EE425(1/2011)

Relaxing the Assumptions of Classical Model

Autocorrelation in Disturbance:

When Error Terms are Correlated.

Autocorrelation

- Nature of autocorrelation
- Consequences of autocorrelation
- Detection/ test for autocorrelation
- How do we remedy this problem?

What is autocorrelation?
(CLRM=Classical Linear Regression Model)

One assumption in CLRM is

$\text{cov}(u_i, u_j \mid X_i, X_j) = E(u_i, u_j) = 0$ for all $i \neq j$

If $E(u_i, u_j) \neq 0$ for some $i \neq j$, meaning that the error terms are correlated, then we have autocorrelation or serial correlation.

The problem of autocorrelation is more common in time series data.

Why autocorrelation exists?

- Specification Bias:
  - Excluded relevant variable
  - Incorrect functional form
- Inertia: the effects of one variable on others do not occurred instantly, but take time, hence change in time series data occurs gradually. Therefore error terms in different period can be related.
- Cobweb Phenomenon: example $S_t = \beta_0 + \beta_1 P_{t-1} + u_i$

  the supply of an agricultural product at time t depends on what happen in year t-1, hence the error terms in different period may be correlated.
- Including lag dependent variable in the model

  $C_t = \beta_0 + \beta_1$ income $+ \beta_2 C_{t-1} + u_t$
- Data transformation

4

- Example: Instead of estimating $Y_t = \beta_0 + \beta_1 X_t + u_t$

  we estimate $\Delta Y_t = \beta_1 \Delta X_t + v_t$

  where $\Delta Y_t = Y_t - Y_{t-1}$ and $\Delta X_t = X_t - X_{t-1}$

- If $Y_t = \beta_0 + \beta_1 X_t + u_t$, then $Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + u_{t-1}$

  $Y_t - Y_{t-1} = \beta_1 (X_t - X_{t-1}) + (u_t - u_{t-1})$

- $E(v_t v_{t-1}) = E[(u_t - u_{t-1})(u_{t-1} - u_{t-2})] \neq 0$

- The transformation causes autocorrelation.

