CHAPTER 5

Violations of the Classical Assumptions

5.1 Introduction

In this chapter, we relax the assumptions made in Chapter 3 one by one and study the effect of that on the OLS estimator. In case the OLS estimator is no longer a viable estimator, we derive an alternative estimator and propose some tests that will allow us to check whether this assumption is violated.

5.2 The Zero Mean Assumption

Violation of assumption 1 implies that the mean of the disturbances is no longer zero. Two cases are considered:

Case 1: $E(u_i) = \mu \neq 0$

The disturbances have a common mean which is not zero. In this case, one can subtract μ from the u_i's and get new disturbances $u_i^* = u_i - \mu$ which have zero mean and satisfy all the other assumptions imposed on the u_i's. Having subtracted μ from u_i we add it to the constant α leaving the regression equation intact:

$$Y_i = \alpha^* + \beta X_i + u_i^* \quad i = 1, 2, \ldots, n \tag{5.1}$$

where $\alpha^* = \alpha + \mu$. It is clear that only α^* and β can be estimated, and not α nor μ. In other words, one cannot retrieve α and μ from an estimate of α^* without additional assumptions or further information, see problem 10. With this reparameterization, equation (5.1) satisfies the four classical assumptions, and therefore OLS gives the BLUE estimators of α^* and β. Hence, a constant non-zero mean for the disturbances affects only the intercept estimate but not the slope. Fortunately, in most economic applications, it is the slope coefficients that are of interest and not the intercept.

Case 2: $E(u_i) = \mu_i$

The disturbances have a mean which varies with every observation. In this case, one can transform the regression equation as in (5.1) by adding and subtracting μ_i. The problem, however, is that $\alpha_i^* = \alpha + \mu_i$ now varies with each observation, and hence we have more parameters than observations. In fact, there are n intercepts and one slope to be estimated with n observations. Unless we have repeated observations like in panel data, see Chapter 12 or we have some prior information on these α_i^*, we cannot estimate this model.

B.H. Baltagi, *Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-20059-5_5,

5.3 Stochastic Explanatory Variables

Sections 5.5 and 5.6 will study violations of assumptions 2 and 3 in detail. This section deals with violations of assumption 4 and its effect on the properties of the OLS estimators. In this case, X is a random variable which may be (i) independent; (ii) contemporaneously uncorrelated; or (iii) simply correlated with the disturbances.

Case 1: If X is independent of u, then all the results of Chapter 3 still hold, but now they are conditional on the particular set of X's drawn in the sample. To illustrate this result, recall that for the simple linear regression:

$$\widehat{\beta}_{OLS} = \beta + \sum_{i=1}^{n} w_i u_i \text{ where } w_i = x_i / \sum_{i=1}^{n} x_i^2 \tag{5.2}$$

Hence, when we take expectations $E(\sum_{i=1}^{n} w_i u_i) = \sum_{i=1}^{n} E(w_i)E(u_i) = 0$. The first equality holds because X and u are independent and the second equality holds because the u's have zero mean. In other words the unbiasedness property of the OLS estimator still holds. However, the

$$\text{var}(\widehat{\beta}_{OLS}) = E(\sum_{i=1}^{n} w_i u_i)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} E(w_i w_j)E(u_i u_j) = \sigma^2 \sum_{i=1}^{n} E(w_i^2)$$

where the last equality follows from assumptions 2 and 3, homoskedasticity and no serial correlation. The only difference between this result and that of Chapter 3 is that we have expectations on the X's rather than the X's themselves. Hence, by conditioning on the particular set of X's that are observed, we can use all the results of Chapter 3. Also, maximizing the likelihood involves both the X's and the u's. But, as long as the distribution of the X's does not involve the parameters we are estimating, i.e., α, β and σ^2, the same maximum likelihood estimators are obtained. Why? Because $f(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_n) = f_1(x_1, x_2, \ldots, x_n)f_2(u_1, u_2, \ldots, u_n)$ since the X's and the u's are independent. Maximizing f with respect to $(\alpha, \beta, \sigma^2)$ is the same as maximizing f_2 with respect to $(\alpha, \beta, \sigma^2)$ as long as f_1 is not a function of these parameters.

Case 2: Consider a simple model of consumption, where Y_t, current consumption, is a function of Y_{t-1}, consumption in the previous period. This is the case for a habit forming consumption good like cigarette smoking. In this case our regression equation becomes

$$Y_t = \alpha + \beta Y_{t-1} + u_t \quad t = 2, \ldots, T \tag{5.3}$$

where we lost one observation due to lagging. It is obvious that Y_t is correlated to u_t, but the question here is whether Y_{t-1} is correlated to u_t. After all, Y_{t-1} is our explanatory variable X_t. As long as assumption 3 is not violated, i.e., the u's are not correlated across periods, u_t represents a freshly drawn disturbance independent of previous disturbances and hence is not correlated with the already *predetermined* Y_{t-1}. This is what we mean by contemporaneously uncorrelated, i.e., u_t is correlated with Y_t, but it is not correlated with Y_{t-1}. The OLS estimator of β is

$$\widehat{\beta}_{OLS} = \sum_{t=2}^{T} y_t y_{t-1} / \sum_{t=2}^{T} y_{t-1}^2 = \beta + \sum_{t=2}^{T} y_{t-1} u_t / \sum_{t=2}^{T} y_{t-1}^2 \tag{5.4}$$

and the expected value of (5.4) is not β because in general,

$$E(\sum_{t=2}^{T} y_{t-1} u_t / \sum_{t=2}^{T} y_{t-1}^2) \neq E(\sum_{t=2}^{T} y_{t-1} u_t)/E(\sum_{t=2}^{T} y_{t-1}^2).$$

The expected value of a ratio is not the ratio of expected values. Also, even if $E(Y_{t-1}u_t) = 0$, one can easily show that $E(y_{t-1}u_t) \neq 0$. In fact, $y_{t-1} = Y_{t-1} - \bar{Y}$, and $\bar{Y}$ contains Y_t in it, and we know that $E(Y_t u_t) \neq 0$. Hence, we lost the unbiasedness property of OLS. However, all the asymptotic properties still hold. In fact, $\widehat{\beta}_{OLS}$ is consistent because

$$\text{plim } \widehat{\beta}_{OLS} = \beta + \text{cov}(Y_{t-1}, u_t)/\text{var}(Y_{t-1}) = \beta \tag{5.5}$$

where the second equality follows from (5.4) and the fact that $\text{plim}(\sum_{t=2}^{T} y_{t-1}u_t/T)$ is $\text{cov}(Y_{t-1}, u_t)$ which is zero, and $\text{plim}(\sum_{t=2}^{T} y_{t-1}^2/T) = \text{var}(Y_{t-1})$ which is positive and finite.

Case 3: X and u are correlated, in this case OLS is biased and inconsistent. This can be easily deduced from (5.2) since $\text{plim}(\sum_{i=1}^{n} x_i u_i/n)$ is the $\text{cov}(X, u) \neq 0$, and $\text{plim}(\sum_{i=1}^{n} x_i^2/n)$ is positive and finite. This means that OLS is no longer a viable estimator, and an alternative estimator that corrects for this bias has to be derived. In fact we will study three specific cases where this assumption is violated. These are: (i) the errors in measurement case; (ii) the case of a lagged dependent variable with correlated errors; and (iii) simultaneous equations.

Briefly, the errors in measurement case involves a situation where the true regression model is in terms of X^*, but X^* is measured with error, i.e., $X_i = X_i^* + \nu_i$, so we observe X_i but not X_i^*. Hence, when we substitute this X_i for X_i^* in the regression equation, we get

$$Y_i = \alpha + \beta X_i^* + u_i = \alpha + \beta X_i + (u_i - \beta\nu_i) \tag{5.6}$$

where the composite error term is now correlated with X_i because X_i is correlated with ν_i. After all, $X_i = X_i^* + \nu_i$ and $E(X_i\nu_i) = E(\nu_i^2)$ if X_i^* and ν_i are uncorrelated.

Similarly, in case (ii) above, if the u's were correlated across time, i.e., u_{t-1} is correlated with u_t, then Y_{t-1}, which is a function of u_{t-1}, will also be correlated with u_t, and $E(Y_{t-1}u_t) \neq 0$. More on this and how to test for serial correlation in the presence of a lagged dependent variable in Chapter 6.

Finally, if one considers a demand and supply equations where quantity Q_t is a function of price P_t in both equations

$$Q_t = \alpha + \beta P_t + u_t \quad \text{(demand)} \tag{5.7}$$

$$Q_t = \delta + \gamma P_t + \nu_t \quad \text{(supply)} \tag{5.8}$$

The question here is whether P_t is correlated with the disturbances u_t and ν_t in both equations. The answer is yes, because (5.7) and (5.8) are two equations in two unknowns P_t and Q_t. Solving for these variables, one gets P_t as well as Q_t as a function of a constant and both u_t and ν_t. This means that $E(P_t u_t) \neq 0$ and $E(P_t \nu_t) \neq 0$ and OLS performed on either (5.7) or (5.8) is biased and inconsistent. We will study this simultaneous bias problem more rigorously in Chapter 11.

For all situations where X and u are correlated, it would be illuminating to show graphically why OLS is no longer a consistent estimator. Let us consider the case where the disturbances are, say, positively correlated with the explanatory variable. Figure 3.3 of Chapter 3 shows the true regression line $\alpha + \beta X_i$. It also shows that when X_i and u_i are positively correlated then an X_i higher than its mean will be associated with a disturbance u_i above its mean, i.e., a positive disturbance. Hence, $Y_i = \alpha + \beta X_i + u_i$ will always be above the true regression line whenever X_i is above its mean. Similarly Y_i would be below the true regression line for every X_i below its mean. This means that not knowing the true regression line, a researcher fitting OLS on this data will have a biased intercept and slope. In fact, the intercept will be understated and the slope will be overstated. Furthermore, this bias does not disappear with more data, since this new data will be generated by the same mechanism described above. Hence these OLS estimates are inconsistent.

Similarly, if X_i and u_i are negatively correlated, the intercept will be overstated and the slope will be understated. This story applies to any equation with at least one of its right hand side variables correlated with the disturbance term. Correlation due to the lagged dependent variable with autocorrelated errors, is studied in Chapter 6, whereas the correlation due to the simultaneous equations problem is studied in Chapter 11.

5.4 Normality of the Disturbances

If the disturbance are not normal, OLS is still BLUE provided assumptions 1–4 still hold. Normality made the OLS estimators minimum variance unbiased MVU and these OLS estimators turn out to be identical to the MLE. Normality allowed the derivation of the distribution of these estimators and this in turn allowed testing of hypotheses using the t and F-tests considered in the previous chapter. If the disturbances are not normal, yet the sample size is large, one can still use the normal distribution for the OLS estimates asymptotically by relying on the Central Limit Theorem, see Theil (1978). Theil's proof is for the case of fixed X's in repeated samples, zero mean and constant variance on the disturbances. A simple asymptotic test for the normality assumption is given by Jarque and Bera (1987). This is based on the fact that the normal distribution has a skewness measure of zero and a kurtosis of 3. Skewness (or lack of symmetry) is measured by

$$S = \frac{[E(X-\mu)^3]^2}{[E(X-\mu)^2]^3} = \frac{\textit{Square of the 3rd moment about the mean}}{\textit{Cube of the variance}}$$

Kurtosis (a measure of flatness) is measured by

$$\kappa = \frac{E(X-\mu)^4}{[E(X-\mu)^2]^2} = \frac{\textit{4th moment about the mean}}{\textit{Square of the variance}}$$

For the normal distribution $S = 0$ and $\kappa = 3$. Hence, the Jarque-Bera (JB) statistic is given by

$$JB = n\left[\frac{S^2}{6} + \frac{(\kappa-3)^2}{24}\right]$$

where S represents skewness and κ represents kurtosis of the OLS residuals. This statistic is asymptotically distributed as χ^2 with two degrees of freedom under H_0. Rejecting H_0, rejects normality of the disturbances but does not offer an alternative distribution. In this sense, the test is non-constructive. In addition, not rejecting H_0 does not necessarily mean that the distribution of the disturbances is normal, it only means we do not reject that the distribution of the disturbances is symmetric and has a kurtosis of 3. See the empirical example in section 5.5 for an illustration. The Jarque-Bera test is part of the standard output using EViews.

5.5 Heteroskedasticity

Violation of assumption 2, means that the disturbances have a varying variance, i.e., $E(u_i^2) = \sigma_i^2$, $i = 1, 2, \ldots, n$. First, we study the effect of this violation on the OLS estimators. For the simple linear regression it is obvious that $\widehat{\beta}_{OLS}$ given in equation (5.2) is still unbiased and consistent because these properties depend upon assumptions 1 and 4, and not assumption 2. However, the variance of $\widehat{\beta}_{OLS}$ is now different

$$\text{var}(\widehat{\beta}_{OLS}) = \text{var}(\textstyle\sum_{i=1}^n w_i u_i) = \sum_{i=1}^n w_i^2 \sigma_i^2 = \sum_{i=1}^n x_i^2 \sigma_i^2 / (\sum_{i=1}^n x_i^2)^2 \qquad (5.9)$$

where the second equality follows from assumption 3 and the fact that var(u_i) is now σ_i^2. Note that if $\sigma_i^2 = \sigma^2$, this reverts back to $\sigma^2/\sum_{i=1}^n x_i^2$, the usual formula for var($\widehat{\beta}_{OLS}$) under homoskedasticity. Furthermore, one can show that $E(s^2)$ will involve all of the σ_i^2's and not one common σ^2, see problem 1. This means that the regression package reporting $s^2/\sum_{i=1}^n x_i^2$ as the estimate of the variance of $\widehat{\beta}_{OLS}$ is committing two errors. One, it is not using the right formula for the variance, i.e., equation (5.9). Second, it is using s^2 to estimate a common σ^2 when in fact the σ_i^2's are different. The bias from using $s^2/\sum_{i=1}^n x_i^2$ as an estimate of var($\widehat{\beta}_{OLS}$) will depend upon the nature of the heteroskedasticity and the regressor. In fact, if σ_i^2 is positively related to x_i^2, one can show that $s^2/\sum_{i=1}^n x_i^2$ understates the true variance and hence the t-statistic reported for $\beta = 0$ is overblown, and the confidence interval for β is tighter than it is supposed to be, see problem 2. This means that the t-statistic in this case is biased towards rejecting $H_0; \beta = 0$, i.e., showing significance of the regression slope coefficient, when it may not be significant.

