1 Introduction

In this chapter we introduce some basic ideas of time series analysis and stochastic processes. Of particular importance are the concepts of stationarity and the autocovariance and sample autocovariance functions. Some standard techniques are described for the estimation and removal of trend and seasonality (of known period) from an observed time series. These are illustrated with reference to the data sets in Section 1.1. The calculations in all the examples can be carried out using the time series package ITSM, the student version of which is supplied on the enclosed CD. The data sets are contained in files with names ending in .TSM. For example, the Australian red wine sales are filed as WINE.TSM. Most of the topics covered in this chapter will be developed more fully in later sections of the book. The reader who is not already familiar with random variables and random vectors should first read Appendix A, where a concise account of the required background is given.

1.1 Examples of Time Series

A **time series** is a set of observations x_t, each one being recorded at a specific time t. A *discrete-time time series* (the type to which this book is primarily devoted) is one in which the set T_0 of times at which observations are made is a discrete set, as is the

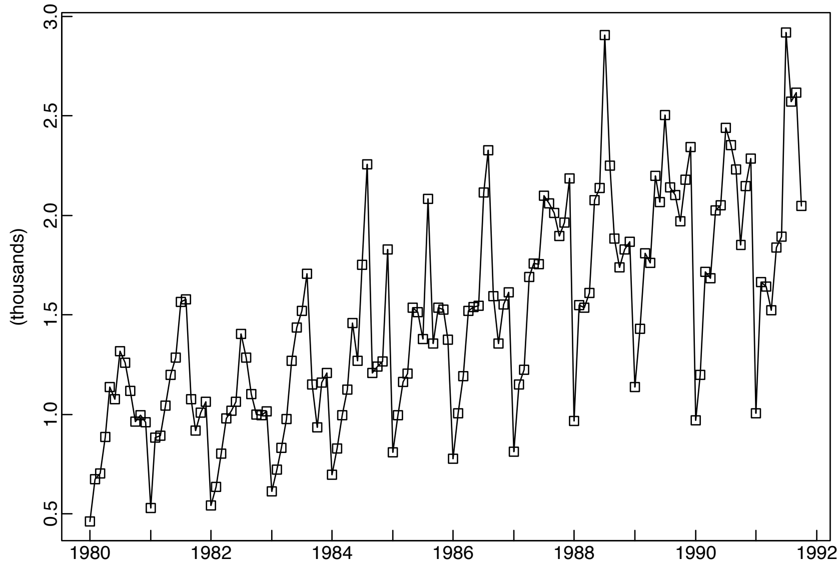

Figure 1-1
The Australian red wine sales, Jan. '80 – Oct. '91.

case, for example, when observations are made at fixed time intervals. *Continuous-time time series* are obtained when observations are recorded continuously over some time interval, e.g., when $T_0 = [0, 1]$.

Example 1.1.1 Australian red wine sales; WINE.TSM

Figure 1.1 shows the monthly sales (in kiloliters) of red wine by Australian winemakers from January 1980 through October 1991. In this case the set T_0 consists of the 142 times {(Jan. 1980), (Feb. 1980), ... ,(Oct. 1991)}. Given a set of n observations made at uniformly spaced time intervals, it is often convenient to rescale the time axis in such a way that T_0 becomes the set of integers $\{1, 2, \ldots, n\}$. In the present example this amounts to measuring time in months with (Jan. 1980) as month 1. Then T_0 is the set $\{1, 2, \ldots, 142\}$. It appears from the graph that the sales have an upward trend and a seasonal pattern with a peak in July and a trough in January. To plot the data using ITSM, run the program by double-clicking on the ITSM icon and then select the option `File>Project>Open>Univariate`, click OK, and select the file WINE.TSM. The graph of the data will then appear on your screen. □

Example 1.1.2 All-star baseball games, 1933–1995

Figure 1.2 shows the results of the all-star games by plotting x_t, where

$$x_t = \begin{cases} 1 & \text{if the National League won in year } t, \\ -1 & \text{if the American League won in year } t. \end{cases}$$

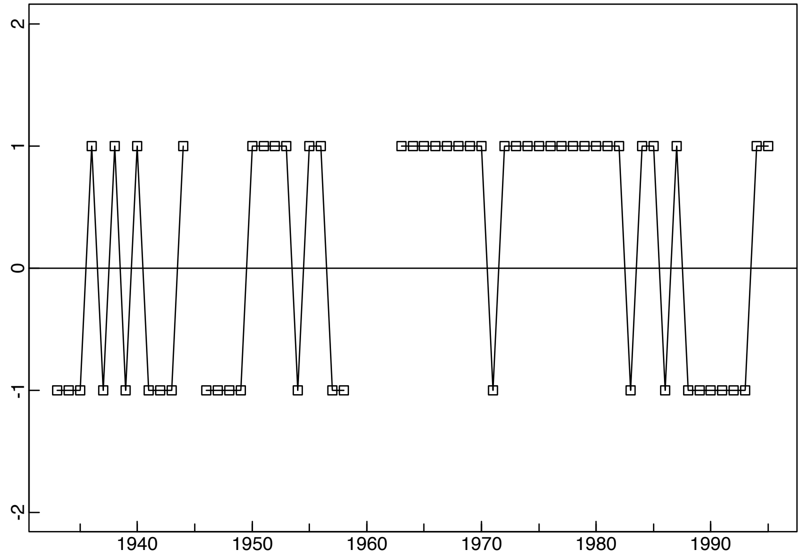

Figure 1-2 Results of the all-star baseball games, 1933–1995.

This is a series with only two possible values, ± 1. It also has some missing values, since no game was played in 1945, and two games were scheduled for each of the years 1959–1962. □

Example 1.1.3 Accidental deaths, U.S.A., 1973–1978; DEATHS.TSM

Like the red wine sales, the monthly accidental death figures show a strong seasonal pattern, with the maximum for each year occurring in July and the minimum for each year occurring in February. The presence of a trend in Figure 1.3 is much less apparent than in the wine sales. In Section 1.5 we shall consider the problem of representing the data as the sum of a trend, a seasonal component, and a residual term. □

Example 1.1.4 A signal detection problem; SIGNAL.TSM

Figure 1.4 shows simulated values of the series

$$X_t = \cos\left(\frac{t}{10}\right) + N_t, \quad t = 1, 2, \ldots, 200,$$

where $\{N_t\}$ is a sequence of independent normal random variables, with mean 0 and variance 0.25. Such a series is often referred to as *signal plus noise*, the signal being the smooth function, $S_t = \cos(\frac{t}{10})$ in this case. Given only the data X_t, how can we determine the unknown signal component? There are many approaches to this general problem under varying assumptions about the signal and the noise. One simple approach is to *smooth* the data by expressing X_t as a sum of sine waves of various frequencies (see Section 4.2) and eliminating the high-frequency components. If we do this to the values of $\{X_t\}$ shown in Figure 1.4 and retain only the lowest

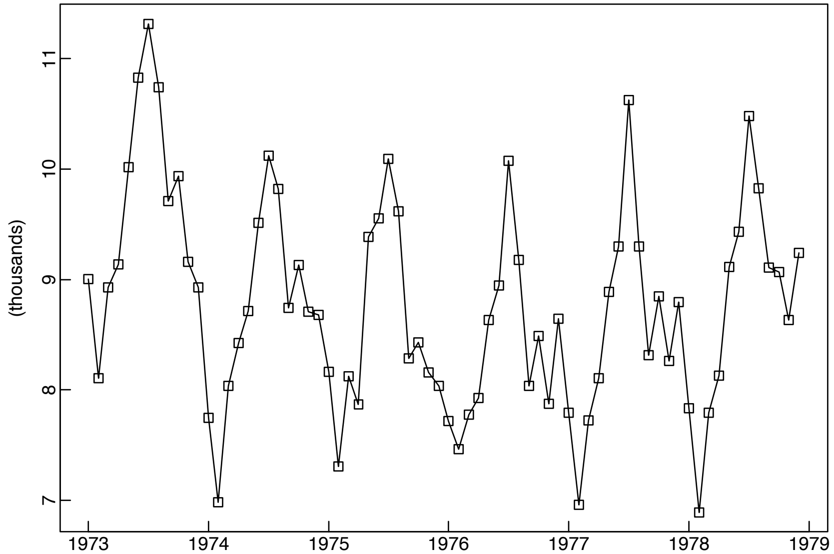

Figure 1-3
The monthly accidental deaths data, 1973–1978.

