2

Simple descriptive techniques

Statistical techniques for analysing time series range from relatively straightforward **descriptive** methods to sophisticated **inferential** techniques. This chapter introduces the former, which will often clarify the main properties of a given series and should generally be tried anyway before attempting more complicated procedures.

2.1 TYPES OF VARIATION

Traditional methods of time-series analysis are mainly concerned with decomposing the variation in a series into trend, seasonal variation, other cyclic changes, and the remaining 'irregular' fluctuations. This approach is not always the best but is particularly valuable when the variation is dominated by trend and/or seasonality. However, it is worth noting that the decomposition is generally not unique unless certain assumptions are made. Thus some sort of modelling, either explicit or implicit, may be involved in these descriptive techniques, and this demonstrates the blurred borderline between descriptive and inferential techniques.

The different sources of variation will now be described in more detail.

(a) *Seasonal effect*

Many time series, such as sales figures and temperature readings, exhibit variation which is annual in period. For example, unemployment is typically 'high' in winter but lower in summer. This yearly variation is easy to understand, and we shall see that it can be measured explicitly and/or removed from the data to give deseasonalized data.

(b) *Other cyclic changes*

Apart from seasonal effects, some time series exhibit variation at a fixed period due to some other physical cause. An example is daily variation in temperature. In addition some time series exhibit oscillations which do not have a fixed period but which are predictable to some extent. For example economic data are sometimes thought to be affected by business cycles with a period varying between about 5 and 7 years, although the existence of such business cycles is the subject of some controversy.

(c) *Trend*

This may be loosely defined as 'long-term change in the mean level'. A difficulty with this definition is deciding what is meant by 'long term'. For example, climatic variables sometimes exhibit cyclic variation over a very long time period such as 50 years. If one just had 20 years' data, this long-term oscillation would appear to be a trend, but if several hundred years' data were available, the long-term oscillation would be visible. Nevertheless in the short term it may still be more meaningful to think of such a long-term oscillation as a trend. Thus in speaking of a 'trend', we must take into account the number of observations available and make a subjective assessment of what is 'long term'. Granger (1966) defines 'trend in mean' as comprising all cyclic components whose wavelength exceeds the length of the observed time series.

(d) *Other irregular fluctuations*

After trend and cyclic variations have been removed from a set of data, we are left with a series of residuals, which may or may not be 'random'. We shall examine various techniques for analysing series of this type to see if some of the apparently irregular variation may be explained in terms of probability models, such as **moving average** or **autoregressive** models which will be introduced in Chapter 3. Alternatively we can see if any cyclic variation is still left in the residuals.

2.2 STATIONARY TIME SERIES

A mathematical definition of a stationary time series will be given later on. However, it is now convenient to introduce the idea of stationarity from an intuitive point of view. Broadly speaking a time series is said to be **stationary** if there is no systematic change in mean (no trend), if there is no systematic change in variance, and if strictly periodic variations have been removed.

Most of the probability theory of time series is concerned with stationary time series, and for this reason time-series analysis often requires one to turn a non-stationary series into a stationary one so as to use this theory. For example it may be of interest to remove the trend and seasonal variation from a set of data and then try to model the variation in the residuals by means of a stationary stochastic process. However, it is also worth stressing that the non-stationary components, such as the trend, may sometimes be of more interest than the stationary residuals.

2.3 THE TIME PLOT

After getting background information and carefully defining objectives, the first, and most important, step in any time-series analysis is to plot the observations against time. This graph should show up important features of the series such as trend, seasonality, outliers and discontinuities. The plot is vital, both to describe the data and to help in formulating a sensible model, and a variety of examples are given throughout this book.

Plotting a time series is not as easy as it sounds. The choice of scales, the size of the intercept, and the way that the points are plotted (e.g. as a continuous line or as separate dots) may substantially affect the way the plot 'looks', and so the analyst must exercise care and judgement. In addition, the usual rules for drawing 'good' graphs should be followed: a clear title must be given, units of measurement should be stated, and axes should be clearly labelled.

Nowadays, graphs are often produced by computers. Some are well done but other packages may produce rather poor graphs and the reader must be prepared to modify them if necessary. For example, one package plotted the data in Figure 5.1(a) with a vertical scale labelled unhelpfully from '.4000E+03' to '.2240E+04'! The hideous E notation was naturally changed before publication. Further advice and examples are given in Appendix D.

2.4 TRANSFORMATIONS

Plotting the data may suggest that it is sensible to consider transforming them, for example by taking logarithms or square roots. The three main reasons for making a transformation are as follows.

(a) *To stabilize the variance*

If there is a trend in the series and the variance appears to increase with the mean then it may be advisable to transform the data. In particular if the standard deviation is directly proportional to the mean, a logarithmic transformation is indicated.