The OLS estimator of β is linear unbiased and consistent, but is it still BLUE? In order to answer this question, we note that the only violation we have is that the var$(u_i) = \sigma_i^2$. Hence, if we divided u_i by σ_i/σ, the resulting $u_i^* = \sigma u_i/\sigma_i$ will have a constant variance σ^2. It is easy to show that u^* satisfies all the classical assumptions including homoskedasticity. The regression model becomes

$$\sigma Y_i/\sigma_i = \alpha\sigma/\sigma_i + \beta\sigma X_i/\sigma_i + u_i^* \qquad (5.10)$$

and OLS on this model (5.10) is BLUE. The OLS normal equations on (5.10) are

$$\begin{aligned} \textstyle\sum_{i=1}^n (Y_i/\sigma_i^2) &= \alpha \textstyle\sum_{i=1}^n (1/\sigma_i^2) + \beta \sum_{i=1}^n (X_i/\sigma_i^2) \\ \textstyle\sum_{i=1}^n (Y_i X_i/\sigma_i^2) &= \alpha \textstyle\sum_{i=1}^n (X_i/\sigma_i^2) + \beta \sum_{i=1}^n (X_i^2/\sigma_i^2) \end{aligned} \qquad (5.11)$$

Note that σ^2 drops out of these equations. Solving (5.11), see problem 3, one gets

$$\widetilde{\alpha} = [\textstyle\sum_{i=1}^n (Y_{1i}/\sigma_i^2)/\sum_{i=1}^n (1/\sigma_i^2)] - \widetilde{\beta}[\sum_{i=1}^n (X_i/\sigma_i^2)/\sum_{i=1}^n (1/\sigma_i^2)] = \bar{Y}^* - \widetilde{\beta}\bar{X}^* \qquad (5.12a)$$

with $\bar{Y}^* = [\sum_{i=1}^n (Y_i/\sigma_i^2)/\sum_{i=1}^n (1/\sigma_i^2)] = \sum_{i=1}^n w_i^* Y_i/\sum_{i=1}^n w_i^*$ and

$$\bar{X}^* = [\textstyle\sum_{i=1}^n (X_i/\sigma_i^2)/\sum_{i=1}^n (1/\sigma_i^2)] = \sum_{i=1}^n w_i^* X_i/\sum_{i=1}^n w_i^*$$

where $w_i^* = (1/\sigma_i^2)$. Similarly,

$$\begin{aligned} \widetilde{\beta} &= \frac{[\sum_{i=1}^n (1/\sigma_i^2)][\sum_{i=1}^n (Y_i X_i/\sigma_i^2)] - [\sum_{i=1}^n (X_i/\sigma_i^2)][\sum_{i=1}^n (Y_i/\sigma_i^2)]}{[\sum_{i=1}^n X_i^2/\sigma_i^2][\sum_{i=1}^n (1/\sigma_i^2)] - [\sum_{i=1}^n (X_i/\sigma_i^2)]^2} \\ &= \frac{(\sum_{i=1}^n w_i^*)(\sum_{i=1}^n w_i^* X_i Y_i) - (\sum_{i=1}^n w_i^* X_i)(\sum_{i=1}^n w_i^* Y_i)}{(\sum_{i=1}^n w_i^*)(\sum_{i=1}^n w_i^* X_i^2) - (\sum_{i=1}^n w_i^* X_i)^2} \qquad (5.12b) \\ &= \frac{\sum_{i=1}^n w_i^*(X_i - \bar{X}^*)(Y_i - \bar{Y}^*)}{\sum_{i=1}^n w_i^*(X_i - \bar{X}^*)^2} \end{aligned}$$

It is clear that the BLU estimators $\widetilde{\alpha}$ and $\widetilde{\beta}$, obtained from the regression in (5.10), are different from the usual OLS estimators $\widehat{\alpha}_{OLS}$ and $\widehat{\beta}_{OLS}$ since they depend upon the σ_i^2's. It is also true that when $\sigma_i^2 = \sigma^2$ for all $i = 1, 2, \ldots, n$, i.e., under homoskedasticity, (5.12) reduces to the usual OLS estimators given by equation (3.4) of Chapter 3. The BLU estimators weight the i-th observation by $(1/\sigma_i)$ which is a measure of precision of that observation. The more precise the observation, i.e., the smaller σ_i, the larger is the weight attached to that observation. $\widetilde{\alpha}$ and $\widetilde{\beta}$ are also known as *Weighted Least Squares* (WLS) estimators which are a specific form of *Generalized Least Squares* (GLS). We will study GLS in details in Chapter 9, using matrix notation.

Under heteroskedasticity, OLS looses efficiency in that it is no longer BLUE. However, because it is still unbiased and consistent and because the true σ_i^2's are never known some researchers compute OLS as an initial consistent estimator of the regression coefficients. It is important to emphasize however, that the standard errors of these estimates as reported by the regression package are biased and any inference based on these estimated variances including the reported t-statistics are misleading. White (1980) proposed a simple procedure that would yield heteroskedasticity consistent standard errors of the OLS estimators. In equation (5.9), this amounts to replacing σ_i^2 by e_i^2, the square of the i-th OLS residual, i.e.,

$$White\text{'s } \mathrm{var}(\widehat{\beta}_{OLS}) = \sum_{i=1}^n x_i^2 e_i^2 / (\sum_{i=1}^n x_i^2)^2 \tag{5.13}$$

Note that we can not consistently estimate σ_i^2 by e_i^2, since there is one observation per parameter estimated. As the sample size increases, so does the number of unknown σ_i^2's. What White (1980) consistently estimates is the $\mathrm{var}(\widehat{\beta}_{OLS})$ which is a weighted average of the e_i^2. The same analysis applies to the multiple regression OLS estimates. In this case, White's (1980) heteroskedasticity consistent estimate of the variance of the k-th OLS regression coefficient β_k, is given by

$$White\text{'s } \mathrm{var}(\widehat{\beta}_k) = \sum_{i=1}^n \widehat{\nu}_{ki}^2 e_i^2 / (\sum_{i=1}^n \widehat{\nu}_{ki}^2)^2$$

where $\widehat{\nu}_k^2$ is the squared OLS residual obtained from regressing X_k on the remaining regressors in the equation being estimated. e_i is the i-th OLS residual from this multiple regression equation. Many regression packages provide White's heteroskedasticity-consistent estimates of the variances and their corresponding robust t-statistics. For example, using EViews, one clicks on Quick, choose Estimate Equation. Now click on Options, a menu appears where one selects White to obtain the heteroskedasticity-consistent estimates of the variances.

While the regression packages correct for heteroskedasticity in the t-statistics they do not usually do that for the F-statistics studied, say in Example 2 in Chapter 4. Wooldridge (1991) suggests a simple way of obtaining a robust LM statistic for H_0; $\beta_2 = \beta_3 = 0$ in the multiple regression (4.1). This involves the following steps:

(1) Run OLS on the restricted model without X_2 and X_3 and obtain the restricted least squares residuals $\widetilde{u}$.

(2) Regress each of the independent variables excluded under the null (i.e., X_2 and X_3) on *all* of the other included independent variables (i.e., $X_4, X_5, \ldots, X_K$) including the constant. Get the corresponding residuals $\widehat{v}_2$ and $\widehat{v}_3$, respectively.

(3) Regress the dependent variable equal to 1 for all observations on $\widehat{v}_2\widetilde{u}, \widehat{v}_3\widetilde{u}$ without a constant and obtain the robust LM statistic equal to the n- the sum of squared residuals of

this regression. This is exactly nR_u^2 of this last regression. Under H_0 this LM statistic is distributed as χ_2^2.

Since OLS is no longer BLUE, one should compute $\widetilde{\alpha}$ and $\widetilde{\beta}$. The only problem is that the σ_i's are rarely known. One example where the σ_i's are known up to a scalar constant is the following simple example of aggregation.

Example 5.1: *Aggregation and Heteroskedasticity.* Let Y_{ij} be the observation on the j-th firm in the i-th industry, and consider the following regression:

$$Y_{ij} = \alpha + \beta X_{ij} + u_{ij} \quad j = 1, 2, \ldots, n_i;\ i = 1, 2, \ldots, m \tag{5.14}$$

If only aggregate observations on each industry are available, then (5.14) is summed over firms, i.e.,

$$Y_i = \alpha n_i + \beta X_i + u_i \qquad i = 1, 2, \ldots, m \tag{5.15}$$

where $Y_i = \sum_{j=1}^{n_i} Y_{ij}$, $X_i = \sum_{j=1}^{n_i} X_{ij}$, $u_i = \sum_{j=1}^{n_i} u_{ij}$ for $i = 1, 2, \ldots, m$. Note that although the u_{ij}'s are IID$(0, \sigma^2)$, by aggregating, we get $u_i \sim (0, n_i\sigma^2)$. This means that the disturbances in (5.15) are heteroskedastic. However, $\sigma_i^2 = n_i\sigma^2$ and is known up to a scalar constant. In fact, σ/σ_i is $1/(n_i)^{1/2}$. Therefore, premultiplying (5.15) by $1/(n_i)^{1/2}$ and performing OLS on the transformed equation results in BLU estimators of α and β. In other words, BLU estimation reduces to performing OLS of $Y_i/(n_i)^{1/2}$ on $(n_i)^{1/2}$ and $X_i/(n_i)^{1/2}$, without an intercept.

There may be other special cases in practice where σ_i is known up to a scalar, but in general, σ_i is usually unknown and will have to be estimated. This is hopeless with only n observations, since there are n σ_i's, so we either have to have repeated observations, or know more about the σ_i's. Let us discuss these two cases.

Case 1: *Repeated Observations*

Suppose that n_i households are selected randomly with income X_i for $i = 1, 2, \ldots, m$. For each household $j = 1, 2, \ldots, n_i$, we observe its consumption expenditures on food, say Y_{ij}. The regression equation is

$$Y_{ij} = \alpha + \beta X_i + u_{ij} \qquad i = 1, 2, \ldots, m\ ;\ j = 1, 2, \ldots, n_i \tag{5.16}$$

where m is the number of income groups selected. Note that X_i has only one subscript, whereas Y_{ij} has two subscripts denoting the repeated observations on households with the same income X_i. The u_{ij}'s are independently distributed $(0, \sigma_i^2)$ reflecting the heteroskedasticity in consumption expenditures among the different income groups. In this case, there are $n = \sum_{i=1}^{m} n_i$ observations and m σ_i^2's to be estimated. This is feasible, and there are two methods for estimating these σ_i^2's. The first is to compute

$$\widehat{s}_i^2 = \textstyle\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2/(n_i - 1)$$

where $\bar{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$. The second is to compute $\widetilde{s}_i^2 = \sum_{j=1}^{n_i} e_{ij}^2/n_i$ where e_{ij} is the OLS residual given by

$$e_{ij} = Y_{ij} - \widehat{\alpha}_{OLS} - \widehat{\beta}_{OLS} X_i$$

Both estimators of σ_i^2 are consistent. Substituting either $\widetilde{s}_i^2$ or $\widehat{s}_i^2$ for σ_i^2 in (5.12) will result in feasible estimators of α and β. However, the resulting estimates are no longer BLUE. The substitution of the consistent estimators of σ_i^2 is justified on the basis that the resulting α and β estimates will be asymptotically efficient, see Chapter 9. Of course, this step could have been replaced by a regression of $Y_{ij}/\widehat{s}_i$ on $(1/\widehat{s}_i)$ and $(X_i/\widehat{s}_i)$ without a constant, or the similar regression in terms of $\widetilde{s}_i$. For this latter estimate, $\widetilde{s}_i^2$, one can iterate, i.e., obtaining new residuals based on the new regression estimates and therefore new $\widetilde{s}_i^2$. The process continues until the estimates obtained from the r-th iteration do not differ from those of the $(r+1)$th iteration in absolute value by more than a small arbitrary positive number chosen as the convergence criterion. Once the estimates converge, the final round estimators are the maximum likelihood estimators, see Oberhofer and Kmenta (1974).

Case 2: *Assuming More Information on the Form of Heteroskedasticity*

If we do not have repeated observations, it is hopeless to try and estimate n variances and α and β with only n observations. More structure on the form of heteroskedasticity is needed to estimate this model, but not necessarily to test it. Heteroskedasticity is more likely to occur with cross-section data where the observations may be on firms with different size. For example, a regression relating profits to sales might have heteroskedasticity, because larger firms have more resources to draw upon, can borrow more, invest more, and loose or gain more than smaller firms. Therefore, we expect the form of heteroskedasticity to be related to the size of the firm, which is reflected in this case by the regressor, sales, or some other variable that measures size, like assets. Hence, for this regression we can write $\sigma_i^2 = \sigma^2 Z_i^2$, where Z_i denotes the sales or assets of firm i. Once again the form of heteroskedasticity is known up to a scalar constant and the BLU estimators of α and β can be obtained from (5.12), assuming Z_i is known. Alternatively, one can run the regression of Y_i/Z_i on $1/Z_i$ and X_i/Z_i without a constant to get the same result. Special cases of Z_i are X_i and $E(Y_i)$. (i) If $Z_i = X_i$ the regression becomes that of Y_i/X_i on $1/X_i$ and a constant. Note that the regression coefficient of $1/X_i$ is the estimate of α, while the constant of the regression is now the estimate of β. But, is it possible to have u_i uncorrelated with X_i when we are assuming $\text{var}(u_i)$ related to X_i? The answer is yes, as long as $E(u_i/X_i) = 0$, i.e., the mean of u_i is zero for every value of X_i, see Figure 3.4 of Chapter 3. This, in turn, implies that the overall mean of the u_i's is zero, i.e., $E(u_i) = 0$ and that $\text{cov}(X_i, u_i) = 0$. If the latter is not satisfied and say $\text{cov}(X_i, u_i)$ is positive, then large values of X_i imply large values of u_i. This would mean that for these values of X_i, we have a non-zero mean for the corresponding u_i's. This contradicts $E(u_i/X_i) = 0$. Hence, if $E(u_i/X_i) = 0$, then $\text{cov}(X_i, u_i) = 0$. (ii) If $Z_i = E(Y_i) = \alpha + \beta X_i$, then σ_i^2 is proportional to the population regression line, which is a linear function of α and β. Since the OLS estimates are consistent one can estimate $E(Y_i)$ by $\widehat{Y}_i = \widehat{\alpha}_{OLS} + \widehat{\beta}_{OLS} X_i$ use $\widehat{Z}_i = \widehat{Y}_i$ instead of $E(Y_i)$. In other words, run the regression of $Y_i/\widehat{Y}_i$ on $1/\widehat{Y}_i$ and $X_i/\widehat{Y}_i$ without a constant. The resulting estimates are asymptotically efficient, see Amemiya (1973).

One can generalize $\sigma_i^2 = \sigma^2 Z_i^2$ to $\sigma_i^2 = \sigma^2 Z_i^{\delta}$ where δ is an unknown parameter to be estimated. Hence rather than estimating n σ_i^2's one has to estimate only σ^2 and δ. Assuming normality one can set up the likelihood function and derive the first-order conditions by differentiating that likelihood with respect to α, β, σ^2 and δ. The resulting equations are highly nonlinear. Alternatively, one can search over possible values for $\delta = 0, 0.1, 0.2, \ldots, 4$, and get the corresponding estimates of α, β, and σ^2 from the regression of $Y_i/Z_i^{\delta/2}$ on $1/Z_i^{\delta/2}$ and $X_i/Z_i^{\delta/2}$

without a constant. This is done for every δ and the value of the likelihood function is reported. Using this search procedure one can get the maximum value of the likelihood and corresponding to it the MLE of α, β, σ^2 and δ. Note that as δ increases so does the degree of heteroskedasticity. Problem 4 asks the reader to compute the relative efficiency of the OLS estimator with respect to the BLU estimator for $Z_i = X_i$ for various values of δ. As expected the relative efficiency of the OLS estimator declines as the degree of heteroskedasticity increases.