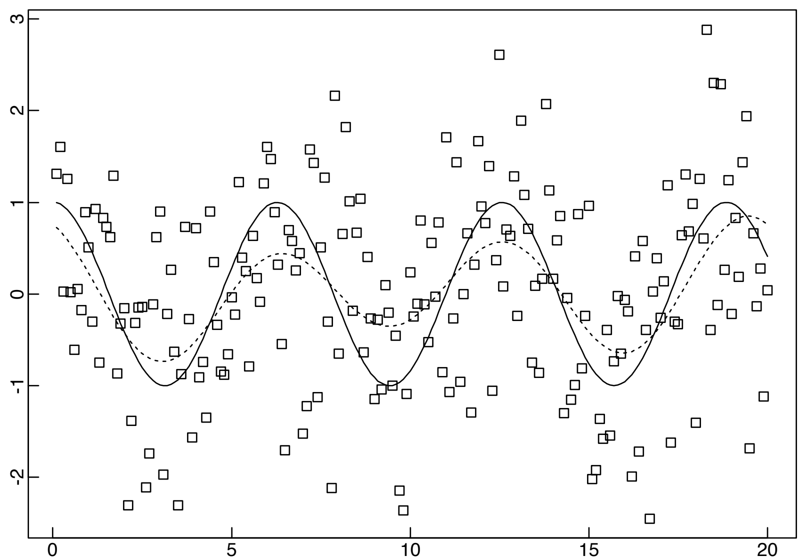

Figure 1-4
The series $\{X_t\}$ of Example 1.1.4.

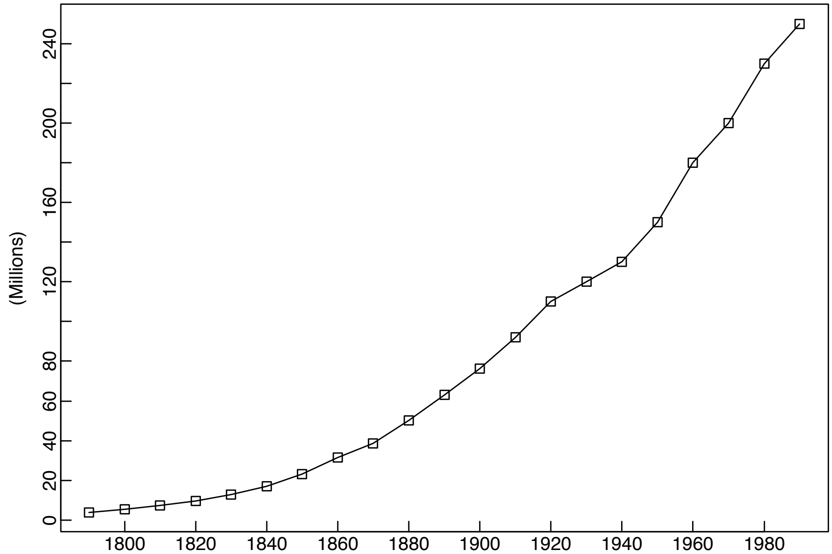

Figure 1-5
Population of the U.S.A. at ten-year intervals, 1790–1990.

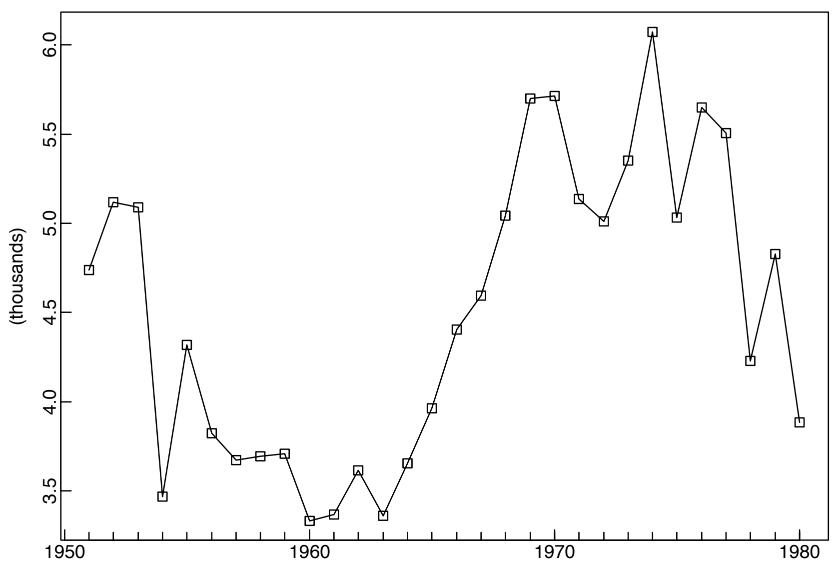

Figure 1-6
Strikes in the U.S.A., 1951–1980.

3.5% of the frequency components, we obtain the estimate of the signal also shown in Figure 1.4. The waveform of the signal is quite close to that of the true signal in this case, although its amplitude is somewhat smaller. □

Example 1.1.5 Population of the U.S.A., 1790–1990; USPOP.TSM

The population of the U.S.A., measured at ten-year intervals, is shown in Figure 1.5. The graph suggests the possibility of fitting a quadratic or exponential trend to the data. We shall explore this further in Section 1.3. □

Example 1.1.6 Number of strikes per year in the U.S.A., 1951–1980; STRIKES.TSM

The annual numbers of strikes in the U.S.A. for the years 1951–1980 are shown in Figure 1.6. They appear to fluctuate erratically about a slowly changing level. □

1.2 Objectives of Time Series Analysis

The examples considered in Section 1.1 are an extremely small sample from the multitude of time series encountered in the fields of engineering, science, sociology, and economics. Our purpose in this book is to study techniques for drawing inferences from such series. Before we can do this, however, it is necessary to set up a hypothetical probability model to represent the data. After an appropriate family of models has been chosen, it is then possible to estimate parameters, check for goodness of fit to the data, and possibly to use the fitted model to enhance our understanding of the mechanism generating the series. Once a satisfactory model has been developed, it may be used in a variety of ways depending on the particular field of application.

The model may be used simply to provide a compact description of the data. We may, for example, be able to represent the accidental deaths data of Example 1.1.3 as the sum of a specified trend, and seasonal and random terms. For the interpretation of economic statistics such as unemployment figures, it is important to recognize the presence of seasonal components and to remove them so as not to confuse them with long-term trends. This process is known as **seasonal adjustment**. Other applications of time series models include separation (or filtering) of noise from signals as in Example 1.1.4, prediction of future values of a series such as the red wine sales in Example 1.1.1 or the population data in Example 1.1.5, testing hypotheses such as global warming using recorded temperature data, predicting one series from observations of another, e.g., predicting future sales using advertising expenditure data, and controlling future values of a series by adjusting parameters. Time series models are also useful in simulation studies. For example, the performance of a reservoir depends heavily on the random daily inputs of water to the system. If these are modeled as a time series, then we can use the fitted model to simulate a large number of independent sequences of daily inputs. Knowing the size and mode of operation of the reservoir, we can determine the fraction of the simulated input sequences that cause the reservoir to run out of water in a given time period. This fraction will then be an estimate of the probability of emptiness of the reservoir at some time in the given period.

1.3 Some Simple Time Series Models

An important part of the analysis of a time series is the selection of a suitable probability model (or class of models) for the data. To allow for the possibly unpredictable nature of future observations it is natural to suppose that each observation x_t is a realized value of a certain random variable X_t.

