.

Postulating that $\sigma_t^2 = \sigma^2(\text{GNP}_t)^2$, they transform (1) and estimate

$$C_t/\text{GNP}_t = \beta_1\,(1/\text{GNP}_t) + \beta_2 + \beta_3\,(D_t/\text{GNP}_t) + u_t/\text{GNP}_t \qquad \textbf{(2)}$$

The empirical results based on the data for 1946–1975 were as follows (standard errors in the parentheses)*:

$$\begin{array}{lllll} \hat{C}_t = 26.19 & + 0.6248\ \text{GNP}_t - & 0.4398\ D_t & \\ \quad (2.73) & (0.0060) & (0.0736) & R^2 = 0.999 \\ \widehat{C_t/\text{GNP}_t} = 25.92\,(1/\text{GNP}_t) & + 0.6246 & -\ 0.4315\,(D_t/\text{GNP}_t) & \\ \quad (2.22) & (0.0068) & (0.0597) & R^2 = 0.875 \end{array}$$

a. What assumption is made by the authors about the nature of heteroscedasticity? Can you justify it?
b. Compare the results of the two regressions. Has the transformation of the original model improved the results, that is, reduced the estimated standard errors? Why or why not?
c. Can you compare the two R^2 values? Why or why not? (*Hint:* Examine the dependent variables.)

11.7. Refer to the estimated regressions (11.6.2) and (11.6.3). The regression results are quite similar. What could account for this outcome?

11.8. Prove that if $w_i = w$ a constant, for each i, β_2^* and $\hat{\beta}_2$ as well as their variance are identical.

11.9. Refer to formulas (11.2.2) and (11.2.3). Assume

$$\sigma_i^2 = \sigma^2 k_i$$

where σ^2 is a constant and where k_i are *known* weights, not necessarily all equal.

Using this assumption, show that the variance given in (11.2.2) can be expressed as

$$\text{var}\,(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \cdot \frac{\sum x_i^2 k_i}{\sum x_i^2}$$

The first term on the right side is the variance formula given in (11.2.3), that is, var $(\hat{\beta}_2)$ under homoscedasticity. What can you say about the nature of the relationship between var $(\hat{\beta}_2)$ under heteroscedasticity and under homoscedasticity? (*Hint:* Examine the second term on the right side of the preceding formula.) Can you draw any general conclusions about the relationships between (11.2.2) and (11.2.3)?

11.10. In the model

$$Y_i = \beta_2 X_i + u_i \quad (\textit{Note:}\text{ there is no intercept})$$

*Eric A. Hanushek and John E. Jackson, *Statistical Methods for Social Scientists,* Academic, New York, 1977, p. 160.

you are told that $\text{var}(u_i) = \sigma^2 X_i^2$. Show that

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2 \sum X_i^4}{\left(\sum X_i^2\right)^2}$$

Problems

11.11. For the data given in Table 11.1, regress average compensation Y on average productivity X, treating employment size as the unit of observation. Interpret your results, and see if your results agree with those given in (11.5.3).

a. From the preceding regression obtain the residuals $\hat{u}_i$.

b. Following the Park test, regress $\ln \hat{u}_i^2$ on $\ln X_i$ and verify the regression (11.5.4).

c. Following the Glejser approach, regress $|\hat{u}_i|$ on X_i and then regress $|\hat{u}_i|$ on $\sqrt{X_i}$ and comment on your results.

d. Find the rank correlation between $|\hat{u}_i|$ and X_i and comment on the nature of heteroscedasticity, if any, present in the data.

11.12. Table 11.6 gives data on the sales/cash ratio in U.S. manufacturing industries classified by the asset size of the establishment for the period 1971–I to 1973–IV. (The data are on a quarterly basis.) The sales/cash ratio may be regarded as a measure of income velocity in the corporate sector, that is, the number of times a dollar turns over.

a. For each asset size compute the mean and standard deviation of the sales/cash ratio.

b. Plot the mean value against the standard deviation as computed in **a**, using asset size as the unit of observation.

c. By means of a suitable regression model decide whether standard deviation of the ratio increases with the mean value. If not, how would you rationalize the result?

