CHAPTER 4

NONSTATIONARY NONSEASONAL MODELS

4.1 INTRODUCTION

When considering annual hydrological and other natural time series of moderate lengths (perhaps a few hundred years), it is often reasonable to assume that a stationary model can adequately model the data. For example, in Section 10.6.2, stationary ARMA models are fitted to 23 time series which are measured from six different types of natural phenomena that vary in length from 96 to 1164 years. The ability to detect statistical characteristics of a time series which change significantly over time may only become possible when the yearly records cover a sufficiently long time horizon. For example, if past climatic records were available or could be constructed for a given location in North America, the results would probably support the hypothesis of climatic nonstationarity over a long time span. Certainly, as the ice sheets advanced and retreated over the North American continent during the past one hundred thousand years, average annual temperatures and other climatic factors changed significantly over time.

Some types of annual time series which are studied in water resources engineering, could be *nonstationary* even over a short time interval. For instance, the average annual cost of hydroelectric power and the annual consumption of water of an expanding metropolis constitute two time series which increase in magnitude over time. In general, time series that reflect the socio - economic aspects of water resources planning may be nonstationary over any time interval being considered.

When modelling nonstationary data, a common procedure is to first remove the nonstationarity by using a suitable technique. Following this, a stationary model can be fit to the resulting stationary time series. This general approach is used in this chapter for nonseasonal models and also in Chapter 12 for a certain class of seasonal models.

4.2 EXPLOSIVE NONSTATIONARITY

If an ARMA(p,q) process is *stationary*, all of the roots of the characteristic equation $\phi(B)=0$ must lie outside the unit circle (see Section 3.2.2). Consequently, when a process is *nonstationary* at least one of the roots of $\phi(B)=0$ must lie on or within the unit circle. If at least one root is inside the unit circle, the process is said to possess *explosive nonstationarity*. When none of the roots are within the unit circle but at least one of the roots lies on the unit circle, this is referred to as *homogeneous nonstationarity*.

For the case of an ARMA(1,1) process in [3.4.1], it is necessary that the root ϕ_1^{-1} of $(1-\phi_1 B)=0$ possess an absolute magnitude which is greater than unity or, equivalently, $|\phi_1| < 1$ in order to have stationarity. On the other hand, when a process with one AR and one MA parameter is nonstationary, the root ϕ_1^{-1} must lie either on or inside the unit circle and hence $|\phi_1| \geq 1$. Suppose, for example, a model is given as

$$z_t - \phi_1 z_{t-1} = a_t - 0.70a_{t-1}$$

or, equivalently,

$$(1-\phi_1 B)z_t = (1-0.70B)a_t \qquad [4.2.1]$$

where a_t is normally independently distributed with a mean of zero and a variance of one [i.e., NID(0,1)]. If $\phi_1 = 1.1$, the root of $1-1.1B=0$ is $1/1.1$ and hence the process possesses explosive nonstationarity. When z_1 is assigned a value of, say, 100, z_t can be simulated using

$$z_2 - 1.1z_1 = a_2 - 0.70a_1 \qquad [4.2.2]$$

where the a_t's are randomly generated on a computer (see Section 9.2). By substituting $t = 3,4,\ldots,20$, into [4.2.1], a sequence of 20 synthetic data points can be obtained where $z_1 = 100$. A plot of 20 simulated values for z_t is shown in Figure 4.2.1. Notice how the series increases greatly over time due to the fact that the root of the characteristic equation lies just inside the unit circle. If ϕ_1 is given a value of 1.5, a simulated series can be even more explosive than that presented in Figure 4.2.1. The simulated sequence of 20 values in Figure 4.2.2 was obtained using [4.2.1] with $\phi_1 = 1.5$ and a starting value of $z_1 = 100$. In that figure, the series increases exponentially with time and the last synthetic data point has a magnitude which is close to 24,000.

4.3 HOMOGENEOUS NONSTATIONARITY

The ARIMA (autoregressive integrated moving average) model is defined in the next subsection for modelling an annual time series possessing homogeneous nonstationarity. As explained in Section 4.3.2, the theoretical ACF for an ARIMA model containing nonstationarity dies off slowly. Consequently, if the sample ACF of a given annual time series attenuates, this may indicate the presence of nonstationarity and the need to fit an ARIMA model to the series. Three kinds of time series are employed in Section 4.3.3 to demonstrate how the sample ACF dies off slowly for a nonstationary series and how to fit an ARIMA model to each series. Finally, Section 4.3.4 describes three equivalent formulations of the ARIMA model.

4.3.1 Autoregressive Integrated Moving Average Model

When at least one of the roots of the characteristic equation lies on the unit circle but none of the roots are inside the unit circle, this produces a milder type of nonstationarity than the explosive case. This is referred to as *homogeneous nonstationarity* because, except for a local level and slope, often portions of a simulated series will be similar to other sections. For example, when ϕ_1 is set equal to unity in [4.2.1] the model becomes

$$z_t - z_{t-1} = a_t - 0.70a_{t-1}$$

or equivalently

$$(1-B)z_t = (1-0.70B)a_t \qquad [4.3.1]$$

where a_t is NID(0,1). Notice that the single root of $(1-B)=0$ is of course unity and hence the model possesses homogeneous nonstationarity. By choosing a starting value of $z_1 = 100$ and

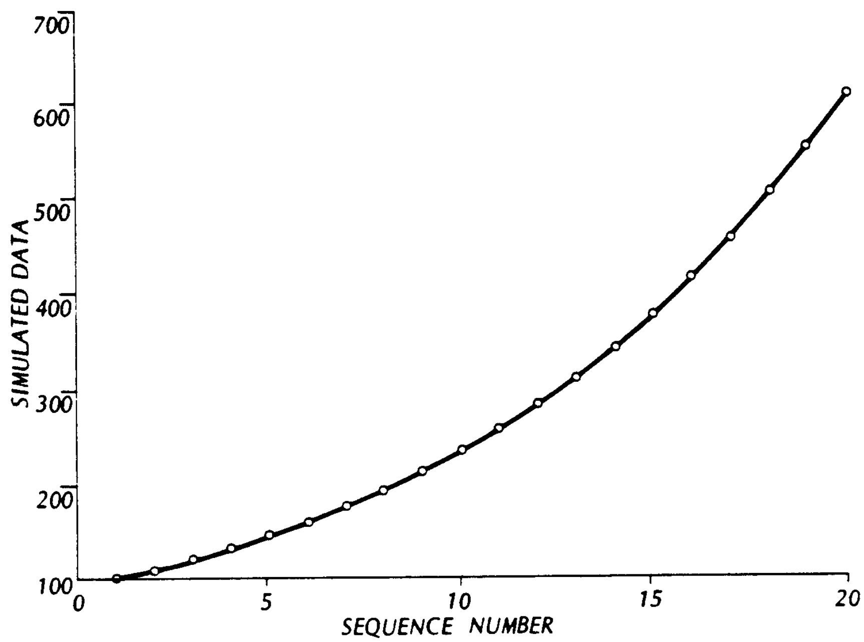

Figure 4.2.1. Simulated data for the model in [4.2.1] with $\phi_1 = 1.1$ and $z_1 = 100$.

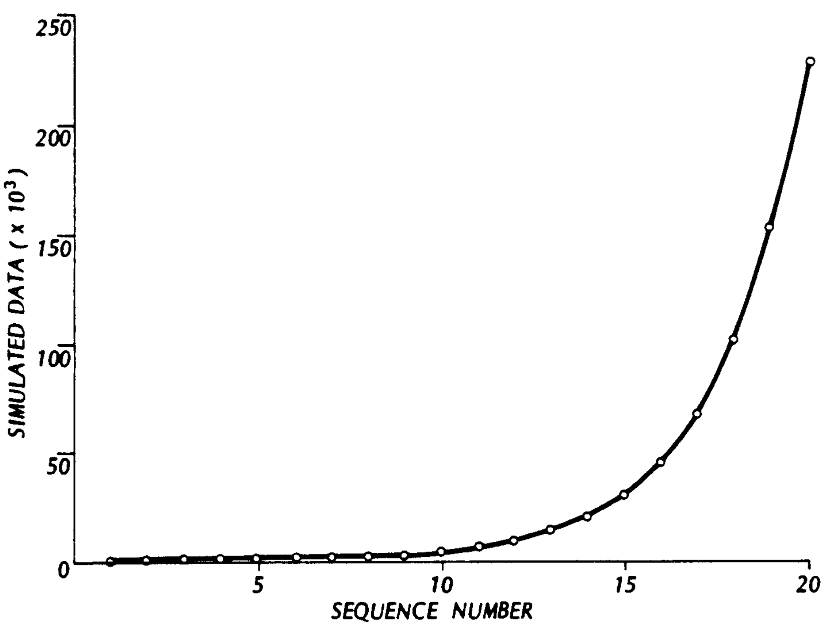

Figure 4.2.2. Simulated data for the model in [4.2.1] with $\phi_1 = 1.5$ and $z_1 = 100$.

having the computer generate the a_t's, a sequence of 20 simulated values can be obtained as shown in Figure 4.3.1. It can be seen that this realization behaves in a much more restrained fashion than those shown in Figures 4.2.1 and 4.2.2. This kind of behaviour is typical of many types of socio - economic series which are encountered in practical applications and therefore the modelling of homogeneous nonstationarity has received widespread attention (Box and Jenkins, 1976).