Definition 1.3.1

A **time series model** for the observed data $\{x_t\}$ is a specification of the joint distributions (or possibly only the means and covariances) of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization.

Remark. We shall frequently use the term *time series* to mean both the data and the process of which it is a realization. □

A complete probabilistic time series model for the sequence of random variables $\{X_1, X_2, \ldots\}$ would specify all of the **joint distributions** of the random vectors $(X_1, \ldots, X_n)'$, $n = 1, 2, \ldots$, or equivalently all of the probabilities

$$P[X_1 \le x_1, \ldots, X_n \le x_n], \quad -\infty < x_1, \ldots, x_n < \infty, \quad n = 1, 2, \ldots.$$

Such a specification is rarely used in time series analysis (unless the data are generated by some well-understood simple mechanism), since in general it will contain far too many parameters to be estimated from the available data. Instead we specify only the **first-** and **second-order moments** of the joint distributions, i.e., the expected values EX_t and the expected products $E(X_{t+h}X_t)$, $t = 1, 2, \ldots$, $h = 0, 1, 2, \ldots$, focusing on properties of the sequence $\{X_t\}$ that depend only on these. Such properties of $\{X_t\}$ are referred to as **second-order properties**. In the particular case where all the joint distributions are multivariate normal, the second-order properties of $\{X_t\}$ completely determine the joint distributions and hence give a complete probabilistic characterization of the sequence. In general we shall lose a certain amount of information by looking at time series "through second-order spectacles"; however, as we shall see in Chapter 2, the theory of minimum mean squared error linear prediction depends only on the second-order properties, thus providing further justification for the use of the second-order characterization of time series models.

Figure 1.7 shows one of many possible realizations of $\{S_t, t = 1, \ldots, 200\}$, where $\{S_t\}$ is a sequence of random variables specified in Example 1.3.3 below. In most practical problems involving time series we see only *one* realization. For example, there is only one available realization of Fort Collins's annual rainfall for the years 1900–1996, but we imagine it to be one of the many sequences that *might* have occurred. In the following examples we introduce some simple time series models. One of our goals will be to expand this repertoire so as to have at our disposal a broad range of models with which to try to match the observed behavior of given data sets.

1.3.1 Some Zero-Mean Models

Example 1.3.1 iid noise

Perhaps the simplest model for a time series is one in which there is no trend or seasonal component and in which the observations are simply independent and identically distributed (iid) random variables with zero mean. We refer to such a sequence of random variables $X_1, X_2, \ldots$ as iid noise. By definition we can write, for any positive integer n and real numbers $x_1, \ldots, x_n$,

$$P[X_1 \le x_1, \ldots, X_n \le x_n] = P[X_1 \le x_1] \cdots P[X_n \le x_n] = F(x_1) \cdots F(x_n),$$

where $F(\cdot)$ is the cumulative distribution function (see Section A.1) of each of the identically distributed random variables $X_1, X_2, \ldots$. In this model there is no dependence between observations. In particular, for all $h \ge 1$ and all $x, x_1, \ldots, x_n$,

$$P[X_{n+h} \le x | X_1 = x_1, \ldots, X_n = x_n] = P[X_{n+h} \le x],$$

showing that knowledge of $X_1, \ldots, X_n$ is of no value for predicting the behavior of X_{n+h}. Given the values of $X_1, \ldots, X_n$, the function f that minimizes the mean squared error $E\left[(X_{n+h} - f(X_1, \ldots, X_n))^2\right]$ is in fact identically zero (see Problem 1.2). Although this means that iid noise is a rather uninteresting process for forecasters, it plays an important role as a building block for more complicated time series models. □

Example 1.3.2 A binary process

As an example of iid noise, consider the sequence of iid random variables $\{X_t, t = 1, 2, \ldots,\}$ with

$$P[X_t = 1] = p, \quad P[X_t = -1] = 1 - p,$$

where $p = \frac{1}{2}$. The time series obtained by tossing a penny repeatedly and scoring $+1$ for each head and -1 for each tail is usually modeled as a realization of this process. A priori we might well consider the same process as a model for the all-star baseball games in Example 1.1.2. However, even a cursory inspection of the results from 1963–1982, which show the National League winning 19 of 20 games, casts serious doubt on the hypothesis $P[X_t = 1] = \frac{1}{2}$. □

Example 1.3.3 Random walk

The random walk $\{S_t, t = 0, 1, 2, \ldots\}$ (starting at zero) is obtained by cumulatively summing (or "integrating") iid random variables. Thus a random walk with zero mean is obtained by defining $S_0 = 0$ and

$$S_t = X_1 + X_2 + \cdots + X_t, \quad \text{for } t = 1, 2, \ldots,$$

where $\{X_t\}$ is iid noise. If $\{X_t\}$ is the binary process of Example 1.3.2, then $\{S_t, t = 0, 1, 2, \ldots,\}$ is called a **simple symmetric random walk**. This walk can be viewed as the location of a pedestrian who starts at position zero at time zero and at each integer time tosses a fair coin, stepping one unit to the right each time a head appears and one unit to the left for each tail. A realization of length 200 of a simple symmetric random walk is shown in Figure 1.7. Notice that the outcomes of the coin tosses can be recovered from $\{S_t, t = 0, 1, \ldots\}$ by differencing. Thus the result of the tth toss can be found from $S_t - S_{t-1} = X_t$. □

1.3.2 Models with Trend and Seasonality

In several of the time series examples of Section 1.1 there is a clear trend in the data. An increasing trend is apparent in both the Australian red wine sales (Figure 1.1) and the population of the U.S.A. (Figure 1.5). In both cases a zero-mean model for the data is clearly inappropriate. The graph of the population data, which contains no apparent periodic component, suggests trying a model of the form

$$X_t = m_t + Y_t,$$

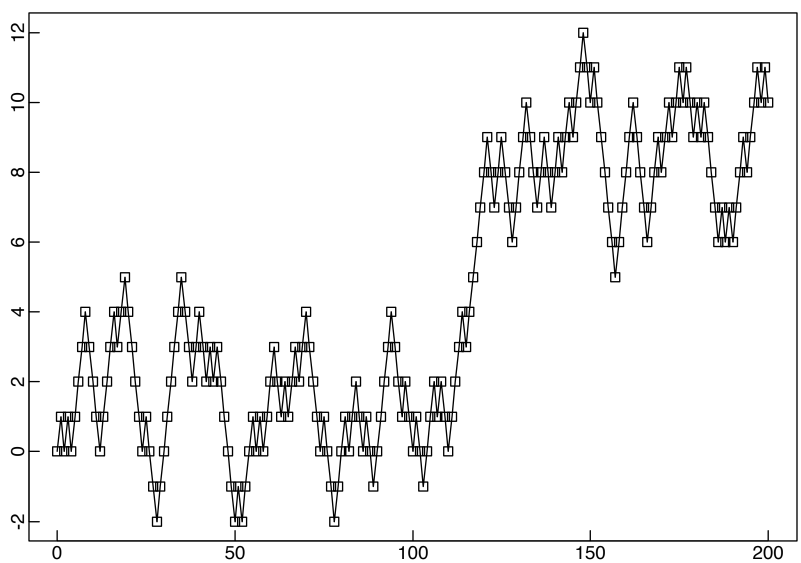

Figure 1-7
One realization of a simple random walk $\{S_t, t = 0, 1, 2, \ldots, 200\}$

where m_t is a slowly changing function known as the **trend component** and Y_t has zero mean. A useful technique for estimating m_t is the method of least squares (some other methods are considered in Section 1.5).

In the least squares procedure we attempt to fit a parametric family of functions, e.g.,

$$m_t = a_0 + a_1 t + a_2 t^2, \tag{1.3.1}$$

to the data $\{x_1, \ldots, x_n\}$ by choosing the parameters, in this illustration a_0, a_1, and a_2, to minimize $\sum_{t=1}^{n}(x_t - m_t)^2$. This method of curve fitting is called **least squares regression** and can be carried out using the program ITSM and selecting the `Regression` option.

