9.1 What Is a Time Series?

In many business and economic studies, the response variable y is measured sequentially in time. For example, we might record the number y of new housing starts for each month in a particular region. This collection of data is called a **time series**. Other examples of time series are data collected on the quarterly number of highway deaths in the United States, the annual sales for a corporation, and the recorded month-end values of the prime interest rate.

> **Definition 9.1**
>
> A **time series** is a collection of data obtained by observing a response variable at periodic points in time.

> **Definition 9.2**
>
> If repeated observations on a variable produce a time series, the variable is called a **time series variable**. We use y_t to denote the value of the variable at time t.

If you were to develop a model relating the number of new housing starts to the prime interest rate over time, the model would be called a **time series model**, because both the dependent variable, new housing starts, and the independent variable, prime interest rate, are measured sequentially over time. Furthermore, time itself would probably play an important role in such a model, because the economic trends and seasonal cycles associated with different points in time would almost certainly affect both time series.

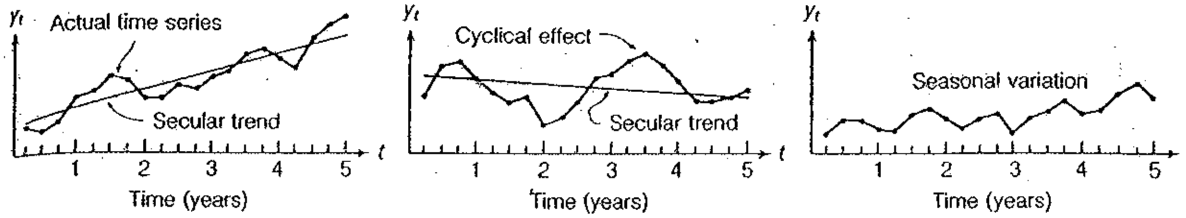
The construction of time series models is an important aspect of business and economic analyses, because many of the variables of most interest to business and economic researchers are time series. This chapter is an introduction to the very complex and voluminous body of material concerned with time series modeling and forecasting future values of a time series.

9.2 Time Series Components

Researchers often approach the problem of describing the nature of a time series y_t by identifying four kinds of change, or variation, in the time series values. These four components are commonly known as (1) secular trend, (2) cyclical effect, (3) seasonal variation, and (4) residual effect. The components of a time series are most easily identified and explained pictorially.

Figure 9.1a (page 496) shows a **secular trend** in the time series values. The secular component describes the tendency of the value of the variable to increase or decrease over a long period of time. Thus, this type of change or variation is also known as the **long-term trend**. In Figure 9.1a, the long-term trend is of an increasing nature. However, this does not imply that the time series has always

FIGURE 9.1 Illustrating the Components of a Time Series



(a) (b) (c)

The cyclical effect in a time series, as shown in Figure 9.1(b), generally describes the fluctuation about the secular trend that is attributable to business and economic conditions at the time. These fluctuations are sometimes called business cycles. During a period of general economic expansion, the business cycle lies above the secular trend, while during a recession, when business activity is likely to slump, the cycle lies below the secular trend. You can see that the cyclical variation does not follow any definite trend, but moves rather unpredictably.

The seasonal variation in a time series describes the fluctuations that recur during specific portions of each year (e.g., monthly or seasonally). In Figure 9.1(c), you can see that the pattern of change in the time series within a year tends to be repeated from year to year, producing a wavelike or oscillating curve.

The final component, the residual effect, is what remains after the secular, cyclical, and seasonal components have been removed. This component is not systematic and may be attributed to unpredictable influences such as wars, hurricanes, presidential assassination, and randomness of human actions. Thus, the residual effect represents the random error component of a time series.

In many practical applications of time series to business, the objective is to *forecast* (predict) some *future value or values* of the series. To obtain forecasts, some type of model that can be projected into the future must be used to describe the time series. One of the most widely used models is the **additive model***

$$y_t = T_t + C_t + S_t + R_t$$

where T_t, C_t, S_t, and R_t represent the secular trend, cyclical effect, seasonal variation, and residual effect, respectively, of the time series variable y_t. Various methods exist for estimating the components of the model and forecasting the time series. These range from simple descriptive techniques, which rely on smoothing the pattern of the time series, to complex inferential models which combine regression analysis with specialized time series models. Several descriptive forecasting techniques are presented in optional Section 9.3, and forecasting using the general linear regression model of Chapter 4 is discussed in Section 9.4. The remainder of the chapter is devoted to the more complex and more powerful time series models.