(b) *To make the seasonal effect additive*

If there is a trend in the series and the size of the seasonal effect appears to increase with the mean then it may be advisable to transform the data so as to make the seasonal effect constant from year to year. The seasonal effect is then said to be additive. In particular if the size of the seasonal effect is directly proportional to the mean, then the seasonal effect is said to be multiplicative and a logarithmic transformation is appropriate to make the effect additive. However, this transformation will only stabilize the variance if the error term is **also** thought to be multiplicative (see Section 2.6), a point which is sometimes overlooked.

(c) *To make the data normally distributed*

Model building and forecasting are usually carried out on the assumption that the data are normally distributed. In practice this is not necessarily the case; there may for example be evidence of skewness in that there tend to be 'spikes' in the time plot which are all in the same direction (up or down). This effect can be difficult to eliminate and it may be necessary to assume a different 'error' distribution.

The logarithmic and square-root transformations are special cases of the class of transformations called the Box-Cox transformation. Given an observed time series $\{x_t\}$ and a transformation parameter λ, the transformed series is given by

$$y_t = \begin{cases} (x_t^{\lambda} - 1)/\lambda & \lambda \neq 0 \\ \log x_t & \lambda = 0 \end{cases}$$

This is effectively just a power transformation when $\lambda \neq 0$, as the constants are introduced to make y_t a continuous function of λ at the value $\lambda = 0$. The 'best' value of λ can be guesstimated, or alternatively estimated by a proper inferential procedure, such as maximum likelihood.

It is instructive to note that Nelson and Granger (1979) found little improvement in forecast performance when a general Box-Cox transformation was tried on a number of series. There are problems in practice with transformations in that a transformation which, say, makes the seasonal effect additive may fail to stabilize the variance and it may be impossible to achieve all requirements at the same time. In any case a model constructed for the transformed data may be less than helpful. For example, forecasts produced by the transformed model may have to be 'transformed back' in order to be of use and this can introduce biasing effects. My personal preference nowadays is to avoid transformations wherever possible except where the transformed variable has a direct physical interpretation. For example, when percentage increases are of interest, then taking logarithms makes sense (see Example D.3). Further general remarks on transformations are given by Granger and Newbold (1986, Section 10.5).

2.5 ANALYSING SERIES WHICH CONTAIN A TREND

In Section 2.1 we loosely defined trend as a 'long-term change in the mean level'. It is much more difficult to give a precise definition and different authors may use the term in different ways. The simplest trend is the familiar 'linear trend + noise', for which the observation at time t is a random variable X_t given by

$$X_t = \alpha + \beta t + \varepsilon_t \qquad (2.1)$$

where α, β are constants and ε_t denotes a random error term with zero mean.

The mean level at time t is given by $m_t = (\alpha + \beta t)$; this is sometimes called 'the trend term'. Other writers prefer to describe the slope β as the trend; the trend is then the **change** in the mean level per unit time. It is usually clear from the context which meaning is intended.

The trend in equation (2.1) is a deterministic function of time and is sometimes called a **global** linear trend. This is generally unrealistic, and there is now more emphasis on **local** linear trends where the parameters α and β in equation (2.1) are allowed to evolve through time. Alternatively, the trend may be of non-linear form such as quadratic growth. Exponential growth can be particularly difficult to handle, even if logarithms are taken to transform the trend to a linear form.

The analysis of a time series which exhibits trend depends on whether one wants to (a) measure the trend and/or (b) remove the trend in order to analyse local fluctuations. It also depends on whether the data exhibit seasonality (see Section 2.6). With seasonal data, it is a good idea to start by calculating successive yearly averages as these will provide a simple description of the underlying trend. An approach of this type is sometimes perfectly adequate, particularly if the trend is fairly small, but sometimes a more sophisticated approach is desired and then the following techniques can be considered.

2.5.1 Curve fitting

A traditional method of dealing with non-seasonal data which contain a trend, particularly yearly data, is to fit a simple function such as a polynomial curve (linear, quadratic etc.), a Gompertz curve or a logistic curve (e.g. see Levenbach and Reuter, 1976; Meade, 1984). The Gompertz curve is given by

$$\log x_t = a + br^t$$

where a, b, r are parameters with $0 < r < 1$, while the logistic curve is given by

$$x_t = a/(1 + b\,e^{-ct})$$

Both these curves are S-shaped and approach an asymptotic value as $t \to \infty$. Fitting the curves to data may lead to non-linear simultaneous equations.

For all curves of this type, the fitted function provides a measure of the trend, and the residuals provide an estimate of local fluctuations, where the residuals are the differences between the observations and the corresponding values of the fitted curve.

