HOMEWORK: CIRCLED

Notice that the observed sales tend to deviate about the least squares line in positive and negative runs. That is, if the difference between the observed sales and predicted sales in year t is positive (or negative), the difference in year $t + 1$ tends to be positive (or negative). Since the variation in the yearly sales is systematic, the implication is that the errors are correlated. (We gave a formal statistical test for correlated errors in Section 6.6.) Violation of this standard regression assumption could lead to unreliable forecasts.

Time series models have been developed specifically for the purpose of making forecasts when the errors are known to be correlated. These models include an autoregressive term for the correlated errors that result from cyclical, seasonal, or other short-term effects. Time series autoregressive models are the subject of Sections 9.6–9.10.

EXERCISES 9.8–9.15

9.8 The accompanying table records the volume of wheat (in thousands of bushels) harvested by members of a farmers' marketing cooperative for the period 1975–1988. The cooperative is interested in detecting the long-term linear trend of the wheat harvest.

YEAR	TIME	WHEAT HARVESTED	YEAR	TIME	WHEAT HARVESTED
1975	1	75	1982	8	91
1976	2	78	1983	9	92
1977	3	82	1984	10	92
1978	4	82	1985	11	93
1979	5	84	1986	12	96
1980	6	85	1987	13	101
1981	7	87	1988	14	102

a. Graph the wheat harvest time series.
b. Propose a model for the long-term linear trend of the time series.
c. Fit the model, using the method of least squares. Plot the least squares line on the graph of part a. Can you identify the long-term trend? Use excel, check all residual boxes.
d. How well does the linear model describe the long-term trend? [*Hint:* Check the value of r^2.]
e. Use the least squares model to forecast the volume of wheat harvested in 1989. Construct a 95% prediction interval for the forecast.

9.9 A realtor working in a large city wants to identify the secular trend in the weekly number of one-family houses sold by her firm. For the past 15 weeks she has collected data on her firm's home sales, as shown in the table.

To summarize, we describe a general approach for constructing a time series:

1. Construct a regression model for the trend, seasonal, and cyclical components of $E(y_t)$. This model may be a polynomial in t for the trend (usually a straight-line or quadratic model) with trigonometric terms or dummy variables for the seasonal (cyclical) effects. The model may also include other time series variables as independent variables. For example, last year's rate of inflation may be used as a predictor of this year's GNP.
2. Next, construct a model for the random component (residual effect) of the model. A model that is widely used in practice is the first-order autoregressive model

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

   When the pattern of autocorrelation is more complex, use the general pth-order autoregressive model

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

3. The two components are then combined so that the model can be used for forecasting:

   $$y_t = E(y_t) + R_t$$

Prediction intervals are calculated to measure the reliability of the forecasts. In the following two sections we will demonstrate how time series models are fit to data and used for forecasting. In Section 9.9 we will present an example in which we fit a seasonal time series model to a set of data.

EXERCISES 9.20–9.24

9.20 Suppose you are interested in buying gold on the commodities market. Your broker has advised you that your best strategy is to sell back the gold at the first substantial jump in price. Hence, you are interested in a short-term investment. Before buying, you would like to model the closing price of gold, y_t, over time (in days), t.

a. Write a first-order model for the deterministic portion of the model, $E(y_t)$.

b. If a plot of the daily closing prices for the past month reveals a quadratic trend, write a plausible model for $E(y_t)$.

c. Since the closing price of gold on day $(t + 1)$ is very highly correlated with the closing price on day t, your broker suggests that the random error components of the model are not white noise. Given this information, postulate a model for the error term, R_t.

9.21 An economist wishes to model the GNP over time (in years) and also as a function of certain personal consumption expenditures. Let t = Time in years and let

y_t = GNP at time t

x_{1t} = Durable goods at time t

x_{2t} = Nondurable goods at time t

x_{3t} = Services at time t

a. The economist believes that y_t is linearly related to the independent variables x_{1t}, x_{2t}, x_{3t}, and t. Write the first-order model for $E(y_t)$.
b. Rewrite the model if interaction between the independent variables and time is present.
c. Postulate a model for the random error component, R_t. Explain why this model is appropriate.

9.22 Airlines sometimes overbook flights because of "no-show" passengers, i.e., passengers who have purchased a ticket but fail to board the flight. An airline supervisor wishes to be able to predict, for a Miami-to-New York flight, the monthly accumulation of no-show passengers during the upcoming year, using data from the past 3 years. Let y_t = Number of no-shows during month t.
a. Using dummy variables, propose a model for $E(y_t)$ that will take into account the seasonal (fall, winter, spring, summer) variation that may be present in the data.
b. Postulate a model for the error term R_t.
c. Write the full time series model for y_t (include random error terms).
d. Suppose the airline supervisor believes that the seasonal variation in the data is not constant from year to year; in other words, that there exists interaction between time and season. Rewrite the full model with the interaction terms added.

9.23 A farmer is interested in modeling the daily price of hogs at a livestock market. The farmer knows that the price varies over time (days) and also is reasonably confident that a seasonal effect is present.
a. Write a seasonal time series model with trigonometric terms for $E(y_t)$, where y_t = Selling price (in dollars) of hogs on day t.
b. Interpret the β parameters.
c. Include in the model an interaction between time and the trigonometric components. What does the presence of interaction signify?
d. Is it reasonable to assume that the random error component of the model, R_t, is white noise? Explain. Postulate a more appropriate model for R_t.

9.24 Suppose a CPA firm wants to model its monthly income, y_t. The firm is growing at an increasing rate, so that the mean income will be modeled as a second-order function of t. In addition, the mean monthly income increases significantly each year from January through April due to processing of tax returns.
a. Write a model for $E(y_t)$ to reflect both the second-order function of time, t, and the January–April jump in mean income.
b. Suppose the size of the January–April jump grows each year. How could this information be included in the model? Assume that 5 years of monthly data are available.