Example 1.3.4 Population of the U.S.A., 1790–1990

To fit a function of the form (1.3.1) to the population data shown in Figure 1.5 we relabel the time axis so that $t = 1$ corresponds to 1790 and $t = 21$ corresponds to 1990. Run ITSM, select `File>Project>Open>Univariate`, and open the file USPOP.TSM. Then select `Regression>Specify`, choose `Polynomial Regression` with order equal to 2, and click OK. Then select `Regression>Estimation>Least Squares`, and you will obtain the following estimated parameter values in the model (1.3.1):

$$\hat{a}_0 = 6.9579 \times 10^6,$$

$$\hat{a}_1 = -2.1599 \times 10^6,$$

and

$$\hat{a}_2 = 6.5063 \times 10^5.$$

A graph of the fitted function is shown with the original data in Figure 1.8. The estimated values of the noise process Y_t, $1 \le t \le 21$, are the residuals obtained by subtraction of $\hat{m}_t = \hat{a}_0 + \hat{a}_1 t + \hat{a}_2 t^2$ from x_t.

The estimated trend component $\hat{m}_t$ furnishes us with a natural predictor of future values of X_t. For example, if we estimate the noise Y_{22} by its mean value, i.e., zero, then (1.3.1) gives the estimated U.S. population for the year 2000 as

$$\hat{m}_{22} = 6.9579 \times 10^6 - 2.1599 \times 10^6 \times 22 + 6.5063 \times 10^5 \times 22^2 = 274.35 \times 10^6.$$

However, if the residuals $\{Y_t\}$ are highly correlated, we may be able to use their values to give a better estimate of Y_{22} and hence of the population X_{22} in the year 2000. □

Example 1.3.5 Level of Lake Huron 1875–1972; LAKE.DAT

A graph of the level in feet of Lake Huron (reduced by 570) in the years 1875–1972 is displayed in Figure 1.9. Since the lake level appears to decline at a roughly linear rate, ITSM was used to fit a model of the form

$$X_t = a_0 + a_1 t + Y_t, \quad t = 1, \ldots, 98 \tag{1.3.2}$$

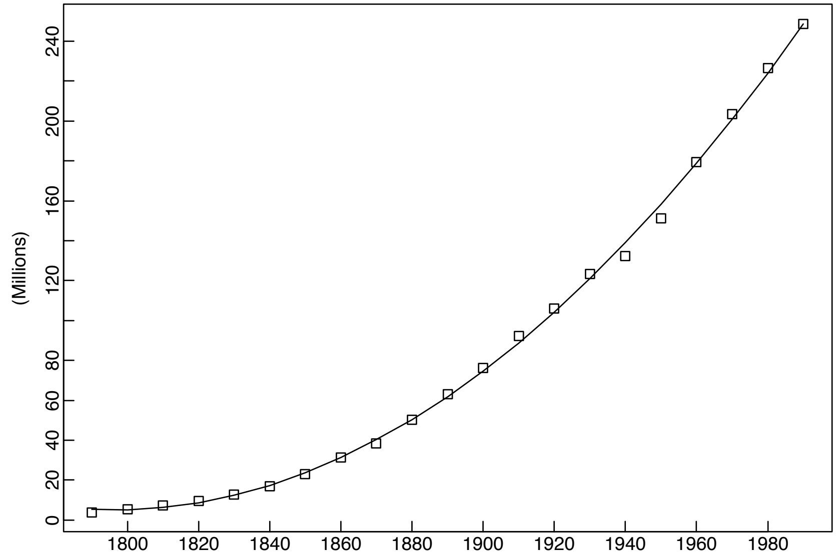

Figure 1-8
Population of the U.S.A. showing the quadratic trend fitted by least squares.

(with the time axis relabeled as in Example 1.3.4). The least squares estimates of the parameter values are

$$\hat{a}_0 = 10.202 \quad \text{and} \quad \hat{a}_1 = -.0242.$$

(The resulting least squares line, $\hat{a}_0 + \hat{a}_1 t$, is also displayed in Figure 1.9.) The estimates of the noise, Y_t, in the model (1.3.2) are the residuals obtained by subtracting the least squares line from x_t and are plotted in Figure 1.10. There are two interesting

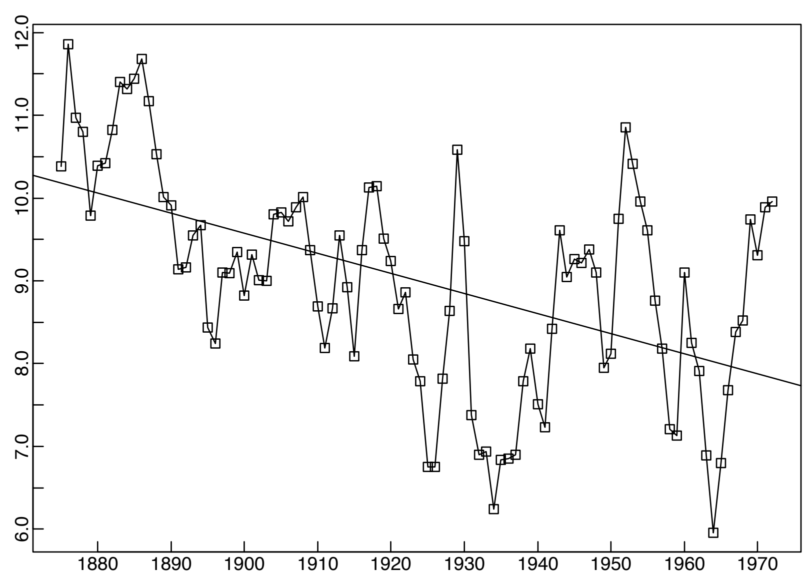

Figure 1-9
Level of Lake Huron 1875–1972 showing the line fitted by least squares.

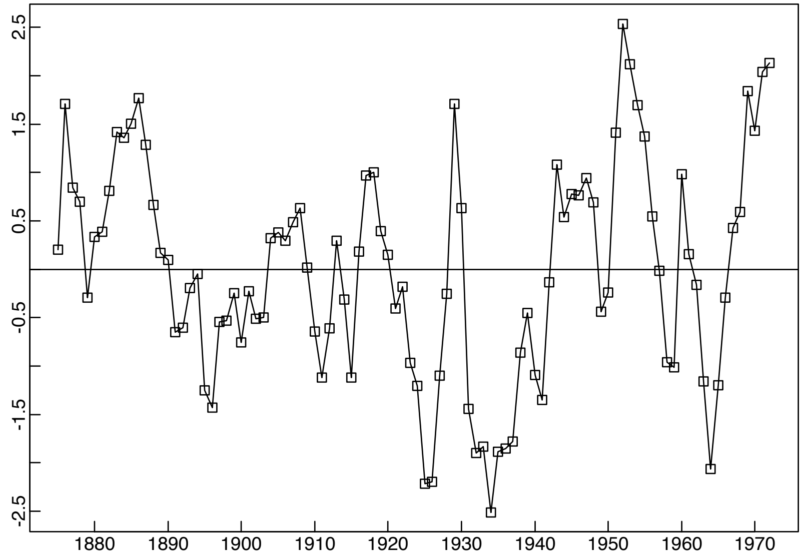

Figure 1-10
Residuals from fitting a line to the Lake Huron data in Figure 1.9.

features of the graph of the residuals. The first is the absence of any discernible trend. The second is the smoothness of the graph. (In particular, there are long stretches of residuals that have the same sign. This would be very unlikely to occur if the residuals were observations of iid noise with zero mean.) Smoothness of the graph of a time series is generally indicative of the existence of some form of dependence among the observations.

Such dependence can be used to advantage in forecasting future values of the series. If we were to assume the validity of the fitted model with iid residuals $\{Y_t\}$, then the minimum mean squared error predictor of the next residual (Y_{99}) would be zero (by Problem 1.2). However, Figure 1.10 strongly suggests that Y_{99} will be positive.