*Another useful model is the multiplicative model $y_t = T_t C_t S_t R_t$. Note that this model can be written in the form of an additive model by taking natural logarithms:

$$\ln y_t = \ln T_t + \ln C_t + \ln S_t + \ln R_t$$

QUARTER	S&P 500
I	151.07
II	165.11
III	168.66
IV	164.12
I	159.18
II	153.18
III	166.10
IV	167.24
I	180.66
II	191.85
III	182.08
IV	211.28
I	232.33
II	245.30
III	238.27
IV	248.61

or Corporation.

ıal components to
ıts $w = .3$, $v = .8$,

›rices are given in

PRICE OF GOLD
374.18
449.03
360.29
317.30
367.87
408.91

rk: Standard &
Commerce.

es. Plot the gold
ı detect the long-

sing a smoothing

ld in 1988.
e price of gold in

SECTION 9.4

FORECASTING: THE REGRESSION APPROACH

Many firms use past sales to forecast future sales. Suppose a wholesale distributor of sporting goods is interested in forecasting its sales revenue for each of the next 5 years. Since an inaccurate forecast may have dire consequences to the distributor, some measure of the forecast's reliability is required. To make such forecasts and assess their reliability, an inferential time series forecasting model must be constructed. The familiar general linear regression model of Chapter 4 represents one type of inferential model since it allows us to calculate prediction intervals for the forecasts.

To illustrate the technique of forecasting with regression, consider the data in Table 9.5. The data are annual sales (in thousands of dollars) for a firm (say, the sporting goods distributor) in each of its 35 years of operation. A plot of the data (Figure 9.5) reveals a linearly increasing trend, so the first-order (straight-line) model

$$E(y_t) = \beta_0 + \beta_1 t$$

TABLE 9.5 A Firm's Yearly Sales Revenue (thousands of dollars)

t	y_t	t	y_t	t	y_t
1	4.8	13	48.4	25	100.3
2	4.0	14	61.6	26	111.7
3	5.5	15	65.6	27	108.2
4	15.6	16	71.4	28	115.5
5	23.1	17	83.4	29	119.2
6	23.3	18	93.6	30	125.2
7	31.4	19	94.2	31	136.3
8	46.0	20	85.4	32	146.8
9	46.1	21	86.2	33	146.1
10	41.9	22	89.9	34	151.4
11	45.5	23	89.2	35	150.9
12	53.5	24	99.1		

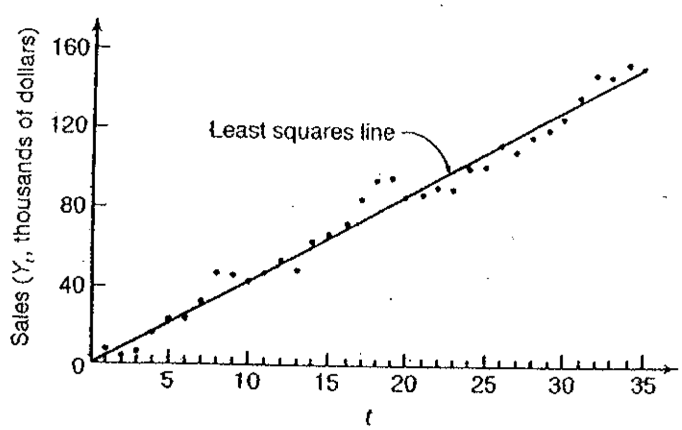


FIGURE 9.5 Plot of Sales Data

seems plausible for describing the secular trend. The SAS computer printout for the model is shown in Figure 9.6. Note that the model apparently provides an excellent fit to the data, with $R^2 = .98$, $F = 1{,}615.72$ (p-value $< .0001$), and $s = 6.39$. The least squares prediction equation, whose coefficients are shaded in Figure 9.6, is

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 t = .4015 + 4.2956t$$

We can obtain sales forecasts and corresponding 95% prediction intervals for years 36–40 by employing the formulas given in Section 3.9. However, these values are given in the bottom portion of the SAS printout shown in Figure 9.6. For example, for $t = 36$, we have $\hat{y}_{36} = 155.0$ with the 95% prediction interval (141.3, 168.8). That is, we predict that sales revenue in year $t = 36$ will fall between \$141,300 and \$168,800 with 95% confidence.

