ANYLYSIS OF RESIDUALS
AUTOCORRELATION

In Chapters 2 and 3, we stated that for the least squares modeling procedure to be best linear unbiased estimator (BLUE), the five fundamental assumptions for a simple regression model and six for a multiple regression model must be satisfied. Four of the assumptions are related to the model's residuals. They are re-numbered and listed below:

1. The residuals are normally distributed with a zero mean, that is, $E(u_i) = 0$.
2. The residuals have a constant variance, σ^2, that is, there is no heteroscedasticity.
3. The successive residuals are not correlated, that is, there is no autocorrelation.
4. The X variables are non-stochastic (that is, Xs are fixed) and are not correlated with the residuals.

It will become apparent that the first assumption of a zero mean is always fulfilled because of the nature of least squares estimation. The assumption of normal distribution of residuals is not a concern as far as the BLUE property is concerned. The Gauss-Markov theorem requires only that the residuals have a zero mean and a constant variance. However, in hypothesis testing, normality of the residuals is a requirement. The other three assumptions, however, are important in OLS estimation, and they do not always hold. A violation of any of these three assumptions could affect the model's forecasting performance.

Assumption 2 is usually more often a problem associated with cross-sectional data than with time series data. The problem of heteroscedasticity will be addressed in Chapter 5 along with cross-sectional modeling. Problems associated with non-stochastic independent variables will be discussed in Chapter 6.

Assumption 3 is more often a problem in time series modeling than in cross-sectional modeling. In Chapters 2 and 3, we developed simple and multiple regression models on the basis of time series data without performing an analysis of the residuals. In this chapter, we will address the problems resulting from the violation of assumptions 3, that is, the problem of autocorrelation.

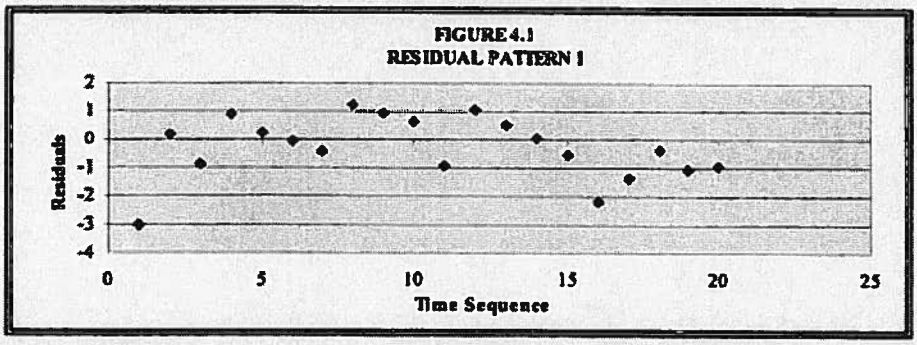
However, in this chapter, we will concentrate on the problems of first order autocorrelation and leave problems associated with higher order autocorrelation to later chapters.

PATTERNS OF RESIDUAL DISTRIBUTION

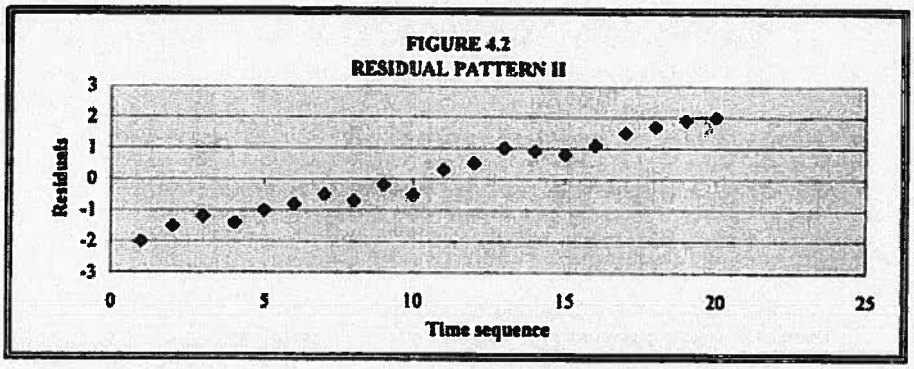
The distribution of residuals of a model can exhibit many different patterns. The following three patterns are probably most common in regression modeling with time series data.

Pattern I. Figure 4.1 is a plot of 20 random numbers generated from the standardized normal distribution, which is a bell-shaped probability distribution described in Appendix 2.2, Chapter 2. These random numbers were generated with the add-in random number generator in EXCEL. The 20 points scatter about the zero line without any particular pattern. This is an ideal situation, and in regression modeling, we want the model residuals to be randomly distributed.

FIGURE 4.1
RESIDUAL PATTERN I

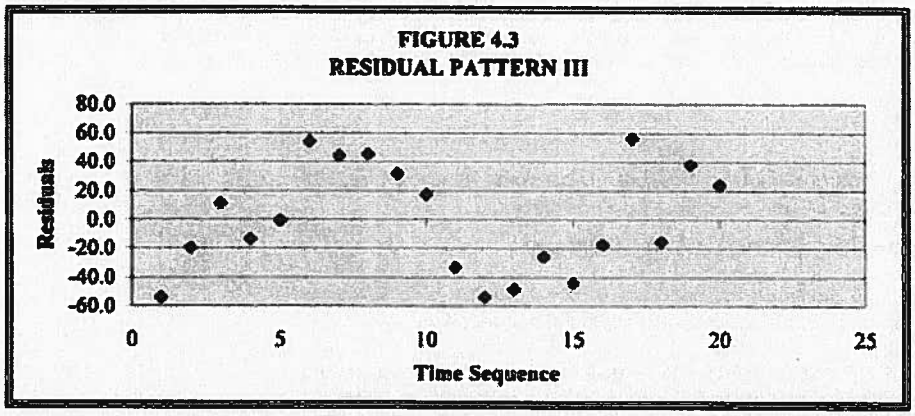


Pattern II. Figure 4.2 is a plot of 20 arbitrarily chosen points, which show an upward trend. This pattern is a special case of autocorrelation, which indicates that a trend variable is missing from the model.