TABLE 11.6 ASSET SIZE (Millions of Dollars)

Year and quarter	1–10	10–25	25–50	50–100	100–250	250–1000	1000 +
1971–I	6.696	6.929	6.858	6.966	7.819	7.557	7.860
–II	6.826	7.311	7.299	7.081	7.907	7.685	7.351
–III	6.338	7.035	7.082	7.145	7.691	7.309	7.088
–IV	6.272	6.265	6.874	6.485	6.778	7.120	6.765
1972–I	6.692	6.236	7.101	7.060	7.104	7.584	6.717
–II	6.818	7.010	7.719	7.009	8.064	7.457	7.280
–III	6.783	6.934	7.182	6.923	7.784	7.142	6.619
–IV	6.779	6.988	6.531	7.146	7.279	6.928	6.919
1973–I	7.291	7.428	7.272	7.571	7.583	7.053	6.630
–II	7.766	9.071	7.818	8.692	8.608	7.571	6.805
–III	7.733	8.357	8.090	8.357	7.680	7.654	6.772
–IV	8.316	7.621	7.766	7.867	7.666	7.380	7.072

Source: Quarterly Financial Report for Manufacturing Corporations, Federal Trade Commission and the Securities and Exchange Commission, U.S. government, various issues (computed).

d. If there is a statistically significant relationship between the two, how would you transform the data so that there is no heteroscedasticity?

11.13. Bartlett's homogeneity-of-variance test.* Suppose there are k independent sample variances $s_1^2, s_2^2, \ldots, s_k^2$ with $f_1, f_2, \ldots, f_k$ df, each from populations which are normally distributed with mean μ and variance σ_i^2. Suppose further that we want to test the null hypothesis H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$; that is, each sample variance is an estimate of the same population variance σ^2.

If the null hypothesis is true, then

$$s^2 = \frac{\sum_{i=1}^{k} f_i s_i^2}{\sum f_i} = \frac{\sum f_i s_i^2}{f}$$

provides an estimate of the common (pooled) estimate of the population variance σ^2, where $f_i = (n_i - 1)$, n_i being the number of observations in the ith group and where $f = \sum_{i=1}^{k} f_i$.

Bartlett has shown that the null hypothesis can be tested by the ratio A/B, which is approximately distributed as the χ^2 distribution with $k - 1$ df, where

$$A = f \ln s^2 - \sum (f_i \ln s_i^2)$$

and

$$B = 1 + \frac{1}{3(k-1)}\left[\sum\left(\frac{1}{f_i}\right) - \frac{1}{f}\right]$$

Apply Bartlett's test to the data of Table 11.1 and verify that the hypothesis that population variances of employee compensation are the same in each employment size of the establishment cannot be rejected at the 5 percent level of significance.

Note: f_i, the df for each sample variance, is 9, since n_i for each sample (i.e., employment class) is 10.

11.14. Consider the following regression-through-the origin model:

$$Y_i = \beta X_i + u_i, \quad \text{for } i = 1, 2$$

You are told that $u_1 \sim N(0, \sigma^2)$ and $u_2 \sim N(0, 2\sigma^2)$ and that they are statistically independent. If $X_1 = +1$ and $X_2 = -1$, obtain the *weighted* least-squares (WLS) estimate of β and its variance. If in this situation you had assumed incorrectly that the two error variances are the same (say, equal to σ^2), what would be the OLS estimator of β? And its variance? Compare these estimates with the estimates obtained by the method of WLS? What general conclusion do you draw?†

11.15. Table 11.7 gives data on 81 cars about MPG (average miles per gallons), HP (engine horsepower), VOL (cubic feet of cab space), SP (top speed, miles per hour), and WT (vehicle weight in 100 lb).

*See "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Society of London A,* vol. 160, 1937, p. 268.

†Adapted from F. A. F. Seber, *Linear Regression Analysis,* John Wiley & Sons, New York, 1977, p. 64.