Figure 12.3 (a) Positive and (b) Negative Autocorrelation

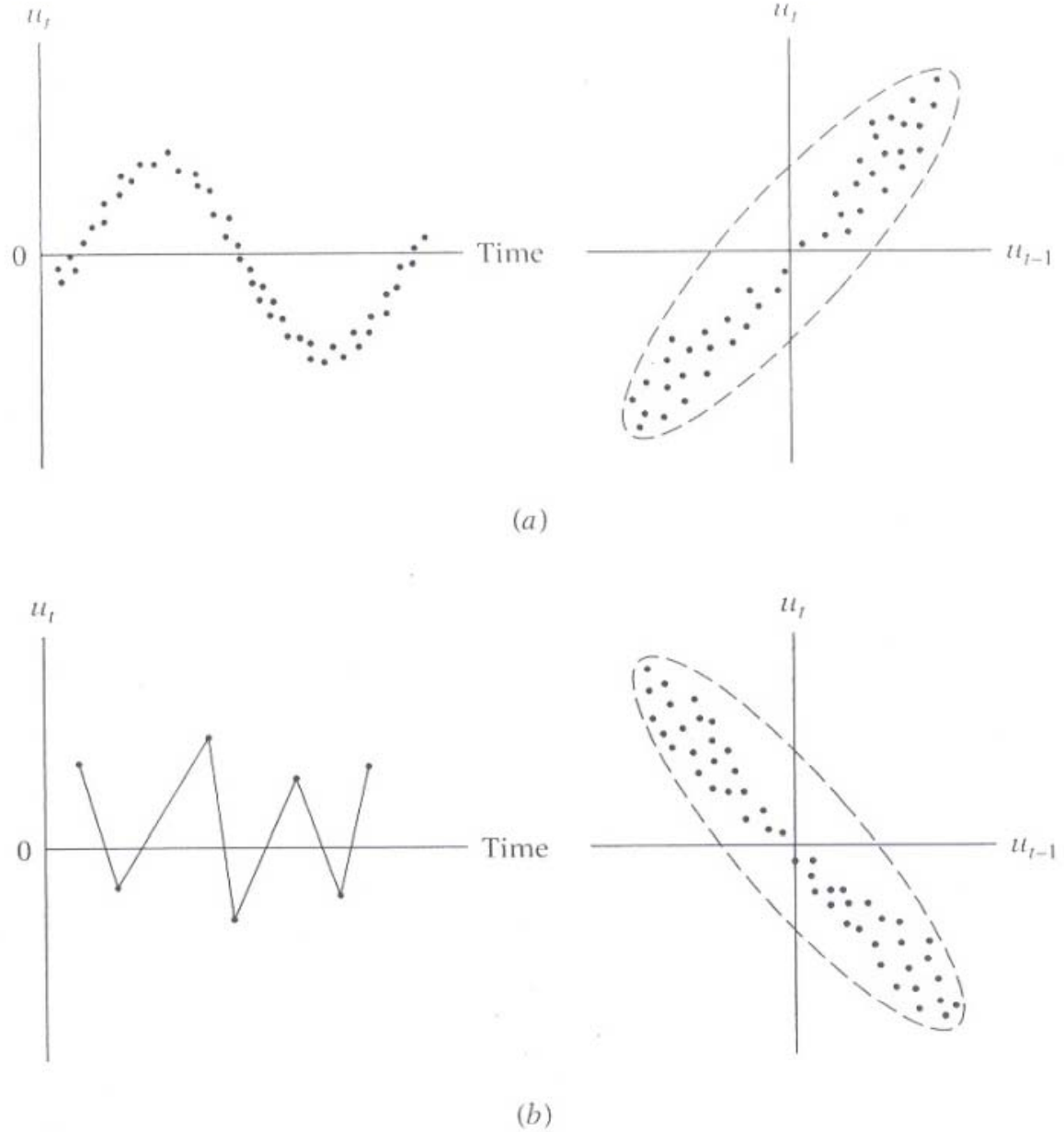
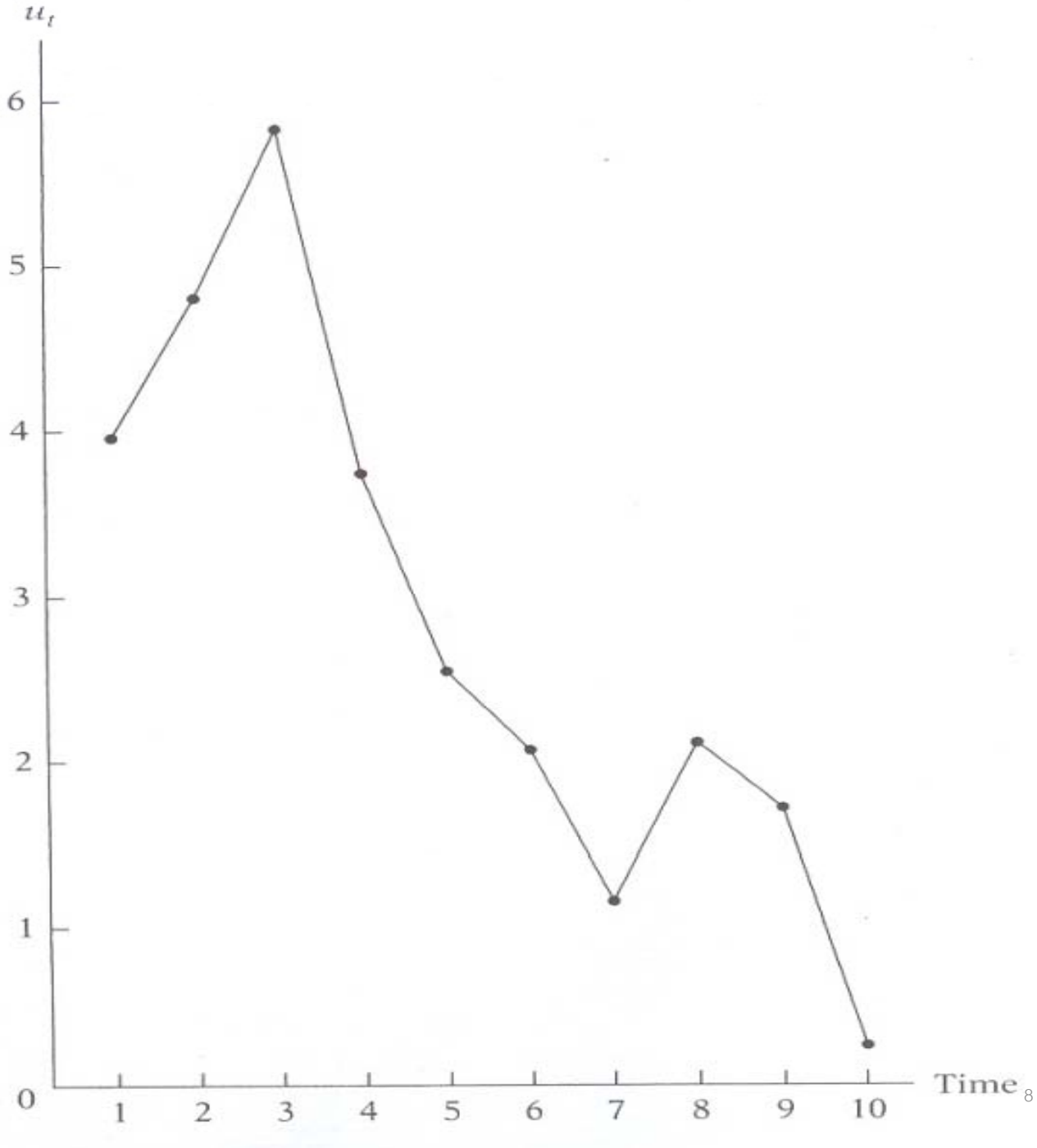