2.5.2 Filtering

A second procedure for dealing with a trend is to use a **linear filter** which converts one time series, $\{x_t\}$, into another, $\{y_t\}$, by the linear operation

$$y_t = \sum_{r=-q}^{+s} a_r x_{t+r}$$

where $\{a_r\}$ is a set of weights. In order to smooth out local fluctuations and estimate the local mean, we should clearly choose the weights so that $\Sigma a_r = 1$, and then the operation is often referred to as a **moving average**. Moving averages are discussed in detail by Kendall, Stuart and Ord (1983, Chapter 46), and we will only provide a brief introduction. Moving averages are often symmetric with $s = q$ and $a_j = a_{-j}$. The simplest example of a symmetric smoothing filter is the simple moving average, for which $a_r = 1/(2q+1)$ for $r = -q, \ldots, +q$, and the smoothed value of x_t is given by

$$\text{Sm}(x_t) = \frac{1}{2q+1} \sum_{r=-q}^{+q} x_{t+r}$$

The simple moving average is not generally recommended by itself for measuring trend, although it can be useful for removing seasonal variation. Another example is provided by the case where the $\{a_r\}$ are successive terms in the expansion of $(\frac{1}{2}+\frac{1}{2})^{2q}$. Thus when $q = 1$, the weights are $a_{-1} = a_{+1} = \frac{1}{4}$, $a_0 = \frac{1}{2}$. As q gets large, the weights approximate to a normal curve.

A third example is Spencer's 15-point moving average, which is used for smoothing mortality statistics to get life tables (Tetley, 1946). This covers 15 consecutive points with $q = 7$, and the symmetric weights are

$$\frac{1}{320}[-3, -6, -5, 3, 21, 46, 67, 74, \ldots]$$

A fourth example, called the Henderson moving average, is described by Kenny and Durbin (1982) and is finding increased use. This average aims to follow a cubic polynomial trend without distortion, and the choice of q depends on the degree of irregularity. The symmetric 9-term moving average, for example, is given by

$$[-0.041, -0.010, 0.119, 0.267, 0.330, \ldots]$$

The general idea is to fit a polynomial curve, not to the whole series, but to different parts. For example a polynomial fitted to the first $(2q+1)$ data points can be used to determine the interpolated value at the middle of the range where $t = (q+1)$, and the procedure can then be repeated using the data from $t = 2$ to $t = (2q+2)$, and so on. A related idea is to use the class of piecewise polynomials called splines (e.g. Wegman and Wright, 1983).

Whenever a symmetric filter is chosen, there is likely to be an **end-effects** problem (e.g. Kendall, Stuart and Ord, 1983, Section 46.11), since $\text{Sm}(x_t)$ is calculated for $t = (q+1)$ to $t = N-q$. In some situations this is not important, but in other situations it is particularly important to get smoothed values up to

$t=N$. The analyst can project the smoothed values by eye, or by some further smoothing procedure, or, alternatively, use an asymmetric filter which only involves present and past values of x_t. For example, the popular technique known as exponential smoothing (see Section 5.2.2) effectively assumes that

$$\mathrm{Sm}(x_t)=\sum_{j=0}^{\infty}\alpha(1-\alpha)^j x_{t-j}$$

where α is a constant such that $0<\alpha<1$. Here we note that the weights $a_j=\alpha(1-\alpha)^j$ decrease geometrically with j.

Having estimated the trend, we can look at the local fluctuations by examining

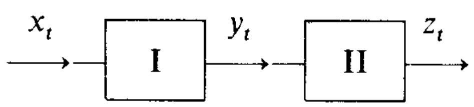
$$\begin{aligned}\mathrm{Res}(x_t)&=\text{residual from smoothed value}\\&=x_t-\mathrm{Sm}(x_t)\\&=\sum_{r=-q}^{+s} b_r x_{t+r}\end{aligned}$$

This is also a linear filter, and if $\Sigma a_r=1$, then $\Sigma b_r=0$, $b_0=1-a_0$, and $b_r=-a_r$ for $r\neq 0$.

How do we choose the appropriate filter? The answer to this question really requires considerable experience plus a knowledge of the frequency aspects of time-series analysis which will be discussed in later chapters. As the name implies, filters are usually designed to produce an output with emphasis on variation at particular frequencies. For example, to get smoothed values we want to remove the local fluctuations which constitute what is called the high-frequency variation. In other words we want what is called a **low-pass** filter. To get $\mathrm{Res}(x_t)$, we want to remove the long-term fluctuations or the low-frequency variation. In other words we want what is called a **high-pass** filter. The Slutsky (or Slutsky-Yule) effect is related to this problem. Slutsky showed that by operating on a completely random series with both averaging and differencing procedures one could induce sinusoidal variation in the data, and he went on to suggest that apparently periodic behaviour in some economic time series might be accounted for by the smoothing procedures used to form the data. We will return to this question later.

