CHAPTER 3

STATIONARY NONSEASONAL MODELS

3.1 INTRODUCTION

Certain types of environmental records are strictly *nonseasonal* while in other situations it may be required to consider a time series of average annual values even if seasonal data were available. For example, tree ring indices and mud varve thicknesses are usually obtainable only in the form of yearly records, whereas mean annual riverflow, temperature and precipitation data can be calculated from average weekly records. Whatever the case, it is often necessary to deal with nonseasonal natural time series.

The yearly data to be analyzed may be approximately stationary or perhaps may possess statistical properties which change over time. As discussed in Section 2.4.2, it is often reasonable to assume that hydrologic and geophysical data having a moderate time span (usually a few hundred years but perhaps more than 1000 years for certain time series) are more or less *stationary*. On the other hand, an annual water demand series for a large city or the yearly economic growth rate of an irrigated farming region, may constitute time series which are *nonstationary* even over a very short time interval. The present chapter deals with the theory of *stationary linear nonseasonal models* while Chapter 4 is concerned with *nonstationary linear nonseasonal models* which can be used for modelling certain types of nonstationary time series.

Nonseasonal models can be fit to yearly records for use in various types of applications. For instance, when *studying changes in the climate* over a specified time span, it may be advantageous to analyze annual time series. Although average annual hydrological data are rarely available for periods greater than two hundred years, longer time series, which reflect past climatic conditions, can be obtained. Some time series records of tree ring indices for the Bristlecone pine in California are longer than 5000 years in length and tree ring data sets for Douglas fir, Ponderosa pine, Jeffrey pine and other types of evergreens are available for periods of time which are often much longer than 500 years (Stokes et al., 1973).

Hurst (1951, 1956) studied the statistical properties of 690 annual time series when he was examining the long-term storage requirements on the Nile River. This research created the need for a stochastic model which could statistically account for what is called the *Hurst phenomenon*. Although the research of Hurst and accompanying academic controversies are assessed in detail in Chapter 10, it should be pointed out here that the linear stationary models of this chapter do in fact statistically explain the Hurst phenomenon (McLeod and Hipel, 1978; Hipel and McLeod, 1978). Consequently, stationary linear nonseasonal models are of great importance in hydrology and as emphasized in Chapter 10, should be employed in preference to fractional Gaussian noise (FGN) and other related models. Moreover, within Section 3.6 it is clearly demonstrated that there is sound *physical justification* for fitting the models of this chapter to yearly riverflow time series.

The current chapter deals with the *mathematical definitions and properties of various types of stationary linear nonseasonal processes*. The processes which are discussed are the *AR* (autoregressive), *MA* (moving average) and *ARMA* (autoregressive-moving average) processes.

For each of the foregoing processes, a simple process is first considered and this is followed by an extension to the general case. Important mathematical properties of the various processes are usually explained by examining a specific case. Furthermore, it is clearly pointed out where the mathematical properties of the processes can be useful for designing a model to fit to a given data set. The procedure of constructing a model by following the identification, estimation and diagnostic check system of *model development*, is discussed in Chapters 5 to 7, respectively, of Part III.

The importance of abiding by *key modelling principles* (see Sections 1.3 and 5.2.4 for general discussions) is addressed at certain locations within this chapter. For example, in order to make the model as simple or parsimonious as possible, some of the model parameters can be constrained to zero (see Section 3.4.4). To satisfy certain underlying modelling assumptions regarding the model residuals, a power transformation such as a *Box-Cox transformation* (Box and Cox, 1964) can be incorporated into the model (see Section 3.4.5).

The mathematical foundations of linear nonseasonal models form the *basic building blocks* for the more complex nonstationary, long memory, seasonal, transfer function-noise, intervention and multivariate models which are dealt with in Chapters 4 and 11, and Parts VI to IX, respectively, later in the book. Consequently, a sound understanding of the models presented in this chapter is essential in order to be able to fully appreciate the flexibility and limitations of the rich array of ARMA-based models which are available for use by engineers. In addition, the basic notation which is developed for the nonseasonal models is simply extended for use with the other classes of models described in the book.

3.2 AUTOREGRESSIVE PROCESSES

The AR model of this section describes how an observation directly depends upon one or more previous measurements plus a white noise term. This form of a time series model is intuitively appealing and has been widely applied to data sets in many different fields. After describing the simplest form of the AR model, the general AR model is defined. Additionally, the theoretical ACF (autocorrelation function) of an AR model is derived and the related Yule-Walker equations are formulated. These equations can be used for obtaining the partial autocorrelation function (PACF) and determining efficient moment estimates for the parameters of an AR model.

3.2.1 Markov Process

When an observation, z_t, measured at time t depends only upon the time series value at time t-1 plus a random shock, a_t, the process describing this relationship is called an AR process of order 1 and is denoted as AR(1). The AR(1) process is commonly called a *Markov process* and is written mathematically as

$$z_t - \mu = \phi_1(z_{t-1} - \mu) + a_t \qquad [3.2.1]$$

where μ is the mean level of the process, ϕ_1 is the nonseasonal AR parameter, a_t is the white noise term at time t that is *identically independently distributed (IID)* with a mean of 0 and variance of σ_a^2 [i.e. IID $(0,\sigma_a^2)$].

The a_t sequence is referred to as *random shocks, disturbances, innovations* or *white noise terms*. After a model has been fit to a given time series and estimates have been obtained for the innovations, the estimates are called estimated innovations or *residuals*.

The most important assumption for the random shocks is that they are independently distributed. This infers that the a_t's are uncorrelated and must satisfy

$$E[a_t a_{t-k}] = \begin{cases} \sigma_a^2, & k = 0 \\ 0, & k \neq 0 \end{cases} \qquad [3.2.2]$$

The a_t's follow the same distribution and sometimes it is convenient to assume that the random shocks are normally distributed. This may be appropriate for estimation purposes, forecasting and simulation. In addition, if normal random variables are uncorrelated then they are also independent.

The difference equation in [3.2.1] can be written more economically by introducing the *backward shift operator B* which is defined by

$$Bz_t = z_{t-1}$$

and

$$B^k z_t = z_{t-k}$$

where k is a positive integer. By using the B operator, the Markov process in [3.2.1] is

$$z_t - \mu = \phi_1(Bz_t - \mu) + a_t$$

or

$$z_t - \mu - \phi_1(Bz_t - \mu) = a_t$$

By treating B as an algebraic operator and factoring, the above equation becomes

$$(1 - \phi_1 B)(z_t - \mu) = a_t$$

where $B\mu = \mu$ since the mean level is a constant at all times. The previous equation can also be given as

$$\phi(B)(z_t - \mu) = a_t \qquad [3.2.3]$$

where $\phi(B) = 1 - \phi_1 B$ is the nonseasonal AR operator or polynomial of order one.

3.2.2 Autoregressive Process of Order p

The Markov process with the single AR parameter, ϕ_1, is a special case of an *AR process of order p [i.e. AR(p)]* which is given as

$$z_t - \mu = \phi_1(z_{t-1} - \mu) + \phi_2(z_{t-2} - \mu) + \cdots + \phi_p(z_{t-p} - \mu) + a_t \qquad [3.2.4]$$

where ϕ_i is the ith *nonseasonal AR parameter*. By introducing the B operator, [3.2.4] can

equivalently be written as

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(z_t - \mu) = a_t$$

or

$$\phi(B)(z_t - \mu) = a_t \qquad [3.2.5]$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is the *nonseasonal AR operator* of order p.

Stationarity

The equation $\phi(B) = 0$ is referred to as the *characteristic equation* for the process. It can be shown (Box and Jenkins, 1976, Ch. 3; Pagano, 1973) that a necessary and sufficient condition for the process to have *stationarity* is that the roots of the characteristic equation must fall outside the *unit circle*. The unit circle is a circle of unit radius centered on the origin of a complex number graph where one axis is the real number component and the other axis forms the imaginary part of the complex number.

Based upon the work of Schur (1917), Pagano (1973) presented an algorithm which can be used to determine whether or not all the roots of a given operator lie outside the unit circle. Consider the situation where it is necessary to ascertain if all the roots of the operator, $\phi(B)$, for an AR(p) process fall outside the unit circle. The first step is to form the Schur matrix **A** of dimension p×p which has (i,j)th element

$$\sum_{k=0}^{min(i,j)} (\phi_{i-k-1}\phi_{j-k-1} - \phi_{p+1+k-i}\phi_{p+1+k-j})$$

where $\phi_{-k} = 0$ and $\phi_0 = 1$. The matrix **A** is actually the inverse of the covariance matrix of p successive observations for an AR(p) process (Siddiqui, 1958). Schur (1917) demonstrated that a necessary and sufficient condition for the roots of $\phi(B) = 0$ to lie outside the unit circle is for **A** to be positive definite. Because **A** is positive definite whenever the covariance matrix is positive definite (Pagano, 1973), to demonstrate that an AR(p) process is stationary it is only necessary to show that **A** is positive definite. A convenient way to do this is to calculate the *Cholesky decomposition* of **A** [see Wilkinson (1965) and Healy (1968)] given by

$$\mathbf{A} = \mathbf{M}\,\mathbf{M}^T$$

where **M** is a lower triangular matrix. If all the diagonal entries of **M** are positive, matrix **A** is positive definite. When there are one or more zero entries on the diagonal of **M** and all other entries are positive, **A** is positive semidefinite. If during the calculation of **M** a diagonal location is encountered where a zero or positive entry cannot be calculated, the Cholesky decomposition does not exist. However, when the Cholesky decomposition shows that **A** is positive definite, then the roots of $\phi(B) = 0$ lie outside the unit circle. For the case of the $\phi(B)$ operator, this property means that the process is stationary.

From [3.2.3] the characteristic equation for the Markov process is

$$(1 - \phi_1 B) = 0$$

By considering B as an algebraic variable, the root of the characteristic equation is $B = \phi_1^{-1}$. In order for ϕ_1^{-1} to lie outside the unit circle to ensure stationarity, then $|\phi_1| < 1$.

The stationarity condition automatically ensures that a process can be written in terms of the a_t's in what is called a pure MA process. For example, the AR(1) process in [3.2.3] can be expressed as

$$z_t - \mu = (1 - \phi_1 B)^{-1} a_t \qquad [3.2.6]$$
$$= (1 + \phi_1 B + \phi_1^2 B^2 + \cdots) a_t$$

Because $|\phi_1| < 1$ due to the stationarity condition, this infers that the infinite series $(1 - \phi_1 B)^{-1}$ will converge for $|B| \leq 1$. The beneficial consequences caused by the restriction upon ϕ_1 can also be explained by writing [3.2.6] as

$$z_t - \mu = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \phi_1^3 a_{t-3} + \cdots$$

If $|\phi_1| < 1$, the dependence of the deviation $(z_t - \mu)$ upon the white noise terms decreases further into the past. Alternatively, if $|\phi_1| \geq 1$, the dependence of $(z_t - \mu)$ upon the white noise would be greater for disturbances which happened well before the more recent shocks. Of course, this type of interpretation would not be meaningful for stationary processes and can be avoided if the stationarity condition is satisfied.

Autocorrelation Function

In order to study the properties of the theoretical ACF for a stationary AR(p) process, firstly multiply [3.2.4] by $(z_{t-k} - \mu)$ to obtain

$$(z_{t-k} - \mu)(z_t - \mu) = \phi_1 (z_{t-k} - \mu)(z_{t-1} - \mu) + \phi_2 (z_{t-k} - \mu)(z_{t-2} - \mu) + \cdots$$
$$+ \phi_p (z_{t-k} - \mu)(z_{t-p} - \mu) + (z_{t-k} - \mu) a_t \qquad [3.2.7]$$

By taking expected values of [3.2.7], the difference equation for the *autocovariance function of the AR(p) process* is

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p}, \quad k > 0 \qquad [3.2.8]$$

The term $E[(z_{t-k} - \mu) a_t]$ is zero for $k > 0$ because z_{t-k} is only a function of the disturbances up to time $t-k$ and a_t is uncorrelated with these shocks. To determine an expression for the theoretical *ACF for the AR(p) process*, divide [3.2.8] by γ_0 to obtain

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \quad k > 0$$

This equation can be equivalently written as

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) \rho_k = \phi(B) \rho_k = 0, \quad k > 0 \qquad [3.2.9]$$

where B operates on k instead of t. The general solution of the difference equation in [3.2.9] is (Box and Jenkins, 1976, p. 55)

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \cdots + A_p G_p^k \qquad [3.2.10]$$

where $G_1^{-1}, G_2^{-1}, \cdots, G_p^{-1}$, are distinct roots of the characteristic equation $\phi(B) = 0$ and the A_i's are constants. If a root G_i^{-1} is real then $|G_i^{-1}| > 1$ due to the stationarity conditions. Hence,

$|G_i|<1$ and $A_iG_i^k$ in [3.2.10] forms a damped exponential which geometrically decays to zero as k increases. Complex roots contribute a damped sine wave to the theoretical ACF in [3.2.10]. Consequently, the theoretical ACF for a stationary AR process will consist of a combination of damped exponential and sine waves.

St. Lawrence River Data: As mentioned in Section 2.5.4, when determining a model to fit to a given data set, it is desirable to have the theoretical ACF of the process to resemble statistically the sample ACF. Consider, for example, the average annual flows of the St. Lawrence River at Ogdensburg, New York. These flows are available from 1860 to 1957 in a report by Yevjevich (1963). The estimated ACF for these yearly flows is calculated using [2.5.9] and is shown in Figure 3.2.1. The 95% confidence limits are determined utilizing [2.5.11] by assuming that the sample ACF is not significantly different from zero after lag 0. As can be seen in Figure 3.2.1, the estimated ACF has significant non-zero values at lower lags and tends to follow a damped exponential curve. Because the theoretical ACF of an AR process behaves in this fashion, this indicates that perhaps some type of model which contains an AR component should be fit to the St. Lawrence flows.

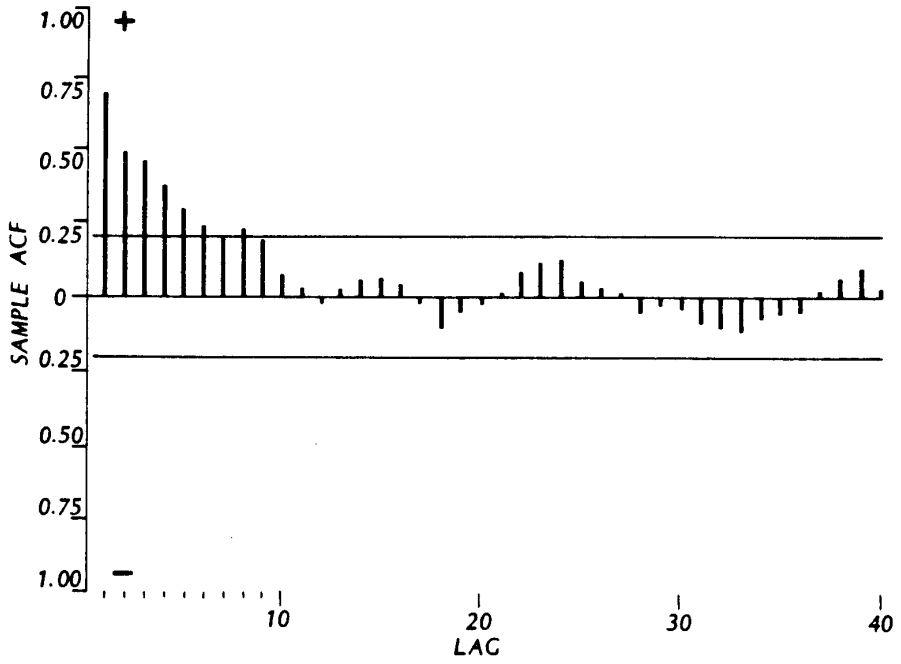

Figure 3.2.1. Sample ACF and 95% confidence limits for the average annual flows of the St. Lawrence River at Ogdensburg, New York.

Yule-Walker Equations

By substituting $k = 1,2, \ldots, p$, into [3.2.9], parameters can be expressed in terms of the theoretical ACF. The resulting set of linear equations are called the *Yule-Walker equations* [after Yule (1927) and Walker (1931)] and are given by

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + \cdots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + \cdots + \phi_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \cdots + \phi_p
\end{aligned} \qquad [3.2.11]$$

By writing the Yule-Walker equations in matrix form, the relationship for the AR parameters is

$$\boldsymbol{\phi} = \mathbf{P}_p^{-1}\boldsymbol{\rho}_p \qquad [3.2.12]$$

where

$$\boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}, \quad \boldsymbol{\rho}_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}, \quad \mathbf{P}_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix}$$

To obtain Yule-Walker estimates for the AR parameters, simply replace the ρ_k's in [3.2.12] by their estimates r_k, $k = 1,2,\ldots,p$, from [2.5.9]. The Yule-Walker estimates possess large sample efficiency and hence have minimum possible variances.

By setting $k = 0$ in [3.2.7] and taking expectations, the expression for the variance is

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma_a^2 \qquad [3.2.13]$$

where $E[z_t a_t] = \sigma_a^2$ since z_t is only correlated with a_t due to the most recent shock a_t. Upon dividing [3.2.13] by $\gamma_0 = \sigma_z^2$, the variance of the process can be expressed as

$$\sigma_z^2 = \frac{\sigma_a^2}{1 - \rho_1\phi_1 - \rho_2\phi_2 - \cdots - \rho_p\phi_p} \qquad [3.2.14]$$

Employing [3.2.13] and [2.5.8], the residual variance can be estimated using

$$\hat{\sigma}_a^2 = c_o - \sum_{i=1}^{p}\hat{\phi}_i c_i$$

In addition to the Yule-Walker estimator, other estimators are available for efficiently estimating the parameters of an AR model. One approach is to employ the maximum likelihood estimator presented in Section 6.2 and Appendix A6.1. A second procedure is to employ the Burg (1975) algorithm which is described by Haykin (1990, pp. 187-192).