Table 12.1 A Hypothetical Example of Positively Autocorrelated Error Terms

	ε_t	$u_t = 0.7u_{t-1} + \varepsilon_t$
0	0	$u_0 = 5$ (assumed)
1	0.464	$u_1 = 0.7(5) + 0.464 = 3.964$
2	2.026	$u_2 = 0.7(3.964) + 2.0262 = 4.8008$
3	2.455	$u_3 = 0.7(4.8010) + 2.455 = 5.8157$
4	−0.323	$u_4 = 0.7(5.8157) - 0.323 = 3.7480$
5	−0.068	$u_5 = 0.7(3.7480) - 0.068 = 2.5556$
6	0.296	$u_6 = 0.7(2.5556) + 0.296 = 2.0849$
7	−0.288	$u_7 = 0.7(2.0849) - 0.288 = 1.1714$
8	1.298	$u_8 = 0.7(1.1714) + 1.298 = 2.1180$
9	0.241	$u_9 = 0.7(2.1180) + 0.241 = 1.7236$
10	−0.957	$u_{10} = 0.7(1.7236) - 0.957 = 0.2495$

Note: ε_t data obtained from *A Million Random Digits and One Hundred Thousand Deviates*, Rand Corporation, Santa Monica, Calif., 1950.

Figure 12.5
Errors u_t Generated by the Scheme from table 12.1



As a starting point, or first approximation, one can assume that the disturbance, or error, terms are generated by the following mechanism.

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad -1 < \rho < 1 \tag{12.2.1}$$

where ρ (= rho) is known as the **coefficient of autocovariance** and where ε_t is the stochastic disturbance term such that it satisfies the standard OLS assumptions, namely,

$$\begin{aligned} E(\varepsilon_t) &= 0 \\ \text{var}\,(\varepsilon_t) &= \sigma_\varepsilon^2 \\ \text{cov}\,(\varepsilon_t, \varepsilon_{t+s}) &= 0 \qquad s \neq 0 \end{aligned} \tag{12.2.2}$$

In the engineering literature, an error term with the preceding properties is often called a **white noise error term.** What Eq. (12.2.1) postulates is that the value of the disturbance term in period t is equal to ρ times its value in the previous period plus a purely random error term.

The scheme (12.2.1) is known as a **Markov first-order autoregressive scheme,** or simply a **first-order autoregressive scheme,** usually denoted as **AR(1).** The name *autoregressive* is appropriate because Eq. (12.2.1) can be interpreted as the regression of u_t on itself lagged one period. It is first order because u_t and its immediate past value are involved; that is, the maximum lag is 1. If the model were $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$, it would be an AR(2), or second-order, autoregressive scheme, and so on

Note that ρ, the coefficient of autocovariance in Eq. (12.2.1), can also be interpreted as the **first-order coefficient of autocorrelation,** or more accurately, the **coefficient of autocorrelation at lag 1.**[9]

Given the AR(1) scheme, it can be shown that (see Appendix 12A, Section 12A.2):

$$\text{var}(u_t) = E(u_t^2) = \frac{\sigma_\varepsilon^2}{1-\rho^2} \tag{12.2.3}$$

$$\text{cov}(u_t, u_{t+s}) = E(u_t u_{t-s}) = \rho^s \frac{\sigma_\varepsilon^2}{1-\rho^2} \tag{12.2.4}$$

$$\text{cor}(u_t, u_{t+s}) = \rho^s \tag{12.2.5}$$

where $\text{cov}(u_t, u_{t+s})$ means covariance between error terms s periods apart and where $\text{cor}(u_t, u_{t+s})$ means correlation between error terms s periods apart. Note that because of the symmetry property of covariances and correlations, $\text{cov}(u_t, u_{t+s}) = \text{cov}(u_t, u_{t-s})$ and $\text{cor}(u_t, u_{t+s}) = \text{cor}(u_t, u_{t-s})$.

two-variable regression model: $Y_t = \beta_1 + \beta_2 X_t + u_t$. We know from Chapter 3 that the OLS estimator of the slope coefficient is

$$\hat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2} \tag{12.2.6}$$

and its variance is given by

$$\text{var}\,(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{12.2.7}$$

where the small letters as usual denote deviation from the mean values.

Now under the AR(1) scheme, it can be shown that the variance of this estimator is

$$\text{var}\,(\hat{\beta}_2)_{\text{AR1}} = \frac{\sigma^2}{\sum x_t^2}\left[1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum x_t^2} + \cdots + 2\rho^{n-1}\frac{x_1 x_n}{\sum x_t^2}\right] \tag{12.2.8}$$

where $\text{var}\,(\hat{\beta}_2)_{\text{AR1}}$ means the variance of $\hat{\beta}_2$ under a first-order autoregressive scheme.

The BLUE Estimator in the Presence of Autocorrelation

Continuing with the two-variable model and assuming the AR(1) process, we can show that the BLUE estimator of β_2 is given by the following expression:[11]

$$\hat{\beta}_2^{\text{GLS}} = \frac{\sum_{t=2}^{n}(x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\sum_{t=2}^{n}(x_t - \rho x_{t-1})^2} + C \qquad (12.3.1)$$

where C is a correction factor that may be disregarded in practice. Note that the subscript t now runs from $t = 2$ to $t = n$. And its variance is given by