Filters in series

Very often a smoothing procedure is carried out in two or more stages – so that one has in effect several linear filters in series. For example two filters in series may be represented as follows:



Filter I with weights $\{a_{j1}\}$ acts on $\{x_t\}$ to produce $\{y_t\}$. Filter II with weights $\{a_{j2}\}$ acts on $\{y_t\}$ to produce $\{z_t\}$. Now

$$\begin{aligned} z_t &= \sum_j a_{j2} y_{t+j} \\ &= \sum_j a_{j2} \sum_r a_{r1} x_{t+j+r} \\ &= \sum_j c_j x_{t+j} \end{aligned}$$

where

$$c_j = \sum_r a_{r1} a_{(j-r)2}$$

are the weights for the overall filter. The weights $\{c_j\}$ are obtained by a procedure called **convolution**, and we write

$$\{c_j\} = \{a_{r1}\} * \{a_{r2}\}$$

where $*$ represents convolution. For example, the filter $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ may be written as

$$(\tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4}) = (\tfrac{1}{2}, \tfrac{1}{2}) * (\tfrac{1}{2}, \tfrac{1}{2})$$

Given a series $x_1, \ldots, x_N$, this smoothing procedure is best done in three stages by adding successive pairs of observations twice and then dividing by 4, as follows:

Observations	*Stage I*	*Stage II*	*Stage III*
x_1			
	$x_1 + x_2$		
x_2		$x_1 + 2x_2 + x_3$	$(x_1 + 2x_2 + x_3)/4$
	$x_2 + x_3$		.
x_3		$x_2 + 2x_3 + x_4$	.
	$x_3 + x_4$		.
x_4		$x_3 + 2x_4 + x_5$	.
	$x_4 + x_5$	.	.
x_5		.	.
	$x_5 + x_6$	.	.
x_6	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.

The Spencer 15-point moving average is actually a convolution of four filters, namely

$$(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}) * (\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}) * (\tfrac{1}{5}, \tfrac{1}{5}, \tfrac{1}{5}, \tfrac{1}{5}, \tfrac{1}{5}) * (-\tfrac{3}{4}, \tfrac{3}{4}, 1, \tfrac{3}{4}, -\tfrac{3}{4})$$

2.5.3 Differencing

A special type of filtering, which is particularly useful for removing a trend, is simply to difference a given time series until it becomes stationary. This method is an integral part of the procedures advocated by Box and Jenkins (1970). For non-seasonal data, first-order differencing is usually sufficient to attain apparent stationarity, so that the new series $\{y_1, \ldots, y_{N-1}\}$ is formed from the original series $\{x_1, \ldots, x_N\}$ by

$$y_t = x_{t+1} - x_t = \nabla x_{t+1}$$

First-order differencing is widely used. Occasionally second-order differencing is required using the operator ∇^2, where

$$\nabla^2 x_{t+2} = \nabla x_{t+2} - \nabla x_{t+1} = x_{t+2} - 2x_{t+1} + x_t$$

2.6 ANALYSING SERIES WHICH CONTAIN SEASONAL VARIATION

In Section 2.1 we introduced seasonal variation which is generally annual in period, while Section 2.4 distinguished between additive seasonality, which is constant from year to year, and multiplicative seasonality. Three seasonal models in common use are

$$\text{A} \quad X_t = m_t + S_t + \varepsilon_t$$

$$\text{B} \quad X_t = m_t S_t + \varepsilon_t$$

$$\text{C} \quad X_t = m_t S_t \varepsilon_t$$

where m_t is the deseasonalized mean level at time t, S_t is the seasonal effect at time t, and ε_t is the random error.

Model A describes the additive case, while models B and C both involve multiplicative seasonality. In model C the error term is also multiplicative, and a logarithmic transformation will turn this into a (linear) additive model which may be easier to handle. The time plot should be examined to see which model is likely to give the better description. The seasonal indices $\{S_t\}$ are usually assumed to change slowly through time so that $S_t \simeq S_{t-s}$, where s is the number of observations per year. The indices are usually normalized so that they sum to zero in the additive case, or average to one in the multiplicative case. Difficulties arise in practice if the seasonal and/or error terms are not exactly multiplicative or additive. For example the seasonal effect may increase with the mean level but not at such a fast rate so that it is somewhere 'in between' being multiplicative or additive. A mixed additive-multiplicative seasonal model is described by Durbin and Murphy (1975).

The analysis of time series which exhibit seasonal variation depends on whether one wants to (a) measure the seasonal effect and/or (b) eliminate seasonality. For series showing little trend, it is usually adequate to estimate the seasonal effect for a particular period (e.g. January) by finding the average of each January observation minus the corresponding yearly average in the additive case, or the January observation divided by the yearly average in the multiplicative case.

For series which do contain a substantial trend, a more sophisticated approach may be required. With monthly data, the commonest way of eliminating the seasonal effect is to calculate

$$\mathrm{Sm}(x_t) = \frac{\frac{1}{2}x_{t-6} + x_{t-5} + x_{t-4} + \cdots + x_{t+5} + \frac{1}{2}x_{t+6}}{12}$$

Note that the sum of the coefficients is 1. A simple moving average cannot be used as this would span 12 months and would not be centred on an integer value of t. A simple moving average over 13 months cannot be used, as this would give twice as much weight to the month appearing at both ends. For quarterly data, the seasonal effect can be eliminated by calculating

$$\mathrm{Sm}(x_t) = \frac{\frac{1}{2}x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2}x_{t+2}}{4}$$

For 4-weekly data, one **can** use a simple moving average over 13 successive observations. The seasonal effect can be estimated by calculating $x_t - \mathrm{Sm}(x_t)$ or $x_t/\mathrm{Sm}(x_t)$ depending on whether the seasonal effect is thought to be additive or multiplicative. A check should be made that the seasonals are reasonably stable, and then the average monthly (or quarterly etc.) effects can be calculated.