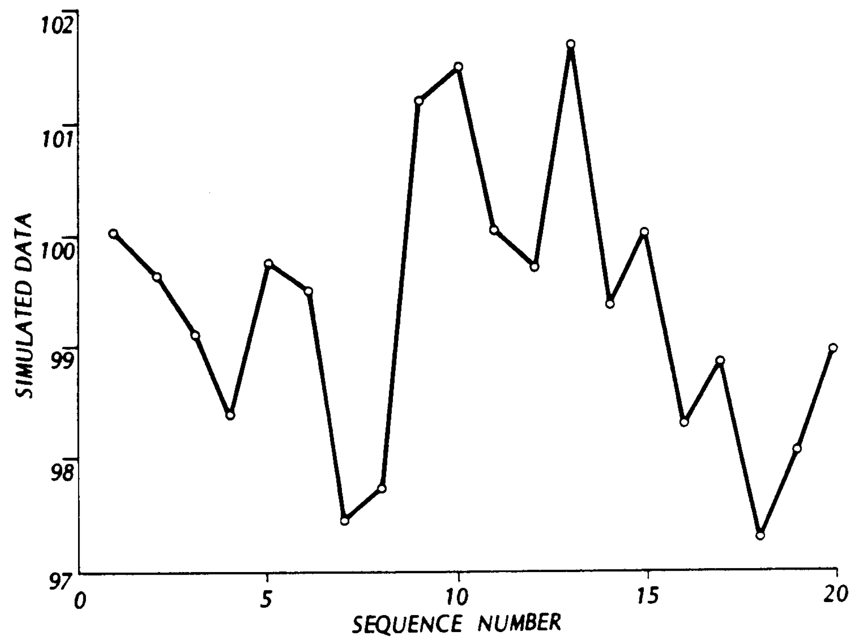

Figure 4.3.1. Simulated data for the model in [4.3.1] that possesses homogeneous nonstationarity.

The operator $\nabla = (1 - B)$ in [4.3.1] is referred to as the *differencing operator* because the root of $(1 - B) = 0$ lies on the unit circle. When ∇ operates on $(z_t - \mu)$ the level μ disappears due to the nonstationarity as is shown by

$$(1 - B)(z_t - \mu) = (z_t - \mu) - (z_{t-1} - \mu) = z_t - z_{t-1} = (1 - B)z_t \quad [4.3.2]$$

When a time series of length N is differenced using [4.3.2], adjacent time series values are subtracted from each other to obtain a sequence of length $N - 1$. This differencing procedure can be repeated just enough times to produce a stationary series labelled w_t. In general, a time series may be differenced d times to produce a stationary series of length $n = N - d$ given by

$$w_t = (1 - B)^d z_t = \nabla^d z_t$$

If the original z_t time series is transformed by a Box-Cox transformation as explained in Section 3.4.5, the stationary w_t series is formed by differencing the transformed series and is calculated using

$$w_t = (1-B)^d z_t^{(\lambda)} \quad [4.3.3]$$

When homogeneous nonstationarity is present, it is reasonable to assume that the w_t series in [4.3.3] can be modelled by the stationary ARMA(p,q) model in [3.4.3] such that

$$\phi(B)w_t = \theta(B)a_t \quad [4.3.4]$$

where the roots of $\phi(B)=0$ lie outside the unit circle for stationarity of the w_t sequence, the d roots of $(1-B)^d$ are on the unit circle due to the homogeneous nonstationarity of the $z_t^{(\lambda)}$ series in [4.3.3], and the roots of $\theta(B)=0$ lie outside the unit circle for invertibility. The process defined by [4.3.3] and [4.3.4] is referred to as an *autoregressive integrated moving average (ARIMA) process*. The reason for the term "integrated" can be found by rewriting [4.3.3] for $d=1$ as

$$z_t^{(\lambda)} = (1-B)^{-1}w_t = (1+B+B^2+\cdots)w_t = \sum_{j=0}^{\infty} w_{t-j} \quad [4.3.5]$$

It can be seen that the $z_t^{(\lambda)}$ series can be obtained by summing or "integrating" the stationary w_t process. When the order of differencing is d then $z_t^{(\lambda)}$ is calculated by "integrating" the w_t process d times. To obtain the original z_t series from the $z_t^{(\lambda)}$ sequence, the inverse of the Box-Cox transformation in [3.4.30] is taken.

The ARIMA (p,d,q) notation is used to indicate the orders of the AR, differencing and MA operators, respectively, which are contained in the ARIMA process given by [4.3.3] and [4.3.4]. When there is no differencing (i.e., $d=0$), the set of ARIMA(p,0,q) processes is the same as the family of stationary ARMA(p,q) processes defined in Section 3.4. However, when dealing with stationary processes it has become common practice to use the term ARMA(p,q), whereas ARIMA(p,d,q) is employed whenever there is a differencing operator (i.e., $d > 0$).

To demonstrate the effects of the differencing operator consider the set of ARIMA(0,d,0) models given by

$$(1-B)^d(z_t - 100) = a_t \quad [4.3.6]$$

where 100 is the mean level of the series for $d=0$ and this level disappears due to differencing when $d>0$. When $d=0$, the model is white noise. In Chapter 9, general procedures are described for simulating with white noise, ARMA, and ARIMA models. Figure 4.3.2 is a plot of 100 simulated terms from the model where the a_t's are randomly generated on a computer as being NID(0,1). It can be seen that the entries in the series appear to be uncorrelated and fluctuate about an overall mean level of 100. The same 100 a_t terms that are used for generating the sequence in Figure 4.3.2 are also employed to simulate series of length 100 for $d = 1,2$ and 3. In Figure 4.3.3, a simulated sequence is shown for an ARIMA(0,1,0) model where a starting value of $z_1 = 100$ is utilized. Notice how the series does not fluctuate about any overall mean level and generally tends to increase in value over time. Using initial values of $z_1 = 100$ and $z_2 = 102$, a synthetic series for an ARIMA(0,2,0) model is generated in Figure 4.3.4. In that figure, the local fluctuations have largely disappeared and the sequence increases dramatically in value with increasing time. Figure 4.3.5 is a simulated trace from an ARIMA(0,3,0) model where starting values of $z_1 = 100$, $z_2 = 102$, and $z_3 = 104$ are employed. The simulated data increases exponentially over time and the right hand portion of the graph seems to mimic a missile trajectory.

4.3.2 Autocorrelation Function

As explained in [3.4.13] in Section 3.4.2, the theoretical ACF for an ARMA(p,q) process satisfies the difference equation

$$\phi(B)\rho_k = 0\,, \quad k > q \qquad [4.3.7]$$

where ρ_k is the theoretical ACF at lag k, and $\phi(B)$ is the AR operator of order p. Assuming distinct roots, the general solution for this difference equation is

$$\rho_k = A_1G_1^k + A_2G_2^k + \cdots + A_pG_p^k \qquad [4.3.8]$$

where $G_1^{-1}, G_2^{-1}, \ldots, G_p^{-1}$, are the roots of the characteristic equation $\phi(B) = 0$ and the A_i's are constants. Due to stationarity conditions, $|G_1^{-1}| > 1$ for a real root G_1^{-1}. Therefore, for increasing lag k, the term $A_iG_i^k$ damps out because $|G_i| < 1$. When all of the roots lie outside the unit circle, the theoretical ACF in [4.3.8] attenuates quickly for moderate and large lags. However, suppose that homogeneous nonstationarity is approached and at least one of the roots G_i^{-1} approaches the unit circle. This, in turn, will cause $|G_i|$ to go towards unity, $A_iG_i^k$ will not die out quickly for larger lags and, hence, ρ_k in [4.3.8] will not damp out fast for moderate and large lags.

The behaviour of the theoretical ACF for a process which is approaching homogeneous nonstationarity has some important practical implications. When the sample ACF in [2.5.9] for a given data set does not die out quickly for larger lags, this may indicate that the data should be differenced to remove homogeneous nonstationarity. For example, the sample ACF along with the 95% confidence limits is displayed in Figure 4.3.6 for the 100 simulated data points in Figure 4.3.3 which were generated by an ARIMA(0,1,0) model. Because the sample ACF attenuates slowly, this indicates the need for differencing. When the simulated sequence from Figure 4.3.3 is differenced to remove nonstationarity, the resulting sample ACF and 95% confidence limits for the differenced data are as shown in Figure 4.3.7. As expected, after differencing only white noise residuals remain. This confirms that the data were originally generated by an ARIMA(0,1,0) model.

In Figure 4.3.6, the sample ACF possesses large values at lower lags that slowly attenuate for increasing lag. However, as noted by Wichern (1973) and Roy (1977), it is not necessary that the sample ACF at the first few lags be rather large if nonstationarity is present. In certain situations, the sample ACF at low lags may in fact be relatively quite small. However, no matter how large the sample ACF values are at the first few lags, when a given data set possesses homogeneous nonstationarity the sample ACF must slowly attenuate for increasing lags.

When it is suspected that a given data set is nonstationary, the time series should be differenced just enough times to cause the sample ACF to attenuate fast for the differenced series. Following this an ARMA(p,q) model can be fitted to the differenced series which is assumed to be stationary. In practice, usually $d = 0$, 1, or 2 for ARIMA models that are fitted to many types of measured series that arise in the natural and social sciences. Furthermore, if the original data set is transformed by a Box-Cox transformation this does not eliminate the need for differencing.