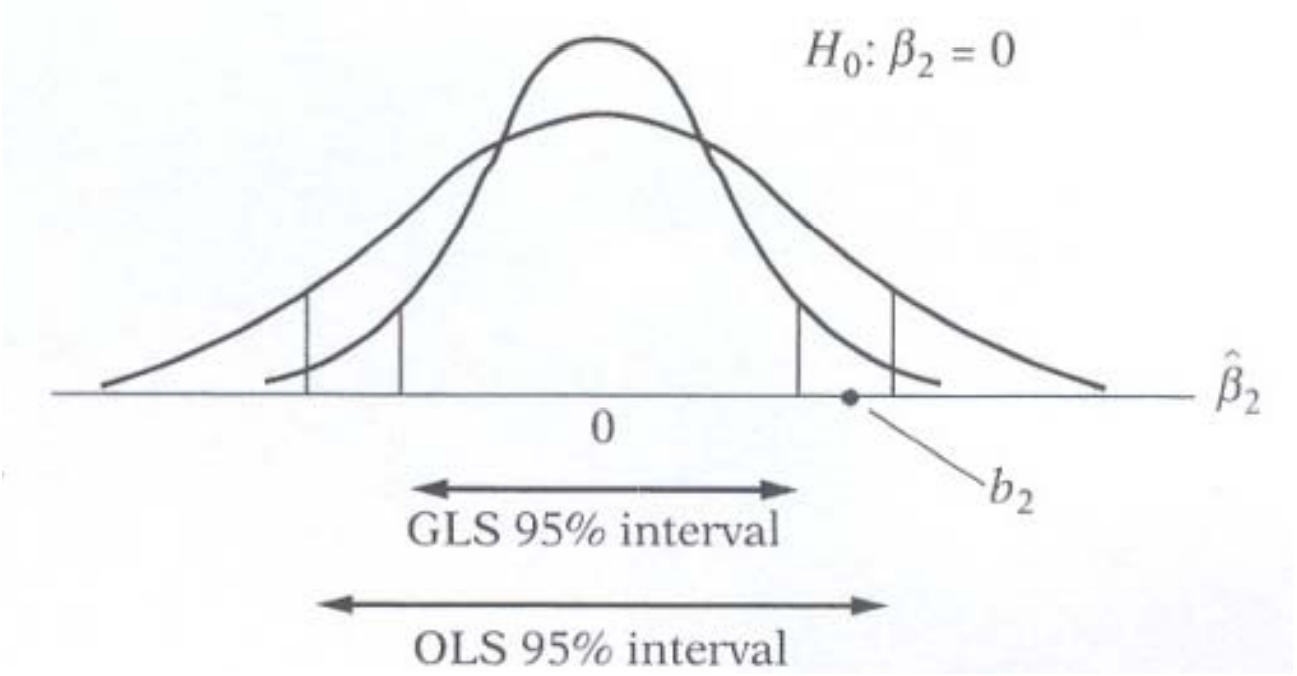
$$\text{var}\,\hat{\beta}_2^{\text{GLS}} = \frac{\sigma^2}{\sum_{t=2}^{n}(x_t - \rho x_{t-1})^2} + D \qquad (12.3.2)$$

where D too is a correction factor that may also be disregarded in practice.

Consequences of Using OLS in the Presence of Autocorrelation

The message is: To establish confidence intervals and to test hypotheses, one should use GLS and not OLS even though the estimators derived from the latter are unbiased and consistent. (However, see Section 12.11 later.)