Markov Process: As shown earlier in this section, in order for an AR(1) process to be stationary $-1<\phi_1<1$. By setting ϕ_2 to ϕ_p equal to zero, equation [3.2.11] becomes

$$\rho_1 = \phi_1$$

$$\rho_2 = \phi_1\rho_1 = \phi_1^2$$

$$\rho_3 = \phi_1\rho_2 = \phi_1^3$$

In general,

$$\rho_k = \phi_1^k. \qquad [3.2.15]$$

Because of the form of [3.2.15], the theoretical ACF attenuates exponentially to zero if ϕ_1 is positive but decays exponentially to zero and oscillates in sign when ϕ_1 is negative. From Figure 3.2.2, it can be seen that when ϕ_1 is assigned a positive value of 0.75, the theoretical ACF only possesses positive values which decay exponentially to zero for increasing lag. However, when ϕ_1 is given a negative value such as -0.75, the theoretical ACF oscillates in sign and decays exponentially to zero as shown in Figure 3.2.3. The variance of an AR(1) process is obtained from [3.2.14] and [3.2.15] as

$$\sigma_z^2 = \frac{\sigma_a^2}{1 - \rho_1\phi_1} = \frac{\sigma_a^2}{1 - \phi_1^2} \qquad [3.2.16]$$

Partial Autocorrelation Function

Because the ACF of an AR process attenuates and does not truncate at a specified lag, it would be advantageous to define a function which does cut off for an AR process. As explained in Chapter 5, such a device would be useful to employ in conjunction with the sample ACF and other tools for identifying the type of model to fit to a given data set.

Let ϕ_{kj} be the jth coefficient in a stationary AR process of order k so that ϕ_{kk} is the last coefficient. The Yule-Walker equations in [3.2.12] can then be equivalently written as

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ . \\ . \\ . \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ . \\ \rho_k \end{bmatrix} \qquad [3.2.17]$$

The coefficient ϕ_{kk} is a function of the lag k and is called the theoretical *partial autocorrelation function (PACF)*. Because of the definition of the theoretical PACF, it must be equal to zero after lag p for an AR(p) process. Furthermore, the possible values of ϕ_{kk} range from -1 to 1.

One method for estimating the PACF is to employ the Yule-Walker equations. By replacing ρ_k in [3.2.17] by its estimate r_k from [2.5.9], the estimates of ϕ_{kk}, $k = 1,2,...$, using Cramer's rule are

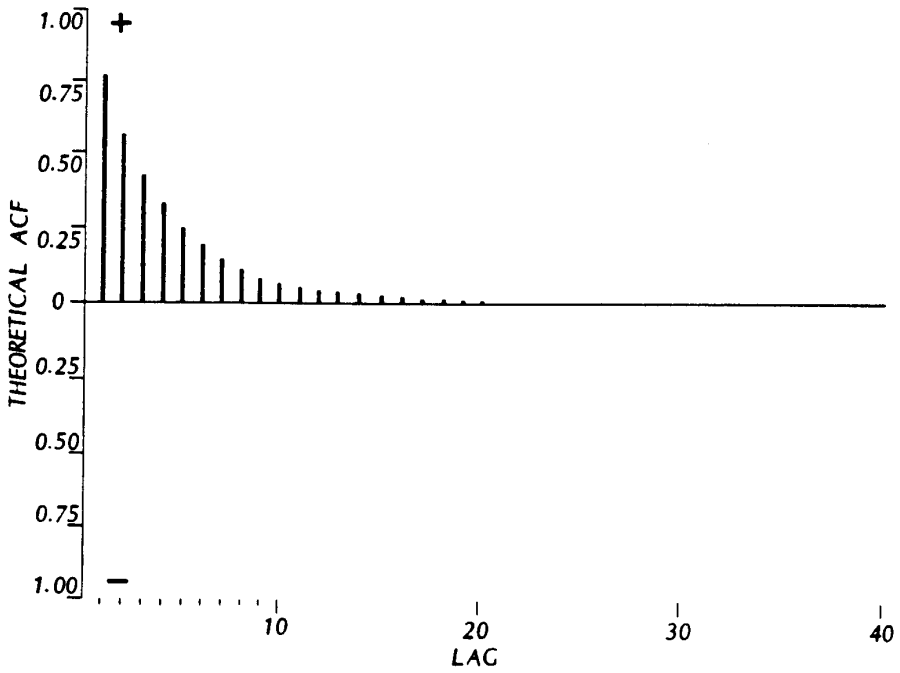

Figure 3.2.2. Theoretical ACF for a Markov Process with $\phi_1 = 0.75$.

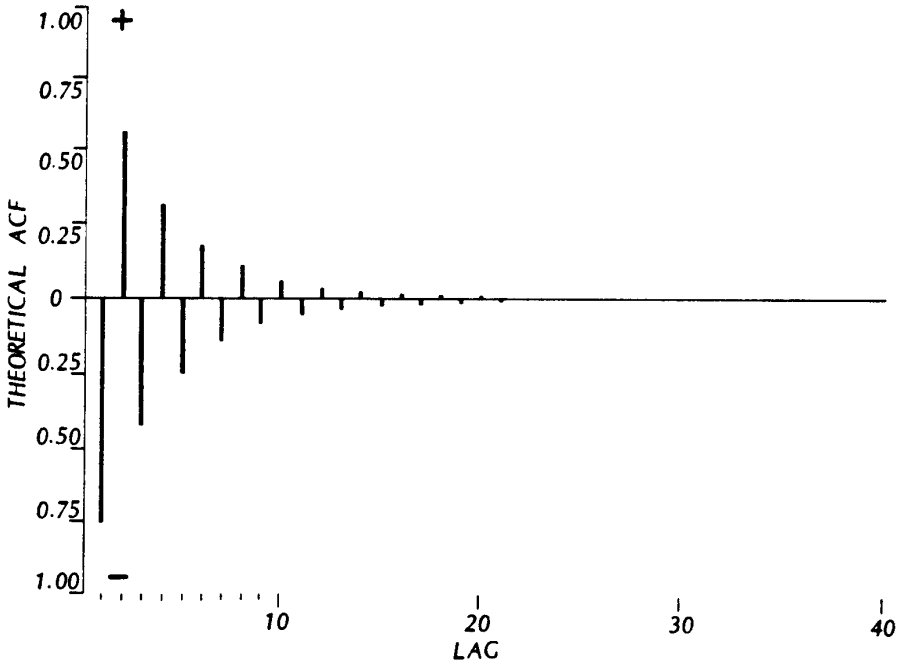

Figure 3.2.3. Theoretical ACF for a Markov Process with $\phi_1 = -0.75$.

$$\hat{\phi}_{11} = r_1,$$

$$\hat{\phi}_{22} = \begin{vmatrix} 1 & r_1 \\ r_1 & r_2 \end{vmatrix} \Big/ \begin{vmatrix} 1 & r_1 \\ r_1 & 1 \end{vmatrix} = \frac{r_2 - r_1^2}{1 - r_1^2}$$

etc. In order to make the estimation of the PACF computationally more economical, the recursive formulae of Durbin (1960) may be employed. However, as noted by Pagano (1972) and Box and Jenkins (1976), Durbin's method is numerically unstable, especially when the process approaches nonstationarity (i.e., the roots of characteristic equation are close to the unit circle).

An alternative procedure for estimating the PACF is to utilize the algorithm devised by Pagano (1972). The *Pagano algorithm* is numerically quite stable because it is based upon the Cholesky decomposition which is known to be stable (Wilkinson, 1965, pp. 231 and 244). The steps required in the Pagano algorithm for calculating the PACF up to lag p are given in Appendix A3.1. Because the algorithm is numerically stable and is also economical with respect to computational requirements, it is amenable for programming on the computer.

When plotting $\hat{\phi}_{kk}$ against lag k, approximate confidence limits must be given in order to decipher values of the estimated PACF which are significantly different from zero. If the process is AR(p), the sample PACF should not be significantly different from zero after lag p. Based upon the hypothesis that the process is AR(p), the estimated values of the PACF at lags greater than p are approximately normally independently distributed with a SE given by (Quenouille, 1949; Barndorff-Nielsen and Schou, 1973)

$$\text{SE}[\hat{\phi}_{kk}] \approx \frac{1}{\sqrt{N}} \qquad [3.2.18]$$

where N is the length of the time series.

St. Lawrence River Data: The graph of the estimated PACF for the average annual flows of the St. Lawrence River is shown in Figure 3.2.4. The 95% confidence limits are calculated by substituting $N = 97$ into [3.2.18] and plotting 1.96 times the SE for $\hat{\phi}_{kk}$ above and below the horizontal axis. It can be seen that there are rather large values for the estimated PACF at lags 1, 3 and 19. The unexpected big value at lag 19 could be due to chance alone or else the limited size of the sample which was used to estimate the PACF at lag 19. Because the estimated PACF cuts off after lag 3, this implies that an AR(3) process should perhaps be fitted to the data. In addition, because the sample PACF at lag 2 is not very large, perhaps the ϕ_2 parameter should be constrained to zero in the AR(3) model in order to reduce the number of model parameters. As shown in Section 6.4.2, the estimated model for the St. Lawrence data is

$$(1 - 0.619B - 0.177B^3)(z_t - 6818.63) = a_t \qquad [3.2.19]$$

where 6818.63 is the maximum likelihood estimate for the mean.

By substituting the values of the AR parameters for the model from [3.2.19] into the Yule-Walker equations in [3.2.11], the theoretical ACF can be determined. It can be seen from Figure 3.2.5 that the theoretical ACF for the St. Lawrence model in [3.2.19] is statistically similar to the

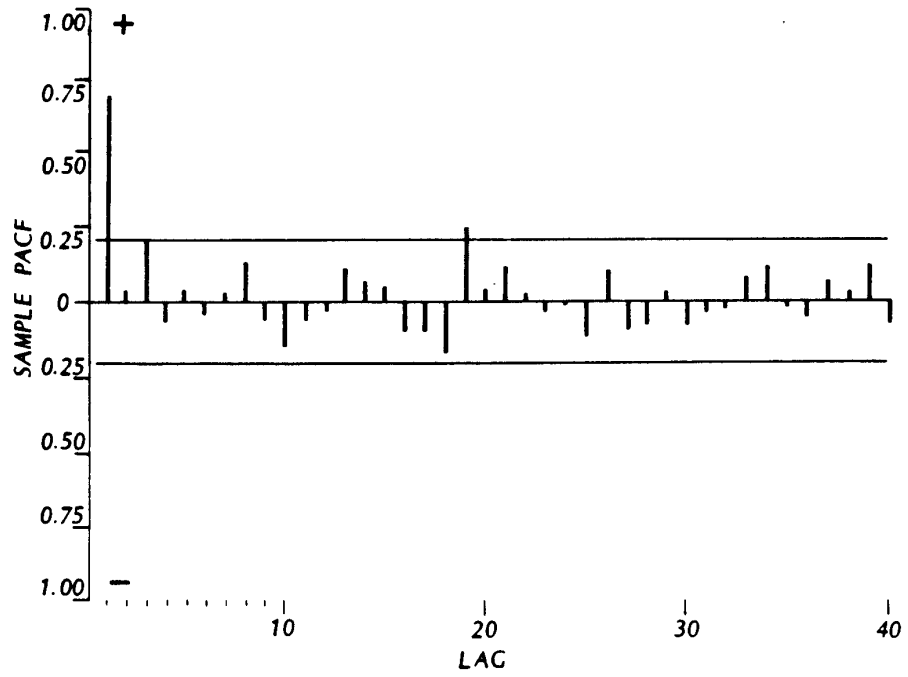

Figure 3.2.4. Sample PACF and 95% confidence limits for the average annual flows of the St. Lawrence River at Ogdensburg, New York.

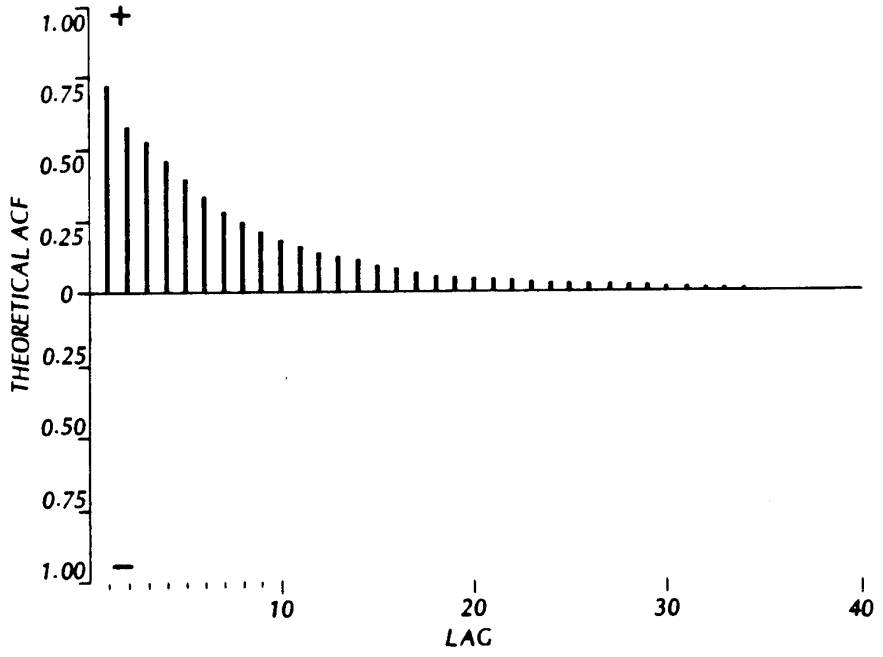

Figure 3.2.5. Theoretical ACF for the AR(3) model without ϕ_2 that is fitted to the average annual flows of the St. Lawrence River at Ogdensburg, New York.

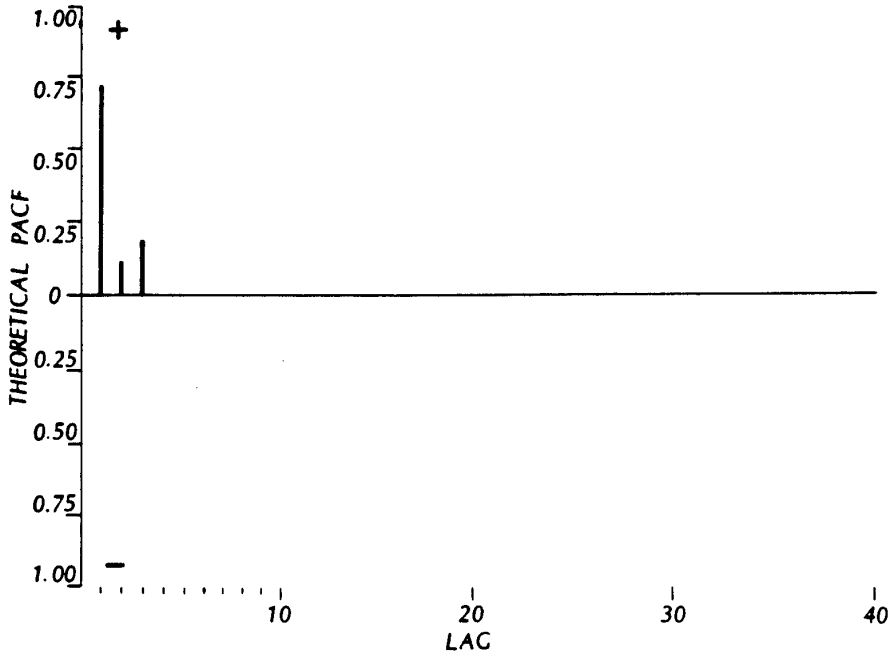

Figure 3.2.6. Theoretical PACF for the AR(3) model without ϕ_2 that is fitted to the average annual flows of the St. Lawrence River at Ogdensburg, New York.

sample ACF given in Figure 3.2.1. This information indicates that an AR model is a reasonable type of model to fit to the St. Lawrence River flows.

To further justify the use of the model in [3.2.19] for modelling the St. Lawrence River flows, the theoretical PACF can be compared to the sample PACF in Figure 3.2.4. In order to calculate the theoretical PACF, the values of theoretical ACF which were determined by substituting the estimates for the AR parameters in [3.2.19] into [3.2.11], are employed in [3.2.17]. The graph of the theoretical PACF for the St. Lawrence River flows is shown in Figure 3.2.6 and it can be seen that this plot is similar to the sample PACF in Figure 3.2.4.

3.3 MOVING AVERAGE PROCESSES

The MA model describes how an observation depends upon the current white noise term as well as one or more previous innovations. After examining the simplest type of MA model, the general form of the MA model is defined and its important theoretical properties are derived.

3.3.1 First Order Moving Average Process

When a time series value, z_t, is dependent only upon the white noise at time $t-1$ plus the current shock, the relationship is written as

$$z_t - \mu = a_t - \theta_1 a_{t-1} \qquad [3.3.1]$$

where θ_1 is the nonseasonal MA parameter. This process is termed a MA process of order one

and is denoted as MA(1). By introducing the B operator, the MA(1) process can be equivalently written as

$$\begin{aligned} z_t - \mu &= a_t - \theta_1 B a_t \\ &= (1 - \theta_1 B) a_t \\ &= \theta(B) a_t \end{aligned} \qquad [3.3.2]$$

where $\theta(B) = 1 - \theta_1 B$ is the nonseasonal MA operator or polynomial of order one.