This type of problem is easy to handle. All modelers need to do is to include a trend indicator in the model. The trend indicator can be either a relevant economic variable or a time sequence, such as T = 1, 2, 3, ...

Pattern III: Figure 4.3 is a plot of another type of autocorrelated residuals. The residuals exhibit a pattern with several consecutive negative numbers followed by several positive numbers like a wave. This is the most common pattern of autocorrelated residuals and is more difficult to handle than Pattern II. If we allow the autocorrelation to remain in the model, it will adversely affect the performance of the model.



SOURCES OF AUTOCORRELATION

Autocorrelation can be found in both cross-sectional and time series data. In cross-sectional modeling, data drawn from one region may reflect the characteristics of the neighboring regions, or data drawn from one income group may reflect the characteristics of other income groups. This type of autocorrelation is called spatial autocorrelation. Autocorrelation in economic time series data is a reflection of the culture and institutional traditions of the population, which produced the series. In other words, what people did in the past would affect their present and future activities. Therefore, autocorrelation would appear if the model were not correctly specified. Following are the two types of misspecification, which can cause autocorrelation:

1. A misspecification of the functional form of a model can result in autocorrelation in the residuals. Suppose that we specified a model as a linear function:

$$Y_t = \alpha + \beta_1 X_t + \beta_2 Z_t + u_t \quad \text{... (4.1)}$$

But the true model is a quadratic function, which is:

$$Y_t = \alpha + \beta_1 X_t + \beta_2 Z_t^2 + u_t \quad \text{... (4.2)}$$

In this case, autocorrelation may appear in the residuals.

2. Often a variable, which is autocorrelated, can be explained by another autocorrelated variable. If this variable is not included in the model, autocorrelation will occur. The reason of missing an important variable from the model is either that data for the variable are not available, or that it is impractical to include too many variables in the model. If the missing variable has a strong autocorrelation, the model will have a problem.

CONSEQUENCES OF AUTOCORRELATION

The presence of autocorrelation does not cause a bias in the estimation of model coefficients, but it reduces the efficiency of a model for forecasting. This is because:

1. It can be mathematically shown that the presence of autocorrelation increases the variance of the residuals as well as the variances of the estimated coefficients. Since the efficiency of a model is inversely related to the variances, an increase in these variances will reduce its efficiency.

2. Direct application of OLS to the autocorrelated data will fail to properly estimate the variances of the model. In fact, OLS will under-estimate these variances invalidating the hypothesis testing. Recall that both F test and t test depend on the variances of the residuals and of the estimated coefficients. If OLS under-estimates these variances, then the calculated F and t values will be larger than they should be. As a result, modelers will be misled to believe that the β coefficients are significantly different from zero when they are not.

Since the OLS procedure is not efficient when autocorrelation is present, it is important to test the OLS model for autocorrelation. If it exists, a generalized least squares (GLS) procedure may be used to correct it.

TESTS FOR DETECTING FIRST ORDER AUTOCORRELATION

There are several statistical tests available for detecting first order autocorrelation in a model. The following two are the ones most often used:

1. ***The Visual Test.*** Plot the residuals of a model and then visually determine if it exhibits any pattern. The autocorrelation exists if the residuals exhibit a specific pattern as shown in Figures 4.2 and 4.3.

2. ***The Durbin-Watson d and the Durbin H Tests.*** The d test is most often used in detecting the first order autocorrelation and the H test is used when the lagged dependent variable is incorporated in a model. Let the estimated model for Equation (4.1) be:

$$Y_t = a + b_1X_t + b_2Z_t + e_t \qquad \ldots (4.3)$$

The Durbin-Watson d statistic for testing first order autocorrelation in e_t is calculated as follows:

$$d = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2} \quad \ldots (4.4)$$

If Z_t is the lagged dependent variable used as an explanatory variable, then Equation (4.3) becomes:

$$Y_t = a + b_1 X_t + b_2 Y_{t-1} + e_t$$

Here the H statistic, instead of d, should be used for the test. The H statistic is calculated as follows:

$$H = \rho \sqrt{\frac{n}{1 - n\text{Var}\ (b_2)}}$$

Here b_2 is the estimated coefficient for the lagged dependent variable, and p is an estimate of the first order autocorrelation coefficient ρ. The H test will be discussed in Chapter 6.

The Durbin-Watson d Test

Most computing software computes the Durbin-Watson d statistic automatically. Although EXCEL does not, it can be calculated easily according to Equation (4.4). Once the d statistic is calculated, it is used for testing the following hypothesis:

H_0: $\rho = 0$

H_A: $\rho \neq 0$

Where ρ is the unknown first order autocorrelation coefficient, the testing procedure is as follows:

1. Select a level of significance. Let it be 5%.

2. Go to Table A-5 in the Appendix of Tables. At the 5% significance level, find the critical values for d_L and d_U corresponding to the number of observations, n, and the number of explanatory variables, k'.

3. If $d \leq d_L$, reject the null hypothesis for $\rho = 0$, meaning that autocorrelation is present in the residuals. If $d_U < d \leq 2$, accept the null hypothesis for $\rho = 0$. If $d_L < d < d_U$, the test is inconclusive.

4. If $d > 2$, test for negative autocorrelation by comparing $(4 - d)$ with d_L and d_U as in step 1 through step 3. If $(4 - d) \leq d_L$, negative autocorrelation is suspected. If $(4 - d) \geq d_U$, accept the null hypothesis for $\rho = 0$, that is, there is no autocorrelation in the residuals.