Figure 12.4 GLS and OLS 95% confidence intervals



OLS Estimation Disregarding Autocorrelation

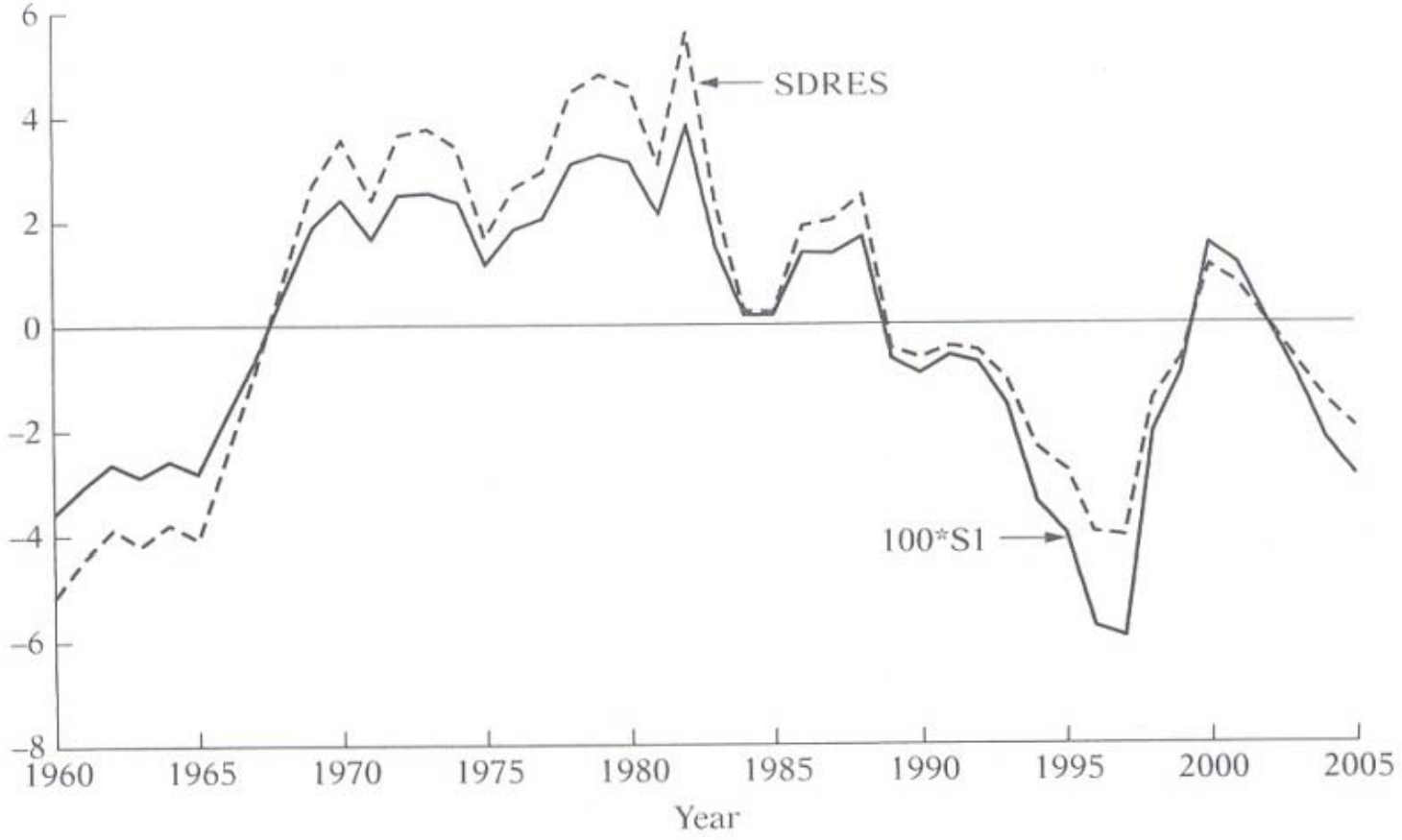
The situation is potentially very serious if we not only use $\hat{\beta}_2$ but also continue to use $\text{var}(\hat{\beta}_2) = \sigma^2 / \sum x_t^2$, which completely disregards the problem of autocorrelation, that is, we mistakenly believe that the usual assumptions of the classical model hold true. Errors will arise for the following reasons:

1. The residual variance $\hat{\sigma}^2 = \sum \hat{u}_t^2/(n-2)$ is likely to underestimate the true σ^2.
2. As a result, we are likely to overestimate R^2.
3. Even if σ^2 is not underestimated, $\text{var}(\hat{\beta}_2)$ may underestimate $\text{var}(\hat{\beta}_2)_{\text{AR1}}$ (Eq. [12.2.8]), its variance under (first-order) autocorrelation, even though the latter is inefficient compared to $\text{var}(\hat{\beta}_2)^{\text{GLS}}$.
4. Therefore, the usual t and F tests of significance are no longer valid, and if applied, are likely to give seriously misleading conclusions about the statistical significance of the estimated regression coefficients.

Detecting Autocorrelation

I. Graphical Method

Figure 12.8 Residuals and Standardized residuals from the wages-productivity regression



II. Durbin-Watson d Test

The most celebrated test for detecting serial correlation is that developed by statisticians Durbin and Watson. It is popularly known as the **Durbin–Watson *d* statistic,** which is defined as

$$d = \frac{\sum_{t=2}^{t=n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n}\hat{u}_t^2} \qquad (12.6.5)$$

which is simply the ratio of the sum of squared differences in successive residuals to the RSS. Note that in the numerator of the d statistic the number of observations is $n - 1$ because one observation is lost in taking successive differences.

A great advantage of the d statistic is that it is based on the estimated residuals, which are routinely computed in regression analysis. Because of this advantage, it is now a common practice to report the Durbin–Watson d along with summary measures, such as R^2, adjusted R^2, t, and F. Although it is now routinely used, it is **important to note the assumptions underlying the *d* statistic.**

1. The regression model includes the intercept term. If it is not present, as in the case of the regression through the origin, it is essential to rerun the regression including the intercept term to obtain the RSS.[22]

2. The explanatory variables, the X's, are nonstochastic, or fixed in repeated sampling.

3. The disturbances u_t are generated by the first-order autoregressive scheme: $u_t = \rho u_{t-1} + \varepsilon_t$. Therefore, it cannot be used to detect higher-order autoregressive schemes.