One can also generalize $\sigma_i^2 = \sigma^2 Z_i^\delta$ to include more Z variables. In fact, a general form of this multiplicative heteroskedasticity is

$$\log\sigma_i^2 = \log\sigma^2 + \delta_1 \log Z_{1i} + \delta_2 \log Z_{2i} + \ldots + \delta_r \log Z_{ri} \tag{5.17}$$

with $r < n$, otherwise one cannot estimate with n observations. $Z_1, Z_2, \ldots, Z_r$ are known variables determining the heteroskedasticity. Note that if $\delta_2 = \delta_3 = \ldots = \delta_r = 0$, we revert back to $\sigma_i^2 = \sigma^2 Z_i^\delta$, where $\delta = \delta_1$. For the estimation of this general multiplicative form of heteroskedasticity, see Harvey (1976).

Another form for heteroskedasticity, is the additive form

$$\sigma_i^2 = a + b_1 Z_{1i} + b_2 Z_{2i} + \ldots + b_r Z_{ri} \tag{5.18}$$

where $r < n$, see Goldfeld and Quandt (1972). Special cases of (5.18) include

$$\sigma_i^2 = a + b_1 X_i + b_2 X_i^2 \tag{5.19}$$

where if a and b_1 are zero we have a simple form of multiplicative heteroskedasticity. In order to estimate the regression model with additive heteroskedasticity of the type given in (5.19), one can get the OLS residuals, the e_i's, and run the following regression

$$e_i^2 = a + b_1 X_i + b_2 X_i^2 + v_i \tag{5.20}$$

where $v_i = e_i^2 - \sigma_i^2$. The v_i's are heteroskedastic, and the OLS estimates of (5.20) yield the following estimates of σ_i^2

$$\widehat{\sigma}_i^2 = \widehat{a}_{OLS} + \widehat{b}_{1,OLS} X_i + \widehat{b}_{2,OLS} X_i^2 \tag{5.21}$$

One can obtain a better estimate of the σ_i^2's by computing the following regression which corrects for the heteroskedasticity in the v_i's

$$(e_i^2/\widehat{\sigma}_i) = a(1/\widehat{\sigma}_i) + b_1(X_i/\widehat{\sigma}_i) + b_2(X_i^2/\widehat{\sigma}_i) + w_i \tag{5.22}$$

The new estimates of σ_i^2 are

$$\widetilde{\sigma}_i^2 = \widetilde{a} + \widetilde{b}_1 X_i + \widetilde{b}_2 X_i^2 \tag{5.23}$$

where $\widetilde{a}$, $\widetilde{b}_1$ and $\widetilde{b}_2$ are the OLS estimates from (5.22). Using the $\widetilde{\sigma}_i^2$'s one can run the regression of $Y_i/\widetilde{\sigma}_i$ on $(1/\widetilde{\sigma}_i)$ and $X_i/\widetilde{\sigma}_i$ without a constant to get asymptotically efficient estimates of α and β. These have the same asymptotic properties as the MLE estimators derived in Rutemiller and Bowers (1968), see Amemiya (1977) and Buse (1984). The problem with this iterative procedure is that there is no guarantee that the $\widetilde{\sigma}_i^2$'s are positive, which means that the square root $\widetilde{\sigma}_i$ may not exist. This problem would not occur if $\sigma_i^2 = (a + b_1 X_i + b_2 X_i^2)^2$ because in this case one regresses $|e_i|$ on a constant, X_i and X_i^2 and the predicted value from this regression would be an estimate of σ_i. It would not matter if this predictor is negative, because we do not have to take its square root and because its sign cancels in the OLS normal equations of the final regression of $Y_i/\widehat{\sigma}_i$ on $(1/\widehat{\sigma}_i)$ and $(X_i/\widehat{\sigma}_i)$ without a constant.

Testing for Homoskedasticity

In the repeated observation's case, one can perform Bartlett's (1937) test. The null hypothesis is $H_0; \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_m^2$. Under the null there is one variance σ^2 which can be estimated by the pooled variance $s^2 = \sum_{i=1}^m \nu_i \widetilde{s}_i^2/\nu$ where $\nu = \sum_{i=1}^m \nu_i$, and $\nu_i = n_i - 1$. Under the alternative hypothesis there are m different variances estimated by $\widetilde{s}_i^2$ for $i = 1, 2, \ldots, m$. The Likelihood Ratio test, which computes the ratio of the likelihoods under the null and alternative hypotheses, reduces to computing

$$B = [\nu \log s^2 - \textstyle\sum_{i=1}^m \nu_i \log s_i^2]/c \tag{5.24}$$

where $c = 1 + \left[\sum_{i=1}^m (1/\nu_i) - 1/\nu\right]/3(m-1)$. Under H_0, B is distributed χ^2_{m-1}. Hence, a large p-value for the B-statistic given in (5.24) means that we do not reject homoskedasticity whereas, a small p-value leads to rejection of H_0 in favor of heteroskedasticity.

In case of no repeated observations, several tests exist in the literature. Among these are the following:

(1) Glejser's (1969) Test: In this case one regresses $|e_i|$ on a constant and Z_i^δ for $\delta = 1, -1, 0.5$ and -0.5. If the coefficient of Z_i^δ is significantly different from zero, this would lead to a rejection of homoskedasticity. The power of this test depends upon the true form of heteroskedasticity. One important result however, is that this power is not seriously impaired if the wrong value of δ is chosen, see Ali and Giaccotto (1984) who confirmed this result using extensive Monte Carlo experiments.

(2) The Goldfeld and Quandt (1965) Test: This is a simple and intuitive test. One orders the observations according to X_i and omits c central observations. Next, two regressions are run on the two separated sets of observations with $(n-c)/2$ observations in each. The c omitted observations separate the low value X's from the high value X's, and if heteroskedasticity exists and is related to X_i, the estimates of σ^2 reported from the two regressions should be different. Hence, the test statistic is s_2^2/s_1^2 where s_1^2 and s_2^2 are the Mean Square Error of the two regressions, respectively. Their ratio would be the same as that of the two residual sums of squares because the degrees of freedom of the two regressions are the same. This statistic is F-distributed with $((n-c)/2)-K$ degrees of freedom in the numerator as well as the denominator. The only remaining question for performing this test is the magnitude of c. Obviously, the larger c is, the more central observations are being omitted and the more confident we feel that the two samples are distant from each other. The loss of c observations should lead to loss of power. However, separating the two samples should give us more confidence that the two variances are in fact the same if we do not reject homoskedasticity. This trade off in power was studied by Goldfeld and Quandt using Monte Carlo experiments. Their results recommend the use of $c = 8$ for $n = 30$ and $c = 16$ for $n = 60$. This is a popular test, but assumes that we know how to order the heteroskedasticity. In this case, using X_i. But what if there are more than one regressor on the right hand side? In that case one can order the observations using $\widehat{Y}_i$.

(3) Spearman's Rank Correlation Test: This test ranks the X_i's and the absolute value of the OLS residuals, the e_i's. Then it computes the difference between these rankings, i.e., $d_i = \text{rank}(|e_i|) - \text{rank}(X_i)$. The Spearman-Correlation coefficient is $r = 1 - \left[6\sum_{i=1}^n d_i^2/(n^3 - n)\right]$. Finally, test H_0; the correlation coefficient between the rankings is zero, by computing $t =$

$[r^2(n-2)/(1-r^2)]^{1/2}$ which is t-distributed with $(n-2)$ degrees of freedom. If this t-statistic has a large p-value we do not reject homoskedasticity. Otherwise, we reject homoskedasticity in favor of heteroskedasticity.

(4) Harvey's (1976) Multiplicative Heteroskedasticity Test: If heteroskedasticity is related to X_i, it looks like the Goldfeld and Quandt test or the Spearman rank correlation test would detect it, and the Glejser test would establish its form. In case the form of heteroskedasticity is of the multiplicative type, Harvey (1976) suggests the following test which rewrites (5.17) as

$$\log e_i^2 = \log\sigma^2 + \delta_1 \log Z_{1i} + \ldots + \delta_r \log Z_{ri} + v_i \tag{5.25}$$

where $v_i = \log(e_i^2/\sigma_i^2)$. This disturbance term has an asymptotic distribution that is $\log\chi_1^2$. This random variable has mean -1.2704 and variance 4.9348. Therefore, Harvey suggests performing the regression in (5.25) and testing H_0; $\delta_1 = \delta_2 = \ldots = \delta_r = 0$ by computing the regression sum of squares divided by 4.9348. This statistic is distributed asymptotically as χ_r^2. This is also asymptotically equivalent to an F-test that tests for $\delta_1 = \delta_2 = \ldots = \delta_r = 0$ in the regression given in (5.25). See the F-test described in example 6 of Chapter 4.

(5) Breusch and Pagan (1979) Test: If one knows that $\sigma_i^2 = f(a + b_1 Z_1 + b_2 Z_2 + .. + b_r Z_r)$ but does not know the form of this function f, Breusch and Pagan (1979) suggest the following test for homoskedasticity, i.e., $H_0; b_1 = b_2 = \ldots = b_r = 0$. Compute $\widehat{\sigma}^2 = \sum_{i=1}^n e_i^2/n$, which would be the MLE estimator of σ^2 under homoskedasticity. Run the regression of $e_i^2/\widehat{\sigma}^2$on the Z variables and a constant, and compute half the regression sum of squares. This statistic is distributed as χ_r^2. This is a more general test than the ones discussed earlier in that f does not have to be specified.

(6) White's (1980) Test: Another general test for homoskedasticity where nothing is known about the form of this heteroskedasticity is suggested by White (1980). This test is based on the difference between the variance of the OLS estimates under homoskedasticity and that under heteroskedasticity. For the case of a simple regression with a constant, White shows that this test compares White's var$(\widehat{\beta}_{OLS})$ given by (5.13) with the usual var$(\widehat{\beta}_{OLS}) = s^2/\sum_{i=1}^n x_i^2$ under homoskedasticity. This test reduces to running the regression of e_i^2 on a constant, X_i and X_i^2 and computing nR^2. This statistic is distributed as χ_2^2 under the null hypothesis of homoskedasticity. The degrees of freedom correspond to the number of regressors without the constant. If this statistic is not significant, then e_i^2 is not related to X_i and X_i^2 and we can not reject that the variance is constant. Note that if there is no constant in the regression, we run e_i^2 on a constant and X_i^2 only, i.e., X_i is no longer in this regression and the degree of freedom of the test is 1. In general, White's test is based on running e_i^2 on the cross-product of all the X's in the regression being estimated, computing nR^2, and comparing it to the critical value of χ_r^2 where r is the number of regressors in this last regression excluding the constant. For the case of two regressors, X_2 and X_3 and a constant, White's test is again based on nR^2 for the regression of e_i^2 on a constant, $X_2, X_3, X_2^2, X_2X_3, X_3^2$. This statistic is distributed as χ_5^2. White's test is standard using EViews. After running the regression, click on residuals tests then choose White. This software gives the user a choice between including or excluding the cross-product terms like X_2X_3 from the regression. This may be useful when there are many regressors.

A modified Breusch-Pagan test was suggested by Koenker (1981) and Koenker and Bassett (1982). This attempts to improve the power of the Breusch-Pagan test, and make it more robust to the non-normality of the disturbances. This amounts to multiplying the Breusch-Pagan statistic (half the regression sum of squares) by $2\widehat{\sigma}^4$, and dividing it by the second sample moment of the squared residuals, i.e., $\sum_{i=1}^n (e_i^2 - \widehat{\sigma}^2)^2/n$, where $\widehat{\sigma}^2 = \sum_{i=1}^n e_i^2/n$. Waldman (1983) showed that if the Z_i's in the Breusch-Pagan test are in fact the X_i's and their cross-products, as in White's test, then this modified Breusch-Pagan test is exactly the nR^2 statistic proposed by White.

White's (1980) test for heteroskedasticity without specifying its form lead to further work on estimators that are more efficient than OLS while recognizing that the efficiency of GLS may not be achievable, see Cragg (1992). Adaptive estimators have been developed by Carroll (1982) and Robinson (1987). These estimators assume no particular form of heteroskedasticity but nevertheless have the same asymptotic distribution as GLS based on the true σ_i^2.

Many Monte Carlo experiments were performed to study the performance of these and other tests of homoskedasticity. One such extensive study is that of Ali and Giaccotto (1984). Six types of heteroskedasticity specifications were considered;

(i) $\sigma_i^2 = \sigma^2$	(ii) $\sigma_i^2 = \sigma^2 \lvert X_i \rvert$	(iii) $\sigma_i^2 = \sigma^2 \lvert E(Y_i) \rvert$
(iv) $\sigma_i^2 = \sigma^2 X_i^2$	(v) $\sigma_i^2 = \sigma^2 [E(Y_i)]^2$	(vi) $\sigma_i^2 = \sigma^2$ for $i \leq n/2$ and $\sigma_i^2 = 2\sigma^2$ for $i > n/2$

Six data sets were considered, the first three were stationary and the last three were nonstationary (Stationary versus non-stationary regressors, are discussed in Chapter 14). Five models were entertained, starting with a model with one regressor and no intercept and finishing with a model with an intercept and 5 variables. Four types of distributions were imposed on the disturbances. These were normal, t, Cauchy and log normal. The first three are symmetric, but the last one is skewed. Three sample sizes were considered, $n = 10, 25, 40$. Various correlations between the disturbances were also entertained. Among the tests considered were tests 1, 2, 5 and 6 discussed in this section. The results are too numerous to summarize, but some of the major findings are the following: (1) The power of these tests increased with sample size and trendy nature or the variability of the regressors. It also decreased with more regressors and for deviations from the normal distribution. The results were mostly erratic when the errors were autocorrelated. (2) There were ten distributionally robust tests using OLS residuals named TROB which were variants of Glejser's, White's and Bickel's tests. The last one being a non-parametric test not considered in this chapter. These tests were robust to both long-tailed and skewed distributions. (3) None of these tests has any significant power to detect heteroskedasticity which deviates substantially from the true underlying heteroskedasticity. For example, none of these tests was powerful in detecting heteroskedasticity of the sixth kind, i.e., $\sigma_i^2 = \sigma^2$ for $i \leq n/2$ and $\sigma_i^2 = 2\sigma^2$ for $i > n/2$. In fact, the maximum power was 9%. (4) Ali and Giaccotto (1984) recommend any of the TROB tests for practical use. They note that the similarity among these tests is the use of squared residuals rather than the absolute value of the residuals. In fact, they argue that tests of the same form that use absolute value rather than squared residuals are likely to be non-robust and lack power.