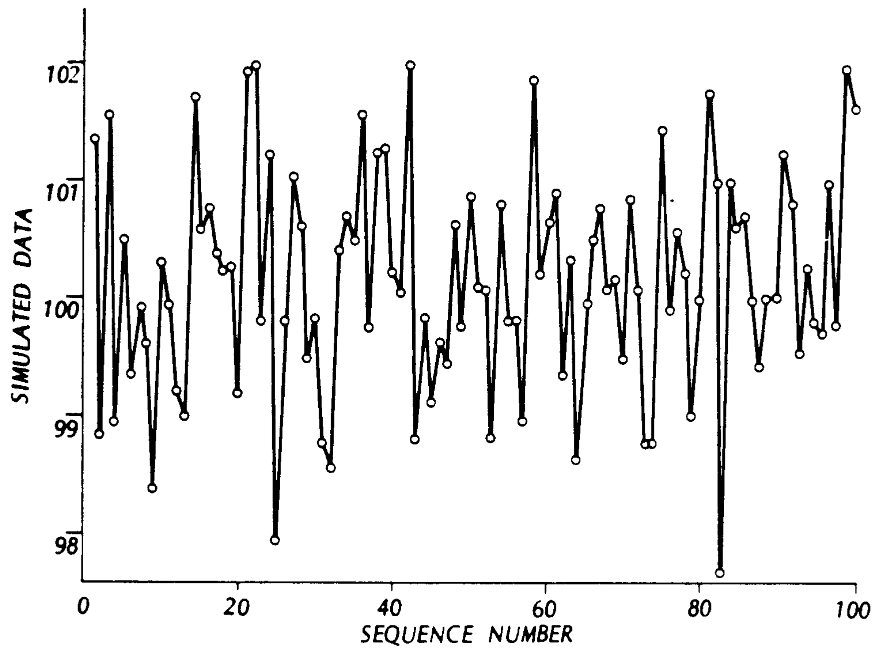

Figure 4.3.2. Simulated sequence for a white noise model.

Figure 4.3.3. Simulated sequence for an ARIMA(0,1,0) model.

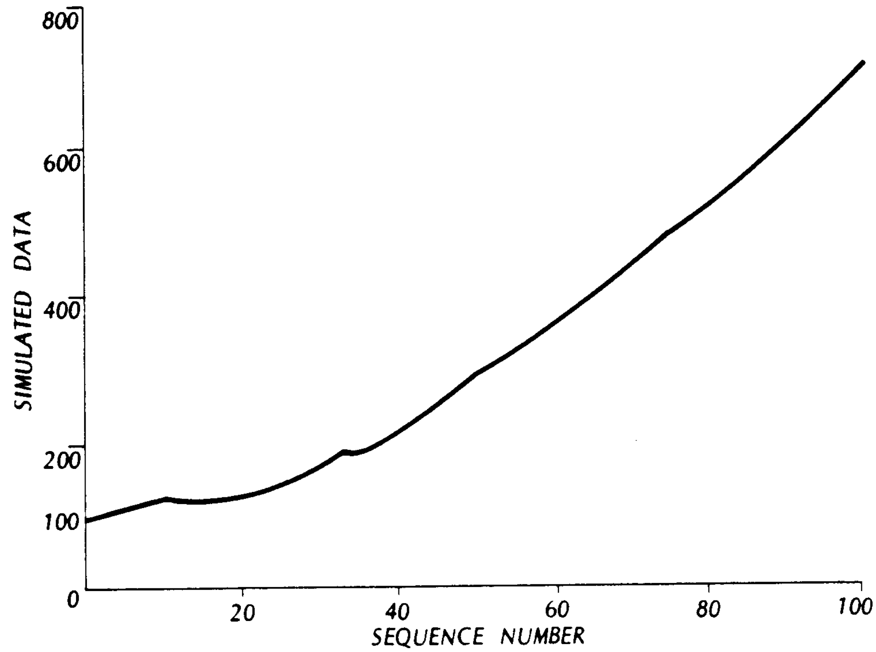

Figure 4.3.4. Simulated sequence for an ARIMA(0,2,0) model.

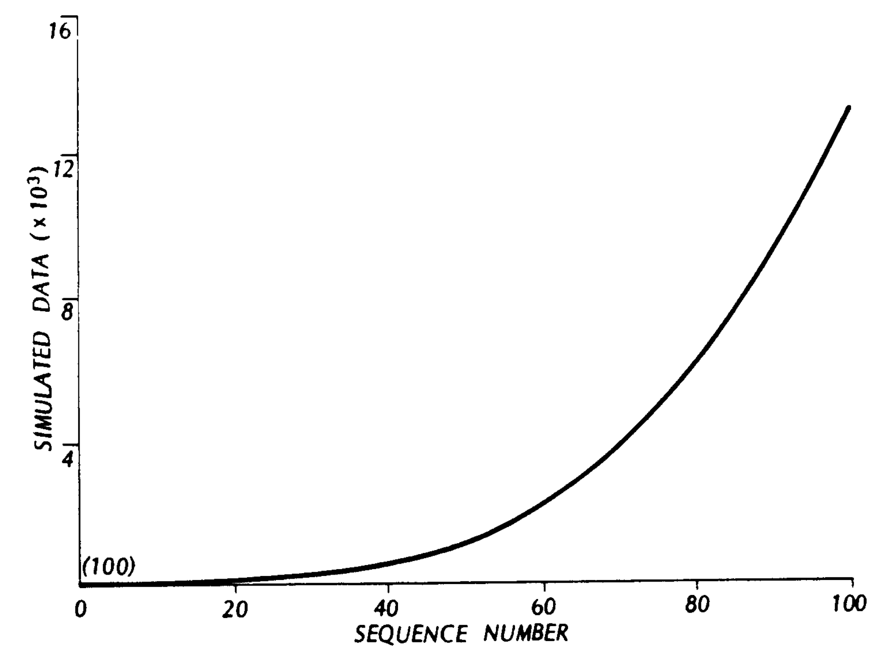

Figure 4.3.5. Simulated sequence for an ARIMA(0,3,0) model.

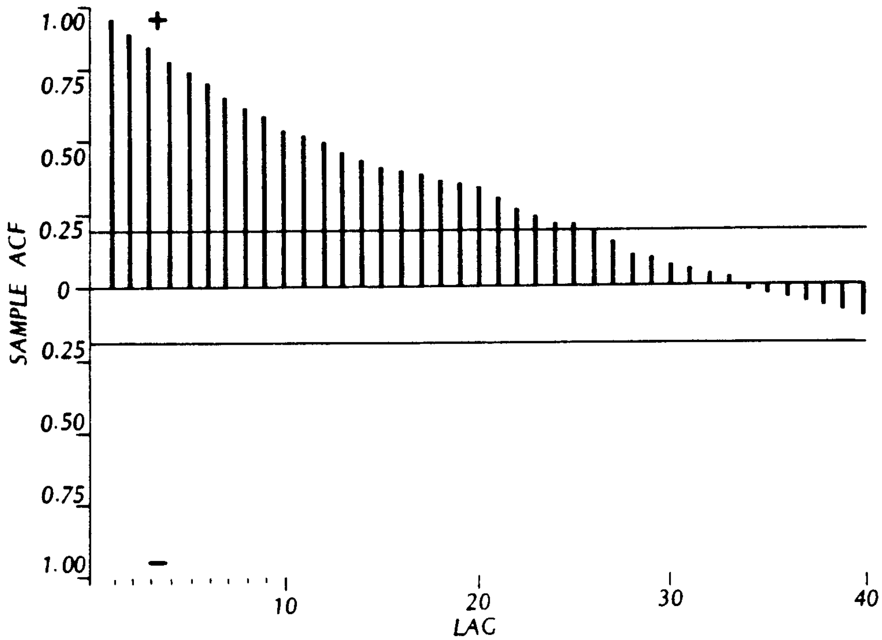

Figure 4.3.6. Sample ACF and 95% confidence limits for simulated data from an ARIMA(0,1,0) model.

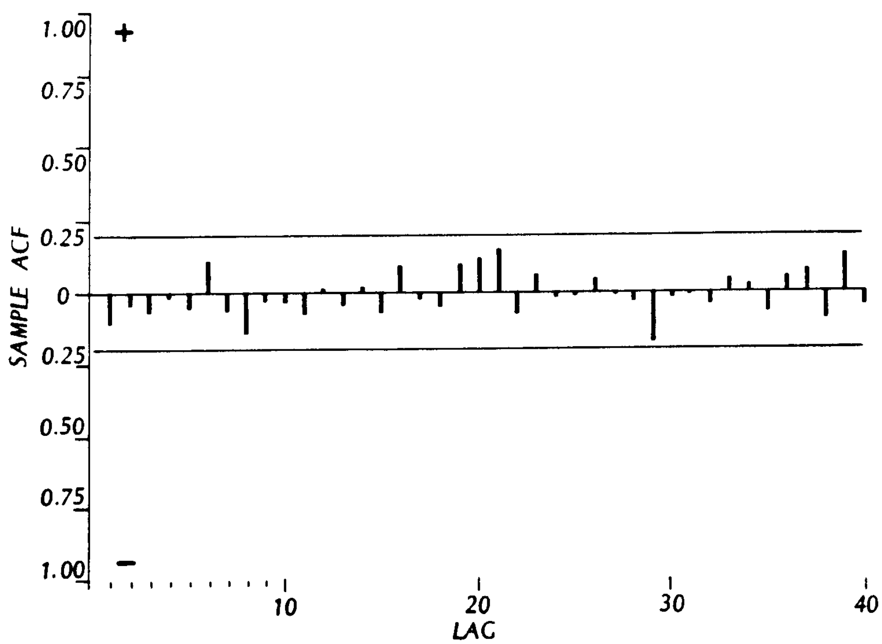

Figure 4.3.7. Sample ACF and 95% confidence limits for the differenced data generated from an ARIMA(0,1,0) model.

Rather, the transformed time series should be differenced as many times as are required to cause the sample ACF of the differenced transformed series to damp out quickly for moderate and large lags.