Note that the prediction intervals for $t = 36, 37, \ldots, 40$ widen as we attempt to forecast further into the future. Intuitively, we know that the farther into the future we forecast, the less certain we are of the accuracy of the forecast since some unexpected change in business and economic conditions may make the model inappropriate. Since we have less confidence in the forecast for, say, $t = 40$ than for $t = 36$, it follows that the prediction interval for $t = 40$ must be wider in order to attain a 95% level of confidence. For this reason, time series forecasting (regardless of the forecasting method) is generally confined to the short term.

Multiple regression models can also be used to forecast future values of a time series with seasonal variation. We illustrate with an example.

EXAMPLE 9.2

Refer to the 1985–1988 quarterly power loads listed in Table 9.1.

a. Propose a model for quarterly power load, y_t, that will account for both the secular trend and seasonal variation present in the series.

b. Fit the model to the data, and use the least squares prediction equation to forecast the utility company's quarterly power loads in 1989. Construct 95% prediction intervals for the forecasts.

SOLUTION

a. A common way to describe seasonal differences in a time series is with dummy variables.* For quarterly data, a model that includes both trend and seasonal components is

$$E(y_t) = \beta_0 + \underbrace{\beta_1 t}_{\substack{\text{Secular}\\ \text{trend}}} + \underbrace{\beta_2 Q_1 + \beta_3 Q_2 + \beta_4 Q_3}_{\text{Seasonal component}}$$

*Another way to account for seasonal variation is with trigonometric (sine and cosine) terms. We discuss seasonal models with trigonometric terms in Section 9.6.

CHOOSING THE DETERMINISTIC COMPONENT

The deterministic portion of the model is chosen in exactly the same manner as for the regression models of the preceding chapters except that some of the independent variables might be time series variables or might be trigonometric functions of time (such as sin t or cos t). It is helpful to think of the deterministic component as consisting of the trend (T_t), cyclical (C_t), and seasonal (S_t) effects described in Section 9.2.

For example, we may want to model the number of new housing starts, y_t, as a function of the prime interest rate, x_t. Then, one model for the mean of y_t is

$$E(y_t) = \beta_0 + \beta_1 x_t$$

for which the mean number of new housing starts is a multiple β_1 of the prime interest rate, plus a constant β_0. Another possibility is a second-order relationship.

$$E(y_t) = \beta_0 + \beta_1 x_t + \beta_2 x_t^2$$

which permits the *rate* of increase in the mean number of housing starts to increase or decrease with the prime interest rate.

Yet another possibility is to model the mean number of new housing starts as a function of both the prime interest rate and the year, t. Thus, the model

$$E(y_t) = \beta_0 + \beta_1 x_t + \beta_2 t + \beta_3 x_t t$$

implies that the mean number of housing starts increases linearly in x_t, the prime interest rate, but the rate of increase depends on the year t. If we wanted to adjust for seasonal (cyclical) effects due to t, we might introduce time into the model using trigonometric functions of t. This topic will be explained in greater detail below.

Another important type of model for $E(y_t)$ is the lagged independent variable model. *Lagging* means that we are pairing observations on a dependent variable and independent variable at two different points in time, with the time corresponding to the independent variable lagging behind the time for the dependent variable. Suppose, for example, we believe that the monthly mean number of new housing starts is a function of the *previous* month's prime interest rate. Thus, we model y_t as a linear function of the lagged independent variable, prime interest rate, x_{t-1}:

$$E(y_t) = \beta_0 + \beta_1 x_{t-1}$$

or, alternatively, as the second-order function,

$$E(y_t) = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-1}^2$$

For this example, the independent variable, prime interest rate x_t, is lagged 1 month behind the response y_t.