How then do we quantify dependence, and how do we construct models for forecasting that incorporate dependence of a particular type? To deal with these questions, Section 1.4 introduces the autocorrelation function as a measure of dependence, and stationary processes as a family of useful models exhibiting a wide variety of dependence structures. $\square$

Harmonic Regression

Many time series are influenced by seasonally varying factors such as the weather, the effect of which can be modeled by a periodic component with fixed known period. For example, the accidental deaths series (Figure 1.3) shows a repeating annual pattern with peaks in July and troughs in February, strongly suggesting a seasonal factor with period 12. In order to represent such a seasonal effect, allowing for noise but assuming no trend, we can use the simple model,

$$X_t = s_t + Y_t,$$

where s_t is a periodic function of t with period d $(s_{t-d} = s_t)$. A convenient choice for s_t is a sum of harmonics (or sine waves) given by

$$s_t = a_0 + \sum_{j=1}^{k}(a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)), \tag{1.3.3}$$

where $a_0, a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ are unknown parameters and $\lambda_1, \ldots, \lambda_k$ are fixed frequencies, each being some integer multiple of $2\pi/d$. To carry out harmonic regression using ITSM, select `Regression>Specify` and check `Include intercept term` and `Harmonic Regression`. Then specify the number of harmonics (k in (1.3.3)) and enter k integer-valued Fourier indices $f_1, \ldots, f_k$. For a sine wave with period d, set $f_1 = n/d$, where n is the number of observations in the time series. (If n/d is not an integer, you will need to delete a few observations from the beginning of the series to make it so.) The other $k - 1$ Fourier indices should be positive integer multiples of the first, corresponding to harmonics of the fundamental sine wave with period d. Thus to fit a single sine wave with period 365 to 365 daily observations we would choose $k = 1$ and $f_1 = 1$. To fit a linear combination of sine waves with periods $365/j$, $j = 1, \ldots, 4$, we would choose $k = 4$ and $f_j = j$, $j = 1, \ldots, 4$. Once k and $f_1, \ldots, f_k$ have been specified, click OK and then select `Regression>Estimation >Least Squares` to obtain the required regression coefficients. To see how well the fitted function matches the data, select `Regression>Show fit`.

Example 1.3.6 Accidental deaths

To fit a sum of two harmonics with periods twelve months and six months to the monthly accidental deaths data $x_1, \ldots, x_n$ with $n = 72$, we choose $k = 2$, $f_1 =$

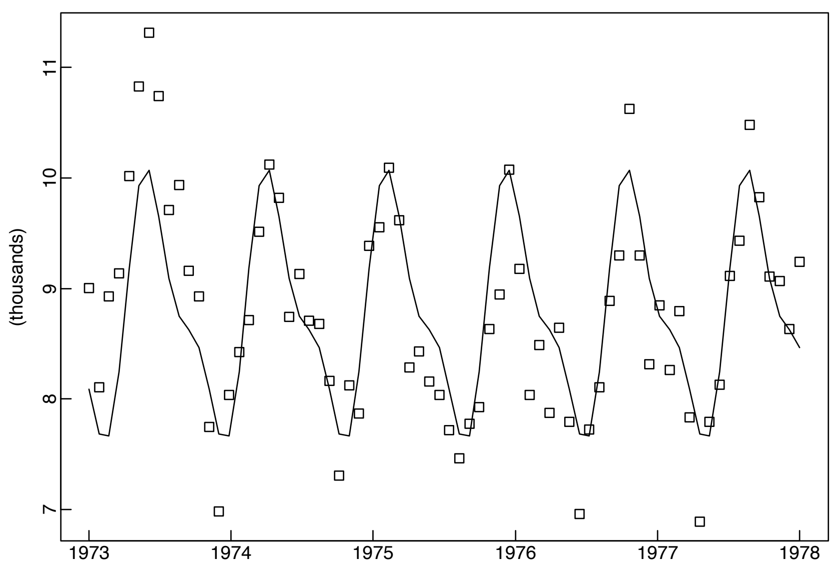

Figure 1-11 The estimated harmonic component of the accidental deaths data from ITSM.

$n/12 = 6$, and $f_2 = n/6 = 12$. Using ITSM as described above, we obtain the fitted function shown in Figure 1.11. As can be seen from the figure, the periodic character of the series is captured reasonably well by this fitted function. In practice, it is worth experimenting with several different combinations of harmonics in order to find a satisfactory estimate of the seasonal component. The program ITSM also allows fitting a linear combination of harmonics and polynomial trend by checking both `Harmonic Regression` and `Polynomial Regression` in the `Regression>Specification` dialog box. Other methods for dealing with seasonal variation in the presence of trend are described in Section 1.5. □

1.3.3 A General Approach to Time Series Modeling

The examples of the previous section illustrate a general approach to time series analysis that will form the basis for much of what is done in this book. Before introducing the ideas of dependence and stationarity, we outline this approach to provide the reader with an overview of the way in which the various ideas of this chapter fit together.

- Plot the series and examine the main features of the graph, checking in particular whether there is
  (a) a trend,
  (b) a seasonal component,
  (c) any apparent sharp changes in behavior,
  (d) any outlying observations.
- Remove the trend and seasonal components to get *stationary* residuals (as defined in Section 1.4). To achieve this goal it may sometimes be necessary to apply a preliminary transformation to the data. For example, if the magnitude of the fluctuations appears to grow roughly linearly with the level of the series, then the transformed series $\{\ln X_1, \ldots, \ln X_n\}$ will have fluctuations of more constant magnitude. See, for example, Figures 1.1 and 1.17. (If some of the data are negative, add a positive constant to each of the data values to ensure that all values are positive before taking logarithms.) There are several ways in which trend and seasonality can be removed (see Section 1.5), some involving estimating the components and subtracting them from the data, and others depending on *differencing* the data, i.e., replacing the original series $\{X_t\}$ by $\{Y_t := X_t - X_{t-d}\}$ for some positive integer d. Whichever method is used, the aim is to produce a stationary series, whose values we shall refer to as residuals.
- Choose a model to fit the residuals, making use of various sample statistics including the sample autocorrelation function to be defined in Section 1.4.
- Forecasting will be achieved by forecasting the residuals and then inverting the transformations described above to arrive at forecasts of the original series $\{X_t\}$.

- An extremely useful alternative approach touched on only briefly in this book is to express the series in terms of its Fourier components, which are sinusoidal waves of different frequencies (cf. Example 1.1.4). This approach is especially important in engineering applications such as signal processing and structural design. It is important, for example, to ensure that the resonant frequency of a structure does not coincide with a frequency at which the loading forces on the structure have a particularly large component.

1.4 Stationary Models and the Autocorrelation Function

Loosely speaking, a time series $\{X_t, t = 0, \pm 1, \ldots\}$ is said to be stationary if it has statistical properties similar to those of the "time-shifted" series $\{X_{t+h}, t = 0, \pm 1, \ldots\}$, for each integer h. Restricting attention to those properties that depend only on the first- and second-order moments of $\{X_t\}$, we can make this idea precise with the following definitions.

Definition 1.4.1 Let $\{X_t\}$ be a time series with $E(X_t^2) < \infty$. The **mean function** of $\{X_t\}$ is

$$\mu_X(t) = E(X_t).$$

The **covariance function** of $\{X_t\}$ is

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

for all integers r and s.

Definition 1.4.2 $\{X_t\}$ is (**weakly**) **stationary** if

(i) $\mu_X(t)$ is independent of t,

and

(ii) $\gamma_X(t + h, t)$ is independent of t for each h.