The d test is more powerful for models based on large samples than on small samples. For small sample models, some authors suggest that as long as d is greater than d_L and less than 2.5 accept the null hypothesis for $\rho = 0$. There is no set criterion for how large is a large sample. Normally, a sample of less than 30 observations is not a large sample as far as the d test is concerned.

REMEDIAL PROCEDURES FOR FIRST ORDER AUTOCORRELATION

Let's begin our discussion with Equation (4.1), which is a linear model. Assume that the u_t term has a first order autocorrelation in the following form:

$$u_t = \rho u_{t-1} + v_t \quad \ldots (4.5a)$$

$$u_t - \rho u_{t-1} = v_t \quad \ldots (4.5b)$$

As stated earlier, ρ is the true but unknown autocorrelation coefficient of a specified model. This condition is a violation of assumption 3 listed in the beginning of this chapter. Equation (4.5a) is an autoregressive model of order one. The residual term in the equation is v_t, which is normally distributed with a zero mean and a constant variance. This type of autocorrelation is often referred to as the first order autoregressive process, or AR (1) process. In order to remove the autocorrelation from a model, we need to transform Equation (4.1) in accordance with the AR (1) scheme. First multiply ρ to both sides of Equation (4.1), with all the variables in the equation lagged by one period.

$$\rho Y_{t-1} = \rho\alpha + \rho\beta_1 X_{t-1} + \rho\beta_2 Z_{t-1} + \rho u_{t-1} \quad \ldots (4.6)$$

Then subtract Equation (4.6) from Equation (4.1) to obtain:

$$Y_t - \rho Y_{t-1} = \alpha(1-\rho) + \beta_1(X_t - \rho X_{t-1}) + \beta_2(Z_t - \rho Z_{t-1}) + (u_t - \rho u_{t-1}) \quad \ldots (4.7)$$

Note that the residual term of Equation (4.7) is simply v_t in Equation (4.5b). By re-arranging terms, Equation (4.7) becomes:

$$Y_t = \alpha(1-\rho) + \beta_1(X_t - \rho X_{t-1}) + \beta_2(Z_t - \rho Z_{t-1}) + \rho Y_{t-1} + v_t \quad \ldots (4.8)$$

This transformation creates an unfortunate situation. Unless the autocorrelation coefficient, ρ, is known, we cannot use OLS to estimate coefficients, α, β_1, β_2 and ρ because Equation (4.8) includes terms like $\beta_1\rho X_{t-1}$, which makes the model non-linear. If we can come up with a ρ value, we can first transform the data, and then apply the OLS to the transformed data to estimate the other coefficients. Let's use the lower case letter to denote the transformed data as follows:

$$y_t = Y_t - \rho Y_{t-1} \quad \ldots (4.9a)$$

$$\text{Constant} = 1 - \rho \quad \ldots (4.9b)$$

$$x_t = X_t - \rho X_{t-1} \quad \ldots (4.9c)$$

$$z_t = Z_t - \rho Z_{t-1} \quad \ldots (4.9d)$$

Note that the constant term may be considered an independent variable in regression estimation except that before the transformation its value is 1 for the entire modeling period. (In this book, we do not call the constant term an independent variable.) This type of transformation is often called generalized differencing. In Chapters 8 and 9, we will encounter modeling situations requiring simple differencing. By substituting the transformed variables into Equation (4.1), it becomes:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + v_t \quad \ldots (4.10)$$

$$\beta_0 = \alpha(1-\rho) \quad \ldots (4.11)$$

The form of Equation (4.10) is similar to that of Equation (4.1), and the coefficients β_0, β_1 and β_2 can be estimated with OLS on the basis of a sample. According to Equation (4.11), the estimate for α is simply:

$$\alpha = \frac{\beta_0}{1-\rho} \qquad \ldots (4.11a)$$

Then, how do we come up with a good ρ value for the transformation? The following two methods are often used in regression modeling when autocorrelation is present. We will first discuss them descriptively and then with the help of an example.

Cochran-Orcutt Iterative Procedure

The Cochran-Orcutt method is a two-step iterative procedure. In step one, an initial estimate of the autocorrelation coefficient, p, is determined; in step two, the data of each variable is transformed with this initial p, and then regression is run on the transformed data. The procedure works as follows:

Step 1. Run an OLS regression on the original data without any transformation. The OLS estimated model is as follow:

$$Y_t = a + b_1X_t + b_2Z_t + e_t$$

or

$$\hat{Y}_t = a + b_1X_t + b_2Z_t \qquad \ldots (4.12)$$

Where $\hat{Y}$ represents the regression line. The residuals, e_t, is calculated as follows:

$$e_t = Y_t - \hat{Y}_t \qquad \ldots (4.13)$$

Estimate ρ according to Equation (4.5a). Let p be an estimate of ρ, then the first order autoregressive equation of the residuals is as follows:

$$\hat{e}_t = pe_{t-1} \qquad \ldots (4.14)$$

Step 2. In this step, do the following:

a. First use the initial p from Equation (4.14) to transform the data according to Equations (4.9a) through (4.9d). Then, apply the OLS to Equation (4.10) along with Equation (4.11a) to obtain a new set of a, b_1 and b_2, which are estimates of α, β_1, and β_2 based on the transformed data.

b. Go back to Step 1 and replace the coefficients in Equation (4.12) with the new set of a, b_1 and b_2 obtained in Step 2 (a). Calculate the residuals, e_t, according to Equation (4.13), and estimate a new p according to Equation (4.14).

c. Go to Step 2 (a) and repeat the procedure until the Durbin-Watson d test indicates that autocorrelation is no longer present or until p changes less than a pre-determined number, say 0.001. Usually it takes one to three full iterations to reach convergence.