In certain situations, it may be difficult to ascertain whether or not a given series is nonstationary. This is because there is often no sharp distinction between stationarity and nonstationarity when the nonstationary boundary is nearby. As one or more of the roots of the characteristic equation approaches the unit circle, an ARMA process gradually changes to a nonstationary process and at the same time the corresponding theoretical ACF attenuates less quickly for increasing lags. Consequently, when examining the sample ACF for a specified data set, it is not always obvious whether or not differencing is required. If the fitted model is to be used for simulation, it may be advantageous to choose a model that does not require differencing so that the simulated data will fluctuate around an overall mean level. On the other hand, a model with a differencing operator may perform better than a stationary model when the model is used for forecasting. If employed judiciously, the Akaike information criterion (AIC) (Akaike, 1974) may be used as a guide to determine if differencing is required (see Sections 1.3.3 and 6.3).

4.3.3 Examples of Nonstationary Time Series

Annual Water Use for New York City

The annual water use for New York City is available from 1898 to 1968 in litres per capita per day (Salas and Yevjevich, 1972) and a graph of the series is portrayed in Figure 4.3.8. Because water use has tended to increase over time, the series is obviously nonstationary. The general patterns in Figure 4.3.8 are quite similar to those in Figure 4.3.3 for data that were simulated from an ARIMA(0,1,0) model. The inherent nonstationarity is also confirmed by the graph in Figure 4.3.9 of the sample ACF and 95% confidence limits of the New York water use data. The estimated ACF in Figure 4.3.9 dies off rather slowly and closely mimics the sample ACF in Figure 4.3.6 for the data that were generated from an ARIMA(0,1,0) model. When the water use data are differenced, the resulting series is white noise since all of the values of the sample ACF for the differenced data fall within the 95% confidence limits. Consequently, the annual New York water use series can be modelled by an ARIMA(0,1,0) model.

Electricity Consumption

The total annual electricity consumption for the U.S. is available from 1920 to 1970 in millions of kiloWatt - hours (United States Bureau of the Census, 1976) and a plot of the series is given in Figure 4.3.10. Due to the increase in electricity demand over time, the series is nonstationary. The behaviour of the electricity consumption series in Figure 4.3.10 closely resembles that in Figure 4.3.4 for data that were simulated from an ARIMA(0,2,0) model. As shown in Figures 4.3.11 and 4.3.12, the sample ACF's attenuate slowly for the given electricity consumption series and also the differenced series, respectively. When the series is differenced twice the nonstationarity is removed as demonstrated by the sample ACF in Figure 4.3.13. The large value at lag one indicates the need for a MA parameter in the model. At lag 9, the sample ACF just crosses the 95% confidence limits and this behaviour may be due to chance alone or could indicate the need for another parameter in the model. The sample PACF in Figure 4.3.14 for the electricity consumption data may be interpreted as attenuating quickly at the first few lags due to the need for a MA component. Based upon this identification information, the most appropriate model to the electricity consumption data is an ARIMA(0,2,1) model. Moreover, when one

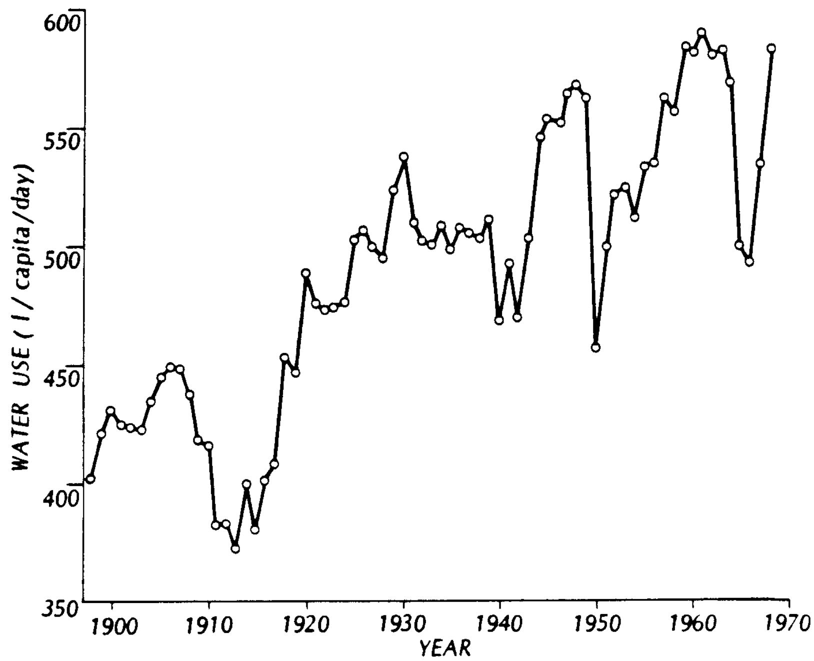

Figure 4.3.8. Annual water use for New York City.

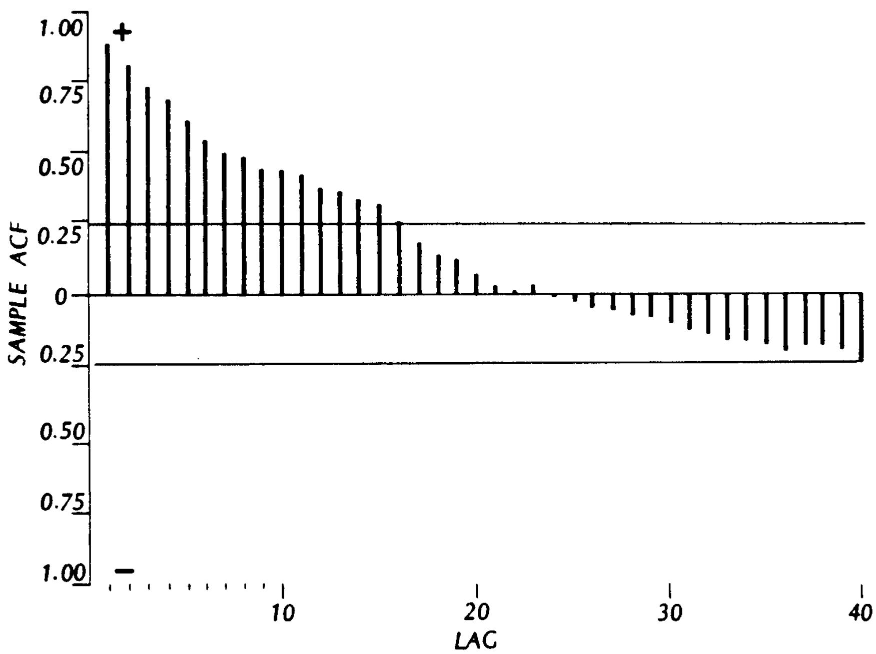

Figure 4.3.9. Sample ACF and 95% confidence limits for the annual water use of New York City.

obtains a maximum likelihood estimate (see Section 6.2) of the Box-Cox parameter λ in [3.4.30], the estimated value is $\lambda = 0.533$, which is essentially a square root transformation (i.e. $\lambda \approx 0.5$). The need for a data transformation can be visually detected by examining the graphs of a smoothing procedure which divides the original graph of the electricity demand series into smooth and rough plots (see Section 22.3).

Beveridge Wheat Price Index

The annual Beveridge wheat price index series which is available from 1500 to 1869 (Beveridge, 1921) is shown in Figure 4.3.15. This series could be closely related to climatic conditions and, therefore, may be of interest to hydrologists and climatologists. For example, during years when the weather is not suitable for abundant grain production, the price of wheat may greatly escalate. If a model can be developed that relates a given hydrologic time series to the Beveridge wheat price indices, this model could be employed to extend the hydrologic record if it were shorter than the other data set (see Sections 17.5.4, 18.5.2 and 19.3.2).

From a plot of the Beveridge wheat price indices in Figure 4.3.15 for the period from 1500 to 1869, it can be seen that the series is nonstationary. Both the level and variance of the time series are increasing over time. A change in variance over time of the original data would eventually be mirrored by variance that is not constant in the residuals of the model fitted to the data. To rectify the situation from the start, natural logarithms are taken of the series so that the variance changes are not as drastic as those shown in Figure 4.3.15. The sample ACF is given for the logarithmic series from 1500 to 1869 in Figure 4.3.16. Because the sample ACF attenuates very slowly for increasing lag, the logarithmic data set should be differenced to remove the inherent nonstationarity. Figure 4.3.17 is a plot of the sample ACF for the differenced logarithmic data along with the 95% confidence limits where it is assumed that the estimated ACF is not significantly different from zero after lag 3. In addition to the large values at low lags, the sample ACF just touches the 95% confidence limits at lag 8. The graph of the sample PACF and 95% confidence limits for the differenced logarithmic series is presented in Figure 4.3.18. A rather large value of the estimated PACF exists at lag 2 while there is a value that crosses the 95% confidence limits at lag 8. Therefore, an AR operator that includes parameters at low lags and also lag 8, may be required in a model that is fitted to the data. After considering a number of possible models, it is found that the most appropriate model to fit to the logarithmic series is a constrained ARIMA(8,1,1) model where ϕ_3 to ϕ_7 are not included in the AR operator.

4.3.4 Three Formulations of the ARIMA Process

In Section 3.4.3, it is shown how the difference equation for the ARMA(p,q) process in [3.4.4] can also be written in the random shock form as an infinite MA process in [3.4.18] or else in the inverted form as an infinite AR process in [3.4.25]. The results in Section 3.4.3 also hold for the stationary w_t process in [4.3.4] which is made stationary by differencing the nonstationary $z_t^{(\lambda)}$ process in [4.3.3]. By using similar procedures, the ARIMA difference equation for the nonstationary $z_t^{(\lambda)}$ process can also be conveniently expressed in either the random shock or inverted forms.