Empirical Example: For the Cigarette Consumption Data given in Table 3.2, the OLS regression yields:

$$\underset{}{\log C} = \underset{(0.91)}{4.30} - \underset{(0.32)}{1.34} \log P + \underset{(0.20)}{0.17} \log Y \qquad \bar{R}^2 = 0.27$$

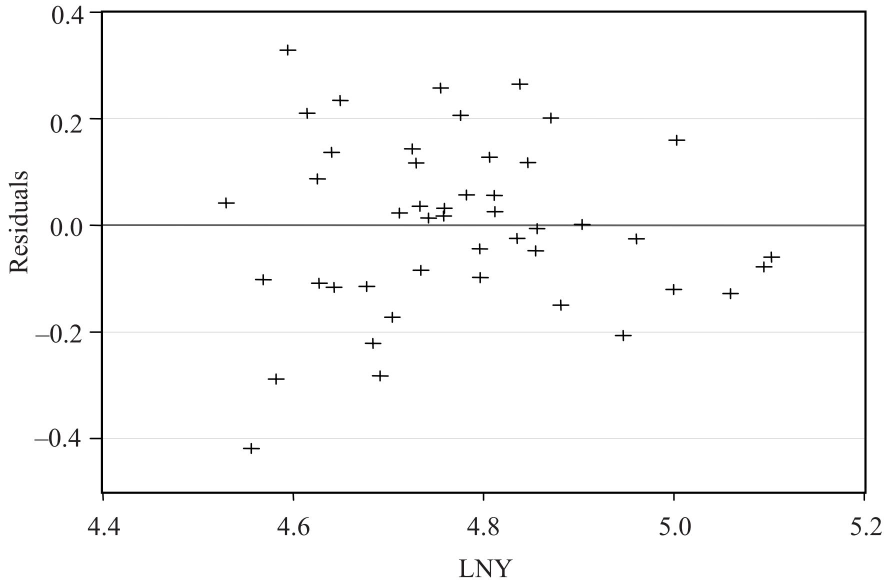

Figure 5.1 Plots of Residuals Versus Log Y

Suspecting heteroskedasticity, we plotted the residuals from this regression versus log$Y$ in Figure 5.1. This figure shows the dispersion of the residuals to decrease with increasing log$Y$. Next, we performed several tests for heteroskedasticity studied in this chapter. The first is Glejser's (1969) test. We ran the following regressions:

$$\underset{}{|e_i|} = \underset{(0.46)}{1.16} - \underset{(0.10)}{0.22} \log Y$$

$$|e_i| = \underset{(0.47)}{-0.95} + \underset{(2.23)}{5.13} (\log Y)^{-1}$$

$$|e_i| = \underset{(0.93)}{-2.00} + \underset{(2.04)}{4.65} (\log Y)^{-0.5}$$

$$|e_i| = \underset{(0.93)}{2.21} - \underset{(0.42)}{0.96} (\log Y)^{0.5}$$

The t-statistics on the slope coefficient in these regressions are -2.24, 2.30, 2.29 and -2.26, respectively. All are significant with p-values of 0.03, 0.026, 0.027 and 0.029, respectively, indicating the rejection of homoskedasticity.

The second test is the Goldfeld and Quandt (1965) test. The observations are ordered according to log$Y$ and $c = 12$ central observations are omitted. Two regressions are run on the first and last 17 observations. The first regression yields $s_1^2 = 0.04881$ and the second regression yields $s_2^2 = 0.01554$. This is a test of equality of variances and it is based on the ratio of two χ^2 random variables with $17 - 3 = 14$ degrees of freedom. In fact, $s_1^2/s_2^2 = 0.04881/0.01554 = 3.141 \sim F_{14,14}$ under H_0. This has a p-value of 0.02 and rejects H_0 at the 5% level. The third test is the Spearman rank correlation test. First one obtains the rank(logY_i) and rank($|e_i|$) and compute $d_i = \text{rank}|e_i| - \text{rank}|\log Y_i|$. From these $r = 1 - \left[6\sum_{i=1}^{46} d_i^2/(n^3 - n)\right] = -0.282$ and

$t = [r^2(n-2)/(1-r^2)]^{1/2} = 1.948$. This is distributed as a t with 44 degrees of freedom. This t-statistic has a p-value of 0.058.

The fourth test is Harvey's (1976) multiplicative heteroskedasticity test which is based upon regressing $\log e_i^2$ on $\log(\log Y_i)$

$$\log e_i^2 = \underset{(17.25)}{24.85} - \underset{(11.03)}{19.08} \log(\log Y)$$

Harvey's (1976) statistic divides the regression sum of squares which is 14.360 by 4.9348. This yields 2.91 which is asymptotically distributed as χ_1^2 under the null. This has a p-value of 0.088 and does not reject the null of homoskedasticity at the 5% significance level.

The fifth test is the Breusch and Pagan (1979) test which is based on the regression of $e_i^2/\widehat{\sigma}^2$ (where $\widehat{\sigma}^2 = \sum_{i=1}^{46} e_i^2/46 = 0.024968$) on $\log Y_i$. The test-statistic is half the regression sum of squares $= (10.971 \div 2) = 5.485$. This is distributed as χ_1^2 under the null hypothesis. This has a p-value of 0.019 and rejects the null of homoskedasticity. This can be generated with Stata after running the OLS regression reported above, and then issuing the command *estat hettest lnrdi*

```
.estat hettest lnrdi

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: lnrdi

         chi2(1)      =   5.49
         Prob > chi2  = 0.0192
```

Finally, White's (1980) test for heteroskedasticity is performed which is based on the regression of e_i^2 on $\log P$, $\log Y$, $(\log P)^2$, $(\log Y)^2$, $(\log P)(\log Y)$ and a constant. This is shown in Table 5.1 using EViews. The test-statistic is $nR^2 = (46)(0.3404) = 15.66$ which is distributed as χ_5^2. This has a p-value of 0.008 and rejects the null of homoskedasticity. Except for Harvey's test, all the tests performed indicate the presence of heteroskedasticity. This is true despite the fact that the data are in logs, and both consumption and income are expressed in per capita terms. White's heteroskedasticity-consistent estimates of the variances are as follows:

$$\log C = \underset{(1.10)}{4.30} - \underset{(0.34)}{1.34} \log P + \underset{(0.24)}{0.17} \log Y$$

These are given in Table 5.2 using EViews. Note that in this case all of the heteroskedasticity-consistent standard errors are larger than those reported using a standard OLS package, but this is not necessarily true for other data sets.

In section 5.4, we described the Jarque and Bera (1987) test for normality of the disturbances. For this cigarette consumption regression, Figure 5.2 gives the histogram of the residuals along with descriptive statistics of these residuals including their mean, median, skewness and kurtosis.

This is done using EViews. The measure of skewness S is estimated to be -0.184 and the measure of kurtosis κ is estimated to be 2.875 yielding a Jarque-Bera statistic of

$$JB = 46\left[\frac{(-0.184)^2}{6} + \frac{(2.875-3)^2}{24}\right] = 0.29.$$

Table 5.1 White Heteroskedasticity Test

F-statistic	4.127779	Probability		0.004073
Obs*R-squared	15.65644	Probability		0.007897
Test Equation:				
Dependent Variable:	RESIDˆ2			
Method:	Least Squares			
Sample:	1 46			
Included observations:	46			
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	18.22199	5.374060	3.390730	0.0016
LNP	9.506059	3.302570	2.878382	0.0064
LNPˆ2	1.281141	0.656208	1.952340	0.0579
LNP*LNY	–2.078635	0.727523	–2.857139	0.0068
LNY	–7.893179	2.329386	–3.388523	0.0016
LNYˆ2	0.855726	0.253048	3.381670	0.0016
R-squared	0.340357	Mean dependent var		0.024968
Adjusted R-squared	0.257902	S.D. dependent var		0.034567
S.E. of regression	0.029778	Akaike info criterion		–4.068982
Sum squared resid	0.035469	Schwarz criterion		–3.830464
Log likelihood	99.58660	F-statistic		4.127779
Durbin-Watson stat	1.853360	Prob (F-statistic)		0.004073

Table 5.2 White Heteroskedasticity-Consistent Standard Errors

Dependent Variable:	LNC			
Method:	Least Squares			
Sample:	1 46			
Included observations:	46			
White Heteroskedasticity-Consistent Standard Errors & Covariance				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.299662	1.095226	3.925821	0.0003
LNP	–1.338335	0.343368	–3.897671	0.0003
LNY	0.172386	0.236610	0.728565	0.4702
R-squared	0.303714	Mean dependent var		4.847844
Adjusted R-squared	0.271328	S.D. dependent var		0.191458
S.E. of regression	0.163433	Akaike info criterion		–0.721834
Sum squared resid	1.148545	Schwarz criterion		–0.602575
Log likelihood	19.60218	F-statistic		9.378101
Durbin-Watson stat	2.315716	Prob (F-statistic)		0.000417

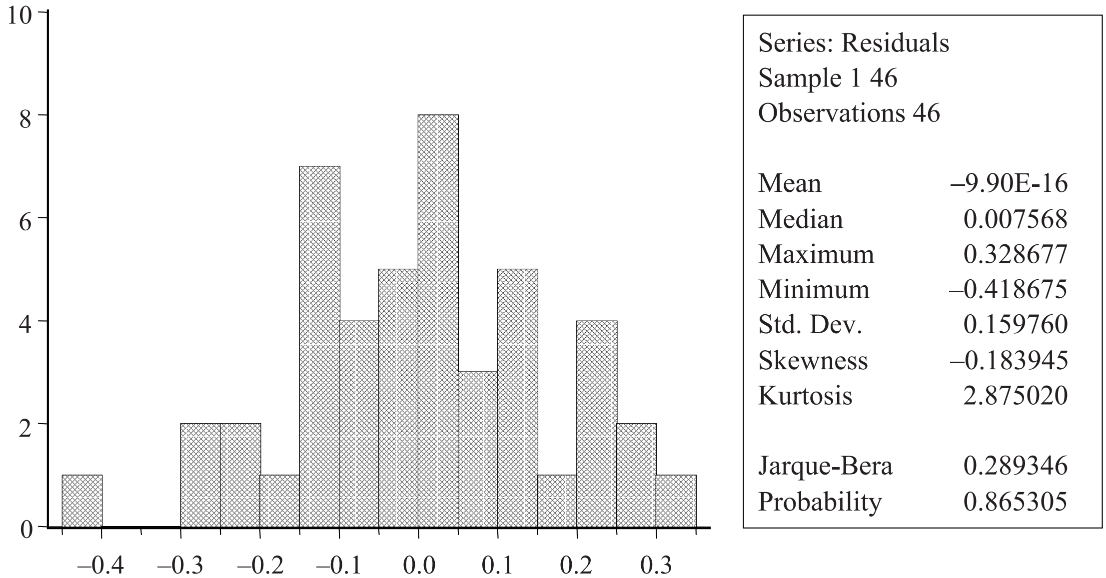

Figure 5.2 Normality Test (Jarque-Bera)

This is distributed as χ^2_2 under the null hypothesis of normality and has a p-value of 0.865. Hence we do not reject that the distribution of the disturbances is symmetric and has a kurtosis of 3.

5.6 Autocorrelation

Violation of assumption 3 means that the disturbances are correlated, i.e., $E(u_i u_j) = \sigma_{ij} \neq 0$, for $i \neq j$, and $i, j = 1, 2, \ldots, n$. Since u_i has zero mean, $E(u_i u_j) = \text{cov}(u_i, u_j)$ and this is denoted by σ_{ij}. This correlation is more likely to occur in time-series than in cross-section studies. Consider estimating the consumption function of a random sample of households. An unexpected event, like a visit of family members will increase the consumption of this household. However, this positive disturbance need not be correlated to the disturbances affecting consumption of other randomly drawn households. However, if we were estimating this consumption function using aggregate time-series data for the U.S., then it is very likely that a recession year affecting consumption negatively this year may have a carry over effect to the next few years. A shock to the economy like an oil embargo in 1973 is likely to affect the economy for several years. A labor strike this year may affect production for the next few years. Therefore, we will switch the i and j subscripts to t and s denoting time-series observations $t, s = 1, 2, \ldots, T$ and the sample size will be denoted by T rather than n. This covariance term is symmetric, so that $\sigma_{12} = E(u_1 u_2) = E(u_2 u_1) = \sigma_{21}$. Hence, only $T(T-1)/2$ distinct σ_{ts}'s have to be estimated. For example, if $T = 3$, then σ_{12}, σ_{13} and σ_{23} are the distinct covariance terms. However, it is hopeless to estimate $T(T-1)/2$ covariances (σ_{ts}) with only T observations. Therefore, more structure on these σ_{ts}'s need to be imposed. A popular assumption is that the u_t's follow a first-order autoregressive process denoted by AR(1):

$$u_t = \rho u_{t-1} + \epsilon_t \qquad t = 1, 2, \ldots, T \tag{5.26}$$

where ϵ_t is IID$(0, \sigma_\epsilon^2)$. It is autoregressive because u_t is related to its lagged value u_{t-1}. One can also write (5.26), for period $t-1$, as

$$u_{t-1} = \rho u_{t-2} + \epsilon_{t-1} \tag{5.27}$$

and substitute (5.27) in (5.26) to get

$$u_t = \rho^2 u_{t-2} + \rho\epsilon_{t-1} + \epsilon_t \tag{5.28}$$

Note that the power of ρ and the subscript of u or ϵ always sum to t. By continuous substitution of this form, one ultimately gets

$$u_t = \rho^t u_0 + \rho^{t-1}\epsilon_1 + .. + \rho\epsilon_{t-1} + \epsilon_t \tag{5.29}$$

This means that u_t is a function of current and past values of ϵ_t and u_0 where u_0 is the initial value of u_t. If u_0 has zero mean, then u_t has zero mean. This follows from (5.29) by taking expectations. Also, from (5.26)

$$\text{var}(u_t) = \rho^2 \text{var}(u_{t-1}) + \text{var}(\epsilon_t) + 2\rho \text{cov}(u_{t-1}, \epsilon_t) \tag{5.30}$$

Using (5.29), u_{t-1} is a function of ϵ_{t-1}, past values of ϵ_{t-1} and u_0. Since u_0 is independent of the ϵ's, and the ϵ's are themselves not serially correlated, then u_{t-1} is independent of ϵ_t. This means that $\text{cov}(u_{t-1}, \epsilon_t) = 0$. Furthermore, for u_t to be homoskedastic, $\text{var}(u_t) = \text{var}(u_{t-1}) = \sigma_u^2$, and (5.30) reduces to $\sigma_u^2 = \rho^2\sigma_u^2 + \sigma_\epsilon^2$, which when solved for σ_u^2 gives:

$$\sigma_u^2 = \sigma_\epsilon^2/(1-\rho^2) \tag{5.31}$$

Hence, $u_0 \sim (0, \sigma_\epsilon^2/(1-\rho^2))$ for the u's to have zero mean and homoskedastic disturbances. Multiplying (5.26) by u_{t-1} and taking expected values, one gets

$$E(u_t u_{t-1}) = \rho E(u_{t-1}^2) + E(u_{t-1}\epsilon_t) = \rho\sigma_u^2 \tag{5.32}$$

since $E(u_{t-1}^2) = \sigma_u^2$ and $E(u_{t-1}\epsilon_t) = 0$. Therefore, $\text{cov}(u_t, u_{t-1}) = \rho\sigma_u^2$, and the correlation coefficient between u_t and u_{t-1} is $\text{correl}(u_t, u_{t-1}) = \text{cov}(u_t, u_{t-1})/\sqrt{\text{var}(u_t)\text{var}(u_{t-1})} = \rho\sigma_u^2/\sigma_u^2 = \rho$. Since ρ is a correlation coefficient, this means that $-1 \le \rho \le 1$. In general, one can show that

$$\text{cov}(u_t, u_s) = \rho^{|t-s|}\sigma_u^2 \qquad t, s = 1, 2, \ldots, T \tag{5.33}$$

see problem 6. This means that the correlation between u_t and u_{t-r} is ρ^r, which is a fraction raised to an integer power, i.e., the correlation is decaying between the disturbances the further apart they are. This is reasonable in economics and may be the reason why this autoregressive form (5.26) is so popular. One should note that this is not the only form that would correlate the disturbances across time. In Chapter 14, we will consider other forms like the Moving Average (MA) process, and higher order Autoregressive Moving Average (ARMA) processes, but these are beyond the scope of this chapter.