Many time series have distinct seasonal patterns. Retail sales are usually highest around Christmas, spring, and fall, with relative lulls in the winter and summer periods. Energy usage is highest in summer and winter, and lowest in in spring and fall. Teenage unemployment rises in the summer months when schools are

FIGURE 9.9 A Seasonal Time Series Model

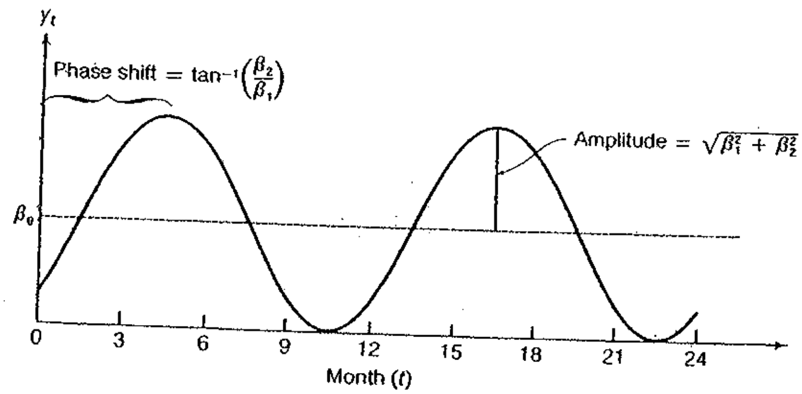
not in session, and falls near Christmas when many businesses hire part-time help.

When a time series' seasonality is exhibited in a relatively consistent pattern from year to year, we can model the pattern using trigonometric terms in the model for $E(y_t)$. For example, the model of a monthly series with mean $E(y_t)$ might be

$$E(y_t) = \beta_0 + \beta_1\left(\cos \frac{2\pi}{12}t\right) + \beta_2\left(\sin\frac{2\pi}{12}t\right)$$

This model would appear as shown in Figure 9.9. Note that the model is cyclic, with a period of 12 months. That is, the mean $E(y_t)$ completes a cycle every 12 months and then repeats the same cycle over the next 12 months. Thus, the expected peaks and valleys of the series remain the same from year to year. The coefficients β_1 and β_2 determine the amplitude and phase shift of the model. The amplitude is the magnitude of the seasonal effect, while the phase shift locates the peaks and valleys in time. For example, if we assume month 1 is January, the mean of the time series depicted in Figure 9.9 has a peak each April and a valley each October.

FIGURE 9.9 A Seasonal Time Series Model



If the data are monthly or quarterly, we can treat the season as a qualitative independent variable (see Example 9.2), and write the model

$$E(y_t) = \beta_0 + \beta_1 S_1 + \beta_2 S_2 + \beta_3 S_3$$

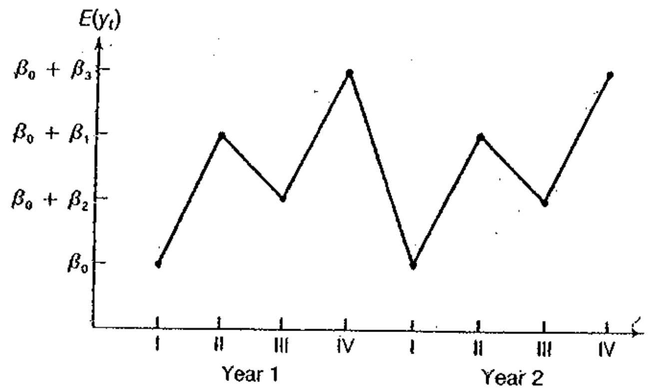
where

$$S_1 = \begin{cases} 1 & \text{if season is spring (II)} \\ 0 & \text{otherwise} \end{cases} \qquad S_2 = \begin{cases} 1 & \text{if season is summer (III)} \\ 0 & \text{otherwise} \end{cases}$$

$$S_3 = \begin{cases} 1 & \text{if season is fall (IV)} \\ 0 & \text{otherwise} \end{cases}$$

Thus, S_1, S_2, and S_3 are dummy variables that describe the four levels of season, letting winter (I) be the base level. The β coefficients determine the mean value of y_t for each season, as shown in Figure 9.10. Note that for the dummy variable model and the trigonometric model we assume the seasonal effects are approximately the same from year to year. If they tend to increase or decrease with time, an interaction of the seasonal effect with time may be necessary. (An example will be given in Section 9.9.)