3.3.2 Moving Average Process of Order q

The MA(1) process can be readily extended to the situation where there are q MA parameters. The *MA process* of order q is denoted by MA(q) and is written as

$$z_t - \mu = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \qquad [3.3.3]$$

where θ_j is the jth nonseasonal MA parameter. By employing the B operator, the MA(q) process can be more economically presented as

$$\begin{aligned} z_t - \mu &= a_t - \theta_1 B a_t - \theta_2 B^2 a_t - \cdots - \theta_q B^q a_t \\ &= (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) a_t \\ &= \theta(B) a_t \end{aligned} \qquad [3.3.4]$$

where $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$, is the *nonseasonal MA operator* or polynomial of order q.

Stationarity

The time series composed of the a_t's is assumed to be stationary. Because z_t in [3.3.4] is formed by a finite linear combination of the a_t's, then z_t must be stationary no matter what values the MA parameters possess. However, it is advantageous to put certain restrictions upon the range of values for the MA parameters. Consider, for example, the MA(1) process in [3.3.2]. By invoking the binomial theorem, this process can be equivalently written as an infinite AR process given as

$$\begin{aligned} a_t &= (1 - \theta_1 B)^{-1} (z_t - \mu) \\ &= (1 + \theta_1 B + \theta_1^2 B^2 + \theta_1^3 B^3 + \cdots)(z_t - \mu) \end{aligned} \qquad [3.3.5]$$

In order for the infinite series $(1 - \theta_1 B)^{-1}$ to converge for $|B| \leq 1$, the parameter θ_1 must be restricted to have an absolute value less than unity. Another way to interpret the restriction upon θ_1 is to write [3.3.5] as

$$z_t - \mu = a_t - \theta_1 (z_{t-1} - \mu) - \theta_1^2 (z_{t-2} - \mu) - \theta_1^3 (z_{t-3} - \mu) - \cdots \qquad [3.3.6]$$

If $|\theta_1| > 1$, it can be seen in [3.3.6] that the current deviation $(z_t - \mu)$ depends more on events that happened further in the past because θ_1^k increases as the lag k gets larger. When $|\theta_1| = 1$,

something that took place a long time ago has as much influence as a recent observation upon the current measurement. In order to avoid these situations, it is necessary that $|\theta_1| < 1$. This is equivalent to stipulating that the root $B = \theta_1^{-1}$ of the characteristic equation $(1 - \theta_1 B) = 0$ must lie outside the unit circle. Consequently, the stationary MA(1) process can only be meaningfully expressed as an infinite AR process if a restriction is placed upon the MA parameter. This restriction is referred to as the invertibility condition and is independent of the stationarity requirements of a process.

Invertibility

The characteristic equation for a MA(q) process is

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q = 0 \qquad [3.3.7]$$

In order for a MA(q) process to be *invertible*, the roots of the characteristic equation must lie outside the unit circle.

An inherent advantage of the invertibility condition is that it does not cause a loss in generality of the MA process. As shown by Fuller (1976, pp. 64-66) and discussed by Anderson (1971, p. 204), any finite MA process whose characteristic equation has some roots greater than one and some less than one can be given a representation whose characteristic equation has all roots greater than one in absolute value. Consequently, the invertibility condition does not limit the ability to identify a suitable invertible model to fit to a given series. In addition, if the invertibility condition is satisfied, a MA process can be expressed as a pure AR process. Finally, when the residuals are being established for a model which is being fitted to a specified time series, the calculation of the residuals will be ill-conditioned if the invertibility condition is not met.

Autocorrelation Function

By using [2.5.3] and [3.3.3], the *autocovariance function of a MA(q) process* is

$$\begin{aligned}\gamma_k &= E[(z_t - \mu)(z_{t-k} - \mu)] \\ &= E[(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q})(a_{t-k} - \theta_1 a_{t-k-1} - \theta_2 a_{t-k-2} \\ &\qquad - \cdots - \theta_q a_{t-k-q})] \end{aligned} \qquad [3.3.8]$$

After multiplication and taking expected values, the autocovariance function is

$$\gamma_k = \begin{cases} (-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q)\sigma_a^2 \quad , \quad k = 1,2,\ldots,q \\ 0 \quad , \quad k>q \end{cases} \qquad [3.3.9]$$

where $\theta_0 = 1$ and $\theta_{-k} = 0$ for $k \geq 1$. When k is set equal to zero in [3.3.8], the variance is

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma_a^2 \qquad [3.3.10]$$

By dividing the autocovariance function by the variance, the theoretical *ACF for a MA(q) process* is found to be

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2} & , \quad k = 1,2,\ldots,q \\ 0 \quad , \quad k > q \end{cases} \qquad [3.3.11]$$

Partial Autocorrelation Function

It is shown in [3.3.5] that a MA(1) process can be equivalently written as an infinite AR process. In general, any finite invertible MA process can be expressed as an infinite AR process. Because the PACF is theoretically defined to be zero after lag p for a finite AR(p) process, the PACF must therefore attenuate at increasing lags for a MA process or equivalently an infinite AR process.

Temperature Data: From [3.3.11], it can be seen that the theoretical ACF for a MA(q) process is exactly zero after lag q. If the sample ACF is tabulated for a given time series using [2.5.9], then the estimated ACF should not be significantly different from zero after lag q if the underlying process is MA(q). For instance, the sample ACF for the average annual temperature data from the English Midlands is shown in Figures 2.5.1 and 2.5.2 in Section 2.5.4. Because the estimated ACF is not significantly different from zero after lag 2, it is reasonable to fit a MA(2) model to the data. Using the estimator described in Appendix A6.1, the estimated model for the temperature data is found to be

$$z_t - 9.216 = (1 + 0.111B + 0.197B^2)a_t \qquad [3.3.12]$$

By substituting the estimates for the MA parameters from [3.3.12] into [3.3.11] (where $\rho_k = 0$ for $k > 2$), the theoretical ACF can be calculated for the MA(2) model. By comparing the theoretical ACF in Figure 2.5.3 to the estimated ACF in Figure 2.5.1, it can be seen that the theoretical ACF for the MA(2) model mimics the estimated ACF.

To calculate the sample PACF for the temperature data from the English Midlands, one first must determine the sample ACF using [2.5.9]. Following this, Pagano's (1972) algorithm, outlined in Appendix A3.1, can be used to solve [3.2.17] in order to determine the sample PACF. The sample PACF along with the 95% confidence limits for the English temperature data, are displayed in Figure 3.3.1. The sample PACF truncates after lag 2 except for a rather large value at lag 15 which is probably due to chance. However, the plot of the sample ACF in Figure 2.5.2 reveals that it also cuts off after lag 2. Hence, either a MA(2) or an AR(2) model may adequately model the temperature data. As is shown in Section 3.4.3 the two models are in fact shown to be almost the same by expressing the AR(2) model as an infinite MA model in which the coefficients after lag 2 are negligible.

After substituting the values of the theoretical ACF for the MA(2) model in [3.3.12] into [3.2.17], one can employ Pagano's (1972) algorithm outlined in Appendix A3.1 to determine the theoretical PACF. It can be seen that the theoretical PACF in Figure 3.3.2 for the estimated MA(2) model closely resembles the sample PACF of the temperature data in Figure 3.3.1.

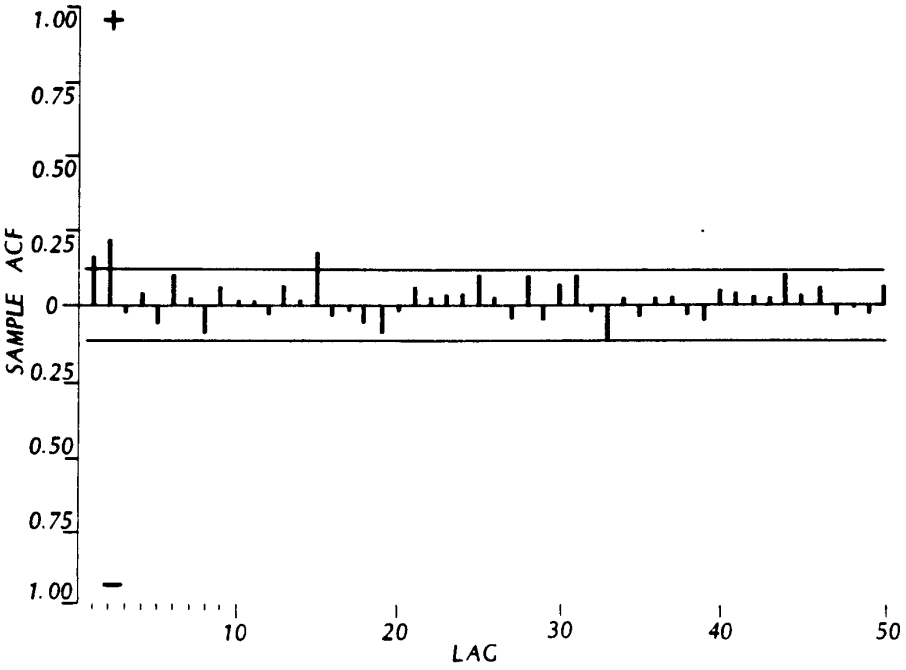

Figure 3.3.1. Sample PACF and 95% confidence limits for the annual temperature data from the English Midlands.

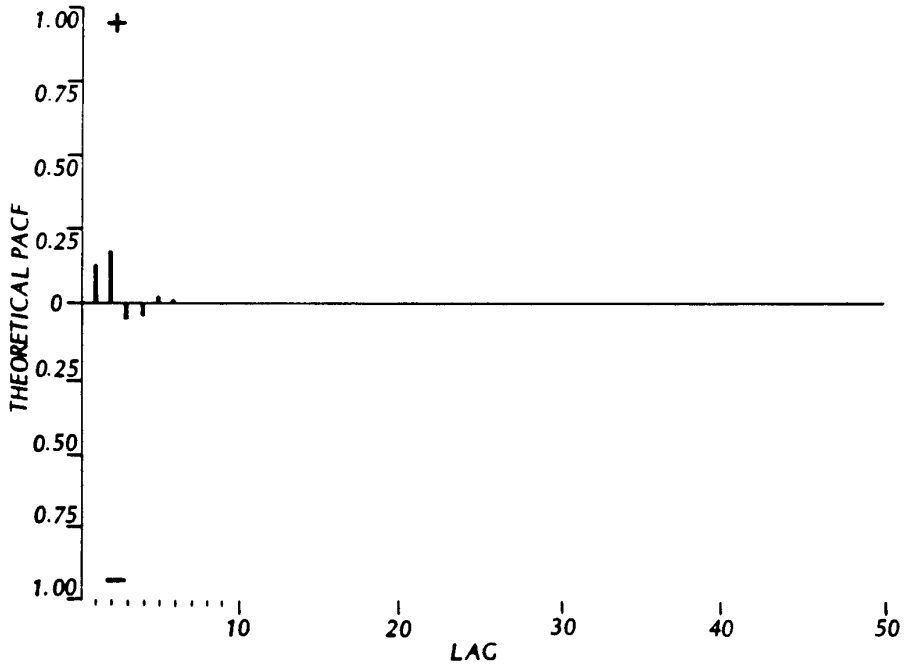

Figure 3.3.2. Theoretical PACF for the MA(2) model fitted to the annual temperature data from the English Midlands.

First Order Moving Average Process

The MA(1) process is given in [3.3.2]. This process is stationary for all values of θ_1 but for invertibility $|\theta_1| < 1$. When the invertibility condition is satisfied, the MA(1) process can be equivalently written as an infinite AR process as is shown in [3.3.5].

By utilizing [3.3.10], the variance of the MA(1) process is

$$\gamma_0 = (1 + \theta_1^2)\sigma_a^2 \qquad [3.3.13]$$

From [3.3.11] the theoretical ACF is

$$\rho_k = \begin{cases} \dfrac{-\theta_1}{1+\theta_1^2} & , \quad k = 1 \\ 0 & , \quad k \geq 2 \end{cases} \qquad [3.3.14]$$

By substituting $\rho_1 = -\theta_1/(1 + \theta_1^2)$ and $\rho_k = 0$, for $k > 1$, into [3.2.17], the PACF can be shown to be

$$\phi_{kk} = -\theta_1^k(1 - \theta_1^2)/(1 - \theta_1^{2(k+1)}) \qquad [3.3.15]$$

Because $|\theta_1| < 1$ for an invertible MA(1) process, the theoretical PACF decreases in value for increasing lag and follows a damped exponential curve. Because of the form of [3.3.15], it can be seen that $|\phi_{kk}| < |\theta_1|^k$. When θ_1 is positive, then from [3.3.14] ρ_1 is negative and the PACF values in [3.3.15] are also negative. On the other hand, if θ_1 is negative, ρ_1 is positive and the PACF values alternate in sign.

3.4 AUTOREGRESSIVE-MOVING AVERAGE PROCESSES

As noted in Sections 1.3 and 5.2.4, a key modelling principle is to have as few parameters as possible in the model. If, for example, the sample ACF for a given data set possesses a value which is significantly different from zero only at lag one, then it may be appropriate to fit a MA(1) model to the data. An AR model may require quite a few AR parameters in order to adequately model the same time series. When the sample PACF for another data set cuts off at lag 2, then the most parsimonious model to fit the time series may be an AR(2) model. In situations where both the sample ACF and PACF attenuate for a certain time series, it may be advantageous to have a model which contains both AR and MA parameters. In this way, the fitted model can be kept as simple as possible by keeping the number of model parameters to a minimum.

3.4.1 First Order Autoregressive-First Order Moving Average Process

If a process consists of both AR and MA parameters, it is called an ARMA process. When there is one AR and one MA parameter the ARMA process is denoted as ARMA(1,1) and the equation for this process is

$$(z_t - \mu) - \phi_1(z_{t-1} - \mu) = a_t - \theta_1 a_{t-1} \qquad [3.4.1]$$

By utilizing the B operator, the ARMA(1,1) process can be equivalently written as

$$(1 - \phi_1 B)(z_t - \mu) = (1 - \theta_1 B)a_t$$

or

$$\phi(B)(z_t - \mu) = \theta(B)a_t \qquad [3.4.2]$$

where $\phi(B) = 1 - \phi_1 B$ and $\theta(B) = 1 - \theta_1 B$ are, respectively, the AR and MA operators of order one.

3.4.2 General Autoregressive-Moving Average Process

In general, an *ARMA process* may consist of p AR parameters and q MA parameters. Such a process is denoted by ARMA(p,q) and is written as

$$(z_t - \mu) - \phi_1(z_{t-1} - \mu) - \phi_2(z_{t-2} - \mu) - \cdots - \phi_p(z_{t-p} - \mu)$$
$$= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \qquad [3.4.3]$$

By implementing the B operator, [3.4.3] can be presented more conveniently as

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(z_t - \mu) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)a_t$$

or

$$\phi(B)(z_t - \mu) = \theta(B)a_t \qquad [3.4.4]$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is the AR operator of order p and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is the MA operator of order q.

As mentioned by Box and Jenkins (1976, p. 74), an ARMA(p,q) process may be interpreted in two ways. It can be considered as a pth order AR process given by

$$\phi(B)(z_t - \mu) = e_t \qquad [3.4.5]$$

where e_t follows the qth order MA process

$$e_t = \theta(B)a_t \qquad [3.4.6]$$

Alternatively, an ARMA(p,q) process can be thought of as a qth order MA process

$$(z_t - \mu) = \theta(B)b_t \qquad [3.4.7]$$

where b_t follows the pth order AR process

$$\phi(B)b_t = a_t \qquad [3.4.8]$$

By substituting either e_t from [3.4.6] into [3.4.5] or else b_t from [3.4.8] into [3.4.7], it follows that

$$\phi(B)(z_t - \mu) = \theta(B)a_t$$

The ARMA(p,q) process contains both the pure AR and MA processes as subsets. Consequently, an AR(p) process is equivalent to an ARMA(p,0) process while a MA(q) process is the same as an ARMA(0,q) process. The ARMA(p,q) family of processes are also sometimes referred to as stationary nonseasonal Box-Jenkins processes because of the comprehensive presentation of these models in the book by Box and Jenkins (1976).

Stationarity and Invertibility

The conditions regarding stationarity and invertibility for AR and MA processes, also hold for ARMA processes. In order for an ARMA(p,q) process to be stationary the roots of the characteristic equation $\phi(B) = 0$ must fall outside the unit circle. Similarly, the roots of $\theta(B) = 0$ must fall outside the unit circle if the process is invertible and can be expressed as a pure AR process.