The Durbin Two-Stage Procedure

The Durbin Two-Stage method is also a two-step procedure, which differs from the Cochran-Orcutt procedure only in the choice of an initial p in the first step. To compare the difference, let's first expand Equation (4.8) as follows:

$$Y_t = \alpha(1-\rho) + \beta_1 X_t - \beta_1\rho X_{t-1} + \beta_2 Z_t - \beta_2\rho Z_{t-1} + \rho Y_{t-1} + v_t \qquad \ldots (4.15)$$

Equation (4.15) is a linear equation if $\alpha(1-\rho)$, $\beta_1\rho$ and $\beta_2\rho$ are considered single numbers. Then, we can apply the OLS to regress Y_t on X_t, X_{t-1}, Z_t, Z_{t-1} and Y_{t-1}. The coefficient of Y_{t-1} of this regression equation will be the initial estimate of ρ. According to Durbin, the two stages are as follows:

1. Apply the OLS to estimate Equation (4.15), considering $\alpha(1-\rho)$, $\beta_1\rho$ and $\beta_2\rho$ as single numbers. Let the estimated coefficient of Y_{t-1} be p; then p is the initial estimate for ρ.

2. Step two is the same as the Cochran-Orcutt procedure. (a) Use p obtained in step 1 to transform Y, X and Z according to Equations (4.9a) through (4.9d), and apply OLS to the transformed data to obtain a new set of a, b_1 and b_2. Then go through Cochran-Orcutt's

step 2 (b) and step 2 (c).

The iterative procedure of Cochran-Orcutt and the Durbin two-stage procedure are both generalized least squares (GLS) procedures. The two procedures differ only in the choice of an initial p for data transformation. Since these procedures use an estimated p for variable transformations, they are called *feasible* generalized least squares (FGLS) procedure.

Effect of Losing the First Observation: The data transformation involved in the two procedures described above will result in a loss of the first data point due to the generalized differencing. For example, $y_1 = Y_2 - pY_1$, so y_1 moves one period forward, and the sample size becomes $n - 1$. For large samples, e.g., $n > 30$, a loss of one point may not make any difference in model efficiency;

but for small samples, e.g. $n < 20$, this loss would adversely affect model efficiency. Prais-Winsten proposed a modification of the two procedures by restoring the first data points of all variables in the sample. The restoring procedure is to multiply the $\sqrt{1-p^2}$ factor to the first observation of each variable including the constant term as follows:

Original variable	Transformed Variable	
Constant term = 1	$\sqrt{1-p^2}$	... (4.16a)
Y_1	$(\sqrt{1-p^2})\,Y_1$	... (4.16b)
X_1	$(\sqrt{1-p^2})\,X_1$	... (4.16c)
Z_1	$(\sqrt{1-p^2})\,Z_1$	... (4.16d)

The two researchers have shown that with the first data points restored, efficiency of the two procedures improves. We will demonstrate how these two procedures work, and how the first observation of each variable may be restored with the Prais-Winsten procedure.

AN ILLUSTRATIVE EXAMPLE

Let's modify the model developed in Chapter 2 for forecasting annual residential electric demand by including in the model another explanatory variable, the summer weather condition in terms of cooling-degree days (CDD).

Let demand be Y; the number of residential customers be X; and summer cooling-degree days be Z. Then the specified demand model becomes:

$$Y_t = \alpha + \beta_1 X_t + \beta_2 Z_t + u_t \qquad \text{... (4.17)}$$

Let the estimated model be:

$$\hat{Y}_t = a + b_1 X_t + b_2 Z_t \qquad \text{... (4.18)}$$

Coefficients a, b_1, and b_2 are the estimates for α, β_1 and β_2. For the observed Y_t Equation (4.18) requires a residual term, e_t, as follows:

$$Y_t = a + b_1 X_t + b_2 Z_t + e_t \qquad \text{... (4.19)}$$

The residuals, e_t are the difference between the observed Y_t and its estimate, $\hat{Y}_t$.

Table 4.1 lists the required modeling data, which includes twenty observations for each variable from 1980 through 1999. The model output is given in Table 4.2. The model predicted values, $\hat{Y}_t$, and the residuals, e_t, are listed in the fifth and sixth columns of Table 4.1. Data given in the last six columns are for later use.

TABLE 4.1
ELECTRIC DEMAND, USERS AND COOLING DEGREE DAYS
($p = 0.487$)