4. The error term u_t is assumed to be normally distributed.

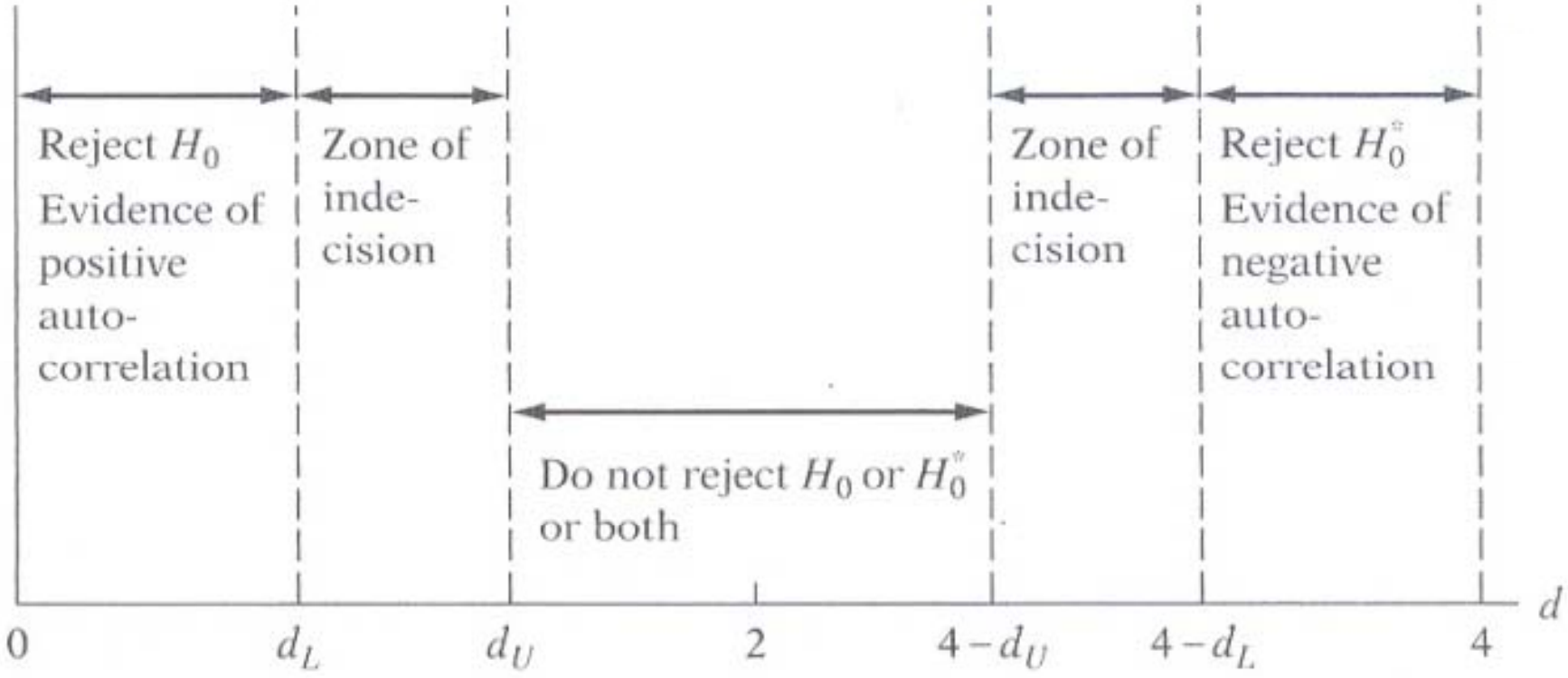
5. The regression model does not include the lagged value(s) of the dependent variable as one of the explanatory variables. Thus, the test is inapplicable in models of the following type:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + \gamma Y_{t-1} + u_t \quad (12.6.6)$$

where Y_{t-1} is the one period lagged value of Y. Such models are known as **autoregressive models,** which we will study in Chapter 17.

6. There are no missing observations in the data. Thus, in our wages–productivity regression for the period 1960–2005, if observations for, say, 1978 and 1982 were missing for some reason, the d statistic would make no allowance for such missing observations.[23]

Figure 12.10 Durbin-Watson d Statistic



Legend

H_0: No positive autocorrelation

H_0^*: No negative autocorrelation

The exact sampling or probability distribution of the d statistic given in Eq. (12.6.5) is difficult to derive because, as Durbin and Watson have shown, it depends in a complicated way on the X values present in a given sample.[24] This difficulty should be understandable because d is computed from $\hat{u}_t$, which are, of course, dependent on the given X's. Therefore, unlike the t, F, or χ^2 tests, there is no unique critical value that will lead to the rejection or the acceptance of the null hypothesis that there is no first-order serial correlation in the disturbances u_t. However, Durbin and Watson were successful in deriving a lower bound d_L and an upper bound d_U such that if the computed d from Eq. (12.6.5) lies outside these critical values, a decision can be made regarding the presence of positive or negative serial correlation. Moreover, these limits depend only on the number of observations n and the number of explanatory variables and do not depend on the values taken by these explanatory variables. These limits, for n going from 6 to 200 and up to 20 explanatory variables, have been tabulated by Durbin and Watson and are reproduced in **Appendix D,** Table D.5 (up to 20 explanatory variables).

The actual test procedure can be explained better with the aid of Figure 12.10, which shows that the limits of d are 0 and 4. These can be established as follows. Expand Eq. (12.6.5) to obtain

$$d = \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \tag{12.6.7}$$

Since $\sum \hat{u}_t^2$ and $\sum \hat{u}_{t-1}^2$ differ in only one observation, they are approximately equal. Therefore, setting $\sum \hat{u}_{t-1}^2 \approx \sum \hat{u}_t^2$, Eq. (12.6.7) may be written as

$$d \approx 2\left(1 - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}\right) \tag{12.6.8}$$

where $\approx$ means approximately.