Autocorrelation Function

The theoretical ACF for an ARMA(p,q) process is derived in a fashion which is similar to that used for an AR process in Section 3.2.2. Multiply both sides of [3.4.3] by $(z_{t-k} - \mu)$ and take expectations to obtain

$$\begin{aligned}\gamma_k - \phi_1\gamma_{k-1} - \phi_2\gamma_{k-2} - \cdots - \phi_p\gamma_{k-p} \\ = \gamma_{za}(k) - \theta_1\gamma_{za}(k-1) - \theta_2\gamma_{za}(k-2) - \cdots - \theta_q\gamma_{za}(k-q)\end{aligned} \qquad [3.4.9]$$

where $\gamma_k = E[(z_{t-k} - \mu)(z_t - \mu)]$ is the theoretical autocovariance function and $\gamma_{za}(k) = E[(z_{t-k} - \mu)a_t]$ is the cross covariance function between z_{t-k} and a_t. Since z_{t-k} is dependent only upon the shocks which have occurred up to time $t-k$, it follows that

$$\begin{aligned}\gamma_{za}(k) &= 0 \quad , \; k > 0 \\ \gamma_{za}(k) &\neq 0 \quad , \; k \leq 0\end{aligned} \qquad [3.4.10]$$

Because of the $\gamma_{za}(k)$ terms in [3.4.9], it is necessary to derive other relationships before it is possible to solve for the autocovariances. This can be effected by multiplying [3.4.3] by a_{t-k} and taking expectations to get

$$\begin{aligned}\gamma_{za}(-k) - \phi_1\gamma_{za}(-k+1) - \phi_2\gamma_{za}(-k+2) - \cdots - \phi_p\gamma_{za}(-k+p) \\ = -[\theta_k]\sigma_a^2\end{aligned} \qquad [3.4.11]$$

where

$$[\theta_k] = \begin{cases} \theta_k & , \; k = 1,2,\ldots,q \\ -1 & , \; k = 0 \\ 0 & , \; \textit{otherwise} \end{cases}$$

and $E[a_{t-k}a_t]$ is defined in [3.2.2].

Equations [3.4.9] and [3.4.11] can be employed to solve for the theoretical *autocovariance function for an ARMA(p,q) process*. For $k > q$, [3.4.9] reduces to

$$\gamma_k - \phi_1\gamma_{k-1} - \phi_2\gamma_{k-2} - \cdots - \phi_p\gamma_{k-p} = 0$$

or

$$\phi(B)\gamma_k = 0 \qquad [3.4.12]$$

If $k > r = \max(p,q)$, [3.4.12] may be used to calculate the γ_k directly from the previous values. For $k = 0,1,2,\ldots,r$, use [3.4.11] to solve for the cross covariances, $\gamma_{za}(k)$, and then substitute the $\gamma_{za}(k)$ into [3.4.9]. By employing the algorithm of McLeod (1975) outlined in Appendix A3.2, the resulting equations can be solved to determine the theoretical autocovariance function for any ARMA(p,q) process where the values of the parameters are known. The theoretical ACF can then be determined by simply dividing by the variance.

By dividing [3.4.12] by γ_0, the difference equation for the theoretical *ACF for an ARMA(p,q) process* is

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\rho_k = \phi(B)\rho_k = 0 \quad , \; k > q \qquad [3.4.13]$$

Except for the fact that [3.4.13] is only valid beyond lag q, the equation is identical to [3.2.9] which is the theoretical ACF for an AR(p) process. Hence, the attenuating behaviour of the ACF beyond lag q for an ARMA(p,q) process is due to the AR component in the model and the starting values for the difference equation. If $q - p < 0$ the entire theoretical ACF, ρ_j, for $j = 0,1,2,...$, will be composed of a mixture of damped exponential and/or damped sine waves which possess characteristics controlled by $\phi(B)$ and the starting values. When $q - p \geq 0$ the $q - p + 1$ initial values $\rho_0,\rho_1,\rho_2,\ldots,\rho_{q-p}$ will not follow this pattern. Furthermore, because of the structure of [3.4.9], the autocorrelations $\rho_1,\rho_2,\ldots,\rho_q$, are a function of both the MA and AR parameters.

Partial Autocorrelation Function

As a result of the MA operator, the ARMA(p,q) process in [3.4.4] can be written as an infinite AR process given by

$$a_t = \theta(B)^{-1}\phi(B)(z_t - \mu) \qquad [3.4.14]$$

where $\theta(B)^{-1}$ is an infinite series in B. Since the definition of the PACF is based upon an AR process, the theoretical PACF is infinite in extent and attenuates with increasing lag. At higher lags, the behaviour of the PACF depends upon the MA parameters and is dominated by a mixture of damped exponentials and/or damped sine waves.

Douglas Fir Tree Ring Data: Because both the ACF and PACF die off for an ARMA(p,q) process, it is sometimes difficult to determine which type of ARMA model to fit to a given data set. Often, it is necessary to study two or three tentative ARMA models. For instance, consider the time series of 700 tree ring indices for Douglas fir at the Navajo National Monument in Arizona. This data is available from 1263 to 1962 and is listed in a report by Stokes et al. (1973). The plots of the sample ACF and PACF are displayed in Figures 3.4.1 and 3.4.2, respectively, along with the 95% confidence limits. Because both plots seem to attenuate, it may be appropriate to

fit some type of ARMA(p,q) model to the data. The large values of both the ACF and PACF at lag one indicate that perhaps an ARMA(1,1) model may adequately model the data, although other ARMA models should perhaps also be examined. Using the estimator described in Appendix A6.1, the estimated ARMA(1,1) model is

$$(1 - 0.682B)(z_t - 99.400) = (1 - 0.424B)a_t \qquad [3.4.15]$$

This calibrated model satisfies the diagnostic checks described in Chapter 7.

By using the parameter values for the tree ring model given in [3.4.15] as input to equations [3.4.9] and [3.4.11], the theoretical ACF can be calculated. The theoretical ACF for the tree ring model shown in Figure 3.4.3 is statistically similar to the sample ACF in Figure 3.4.1. To calculate the theoretical PACF using Pagano's algorithm in Appendix A3.1, the values of the theoretical ACF are substituted for the ρ_k's in [3.2.17]. The theoretical PACF in Figure 3.4.4 has the same form as the sample PACF in Figure 3.4.2. Because the fitted ARMA model appears to statistically preserve both the historical ACF and PACF, this fact enhances the desirability of ARMA models for use in the natural sciences. Additionally, Section 3.6 explains why ARMA models are suitably designed for capturing the physical chracteristics of annual streamflows.

ARMA(1,1) Process

The ARMA(1,1) process is given in [3.4.2]. As is the case for the AR(1) process (see Section 3.2.2), in order for the ARMA(1,1) process to be stationary, $|\phi_1| < 1$. Similarly, because the MA(1) process is invertible if $|\theta_1| < 1$ (see Section 3.3.2), the ARMA(1,1) process is invertible when the same conditions are placed upon θ_1.

To derive the autocovariance function for an ARMA(1,1) model, first use [3.4.9] to obtain

$$\gamma_0 = \phi_1\gamma_1 + \sigma_a^2 - \theta_1\gamma_{za}(-1)$$

$$\gamma_1 = \phi_1\gamma_0 - \theta_1\sigma_a^2$$

$$\gamma_k = \phi_1\gamma_{k-1}\,, \quad k \geq 2$$

Next, after setting $k = 1$, employ [3.4.11] to get

$$\gamma_{za}(-1) = (\phi_1 - \theta_1)\sigma_a^2$$

where $\gamma_{za}(0) = \sigma_a^2$ in both [3.4.9] and [3.4.11]. Upon substituting $\gamma_{za}(-1)$ into the previous equation for γ_0, the autocovariances for an ARMA(1,1) process are found to be

$$\gamma_0 = \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2}\sigma_a^2$$

$$\gamma_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2}\sigma_a^2$$

$$\gamma_k = \phi_1\gamma_{k-1}\,, \quad k \geq 2 \qquad [3.4.16]$$

By dividing by γ_0, the theoretical ACF of an ARMA(1,1) process is

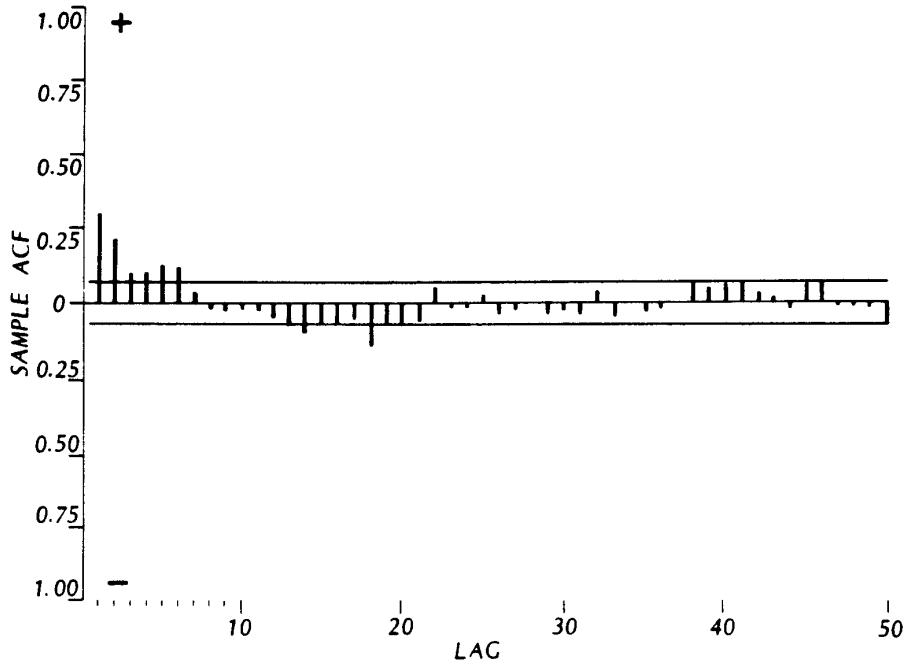

Figure 3.4.1. Sample ACF and 95% confidence limits for the Douglas Fir tree ring series at Navajo National Monument in Arizona.

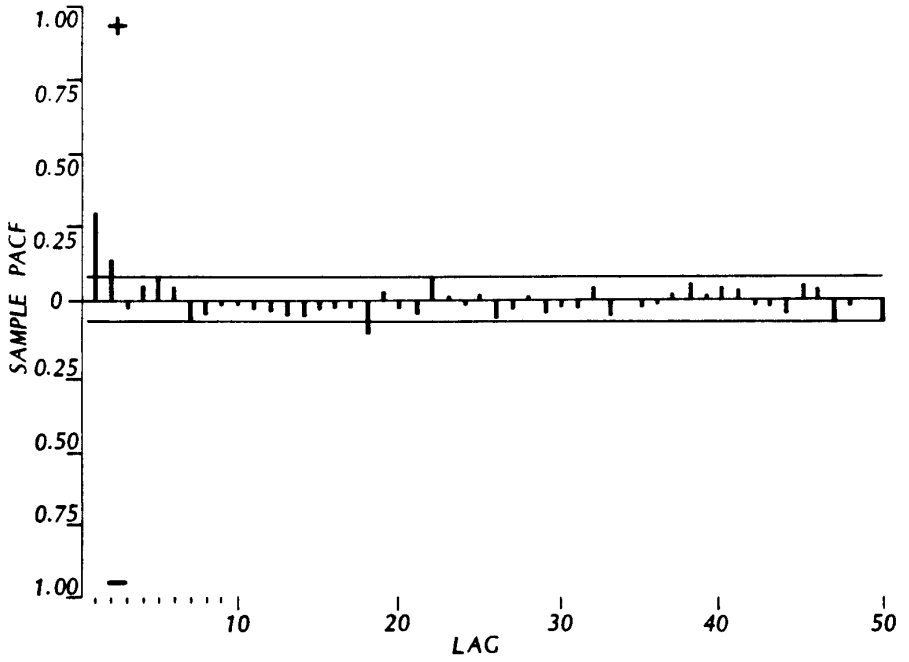

Figure 3.4.2. Sample PACF and 95% confidence limits for the Douglas Fir tree ring series at Navajo National Monument in Arizona.

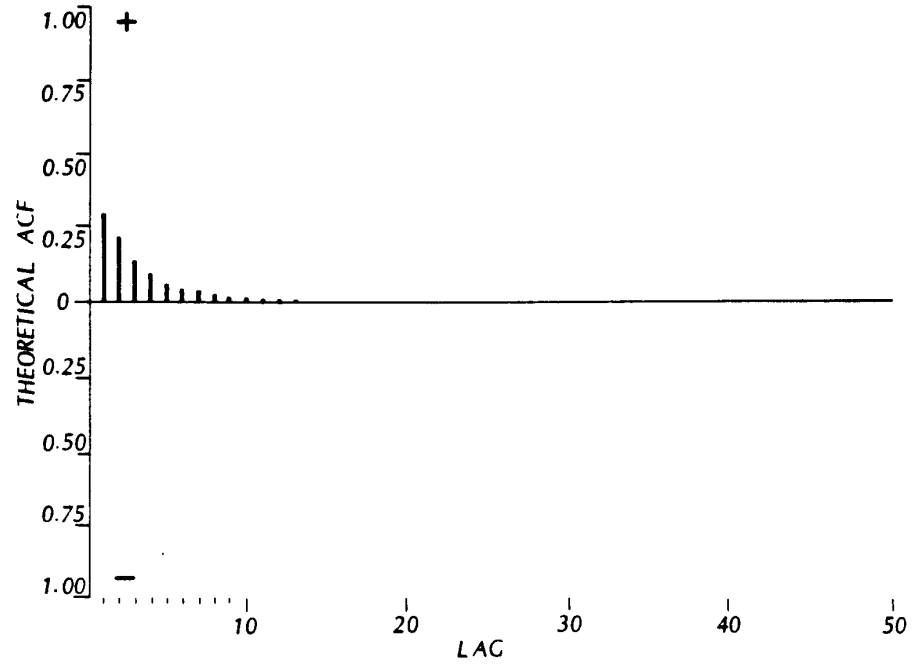

Figure 3.4.3. Theoretical ACF for the ARMA(1,1) model fitted to the Douglas Fir tree ring series at Navajo National Monument in Arizona.

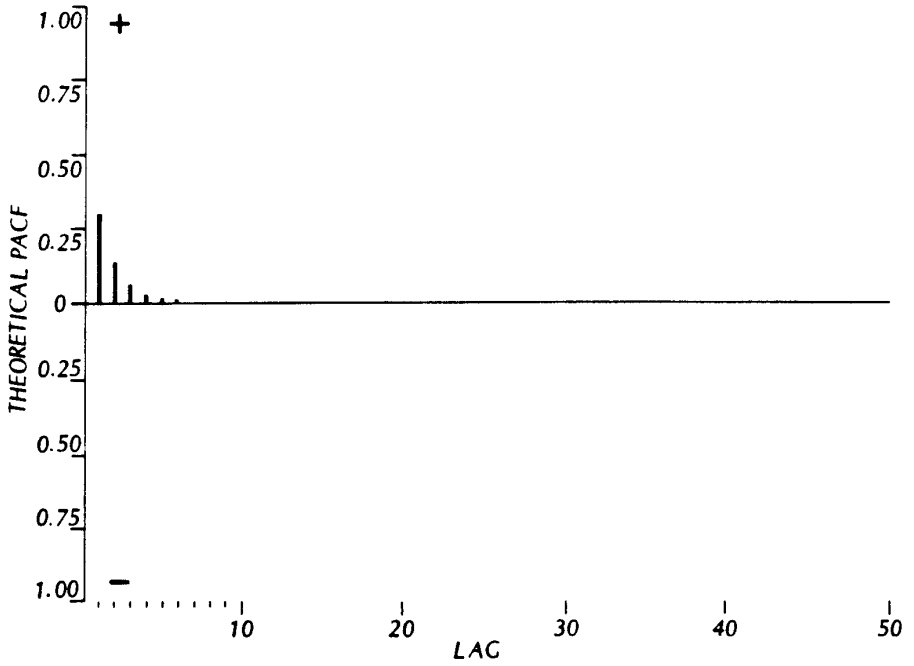

Figure 3.4.4. Theoretical PACF for the ARMA(1,1) model fitted to the Douglas Fir tree ring series at Navajo National Monument in Arizona.

$$\rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1}$$

$$\rho_k = \phi_1\rho_{k-1} \ , \quad k \geq 2 \qquad [3.4.17]$$

From [3.4.17], it can be seen that ρ_1 is a function of both the MA and AR parameters. The ACF at lag 2 depends only upon ϕ_1 and also the starting value ρ_1. Furthermore, the theoretical ACF decays exponentially from the starting value ρ_1. This exponential decay is even when ϕ_1 is positive and is oscillatory whenever ϕ_1 is negative. In addition, the sign of $(\phi_1 - \theta_1)$ dictates the sign of ρ_1 and also from which side of zero the exponential decay commences.

By substituting the theoretical ACF in [3.4.17] into the Yule-Walker equations which are given in [3.2.17], the theoretical PACF can be determined for the ARMA(1,1) process. At lag 1, $\phi_{11} = \rho_1$, while for lags greater than one the PACF of an ARMA(1,1) process behaves like the PACF of a MA(1) process (see Section 3.3.2) and hence follows the form of a damped exponential. When θ_1 is positive, the PACF consists of an evenly damped exponential which decays from ρ_1, where the sign of ρ_1 is determined by the sign of $(\phi_1 - \theta_1)$. If θ_1 is negative, the PACF is dominated by an oscillating exponential which attenuates from $\phi_{11} = \rho_1$, where the sign of ρ_1 is determined by $(\phi_1 - \theta_1)$.

3.4.3 Three Formulations of the Autoregressive-Moving Average Process

An ARMA(p,q) process can be expressed in three explicit forms. One formulation is to use the difference equation given in [3.4.4]. A second method is to express the process as a pure MA process. This is also referred to as the random shock form of the process. Finally, the third option is to formulate the process as a pure AR process which is also called the inverted form of the process.