Consequences for OLS

How is the OLS estimator affected by the violation of the no autocorrelation assumption among the disturbances? The OLS estimator is still unbiased and consistent since these properties rely on assumptions 1 and 4 and have nothing to do with assumption 3. For the simple linear regression, using (5.2), the variance of $\widehat{\beta}_{OLS}$ is now

$$\begin{aligned} \text{var}(\widehat{\beta}_{OLS}) &= \text{var}(\textstyle\sum_{t=1}^T w_t u_t) = \sum_{t=1}^T \sum_{s=1}^T w_t w_s \text{cov}(u_t, u_s) \\ &= \sigma_u^2 / \textstyle\sum_{t=1}^T x_t^2 + \sum\limits_{t \neq s}\sum w_t w_s \rho^{|t-s|} \sigma_u^2 \end{aligned} \quad (5.34)$$

where $\text{cov}(u_t, u_s) = \rho^{|t-s|}\sigma_u^2$ as explained in (5.33). Note that the first term in (5.34) is the usual variance of $\widehat{\beta}_{OLS}$ under the classical case. The second term in (5.34) arises because of the correlation between the u_t's. Hence, the variance of OLS computed from a regression package, i.e., $s^2/\sum_{t=1}^T x_t^2$ is a wrong estimate of the variance of $\widehat{\beta}_{OLS}$ for two reasons. First, it is using the wrong formula for the variance, i.e., $\sigma_u^2/\sum_{t=1}^T x_t^2$ rather than (5.34). The latter depends on ρ through the extra term in (5.34). Second, one can show, see problem 7, that $E(s^2) \neq \sigma_u^2$ and will involve ρ as well as σ_u^2. Hence, s^2 is not unbiased for σ_u^2 and $s^2/\sum_{t=1}^T x_t^2$ is a biased estimate of $\text{var}(\widehat{\beta}_{OLS})$. The direction and magnitude of this bias depends on ρ and the regressor. In fact, if ρ is positive, and the x_t's are themselves positively autocorrelated, then $s^2/\sum_{t=1}^T x_t^2$ understates the true variance of $\widehat{\beta}_{OLS}$. This means that the confidence interval for β is tighter than it should be and the t-statistic for H_0; $\beta = 0$ is overblown, see problem 8. As in the heteroskedastic case, but for completely different reasons, any inference based on $\text{var}(\widehat{\beta}_{OLS})$ reported from the standard regression packages will be misleading if the u_t's are serially correlated.

Newey and West (1987) suggested a simple heteroskedasticity and autocorrelation-consistent covariance matrix for the OLS estimator without specifying the functional form of the serial correlation. The basic idea extends White's (1980) replacement of heteroskedastic variances with squared OLS residuals e_t^2 by additionally including products of least squares residuals $e_t e_{t-s}$ for $s = 0, \pm 1, \ldots, \pm p$ where p is the maximum order of serial correlation we are willing to assume. The consistency of this procedure relies on p being very small relative to the number of observations T. This is consistent with popular serial correlation specifications considered in this chapter where the autocorrelation dies out quickly as j increases. Newey and West (1987) allow the higher order covariance terms to receive diminishing weights. This Newey-West option for the least squares estimator is available using EViews. Andrews (1991) warns about the unreliability of such standard error corrections in some circumstances. Wooldridge (1991) shows that it is possible to construct serially correlated robust F-statistics for testing joint hypotheses as considered in Chapter 4. However, these are beyond the scope of this book.

Is OLS still BLUE? In order to determine the BLU estimator in this case, we lag the regression equation once, multiply it by ρ, and subtract it from the original regression equation, we get

$$Y_t - \rho Y_{t-1} = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + \epsilon_t \quad t = 2, 3, \ldots, T \quad (5.35)$$

This transformation, known as the Cochrane-Orcutt (1949) transformation, reduces the disturbances to classical errors. Therefore, OLS on the resulting regression renders the estimates BLU, i.e., run $\widetilde{Y}_t = Y_t - \rho Y_{t-1}$ on a constant and $\widetilde{X}_t = X_t - \rho X_{t-1}$, for $t = 2, 3, \ldots, T$. Note that we have lost one observation by lagging, and the resulting estimators are BLUE only for linear combinations of $(T-1)$ observations in Y.[1] Prais and Winsten (1954) derive the BLU estimators for linear combinations of T observations in Y. This entails recapturing the initial observation as follows: (i) Multiply the first observation of the regression equation by $\sqrt{1-\rho^2}$;

$$\sqrt{1-\rho^2}Y_1 = \alpha\sqrt{1-\rho^2} + \beta\sqrt{1-\rho^2}X_1 + \sqrt{1-\rho^2}u_1$$

(ii) add this transformed initial observation to the Cochrane-Orcutt transformed observations for $t = 2, \ldots, T$ and run the regression on the T observations rather than the $(T-1)$ observations. See Chapter 9, for a formal proof of this result. Note that

$$\widetilde{Y}_1 = \sqrt{1-\rho^2}Y_1$$

and

$$\widetilde{Y}_t = Y_t - \rho Y_{t-1} \quad \text{for } t = 2, \ldots, T$$

Similarly, $\widetilde{X}_1 = \sqrt{1-\rho^2}X_1$ and $\widetilde{X}_t = X_t - \rho X_{t-1}$ for $t = 2, \ldots, T$. The constant variable $C_t = 1$ for $t = 1, \ldots, T$ is now a new variable $\widetilde{C}_t$ which takes the values $\widetilde{C}_1 = \sqrt{1-\rho^2}$ and $\widetilde{C}_t = (1-\rho)$ for $t = 2, \ldots, T$. Hence, the Prais-Winsten procedure is the regression of $\widetilde{Y}_t$ on $\widetilde{C}_t$ and $\widetilde{X}_t$ without a constant. It is obvious that the resulting BLU estimators will involve ρ and are therefore, different from the usual OLS estimators except in the case where $\rho = 0$. Hence, OLS is no longer BLUE. Furthermore, we need to know ρ in order to obtain the BLU estimators. In applied work, ρ is not known and has to be estimated, in which case the Prais-Winsten regression is no longer BLUE since it is based on an estimate of ρ rather than the true ρ itself. However, as long as $\widehat{\rho}$ is a consistent estimate for ρ then this is a sufficient condition for the corresponding estimates of α and β in the next step to be asymptotically efficient, see Chapter 9. We now turn to various methods of estimating ρ.

(1) The Cochrane-Orcutt (1949) Method: This method starts with an initial estimate of ρ, the most convenient is 0, and minimizes the residual sum of squares in (5.35). This gives us the OLS estimates of α and β. Then we substitute $\widehat{\alpha}_{OLS}$ and $\widehat{\beta}_{OLS}$ in (5.35) and we get

$$e_t = \rho e_{t-1} + \epsilon_t \quad t = 2, \ldots, T \tag{5.36}$$

where e_t denotes the OLS residual. An estimate of ρ can be obtained by minimizing the residual sum of squares in (5.36) or running the regression of e_t on e_{t-1} without a constant. The resulting estimate of ρ is $\widehat{\rho}_{co} = \sum_{t=2}^{T} e_t e_{t-1} / \sum_{t=2}^{T} e_{t-1}^2$ where both summations run over $t = 2, 3, \ldots, T$. The second step of the Cochrane-Orcutt procedure (2SCO) is to perform the regression in (5.35) with $\widehat{\rho}_{co}$ instead of ρ. One can iterate this procedure (ITCO) by computing new residuals based on the new estimates of α and β and hence a new estimate of ρ from (5.36), and so on, until convergence. Both the 2SCO and the ITCO are asymptotically efficient, the argument for iterating must be justified in terms of small sample gains.

(2) The Hilderth-Lu (1960) Search Procedure: ρ is between -1 and 1. Therefore, this procedure searches over this range, i.e., using values of ρ say between -0.9 and 0.9 in intervals of 0.1. For each ρ, one computes the regression in (5.35) and reports the residual sum of squares corresponding to that ρ. The minimum residual sum of squares gives us our choice of ρ and the corresponding regression gives us the estimates of α, β and σ^2. One can refine this procedure around the best ρ found in the first stage of the search. For example, suppose that $\rho = 0.6$ gave the minimum residual sum of squares, one can search next between 0.51 and 0.69 in intervals of 0.01. This search procedure guards against a local minimum. Since the likelihood in this case contains ρ as well as σ^2 and α and β, this search procedure can be modified to maximize the likelihood rather than minimize the residual sum of squares, since the two criteria are no longer equivalent. The maximum value of the likelihood will give our choice of ρ and the corresponding estimates of α, β and σ^2.

(3) Durbin's (1960) Method: One can rearrange (5.35) by moving Y_{t-1} to the right hand side, i.e.,

$$Y_t = \rho Y_{t-1} + \alpha(1-\rho) + \beta X_t - \rho\beta X_{t-1} + \epsilon_t \tag{5.37}$$

and running OLS on (5.37). The error in (5.37) is classical, and the presence of Y_{t-1} on the right hand side reminds us of the contemporaneously uncorrelated case discussed under the violation of assumption 4. For that violation, we have shown that unbiasedness is lost, but not consistency. Hence, the estimate of ρ as a coefficient of Y_{t-1} is biased but consistent. This is the Durbin estimate of ρ, call it $\widehat{\rho}_D$. Next, the second step of the Cochrane-Orcutt procedure is performed using this estimate of ρ.

(4) Beach-MacKinnon (1978) Maximum Likelihood Procedure: Beach and MacKinnon (1978) derived a cubic equation in ρ which maximizes the likelihood function concentrated with respect to α, β, and σ^2. With this estimate of ρ, denoted by $\widehat{\rho}_{BM}$, one performs the Prais-Winsten procedure in the next step.

Correcting for serial correlation is not without its critics. Mizon (1995) argues this point forcefully in his article entitled "A simple message for autocorrelation correctors: Don't." The main point being that serial correlation is a symptom of dynamic misspecification which is better represented using a general unrestricted dynamic specification.

Monte Carlo Results

Rao and Griliches (1969) performed a Monte Carlo study using an autoregressive X_t, and various values of ρ. They found that OLS is still a viable estimator as long as $|\rho| < 0.3$, but if $|\rho| > 0.3$, then it pays to perform procedures that correct for serial correlation based on an estimator of ρ. Their recommendation was to compute a Durbin's estimate of ρ in the first step and to do the Prais-Winsten procedure in the second step. Maeshiro (1976, 1979) found that if the X_t series is trended, which is usual with economic data, then OLS outperforms 2SCO, but not the two-step Prais-Winsten (2SPW) procedure that recaptures the initial observation. These results were confirmed by Park and Mitchell (1980) who performed an extensive Monte Carlo using trended and untrended X_t's. Their basic findings include the following: (i) For trended X_t's, OLS beats 2SCO, ITCO and even a Cochrane-Orcutt procedure that is based on the true ρ. However, OLS was beaten by 2SPW, iterative Prais-Winsten (ITPW), and Beach-MacKinnon (BM). Their conclusion is that one should not use regressions based on $(T-1)$ observations as in Cochrane and Orcutt. (ii) Their results find that the ITPW procedure is the recommended estimator beating 2SPW and BM for high values of true ρ, for both trended as well as nontrended X_t's. (iii) Test of hypotheses regarding the regression coefficients performed miserably for all estimators based on an estimator of ρ. The results indicated less bias in standard error estimation for ITPW, BM and 2SPW than OLS. However, the tests based on these standard errors still led to a high probability of type I error for all estimation procedures.

Testing for Autocorrelation

So far, we have studied the properties of OLS under the violation of assumption 3. We have derived asymptotically efficient estimators of the coefficients based on consistent estimators of ρ and studied their small sample properties using Monte Carlo experiments. Next, we focus on the problem of detecting this autocorrelation between the disturbances. A popular diagnostic for detecting such autocorrelation is the Durbin and Watson (1951) statistic[2]

$$d = \sum_{t=2}^{T}(e_t - e_{t-1})^2 / \sum_{t=1}^{T} e_t^2 \tag{5.38}$$

If this was based on the true u_t's and T was very large then d can be shown to tend in the limit as T gets large to $2(1-\rho)$, see problem 9. This means that if $\rho \rightarrow 0$, then $d \rightarrow 2$; if $\rho \rightarrow 1$, then $d \rightarrow 0$ and if $\rho \rightarrow -1$, then $d \rightarrow 4$. Therefore, a test for H_0; $\rho = 0$, can be based on whether d is close to 2 or not. Unfortunately, the critical values of d depend upon the X_t's, and these vary from one data set to another. To get around this, Durbin and Watson established upper (d_U) and lower (d_L) bounds for this critical value. If the observed d is less than d_L, or larger than $4 - d_L$, we reject H_0. If the observed d is between d_U and $4 - d_U$, then we do not reject H_0. If d lies in any of the two indeterminant regions, then one should compute the exact critical values which depend on the data. Most regression packages report the Durbin-Watson statistic, but few give the exact p-value for this d-statistic. If one is interested in a single sided test, say H_0; $\rho = 0$ versus $H_1; \rho > 0$ then one would reject H_0 if $d < d_L$, and not reject H_0 if $d > d_U$. If $d_L < d < d_U$, then the test is inconclusive. Similarly for testing $H_0; \rho = 0$ versus $H_1; \rho < 0$, one computes $(4-d)$ and follow the steps for testing against positive autocorrelation. Durbin and Watson tables for d_L and d_U covered samples sizes from 15 to 100 and a maximum of 5 regressors. Savin and White (1977) extended these tables for $6 \leq T \leq 200$ and up to 10 regressors.

The Durbin-Watson statistic has several limitations. We discussed the inconclusive region and the computation of exact critical values. The Durbin-Watson statistic is appropriate when there is a constant in the regression. In case there is no constant in the regression, see Farebrother (1980). Also, the Durbin-Watson statistic is inappropriate when there are lagged values of the dependent variable among the regressors. We now turn to an alternative test for serial correlation that does not have these limitations and that is also easy to apply. This test was derived by Breusch (1978) and Godfrey (1978) and is known as the Breusch-Godfrey test for zero first-order serial correlation. This is a Lagrange Multiplier test that amounts to running the regression of the OLS residuals e_t on e_{t-1} and the original regressors in the model. The test statistic is TR^2. Its distribution under the null is χ_1^2. In this case, the regressors are a constant and X_t, and the test checks whether the coefficient of e_{t-1} is significant. The beauty of this test is that (i) it is the same test for first-order serial correlation, whether the disturbances are Moving Average of order one MA(1) or AR(1). (ii) This test is easily generalizable to higher autoregressive or Moving Average schemes. For second-order serial correlation, like MA(2) or AR(2) one includes two lags of the residuals on the right hand side; i.e., both e_{t-1} and e_{t-2}. (iii) This test is still valid even when lagged values of the dependent variable are present among the regressors, see Chapter 6. The Breusch and Godfrey test is standard using EViews and it prompts the user with a choice of the number of lags of the residuals to include among the regressors to test for serial correlation. You click on residuals, then tests and choose Breusch-Godfrey. Next, you input the number of lagged residuals you want to include.