Year	Demand Y_t	Users X_t	CDD Z_t	Estimated Demand	Residuals e_t	X_{t-1}	Z_{t-1}	Y_{t-1}	$y_t = (Y_t - pY_{t-1})$	$x_t = (X_t - pX_{t-1})$	$z_t = (Z_t - pZ_{t-1})$
1	2	3	4	5	6	7	8	9	10	11	12
1980	75.93	23.68	16.41	74.90	1.03						
1981	75.57	23.77	15.91	75.75	-0.18	23.68	16.41	75.93	38.59	12.24	7.92
1982	75.21	23.89	14.41	75.85	-0.64	23.77	15.91	75.57	38.41	12.31	6.66
1983	81.19	24.02	17.16	81.41	-0.22	23.89	14.41	75.21	44.56	12.39	10.14
1984	80.83	24.20	15.33	82.08	-1.25	24.02	17.16	81.19	41.29	12.50	6.97
1985	84.00	24.35	16.51	86.00	-2.00	24.20	15.33	80.83	44.64	12.56	9.04
1986	84.75	24.47	15.23	86.38	-1.63	24.35	16.51	84.00	43.84	12.61	7.19
1987	90.07	24.61	15.34	88.81	1.26	24.47	15.23	84.75	48.80	12.69	7.92
1988	95.04	24.75	15.61	91.44	3.60	24.61	15.34	90.07	51.18	12.76	8.14
1989	95.36	24.89	15.26	93.30	2.06	24.75	15.61	95.04	49.08	12.84	7.66
1990	97.19	25.05	15.84	96.64	0.55	24.89	15.26	95.36	50.75	12.93	8.41
1991	102.39	25.16	17.72	100.79	1.60	25.05	15.84	97.19	55.06	12.96	10.01
1992	96.59	25.29	12.67	96.63	-0.04	25.16	17.72	102.39	46.73	13.04	4.04
1993	103.20	25.55	16.03	105.08	-1.88	25.29	12.67	96.59	56.16	13.23	9.86
1994	104.59	25.62	15.76	105.89	-1.30	25.55	16.03	03.20	54.33	13.18	7.95
1995	106.62	25.74	16.72	109.05	-2.43	25.62	15.76	104.59	55.68	13.26	9.04
1996	106.79	25.81	14.81	107.82	-1.03	25.74	16.72	106.62	54.87	13.27	6.67
1997	108.46	25.88	13.49	107.32	1.14	25.81	14.81	106.79	56.45	13.31	6.28
1998	113.01	26.06	16.35	113.83	-0.82	25.88	13.49	108.46	60.19	13.46	9.78
1999	120.05	26.26	16.97	117.88	2.17	26.06	16.35	113.01	65.01	13.57	9.01

The estimated coefficients in Equation (4.20) and the relevant statistics are shown in Table 4.2. The model residuals, $e_t = Y_t - \hat{Y}_t$, which are shown in the sixth column of Table 4.1, are plotted in Figures 4.4 and 4.5.

The estimated model is as follows:

$$\hat{Y}_t = -333.648 + 16.390X_t + 1.245Z_t \qquad \ldots (4.20)$$

TABLE 4.2
ELECTRIC DEMAND VERSUS CUSTOMERS AND WEATHER
(1980 – 1999)
(Summary Of Regression Output)

Regression Statistics	
Multiple R	0.993
R Square	0.985
Adjusted R Square	0.983
Standard Error	1.719
Observations	20

ANOVA

	Df	SS	MS	F	Significance F
Regression	2	3352.283	1676.14	566.917	2.7555E-16
Residual	17	50.262	2.96		
Total	19	3402.545			

	Coefficients	Stand. Error	T Stat	p-Value	Lower 95%	Upper 95%
Intercept	-333.648	13.443	-24.820	8.5573E-15	-362.010	-305.286
Customers	16.390	0.488	33.605	5.4694E-17	15.361	17.419
Weather	1.245	0.326	3.815	0.00138509	0.557	1.934

Durbin-Watson d = 1.002

In accordance with the hypothesis testing procedure outlined in Chapters 2 and 3, both F and t statistics indicate that the coefficients of the model are statistically significant. The adjusted R^2 is 98.3%, which is fairly high, indicating a good fit. What about the model residuals? Are they randomly distributed? Let's perform the two tests for first order autocorrelation as described earlier.

Residual Analysis: Test for First Order Autocorrelation

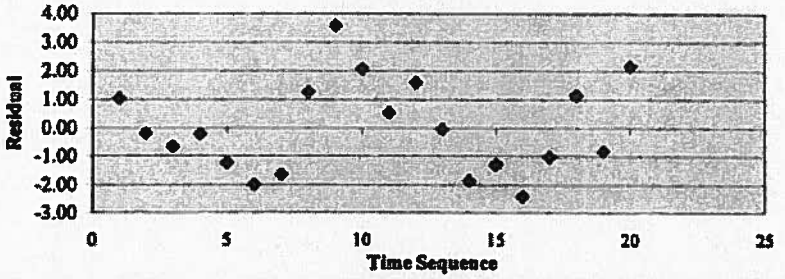
On the basis of F and t tests and R^2, the proposed linear model is a good fit to the data. However, as stated earlier, if autocorrelation is present in a model, these statistics can be misleading. Let's test the residuals for autocorrelation.

The Visual Test: The residuals of the model have been plotted in Figure 4.4. The figure is clearly a Pattern III type, indicating the presence of autocorrelation. The other way of looking at the residual pattern is to plot the current residual, e_t, against the residual of lag 1, e_{t-1}. This plot is shown in Figure 4.5. If the plot shows that there is a linear relationship between them, then it means that first order autocorrelation is present. It appears that a first order autoregressive model AR (1) like Equation (4.5a) can be fitted to the residuals listed in the sixth column of Table 4.1. The estimated AR (1) model is as follows:

$$\hat{e}_t = \underset{(2.252)}{0.4870}e_{t-1} \qquad \ldots (4.21)$$

Like $\hat{Y}$, $\hat{e}_t$ is an estimate of e_t according to the AR (1) scheme. The number in parenthesis underneath the AR (1) coefficient is its t-value.