A seasonal effect can also be eliminated by differencing (see Sections 4.6, 5.2.4; Box and Jenkins, 1970). For example with monthly data one can employ the operator ∇_{12} where

$$\nabla_{12}x_t = x_t - x_{t-12}$$

Alternative methods of seasonal adjustment are reviewed by Butter and Fase (1991) and Hylleberg (1992). These include the widely used X-11 method which employs a series of linear filters (e.g. see Wallis, 1982; Kendall, Stuart and Ord, 1983, Section 46.41). The possible presence of calendar effects should also be considered (Cleveland and Devlin, 1982; Cleveland, 1983). For example if Easter falls in March one year, rather than April, then this may alter the seasonal effect on sales for both months.

2.7 AUTOCORRELATION

An important guide to the properties of a time series is provided by a series of quantities called sample autocorrelation coefficients, which measure the correlation between observations at different distances apart. These coefficients often provide insight into the probability model which generated the data. We assume that the reader is familiar with the ordinary correlation coefficient, namely that given N pairs of observations on two variables x and y, the correlation coefficient is given by

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2]}} \tag{2.2}$$

A similar idea can applied to time series to see if successive observations are correlated.

Given N observations $x_1, \ldots, x_N$, on a discrete time series we can form $N-1$ pairs of observations, namely (x_1, x_2), $(x_2, x_3), \ldots,$ (x_{N-1}, x_N). Regarding the first observation in each pair as one variable, and the second observation as a second variable, the correlation coefficient between x_t and x_{t+1} is given by

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\sqrt{\left[\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})^2 \sum_{t=1}^{N-1} (x_{t+1} - \bar{x}_{(2)})^2\right]}} \tag{2.3}$$

by analogy with equation (2.2), where

$$\bar{x}_{(1)} = \sum_{t=1}^{N-1} x_t/(N-1)$$

is the mean of the first $N-1$ observations and

$$\bar{x}_{(2)} = \sum_{t=2}^{N} x_t/(N-1)$$

is the mean of the last $N-1$ observations. As the coefficient given by equation (2.3) measures correlation between successive observations it is called an **autocorrelation** coefficient or serial correlation coefficient.

Equation (2.3) is rather complicated, and so, as $\bar{x}_{(1)} \simeq \bar{x}_{(2)}$, it is usually approximated by

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{(N-1) \sum_{t=1}^{N} (x_t - \bar{x})^2 / N} \tag{2.4}$$

where $\bar{x} = \sum_{t=1}^{N} x_t/N$ is the overall mean. Some authors also drop the factor $N/(N-1)$, which is close to one for large N, to give the even simpler formula

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2} \tag{2.5}$$

and this is the form that will be used in this book.

In a similar way we can find the correlation between observations a distance k apart, which is given by

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2} \tag{2.6}$$

This is called the autocorrelation coefficient at lag k.

In practice the autocorrelation coefficients are usually calculated by computing the series of autocovariance coefficients, $\{c_k\}$, which we define by analogy with the usual covariance formula as

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \tag{2.7}$$

This is the autocovariance coefficient at lag k.

We then compute

$$r_k = c_k / c_0 \tag{2.8}$$

for $k = 1, 2, \ldots, m$, where $m < N$. There is often little point in calculating r_k for values of k greater than about $N/4$.

Note that some authors prefer to use

$$c_k = \frac{1}{N-k} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$$

rather than equation 2.7, but there is little difference for large N (see Section 4.1).

2.7.1 The correlogram

A useful aid in interpreting a set of autocorrelation coefficients is a graph called a correlogram in which r_k is plotted against the lag k. Examples are given in Figures 2.1–2.3. A visual inspection of the correlogram is often very helpful.

2.7.2 Interpreting the correlogram

Interpreting the meaning of a set of autocorrelation coefficients is not always easy. Here we offer some general advice.

(a) *A random series*

If a time series is completely random, then for large N, $r_k \simeq 0$ for all non-zero values of k. In fact we will see later that for a random time series r_k is approximately N(0, $1/N$), so that, if a time series is random, 19 out of 20 of the values of r_k can be expected to lie between $\pm 2/\sqrt{N}$. However if one plots say the first 20 values of r_k then one can expect to find one 'significant' value on average even when the time series really is random. This spotlights one of the difficulties in interpreting the correlogram, in that a large number of coefficients is quite likely to contain one (or more) 'unusual' results, even when no real effects are present. (See also Section 4.1.)