Random Shock Form

Because $\phi(B)$ and $\theta(B)$ can be treated as algebraic operators, the ARMA(p,q) process can be written in *random shock form* as

$$\begin{aligned}(z_t - \mu) &= \phi(B)^{-1}\theta(B)a_t \\ &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \\ &= a_t + \psi_1 B a_t + \psi_2 B^2 a_t + \cdots \\ &= (1 + \psi_1 B + \psi_2 B^2 + \ldots)a_t \\ &= \psi(B)a_t \end{aligned} \qquad [3.4.18]$$

where $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots$, is the *random shock or infinite MA operator* and ψ_i is the ith parameter, coefficient or weight of $\psi(B)$. It is often convenient to express an ARMA process in the form given in [3.4.18] for both theoretical and application purposes. For instance, the ψ weights are required in Section 8.2.3 to calculate the variance of the forecasts. As explained in Section 9.3, one way to simulate data is to first express an ARMA model in random shock form

and then use this format of the model for simulation purposes. By writing each member of a set of ARMA models in random shock form, the models can be conveniently compared by looking at the magnitude and sign of the ψ parameters. Furthermore, nonstationary processes (see Chapter 4) and seasonal processes (see Part VI) can also be written in the general random shock form of the process.

As is noted in Section 3.2.2, if an AR(p) process is stationary the roots of $\phi(B) = 0$ must lie outside the unit circle and this insures that the process can also be written as an infinite MA process which will converge for $|B| \leq 1$. Consequently, a necessary condition for stationarity for an ARMA(p,q) process, is that the weights $\psi_1,\psi_2,...$, in $\psi(B) = \phi(B)^{-1}\theta(B)$, form a convergent series for $|B| \leq 1$ [see Box and Jenkins (1976, Appendix A3.1, pp. 80-82) for a mathematical proof]. The stationarity requirement is proven by examining the theoretical autocovariance which is given by

$$\gamma_k = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \quad , \quad k = 0,1,2,... \qquad [3.4.19]$$

By substituting $k = 0$ into [3.4.19], the variance is found to be

$$\gamma_0 = \sigma_z^2 = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2 \qquad [3.4.20]$$

In order to have a finite variance and hence stationarity, the ψ weights must decrease in a manner which allows the right side of [3.4.20] to converge.

To develop a relationship for determining the ψ parameters or weights, first multiply [3.4.18] by $\phi(B)$ to obtain

$$\phi(B)(z_t - \mu) = \phi(B)\psi(B)a_t$$

From [3.4.4], $\theta(B)a_t$ can be substituted for $\phi(B)(z_t - \mu)$ in the previous equation to get

$$\phi(B)\psi(B) = \theta(B)$$

The ψ weights can be conveniently determined by expressing the above equation as

$$\phi(B)\psi_k = -\theta_k \qquad [3.4.21]$$

where B operates on k, $\psi_0 = 1$, $\psi_k = 0$ for $k < 0$, and $\theta_k = 0$ if $k > q$. When $k > q$ then $\theta_k = 0$ and ψ_k in [3.4.21] satisfies the same difference equation as the theoretical ACF of an AR process and also an ARMA process in [3.2.9] and [3.4.13], respectively. Consequently, when ψ_k is plotted against lag k it will follow the same type of patterns as the theoretical ACF of the process. For increasing lags, the graph may follow a smooth exponential decay, an exponential decline that alternates in sign, or a damped sinusoidal decay.

Given the AR and MA parameters, one can employ [3.4.21] to calculate the random shock parameters. To decide upon how many ψ coefficients to estimate, one can calculate enough ψ coefficients to keep the relative error in the variance of the random shock process less than a specified error level. When σ_a^2 is assumed to be one and the ψ's are only considered up to lag q', from [3.4.20] the variance of the random shock process is approximately given by

$$\sum_{j=0}^{q'} \psi_j^2$$

When it is assumed that $\sigma_a^2 = 1$, one can calculate the variance, γ_0, of the given ARMA(p,q) process by solving [3.4.9] and [3.4.11]. Consequently, the relative absolute error due to the random shock approximation is

$$\left| \frac{\gamma_0 - \sum_{j=0}^{q'} \psi_j^2}{\gamma_0} \right|$$

One can choose q' to be just large enough to cause the above expression to have a value less than the specified error level. To demonstrate how the ψ coefficients are calculated using [3.4.21], two examples are now given.

Example Using the Temperature Model: In Section 3.3.2 it is noted that it may be appropriate to fit either a MA(2) or an AR(2) model to the annual temperature data from the English Midlands. Because the sample ACF in Figure 2.5.1 seems to truncate after lag 2, a MA(2) model may be needed. However, since the sample PACF in Figure 3.3.1 cuts off after lag 2, an AR(2) model may be suitable for modelling the series. In reality, either of these models may be employed, since they are quite similar. This can be demonstrated by expanding the AR(2) model as an infinite MA model and then comparing the results to the MA(2) model in [3.3.12].

The estimated AR(2) model for the temperature data is

$$(1 - 0.119B - 0.200B^2)(z_t - 9.216) = a_t \qquad [3.4.22]$$

where 9.216 is the MLE of the mean level. For the model in the above equation, [3.4.21] becomes

$$(1 - 0.119B - 0.200B^2)\psi_k = 0$$

When $k = 1$

$$(1 - 0.119B - 0.200B^2)\psi_1 = 0 \text{ or } \psi_1 - 0.119\psi_0 - 0.200\psi_{-1} = 0$$

Since $\psi_0 = 1$ and $\psi_{-1} = 0$ the expression reduces to

$$\psi_1 = 0.119$$

For $k = 2$

$$(1 - 0.119B - 0.200B^2)\psi_2 = 0 \text{ or } \psi_2 - 0.119\psi_1 - 0.200\psi_0 = 0$$

Therefore, $\psi_2 = 0.119(0.119) + 0.200 = 0.214$.

When $k = 3$

$$(1 - 0.119B - 0.200B^2)\psi_3 = 0 \text{ or } \psi_3 - 0.119\psi_2 - 0.200\psi_1 = 0$$

Therefore, $\psi_3 = 0.119(0.214) + 0.200(0.119) = 0.049$.

In general, the expression for ψ_k is

$$\psi_k = 0.119\psi_{k-1} + 0.200\psi_{k-2},\ k > 0$$

Because of the form of this equation, ψ_k decays towards zero rather quickly for increasing lag after lag 2.

Using the results for the ψ coefficients, the random shock form of the AR(2) model in [3.4.22] is

$$\begin{aligned} z_t - 9.216 &= (1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots)a_t \\ &= (1 + 0.119B + 0.214B^2 + 0.049B^3 + \cdots)a_t \end{aligned} \qquad [3.4.23]$$

The SE of estimation for both MA parameters in [3.3.12] is 0.062 and it can be seen that each MA parameter in [3.4.23] is within one SE of the corresponding MA parameter in [3.3.12]. Consequently, for practical purposes the AR(2) model in [3.4.22] is actually the same as the MA(2) model in [3.3.12].

Example Using the Tree Ring Model: The sample ACF and PACF are shown in Figures 3.4.1 and 3.4.2, respectively, for the Douglas Fir tree ring series at Navajo National Monument in Arizona. Because both of these plots attenuate, it may be appropriate to fit an ARMA(1,1) model to this series. The fitted model for this data is given in [3.4.15].

For the ARMA(1,1) model, [3.4.21] becomes

$$(1 - 0.682B)\psi_k = -\theta_k$$

where $\theta_k = 0$ for $k > 1$. When $k = 1$

$$(1 - 0.682B)\psi_1 = -0.424 \text{ or } \psi_1 - 0.682\psi_0 = -0.424$$

Therefore, $\psi_1 = 0.682 - 0.424 = 0.258$.

For $k = 2$

$$(1 - 0.682B)\psi_2 = 0 \text{ or } \psi_2 - 0.682\psi_1 = 0$$

Hence, $\psi_2 = 0.682\psi_1 = 0.682(0.258) = 0.176$.

When $k = 3$

$$(1 - 0.682B)\psi_3 = 0 \text{ or } \psi_3 - 0.682\psi_2 = 0$$

Therefore, $\psi_3 = 0.682\psi_2 = 0.682(0.176) = 0.120$.

The general expression for ψ_k is

$$\psi_k = 0.682\psi_{k-1} = (0.682)^{k-1}\psi_1 \ ,\ k > 0$$

Due to the form of this equation, ψ_k will decrease in absolute value for increasing lag. When the ARMA(1,1) model is expressed using the ψ coefficients, the random shock form of the models is

$$z_t - 99.400 = (1 + 0.258B + 0.176B^2 + 0.120B^3 + ...)a_t \qquad [3.4.24]$$

Inverted Form

To express the ARMA(p,q) process in *inverted form* as a pure AR process, [3.4.4] can be written as

$$\begin{aligned} a_t &= \theta(B)^{-1}\phi(B)(z_t - \mu) \\ &= (z_t - \mu) - \pi_1(z_{t-1} - \mu) - \pi_2(z_{t-2} - \mu) - \cdots \\ &= (z_t - \mu) - \pi_1 B(z_t - \mu) - \pi_2 B^2(z_t - \mu) - \cdots \\ &= (1 - \pi_1 B - \pi_2 B^2 - ...)(z_t - \mu) \\ &= \pi(B)(z_t - \mu) \end{aligned} \qquad [3.4.25]$$

where $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \cdots$, is the *inverted or infinite AR operator* and π_i is the ith parameter, coefficient or weight of $\pi(B)$. Besides ARMA models, it is often convenient to write nonstationary processes and also various types of seasonal processes in the above format. Furthermore, by comparing [3.4.18] and [3.4.25] it is evident that

$$\psi(B)^{-1} = \pi(B) \qquad [3.4.26]$$

In Section 3.3.2, it is pointed out that in order for a MA(q) process to be expressed as a meaningful infinite AR process the roots of $\theta(B) = 0$ must lie outside the unit circle. Invertibility is also achieved for an ARMA(p,q) process when the roots of $\theta(B) = 0$ lie outside the unit circle. This is because the weights $\pi_1, \pi_2, ...$, in the inverted operator $\pi(B) = \theta(B)^{-1}\phi(B)$ constitute a convergent series for $|B| \leq 1$. The invertibility condition is independent of the stationarity condition and can also be used with nonstationary processes.

To determine a relationship for computing the π parameters, multiply [3.4.25] by $\theta(B)$ to get

$$\theta(B)a_t = \theta(B)\pi(B)(z_t - \mu)$$

Using [3.4.4], $\phi(B)(z_t - \mu)$ can be substituted for $\theta(B)a_t$ in the above equation to obtain

$$\phi(B) = \theta(B)\pi(B)$$

The π weights can be readily ascertained by expressing the above equation as

$$\theta(B)\pi_k = \phi_k \qquad [3.4.27]$$

where B operates on k, $\pi_0 = -1$ when using [3.4.27] to calculate π_k for $k > 0$, $\pi_k = 0$ for $k < 0$, and $\phi_k = 0$ if $k > p$. When $k > p$, π_k satisfies the same difference equation as the inverse autocorrelation function (IACF) that is discussed in Section 5.3.6. Consequently, when π_k is plotted against lag k it will possess the same behaviour as the IACF. For increasing lags, the graph may consist of a smooth exponential decay, an exponential decline where the values alternate in sign or a damped sinusoidal decay. Some examples are now presented to demonstrate how to employ [3.4.27] for calculating the π parameters by hand.

Example Using the Temperature Model: The MA(2) model in [3.3.12] for the average annual temperature data from the English Midlands can be equivalently expressed as an infinite AR model. To determine the π weights for the MA(2) model, [3.4.27] becomes

$$(1 + 0.111B + 0.197B^2)\pi_k = 0$$

When $k = 1$

$$(1 + 0.111B + 0.197B^2)\pi_1 = 0 \text{ or } \pi_1 + 0.111\pi_0 + 0.197\pi_{-1} = 0$$

Since $\pi_0 = -1$ and $\pi_{-1} = 0$, the expression reduces to $\pi_1 = 0.111$.

For $k = 2$

$$(1 + 0.111B + 0.197B^2)\pi_2 = 0 \text{ or } \pi_2 + 0.111\pi_1 + 0.197\pi_0 = 0$$

Therefore, $\pi_2 = -0.111(0.111) + 0.197 = 0.185$.

When $k = 3$

$$(1 + 0.111B + 0.197B^2)\pi_3 = 0 \text{ or } \pi_3 + 0.111\pi_2 + 0.197\pi_1 = 0$$

Hence, $\pi_3 = -0.111(0.185) - 0.197(0.111) = -0.042$.

In general, the expression for π_k is

$$\pi_k = -0.111\pi_{k-1} - 0.197\pi_{k-2}, \quad k > 0$$

Because of the structure of the above equation, π_k attenuates quickly in absolute value after lag 2.

By employing the results for the π weights, the inverted form of the MA(2) model in [3.3.12] is

$$(1 - 0.111B - 0.185B^2 + 0.042B^3 + ...)(z_t - 9.216) = a_t \qquad [3.4.28]$$

It can be seen that inverted form of the MA(2) model in [3.4.28] is almost the same as the AR(2) model in [3.4.22] for the temperature data. The SE of estimation for both AR parameters in [3.4.22] is 0.062 and each AR parameter in [3.4.28] is within one SE of the corresponding AR parameter in [3.4.22]. This confirms that the MA(2) model in [3.3.12] is statistically the same as the AR(2) model in [3.4.22].

Example Using the Tree Ring Model: The ARMA(1,1) model which is fitted to the Douglas Fir tree ring series at Navajo National Monument in Arizona is given in [3.4.15]. This model can be equivalently expressed as an inverted model by using [3.4.27] to obtain the π weights. For the case of the ARMA(1,1) model, [3.4.27] becomes

$$(1 - 0.424B)\pi_k = \phi_k$$

where $\phi_k = 0$ for $k > 1$ and $\pi_0 = -1$.

When $k = 1$

$$(1 - 0.424B)\pi_1 = 0.682 \text{ or } \pi_1 - 0.424\pi_0 = 0.682$$

Therefore, $\pi_1 = -0.424 + 0.682 = 0.258$.

For $k = 2$

$$(1 - 0.424B)\pi_2 = 0 \text{ or } \pi_2 - 0.424\pi_1 = 0$$

Hence, $\pi_2 = 0.424\pi_1 = 0.424(0.258) = 0.109$.

When $k = 3$

$$(1 - 0.424B)\pi_3 = 0 \text{ or } \pi_3 - 0.424\pi_2 = 0$$

Therefore, $\pi_3 = 0.424\pi_2 = 0.424(0.109) = 0.046$.

The general expression for π_k is

$$\pi_k = 0.424\pi_{k-1} = (0.424)^{k-1}\pi_1, \quad k > 0$$

It can be seen from this equation that π_k will decrease in absolute value for increasing lag. When the ARMA(1,1) model is written using the π parameters, the inverted form of the model is

$$(1 - 0.258B - 0.109B^2 - 0.046B^3 - \cdots)(z_t - 99.40) = a_t \qquad [3.4.29]$$

Linear Filter Interpretation

The random shock form of the process in [3.4.18] can be considered in terms of a linear filter. As shown in Figure 3.4.5, the white noise input passes through the *linear filter* $\psi(B)$ which transforms the white noise into the output $(z_t - \mu)$. Because of this, the random shock operation $\psi(B)$ is referred to as the transfer function or the filter. When the sequence formed by the ψ weights is either finite or infinite and convergent, the filter is stable because the process z_t is stationary. For stationary processes μ is the mean level about which the process varies. However, when the filter is unstable and the process is not stationary, by definition the process does not fluctuate about any mean level and μ can be considered as a reference point.

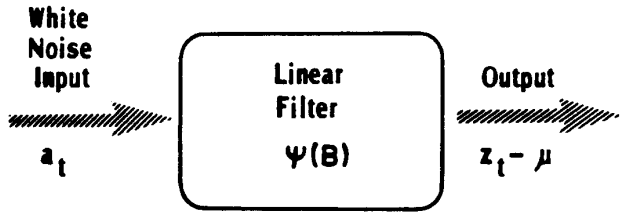

Figure 3.4.5. Linear filter interpretation of the random shock model.

Linear Difference Equations

Equation [3.4.4] for an ARMA(p,q) process constitutes what is called a *linear difference equation* because the process is linear with respect to the AR and MA parameters. Similarly, the random shock and inverted forms of the ARMA model in [3.4.18] and [3.4.25], respectively, are also linear difference equations. Another example of a linear difference equation is [3.2.9] for the theoretical ACF of an AR(p) process. Difference equations arise in time series modelling because it is necessary to model time series which have values at discrete and evenly spaced time intervals. On the other hand, differential equations are employed for modelling systems which evolve over continuous time.

The solution of a linear difference equation is analogous to that for a *linear differential equation*. The final solution for a linear differential equation consists of an equation which does not possess any differentials. Similarly, the solution to a linear difference equation is an equation which does not contain entries which can be written in terms of the B operator. As is the case for a linear differential equation, the general solution for a linear difference equation consists of the summation of a complementary function plus a particular integral. For a brief description of how to solve linear difference equations, the reader may wish to refer to Box and Jenkins (1976, Appendix A4.1, pp. 114-119). Pandit and Wu (1983) make extensive comparisons between linear stochastic differential and difference equations. In fact, these authors explain how to obtain both difference and differential equations from a time series to represent the underlying dynamic system and how to employ these equations for prediction, control and other applications.

3.4.4 Constrained Models

As mentioned earlier, a primary objective in stochastic modelling is to adequately model the data using a model which possesses as few parameters as possible. The principle of model parsimony can be achieved in practice by using a discrimination procedure such as the Akaike information criterion (Akaike, 1974) (see Sections 1.3.3 and 6.3) in conjunction with diagnostic checks (see Chapter 7). This can result in selecting an ARMA(p,q) model where some of the AR and MA parameters which are less than order p and q respectively, are omitted from the model. For instance, as shown in Section 3.2.2, the most appropriate model to fit to the average annual flows of the St. Lawrence River at Ogdensburg, New York is an AR(3) model without the ϕ_2 parameter. The difference equation for this model is given in [3.2.19]. Models which have some of the parameters constrained to zero are referred to as *constrained models*. The option for omitting model parameters can be done with both nonseasonal and seasonal models.