FIGURE 4.4
RESIDUAL ANALYSIS

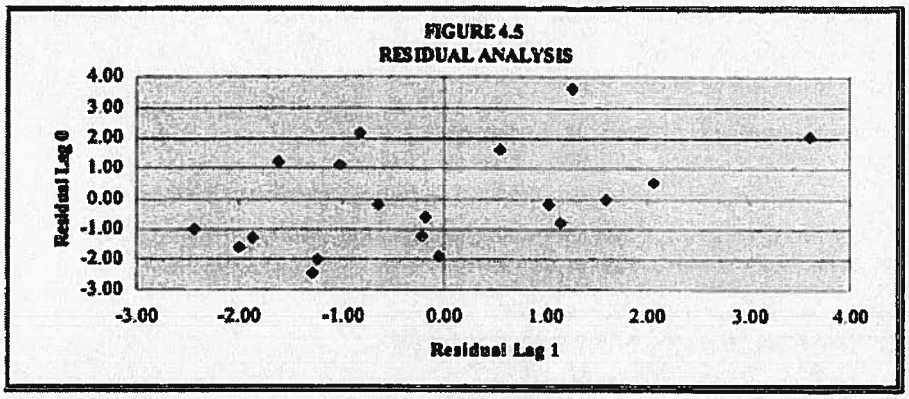


Note that Equation (4.21) does not have a constant term because the mean of the residuals is zero (assumption 1). The calculated t-value is for testing the null hypothesis for $\rho = 0$. (Note that ρ is unknown, and p is an estimate for ρ.) According to our rule of thumb, if $t > 2.0$, which is the case ($t = 2.252$), we reject the null hypothesis for $\rho = 0$ and conclude that ρ of the specified AR (1) process is significantly different from zero, meaning that there is a significant autocorrelation of the first order present in the residuals.

Durbin-Watson d Test: The d statistic is calculated according to Equation (4.4). The regression program in EXCEL does not calculate d automatically, but it can be easily calculated in the

spreadsheet environment. The calculated d is shown at the bottom of Table 4.2, which is, $d = 1.002$. Let's follow the d test procedure to perform the test.

1. Select the level of significance at 5%.
2. Go to Table A-5 in the Appendix of Tables. At $n = 20$ and $k' = 2$, we find $d_L = 1.100$ and $d_U = 1.537$. Note that k' is the number of explanatory variables in the model.
3. Test the null hypothesis for $\rho = 0$. Since $d = 1.002 < d_L = 1.100$, we reject the null hypothesis and conclude that autocorrelation is present. The d test confirms our suspicion from visual inspection of the residual plot that autocorrelation is present.



REMEDIAL PROCEDURES

We have described two remedial procedures for removing autocorrelation from a model: (1) the Cochran-Orcutt iterative procedure and (2) the Durbin two-stage method. Let's demonstrate below how to apply these two procedures to remove autocorrelation from the model.

Cochran-Orcutt Iterative Procedure

Computing software, like SAS and others, can automatically go through the iterative procedure to remove autocorrelation from a model. Although EXCEL does not have this feature, we can use it to demonstrate how the Cochran-Orcutt procedure works. The iterative procedure requires a "Stopping Rule." Let's stop iteration when p changes by less than 0.001.

Step 1. Run an OLS model with the original data and estimate a p value by fitting an AR (1) model to the residuals, which we have done. Equation (4.20) is the OLS model, and Equation (4.21) is the autoregressive, AR (1), model of the residuals. These two models are reproduced below for our demonstration.

$$\hat{Y}_t = -333.648 + 16.390X_t + 1.245Z_t$$

$\hat{\varepsilon}_t = 0.4870e_{t-1}$

The AR (1) model of the residuals provides us with an initial p = 0.487.

Step 2. (a) Use p = 0.487 to transform the data according to Equations (4.9a) through (4.9d). The original data and the transformed data are shown in Columns 10 through 12 of Table 4.1.

Note that the transformed constant term is not shown in the table because it is still a constant for all the sampling years. As noted earlier, without transformation, the constant term is "1" for all years; with transformation it becomes (1 – p) = 1 – 0.487 = 0.513. In addition, the transformation, as mentioned earlier, results in a loss of the first point of the sample. By regressing y_t on x_t and z_t, we obtain the following model,

$$\hat{y}_t = -176.654 + 16.838x_t + 1.209z_t \qquad \ldots (4.22)$$

In Equation (4.22), $b_1 = 16.838$, $b_2 = 1.209$, but the value of the constant, –176.654, is not an estimate of α. That is, $a \neq -176.654$, rather

$$a(1 - p) = a(1 - 0.487) = -176.654$$

$$a = \frac{-176.654}{1 - 0.487} = -344.355 \qquad \ldots (4.23)$$

(b) Replace the coefficients in Equation (4.20), which is the model based on the original data, with the newly estimated coefficients in Equations (4.22) and (4.23). Equation (4.20) becomes:

$$\hat{Y}_t = -344.355 + 16.838X_t + 1.209Z_t \qquad \ldots (4.24)$$

On the basis of Equation (4.24), we re-calculate the residual series as follows:

$$e_t = Y_t - \hat{Y}_t \qquad \ldots (4.25)$$

$$\hat{y}_t = -335.435C_t + 16.481x_t + 1.225z_t \quad \ldots (4.30)$$
$$(-15.877) \quad (19.618) \quad (5.536)$$

p	= 0.576
Adjusted R^2	= 0.960
F	= 154.008
Standard Error of Residuals	= 1.518
Number of Observations	= 20
Durbin-Watson d	= 1.766

The constant term C_t in Equation (4.30) takes on two values instead of the usual one. For year 1980, $C_t = 0.817$, and for all other years, $C_t = 0.424$ or $(1 - .576)$.

Inspecting the t statistics, which are shown underneath the estimated coefficients, and the Durbin-Watson d statistic, we can conclude that Equation (4.30) is a good fit to the data, and the model no longer has first order autocorrelation. There is no need to go through another iteration. Then, for forecasting, the model becomes:

$$\hat{Y}_t - 0.576Y_{t-1} = -335.435C_t + 16.481\,(X_t - 0.576X_{t-1}) + 1.225\,(Z_t - 0.576Z_{t-1})$$

Moving Y_{t-1} to the right hand side of the equation, we obtain:

$$\hat{Y}_t = -335.435C_t + 16.481\{X_t - 0.576X_{t-1}\} + 1.225(Z_t - 0.576Z_{t-1}) + 0.576Y_{t-1} \quad \ldots (4.31)$$

The coefficient, 0.576, for Y_{t-1} is the estimated p required for the first step of the Durbin method. Note that in the first step, the first data points of the lagged variables are not restored.