What about first differencing the data as a possible solution for getting rid of serial correlation? Some economic behavioral equations are specified with variables in first difference form, like GDP growth, but other equations are first differenced for estimation purposes. In the latter case, if the original disturbances were not autocorrelated, (or even correlated, with $\rho \neq 1$), then the transformed disturbances are serially correlated. After all, first differencing the disturbances is equivalent to setting $\rho = 1$ in $u_t - \rho u_{t-1}$, and this new disturbance $u_t^* = u_t - u_{t-1}$ has u_{t-1} in common with $u_{t-1}^* = u_{t-1} - u_{t-2}$, making $E(u_t^* u_{t-1}^*) = -E(u_{t-1}^2) = -\sigma_u^2$. However, one could argue that if ρ is large and positive, first differencing the data may not be a bad solution. Rao and Miller (1971) calculated the variance of the BLU estimator correcting for serial correlation, for various guesses of ρ. They assume a true ρ of 0.2, and an autoregressive X_t

$$X_t = \lambda X_{t-1} + w_t \quad \text{with } \lambda = 0, 0.4, 0.8. \tag{5.39}$$

They find that OLS (or a guess of $\rho = 0$), performs better than first differencing the data, and is pretty close in terms of efficiency to the true BLU estimator for trended X_t ($\lambda = 0.8$). However, the performance of OLS deteriorates as λ declines to 0.4 and 0, with respect to the true BLU estimator. This supports the Monte Carlo finding by Rao and Griliches that for $|\rho| < 0.3$, OLS performs reasonably well relative to estimators that correct for serial correlation. However, the first-difference estimator, i.e., a guess of $\rho = 1$, performs badly for trended X_t ($\lambda = 0.8$) giving the worst efficiency when compared to any other guess of ρ. Only when the X_t's are less trended ($\lambda = 0.4$) or random ($\lambda = 0$), does the efficiency of the first-difference estimator improve. However, even for those cases one can do better by guessing ρ. For example, for $\lambda = 0$, one can always do better than first differencing by guessing any positive ρ less than 1. Similarly, for true $\rho = 0.6$, a higher degree of serial correlation, Rao and Miller (1971) show that the performance of OLS deteriorates, while that of the first difference improves. However, one can still do better than first differencing by guessing in the interval $(0.4, 0.9)$. This gain in efficiency increases with trended X_t's.

Empirical Example: Table 5.3 gives the U.S. Real Personal Consumption Expenditures (C) and Real Disposable Personal Income (Y) from the Economic Report of the President over the period 1959–2007. This data set is available as CONSUMP.DAT on the Springer web site.

The OLS regression yields:

$$\underset{}{C_t} = \underset{(219.56)}{-1343.31} + \underset{(0.011)}{0.979}\; Y_t + \text{residuals}$$

Figure 5.3 plots the actual, fitted and residuals using EViews 6.0. This shows positive serial correlation with a string of positive residuals followed by a string of negative residuals followed by positive residuals. The Durbin-Watson statistic is $d = 0.181$ which is much smaller than the lower bound $d = 1.497$ for $T = 49$ and one regressor. Therefore, we reject the null hypothesis of $H_0; \rho = 0$ at the 5% significance level.

The Breusch (1978) and Godfrey (1978) regression that tests for first-order serial correlation is given in Table 5.4. This is done using EViews 6.0.

This yields

$$e_t = \underset{(102.77)}{-54.41} + \underset{(0.005)}{0.004}\; Y_t + \underset{(0.070)}{0.909}\; e_{t-1} + \text{residuals}$$

The test statistic is TR^2 which yields $49 \times (0.786) = 38.5$. This is distributed as χ_1^2 under H_0; $\rho = 0$. This rejects the null hypothesis of no first order serial correlation with a p-value of 0.0000 shown in Table 5.4.

Table 5.3 U.S. Consumption Data, 1959–2007

C = Real Personal Consumption Expenditures (in 1987 dollars)
Y = Real Disposable Personal Income (in 1987 dollars)

YEAR	Y	C
1959	8776	9685
1960	8837	9735
1961	8873	9901
1962	9170	10227
1963	9412	10455
1964	9839	11061
1965	10331	11594
1966	10793	12065
1967	10994	12457
1968	11510	12892
1969	11820	13163
1970	11955	13563
1971	12256	14001
1972	12868	14512
1973	13371	15345
1974	13148	15094
1975	13320	15291
1976	13919	15738
1977	14364	16128
1978	14837	16704
1979	15030	16931
1980	14816	16940
1981	14879	17217
1982	14944	17418
1983	15656	17828

YEAR	Y	C
1984	16343	19011
1985	17040	19476
1986	17570	19906
1987	17994	20072
1988	18554	20740
1989	18898	21120
1990	19067	21281
1991	18848	21109
1992	19208	21548
1993	19593	21493
1994	20082	21812
1995	20382	22153
1996	20835	22546
1997	21365	23065
1998	22183	24131
1999	23050	24564
2000	23862	25472
2001	24215	25697
2002	24632	26238
2003	25073	26566
2004	25750	27274
2005	26290	27403
2006	26835	28098
2007	27319	28614

Source: Economic Report of the President

Regressing the OLS residuals on their lagged values yields

$$\begin{aligned} e_t = \ & \underset{(0.062)}{0.906}\ e_{t-1} + \text{residuals} \end{aligned}$$

The two-step Cochrane-Orcutt (1949) procedure based on $\widehat{\rho} = 0.906$ using Stata 11 yields the results given in Table 5.5.

The Prais-Winsten (1954) procedure using Stata 11 yields the results given in Table 5.6. The estimate of the marginal propensity to consume is 0.979 for OLS, 0.989 for two-step Cochrane-Orcutt, and 0.912 for iterative Prais-Winsten. All of these estimates are significant.

The Newey-West heteroskedasticity and autocorrelation-consistent standard errors for least squares with a three-year lag truncation are given in Table 5.7 using EViews 6. Note that both standard errors are now larger than those reported by least squares. But once again, this is not necessarily the case for other data sets.

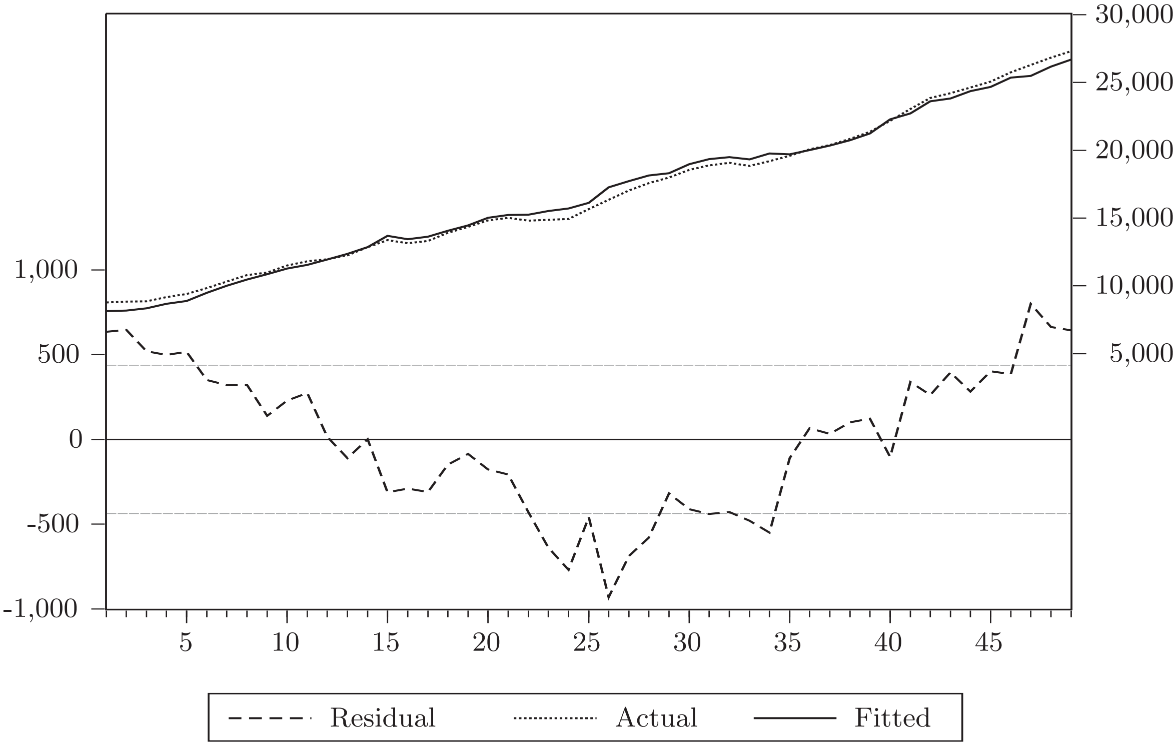

Figure 5.3 Residual Plot: Consumption Regression

Table 5.4 Breusch-Godfrey LM Test

F-statistic	168.9023	Prob. F(1,46)		0.0000
Obs*R-squared	38.51151	Prob. Chi-Square(1)		0.0000
Test Equation:				
Dependent Variable:	RESID			
Method:	Least Squares			
Sample	1959 2007			
Included observations:	49			
Presample missing value lagged residuals set to zero				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	–54.41017	102.7650	–0.529462	0.5990
Y	0.003590	0.005335	0.673044	0.5043
RESID(–1)	0.909272	0.069964	12.99624	0.0000
R-squared	0.785949	Mean dependent var		–5.34E-13
Adjusted R-squared	0.776643	S.D. dependent var		433.0451
S.E. of regression	204.6601	Akaike info criterion		13.53985
Sum squared resid	1926746.	Schwarz criterion		13.65567
Log likelihood	–328.7263	Hannan-Quinn criter.		13.58379
F-statistic	84.45113	Durbin-Watson stat		2.116362
Prob(F-statistic)	0.000000			

Table 5.5 Cochrane-Orcutt AR(1) Regression – Twostep

. prais c y, corc two

Iteration 0: rho = 0.0000
Iteration 1: rho = 0.9059

Cochrane-Orcutt AR(1) regression – twostep estimates

Source	SS	df	MS			
				Number of obs	=	48
				F(1, 46)	=	519.58
Model	17473195	1	17473195	Prob > F	=	0.0000
Residual	1546950.74	46	33629.364	R-squared	=	0.9187
				Adj R-squared	=	0.9169
Total	19020145.7	47	404683.951	Root MSE	=	183.38

c	Coef.	Std. Err.	t	$P > \|t\|$	[95% Conf. Interval]	
y	.9892295	.0433981	22.79	0.000	.9018738	1.076585
_cons	–1579.722	1014.436	–1.56	0.126	–3621.676	462.2328
rho	.9059431					

Durbin-Watson statistic (original) 0.180503
Durbin-Watson statistic (transformed) 2.457550

Table 5.6 The Iterative Prais-Winsten AR(1) Regression

. prais c y

Prais-Winsten AR(1) regression – iterated estimates

Source	SS	df	MS			
				Number of obs	=	49
				F(1, 47)	=	119.89
Model	3916565.48	1	3916565.48	Prob > F	=	0.0000
Residual	1535401.45	47	32668.1159	R-squared	=	0.7184
				Adj R-squared	=	0.7124
Total	5451966.93	48	113582.644	Root MSE	=	180.74

c	Coef.	Std. Err.	t	P > |t|	[95% Conf. Interval]	
y	.912147	.047007	19.40	0.000	.8175811	1.006713
_cons	358.9638	1174.865	0.31	0.761	–2004.56	2722.488
rho	.9808528					

Durbin-Watson statistic (original) 0.180503
Durbin-Watson statistic (transformed) 2.314703

Table 5.7 The Newey-West HAC Standard Errors

Dependent Variable:	CONSUM			
Method:	Least Squares			
Sample:	1959 2007			
Included observations:	49			
Newey-West HAC Standard Errors & Covariance (lag truncation=3)				

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	–1343.314	422.2947	–3.180987	0.0026
Y	0.979228	0.022434	43.64969	0.0000

R-squared	0.993680	Mean dependent var	16749.10
Adjusted R-squared	0.993545	S.D. dependent var	5447.060
S.E. of regression	437.6277	Akaike info criterion	15.04057
Sum squared resid	9001348.	Schwarz criterion	15.11779
Log likelihood	–366.4941	Hannan-Quinn criter.	15.06987
F-statistic	7389.281	Durbin-Watson stat	0.180503
Prob(F-statistic)	0.000000		

Notes

1. A computational warning is in order when one is applying the Cochrane-Orcutt transformation to cross-section data. Time-series data has a natural ordering which is generally lacking in cross-section data. Therefore, one should be careful in applying the Cochrane-Orcutt transformation to cross-section data since it is not invariant to the ordering of the observations.

2. Another test for serial correlation can be obtained as a by-product of maximum likelihood estimation. The maximum likelihood estimator of ρ has a normal limiting distribution with mean ρ and variance $(1-\rho^2)/T$. Hence, one can compute $\widehat{\rho}_{MLE}/[(1-\widehat{\rho}^2_{MLE})/T]^{1/2}$and compare it to critical values from the normal distribution.

Problems

1. *s^2 Is Biased Under Heteroskedasticity.* For the simple linear regression with heteroskedasticity, i.e., $E(u_i^2)=\sigma_i^2$, show that $E(s^2)$ is a function of the σ_i^2's?

2. *OLS Variance Is Biased Under Heteroskedasticity.* For the simple linear regression with heteroskedasticity of the form $E(u_i^2)=\sigma_i^2=bx_i^2$ where $b>0$, show that $E(s^2/\sum_{i=1}^n x_i^2)$ understates the variance of $\widehat{\beta}_{OLS}$ which is

   $\sum_{i=1}^n x_i^2\sigma_i^2/(\sum_{i=1}^n x_i^2)^2$.

3. *Weighted Least Squares.* This is based on Kmenta (1986).

    (a) Solve the two equations in (5.11) and show that the solution is given by (5.12).

    (b) Show that

$$\begin{aligned} \mathrm{var}(\widetilde{\beta}) &= \frac{\sum_{i=1}^{n}(1/\sigma_i^2)}{[\sum_{i=1}^{n} X_i^2/\sigma_i^2][\sum_{i=1}^{n}(1/\sigma_i^2)] - [\sum_{i=1}^{n}(X_i/\sigma_i^2)]^2} \\ &= \frac{\sum_{i=1}^{n} w_i^*}{(\sum_{i=1}^{n} w_i^* X_i^2)(\sum_{i=1}^{n} w_i^*) - (\sum_{i=1}^{n} w_i^* X_i)^2} \\ &= \frac{1}{\sum_{i=1}^{n} w_i^*(X_i - \bar{X}^*)^2} \end{aligned}$$

    where $w_i^* = (1/\sigma_i^2)$ and $\bar{X}^* = \sum_{i=1}^{n} w_i^* X_i / \sum_{i=1}^{n} w_i^*$.

4. *Relative Efficiency of OLS Under Heteroskedasticity.* Consider the simple linear regression with heteroskedasticity of the form $\sigma_i^2 = \sigma^2 X_i^\delta$ where $X_i = 1, 2, \ldots, 10$.