An interesting constrained model is the one which is fitted to the yearly Wolfer sunspot number series in Section 6.4.3. This sunspot series is available from 1700 to 1960 in the work of Waldmeier (1961). If it is deemed appropriate to fit an ARMA model to the sunspot series, it turns out that the best ARMA model is an AR(9) model with ϕ_3 to ϕ_8 left out of the model. In addition, as is shown in Section 6.4.3, it is first necessary to take a square root transformation of the data before fitting the constrained AR(9) model.

A constrained AR model is also referred to as a subset AR model. Research on this topic is provided by authors such as Haggan and Oyetunji (1984) as well as Yu and Lin (1991). Moreover, subset autoregression is also discussed in Section 6.3.6.

3.4.5 Box-Cox Transformation

As noted in Section 3.2.1, the a_t series is always assumed to be independently distributed and possess a constant variance about a zero mean level. In addition, it is often appropriate to invoke the normality assumption for the residuals in order to obtain MLE's for the model parameters (Chapter 6) and subsequently to carry out diagnostic checks (Chapter 7). When fitting ARMA models to a given data set, the model residuals can be estimated along with the model parameters at the estimation stage and model adequacy can be ascertained by checking that the residual assumptions are satisfied. The independence assumption is the most important of all and its violation can cause drastic consequences (Box and Tiao, 1973, p. 522). In fact, when the independence assumption is violated it is necessary to design another model to fit to the data (see Chapter 7). However, if the constant variance and/or normality assumptions are not true, they are often reasonably well fulfilled when the observations are transformed by a Box-Cox transformation (Hipel et al., 1977; McLeod et al., 1977).

A *Box-Cox transformation* (Box and Cox, 1964) is defined by

$$z_t^{(\lambda)} = \begin{cases} \lambda^{-1}[(z_t + c)^{\lambda} - 1] & , \ \lambda \neq 0 \\ ln(z_t + c) & , \ \lambda = 0 \end{cases} \qquad [3.4.30]$$

where c is a constant. The power transformation in [3.4.30] is valid for $z_t + c > 0$. Consequently, if all of the values in the time series are greater than zero usually the constant is set equal to zero. When negative and/or zero values of z_t are present it is usually most convenient to select the constant to be slightly larger than the absolute value of the largest non-positive entry in the time series.

Because the parameter values of an ARMA model fitted to a given time series are unchanged by a linear transformation, the transformation in [3.4.30] is equivalent to

$$z_t^{(\lambda)} = \begin{cases} z_t^{\lambda} & , \ \lambda \neq 0 \\ lnz_t & , \ \lambda = 0 \end{cases} \qquad [3.4.31]$$

where the entries of the z_t series are all greater than zero. The form of the Box-Cox transformation in [3.4.30] is preferable theoretically to that in [3.4.31] because the transformation in [3.4.30] is continuous at $\lambda = 0$. By invoking L'Hopital's rule, it can be shown that the transformation for $\lambda \neq 0$ in [3.4.30] reduces to $ln(z_t + c)$ in the limit as λ approaches zero. When $\lambda = 1$, this means that there is no power transformation.

After the entries in a time series have been changed by a transformation such as that given in [3.4.30], or others discussed by Jain and Singh (1986), an appropriate ARMA model can be fitted to the transformed data. The equation for an ARMA(p,q) model for the $z_t^{(\lambda)}$ series is

$$\phi(B)(z_t^{(\lambda)} - \mu) = \theta(B)a_t \qquad [3.4.32]$$

where μ is the mean level of the $z_t^{(\lambda)}$ sequence. Box-Cox transformations are useful when dealing with both nonseasonal and seasonal time series. For notational convenience in later chapters

often $z^{(\lambda)}$ is simply written as z_t where it is assumed that the series to which the model is fitted is transformed using [3.4.30] whenever necessary. Finally, data transformations that can be considered when dealing with extreme values are referred to in Section 5.3.3.

3.5 THEORETICAL SPECTRUM

As noted in Section 2.6, most of the time series modelling and analysis methods presented in this book are defined and used within the time domain. For example, the theoretical ACF and PACF for an ARMA model constitute time domain properties which are needed for model identification in Chapter 5. Based upon a knowledge of the general properties of the theoretical ACF and PACF, one can examine the characteristics of the sample ACF and PACF for deciding upon which parameters to include in an ARMA model to fit to a given data set.

The objective of this section is to define the theoretical spectrum for ARMA models and present some graphs of the spectrum for specific kinds of ARMA models. As explained below, the spectral density is simply the Fourier transform of the theoretical autocovariance function. Consequently, the spectral density is simply the representation of the autocovariance function within the frequency domain.

3.5.1 Definitions

Any stationary time series, z_t, can be viewed as being composed of a limiting sum of sinusoids of the form

$$A_i\cos(2\pi f_i t + \alpha_i)$$

where f_i is the frequency, A_i is the amplitude and α_i is the phase. The frequency varies from -1/2 to 1/2 in cycles per unit time. The amplitude and phase components at frequency f_i are uncorrelated random variables with a mean of zero in each different realization of the time series. The variance of the amplitude is determined by the spectrum which is defined in the next paragraph. Those frequencies for which the spectrum, $S(f)$, is large will contribute sinusoids with greater amplitudes and thus represent more important sources of variation in the time series.

The Cramer spectral representation expresses the aforesaid facts in a more precise fashion. Every covariance stationary time series with a mean of zero has the Cramer spectral representation [see for example, Kleiner et al. (1979, p. 319)]

$$z_t = \int_{-1/2}^{1/2} e^{i2\pi ft}\, dZ(f) \qquad [3.5.1]$$

where $Z(f)$ for $|f| \le 1/2$ is a continuous stochastic process with orthogonal increments (so that $Z(f_2) - Z(f_1)$ and $Z(f_4) - Z(f_3)$ are uncorrelated whenever $f_1 < f_2 \le f_3 < f_4$). The process $Z(f)$ defines the cumulative spectral density function $F(f)$, by

$$F(f) = E\left[|Z(f)|^2\right]$$

$$dF(f) = E\left[|dZ(f)|^2\right]$$

$$F(-1/2) = 0$$

$$F(1/2) = var(z_t) = \gamma_0 \qquad [3.5.2]$$

For most types of time series, the derivative of $F(f)$ exists and the *spectrum* may be defined as

$$S(f) = 2F'(f), \quad 0 \le f \le 1/2 \qquad [3.5.3]$$

The factor of 2 on the right hand side of [3.5.3] allows for the fact that the spectrum is symmetric about zero and hence only the spectrum in the range $0 \le f \le 1/2$ needs to be considered. In addition to spectrum, other commonly used names for $S(f)$ are spectral density, power spectral density, spectral density function, power spectral density function and power spectrum.

It follows from [3.5.1] and the orthogonal increment property of $A(f)$ that

$$\begin{aligned}\gamma_k &= 1/2 \int_{-1/2}^{1/2} e^{i2\pi fk}\, S(f)\, df \\ &= 1/2 \int_{-1/2}^{1/2} (\cos 2\pi fk + i \sin 2\pi fk)\, S(f)\, df \\ &= \int_{0}^{1/2} \cos 2\pi fk\, S(f)\, df \end{aligned} \qquad [3.5.4]$$

For $k = 0$,

$$\gamma_0 = \int_{0}^{1/2} S(f)\, df \qquad [3.5.5]$$

Because of [3.5.5], the spectrum gives the distribution of the variance of the process over frequency and the area under the spectral curve is the variance.

By taking the inverse transformation of [3.5.4], it follows that the spectral density function is given by

$$S(f) = 2 \sum_{k=-\infty}^{\infty} \gamma_k \cos 2\pi fk \qquad [3.5.6]$$

The above equation shows that the spectrum is simply the Fourier transform of the autocovariance function.

The spectrum can conveniently be written in terms of the autocovariance generating function. When an ARMA process is expressed as the random shock form of the process in [3.4.18], the *autocovariance generating function* is given as (Box and Jenkins, 1976, p. 81)

$$\gamma(B) = \sigma_a^2 \psi(B)\psi(B^{-1}) \qquad [3.5.7]$$

where B^{-1} is the *forward shift operator* defined by

$$B^{-1}z_t = z_{t+1} \text{ and } B^{-k}z_t = z_{t+k}$$

Because the spectrum is the Fourier transform of the autocovariance function, it can be written in terms of the autocovariance generating function in [3.5.4] as

$$S(f) = 2\gamma(e^{i2\pi f}) = 2\sigma_a^2\psi(e^{i2\pi f})\psi(e^{-i2\pi f})$$
$$= 2\sigma_a^2|\psi(e^{-i2\pi f})|^2 \qquad [3.5.8]$$

When utilizing the AR and MA operators, [3.5.8] for an ARMA process is given as

$$S(f) = 2\sigma_a^2\left|\frac{\theta(e^{-i2\pi f})}{\phi(e^{-i2\pi f})}\right|^2 \qquad [3.5.9]$$

To calculate the theoretical spectrum for an ARMA process, the sum of angles method (Robinson, 1967, p. 64; Otnes and Enochson, 1972, p. 139) can be used to recursively calculate the sine and cosine terms (see [2.6.3] and [2.6.4]).

The *normalized spectral density function* is given by

$$s(f) = \frac{S(f)}{\gamma_0} \qquad [3.5.10]$$

Because $s(f)$ is not a function of σ_a^2, it is often used instead of $S(f)$. For the applications in this section, the normalized spectral density is employed.

Examples: Consider obtaining the autocovariance function and the normalized spectrum for a MA(1) process by employing [3.5.7] and [3.5.10], respectively. When using the autocovariance generating function to ascertain γ_k, the coefficient of either B^k or B^{-k} are examined in [3.5.7]. For a MA(1) process, $\psi(B) = 1 - \theta_1 B$ and the autocovariance generating function is

$$\gamma(B) = \sigma_a^2(1 - \theta_1 B)(1 - \theta_1 B^{-1})$$
$$= \sigma_a^2(-\theta_1 B^{-1} + (1 + \theta_1^2) - \theta_1 B)$$

From the coefficients of the backward shift operator, the autocovariances are found to be

$$\gamma_0 = (1 + \theta_1^2)\sigma_a^2$$
$$\gamma_1 = -\theta_1\sigma_a^2$$
$$\gamma_k = 0 \quad k \geq 2$$

By utilizing [3.5.9], the spectrum for a MA(1) process is

$$S(f) = 2\sigma_a^2|1 - \theta_1(e^{-i2\pi f})|^2$$
$$= 2\sigma_a^2[(1 - \theta_1\cos 2\pi f)^2 + (\theta_1\sin 2\pi f)^2]$$
$$= 2\sigma_a^2(1 - 2\theta_1\cos 2\pi f + \theta_1^2)$$

From [3.3.10], the variance of a MA(1) process is $(1 + \theta_1^2)\sigma_a^2$ and, consequently, the normalized spectrum is calculated using [3.5.10] as

$$s(f) = \frac{2(1 - 2\theta_1\cos 2\pi f + \theta_1^2)}{(1 + \theta_1^2)} \qquad [3.5.11]$$

3.5.2 Plots of the Log Normalized Spectrum

For a white noise process, the normalized spectrum in [3.5.10] reduces to

$$S(f) = 2 \qquad [3.5.12]$$

Consequently, all frequencies are equally important for explaining the process and a graph of $S(f)$ against frequency would simply be a straight line.

When considering an AR(1) process the variance of the process is given in [3.2.16] as $\sigma_a^2/(1-\phi_1^2)$. The normalized spectrum is calculated from [3.5.10] as

$$\begin{aligned} s(f) &= \frac{2(1-\phi_1^2)}{|1-\phi_1 e^{-i2\pi f}|^2} \\ &= \frac{2(1-\phi_1^2)}{(1-\phi_1\cos 2\pi f)^2 + (\phi_1 \sin 2\pi f)^2} \\ &= \frac{2(1-\phi_1^2)}{1-2\phi_1\cos 2\pi f + \phi_1^2} \end{aligned} \qquad [3.5.13]$$

For $\phi_1 > 0$, the normalized spectrum in [3.5.13] is easily seen to be a steadily decreasing function for increasing frequency. This means that most of the variance of the time series can be represented as low frequency sinusoids. When the natural logarithms of the normalized spectrum are plotted against frequency, this may improve the ability to distinguish important features of the graph. The log normalized spectrums for AR(1) processes with $\phi_1 = 0.3$ and $\phi_1 = 0.8$ are displayed in Figures 3.5.1 and 3.5.2, respectively. The spectrums are calculated at enough points to cause the curves to appear to be smooth. As can be seen, low frequencies are dominant in both of these figures and consequently the spectrums are said to be "red" (this is because red is on the low frequency end of visible light in the electromagnetic spectrum). Furthermore, because the process with $\phi_1 = 0.3$ is closer to white noise than the process with $\phi_1 = 0.8$, the log normalized spectrum in Figure 3.5.1 is "flatter" than the plot in Figure 3.5.2.

When ϕ_1 for an AR(1) process is negative, the spectrum is dominated by high frequencies. Figures 3.5.3 and 3.5.4 are plots of the log normalized spectrum for $\phi_1 = -0.3$ and $\phi_1 = -0.8$, respectively. As shown in these graphs, most of the variance is explained by high frequencies terms in the "blue" end of the frequency scale. In addition, the upswing in the log normalized spectrum in the high frequencies is more pronounced for the process with $\phi_1 = -0.80$ as compared to the case when $\phi_1 = -0.30$.

Other theoretical spectrums can be readily examined by employing [3.5.10]. Of particular interest are the spectrums of the models that have been fitted to various geophysical time series. This is because a plot of the spectral density of a fitted ARMA model can be useful in obtaining insight into important properties of the original time series.

Figure 3.5.5 shows a plot of the log normalized spectrum for the constrained AR(3) model without ϕ_2 that is fitted to the average annual flows of the St. Lawrence River at Ogdensburg, New York. The difference equation for this model is given in [3.2.19]. As can be observed in Figure 3.5.5, the low frequencies are most important for explaining the variance. From a

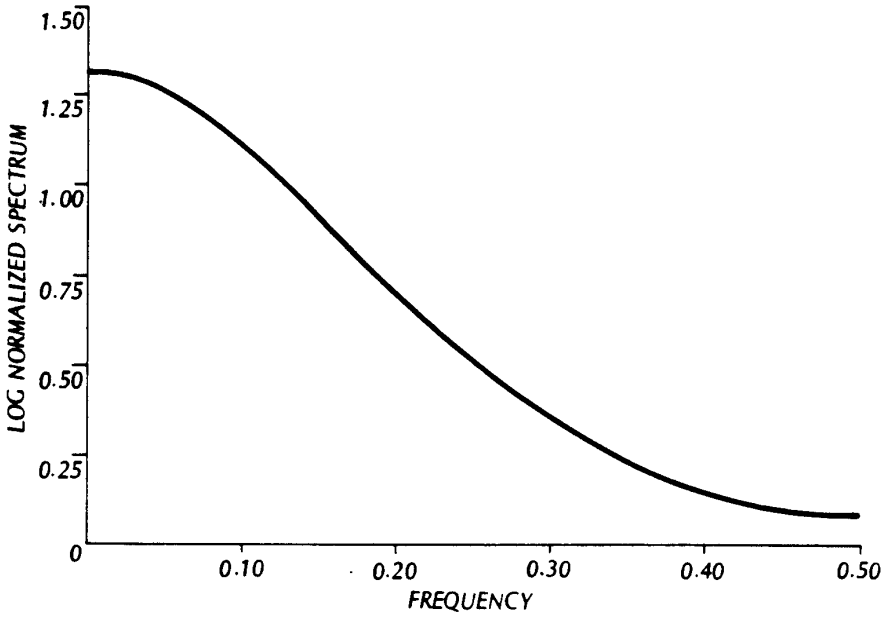

Figure 3.5.1. Log normalized spectrum for an AR(1) process with $\phi_1 = 0.3$.

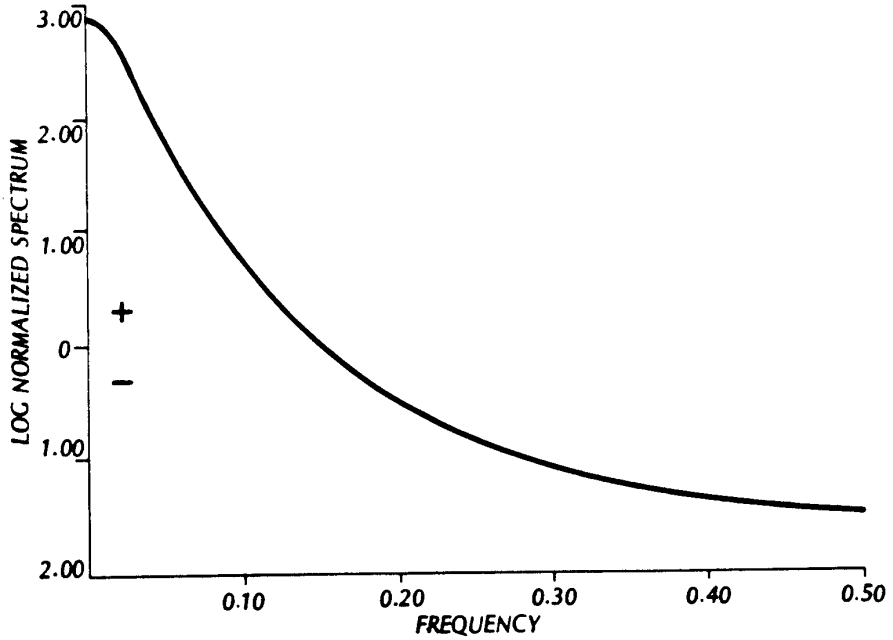

Figure 3.5.2. Log normalized spectrum for an AR(1) process with $\phi_1 = 0.8$.