TABLE 11.7 PASSENGER CAR MILAGE DATA

Observation	MPG	SP	HP	VOL	WT	Observation	MPG	SP	HP	VOL	WT
1	65.4	96	49	89	17.5	42	32.2	106	95	106	30.0
2	56.0	97	55	92	20.0	43	32.2	109	102	92	30.0
3	55.9	97	55	92	20.0	44	32.2	106	95	88	30.0
4	49.0	105	70	92	20.0	45	31.5	105	93	102	30.0
5	46.5	96	53	92	20.0	46	31.5	108	100	99	30.0
6	46.2	105	70	89	20.0	47	31.4	108	100	111	30.0
7	45.4	97	55	92	20.0	48	31.4	107	98	103	30.0
8	59.2	98	62	50	22.5	49	31.2	120	130	86	30.0
9	53.3	98	62	50	22.5	50	33.7	109	115	101	35.0
10	43.4	107	80	94	22.5	51	32.6	109	115	101	35.0
11	41.1	103	73	89	22.5	52	31.3	109	115	101	35.0
12	40.9	113	92	50	22.5	53	31.3	109	115	124	35.0
13	40.9	113	92	99	22.5	54	30.4	133	180	113	35.0
14	40.4	103	73	89	22.5	55	28.9	125	160	113	35.0
15	39.6	100	66	89	22.5	56	28.0	115	130	124	35.0
16	39.3	103	73	89	22.5	57	28.0	102	96	92	35.0
17	38.9	106	78	91	22.5	58	28.0	109	115	101	35.0
18	38.8	113	92	50	22.5	59	28.0	104	100	94	35.0
19	38.2	106	78	91	22.5	60	28.0	105	100	115	35.0
20	42.2	109	90	103	25.0	61	27.7	120	145	111	35.0
21	40.9	110	92	99	25.0	62	25.6	107	120	116	40.0
22	40.7	101	74	107	25.0	63	25.3	114	140	131	40.0
23	40.0	111	95	101	25.0	64	23.9	114	140	123	40.0
24	39.3	105	81	96	25.0	65	23.6	117	150	121	40.0
25	38.8	111	95	89	25.0	66	23.6	122	165	50	40.0
26	38.4	110	92	50	25.0	67	23.6	122	165	114	40.0
27	38.4	110	92	117	25.0	68	23.6	122	165	127	40.0
28	38.4	110	92	99	25.0	69	23.6	122	165	123	40.0
29	46.9	90	52	104	27.5	70	23.5	148	245	112	40.0
30	36.3	112	103	107	27.5	71	23.4	160	280	50	40.0
31	36.1	103	84	114	27.5	72	23.4	121	162	135	40.0
32	36.1	103	84	101	27.5	73	23.1	121	162	132	40.0
33	35.4	111	102	97	27.5	74	22.9	110	140	160	45.0
34	35.3	111	102	113	27.5	75	22.9	110	140	129	45.0
35	35.1	102	81	101	27.5	76	19.5	121	175	129	45.0
36	35.1	106	90	98	27.5	77	18.1	165	322	50	45.0
37	35.0	106	90	88	27.5	78	17.2	140	238	115	45.0
38	33.2	109	102	86	30.0	79	17.0	147	263	50	45.0
39	32.9	109	102	86	30.0	80	16.7	157	295	119	45.0
40	32.3	120	130	92	30.0	81	13.2	130	236	107	55.0
41	32.2	106	95	113	30.0						

Note:
VOL = cubic feet of cab space
HP = engine horsepower
MPG = average miles per gallon
SP = top speed, miles per hour
WT = vehicle weight, hundreds of pounds
Observation = car observation number (Names of cars not disclosed)
Source: U.S. Environmental Protection Agency, 1991, Report EPA/AA/CTAB/91-02.

a. Consider the following model:

$$\text{MPG}_i = \beta_1 + \beta_2\text{SP} + \beta_3\text{HP} + \beta_4\text{WT} + u_i$$

Estimate the parameters of this model and interpret the results. Do they make economic sense?

b. Would you expect the error variance in the preceding model to be heteroscedastic? Why?

c. Use the White test to find out if the error variance is heteroscedastic.

d. Obtain White's heteroscedasticity-consistent standard errors and t values and compare your results with those obtained from OLS.

e. If heteroscedasticity is established, how would you transform the data so that in the transformed data the error variance is homoscedastic? Show the necessary calculations.

11.16. Food expenditure in India. In Table 2.8 we have given data on expenditure on food and total expenditure for 55 families in India.

a. Regress expenditure on food on total expenditure, and examine the residuals obtained from this regression.

b. Plot the residuals obtained in **a** against total expenditure and see if you observe any systematic pattern.

c. If the plot in **b** suggests that there is heteroscedasticity, apply the Park, Glejser, and White tests to find out if the impression of heteroscedasticity observed in **b** is supported by these tests.

d. Obtain White's heteroscedasticity-consistent standard errors and compare those with the OLS standard errors. Decide if it is worth correcting for heteroscedasticity in this example.