FIGURE 9.10 Seasonal Model for Quarterly Data Using Dummy Variables



The appropriate form of the deterministic time series model will depend on both theory and data. Economic theory often provides several plausible models relating the mean response to one or more independent variables. The data can then be used to determine which, if any, of the models is best supported. The process is often an iterative one, beginning with preliminary models based on theoretical notions, using data to refine and modify these notions, collecting additional data to test the modified theories, and so forth.

CHOOSING THE RESIDUAL COMPONENT

The appropriate form of the residual component R_t will depend on the pattern of autocorrelation in the residuals (see Section 9.5). The autoregressive model is very useful for this aspect of time series modeling, with the general form

$$R_t = \phi_1 R_{t-1} + \phi_2 R_{t-2} + \cdots + \phi_p R_{t-p} + \varepsilon_t$$

which is called an autoregressive model of order p. Recall that the name *autoregressive* comes from the fact that R_t is regressed on its own past values. As the order p is increased, more complex autocorrelation functions can be modeled. There are several other types of models that can be used for the random component, but the autoregressive model is very flexible and receives more application in business forecasting than the others.

The simplest autoregressive model is the first-order autoregressive model

$$R_t = \phi R_{t-1} + \varepsilon_t$$

Recall that the autocorrelation between residuals at two different points in time diminishes as the distance between the time points increases. Since many business and economic time series exhibit this property, the first-order autoregressive model is a popular choice for the residual component.

EXERCISES 9.2

Thus, we use the pair of models

$$y_t = \beta_0 + \beta_1 t + R_t$$
$$R_t = \phi R_{t-1} + \varepsilon_t$$

to describe the yearly sales of the firm. In order to estimate the parameters of the time series model (β_0, β_1, and ϕ), a modification of the least squares method is required. To do this, we use a *transformation* that is much like the variance-stabilizing transformations discussed in Chapter 6.

First, we multiply the model

$$y_t = \beta_0 + \beta_1 t + R_t \tag{1}$$

by ϕ at time $(t - 1)$ to obtain

$$\phi y_{t-1} = \phi\beta_0 + \phi\beta_1(t - 1) + \phi R_{t-1} \tag{2}$$

Taking the difference between equations (1) and (2), we have

$$y_t - \phi y_{t-1} = \beta_0(1 - \phi) + \beta_1[t - \phi(t - 1)] + (R_t - \phi R_{t-1})$$

or, since $R_t = \phi R_{t-1} + \varepsilon_t$, then

$$y_t^* = \beta_0^* + \beta_1 t^* + \varepsilon_t$$

where $y_t^* = y_t - \phi y_{t-1}$, $t^* = t - \phi(t - 1)$ and $\beta_0^* = \beta_0(1 - \phi)$. Thus, we can use the transformed dependent variable y_t^* and transformed independent variable t^* and obtain least squares estimates of β_0^* and β_1. The residual ε_t is uncorrelated, so that the assumptions necessary for the least squares estimators are all satisfied. The estimator of the original intercept, β_0, can be calculated by

$$\hat{\beta}_0 = \frac{\hat{\beta}_0^*}{1 - \phi}$$

This transformed model appears to solve the problem of first-order autoregressive residuals. However, making the transformation requires knowing the value of the parameter ϕ. Also, we lose the initial observation, since the values of y_t^* and t^* can be calculated only for $t \geq 2$. The methods for estimating ϕ and adjustments for the values at $t = 1$ will not be detailed here. Instead, we will present output from the SAS computer package, which both performs the transformation and estimates the model parameters, β_0, β_1, and ϕ.

The SAS printout of the straight-line, autoregressive time series model fit to the sales data is shown in Figure 9.13 (page 528). Note that the format of the time series printout is different from that of the standard regression printout. The estimates of β_0 and β_1 in the deterministic component $E(y_t)$ appear at the bottom of the printout under the column heading B VALUE. The estimate of the first-order autoregressive parameter ϕ is given in the middle portion of the printout titled ESTIMATES OF THE AUTOREGRESSIVE PARAMETERS beneath the column heading COEFFICIENT.