Stage 2: Set p = 0.576. Transform the data of all variables according to Equations (4.9a) through (4.9d) in the same manner as was done for the Cochran-Orcutt procedure. Then, restore the first data points of all variables including the constant term according to Equations (4.16a) through (4.16d). The transformed data for the Durbin procedure are given in Table 4.4. The restored first data points for y_1, the constant term, x_1 and z_1 are calculated as follows:

$y_1 = \sqrt{1-p^2}\,Y_1 = \sqrt{1-(0.576)^2} \times 75.93 = 62.03$

Consant $= \sqrt{1-p^2} = \sqrt{1-(0.576)^2} = 0.817$

$x_1 = \sqrt{1-p^2}\,X_1 = \sqrt{1-(0.576)^2} \times 23.68 = 19.35$

$z_1 = \sqrt{1-p^2}\,Z_1 = \sqrt{1-(0.576)^2} \times 16.41 = 13.41$

TABLE 4.4
DATA FOR DURBIN'S TWO-STAGE PROCEDURE
P = 0.576

Year	Demand Y_t	Customer X_t	Weather Z_t	$y_t = Y_t - pY_{t-1}$	Constant	$x_t = X_t - pY_{t-1}$	$z_t = Z_t - Z_{t-1}$
1980	75.93	23.68	16.41	62.03	0.817	19.35	13.41
1981	75.57	23.77	15.91	31.83	0.424	10.13	6.46
1982	75.21	23.89	14.41	31.68	0.424	10.20	5.25
1983	81.19	24.02	17.16	37.87	0.424	10.26	8.86
1984	80.83	24.20	15.33	34.06	0.424	10.36	5.45
1985	84.00	24.35	16.51	37.44	0.424	10.41	7.68
1986	84.75	24.47	15.23	36.37	0.424	10.44	5.72
1987	90.07	24.61	15.34	41.25	0.424	10.52	6.57
1988	95.04	24.75	15.61	43.16	0.424	10.57	6.77
1989	95.36	24.89	15.26	40.62	0.424	10.63	6.27
1990	97.19	25.05	15.84	42.26	0.424	10.71	7.05
1991	102.39	25.16	17.72	46.41	0.424	10.73	8.60
1992	96.59	25.29	12.67	37.61	0.424	10.80	2.46
1993	103.20	25.55	16.03	47.56	0.424	10.98	8.73
1994	104.59	25.62	15.76	45.15	0.424	10.90	6.53
1995	106.62	25.74	16.72	46.38	0.424	10.98	7.64
1996	106.79	25.81	14.81	45.38	0.424	10.98	5.18
1997	108.46	25.88	13.49	46.95	0.424	11.01	4.96
1998	113.01	26.06	16.35	50.54	0.424	11.15	8.58
1999	120.05	26.26	16.97	54.96	0.424	11.25	7.55

Note that the first data point in the Constant column of Table 4.4 is different from all other points, which are constant. Care should be exercised when computing software is used for modeling. The constant term in this case should be treated as a variable, and the software should be instructed not to estimate another coefficient for the constant term. The estimated model based on the transformed data is as follows:

$$Y_t = -172.393 + 9.808X_t - 2.524X_{t-1} + 1.377Z_t - 0.627Z_{t-1} + 0.576Y_{t-1} \quad \ldots(4.29)$$

TABLE 4.3
DATA FOR COCHRAN-ORCUTT MODELING
(p = 0.4951)

Year	Demand Y_t	Customers X_t	CDD Z_t	$y_t = (Y_t - pY_{t-1})$	$x_t = (X_t - pX_{t-1})$	$z_t = (Z_t - pZ_{t-1})$
1980	75.93	23.68	16.41			
1981	75.57	23.77	15.91	37.98	12.05	7.79
1982	75.21	23.89	14.41	37.80	12.12	6.53
1983	81.19	24.02	17.16	43.95	12.19	10.03
1984	80.83	24.20	15.33	40.63	12.31	6.83
1985	84.00	24.35	16.51	43.98	12.37	8.92
1986	84.75	24.47	15.23	43.16	12.41	7.06
1987	90.07	24.61	15.34	48.11	12.50	7.80
1988	95.04	24.75	15.61	50.45	12.57	8.02
1989	95.36	24.89	15.26	48.31	12.64	7.53
1990	97.19	25.05	15.84	49.98	12.73	8.29
1991	102.39	25.16	17.72	54.27	12.76	9.88
1992	96.59	25.29	12.67	45.90	12.83	3.90
1993	103.20	25.55	16.03	55.38	13.03	9.76
1994	104.59	25.62	15.76	53.50	12.97	7.82
1995	106.62	25.74	16.72	54.84	13.06	8.92
1996	106.79	25.81	14.81	54.00	13.07	6.53
1997	108.46	25.88	13.49	55.59	13.10	6.16
1998	113.01	26.06	16.35	59.31	13.25	9.67
1999	120.05	26.26	16.97	64.10	13.36	8.886

(c) Go to step 1 to estimate a new p by regressing e_t on e_{t-1}.