Remark 1. Strict stationarity of a time series $\{X_t, t = 0, \pm 1, \ldots\}$ is defined by the condition that $(X_1, \ldots, X_n)$ and $(X_{1+h}, \ldots, X_{n+h})$ have the same joint distributions for all integers h and $n > 0$. It is easy to check that if $\{X_t\}$ is strictly stationary and $EX_t^2 < \infty$ for all t, then $\{X_t\}$ is also weakly stationary (Problem 1.3). Whenever we use the term *stationary* we shall mean weakly stationary as in Definition 1.4.2, unless we specifically indicate otherwise. □

Remark 2. In view of condition (ii), whenever we use the term covariance function with reference to a *stationary* time series $\{X_t\}$ we shall mean the function γ_X of *one* variable, defined by

$$\gamma_X(h) := \gamma_X(h,0) = \gamma_X(t+h,t).$$

The function $\gamma_X(\cdot)$ will be referred to as the autocovariance function and $\gamma_X(h)$ as its value at *lag* h. □

Definition 1.4.3

Let $\{X_t\}$ be a stationary time series. The **autocovariance function** (ACVF) of $\{X_t\}$ at lag h is

$$\gamma_X(h) = \mathrm{Cov}(X_{t+h}, X_t).$$

The **autocorrelation function** (ACF) of $\{X_t\}$ at lag h is

$$\rho_X(h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)} = \mathrm{Cor}(X_{t+h}, X_t).$$

In the following examples we shall frequently use the easily verified **linearity property of covariances**, that if $\mathrm{E}X^2 < \infty$, $\mathrm{E}Y^2 < \infty$, $\mathrm{E}Z^2 < \infty$ and a, b, and c are any real constants, then

$$\mathrm{Cov}(aX + bY + c, Z) = a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z).$$

Example 1.4.1 iid noise

If $\{X_t\}$ is iid noise and $E(X_t^2) = \sigma^2 < \infty$, then the first requirement of Definition 1.4.2 is obviously satisfied, since $E(X_t) = 0$ for all t. By the assumed independence,

$$\gamma_X(t+h,t) = \begin{cases} \sigma^2, & \text{if } h = 0, \\ 0, & \text{if } h \neq 0, \end{cases}$$

which does not depend on t. Hence iid noise with finite second moment is stationary. We shall use the notation

$$\{X_t\} \sim \mathrm{IID}\left(0, \sigma^2\right)$$

to indicate that the random variables X_t are independent and identically distributed random variables, each with mean 0 and variance σ^2. □

Example 1.4.2 White noise

If $\{X_t\}$ is a sequence of uncorrelated random variables, each with zero mean and variance σ^2, then clearly $\{X_t\}$ is stationary with the same covariance function as the iid noise in Example 1.4.1. Such a sequence is referred to as **white noise** (with mean 0 and variance σ^2). This is indicated by the notation

$$\{X_t\} \sim \mathrm{WN}\left(0, \sigma^2\right).$$

Clearly, every IID$\left(0, \sigma^2\right)$ sequence is WN$\left(0, \sigma^2\right)$ but not conversely (see Problem 1.8 and the ARCH(1) process of Section 10.3). □

Example 1.4.3 The random walk

If $\{S_t\}$ is the random walk defined in Example 1.3.3 with $\{X_t\}$ as in Example 1.4.1, then $ES_t = 0$, $E(S_t^2) = t\sigma^2 < \infty$ for all t, and, for $h \geq 0$,

$$\begin{aligned}\gamma_S(t+h, t) &= \text{Cov}(S_{t+h}, S_t)\\ &= \text{Cov}(S_t + X_{t+1} + \cdots + X_{t+h}, S_t)\\ &= \text{Cov}(S_t, S_t)\\ &= t\sigma^2.\end{aligned}$$

Since $\gamma_S(t+h, t)$ depends on t, the series $\{S_t\}$ is *not* stationary. □

Example 1.4.4 First-order moving average or MA(1) process

Consider the series defined by the equation

$$X_t = Z_t + \theta Z_{t-1}, \quad t = 0, \pm 1, \ldots, \tag{1.4.1}$$

where $\{Z_t\} \sim \text{WN}\left(0, \sigma^2\right)$ and θ is a real-valued constant. From (1.4.1) we see that $EX_t = 0$, $EX_t^2 = \sigma^2(1+\theta^2) < \infty$, and

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2\left(1+\theta^2\right), & \text{if } h = 0,\\ \sigma^2\theta, & \text{if } h = \pm 1,\\ 0, & \text{if } |h| > 1.\end{cases}$$

Thus the requirements of Definition 1.4.2 are satisfied, and $\{X_t\}$ is stationary. The autocorrelation function of $\{X_t\}$ is

$$\rho_X(h) = \begin{cases} 1, & \text{if } h = 0,\\ \theta/\left(1+\theta^2\right), & \text{if } h = \pm 1,\\ 0, & \text{if } |h| > 1.\end{cases}$$ □

Example 1.4.5 First-order autoregression or AR(1) process

Let us *assume* now that $\{X_t\}$ is a stationary series satisfying the equations

$$X_t = \phi X_{t-1} + Z_t, \quad t = 0, \pm 1, \ldots, \tag{1.4.2}$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, $|\phi| < 1$, and Z_t is uncorrelated with X_s for each $s < t$. (We shall show in Section 2.2 that there is in fact exactly one such solution of (1.4.2).) By taking expectations on each side of (1.4.2) and using the fact that $EZ_t = 0$, we see at once that

$$EX_t = 0.$$

To find the autocorrelation function of $\{X_t\}$ we multiply each side of (1.4.2) by X_{t-h} $(h > 0)$ and then take expectations to get

$$\begin{aligned}\gamma_X(h) &= \mathrm{Cov}(X_t, X_{t-h})\\ &= \mathrm{Cov}(\phi X_{t-1}, X_{t-h}) + \mathrm{Cov}(Z_t, X_{t-h})\\ &= \phi\gamma_X(h-1) + 0 = \cdots = \phi^h \gamma_x(0).\end{aligned}$$

Observing that $\gamma(h) = \gamma(-h)$ and using Definition 1.4.3, we find that

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \phi^{|h|}, \quad h = 0, \pm 1, \ldots.$$

It follows from the linearity of the covariance function in each of its arguments and the fact that Z_t is uncorrelated with X_{t-1} that

$$\gamma_X(0) = \mathrm{Cov}(X_t, X_t) = \mathrm{Cov}(\phi X_{t-1} + Z_t, \phi X_{t-1} + Z_t) = \phi^2\gamma_X(0) + \sigma^2$$

and hence that $\gamma_X(0) = \sigma^2 / \left(1 - \phi^2\right)$. □

1.4.1 The Sample Autocorrelation Function

Although we have just seen how to compute the autocorrelation function for a few simple time series models, in practical problems we do not start with a model, but with *observed data* $\{x_1, x_2, \ldots, x_n\}$. To assess the degree of dependence in the data and to select a model for the data that reflects this, one of the important tools we use is the **sample autocorrelation function** (sample ACF) of the data. If we believe that the data are realized values of a stationary time series $\{X_t\}$, then the sample ACF will provide us with an estimate of the ACF of $\{X_t\}$. This estimate may suggest which of the many possible stationary time series models is a suitable candidate for representing the dependence in the data. For example, a sample ACF that is close to zero for all nonzero lags suggests that an appropriate model for the data might be iid noise. The following definitions are natural *sample* analogues of those for the autocovariance and autocorrelation functions given earlier for stationary time series *models*.