FIGURE 9.13 SAS Computer Printout for the Combined Straight-Line Autoregressive Residual Fit to the Sales Data

```
                          A U T O R E G   P R O C E D U R E


                      ORDINARY LEAST SQUARES ESTIMATES

            SSE              1345.454     DFE                    33
            MSE              40.77132     ROOT MSE         6.385242
            SBC              234.1562     AIC              231.0455
            REG RSQ            0.9800     TOTAL RSQ          0.9800
            DURBIN-WATSON      0.8207

     VARIABLE DF          B VALUE         STD ERROR    T RATIO APPROX PROB

     INTERCPT  1       0.40151261        2.20570829      0.182      0.8567
     T         1       4.29563025        0.10686692     40.196      0.0001

                       ESTIMATES OF AUTOCORRELATIONS

LAG   COVARIANCE   CORRELATION   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
 0      38.4415       1.000000    |                   |********************|
 1      22.6661       0.589624    |                   |************        |

                      PRELIMINARY MSE=       25.07708

                ESTIMATES OF THE AUTOREGRESSIVE PARAMETERS
            [illegible]           STD ERROR           T RATIO
            [illegible]          0.14277861         -4.129639


                          YULE-WALKER ESTIMATES

            SSE               877.6854     DFE                   32
            MSE               27.42767     ROOT MSE        5.237143
            SBC               223.1868     AIC             218.5208
            REG RSQ             0.9412     TOTAL RSQ    [illegible]

     VARIABLE DF    B VALUE            STD ERROR    T RATIO APPROX PROB

     INTERCPT  1    [illegible]       3.99697517      0.102      0.9198
     T         1    [illegible]       0.18983105     22.630      0.0001
```

The interpretations of two quantities shown on the SAS printout for the time series procedure differ from those described in the preceding sections. First, the quantity printed as REG RSQ (in the lower portion of the printout) is not the value of R^2 based on the values of y_t. Instead, it is based on the values of the transformed variable, y_t^*. When we refer to R^2 in this chapter, we will always mean the value R^2 based on the original time series variable y_t. This value, which will usually be larger than the REG RSQ value, is given on the printout as TOTAL RSQ (shaded). Second, the SAS time series model is defined so that ϕ takes the *opposite* sign from the value contained in our model. Consequently, you must multiply the estimate of ϕ shown in the printout by (-1) to obtain the estimate of ϕ for our model.

TABLE 9.6

Therefore, the fitted models are

$$\hat{y}_t = .4058 + 4.2959t + \hat{R}_t$$
$$\hat{R}_t = .5896\hat{R}_{t-1}$$

with

$$\text{MSE} = 27.43$$

and

$$R^2 = .9869 \quad \text{(TOTAL RSQ on the printout, Figure 9.13)}$$

A comparison of the least squares (Figure 9.6) and autoregressive (Figure 9.13) computer printouts is given in Table 9.6. Note that the autoregressive model reduces MSE and increases R^2. The values of the estimators β_0 and β_1 change very little, but the estimated standard errors are considerably increased, thereby decreasing the t-value for testing H_0: $\beta_1 = 0$. Note that the implication that the linear relationship between sales y_t and year t is of significant predictive value is the same using either method. However, you can see that the underestimation of standard errors by using least squares in the presence of residual autocorrelation could result in the inclusion of unimportant independent variables in the model, since the t-values will usually be inflated.

TABLE 9.6

	LEAST SQUARES	AUTOREGRESSIVE
R^2	.98	.99
MSE	40.77	27.43
$\hat{\beta}_0$	.4015	.4058
$\hat{\beta}_1$	4.2956	4.2959
Standard error ($\hat{\beta}_0$)	2.2057	3.9970
Standard error ($\hat{\beta}_1$)	.1069	.1898
t-statistic for H_0: $\beta_1 = 0$	40.20 ($p < .0001$)	22.63 ($p < .0001$)
$\hat{\phi}$	—	.5896
t-statistic for H_0: $\phi = 0$	—	4.13

The estimated value of ϕ is .5896, and an approximate t-test* of the hypothesis H_0: $\phi = 0$ yields a t-value of 4.13. With 32 df, this value is significant at less than $\alpha = .01$. Thus, the result of the Durbin–Watson d-test is confirmed: There is adequate evidence of positive residual autocorrelation.† Furthermore, the first-order autoregressive model appears to describe this residual correlation well.