(b) *Short-term correlation*

Stationary series often exhibit short-term correlation characterized by a fairly large value of r_1 followed by a few further coefficients which, while greater than zero, tend to get successively smaller. Values of r_k for longer lags tend to be approximately zero. An example of such a correlogram is shown in Figure 2.1. A time series which gives rise to such a correlogram is one for which an observation above the mean tends to be followed by one or more further observations above the mean, and similarly for observations below the mean.

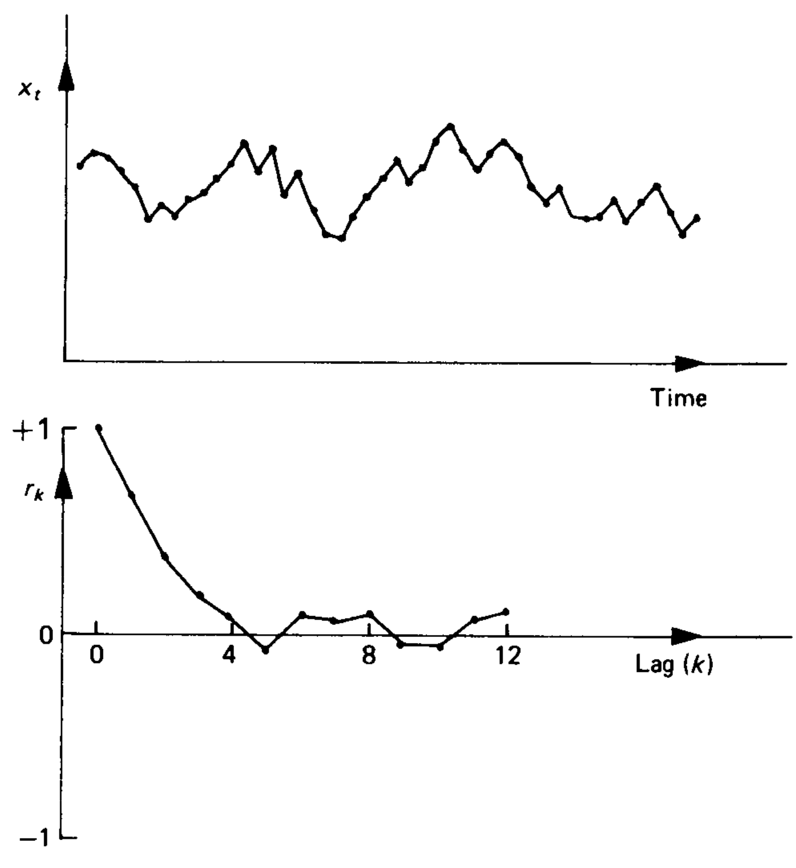


Figure 2.1 A time series showing short-term correlation together with its correlogram.

(c) *Alternating series*

If a time series has a tendency to alternate, with successive observations on different sides of the overall mean, then the correlogram also tends to alternate. The value of r_1 will be negative. However the value of r_2 will be positive, as observations at lag 2 will tend to be on the same side of the mean. A typical alternating time series together with its correlogram is shown in Figure 2.2.

(d) *Non-stationary series*

If a time series contains a trend, then the values of r_k will not come down to zero except for very large values of the lag. This is because an observation on one side of the overall mean tends to be followed by a large number of further observations on the same side of the mean because of the trend. A typical non-stationary time series together with its correlogram is shown in Figure 2.3. Little can be inferred from a correlogram of this type as the trend dominates all other features. In fact the sample autocorrelation function, $\{r_k\}$, is only meaningful for **stationary** time series (see Chapters 3 and 4) and so any trend should be removed before calculating $\{r_k\}$.

(e) *Seasonal fluctuations*

If a time series contains a seasonal fluctuation, then the correlogram will also exhibit an oscillation at the same frequency. For example with monthly observations, r_6 will be 'large' and negative, while r_{12} will be 'large' and