11.17. Repeat exercise 11.16, but this time regress the logarithm of expenditure on food on the logarithm of total expenditure. If you observe heteroscedasticity in the linear model of exercise 11.16 but not in the log–linear model, what conclusion do you draw? Show all the necessary calculations.

11.18. A shortcut to White's test. As noted in the text, the White test can consume degrees of freedom if there are several regressors and if we introduce all the regressors, their squared terms, and their cross products. Therefore, instead of estimating regressions like (11.5.22), why not simply run the following regression:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 \hat{Y}_i + \alpha_2 \hat{Y}_i^2 + \nu_i$$

where $\hat{Y}_i$ are the estimated Y (i.e., regressand) values from whatever model you are estimating? After all, $\hat{Y}_i$ is simply the weighted average of the regressors, with the estimated regression coefficients serving as the weights.

Obtain the R^2 value from the preceding regression and use (11.5.22) to test the hypothesis that there is no heteroscedasticity.

Apply the preceding test to the food expenditure example of exercise 11.16.

11.19. Return to the R&D example discussed in Section 11.7. Repeat the example using profits as the regressor. A priori, would you expect your

TABLE 11.8 MEDIAN SALARIES OF FULL PROFESSORS IN STATISTICS, 2000–2001

Years in rank	Count	Median
0 to 1	11	$69,000
2 to 3	20	$70,500
4 to 5	26	$74,050
6 to 7	33	$82,600
8 to 9	18	$91,439
10 to 11	26	$83,127
12 to 13	31	$84,700
14 to 15	15	$82,601
16 to 17	22	$93,286
18 to 19	23	$90,400
20 to 21	13	$98,200
22 to 24	29	$100,000
25 to 27	22	$99,662
28 to 32	22	$116,012
33 or more	11	$85,200

Source: American Statistical Association, "2000–2001 Salary Report of Academic Statisticians," *Amstat News*, Issue 282, December 2000, p. 4.

results to be different from those using sales as the regressor? Why or why not?

11.20. Table 11.8 gives data on median salaries of full professors in statistics in research universities in the United States for the academic year 2000–2001.

a. Plot median salaries against years in rank (as a measure of years of experience). For the plotting purposes, assume that the median salaries refer to the midpoint of years in rank. Thus, the salary $74,050 in the range 4–5 refers to 4.5 years in the rank, and so on. For the last group, assume that the range is 33–35.

b. Consider the following regression models:

$$Y_i = \alpha_1 + \alpha_2 X_i + u_i \quad \textbf{(1)}$$

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + v_i \quad \textbf{(2)}$$

where Y = median salary, X = year in rank (measured at midpoint of the range), and u and v are the error terms. Can you argue why model (2) might be preferable to model (1)? From the data given, estimate both the models.

c. If you observe heteroscedasticity in model (1) but not in model (2), what conclusion would you draw? Show the necessary computations.

d. If heteroscedasticity is observed in model (2), how would you transform the data so that in the transformed model there is no heteroscedasticity?

11.21. You are given the following data:

$$\text{RSS}_1 \text{ based on the first 30 observations} = 55, \text{ df} = 25$$

$$\text{RSS}_2 \text{ based on the last 30 observations} = 140, \text{ df} = 25$$

Carry out the Goldfeld–Quandt test of heteroscedasticity at the 5 percent level of significance.

11.22. Table 11.9 gives data on percent change per year for stock prices (Y) and consumer prices (X) for a cross section of 20 countries.

a. Plot the data in a scattergram.

b. Regress Y on X and examine the residuals from this regression. What do you observe?

c. Since the data for Chile seem atypical (outlier?), repeat the regression in **b**, dropping the data on Chile. Now examine the residuals from this regression. What do you observe?

d. If on the basis of the results in **b** you conclude that there was heteroscedasticity in error variance but on the basis of the results in **c** you reverse your conclusion, what general conclusions do you draw?

TABLE 11.9 STOCK AND CONSUMER PRICES, POST–WORLD WAR II PERIOD (Through 1969)

	Rate of change, % per year	
Country	Stock prices, *Y*	Consumer prices, *X*
1. Australia	5.0	4.3
2. Austria	11.1	4.6
3. Belgium	3.2	2.4
4. Canada	7.9	2.4
5. Chile	25.5	26.4
6. Denmark	3.8	4.2
7. Finland	11.1	5.5
8. France	9.9	4.7
9. Germany	13.3	2.2
10. India	1.5	4.0
11. Ireland	6.4	4.0
12. Israel	8.9	8.4
13. Italy	8.1	3.3
14. Japan	13.5	4.7
15. Mexico	4.7	5.2
16. Netherlands	7.5	3.6
17. New Zealand	4.7	3.6
18. Sweden	8.0	4.0
19. United Kingdom	7.5	3.9
20. United States	9.0	2.1

Source: Phillip Cagan, *Common Stock Values and Inflation: The Historical Record of Many Countries,* National Bureau of Economic Research, Suppl., March 1974, Table 1, p. 4.