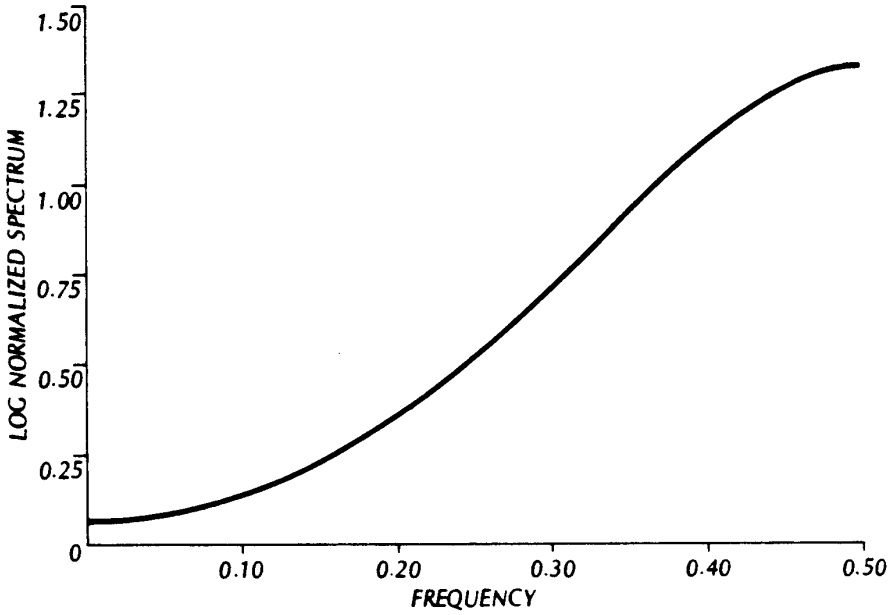

Figure 3.5.3. Log normalized spectrum for an AR(1) process with $\phi_1 = -0.3$.

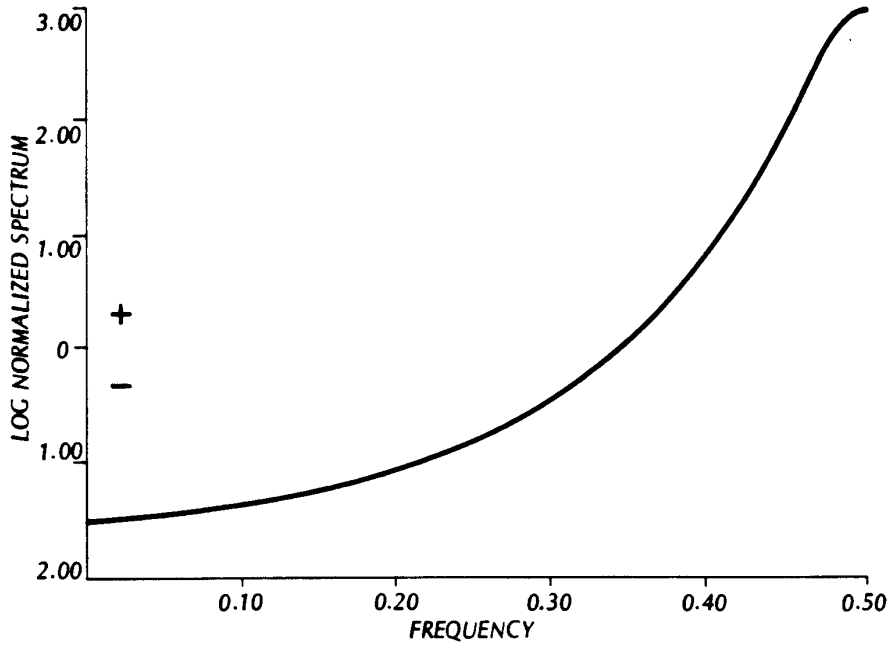

Figure 3.5.4. Log normalized spectrum for an AR(1) process with $\phi_1 = -0.8$.

physical point of view, this makes sense because the Great Lakes, that are located upstream from Ogdensburg, have a dampening effect upon extreme weather conditions that may occur in a given year. The enormous storage capacity of the Great Lakes produces a long term influence upon the flows of the St. Lawrence River and, hence, low frequency terms are dominant in the spectrum.

The MA(2) model that is fitted to the average annual temperature data in the English Midlands is given in [3.3.12]. The plot of the log normalized spectrum for this model is presented in Figure 3.5.6. The low frequency end of the spectrum is most important but the high frequency terms also assist in explaining the variability in the series. Since the fitted model is in fact close to white noise, the high points in the log normalized spectrum are spread over a wide range of frequencies.

Figure 3.5.7 is a graph of the log normalized spectrum for the ARMA(1,1) model fitted to the annual tree ring indices for Douglas Fir at the Navajo National Monument in Arizona. The difference equation for the fitted model is given in [3.4.15]. As shown in Figure 3.5.7 the spectrum is red. This could be due to the fact that the growth of a tree for a given year may be highly dependent upon the weather conditions over a long time span. For example, if the climate is favourable for healthy growth over a rather long period of time, the tree may be hardy enough to withstand severe weather patterns when they do arise without having its growth seriously retarded.

A series of 5405 tree ring widths for Bristlecone Pine at Campito Mountain in Eastern California from 3435 B.C. to 1969 A.D., is listed in units of 0.01 mm. The most appropriate ARMA model to fit to the first 500 years of this series is an ARMA(4,3) model. As can be seen for the log normalized spectrum for this model in Figure 3.5.8, there is a strong low frequency component. The peak at 0.275 cycles/year corresponds to a period of 1/0.275 = 3.6 years. When a plot of the first 500 years of the series is examined, it appears that a weak periodic component may be present in the data.

After transforming the data using a square root transformation, the most appropriate model to fit to the annual sunspot numbers is a constrained AR(9) model with ϕ_3 to ϕ_8 left out of the model. This model is given in [6.4.3]. The log normalized spectrum in Figure 3.5.9 for the sunspot model shows that the low frequencies are the most crucial for explaining the variance in the series. As noted by Granger (1957), the periodicity of the sunspot data follows a uniform distribution with a mean of about 11 years. This is confirmed by the peak in Figure 3.5.9 at a frequency of slightly less than 0.1. The cumulative periodogram for the sunspot data in Figure 2.6.3 also possesses a dramatic jump at a frequency of about 1/11.

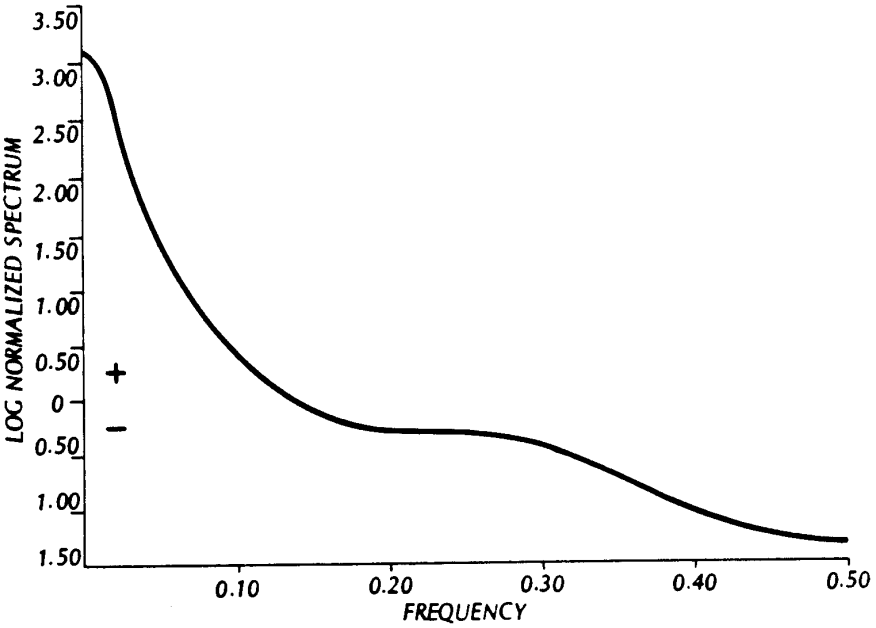

Figure 3.5.5. Log normalized spectrum for the constrained AR(3) model without ϕ_2 that is fitted to the average annual flows of the St. Lawrence River at Ogdensburg, New York.

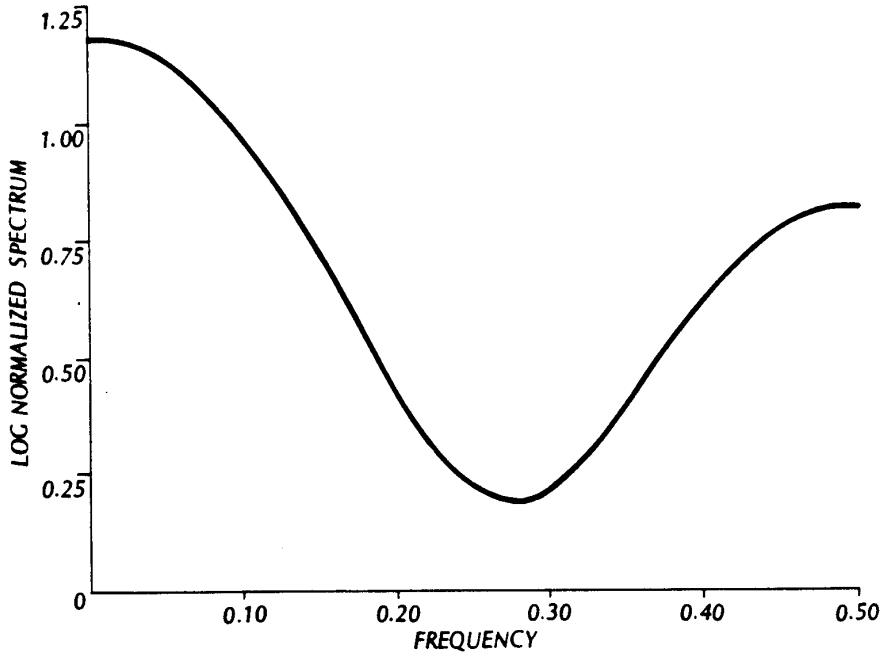

Figure 3.5.6. Log normalized spectrum for the MA(2) model fitted to the average yearly temperature data in the English Midlands.

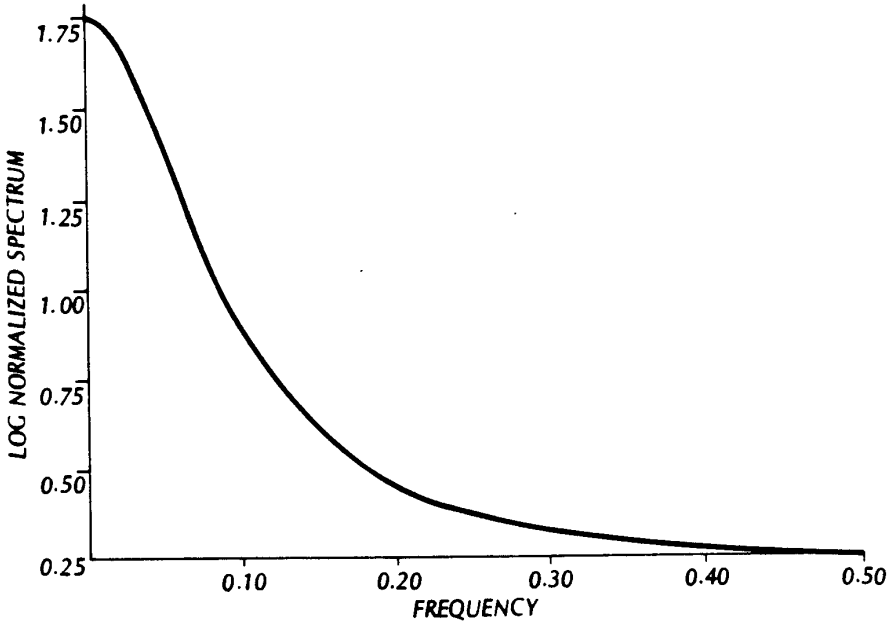

Figure 3.5.7. Log normalized spectrum for the ARMA(1,1) model fitted to the annual tree ring indices for Douglas Fir at Navajo National Monument in Arizona.

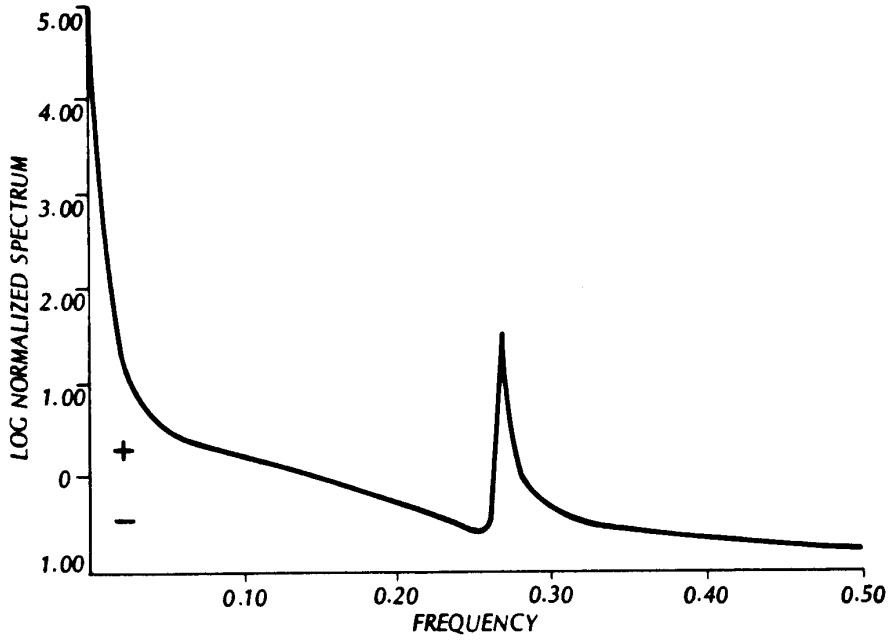

Figure 3.5.8. Log normalized spectrum for the ARMA(4,3) model fitted to the first 500 years of the Bristlecone Pine tree ring series at Campito Mountain, California.

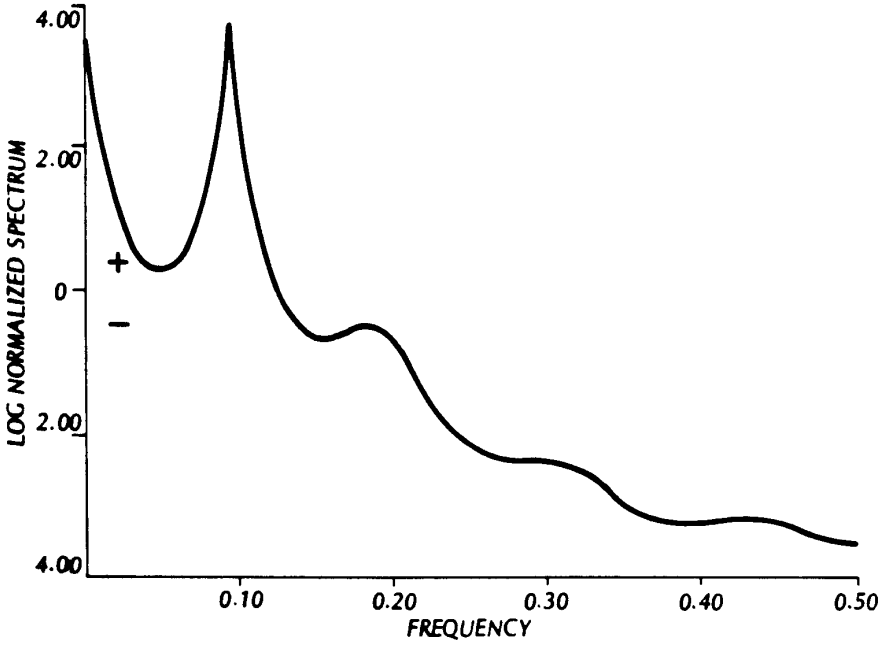

Figure 3.5.9. Log normalized spectrum for the constrained AR(9) model without ϕ_3 to ϕ_8 that is fitted to the sunspot numbers series transformed using square roots.

3.6 PHYSICAL JUSTIFICATION OF ARMA MODELS

3.6.1 Environmental Systems Model of a Watershed

The main physical components of the hydrological cycle are shown in Figure 1.4.1. As explained in Section 1.4.2, the hydrological cycle is the environmental system describing the distribution and circulation of water in all its forms on the surface of the land, underground and in the atmosphere. When modelling any part of the hydrological cycle, one would like to employ models that encapsulate the key physical characteristics of the subsystem being modelled. In other words, one would desire to use models that are physically founded and thereby properly describe the essential elements of the physical system.