    (a) Compute $\mathrm{var}(\widehat{\beta}_{OLS})$ for $\delta = 0.5, 1, 1.5$ and 2.

    (b) Compute $\mathrm{var}(\widetilde{\beta}_{BLUE})$ for $\delta = 0.5, 1, 1.5$ and 2.

    (c) Compute the efficiency of $\widehat{\beta}_{OLS} = \mathrm{var}(\widetilde{\beta}_{BLUE})/\mathrm{var}(\widehat{\beta}_{OLS})$ for $\delta = 0.5, 1, 1.5$ and 2. What happens to this efficiency measure as δ increases?

5. Consider the *simple regression with only a constant* $y_i = \alpha + u_i$ for $i = 1, 2, \ldots, n$; where the u_i's are independent with mean zero and $\mathrm{var}(u_i) = \sigma_1^2$ for $i = 1, 2, \ldots, n_1$; and $\mathrm{var}(u_i) = \sigma_2^2$ for $i = n_1 + 1, \ldots, n_1 + n_2$ with $n = n_1 + n_2$.

    (a) Derive the OLS estimator of α along with its mean and variance.

    (b) Derive the GLS estimator of α along with its mean and variance.

    (c) Obtain the relative efficiency of OLS with respect to GLS. Compute their relative efficiency for various values of $\sigma_2^2/\sigma_1^2 = 0.2, 0.4, 0.6, 0.8, 1, 1.25, 1.33, 2.5, 5$; and $n_1/n = 0.2, 0.3, 0.4, \ldots,$ 0.8. Plot this relative efficiency.

    (d) Assume that u_i is $N(0, \sigma_1^2)$ for $i = 1, 2, \ldots, n_1$; and $N(0, \sigma_1^2)$ for $i = n_1 + 1, \ldots, n_1 + n_2$; with u_i's being independent. What is the maximum likelihood estimator of α, σ_1^2 and σ_2^2?

    (e) Derive the LR test for testing $H_0; \sigma_1^2 = \sigma_2^2$ in part (d).

6. Show that for an *AR(1) model* given in (5.26), $E(u_t u_s) = \rho^{|t-s|}\sigma_u^2$ for $t, s = 1, 2, \ldots, T$.

7. *Relative Efficiency of OLS Under the AR(1) Model.* This problem is based on Johnston (1984, pp. 310–312). For the simple regression without a constant $y_t = \beta x_t + u_t$ with $u_t = \rho u_{t-1} + \epsilon_t$ and $\epsilon_t \sim \mathrm{IID}(0, \sigma_\epsilon^2)$

    (a) Show that

$$\begin{aligned} \mathrm{var}(\widehat{\beta}_{OLS}) &= \frac{\sigma_u^2}{\sum_{t=1}^{T} x_t^2}\left(1 + 2\rho\frac{\sum_{t=1}^{T-1} x_t x_{t+1}}{\sum_{t=1}^{T} x_t^2} + 2\rho^2\frac{\sum_{t=1}^{T-2} x_t x_{t+2}}{\sum_{t=1}^{T} x_t^2}\right. \\ &\quad \left. + \ldots + 2\rho^{T-1}\frac{x_1 x_T}{\sum_{t=1}^{T} x_t^2}\right) \end{aligned}$$

    and that the Prais-Winsten estimator $\widehat{\beta}_{PW}$ has variance

$$\mathrm{var}(\widehat{\beta}_{PW}) = \frac{\sigma_u^2}{\sum_{t=1}^{T} x_t^2}\left[\frac{1-\rho^2}{1 + \rho^2 - 2\rho\sum_{t=1}^{T-1} x_t x_{t+1}/\sum_{t=1}^{T} x_t^2}\right]$$

    These expressions are easier to prove using matrix algebra, see Chapter 9.

(b) Let x_t itself follow an AR(1) scheme with parameter λ, i.e., $x_t = \lambda x_{t-1} + v_t$, and let $T \to \infty$. Show that

$$\begin{aligned} \text{asy eff}(\widehat{\beta}_{OLS}) &= \lim_{T\to\infty} \frac{\text{var}(\widehat{\beta}_{PW})}{\text{var}(\widehat{\beta}_{OLS})} = \frac{1-\rho^2}{(1+\rho^2-2\rho\lambda)(1+2\rho\lambda+2\rho^2\lambda^2+\ldots)} \\ &= \frac{(1-\rho^2)(1-\rho\lambda)}{(1+\rho^2-2\rho\lambda)(1+\rho\lambda)} \end{aligned}$$

(c) Tabulate this asy eff$(\widehat{\beta}_{OLS})$ for various values of ρ and λ where ρ varies between -0.9 to $+0.9$ in increments of 0.1, while λ varies between 0 and 0.9 in increments of 0.1. What do you conclude? How serious is the loss in efficiency in using OLS rather than the PW procedure?

(d) Ignoring this autocorrelation one would compute $\sigma_u^2/\sum_{t=1}^T x_t^2$ as the var$(\widehat{\beta}_{OLS})$. The difference between this wrong formula and that derived in part (a) gives us the bias in estimating the variance of $\widehat{\beta}_{OLS}$. Show that as $T \to \infty$, this asymptotic proportionate bias is given by $-2\rho\lambda/(1+\rho\lambda)$. Tabulate this asymptotic bias for various values of ρ and λ as in part (c). What do you conclude? How serious is the asymptotic bias of using the wrong variances for $\widehat{\beta}_{OLS}$ when the disturbances are first-order autocorrelated?

(e) Show that

$$\begin{aligned} E(s^2) &= \sigma_u^2 \left\{ T - \left(1 + 2\rho\frac{\sum_{t=1}^{T-1} x_t x_{t+1}}{\sum_{t=1}^T x_t^2} + 2\rho^2 \frac{\sum_{t=1}^{T-2} x_t x_{t+2}}{\sum_{t=1}^T x_t^2} \right.\right. \\ &\quad \left.\left. + \ldots + 2\rho^{T-1}\frac{x_1 x_T}{\sum_{t=1}^T x_t^2} \right) \right\} /(T-1) \end{aligned}$$

Conclude that if $\rho = 0$, then $E(s^2) = \sigma_u^2$. If x_t follows an AR(1) scheme with parameter λ, then for a large T, we get

$$E(s^2) = \sigma_u^2 \left(T - \frac{1+\rho\lambda}{1-\rho\lambda} \right) /(T-1)$$

Compute this $E(s^2)$ for $T = 101$ and various values of ρ and λ as in part (c). What do you conclude? How serious is the bias in using s^2 as an unbiased estimator for σ_u^2?

8. *OLS Variance Is Biased Under Serial Correlation.* For the AR(1) model given in (5.26), show that if $\rho > 0$ and the x_t's are positively autocorrelated that $E(s^2/\sum x_t^2)$ understates the var$(\widehat{\beta}_{OLS})$ given in (5.34).

9. Show that for the AR(1) model, the *Durbin-Watson statistic* has plim$d \to 2(1-\rho)$.

10. *Regressions with Non-zero Mean Disturbances.* Consider the simple regression with a constant

$$Y_i = \alpha + \beta X_i + u_i \quad i = 1, 2, \ldots, n$$

where α and β are scalars and u_i is independent of the X_i's. Show that:

(a) If the u_i's are independent and identically gamma distributed with $f(u_i) = \frac{1}{\Gamma(\theta)} u_i^{\theta-1} e^{-u_i}$ where $u_i \geq 0$ and $\theta > 0$, then $\widehat{\alpha}_{OLS} - s^2$ is unbiased for α.

(b) If the u_i's are independent and identically χ^2 distributed with ν degrees of freedom, then $\widehat{\alpha}_{OLS} - s^2/2$ is unbiased for α.

(c) If the u_i's are independent and identically exponentially distributed with f$(u_i) = \frac{1}{\theta} e^{-u_i/\theta}$ where $u_i \geq 0$ and $\theta > 0$, then $\widehat{\alpha}_{OLS} - s$ is consistent for α.

11. *The Heteroskedastic Consequences of an Arbitrary Variance for the Initial Disturbance of an AR(1) Model*. This is based on Baltagi and Li (1990, 1992). Consider a simple AR(1) model

$$u_t = \rho u_{t-1} + \epsilon_t \quad t = 1, 2, \ldots, T \quad |\rho| < 1$$

with $\epsilon_t \sim \text{IID}(0, \sigma_\epsilon^2)$ independent of $u_0 \sim (0, \sigma_\epsilon^2/\tau)$, and τ is an arbitrary positive parameter.

(a) Show that this arbitrary variance on the initial disturbance u_0 renders the disturbances, in general, heteroskedastic.

(b) Show that $\text{var}(u_t) = \sigma_t^2$ is increasing if $\tau > (1 - \rho^2)$ and decreasing if $\tau < (1 - \rho^2)$. When is the process homoskedastic?

(c) Show that $\text{cov}(u_t, u_{t-s}) = \rho^s \sigma_{t-s}^2$ for $t \geq s$. **Hint:** See the solution by Kim (1991).

(d) Consider the simple regression model

$$y_t = \beta x_t + u_t \quad t = 1, 2 \ldots, T$$

with u_t following the AR(1) process described above. Consider the common case where $\rho > 0$ and the x_t's are positively autocorrelated. For this case, it is a standard result that the $\text{var}(\widehat{\beta}_{OLS})$ is understated under the stationary case (i.e., $(1 - \rho^2) = \tau$), see problem 8. This means that OLS rejects too often the hypothesis $H_0; \beta = 0$. Show that OLS will reject more often than the stationary case if $\tau < 1 - \rho^2$ and less often than the stationary case if $\tau > (1 - \rho^2)$. **Hint:** See the solution by Koning (1992).

12. *ML Estimation of Linear Regression Model with AR(1) Errors and Two Observations*. This is based on Magee (1993). Consider the regression model $y_i = x_i \beta + u_i$, with only two observations $i = 1, 2$, and the nonstochastic $|x_1| \neq |x_2|$ are scalars. Assume that $u_i \sim N(0, \sigma^2)$ and $u_2 = \rho u_1 + \epsilon$ with $|\rho| < 1$. Also, $\epsilon \sim N[0, (1 - \rho^2)\sigma^2]$ where ϵ and u_1 are independent.

(a) Show that the OLS estimator of β is $(x_1 y_1 + x_2 y_2)/(x_1^2 + x_2^2)$.

(b) Show that the ML estimator of β is $(x_1 y_1 - x_2 y_2)/(x_1^2 - x_2^2)$.

(c) Show that the ML estimator of ρ is $2x_1 x_2/(x_1^2 + x_2^2)$ and thus is nonstochastic.

(d) How do the ML estimates of β and ρ behave as $x_1 \to x_2$ and $x_1 \to -x_2$? Assume $x_2 \neq 0$. **Hint:** See the solution by Baltagi and Li (1995).

13. For the empirical example in section 5.5 based on the Cigarette Consumption Data in Table 3.2.

(a) Replicate the OLS regression of log$C$ on log$P$, log$Y$ and a constant. Plot the residuals versus log$Y$ and verify Figure 5.1.

(b) Run Glejser's (1969) test by regressing $|e_i|$ the absolute value of the residuals from part (a), on $(\text{log}Y_i)^\delta$ for $\delta = 1, -1, -0.5$ and 0.5. Verify the t-statistics reported in the text.

(c) Run Goldfeld and Quandt's (1965) test by ordering the observations according to logY_i and omitting 12 central observations. Report the two regressions based on the first and last 17 observations and verify the F-test reported in the text.

(d) Verify the Spearman rank correlation test based on the rank (logY_i) and rank $|e_i|$.

(e) Verify Harvey's (1976) multiplicative heteroskedasticity test based on regressing loge_i^2 on log(logY_i).

(f) Run the Breusch and Pagan (1979) test based on the regression of $e_i^2/\widehat{\sigma}^2$ on logY_i, where $\widehat{\sigma}^2 = \sum_{i=1}^{46} e_i^2/46$.

(g) Run White's (1980) test for heteroskedasticity.

(h) Run the Jarque and Bera (1987) test for normality of the disturbances.

(i) Compute White's (1980) heteroskedasticity robust standard errors for the regression in part (a).

14. *A Simple Linear Trend Model with AR(1) Disturbances*. This is based on Krämer (1982).

    (a) Consider the following simple linear trend model

    $$Y_t = \alpha + \beta t + u_t$$

    where $u_t = \rho u_{t-1} + \epsilon_t$ with $|\rho| < 1$, $\epsilon_t \sim \text{IID}(0, \sigma_\epsilon^2)$ and $\text{var}(u_t) = \sigma_u^2 = \sigma_\epsilon^2/(1-\rho^2)$. Our interest is focused on the estimates of the trend coefficient, β, and the estimators to be considered are OLS, CO (assuming that the true value of ρ is known), the first-difference estimator (FD), and the Generalized Least Squares (GLS), which is Best Linear Unbiased (BLUE) in this case.

    In the context of the simple linear trend model, the formulas for the variances of these estimators reduce to

    $$\begin{aligned} V(OLS) &= 12\sigma^2\{-6\rho^{T+1}[(T-1)\rho-(T+1)]^2-(T^3-T)\rho^4 \\ &+2(T^2-1)(T-3)\rho^3+12(T^2+1)\rho^2-2(T^2-1)(T+3)\rho \\ &+(T^3-T)\}/(1-\rho^2)(1-\rho)^4(T^3-T)^2 \\ V(CO) &= 12\sigma^2(1-\rho)^2(T^3-3T^2+2T), \\ V(FD) &= 2\sigma^2(1-\rho^{T-1})/(1-\rho^2)(T-1)^2, \\ V(GLS) &= 12\sigma^2/(T-1)[(T-3)(T-2)\rho^2-2(T-3)(T-1)\rho+T(T+1)]. \end{aligned}$$

    (b) Compute these variances and their relative efficiency with respect to the GLS estimator for $T = 10, 20, 30, 40$ and ρ between -0.9 and 0.9 in 0.1 increments.

    (c) For a given T, show that the limit of $\text{var}(OLS)/\text{var}(CO)$ is zero as $\rho \to 1$. Prove that $\text{var}(FD)$ and $\text{var}(GLS)$ both tend in the limit to $\sigma_\epsilon^2/(T-1) < \infty$ as $\rho \to 1$. Conclude that $\text{var}(GLS)/\text{var}(FD)$ tend to 1 as $\rho \to 1$. Also, show that $\lim_{\rho\to 1}[\text{var}(GLS)/\text{var}(OLS)] = 5(T^2+T)/6(T^2+1) < 1$ provided $T > 3$.

    (d) For a given ρ, show that $\text{var}(FD) = O(T^{-2})$ whereas the variance of the remaining estimators is $O(T^{-3})$. Conclude that $\lim_{T\to\infty}[\text{var}(FD)/\text{var}(CO)] = \infty$ for any given ρ.

15. Consider the empirical example in section 5.6, based on the Consumption-Income data in Table 5.3. Obtain this data set from the CONSUMP.DAT file on the Springer web site.

    (a) Replicate the OLS regression of C_t on Y_t and a constant, and compute the Durbin-Watson statistic. Test $H_0; \rho = 0$ versus $H_1; \rho > 0$ at the 5% significance level.

    (b) Test for first-order serial correlation using the Breusch and Godfrey test.

    (c) Perform the two-step Cochrane-Orcutt procedure and verify Table 5.5. What happens if we iterate the Cochrane-Orcutt procedure?