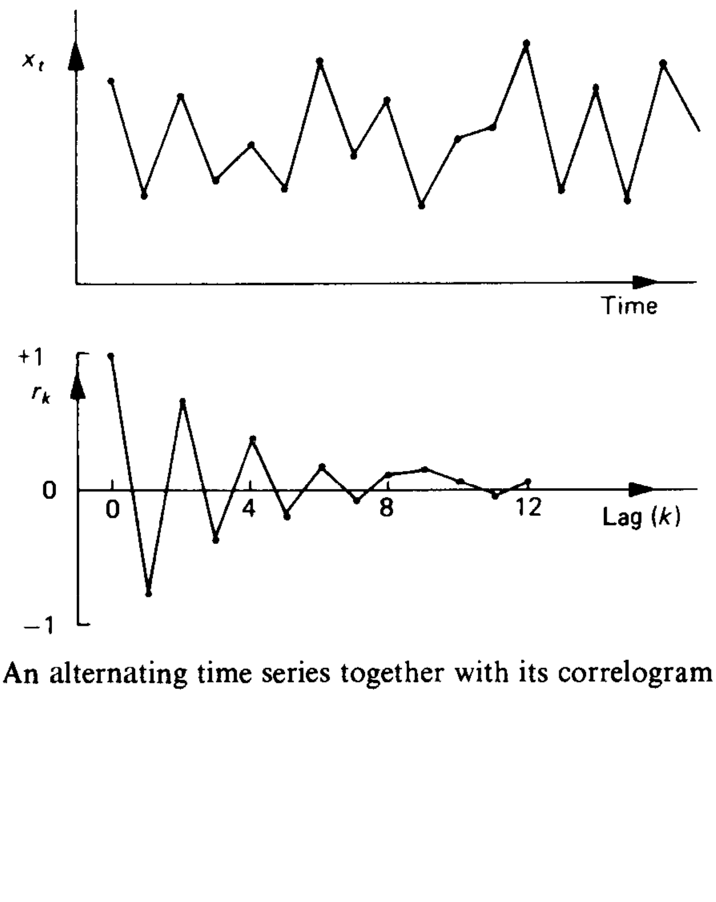


Figure 2.2 An alternating time series together with its correlogram.

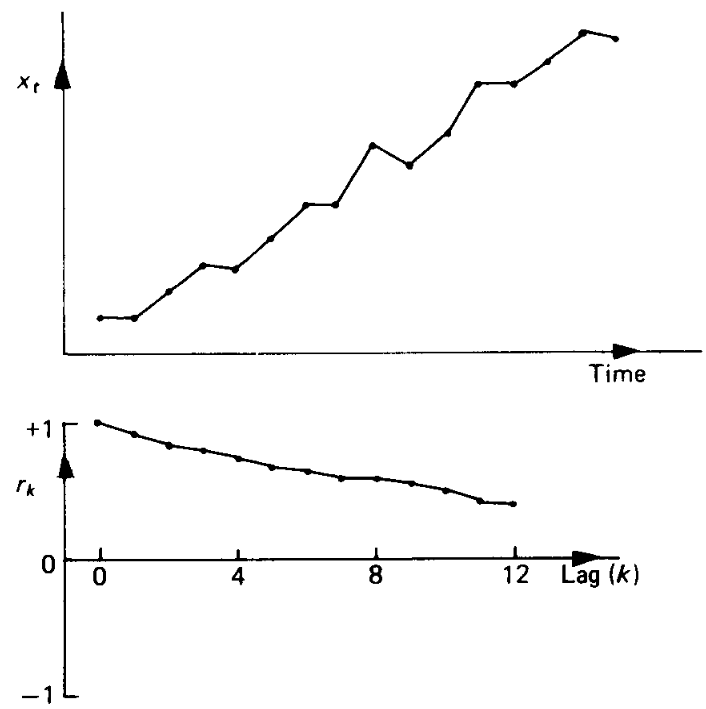


Figure 2.3 A non-stationary time series together with its correlogram.

positive. In particular if x_t follows a sinusoidal pattern, then so does r_k. For example, if

$$x_t = a \cos t\omega$$

where a is a constant and the frequency ω is such that $0 < \omega < \pi$, then it can be shown (see Exercise 2.3) that

$$r_k \simeq \cos k\omega \qquad \text{for large } N$$

Figure 2.4(a) shows the correlogram of the monthly air temperature data shown in Figure 1.2. The sinusoidal pattern of the correlogram is clearly evident, but for seasonal data of this type the correlogram provides little extra information as the seasonal pattern is clearly evident in the time plot of the data.

If the seasonal variation is removed from seasonal data, then the correlogram may provide useful information. The seasonal variation was removed from the air temperature data by the simple procedure of calculating the 12 monthly averages and subtracting the appropriate one from each individual observation. The correlogram of the resulting series (Figure 2.4(b)) shows that the first three coefficients are significantly different from zero. This indicates short-term correlation in that a month which is say colder than the average for that month will tend to be followed by one or two further months which are colder than average.