Now let us define

$$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \tag{12.6.9}$$

as the sample first-order coefficient of autocorrelation, an estimator of ρ. (See footnote 9.) Using Eq. (12.6.9), we can express Eq. (12.6.8) as

$$d \approx 2(1 - \hat{\rho}) \tag{12.6.10}$$

But since $-1 \leq \rho \leq 1$, Eq. (12.6.10) implies that

$$0 \leq d \leq 4 \tag{12.6.11}$$

The mechanics of the Durbin–Watson test are as follows, assuming that the assumptions underlying the test are fulfilled:

1. Run the OLS regression and obtain the residuals.
2. Compute d from Eq. (12.6.5). (Most computer programs now do this routinely.)
3. For the given sample size and given number of explanatory variables, find out the critical d_L and d_U values.
4. Now follow the decision rules given in Table 12.6. For ease of reference, these decision rules are also depicted in Figure 12.10.

To illustrate the mechanics, let us return to our wages–productivity regression. From the data given in Table 12.5 the estimated d value can be shown to be 0.2175, suggesting that there is positive serial correlation in the residuals. From the Durbin–Watson tables, we find that for 46 observations and one explanatory variable, $d_L = 1.475$ and $d_U = 1.566$ at the 5 percent level. Since the computed d of 0.2175 lies below d_L, we cannot reject the hypothesis that there is positive serial correlation in the residuals.

Although extremely popular, the d test has one great drawback in that, if it falls in the **indecisive zone,** one cannot conclude that (first-order) autocorrelation does or does not exist.

Table 12.6 Durbin-Watson d Test: Decision Rules

Null Hypothesis	Decision	If
No positive autocorrelation	Reject	$0 < d < d_L$
No positive autocorrelation	No decision	$d_L \leq d \leq d_U$
No negative correlation	Reject	$4 - d_L < d < 4$
No negative correlation	No decision	$4 - d_U \leq d \leq 4 - d_L$
No autocorrelation, positive or negative	Do not reject	$d_U < d < 4 - d_U$

To solve this problem, several authors have proposed modifications of the d test but they are rather involved and beyond the scope of this book.[25] In many situations, however, it has been found that the upper limit d_U is approximately the true significance limit and therefore in case d lies in the indecisive zone, one can use the following **modified *d* test:** Given the level of significance α,

1. H_0: $\rho = 0$ versus H_1: $\rho > 0$. Reject H_0 at α level if $d < d_U$. That is, there is statistically significant positive autocorrelation.
2. H_0: $\rho = 0$ versus H_1: $\rho < 0$. Reject H_0 at α level if the estimated $(4 - d) < d_U$, that is, there is statistically significant evidence of negative autocorrelation.
3. H_0: $\rho = 0$ versus H_1: $\rho \neq 0$. Reject H_0 at 2α level if $d < d_U$ or $(4 - d) < d_U$, that is, there is statistically significant evidence of autocorrelation, positive or negative.

III. A General Test of Autocorrelation: The Breusch-Godfrey (BG) Test

To avoid some of the pitfalls of the Durbin–Watson d test of autocorrelation, statisticians Breusch and Godfrey have developed a test of autocorrelation that is general in the sense that it allows for (1) nonstochastic regressors, such as the lagged values of the regressand; (2) higher-order autoregressive schemes, such as AR(1), AR(2), etc.; and (3) simple or higher-order **moving averages** of white noise error terms, such as ε_t in Eq. (12.2.1).[31]

Without going into the mathematical details, which can be obtained from the references, the **BG test**, which is also known as the **LM test**,[32] proceeds as follows: We use the two-variable regression model to illustrate the test, although many regressors can be added to the model. Also, lagged values of the regressand can be added to the model. Let

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{12.6.14}$$

Assume that the error term u_t follows the pth-order autoregressive, AR(p), scheme as follows:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t \tag{12.6.15}$$

where ε_t is a white noise error term as discussed previously. As you will recognize, this is simply the extension of the AR(1) scheme.

The null hypothesis H_0 to be tested is that

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_p = 0 \qquad (12.6.16)$$

That is, there is no serial correlation of any order. The BG test involves the following steps:

1. Estimate Eq. (12.6.14) by OLS and obtain the residuals, $\hat{u}_t$.

2. Regress $\hat{u}_t$ on the original X_t (if there is more than one X variable in the original model, include them also) and $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-p}$, where the latter are the lagged values of the estimated residuals in step 1. Thus, if $p = 4$, we will introduce four lagged values of the residuals as additional regressors in the model. Note that to run this regression we will have only $(n - p)$ observations (why?). In short, run the following regression:

$$\hat{u}_t = \alpha_1 + \alpha_2 X_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \cdots + \hat{\rho}_p \hat{u}_{t-p} + \varepsilon_t \qquad (12.6.17)$$

and obtain R^2 from this (auxiliary) regression.[33]

3. If the sample size is large (technically, infinite), Breusch and Godfrey have shown that

$$(n - p)R^2 \sim \chi^2_p \qquad (12.6.18)$$

That is, asymptotically, $n - p$ times the R^2 value obtained from the auxiliary regression (12.6.17) follows the chi-square distribution with p df. If in an application, $(n - p)R^2$ exceeds the critical chi-square value at the chosen level of significance, we reject the null hypothesis, in which case at least one ρ in Eq. (12.6.15) is statistically significantly different from zero.