For Step 1 in the second round of the procedure, the estimated AR (1) equation is as follows:

$$\hat{e}_t = 0.4951 e_{t-1} \qquad \ldots (4.26)$$

The estimated p for the second round of iteration is 0.4951. The new p is greater than the last estimate by 0.0081 (0.4951 – 0.4870), which is larger than our stopping rule value of 0.001. For that reason, the iteration continues. The transformed variables on the basis of p = 0.4951 are given

in the last three columns of Table 4.3. By running a regression on the newly transformed data, we obtain:

$$\hat{y}_t = -173.980 + 16.838 x_t + 1.209 z_t \qquad \ldots (4.27)$$
$$(-14.85) \quad (18.21) \quad (5.13)$$

Adjusted R^2	= 0.955
F	= 190.154
Standard Error of Residuals	= 1.537
Number of Observations	= 19
Durbin-Watson d	= 1.664

After the second full iteration, the third p = 0.4957, which is only 0.0006 larger than the last p, and the Durbin-Watson d statistic is 1.664, greater than d_U = 1.537 and less than 2, indicating no autocorrelation in the residuals. We stop the iteration. In addition, all other modeling statistics indicate that Equation (4.27) is a good fit of the data. Equation (4.27) can be re-written in its differenced form for forecasting as follows:

$$\hat{Y}_t - 0.4951 Y_{t-1} = -173.980 + 16.838 (X_t - 0.4951 X_{t-1}) + 1.209 (Z_t - 0.4951 Z_{t-1})$$

By re-arranging terms, the forecasting model becomes:

$$\hat{Y}_t = -173.980 + 16.838 (X_t - 0.4951 X_{t-1}) + 1.209 (Z_t - 0.4951 Z_{t-1}) + 0.4951 Y_{t-1} \qquad \ldots (4.28)$$

Had we used a more sophisticated computing software than EXCEL, all the steps would have been done automatically.

Durbin's Two-Stage Procedure with Prais-Winsten Transformation

As shown in the last three columns of Table 4.3, the generalized differencing of the data results in a loss of one data point for each variable. For a small sample size, e.g., n < 20, it can mean a loss of model efficiency. According to Prais-Winsten (PW), the lost data points may be restored on the basis of Equations (4.16a) through (4.16d). Studies have found that, for small samples, incorporating the PW transformation into Durbin's two-stage method for removing autocorrelation tends to improve model efficiency. For large samples, however, PW transformation is unlikely to make much difference. We will use the same modeling data, (i.e., data given in Table 4.1), to demonstrate how Durbin's two-stage method with PW transformation of the first data points works in removing autocorrelation.

Stage 1: Estimate p by applying the OLS to Equation (4.15), which is considered a linear equation as discussed earlier. The required data for each variable and its lagged values are all given in Table 4.1. Regress Y_t on X_t, X_{t-1}, Z_t, Z_{t-1}, and Y_{t-1}, the estimated equation comes to:

Equation (4.20) is the OLS model based on the original data without considering the autocorrelation problem. Equation (4.28) reflects the treatment of autocorrelation in accordance with the Cochran-Orcutt procedure, whereas Equation (4.31) was estimated by applying the Durbin two-stage method with PW transformation of the first data points.

EVALUATION OF MODEL EFFICIENCY

The model evaluation statistics of the three modeling procedures are listed in Table 4.5. The values of R^2 for the Cochran-Orcutt and the Durbin procedures are lower than the R^2 of the OLS model. This change should not be taken as a decrease in the goodness of fit. Differencing of data usually causes R^2 to go down, and here R^2 measures the goodness of fit of a model fitted to the differenced data and not to the original data. The standard errors of the models are the true indicators of model efficiency.

TABLE 4.5
MODEL EFFICIENCY COMPARISON

Estimation Method	Adjusted R^2	Residual Std Error	Std Error X-Coef.	Std Error Z-Coef
OLS: Eq. (4.20)	0.983	1.719	0.488	0.326
Cochran-Orcutt: Eq. (4.28)	0.955	1.537	0.925	0.235
Durbin with PW Eq. (4.31)	0.902	1.518	0.840	0.221

The standard errors of residuals of the two modified models are lower than that of the OLS model indicating an improvement in the fit. The standard error of the estimated coefficient for X in the OLS model is lower than those of the other two models. This should not be a surprise. As stated earlier, when autocorrelation is present in a model, we might under-estimate the standard errors of the OLS coefficients, which is the case here. Comparing the Cochran-Orcutt model with

the Durbin two-stage model, we conclude that the Durbin procedure with PW transformation of the first data point produced a better model for this example.

REMARKS ON DEALING WITH AUTOCORRELATION

We have demonstrated how to test the model for autocorrelation and how to remove it when it is present. Here are a few tips for dealing with autocorrelation problems.

1. We have stated that the first order autocorrelation is an AR (1) process. For the AR (1) to be stable, we require that $|\rho| < 1$. If the estimated $|\rho| \geq 1$, we should not use either of the two remedial procedures discussed above; instead, we should look for a trend variable to rectify the problem of autocorrelation.

2. We have stated that the presence of autocorrelation increases the variances of the residuals and the estimated coefficients, and the OLS procedure will produce biased estimates of these variances. The larger the autocorrelation coefficient, the larger the bias will be. It is important that modelers test their models for autocorrelation and take appropriate actions if autocorrelation is present.

3. When the sample size is small, say $n < 20$, and the estimated ρ is less than 0.3 in absolute terms, any attempt to remove autocorrelation may make the model less efficient.

QUESTIONS FOR REVIEW

1. Define the following terms:

   a. Autocorrelation
   b. Autocorrelation coefficient
   c. Generalized differencing
   d. Spatial autocorrelation

2. What are the causes of autocorrelation in a model based on time series data? Explain.

3. Can autocorrelation occur in cross-sectional models? Explain.

4. What are the consequences if autocorrelation is present in a model?

5. Describe the visual test procedure for detecting autocorrelation.

6. Describe the Durbin-Watson test procedure for autocorrelation.

7. Describe the Cochran-Orcutt estimation procedure for handling autocorrelation problems.

8. Describe the Prais-Winsten transformation for restoring the lost data points. Under what circumstances the PW transformation may improve model efficiency?

9. Describe the Durbin two-stage estimation procedure with Prais-Winsten transformation for handling the autocorrelation problem.