Definition 1.4.4

Let $x_1, \ldots, x_n$ be observations of a time series. The **sample mean** of $x_1, \ldots, x_n$ is

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

The **sample autocovariance function** is

$$\hat{\gamma}(h) := n^{-1}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad -n < h < n.$$

The **sample autocorrelation function** is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad -n < h < n.$$

Remark 3. For $h \ge 0$, $\hat{\gamma}(h)$ is approximately equal to the sample covariance of the $n - h$ pairs of observations $(x_1, x_{1+h}), (x_2, x_{2+h}), \ldots, (x_{n-h}, x_n)$. The difference arises from use of the divisor n instead of $n - h$ and the subtraction of the *overall* mean, $\bar{x}$, from each factor of the summands. Use of the divisor n ensures that the sample covariance matrix $\hat{\Gamma}_n := [\hat{\gamma}(i - j)]_{i,j=1}^{n}$ is nonnegative definite (see Section 2.4.2). □

Remark 4. Like the sample covariance matrix defined in Remark 3, the sample correlation matrix $\hat{R}_n := [\hat{\rho}(i - j)]_{i,j=1}^{n}$ is nonnegative definite. Each of its diagonal elements is equal to 1, since $\hat{\rho}(0) = 1$. □

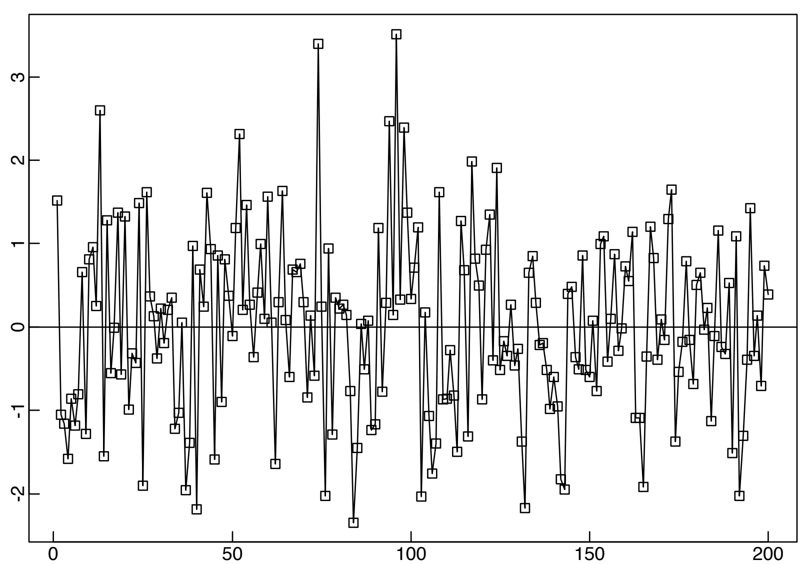

Figure 1-12
200 simulated values of iid N(0,1) noise.

Example 1.4.6 Figure 1.12 shows 200 simulated values of normally distributed iid (0, 1), denoted by IID N(0, 1), noise. Figure 1.13 shows the corresponding sample autocorrelation function at lags 0, 1, ..., 40. Since $\rho(h) = 0$ for $h > 0$, one would also expect the corresponding sample autocorrelations to be near 0. It can be shown, in fact, that for iid noise with finite variance, the sample autocorrelations $\hat{\rho}(h), h > 0$, are approximately IID N(0, $1/n$) for n large (see TSTM p. 222). Hence, approximately 95% of the sample autocorrelations should fall between the bounds $\pm 1.96/\sqrt{n}$ (since 1.96 is the .975 quantile of the standard normal distribution). Therefore, in Figure 1.13 we would expect roughly $40(.05) = 2$ values to fall outside the bounds. To simulate 200 values of IID N(0, 1) noise using ITSM, select `File>Project>New>Univariate` then `Model>Simulate`. In the resulting dialog box, enter 200 for the required `Number of Observations`. (The remaining entries in the dialog box can be left as they are, since the model assumed by ITSM, until you enter another, is IID N(0, 1) noise. If you wish to reproduce exactly the same sequence at a later date, record the `Random Number Seed` for later use. By specifying different values for the random number seed you can generate independent realizations of your time series.) Click on `OK` and you will see the graph of your simulated series. To see its sample autocorrelation function together with the autocorrelation function of the model that generated it, click on the third yellow button at the top of the screen and you will see the two graphs superimposed (with the latter in red.) The horizontal lines on the graph are the bounds $\pm 1.96/\sqrt{n}$. □

Remark 5. The sample autocovariance and autocorrelation functions can be computed for *any* data set $\{x_1, \ldots, x_n\}$ and are not restricted to observations from a

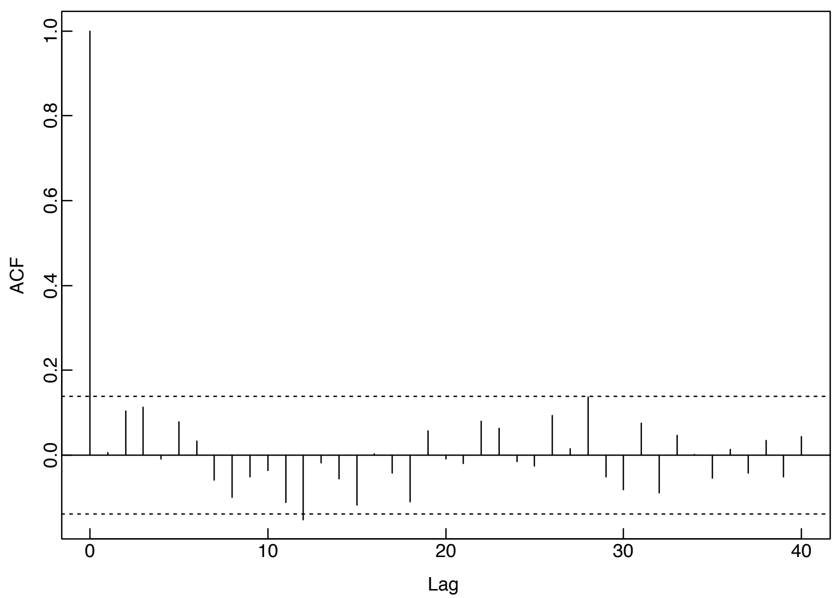

Figure 1-13 The sample autocorrelation function for the data of Figure 1.12 showing the bounds $\pm 1.96/\sqrt{n}$.

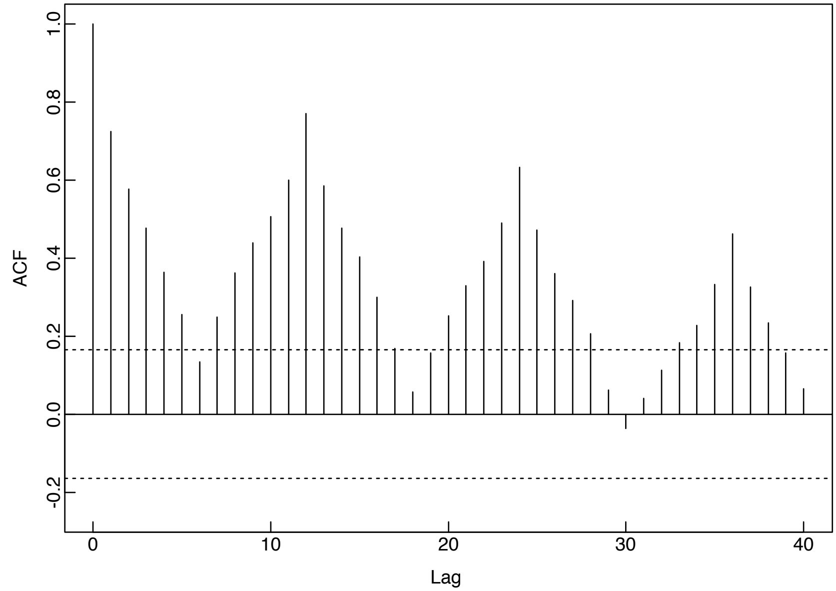

Figure 1-14
The sample autocorrelation function for the Australian red wine sales showing the bounds $\pm 1.96/\sqrt{n}$.

stationary time series. For data containing a trend, $|\hat{\rho}(h)|$ will exhibit slow decay as h increases, and for data with a substantial deterministic periodic component, $|\hat{\rho}(h)|$ will exhibit similar behavior with the same periodicity. (See the sample ACF of the Australian red wine sales in Figure 1.14 and Problem 1.9.) Thus $\hat{\rho}(\cdot)$ can be useful as an indicator of nonstationarity (see also Section 6.1). □

1.4.2 A Model for the Lake Huron Data

As noted earlier, an iid noise model for the residuals $\{y_1, \ldots, y_{98}\}$ obtained by fitting a straight line to the Lake Huron data in Example 1.3.5 appears to be inappropriate. This conclusion is confirmed by the sample ACF of the residuals (Figure 1.15), which has three of the first forty values well outside the bounds $\pm 1.96/\sqrt{98}$.