Treating $\phi(B)$, $\theta(B)$, and $(1-B)^d$ as algebraic operators, the *random shock form* of the ARIMA process is

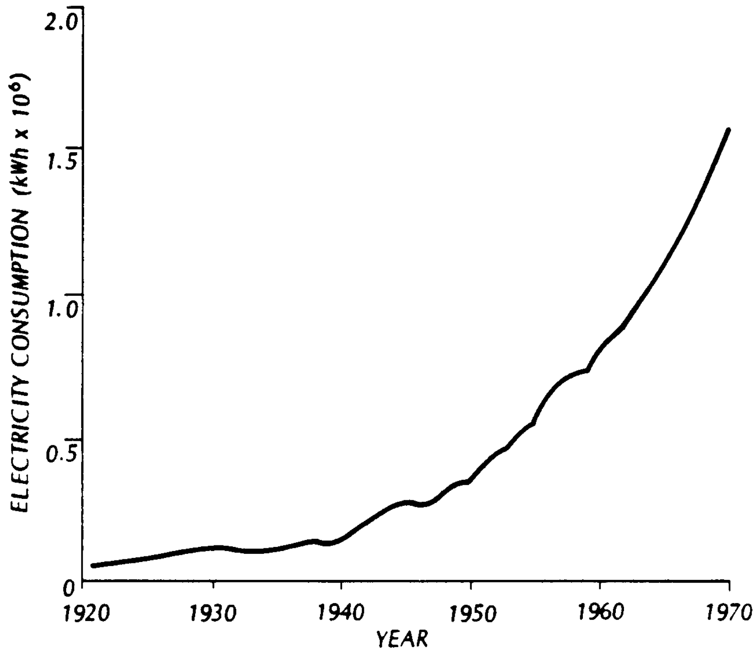

Figure 4.3.10. Total annual electricity consumption in the U.S.A.

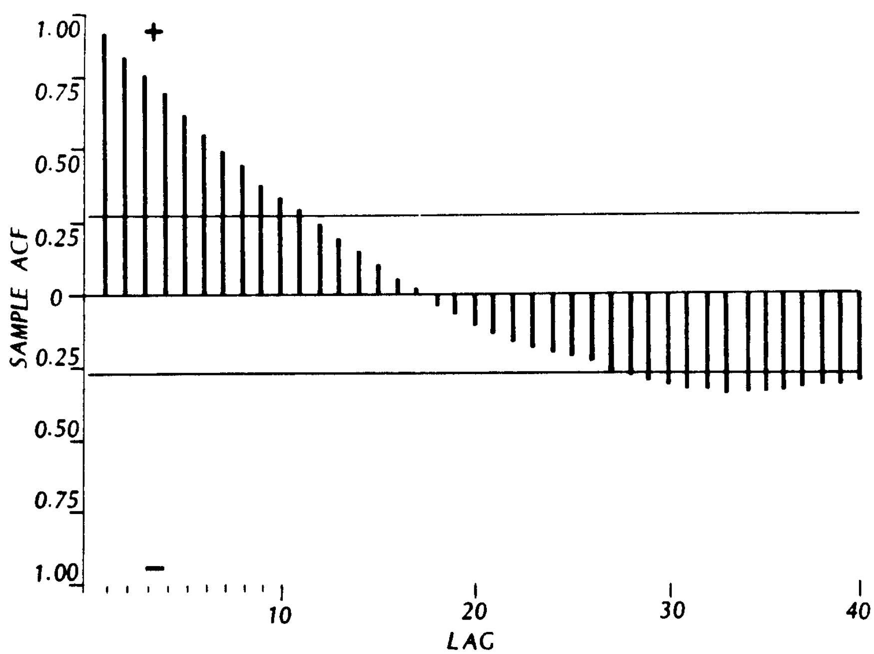

Figure 4.3.11. Sample ACF and 95% confidence limits for the annual American electricity consumption.

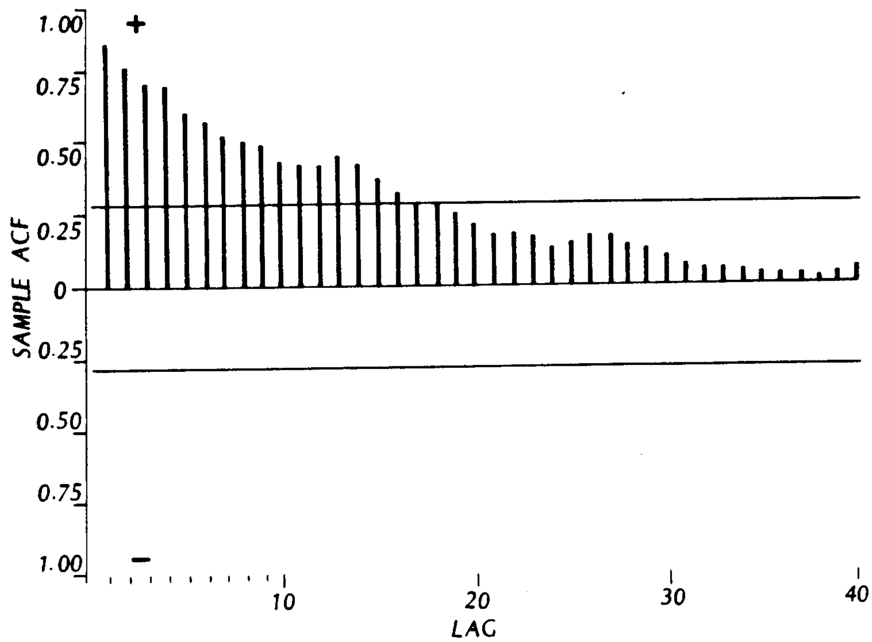

Figure 4.3.12. Sample ACF and 95% confidence limits for the differenced annual American electricity consumption series.

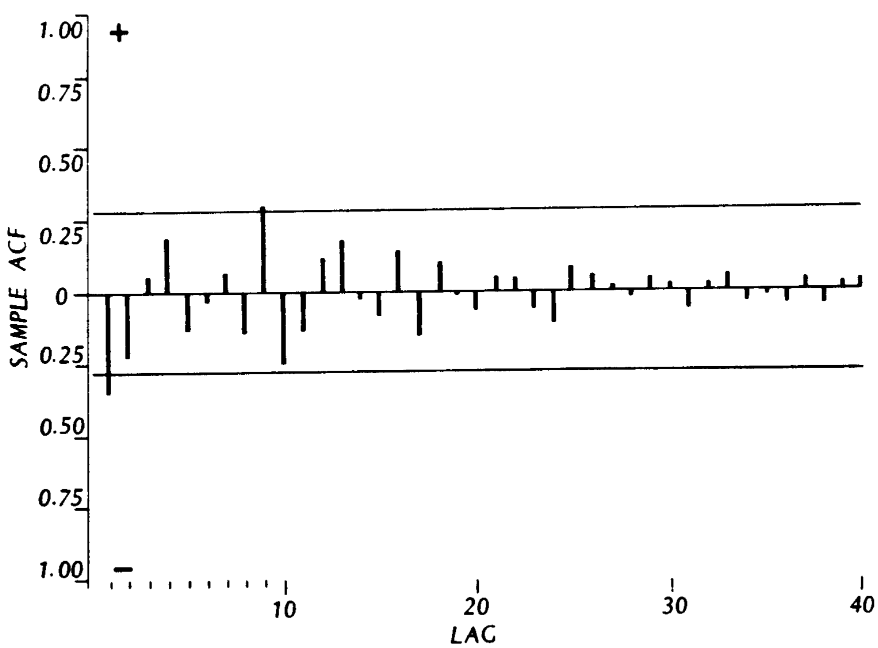

Figure 4.3.13. Sample ACF and 95% confidence limits for the annual American electricity consumption series that is differenced twice.

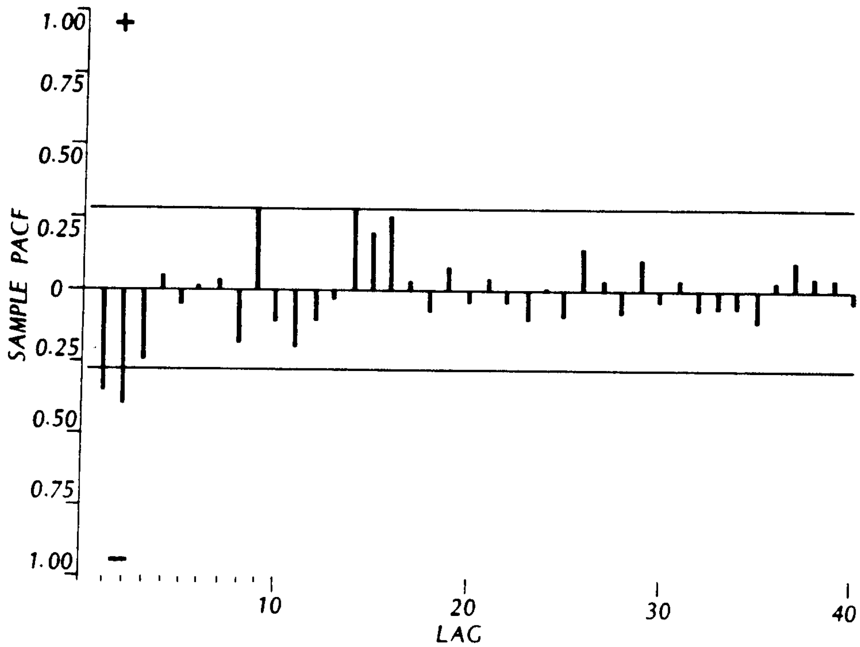

Figure 4.3.14. Sample PACF and 95% confidence limits for the annual American electricity consumption series that is differenced twice.