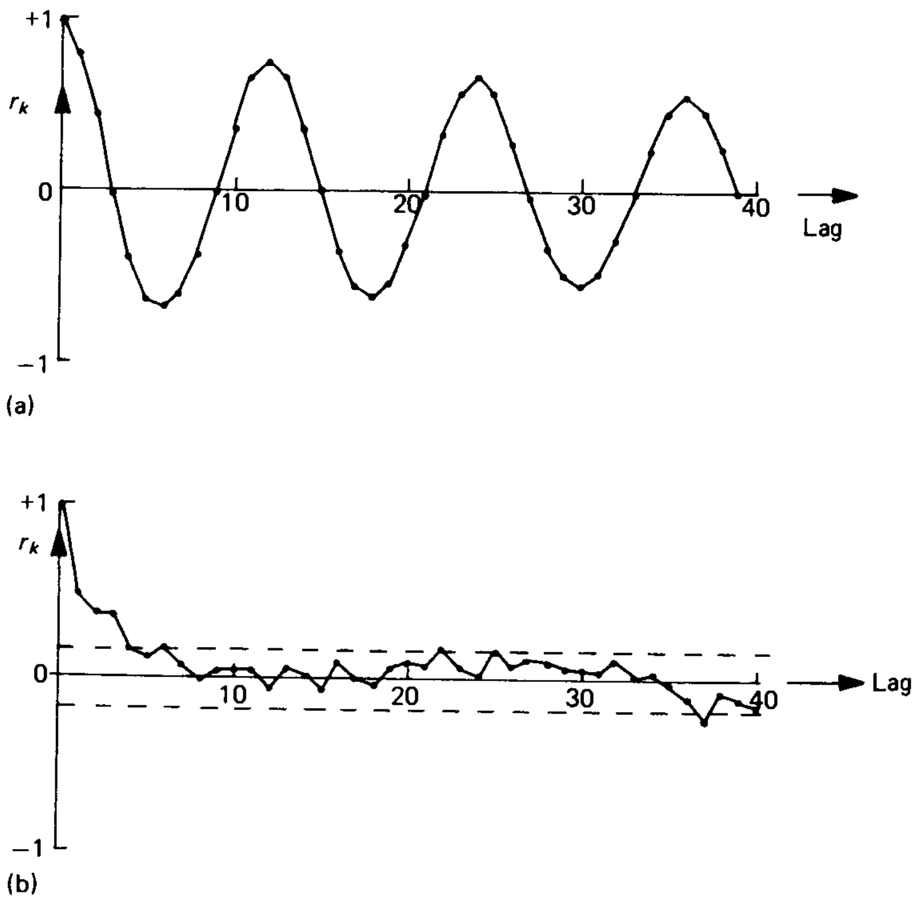


Figure 2.4 The correlogram of monthly observations on air temperature at Recife: (a) for the raw data; (b) for the seasonally adjusted data. The dotted lines in (b) are at $\pm 2/\sqrt{N}$. Values outside these lines are significantly different from zero.

(f) *Outliers*

If a time series contains one or more outliers, the correlogram may be seriously affected and it may be advisable to adjust outliers in some way before starting the formal analysis. For example, if there is one outlier in the time series and it is not adjusted, then the plot of x_t against x_{t+k} will contain **two** 'extreme' points which will tend to depress the sample correlation coefficients towards zero. If there are two outliers this effect is even more noticeable, except when the lag equals the distance between the outliers when a spuriously large correlation may occur.

(g) *General remarks*

Clearly considerable experience is required in interpreting autocorrelation coefficients. In addition we need to study the probability theory of stationary series and discuss the classes of model which may be appropriate. We must also discuss the sampling properties of r_k. These topics will be covered in the next two chapters and we shall then be in a better position to interpret the correlogram of a given time series.

2.8 OTHER TESTS OF RANDOMNESS

In most cases, a visual examination of the graph of a time series is enough to see that it is **not** random. However it is occasionally desirable to test a stationary time series for 'randomness'. In other words one wants to test if $x_1, \ldots, x_N$ could have arisen in that order by chance by taking a simple random sample size N from a population with unknown characteristics. A variety of tests exist for this purpose and they are described by Kendall, Stuart and Ord (1983, Section 45.15). For example one can examine the number of times there is a local maximum or minimum in the time series. A local maximum is defined to be any observation x_t such that $x_t > x_{t-1}$ and $x_t > x_{t+1}$. A converse definition applies to local minima. If the series really is random one can work out the expected number of turning points and compare it with the observed value. Tests of this type will not be described here, as I have always found it more convenient to simply examine the correlogram (and possibly the spectral density function) of a given time series to see if it is random.

Having fitted a model to a non-random series, one often wants to see if the residuals are random. Testing residuals for randomness is a somewhat different problem and will be discussed in Section 4.7.

EXERCISES

2.1 The following data show the sales of company X in successive 4-week periods over 1967–1970.

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
1967	153	189	221	215	302	223	201	173	121	106	86	87	108
1968	133	177	241	228	283	255	238	164	128	108	87	74	95
1969	145	200	187	201	292	220	233	172	119	81	65	76	74
1970	111	170	243	178	248	202	163	139	120	96	95	53	94

(a) Plot the data.
(b) Assess the trend and seasonal effects.

2.2 Sixteen successive observations on a stationary time series are as follows:
1.6, 0.8, 1.2, 0.5, 0.9, 1.1, 1.1, 0.6, 1.5, 0.8, 0.9, 1.2, 0.5, 1.3, 0.8, 1.2

(a) Plot the observations.
(b) Looking at the graph, guess an approximate value for the autocorrelation coefficient at lag 1.
(c) Plot x_t against x_{t+1}, and again try to guess the value of r_1.
(d) Calculate r_1.