APPENDIX 11A

11A.1 PROOF OF EQUATION (11.2.2)

From Appendix 3A, Section 3A.3, we have

$$\begin{aligned}\operatorname{var}(\hat{\beta}_2) &= E\left(k_1^2 u_1^2 + k_2^2 u_2^2 + \cdots + k_n^2 u_n^2 + 2 \text{ cross-product terms}\right)\\ &= E\left(k_1^2 u_1^2 + k_2^2 u_2^2 + \cdots + k_n^2 u_n^2\right)\end{aligned}$$

since the expectations of the cross-product terms are zero because of the assumption of no serial correlation,

$$\operatorname{var}(\hat{\beta}_2) = k_1^2 E\left(u_1^2\right) + k_2^2 E\left(u_2^2\right) + \cdots + k_n^2 E\left(u_n^2\right)$$

since the k_i are known. (Why?)

$$\operatorname{var}(\hat{\beta}_2) = k_1^2 \sigma_1^2 + k_2^2 \sigma_2^2 + \cdots + k_n^2 \sigma_n^2$$

since $E(u_i^2) = \sigma_i^2$.

$$\begin{aligned}\operatorname{var}(\hat{\beta}_2) &= \sum k_i^2 \sigma_i^2\\ &= \sum \left[\left(\frac{x_i}{\sum x_i^2}\right)^2 \sigma_i^2\right] \qquad \text{since } k_i = \frac{x_i}{\sum x_i^2} \qquad \textbf{(11.2.2)}\\ &= \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}\end{aligned}$$

11A.2 THE METHOD OF WEIGHTED LEAST SQUARES

To illustrate the method, we use the two-variable model $Y_i = \beta_1 + \beta_2 X_i + u_i$. The unweighted least-squares method minimizes

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \tag{1}$$

to obtain the estimates, whereas the weighted least-squares method minimizes the weighted residual sum of squares:

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)^2 \tag{2}$$

where β_1^* and β_2^* are the weighted least-squares estimators and where the weights w_i are such that

$$w_i = \frac{1}{\sigma_i^2} \tag{3}$$

that is, the weights are inversely proportional to the variance of u_i or Y_i conditional upon the given X_i, it being understood that $\text{var}(u_i \mid X_i) = \text{var}(Y_i \mid X_i) = \sigma_i^2$.

Differentiating (2) with respect to $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$, we obtain

$$\frac{\partial \sum w_i \hat{u}_i^2}{\partial \beta_1^*} = 2\sum w_i(Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)(-1)$$

$$\frac{\partial \sum w_i \hat{u}_i^2}{\partial \beta_2^*} = 2\sum w_i(Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)(-X_i)$$

Setting the preceding expressions equal to zero, we obtain the following two normal equations:

$$\sum w_i Y_i = \hat{\beta}_1^* \sum w_i + \hat{\beta}_2^* \sum w_i X_i \tag{4}$$

$$\sum w_i X_i Y_i = \hat{\beta}_1^* \sum w_i X_i + \hat{\beta}_2^* \sum w_i X_i^2 \tag{5}$$

Notice the similarity between these normal equations and the normal equations of the unweighted least squares.

Solving these equations simultaneously, we obtain

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^* \tag{6}$$

and

$$\hat{\beta}_2^* = \frac{\left(\sum w_i\right)\left(\sum w_i X_i Y_i\right) - \left(\sum w_i X_i\right)\left(\sum w_i Y_i\right)}{\left(\sum w_i\right)\left(\sum w_i X_i^2\right) - \left(\sum w_i X_i\right)^2} \qquad \textbf{(11.3.8) = (7)}$$

The variance of $\hat{\beta}_2^*$ shown in (11.3.9) can be obtained in the manner of the variance of $\hat{\beta}_2$ shown in Appendix 3A, Section 3A.3.