The following *practical points* about the BG test may be noted:

1. The regressors included in the regression model may contain lagged values of the regressand Y, that is, Y_{t-1}, Y_{t-2}, etc., may appear as explanatory variables. Contrast this model with the Durbin–Watson test restriction that there may be no lagged values of the regressand among the regressors.

2. As noted earlier, the BG test is applicable even if the disturbances follow a pth-order **moving average (MA)** process, that is, the u_t are generated as follows:

$$u_t = \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \lambda_2 \varepsilon_{t-2} + \cdots + \lambda_p \varepsilon_{t-p} \qquad (12.6.19)$$

where ε_t is a white noise error term, that is, the error term that satisfies all the classical assumptions.

In the chapters on time series econometrics, we will study in some detail the pth-order autoregressive and moving average processes.

3. If in Eq. (12.6.15) $p = 1$, meaning first-order autoregression, then the BG test is known as **Durbin's *M* test.**

4. A drawback of the BG test is that the value of p, the length of the lag, cannot be specified a priori. Some experimentation with the p value is inevitable. Sometimes one can use the so-called **Akaike** and **Schwarz** information criteria to select the lag length. We will discuss these criteria in Chapter 13 and later in the chapters on time series econometrics.

5. Given the values of the X variable(s) and the lagged values of u, the test assumes that the variance of u in Eq. (12.6.15) is homoscedastic.

Correcting for Autocorrelation: The Method of Generalized Least Squares

Knowing the consequences of autocorrelation, especially the lack of efficiency of OLS estimators, we may need to remedy the problem. The remedy depends on the knowledge one has about the nature of interdependence among the disturbances, that is, knowledge about the structure of autocorrelation.

As a starter, consider the two-variable regression model:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (12.9.1)$$

and assume that the error term follows the AR(1) scheme, namely,

$$u_t = \rho u_{t-1} + \varepsilon_t \quad -1 < \rho < 1 \qquad (12.9.2)$$

Now we consider two cases: (1) ρ is known and (2) ρ is not known but has to be estimated.

When ρ Is Known

If the coefficient of first-order autocorrelation is known, the problem of autocorrelation can be easily solved. If Eq. (12.9.1) holds true at time t, it also holds true at time $(t-1)$. Hence,

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \tag{12.9.3}$$

Multiplying Eq. (12.9.3) by ρ on both sides, we obtain

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1} \tag{12.9.4}$$

Subtracting Eq. (12.9.4) from Eq. (12.9.1) gives

$$(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t \tag{12.9.5}$$

where $\varepsilon_t = (u_t - \rho u_{t-1})$

We can express Eq. (12.9.5) as

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t \tag{12.9.6}$$

where $\beta_1^* = \beta_1(1-\rho)$, $Y_t^* = (Y_t - \rho Y_{t-1})$, $X_t^* = (X_t - \rho X_{t-1})$, and $\beta_2^* = \beta_2$.

When ρ Is Not Known

Although conceptually straightforward to apply, the method of generalized difference given in Eq. (12.9.5) is difficult to implement because ρ is rarely known in practice. Therefore, we need to find ways of estimating ρ. We have several possibilities.

The First-Difference Method

Since ρ lies between 0 and ± 1, one could start from two extreme positions. At one extreme, one could assume that $\rho = 0$, that is, no (first-order) serial correlation, and at the other extreme we could let $\rho = \pm 1$, that is, perfect positive or negative correlation. As a matter of fact, when a regression is run, one generally assumes that there is no autocorrelation and then lets the Durbin–Watson or other test show whether this assumption is justified. If, however, $\rho = +1$, the generalized difference equation (12.9.5) reduces to the **first-difference equation:**

$$Y_t - Y_{t-1} = \beta_2(X_t - X_{t-1}) + (u_t - u_{t-1})$$

or

$$\Delta Y_t = \beta_2 \Delta X_t + \varepsilon_t \tag{12.9.7}$$

where Δ is the first-difference operator introduced in Eq. (12.1.10).

Since the error term in Eq. (12.9.7) is free from (first-order) serial correlation (why?), to run the regression (12.9.7) all one has to do is form the first differences of both the regressand and regressor(s) and run the regression on these first differences.

Iterative Methods of Estimating ρ

All the methods of estimating ρ discussed previously provide us with only a single estimate of ρ. But there are the so-called **iterative methods** that estimate ρ iteratively, that is, by successive approximation, starting with some initial value of ρ. Among these methods the following may be mentioned: the **Cochrane–Orcutt iterative procedure,** the **Cochrane–Orcutt two-step procedure,** the **Durbin two–step procedure,** and the **Hildreth–Lu scanning or search procedure.** Of these, the most popular is the Cochran–Orcutt iterative method. To save space, the iterative methods are discussed by way of exercises. Remember that the ultimate objective of these methods is to provide an estimate of ρ that may be used to obtain GLS estimates of the parameters. One advantage of the Cochrane–Orcutt iterative method is that it can be used to estimate not only an AR(1) scheme, but also higher-order autoregressive schemes, such as $\hat{u}_t = \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + v_t$, which is AR(2). Having obtained the two ρs, one can easily extend the generalized difference equation (12.9.6). Of course, the computer can now do all this.