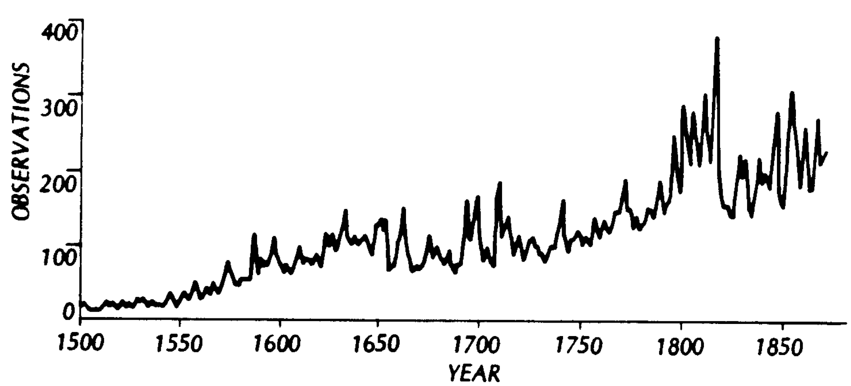

Figure 4.3.15. Beveridge wheat price indices from 1500 to 1869.

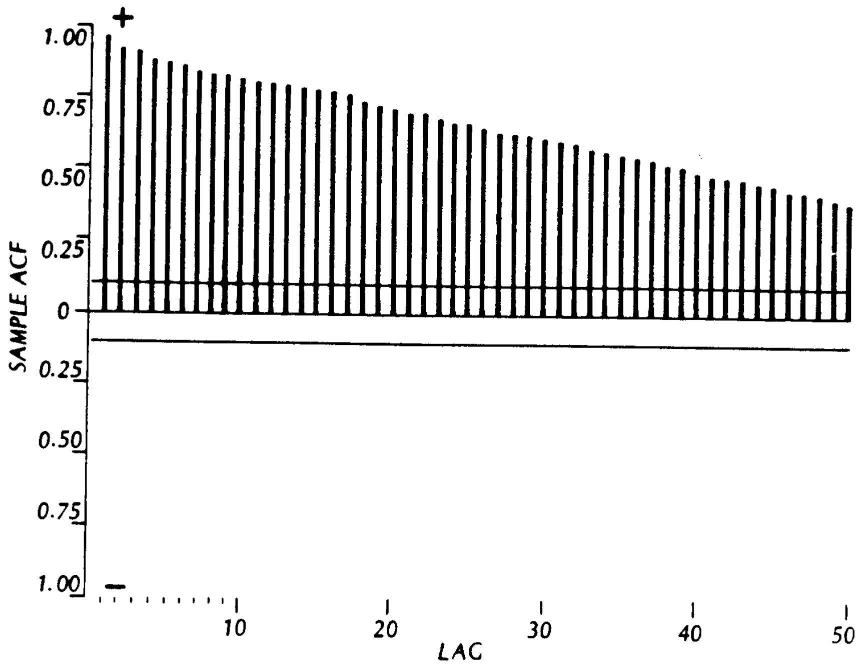

Figure 4.3.16. Sample ACF and 95% confidence limits for the logarithmic Beveridge wheat price index series.

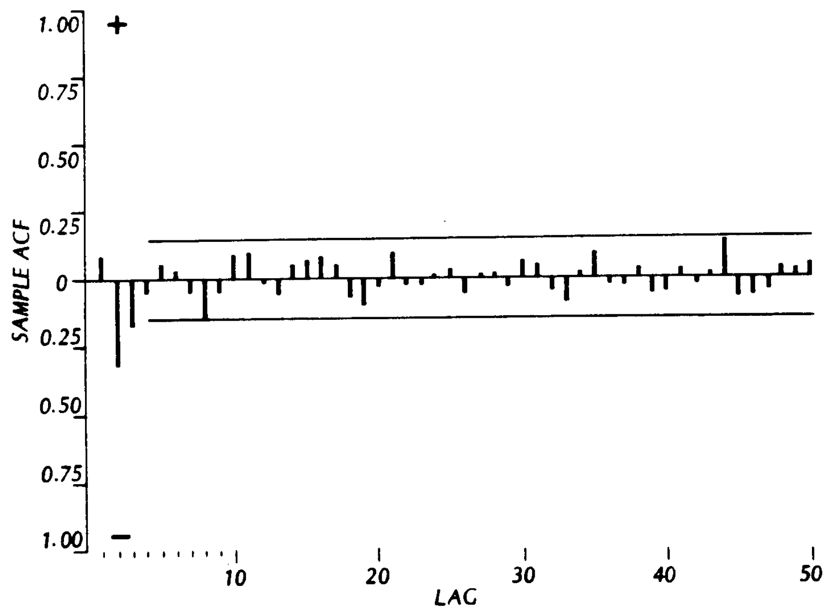

Figure 4.3.17. Sample ACF and 95% confidence limits for the differenced logarithmic Beveridge wheat price index series when ρ_k is zero after lag 3.

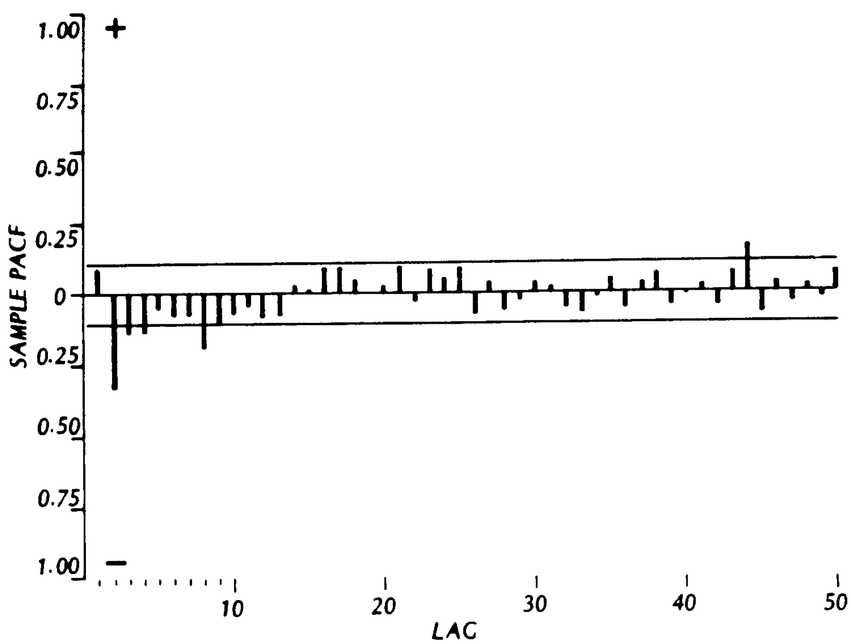

Figure 4.3.18. Sample PACF and 95% confidence limits for the differenced logarithmic Beveridge wheat price index series.

$$
\begin{aligned}
z_t^{(\lambda)} &= [\phi(B)(1-B)^d]^{-1}\theta(B)a_t \\
&= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \\
&= (1 + \psi_1 B + \psi_2 B^2 + \cdots)a_t \\
&= \psi(B)a_t
\end{aligned}
\qquad [4.3.9]
$$

where $\psi(B)$ is the *random shock or infinite MA operator* and ψ_i is the ith parameter or weight of $\psi(B)$. To develop a relationship for ascertaining the ψ parameters, first multiply [4.3.9] by $\phi(B)(1-B)^d$ to obtain

$$\phi(B)(1-B)^d z_t^{(\lambda)} = \phi(B)(1-B)^d \psi(B)a_t$$

From [4.3.3] and [4.3.4], $\theta(B)a_t$ can be exchanged for $\phi(B)(1-B)^d z_t^{(\lambda)}$ in the previous equation to get

$$\theta(B) = \phi(B)(1-B)^d \psi(B) \qquad [4.3.10]$$

The ψ weights can be readily determined by expressing [4.3.10] as

$$\phi(B)(1-B)^d \psi_k = -\theta_k \qquad [4.3.11]$$

where B operates on k, $\psi_0 = 1$, $\psi_k = 0$ for $k < 0$ and $\theta_k = 0$ if $k > q$. As is done for the examples in Section 3.4.3, the ψ weights can be recursively calculated by solving [4.3.11] for $k = 1,2,\ldots,q'$, where q' is the number of ψ parameters that are required.

In order to write the *inverted form* of the process, the ARIMA process is reformulated as

$$a_t = \theta(B)^{-1}\phi(B)(1-B)^d z_t^{(\lambda)}$$
$$= z_t^{(\lambda)} - \pi_1 z_{t-1}^{(\lambda)} - \pi_2 z_{t-2}^{(\lambda)} - \cdots$$
$$= (1 - \pi_1 B - \pi_2 B^2 - \cdots)z_t^{(\lambda)}$$
$$= \pi(B) z_t^{(\lambda)} \qquad [4.3.12]$$

where $\pi(B)$ is the *inverted or infinite AR operator* and π_i is the ith parameter or weight of $\pi(B)$. To determine a relationship for calculating the π parameters, multiply [4.3.12] by $\theta(B)$ to get

$$\theta(B)a_t = \theta(B)\pi(B)z_t^{(\lambda)}$$

By employing [4.3.3] and [4.3.4], $\phi(B)(1-B)^d z_t^{(\lambda)}$ can be substituted for $\theta(B)a_t$ in the above equation to obtain

$$\phi(B)(1-B)^d = \theta(B)\pi(B) \qquad [4.3.13]$$

The π coefficients can be easily ascertained by expressing the above equation as

$$\theta(B)\pi_k = (1-B)^d \phi_k \qquad [4.3.14]$$

where $\pi_0 = -1$ and $\phi_0 = -1$ when using [4.3.14] to calculate π_k for $k > 0$, $\pi_k = 0$ for $k < 0$, and $\phi_k = 0$ if $k > p$ or $k < 0$. By solving [4.3.14] for $k = 1, 2, \ldots, p'$, where p' is the number of π parameters that are needed, the π weights can be recursively calculated in the same fashion as the examples in Section 3.4.3.