The roughly geometric decay of the first few sample autocorrelations (with $\hat{\rho}(h+1)/\hat{\rho}(h) \approx 0.7$) suggests that an AR(1) series (with $\phi \approx 0.7$) might provide a reasonable model for these residuals. (The form of the ACF for an AR(1) process was computed in Example 1.4.5.)

To explore the appropriateness of such a model, consider the points (y_1, y_2), $(y_2, y_3), \ldots, (y_{97}, y_{98})$ plotted in Figure 1.16. The graph does indeed suggest a linear relationship between y_t and y_{t-1}. Using simple least squares estimation to fit a straight line of the form $y_t = ay_{t-1}$, we obtain the model

$$Y_t = .791Y_{t-1} + Z_t, \tag{1.4.3}$$

where $\{Z_t\}$ is iid noise with variance $\sum_{t=2}^{98}(y_t - .791y_{t-1})^2/97 = .5024$. The sample ACF of the estimated noise sequence $z_t = y_t - .791y_{t-1}$, $t = 2, \ldots, 98$, is slightly

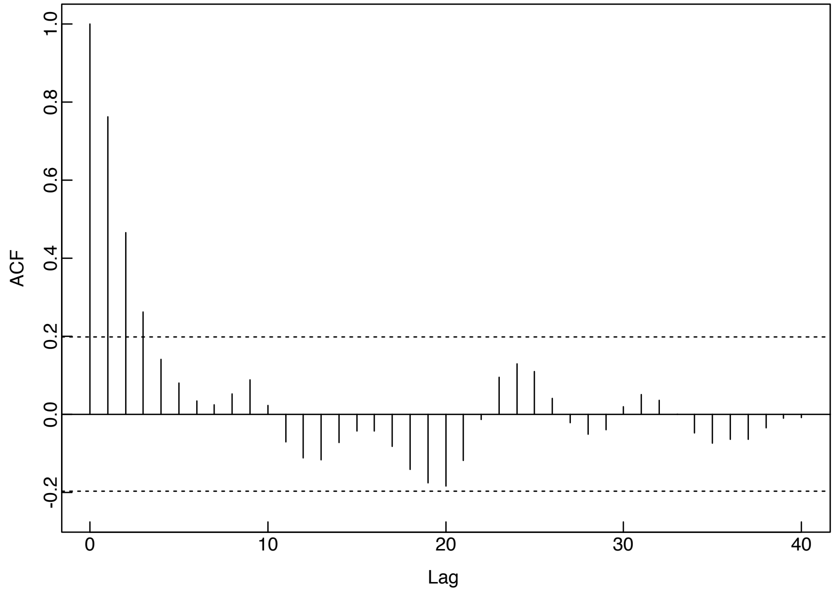

Figure 1-15
The sample autocorrelation function for the Lake Huron residuals of Figure 1.10 showing the bounds $\pm 1.96/\sqrt{n}$.

outside the bounds $\pm 1.96/\sqrt{97}$ at lag 1 ($\hat{\rho}(1) = .216$), but it is inside the bounds for all other lags up to 40. This check that the estimated noise sequence is consistent with the iid assumption of (1.4.3) reinforces our belief in the fitted model. More *goodness of fit* tests for iid noise sequences are described in Section 1.6. The estimated noise sequence $\{z_t\}$ in this example passes them all, providing further support for the model (1.4.3).

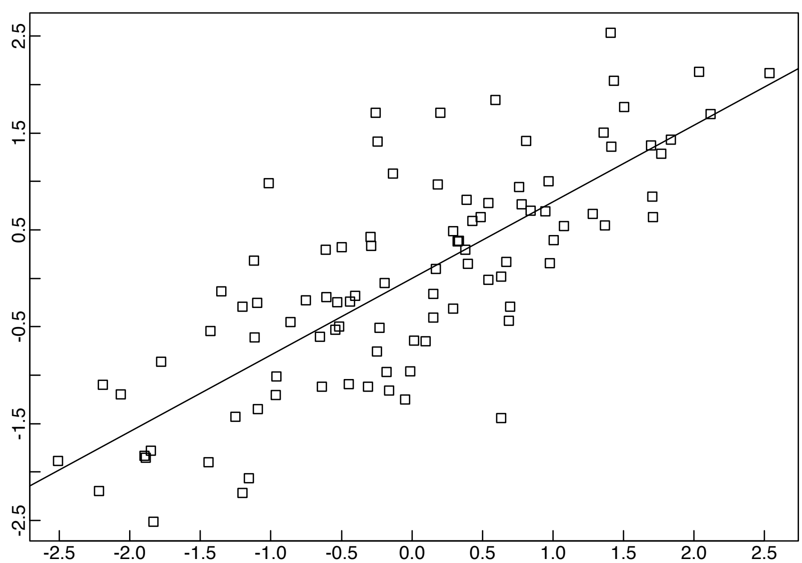

Figure 1-16
Scatter plot of (y_{t-1}, y_t), $t = 2, \ldots, 98$, for the data in Figure 1.10 showing the least squares regression line $y = .791x$.

A better fit to the residuals in equation (1.3.2) is provided by the second-order autoregression

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t, \tag{1.4.4}$$

where $\{Z_t\}$ is iid noise with variance σ^2. This is analogous to a linear model in which Y_t is regressed on the previous *two* values Y_{t-1} and Y_{t-2} of the time series. The least squares estimates of the parameters ϕ_1 and ϕ_2, found by minimizing $\sum_{t=3}^{98}(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2$, are $\hat{\phi}_1 = 1.002$ and $\hat{\phi}_2 = -.2834$. The estimate of σ^2 is $\hat{\sigma}^2 = \sum_{t=3}^{98}(y_t - \hat{\phi}_1 y_{t-1} - \hat{\phi}_2 y_{t-2})^2/96 = .4460$, which is approximately 11% smaller than the estimate of the noise variance for the AR(1) model (1.4.3). The improved fit is indicated by the sample ACF of the estimated residuals, $y_t - \hat{\phi}_1 y_{t-1} - \hat{\phi}_2 y_{t-2}$, which falls well within the bounds $\pm 1.96/\sqrt{96}$ for *all* lags up to 40.

1.5 Estimation and Elimination of Trend and Seasonal Components

The first step in the analysis of any time series is to plot the data. If there are any apparent discontinuities in the series, such as a sudden change of level, it may be advisable to analyze the series by first breaking it into homogeneous segments. If there are outlying observations, they should be studied carefully to check whether there is any justification for discarding them (as for example if an observation has been incorrectly recorded). Inspection of a graph may also suggest the possibility of representing the data as a realization of the process (the **classical decomposition** model)

$$X_t = m_t + s_t + Y_t, \tag{1.5.1}$$

where m_t is a slowly changing function known as a **trend component**, s_t is a function with known period d referred to as a **seasonal component**, and Y_t is a **random noise component** that is stationary in the sense of Definition 1.4.2. If the seasonal and noise fluctuations appear to increase with the level of the process, then a preliminary transformation of the data is often used to make the transformed data more compatible with the model (1.5.1). Compare, for example, the red wine sales in Figure 1.1 with the transformed data, Figure 1.17, obtained by applying a logarithmic transformation. The transformed data do not exhibit the increasing fluctuation with increasing level that was apparent in the original data. This suggests that the model (1.5.1) is more appropriate for the transformed than for the original series. In this section we shall assume that the model (1.5.1) is appropriate (possibly after a preliminary transformation of the data) and examine some techniques for estimating the components m_t, s_t, and Y_t in the model.

Our aim is to estimate and extract the deterministic components m_t and s_t in the hope that the residual or noise component Y_t will turn out to be a stationary time series. We can then use the theory of such processes to find a satisfactory probabilistic