*An explanation of this t-test has been omitted. Consult the references at the end of the chapter for details of this test.

†This result is to be expected since it can be shown (proof omitted) that $\hat{\phi} \approx 1 - d/2$, where d is the value of the Durbin–Watson statistic.

The steps for fitting a time series model to a set of data are summarized in the box. Once the model is estimated, the model can be used to forecast future values of the time series y_t.

STEPS FOR FITTING TIME SERIES MODELS

1. Use the least squares approach to obtain initial estimates of the β parameters. Do *not* use the t- or F-tests to assess the importance of the parameters, since the estimates of their standard errors may be biased (often underestimated).
2. Analyze the residuals to determine whether they are autocorrelated. The Durbin–Watson test is one technique for making this determination.
3. If there is evidence of autocorrelation, construct a model for the residuals. The autoregressive model is one useful model. Consult the references at the end of the chapter for more types of residual models and for methods of identifying the most suitable model.
4. Reestimate the β parameters, taking the residual model into account. This involves a simple transformation if an autoregressive model is used, and several statistical packages have computer routines to accomplish this.

EXERCISES 9.25–9.29

9.25 The Gross National Product (GNP) is a measure of total U.S. output, and is therefore an important indicator of the U.S. economy. The quarterly GNP values (in billions of dollars) from 1975–1985 are given in the table. Let y_t be the GNP in quarter t, $t = 1, 2, 3, \ldots, 44$.

YEAR	QUARTER 1	2	3	4
1975	1,479.8	1,516.7	1,578.5	1,621.8
1976	1,672.0	1,698.6	1,729.0	1,772.5
1977	1,834.8	1,895.1	1,954.4	1,988.9
1978	2,031.7	2,139.5	2,202.5	2,281.6
1979	2,335.5	2,377.9	2,454.8	2,502.9
1980	2,572.9	2,578.8	2,639.1	2,736.0
1981	2,866.6	2,912.5	3,004.9	3,032.2
1982	3,021.4	3,070.2	3,090.7	3,109.6
1983	3,171.5	3,272.0	3,362.2	3,432.0
1984	3,676.5	3,757.5	3,812.2	3,852.5
1985	3,909.3	3,965.0	4,030.5	4,087.7

Source: *Standard & Poor's Trade and Securities Statistics* (Annual). New York: Standard & Poor's Corporation.

Combining these, we obtain

$$y_{n+1} = \beta_0 + \beta_1 x_{n+1} + \overbrace{\phi R_n + \varepsilon_{n+1}}^{R_{n+1}}$$

The forecast of y_{n+1} is obtained by estimating each of the unknown quantities in this equation:*

$$\hat{y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1} + \hat{\phi}\hat{R}_n$$

where $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\phi}$ are the estimates based on the time series model-fitting approach presented in Section 9.7 and ε_{n+1} is estimated by its expected value 0. The estimate $\hat{R}_n$ of the residual R_n is obtained by noting that

$$R_n = y_n - (\beta_0 + \beta_1 x_n)$$

so that

$$\hat{R}_n = y_n - (\hat{\beta}_0 + \hat{\beta}_1 x_n)$$

The two-step-ahead forecast of y_{n+2} is similarly obtained. The true value of y_{n+2} is

$$\begin{aligned} y_{n+2} &= \beta_0 + \beta_1 x_{n+2} + R_{n+2} \\ &= \beta_0 + \beta_1 x_{n+2} + \phi R_{n+1} + \varepsilon_{n+2} \end{aligned}$$

and the forecast at $t = n + 2$ is

$$\hat{y}_{n+2} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+2} + \hat{\phi}\hat{R}_{n+1}$$

The residual R_{n+1} (and all future residuals) can now be obtained from the recursive relation

$$R_{n+1} = \phi R_n + \varepsilon_{n+1}$$

so that

$$\hat{R}_{n+1} = \hat{\phi}\hat{R}_n$$

Thus, the forecasting of future y-values is an iterative process, with each new forecast making use of the previous residual to obtain the estimated residual for the future time period. The general forecasting procedure using time series models with first-order autoregressive residuals is outlined in the next box.