An interesting property of the π weights is when $d \geq 1$ the parameters in the inverted operator sum to unity. This fact can be proven by substituting $B = 1$ into [4.3.13]. In that equation, $\phi(1)$ and $\theta(1)$ are not zero since the roots of the characteristic equations for the AR and MA operators lie outside the unit circle. However, $(1-B)^d = 0$ for $B = 1$ and therefore [4.3.13] reduces to

$$\pi(1) = 0$$

or

$$\sum_{j=1}^{\infty} \pi_j = 1 \qquad [4.3.15]$$

Consequently, for $d \geq 1$ equation [4.3.12] can be written as

$$z_t^{(\lambda)} = \sum_{j=1}^{\infty} \pi_j z_{t-j}^{(\lambda)} + a_t \qquad [4.3.16]$$

where the summation term on the right hand side constitutes a weighted average of the previous values of the process.

4.4 INTEGRATED MOVING AVERAGE PROCESSES

In Section 4.3.3, it was found that the most appropriate model to fit to the total annual electricity consumption in the U.S.A. is an ARIMA(0,2,1) model. When modelling time series from economics and other fields of study, it often turns out that ARIMA models are needed where $p=0$ and both d and q are greater than zero. Because no AR operator is present, an ARIMA(0,d,q) process is often referred to as an *integrated moving average (IMA) process* and is denoted by IMA(0,d,q). For a detailed description of IMA processes, the reader may wish to refer to the book of Box and Jenkins (1976, Ch. 4, pp. 103-114).

A special case of the IMA(0,d,q) family of processes is the IMA(0,1,1) process given by

$$(1-B)z_t^{(\lambda)} = (1-\theta_1 B)a_t$$

or

$$z_t^{(\lambda)} = z_{t-1}^{(\lambda)} + a_t - \theta_1 a_{t-1}$$

Keeping in mind that the data, z_t, may require a Box-Cox transformation, the above equation can be more conveniently written by dropping the λ superscript and writing it as

$$z_t = z_{t-1} + a_t - \theta_1 a_{t-1} \qquad [4.4.1]$$

The minimum mean square error forecasts (see Section 8.2) obtained from an IMA(0,1,1) process, are the same forecasts that are produced when using *single exponential smoothing* [see, for example, Gilchrist (1976, p. 108)]. Because exponential smoothing has been used extensively for forecasting economic time series [see, for instance, Makridakis and Wheelwright (1978) and Gilchrist (1976)], the IMA(0,1,1) process has received widespread attention. Important original research regarding the optimal properties of exponentially weighted forecasts is given by Muth (1960).

To appreciate the inherent structure of the IMA(0,1,1) process in [4.4.1], the random shock form of the process in [4.3.9] is useful. The ψ coefficients can be obtained by employing [4.3.11] for positive values of k. For $k=1$

$$(1-B)\psi_1 = -\theta_1 \text{ or } \psi_1 - \psi_0 = -\theta_1$$

But $\psi_0 = 1$ and, therefore, $\psi_1 = 1-\theta_1$.

When $k=2$

$$(1-B)\psi_2 = 0 \text{ or } \psi_2 - \psi_1 = 0$$

Therefore, $\psi_2 = \psi_1 = (1-\theta_1)$.

For $k=3$

$$(1-B)\psi_3 = 0 \text{ or } \psi_3 - \psi_2 = 0$$

Therefore, $\psi_3 = \psi_2 = (1-\theta_1)$.

In general,

$$(1 - B)\psi_k = 0 \text{ or } \psi_k - \psi_{k-1} = 0$$

Therefore, $\psi_k = \psi_{k-1} = \cdots = \psi_1 = (1 - \theta_1)$.

By substituting for the ψ parameters into [4.3.9], the random shock form of the model is

$$z_t = (1 - \theta_1)\sum_{j=1}^{\infty} a_{t-j} + a_t \qquad [4.4.2]$$

From [4.4.2] it can be seen that the present value of the process depends upon the current random shock, a_t, plus the summation of an equal weighting of all previous disturbances. Consequently, part of the random shock in any period has a permanent effect due to the weight, $(1 - \theta_1)$, while the rest affects the system only in the current time period.

The inverted form of the process can be employed for understanding the properties of an IMA(0,1,1) process. By examining [4.3.14] for positive values of k, the π coefficient can be ascertained. For $k = 1$

$$(1 - \theta_1 B)\pi_1 = (1 - B)\phi_1 \text{ or } \pi_1 - \theta_1\pi_0 = \phi_1 - \phi_0$$

But $\pi_0 = \phi_0 = -1$ when determining the π weights and $\phi_1 = 0$ since $p = 0$. Therefore, $\pi_1 = 1 - \theta_1$. When $k = 2$

$$(1 - \theta_1 B)\pi_2 = (1 - B)\phi_2 \text{ or } \pi_2 - \theta_1\pi_1 = \phi_2 - \phi_1$$

Because no AR parameters are present in the IMA(0,1,1) process, $\phi_1 = \phi_2 = 0$ and, therefore, $\pi_2 = \theta_1\pi_1 = \theta_1(1 - \theta_1)$. For $k = 3$

$$(1 - \theta_1 B)\pi_3 = (1 - B)\phi_3 \text{ or } \pi_3 - \theta_1\pi_2 = \phi_3 - \phi_2 = 0$$

Hence, $\pi_3 = \theta_1\pi_2 = \theta_1^2(1 - \theta_1)$.

In general, the π coefficient at lag k is determined by

$$(1 - \theta_1 B)\pi_k = (1 - B)\phi_k \text{ or } \pi_k - \theta_1\pi_{k-1} = \theta_1^{k-1}(1 - \theta_1)$$

By substituting for the π parameters into [4.3.16], the inverted form of the process is

$$z_t = (1 - \theta_1)\sum_{j=1}^{\infty} \theta_1^{j-1} z_{t-j} + a_t \qquad [4.4.3]$$

The summation term on the right hand side of [4.4.3] constitutes an *exponentially weighted moving average (EWMA)* of the previous values of the process and is denoted as

$$\bar{z}_{t-1}(\theta_1) = (1 - \theta_1)\sum_{j=1}^{\infty} \theta_1^{j-1} z_{t-j} \qquad [4.4.4]$$

The weights in [4.4.4] are formed by the sequence of π parameters given by $(1 - \theta_1), (1 - \theta_1)\theta_1, (1 - \theta_1)\theta_1^2, (1 - \theta_1)\theta_1^3, \ldots$. When θ_1 has a value of zero, the IMA(0,1,1) process in [4.4.2] reduces to an IMA(0,1,0) process where $\pi_1 = 1$ and $\pi_k = 0$ for $k>1$. As the value of θ_1 approaches unity, the π weights attenuate more slowly and the EWMA in [4.4.4] stretches further into the past of the process. When θ_1 is equal to one, the MA and differencing operators cancel in [4.4.1] and the process is a white noise IMA(0,0,0) process.

From its definition in [4.4.4], the recursion formula for the EWMA can be written as

$$\bar{z}_t(\theta_1) = (1-\theta_1)z_t + \theta_1 \bar{z}_{t-1}(\theta_1) \qquad [4.4.5]$$

This expression is what is employed for obtaining forecasts using single exponential smoothing [see, for example, Makridakis and Wheelwright (1978, Ch. 5)]. Although the IMA(0,1,1) process possesses no mean due to the fact that it is nonstationary, the EWMA in [4.4.4] can be regarded as being the location or level of the process. From [4.4.5] it can be seen that each new level is calculated by interpolating between the new observation and the previous level. When θ_1 is equal to zero, the process is actually an IMA(0,1,0) process and the current level in [4.4.5] would be solely due to the present observation. If θ_1 were close to unity, the current level, $\bar{z}_t(\theta_1)$, in [4.4.5] would depend heavily upon the previous level, $\bar{z}_{t-1}(\theta_1)$, while the current observation, z_t, would be given a small weight of $(1-\theta_1)$.

Muth (1960) suggests an intuitive approach for interpreting the generation of the single exponential smoothing procedure or equivalently the IMA(0,1,1) process. From [4.4.3] and [4.4.4]

$$z_t = \bar{z}_{t-1}(\theta_1) + a_t$$

By substituting [4.4.3] into [4.4.5] it turns out that

$$\bar{z}_t(\theta_1) = \bar{z}_{t-1}(\theta_1) + (1-\theta_1)a_t \qquad [4.4.6]$$

The first of the previous two equations demonstrates how the current value z_t is produced by the level of the system at time $t-1$ plus a random shock added at time t. However, [4.4.6] shows that only a proportion, $(1-\theta_1)$, of the innovation has a lasting influence by being absorbed into the current level of the process.

4.5 DIFFERENCING ANALOGIES

When dealing with discrete data, the differencing operator $\nabla^d = (1-B)^d$ can be employed to remove homogeneous nonstationarity. It turns out that the differencing operator is analogous to differentiation when continuous functions are being studied. Consider, for example, a discrete process which is defined by

$$z_t = \begin{cases} a_t & \text{for } t < T \\ c + a_t & \text{for } t \geq T \end{cases} \qquad [4.5.1]$$

where c is a constant which reflects a local level for $t \geq T$. When a_t is assumed to be IID$(0,\sigma_a^2)$, the mean level of the z_t process before time T is zero while the mean of the process is c for $t \geq T$. The effect of differencing the data once is to remove the local level due to the constant c in [4.5.1]. For $t>T$ the differenced series is calculated as

$$\nabla z_t = (1 - B)z_t = z_t - z_{t-1} = (c + a_t) - (c + a_{t-1})$$
$$= a_t - a_{t-1}$$

The above operation is analogous to taking the first derivative of a continuous function of time which is given as

$$y = \begin{cases} 0 & \text{for } t < T \\ c & \text{for } t \geq T \end{cases}$$

The derivative $\frac{dy}{dt}$ is of course zero for $t>T$ and the local level drops out due to differentiation.