SECTION 9.7

FITTING TIME SERIES REGRESSION MODELS

We have proposed a general form for a time series model:

$$y_t = E(y_t) + R_t$$

where

$$E(y_t) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

and, using an autoregressive model for R_t,

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

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.

a. Hypothesize a time series model for quarterly GNP that includes a straight-line long-term trend and autocorrelated residuals.

b. The SAS printout for the pair of models

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

is shown here. Write the least squares prediction equation.

SAS Printout for Exercise 9.25

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


                    ORDINARY LEAST SQUARES ESTIMATES

            SSE              246375.5    DFE                 42
            MSE              5866.083    ROOT MSE      76.59036
            SBC              512.1736    AIC           508.6052
            REG RSQ            0.9909    TOTAL RSQ       0.9909
            DURBIN-WATSON      0.2649

      VARIABLE DF        B VALUE       STD ERROR    T RATIO APPROX PROB

      INTERCPT  1     1302.76522      23.4921909     55.455      0.0001
      T         1       61.32387       0.9092805     67.442      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       5599.44     1.000000   |                    |********************|
 1       4620.35     0.825145   |                    |****************    |

                    PRELIMINARY MSE=        1786.978

              ESTIMATES OF THE AUTOREGRESSIVE PARAMETERS
            LAG     COEFFICIENT       STD ERROR         T RATIO
              1     -0.82514548      0.08822573       -9.352663


                    YULE-WALKER ESTIMATES

            SSE              63839.13    DFE                 41
            MSE              1557.052    ROOT MSE       39.4595
            SBC              457.6783    AIC           452.3257
            REG RSQ            0.9575    TOTAL RSQ       0.9976

      VARIABLE DF        B VALUE       STD ERROR    T RATIO APPROX PROB

      INTERCPT  1     1327.28665      54.7378551     24.248      0.0001
      T         1       61.02839       2.0089776     30.378      0.0001
```

c. Interpret the estimates of the model parameters, β_0, β_1, and ϕ.

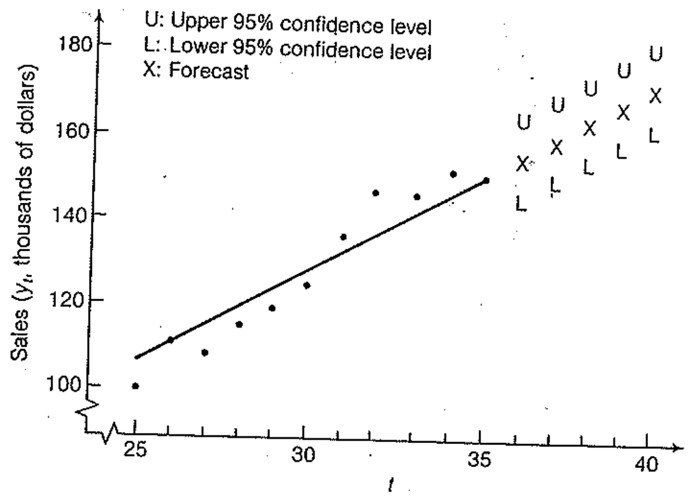
d. Interpret the value of R^2.

9.26 Refer to Exercise 9.8.

a. Hypothesize a time series model for annual volume of wheat harvested, y_t, that takes into account the residual autocorrelation.

b. If you have access to a computer package that uses the modified least squares method, fit the autoregressive time series model. Interpret the estimates of the model parameters.

FIGURE 9.14
Forecasts and Prediction Intervals for Years 36–40: Straight-Line Model with Autoregressive Residual



It is important to note that the forecasting procedure makes explicit use of the residual autocorrelation. The result is a better forecast than would be obtained using the standard least squares procedure of Chapter 4 (which ignores residual correlation). Generally, this is reflected by narrower prediction intervals for the time series forecasts than for the least squares prediction.* The end result, then, of using a time series model when autocorrelation is present is that you obtain more reliable estimates of the β coefficients, smaller residual variance, and more accurate prediction intervals for future values of the time series. ■

EXERCISES 9.30–9.37

9.30 The annual time series model $y_t = \beta_0 + \beta_1 t + \phi R_{t-1} + \varepsilon_t$ was fit to data collected for $n = 30$ years with the following results:

$$\hat{y}_t = 10 + 2.5t + .64\hat{R}_{t-1}$$

$$y_{30} = 82 \qquad \text{MSE} = 4.3$$

a. Calculate forecasts for y_t for $t = 31$, $t = 32$, and $t = 33$.

b. Construct approximate 95% prediction intervals for the forecasts obtained in part a.

9.31 The quarterly time series model $y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \phi R_{t-1} + \varepsilon_t$ was fit to data collected for $n = 48$ quarters, with the following results:

$$\hat{y}_t = 220 + 17t - .3t^2 + .82\hat{R}_{t-1}$$

$$y_{48} = 350 \qquad \text{MSE} = 10.5$$

a. Calculate forecasts for y_t for $t = 49$, $t = 50$, and $t = 51$.

b. Construct approximate 95% prediction intervals for the forecasts obtained in part a.

YEAR	DJA
1984	1,187.
1985	1,500.
1986	1,926.

*When n is large, approximate 95% prediction intervals obtained from the standard least squares procedure reduce to $\hat{y}_t \pm 2\sqrt{\text{MSE}}$ for *all* future values of the time series. These intervals may actually be narrower than the more accurate prediction intervals produced from the time series analysis.