EXAMPLE 9.3

Suppose we want to forecast the sales of the company for the data analyzed in Section 9.7. Recall that we fit the regression–autoregression pair of models

$$y_t = \beta_0 + \beta_1 t + R_t \qquad R_t = \phi R_{t-1} + \varepsilon_t$$

Using 35 years of sales data, we obtained the estimated models

$$\hat{y}_t = .4058 + 4.2959t + \hat{R}_t \qquad \hat{R}_t = .5896\hat{R}_{t-1}$$

*Note that the forecast requires the value of x_{n+1}. When x_t is itself a time series, the future value x_{n+1} will generally be unknown and must also be estimated. Often, $x_t = t$ (as in Example 9.3). In this case, the future time period (e.g., $t = n + 1$) is known and no estimate is required.

FORECASTING USING TIME SERIES MODELS WITH FIRST-ORDER AUTOREGRESSIVE RESIDUALS

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + R_t$$
$$R_t = \phi R_{t-1} + \varepsilon_t$$

STEP 1 Use a statistical software package to obtain the estimated model

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + \cdots + \hat{\beta}_k x_{kt} + \hat{R}_t, \quad t = 1, 2, \ldots, n$$
$$\hat{R}_t = \hat{\phi}\hat{R}_{t-1}$$

STEP 2 Compute the estimated residual for the last time period in the data (i.e., $t = n$) as follows:

$$\hat{R}_n = y_n - \hat{y}_n$$
$$= y_n - (\hat{\beta}_0 + \hat{\beta}_1 x_{1n} + \hat{\beta}_2 x_{2n} + \cdots + \hat{\beta}_k x_{kn})$$

STEP 3 To forecast the value y_{n+1}, compute

$$\hat{R}_{n+1} = \hat{\phi}\hat{R}_n \quad \text{(where } \hat{R}_n \text{ is obtained from step 2)}$$
$$\hat{y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{1,n+1} + \hat{\beta}_2 x_{2,n+1} + \cdots + \hat{\beta}_k x_{k,n+1} + \hat{R}_{n+1}$$

STEP 4 To forecast the value y_{n+2}, compute

$$\hat{R}_{n+2} = \hat{\phi}\hat{R}_{n+1} \quad \text{(where } \hat{R}_{n+1} \text{ is obtained from step 3)}$$
$$\hat{y}_{n+2} = \hat{\beta}_0 + \hat{\beta}_1 x_{1,n+2} + \hat{\beta}_2 x_{2,n+2} + \cdots + \hat{\beta}_k x_{k,n+2} + \hat{R}_{n+2}$$

Future forecasts are obtained in a similar manner.

Combining these, we have

$$\hat{y}_t = .4058 + 4.2959t + .5896\hat{R}_{t-1}$$

a. Use the fitted model to forecast sales in years $t = 36$, 37, and 38.

b. Calculate approximate 95% prediction intervals for the forecasts.

SOLUTION

a. The forecast for the 36th year requires an estimate of the residual R_{35},

$$\hat{R}_{35} = y_{35} - [\hat{\beta}_0 + \hat{\beta}_1(35)]$$
$$= 150.9 - [.4058 + 4.2959(35)]$$
$$= .1377$$

Then

$$\hat{R}_{36} = \hat{\phi}\hat{R}_{35} = (.5896)(.1377) = .0812$$

and

$$\hat{y}_{36} = \hat{\beta}_0 + \hat{\beta}_1(36) + \hat{R}_{36}$$
$$= .4058 + 4.2959(36) + .0812$$
$$= 155.14$$

$\hat{R}_{37} = \hat{\phi}\hat{R}_{36}$

$\hat{y}_{37} = \hat{\beta}_0 + \hat{\beta}_1(37) + \hat{R}_{37}$