Next, consider the analogous effects of differencing operators of order two for the discrete case and second order derivatives for a continuous function. Suppose that a discrete process is given as

$$z_t = c + bt + a_t \qquad [4.5.2]$$

where b and c are constants. The term, $(c + bt)$, forms a linear deterministic trend while the white noise, a_t, constitutes the probabilistic component of the process, z_t. By using a differencing operator of order one, the constant c in [4.5.2] can be removed as is shown by

$$\nabla z_t = (1 - B)z_t = z_t - z_{t-1}$$
$$= (c + bt + a_t) - (c + b(t - 1) + a_{t-1})$$
$$= b + a_t - a_{t-1}$$

By employing a differencing operator of order two, the entire deterministic trend can be eliminated.

$$\nabla^2 z_t = \nabla(\nabla z_t)$$
$$= (b + a_t - a_{t-1}) - (b + a_{t-1} - a_{t-2})$$
$$= a_t - 2a_{t-1} + a_{t-2}$$

For the continuous case, a function of t may be given as

$$y = c + bt$$

The value of the first derivative is $\frac{dy}{dt} = b$ while $\frac{d^2y}{dt^2} = 0$. Hence, the first order derivative removes the intercept, c, while the second order derivative completely eliminates the linear function.

4.6 DETERMINISTIC AND STOCHASTIC TRENDS

The component $c + bt$ in [4.5.2] is an example of a deterministic linear trend component. In general, the *deterministic trend* component could be any function $f(t)$ and after the trend component is removed from the time series being studied, the residual could be modelled by an appropriate stochastic model. For example, suppose that the series is transformed by a Box-Cox transformation and following this a trend component $f(t)$ and perhaps also an overall mean level μ are subtracted from the transformed series. If the resulting series were modelled by an ARMA(p,q) model, the model would be written as

$$\phi(B)(z_t^{(\lambda)} - f(t) - \mu) = \theta(B)a_t \qquad [4.6.1]$$

This type of procedure is similar to what is used with the deseasonalized models in Chapter 13. Due to the annual rotation of the earth around the sun, there is a physical justification for including a sinusoidal deterministic component when modelling certain kinds of natural seasonal time series. Consequently, the data are deseasonalized by removing a deterministic sinusoidal component and following this the resulting nonseasonal series is modelled using an ARMA(p,q) model.

The model in [4.6.1] possesses a deterministic trend component. In certain types of series with linear trends, the trends may not be restricted to occur at a specified time nor have approximately the same slope or duration. Rather, the trends may occur stochastically and there may be no physical basis for justifying the use of a deterministic trend. As was demonstrated in the previous section, a differencing operator of order two could account for linear trends if they were known or expected to be present. Consequently, to allow for stochastic linear trends, the series which may have first been changed by a Box-Cox transformation could be differenced twice before an ARMA(p,q) model is fitted. In general, *stochastic trends* of order $d - 1$ are automatically incorporated into the ARIMA(p,d,q) model

$$\phi(B)\nabla^d z_t^{(\lambda)} = \theta(B)a_t \qquad [4.6.2]$$

In certain instances, it may not be clear as to whether or not one should include a deterministic trend component in the model. Recall that the w_t sequence in [4.3.3] is assumed to have a mean of $\mu_w = 0$ after the $z_t^{(\lambda)}$ series is differenced d times. However, if the estimated mean of μ_w were significantly different from zero this may indicate that differencing cannot remove all of the nonstationarity in the data and perhaps a deterministic trend is present. When estimating the parameters of an ARIMA model which is fit to a given data set, the MLE (maximum likelihood estimate) $\hat{w}$ of μ_w can be obtained (see Chapter 6). Because a MLE possesses a limiting normal distribution, by using the estimated SE (standard error) and subjectively choosing a level of significance, significance testing can be done for the estimated model parameter. For instance, if the absolute value of $\hat{w}$ is less than twice its SE, it can be argued that $\hat{w}$ is not significantly different from zero and should be omitted from the model. Likewise, when estimating the sample ACF, the mean of the differenced series can be set equal to zero when it is thought that a deterministic trend component is not present. For the sample ACF's of differenced series that are examined in this chapter (see, for instance, Figure 4.3.13), it is assumed that the mean of the differenced series is zero. On the other hand, when a deterministic trend is contained in the data, the mean of the differenced series should be removed when estimating the sample ACF. This will preclude the masking of information in the plot of the sample ACF that can assist in identifying the AR and MA parameters that are required in a model which is fitted to the series.

4.7 CONCLUSIONS

As demonstrated by the interesting applications of Section 4.3.3, the ARIMA(p,d,q) model in [4.3.4] is capable of modelling a variety of time series containing stochastic trends. The first step in modelling a given time series is to ascertain if the data are nonstationary. If, for example, the sample ACF attenuates slowly, this may indicate the presence of nonstationarity and the need for differencing to remove it. Subsequent to obtaining a stationary series, an ARMA(p,q) model can be fitted to the differenced data. If the model residuals are not homoscedastic (i.e., have constant variance) and/or normally distributed, the original time series can be transformed using the Box-Cox transformation in [3.4.30] in order to rectify the situation. Following this, an ARIMA(p,d,q) model can be fitted to the transformed time series by using the procedure just described for the untransformed one.

For the ARIMA(p,d,q) model in [4.3.4], it is assumed that d can only have values that are non-negative integers. A generalization of the ARIMA model is to allow d to be a real number. For a specified range of the parameter d, the resulting process will possess long memory (see Section 2.5.3 for a definition of long and short memory processes) and, consequently, this process is discussed in more detail with other long memory processes in Part V. As explained in Chapter 11 in Part V, when d is allowed to take on real values, the resulting model is referred to as a fractional ARMA or FARMA process. However, before presenting some long memory processes in Part V, the identification estimation, and diagnostic check stages of model construction are described in Part III for use with the stationary and nonstationary linear time series models of Part II. Many of the model building tools of part III are modified and extended for employment with the FARMA models of Chapter 11 as well as the many other types of models presented in the book and listed in Table 1.6.2.

PROBLEMS

4.1 List the names of five types of yearly time series which you expect would be nonstationary. Give reasons for your suspicions. Refer to a journal such as Water Resources Bulletin, Stochastic Hydrology and Hydraulics, Journal of Hydrology, Environmetrics or Water Resources Research and find three examples of yearly nonstationary series. How did the authors of the paper, in which a given series appeared, model the nonstationarity?

4.2 In Section 4.3.1, it is pointed out that a time series should be differenced just enough times to remove homogeneous nonstationarity. What happens if you do not difference the series enough times before fitting an ARMA model to it? What problems can arise if the series is differenced too many times?

4.3 By referring to the paper of Roy (1977), explain why the values of the sample ACF at the first few lags do not have to be large if nonstationarity is present.

4.4 An ARIMA(1,2,1) model is written as

$$(1-B)^2(1-0.8B)z_t = (1-0.5B)a_t$$

Write this model in the random shock and inverted forms. Determine at least seven random shock and inverted parameters.

4.5 For the model in question 4.4, simulate a sequence of 20 values assuming that the innovations are NID(0,1). Simulate another sequence of 20 values using innovations that are NID(0,25). Plot the two simulated sequences and compare the results. To obtain each synthetic data set, you can use a computer programming package such as the McLeod-Hipel Time Series package referred to in Section 1.7. Moreover, you may wish to examine synthetic data generated from other types of ARIMA models.

4.6 Write down the definition of a single exponential smoothing model. Show why the forecasts from this model are the same as the minimum mean square error forecasts obtained from an IMA(0,1,1) model.

4.7 Give the definition of a random walk process. What is the relationship between a random walk process and an IMA(0,1,1) process?

4.8 For each of the series found in question 4.1, explain what type of trend do you think is contained in the data? How would you model each series?

4.9 Outline the approaches that Pandit and Wu (1983) suggest for modelling stochastic and deterministic trends in Chapters 9 and 10, respectively, in their book. Compare these to the procedures described in Section 4.6 and elsewhere in this book.

4.10 Describe the procedure of Abraham and Wu (1978) for detecting the need for a deterministic component when modelling a given time series. Discuss the advantages and drawbacks of their approach.

REFERENCES

DATA SETS

Beveridge, W. H. (1921). Weather and harvest cycles. *Economics Journal*, 31:429-552.

Salas, J. D. and Yevjevich, V. (1972). Stochastic structure of water use time series. Hydrology Paper No. 52, Colorado State University, Fort Collins, Colorado.

United States Bureau of the Census (1976). *The Statistical History of the United States from Colonial Times to the Present*. Washington, D.C.

FORECASTING

Gilchrist, W. (1976). *Statistical Forecasting*. Wiley, New York.

Makridakis, S. and Wheelright, S. C. (1978). *Forecasting - Methods and Applications*. Wiley, New York.

Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, 55:299-306.

TIME SERIES ANALYSIS

Abraham, B. and Box, G. E. P. (1978). Deterministic and forecast-adaptive time-dependent models. *Applied Statistics*, 27(2):120-130.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716-723.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, Oakland, California, revised edition.

Pandit, S.M. and Wu, S.M. (1983). *Time Series and System Analysis with Applications*. Wiley, New York.

Roy, R. (1977). On the asymptotic behaviour of the sample autocovariance function for an integrated moving average process. *Biometrika*, 64(2):419-421.

Wichern, D. W. (1973). The behaviour of the sample autocorrelation function for an integrated moving average process. *Biometrika*, 60(2):235-239.