Note: $\bar{Y}^* = \sum w_i Y_i / \sum w_i$ and $\bar{X}^* = \sum w_i X_i / \sum w_i$. As can be readily verified, these weighted means coincide with the usual or unweighted means $\bar{Y}$ and $\bar{X}$ when $w_i = w$, a constant, for all i.

11A.3 PROOF THAT $E(\hat{\sigma}^2) \neq \sigma^2$ IN THE PRESENCE OF HETEROSCEDASTICITY

Consider the two-variable model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{1}$$

where $\text{var}(u_i) = \sigma_i^2$

Now

$$\begin{aligned}\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2} &= \frac{\sum [\beta_1 + \beta_2 X_i + u_i - \hat{\beta}_1 - \hat{\beta}_2 X_i]^2}{n-2} \\ &= \frac{\sum [-(\hat{\beta}_1 - \beta_1) - (\hat{\beta}_2 - \beta_2) X_i + u_i]^2}{n-2}\end{aligned} \tag{2}$$

Noting that $(\hat{\beta}_1 - \beta_1) = -(\hat{\beta}_2 - \beta_2)\overline{X} + \overline{u}$, and substituting this into (2) and taking expectations on both sides, we get:

$$E(\hat{\sigma}^2) = \frac{1}{n-2}\left\{-\sum x_i^2 \operatorname{var}(\hat{\beta}_2) + E\left[\sum (u_i - \bar{u})^2\right]\right\}$$
$$= \frac{1}{n-2}\left[-\frac{\sum x_i^2\sigma_i^2}{\sum x_i^2} + \frac{(n-1)\sum \sigma_i^2}{n}\right] \tag{3}$$

where use is made of (11.2.2).

As you can see from (3), if there is homoscedasticity, that is, $\sigma_i^2 = \sigma^2$ for each i, $E(\hat{\sigma}^2) = \sigma^2$. Therefore, the expected value of the conventionally computed $\hat{\sigma}^2 = \sum \hat{u}^2/(n-2)$ will not be equal to the true σ^2 in the presence of heteroscedasticity.[1]

11A.4 WHITE'S ROBUST STANDARD ERRORS

To give you some idea about White's heteroscedasticity-corrected standard errors, consider the two-variable regression model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \operatorname{var}(u_i) = \sigma_i^2. \tag{1}$$

As shown in (11.2.2),

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2} \tag{2}$$

Since σ_i^2 are not directly observable, White suggests using $\hat{u}_i^2$, the squared residual for each i, in place of σ_i^2 and estimate the var $(\hat{\beta}_2)$ as follows:

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \hat{u}_i^2}{\left(\sum x_i^2\right)^2} \tag{3}$$

White has shown that (3) is a consistent estimator of (2), that is, as the sample size increases indefinitely, (3) converges to (2).[2]

Incidentally, note that if your software package does not contain White's robust standard error procedure, you can do it as shown in (3) by first running the usual OLS regression, obtaining the residuals from this regression and then using formula (3).

White's procedure can be generalized to the k-variable regression model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{4}$$

---

[1]Further details can be obtained from Jan Kmenta, *Elements of Econometrics,* 2d. ed., Macmillan, New York, 1986, pp. 276–278.

[2]To be more precise, n times (3) converges in probability to $E[(X_i - \mu_X)^2 u_i^2]/(\sigma_X^2)^2$, which is the probability limit of n times (2), where n is the sample size, μ_x is the expected value of X, and σ_X^2 is the (population) variance of X. For more details, see Jeffrey M. Wooldridge, *Introductory Econometrics; A Modern Approach,* South-Western Publishing, 2000, p. 250.

The variance of any partial regression coefficient, say $\hat{\beta}_j$, is obtained as follows:

$$\operatorname{var}(\hat{\beta}_j) = \frac{\sum \hat{w}_{ji}^2 \hat{u}_i^2}{\left(\sum \hat{w}_{ji}^2\right)^2} \tag{5}$$

where $\hat{u}_i$ are the residuals obtained from the (original) regression (4) and $\hat{w}_j$ are the residuals obtained from the (auxiliary) regression of the regressor X_j on the remaining regressors in (4).

Obviously, this is a time-consuming procedure, for you will have to estimate (5) for each X variable. Of course, all this labor can be avoided if you have a statistical package that does this routinely. Packages such as PcGive, Eviews, Microfit, Shazam, Stata, and Limdep now obtain White's heteroscedasticity-robust standard errors very easily.