    (d) Perform the Prais-Winsten procedure and verify Table 5.6.

    (e) Compute the Newey-West heteroskedasticity and autocorrelation-consistent standard errors for the least squares estimates in part (a).

16. Benderly and Zwick (1985) considered the following equation

    $$RS_t = \alpha + \beta Q_{t+1} + \gamma P_t + u_t$$

    where RS_t = the real return on stocks in year t, Q_{t+1} = the annual rate of growth of real GNP in year $t+1$, and P_t = the rate of inflation in year t. The data is provided on the Springer web site and labeled BENDERLY.ASC. This data covers 31 annual observations for the U.S. over the period 1952–1982. This was obtained from Lott and Ray (1991). This equation is used to test the significance of the inflation rate in explaining real stock returns. Use the sample period 1954–1976 to answer the following questions:

(a) Run OLS to estimate the above equation. Remember to use Q_{t+1}. Is P_t significant in this equation? Plot the residuals against time. Compute the Newey-West heteroskedasticity and autocorrelation-consistent standard errors for these least squares estimates.

(b) Test for serial correlation using the D.W. test.

(c) Would your decision in (b) change if you used the Breusch-Godfrey test for first-order serial correlation?

(d) Run the Cochrane-Orcutt procedure to correct for first-order serial correlation. Report your estimate of ρ.

(e) Run a Prais-Winsten procedure accounting for the first observation and report your estimate of ρ. Plot the residuals against time.

17. Using our cross-section Energy/GDP data set in Chapter 3, problem 3.16 consider the following two models:

$$\begin{aligned} &Model\ 1: \quad \log En = \alpha + \beta \log RGDP + u \\ &Model\ 2: \quad En = \alpha + \beta RGDP + v \end{aligned}$$

Make sure you have corrected the W. Germany observation on EN as described in problem 3.16 part (d).

(a) Run OLS on both Models 1 and 2. Test for heteroskedasticity using the Goldfeldt/Quandt Test. Omit $c = 6$ central observations. Why is heteroskedasticity a problem in Model 2, but not Model 1?

(b) For Model 2, test for heteroskedasticity using the Glejser Test.

(c) Now use the Breusch-Pagan Test to test for heteroskedasticity on Model 2.

(d) Apply White's Test to Model 2.

(e) Do all these tests give the same decision?

(f) Propose and estimate a simple transformation of Model 2, assuming heteroskedasticity of the form $\sigma_i^2 = \sigma^2 RGDP^2$.

(g) Propose and estimate a simple transformation of Model 2, assuming heteroskedasticity of the form $\sigma_i^2 = \sigma^2(a + bRGDP)^2$.

(h) Now suppose that heteroskedasticity is of the form $\sigma_i^2 = \sigma^2 RGDP^\gamma$ where γ is an unknown parameter. Propose and estimate a simple transformation for Model 2. **Hint:** You can write σ_i^2 as $\exp\{\alpha + \gamma \log RGDP\}$ where $\alpha = \log\sigma^2$.

(i) Compare the standard errors of the estimates for Model 2 from OLS, also obtain White's heteroskedasticity-consistent standard errors. Compare them with the simple Weighted Least Squares estimates of the standard errors in parts (f), (g) and (h). What do you conclude?

18. You are given quarterly data from the first quarter of 1965 (1965.1) to the fourth quarter of 1983 (1983.4) on employment in Orange County California (EMP) and real gross national product (RGNP). The data set is in a file called ORANGE.DAT on the Springer web site.

(a) Generate the lagged variable of real GNP, call it $RGNP_{t-1}$ and estimate the following model by OLS: $EMP_t = \alpha + \beta RGNP_{t-1} + u_t$.

(b) What does inspection of the residuals and the Durbin-Watson statistic suggest?

(c) Assuming $u_t = \rho u_{t-1} + \epsilon_t$ where $|\rho| < 1$ and $\epsilon_t \sim \text{IIN}(0, \sigma_\epsilon^2)$, use the Cochrane-Orcutt procedure to estimate ρ, α and β. Compare the latter estimates and their standard errors with those of OLS.

(d) The Cochrane-Orcutt procedure omits the first observation. Perform the Prais-Winsten adjustment. Compare the resulting estimates and standard error with those in part (c).

(e) Apply the Breusch-Godfrey test for first and second order autoregression. What do you conclude?

(f) Compute the Newey-West heteroskedasticity and autocorrelation-consistent covariance standard errors for the least squares estimates in part (a).

19. Consider the earning data underlying the regression in Table 4.1 and available on the Springer web site as EARN.ASC.

(a) Apply White's test for heteroskedasticity to the regression residuals.

(b) Compute White's heteroskedasticity-consistent standard errors.

(c) Test the least squares residuals for normality using the Jarque-Bera test.

20. *Hedonic Housing*. Harrison and Rubinfield (1978) collected data on 506 census tracts in the Boston area in 1970 to study hedonic housing prices and the willingness to pay for clean air. This data is available on the Springer web site as HEDONIC.XLS. The dependent variable is the *Median Value* (MV) of owner-occupied homes. The regressors include two structural variables, RM the average number of rooms, and AGE representing the proportion of owner units built prior to 1940. In addition there are eight neighborhood variables: B, the proportion of blacks in the population; LSTAT, the proportion of population that is lower status; CRIM, the crime rate; ZN, the proportion of 25000 square feet residential lots; INDUS, the proportion of nonretail business acres; TAX, the full value property tax rate ($/$10000); PTRATIO, the pupil-teacher ratio; and CHAS represents the dummy variable for Charles River: = 1 if a tract bounds the Charles. There are also two accessibility variables, DIS the weighted distances to five employment centers in the Boston region, and RAD the index of accessibility to radial highways. One more regressor is an air pollution variable NOX, the annual average nitrogen oxide concentration in parts per hundred million.

(a) Run OLS of MV on the 13 independent variables and a constant. Plot the residuals.

(b) Apply White's tests for heteroskedasticity.

(c) Obtain the White heteroskedasticity-consistent standard errors.

(d) Test the least squares residuals for normality using the Jarque-Bera test.

21. *Agglomeration Economies, Diseconomies, and Growth.* Wheeler (2003) uses data on 3106 counties of the contiguous USA to fit a fourth-order polynomial relating County population (employment) growth (over the period 1980 to 1990) as a function of log(size), where size is measured as total resident population or total civilian employment. Other control variables include the proportion of the adult resident population (i.e. of age 25 or older) with a bachelor's degree or more; the proportion of total employment in manufacturing; and the unemployment rate, all for the year 1980; Per capita income in 1979; the proportion of the resident population belonging to non-white racial categories in 1980, and the share of local government expenditures going to each of three public goods-education, roads and highways, police protection-in 1982. This data can be downloaded from the JAE archive data web site.

(a) Replicate the OLS regressions reported in Tables VIII and IX of Wheeler (2003, pp. 88–89).

(b) Apply White's and Breusch-Pagan tests for heteroskedasticity.

(c) Test the least squares residuals for normality using the Jarque-Bera test.

References

For additional readings consult the econometrics books cited in the Preface. Also the chapter on heteroskedasticity by Griffiths (2001), and the chapter on serial correlation by King (2001):

Ali, M.M. and C. Giaccotto (1984), "A study of Several New and Existing Tests for Heteroskedasticity in the General Linear Model," *Journal of Econometrics*, 26: 355–373.

Amemiya, T. (1973), "Regression Analysis When the Variance of the Dependent Variable is Proportional to the Square of its Expectation," *Journal of the American Statistical Association*, 68: 928–934.

Amemiya, T. (1977), "A Note on a Heteroskedastic Model," *Journal of Econometrics*, 6: 365–370.

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59: 817–858.

Baltagi, B. and Q. Li (1990), "The Heteroskedastic Consequences of an Arbitrary Variance for the Initial Disturbance of an AR(1) Model," *Econometric Theory*, Problem 90.3.1, 6: 405.

Baltagi, B. and Q. Li (1992), "The Bias of the Standard Errors of OLS for an AR(1) process with an Arbitrary Variance on the Initial Observations," *Econometric Theory*, Problem 92.1.4, 8: 146.

Baltagi, B. and Q. Li (1995), "ML Estimation of Linear Regression Model with AR(1) Errors and Two Observations," *Econometric Theory*, Solution 93.3.2, 11: 641–642.

Bartlett's test, M.S. (1937), "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Statistical Society*, A, 160: 268–282.

Beach, C.M. and J.G. MacKinnon (1978), "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors," *Econometrica*, 46: 51–58.

Benderly, J. and B. Zwick (1985), "Inflation, Real Balances, Output and Real Stock Returns," *American Economic Review*, 75: 1115–1123.

Breusch, T.S. (1978), "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers*, 17: 334–355.

Breusch, T.S. and A.R. Pagan (1979), "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica*, 47: 1287–1294.

Buse, A. (1984), "Tests for Additive Heteroskedasticity: Goldfeld and Quandt Revisited," *Empirical Economics*, 9: 199–216.

Carroll, R.H. (1982), "Adapting for Heteroskedasticity in Linear Models," *Annals of Statistics*, 10: 1224–1233.

Cochrane, D. and G. Orcutt (1949), "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, 44: 32–61.

Cragg, J.G. (1992), "Quasi-Aitken Estimation for Heteroskedasticity of Unknown Form," *Journal of Econometrics*, 54: 197–202.

Durbin, J. (1960), "Estimation of Parameters in Time-Series Regression Model," *Journal of the Royal Statistical Society*, Series B, 22: 139–153.

Durbin, J. and G. Watson (1950), "Testing for Serial Correlation in Least Squares Regression-I," *Biometrika*, 37: 409–428.

Durbin, J. and G. Watson (1951), "Testing for Serial Correlation in Least Squares Regression-II," *Biometrika*, 38: 159–178.

Evans, M.A., and M.L. King (1980) "A Further Class of Tests for Heteroskedasticity," *Journal of Econometrics*, 37: 265–276.

Farebrother, R.W. (1980), "The Durbin-Watson Test for Serial Correlation When There is No Intercept in the Regression," *Econometrica*, 48: 1553–1563.

Glejser, H. (1969), "A New Test for Heteroskedasticity," *Journal of the American Statistical Association*, 64: 316–323.

Godfrey, L.G. (1978), "Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46: 1293–1302.

Goldfeld, S.M. and R.E. Quandt (1965), "Some Tests for Homoscedasticity," *Journal of the American Statistical Association*, 60: 539–547.

Goldfeld, S.M. and R.E. Quandt (1972), *Nonlinear Methods in Econometrics* (North-Holland: Amsterdam).

Griffiths, W.E. (2001), "Heteroskedasticity," Chapter 4 in B.H. Baltagi, (ed.), *A Companion to Theoretical Econometrics* (Blackwell: Massachusetts).

Harrison, M. and B.P. McCabe (1979), "A Test for Heteroskedasticity Based On Ordinary Least Squares Residuals," *Journal of the American Statistical Association*, 74: 494–499.

Harrison, D. and D.L. Rubinfeld (1978), "Hedonic Housing Prices and the Demand for Clean Air," *Journal of Environmental Economics and Management*, 5: 81–102.

Harvey, A.C. (1976), "Estimating Regression Models With Multiplicative Heteroskedasticity," *Econometrica*, 44: 461–466.

Hilderth, C. and J. Lu (1960), "Demand Relations with Autocorrelated Disturbances," Technical Bulletin 276 (Michigan State University, Agriculture Experiment Station).

Jarque, C.M. and A.K. Bera (1987), "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, 55: 163–177.

Kim, J.H. (1991), "The Heteroskedastic Consequences of an Arbitrary Variance for the Initial Disturbance of an AR(1) Model," *Econometric Theory*, Solution 90.3.1, 7: 544–545.

King, M. (2001), "Serial Correlation," Chapter 2 in B.H. Baltagi, (ed.), *A Companion to Theoretical Econometrics* (Blackwell: Massachusetts).

Koenker, R. (1981), "A Note on Studentizing a Test for Heteroskedasticity," *Journal of Econometrics*, 17: 107–112.

Koenker, R. and G.W. Bassett, Jr. (1982), "Robust Tests for Heteroskedasticity Based on Regression Quantiles," *Econometrica*, 50:43–61.

Koning, R.H. (1992), "The Bias of the Standard Errors of OLS for an AR(1) process with an Arbitrary Variance on the Initial Observations," *Econometric Theory*, Solution 92.1.4, 9: 149–150.

Krämer, W. (1982), "Note on Estimating Linear Trend When Residuals are Autocorrelated," *Econometrica*, 50: 1065–1067.

Lott, W.F. and S.C. Ray (1992), *Applied Econometrics: Problems With Data Sets* (The Dryden Press: New York).

Maddala, G.S. (1977), *Econometrics* (McGraw-Hill: New York).

Maeshiro, A. (1976), "Autoregressive Transformation, Trended Independent Variables and Autocorrelated Disturbance Terms," *The Review of Economics and Statistics*, 58: 497–500.

Maeshiro, A. (1979), "On the Retention of the First Observations in Serial Correlation Adjustment of Regression Models," *International Economic Review*, 20: 259–265.

Magee L. (1993), "ML Estimation of Linear Regression Model with AR(1) Errors and Two Observations," *Econometric Theory*, Problem 93.3.2, 9: 521–522.

Mizon, G.E. (1995), "A Simple Message for Autocorrelation Correctors: Don't," *Journal of Econometrics* 69: 267–288.

Newey, W.K. and K.D. West (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55: 703–708.

Oberhofer, W. and J. Kmenta (1974), "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models," *Econometrica*, 42: 579–590.

Park, R.E. and B.M. Mitchell (1980), "Estimating the Autocorrelated Error Model With Trended Data," *Journal of Econometrics*, 13: 185–201.

Prais, S. and C. Winsten (1954), "Trend Estimation and Serial Correlation," Discussion Paper 383 (Cowles Commission: Chicago).

Rao, P. and Z. Griliches (1969), "Some Small Sample Properties of Several Two-Stage Regression Methods in the Context of Autocorrelated Errors," *Journal of the American Statistical Association*, 64: 253–272.

Rao, P. and R. L. Miller (1971), *Applied Econometrics* (Wadsworth: Belmont).

Robinson, P.M. (1987), "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 55: 875–891.

Rutemiller, H.C. and D.A. Bowers (1968), "Estimation in a Heteroskedastic Regression Model," *Journal of the American Statistical Association*, 63: 552–557.

Savin, N.E. and K.J. White (1977), "The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors," *Econometrica*, 45: 1989–1996.

Szroeter, J. (1978), "A Class of Parametric Tests for Heteroskedasticity in Linear Econometric Models," *Econometrica*, 46: 1311–1327.

Theil, H. (1978), *Introduction to Econometrics* (Prentice-Hall: Englewood Cliffs, NJ).

Waldman, D.M. (1983), "A Note on Algebraic Equivalence of White's Test and a Variation of the Godfrey/Breusch-Pagan Test for Heteroskedasticity," *Economics Letters*, 13: 197–200.

Wheeler, C. (2003), "Evidence on Agglomeration Economies, Diseconomies, and Growth," *Journal of Applied Econometrics*, 18: 79–104.

White, H. (1980), "A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48: 817–838.

Wooldridge, J.M. (1991), "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics*, 47: 5–46.