For a substantial period of time, hydrologists as well as other environmental scientists have been concerned with developing a physical basis for stochastic modelling. In 1963, for example, Yevjevich examined the physical justification for using the AR(1) model in [3.2.3]. Moss and Bryson (1974) looked at the physical basis of seasonal stochastic models, which are described in Part VI of this book. Klemes (1978) as well as Salas and Smith (1981) provided a review of research on the physical foundations of stochastic models used in hydrology. Moreover, Parlange et al. (1992) explained how an AR(1) model can be formulated on the basis of the hydrologic budget and soil water transport equation, and demonstrate that the model predictions compare well with experimental results.

Fiering (1967) entertained a watershed in which the annual precipitation is decomposed into evaporation, infiltration and surface runoff. By employing the mass balance equation for the groundwater storage, he found the correlation structure of annual streamflow as a function of the correlation structure of precipitation which was assumed to be independent or else AR(1). Salas and Smith (1981) demonstrated that the conceptual watershed model of Thomas and Fiering leads to ARMA streamflows and ARMA groundwater storage. The objective of this section is to point out some of the main findings of Salas and Smith (1981) so that the reader can fully appreciate the *physical justification* for employing ARMA models in hydrology.

Figure 3.6.1 displays the environmental systems model for a *watershed* that Salas and Smith (1981) utilize in their research. This systems model is, of course, a component of the overall hydrological cycle depicted in Figure 1.4.1. In essence, the physical systems model in Figure 3.6.1 shows how precipitation is transformed into runoff or annual riverflow.

Following the notation provided by Salas and Smith (1981) for the environmental model of the watershed shown in Figure 3.6.1, let x_t represent the precipitation in year t. Assume that an amount bx_t of the precipitation evaporates and an amount ax_t infiltrates through the soil into groundwater storage. Therefore, $(1 - a - b)x_t = dx_t$ represents the surface runoff that flows into the rivers and streams. Moreover, let S_{t-1} be the groundwater storage at the start of year t and assume that cS_{t-1} is the groundwater contribution to runoff. In the above algebraic description of the watershed model, it is necessary that $0 \leq a,b,c,d \leq 1$ and $0 \leq a + b \leq 1$.

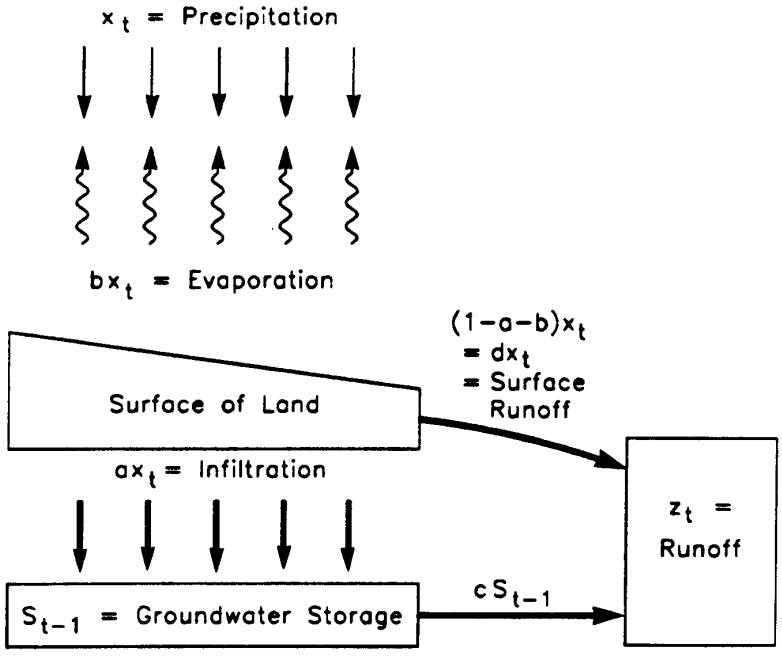

Figure 3.6.1. Environmental systems model of a watershed.

As shown in Figure 3.6.1, the total runoff or riverflow z_t is composed of the direct surface runoff dx_t plus the groundwater contribution cS_{t-1}. Accordingly,

$$z_t = cS_{t-1} + dx_t \qquad [3.6.1]$$

Furthermore, the *mass balance equation* for the groundwater storage is

$$S_t = (1-c)S_{t-1} + ax_t \qquad [3.6.2]$$

The above two equations can be combined to obtain (Salas and Smith, 1981)

$$z_t = (1-c)z_{t-1} + dx_t - [d(1-c) - ac]x_{t-1} \qquad [3.6.3]$$

When writing down the difference equations for the AR, MA, and ARMA models in Sections 3.2 to 3.4, respectively, the mean level μ is subtracted from the variable z_t being modelled. Because of this, the theoretical mean of the a_t innovations in these models is zero. In order to compare the results of this section to the ARMA models, it is convenient to write equations [3.6.1] to [3.6.3] in a similar fashion. More specifically, let μ, μ_x and μ_s be the theoretical means for the variables z_t, x_t and S_t, respectively. By replacing z_t, x_t and S_t by $(z_t - \mu)$, $(x_t - \mu_x)$ and $(S_t - \mu_s)$, respectively, equations [3.6.1] to [3.6.3] can be equivalently rewritten as

$$z_t - \mu = c(S_{t-1} - \mu_s) + d(x_t - \mu_x) \qquad [3.6.4]$$

$$S_t - \mu_s = (1-c)(S_{t-1} - \mu_s) + a(x_t - \mu_x) \qquad [3.6.5]$$

$$z_t - \mu = (1-c)(z_{t-1} - \mu) + d(x_t - \mu_x) - [d(1-c) - ac](x_{t-1} - \mu_x) \qquad [3.6.6]$$

Based upon three different models for the precipitation, Salas and Smith (1981) derive the models for the corresponding groundwater storage and riverflows. Below, the results of their research are summarized for the three cases of independent, AR(1) and ARMA(1,1) precipitation.

3.6.2 Independent Precipitation

If the precipitation is independent, it can be written as

$$(x_t - \mu_x) = a_t \qquad [3.6.7]$$

where μ_x is the mean of the total amount of precipitation x_t falling in year t and a_t is IID$(0,\sigma_a^2)$ as in [3.2.2]. Substituting $(x_t - \mu_x)$ from [3.6.7] into [3.6.5] produces

$$S_t - \mu_s = (1-c)(S_{t-1} - \mu_s) + a(a_t) \qquad [3.6.8]$$

In terms of the groundwater storage variable S_t, the above relationship is simply an AR(1) model. When the AR(1) model in [3.6.8] is compared to the one in [3.2.1] notice that S_t replaces z_t, $(1-c) = \phi_1$ and a constant a instead of unity is in front of the innovation term.

To find the relationship for riverflow, replace $(x_t - \mu_x)$ by a_t in [3.6.6] to obtain

$$z_t - \mu = (1-c)(z_{t-1} - \mu) + da_t - [d(1-c) - ac]a_{t-1} \quad [3.6.9]$$

The above expression for total yearly flow z_t is simply an ARMA(1,1) model defined in [3.4.1]. One can employ [3.4.17] to write the theoretical ACF for the ARMA(1,1) model as

$$\rho_k = (1-c)\rho_{k-1} \text{ for } k > 1 \quad [3.6.10]$$

where ρ_k is the theoretical ACF at lag k.

In summary, independent precipitation produces AR(1) storage as shown in [3.6.8]. Additionally, this kind of precipitation causes the ARMA(1,1) flow given in [3.6.9].

3.6.3 AR(1) Precipitation

As is demonstrated below for the watershed model in Figure 3.6.1, AR(1) precipitation causes AR(2) groundwater storage and ARMA(2,1) runoff. From [3.2.1], an AR(1) model for the precipitation x_t is written as

$$(x_t - \mu_x) = \phi_1(x_{t-1} - \mu_x) + a_t \quad [3.6.11]$$

To determine the type of groundwater storage that this precipitation creates, substitute [3.6.11] into [3.6.5] to get

$$S_t - \mu_s = (1 - c + \phi_1)(S_{t-1} - \mu_s) - (1-c)\phi_1(S_{t-2} - \mu_s) + (a)a_t \quad [3.6.12]$$

In terms of storage, [3.6.12] is an ARMA(1,1) model.

By combining [3.6.11] and [3.6.6], the riverflow generated by AR(1) precipitation is

$$z_t - \mu = (1 - c + \phi_1)(z_{t-1} - \mu) - (1-c)\phi_1(z_{t-2} - \mu) + (d)a_t - [d(1-c) - ac]a_{t-1} \quad [3.6.13]$$

Hence, AR(1) precipitation causes ARMA(2,1) riverflows. From [3.4.13], the theoretical ACF for this ARMA(2,1) model is

$$\rho_k = (1 - c + \phi_1)\rho_{k-1} - (1-c)\phi_1\rho_{k-2} \quad \text{for } k > 1 \quad [3.6.14]$$

3.6.4 ARMA(1,1) Precipitation

The ARMA(1,1) model for the precipitation x_t in Figure 3.6.1 is written as

$$(x_t - \mu_x) = \phi_1(x_{t-1} - \mu_x) + a_t - \theta_1 a_{t-1} \quad [3.6.15]$$

By substituting [3.6.15] into [3.6.5], the resulting groundwater storage is found to be

$$S_t - \mu_s = (1 - c + a\phi_1)(S_{t-1} - \mu_s) + (1-c)a\phi_1(S_{t-2} - \mu_s) + (a)a_t - a\theta_1 a_{t-1} \quad [3.6.16]$$

which is an ARMA(2,1) model. When [3.6.15] is combined with [3.6.6], the model for riverflow is

$$z_t - \mu = (1 - c + \phi_1)(z_{t-1} - \mu) - (1 - c)\phi_1(z_{t-2} - \mu)$$
$$+ da_t - [d(1 - c + \theta_1) - ac]a_{t-1} - \theta_1[ac - d(1 - c)]a_{t-2} \quad [3.6.17]$$

which corresponds to an ARMA(2,2) model. The theoretical ACF for the model in [3.6.17] is obtained from [3.4.13] as

$$\rho_k = (1 - c + \phi_1)\rho_{k-1} - (1 - c)\phi_1\rho_{k-2} \text{ for } k > 2 \quad [3.6.18]$$

Table 3.6.1 summarizes the kinds of groundwater storage and streamflow models that are created by the three different types of precipitation investigated in Sections 3.6.2 to 3.6.4. As noted earlier, these results were originally derived by Salas and Smith (1981) for the environmental systems model of the watershed displayed in Figure 3.6.1. The findings clearly demonstrate that ARMA models possess a valid physical basis for modelling this kind of hydrologic system. Consequently, in Parts III, IV and V of the book, ARMA models are fitted directly to annual riverflow and other types of yearly environmental time series.

Table 3.6.1. Physical basis of ARMA models in hydrology.

Types of Models for Precipitation	Resulting Models	
	Groundwater Storage	Streamflow Runoff
Independent	AR(1)	ARMA(1,1)
AR(1)	AR(2)	ARMA(2,1)
ARMA(1,1)	ARMA(2,1)	ARMA(2,2)

3.7 CONCLUSIONS

The AR and MA classes of models of Sections 3.2 and 3.3, respectively, are members of the general family of ARMA models defined in Section 3.4. These models possess sound theoretical designs and their important theoretical properties are known. For example, the theoretical ACF's for AR, MA, and ARMA models are derived in this chapter and a simple algorithm for calculating the theoretical ACF of any ARMA model is given in Appendix A3.2. Knowledge of the theoretical ACF structure of ARMA models is required for identifying the most appropriate type of ARMA model to fit to a given data set. As explained in Part III, well developed model construction tools are available for fitting ARMA models to stationary nonseasonal time series by following the identification, estimation and diagnostic check stages of model building. Practical applications in Part III clearly demonstrate that ARMA models are ideally suited for describing stationary annual riverflow series as well as other kinds of environmental data sets.

In addition to having a rigorous theoretical design and possessing comprehensive model building tools, ARMA models possess other inherent assets for ensuring their successful application in the environmental sciences. Firstly, the results of Section 3.6 confirm that there is valid physical justification for employing ARMA models for fitting to yearly hydrologic time series. For example, from Table 3.6.1 one can see that if the annual precipitation is ARMA(1,1), then the groundwater storage must be ARMA(2,1) and the yearly streamflow runoff is ARMA(2,2).

Secondly, in Chapter 10 it is clearly demonstrated using annual hydrologic data and simulation experiments that ARMA models provide a logical explanation for the famous Hurst phenomenon. More specifically, ARMA models are shown to preserve statistically what are called the Hurst statistics, which are statistics that reflect the long term storage capacity of reservoirs. Thirdly, forecasting experiments using yearly hydrologic and other kinds of time series in Chapter 8, show that ARMA models forecast at least as well and usually better than their competitors. Finally, the basic ARMA model of Chapter 3 provides the solid foundations for developing the long memory, seasonal, transfer function-noise, intervention and multivariate models of Chapter 10, and Parts VI to IX, respectively. In fact, by introducing what is called the differencing operator to remove nonstationarity, the ARMA model is extended in Chapter 4 so that it can handle nonstationary annual time series.

APPENDIX A3.1

ALGORITHM FOR ESTIMATING

THE PARTIAL AUTOCORRELATION FUNCTION

The *Pagano algorithm* (1972) uses the following steps to estimate the PACF up to lag p for a specified time series.

1. Determine the modified Cholesky decomposition (Wilkinson, 1965, p. 229) of the estimated autocorrelation matrix $\mathbf{R}_p$ given by

$$\mathbf{R}_p = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & 1 & r_1 & \cdots & r_{p-2} \\ r_2 & r_1 & 1 & \cdots & r_{p-3} \\ . & . & . & & . \\ . & . & . & \cdots & . \\ . & . & . & & . \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & 1 \end{bmatrix} \qquad [A3.1.1]$$

where r_k is estimated using [2.5.9] and the theoretical ACF is defined in [2.5.4]. The modified Cholesky decomposition of $\mathbf{R}_p$ is

$$\mathbf{R}_p = \mathbf{L}_p \mathbf{D}_p \mathbf{L}_p^T \qquad [A3.1.2]$$

where $\mathbf{L}_p$ is a unit lower triangular matrix defined by

$$\mathbf{L}_p = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ l_{21} & 1 & 0 & \cdots & 0 \\ l_{31} & l_{32} & 1 & \cdots & 0 \\ . & . & . & & . \\ . & . & . & \cdots & . \\ . & . & . & & . \\ l_{p1} & l_{p2} & l_{p3} & \cdots & 1 \end{bmatrix},$$

$\mathbf{D}_p$ is a diagonal matrix where d_{kk} is the kth typical diagonal entry which is obtained from

$$d_{ii}l_{ki} = r_{k-i} - \sum_{j=1}^{i} b_{kj}l_{jj} = b_{ki}, \quad i = 1,2,\ldots,k-1$$

$$d_{kk} = 1 - \sum_{j=1}^{k-1} b_{kj}l_{kj}$$

and

$$d_{11} = 1$$

where the sequence b_{kj} is defined by the algorithm.

2. Solve the triangular system of equations given by

$$\mathbf{L}_p\alpha_p = \mathbf{r}_p \qquad [A3.1.3]$$

where the unknown vector is

$$\alpha_p^T = (\alpha_1,\alpha_2,\ldots,\alpha_p)$$

and

$$\mathbf{r}_p^T = (r_1,r_2,\ldots,r_p)$$

3. Calculate the estimates $\hat{\phi}_{kk}$ of the PACF using

$$\hat{\phi}_{kk} = \frac{\alpha_k}{d_{kk}}, \quad k = 1,2,\ldots,p \qquad [A3.1.4]$$

4. If the $\hat{\phi}_{kj}(j = 1,2,\ldots,k)$ are required for some $k \leq p$, they can be determined by solving the triangular system of equations

$$\mathbf{L}_k^T = \begin{pmatrix} \hat{\phi}_{k1} \\ \hat{\phi}_{k2} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\phi}_{kk} \end{pmatrix} = \begin{pmatrix} \hat{\phi}_{11} \\ \hat{\phi}_{22} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\phi}_{kk} \end{pmatrix} \qquad [A3.1.5]$$

From [3.2.13], the estimate for the variance of the white noise sequence for an AR model of

order k is given by

$$\hat{\sigma}_a^2(k) = c_0 - \hat{\phi}_{k1}c_1 - \hat{\phi}_{k2}c_2 - \cdots - \hat{\phi}_{kk}c_k \qquad \text{[A3.1.6]}$$

Alternatively, the white noise variance for an AR(k) model may be estimated recursively by employing

$$\sigma_a^2(k) = \sigma_a^2(k-1)(1-\phi_{kk}^2) \qquad \text{[A3.1.7]}$$

where

$$\sigma_a^2(0) = c_0$$

which is the sample variance calculated using [2.5.2] for the given series.

APPENDIX A3.2

THEORETICAL ACF FOR AN ARMA PROCESS

When the parameters of either a nonseasonal or seasonal ARMA process are known, the following algorithm of McLeod (1975) can be employed to determine the theoretical autocovariance, γ_k, and also the theoretical ACF, ρ_k. For the case of a nonseasonal ARMA(p,q) model, the algorithm is as follows:

1. Set $r = max(p,q)$ and $\phi_0 = \theta_0 = -1$, $c_0 = 1$.

2. Then calculate

$$c_k = -\theta_k + \sum_{i=1}^{min(p,k)} \phi_i c_{k-i} \text{ for } k = 1,2,\ldots,q\ .$$

3. Set $b_k = -\sum_{i=k}^{q} \theta_i c_{i-k}$ for $k = 0,1,\ldots,q$, and set $b_k = 0$ if $k > q$.

4. If $p = 0$ then set $\gamma_k = b_k\sigma_a^2$ for $k = 0,1,\ldots,q$; otherwise

$$\begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_r \end{pmatrix} = -\mathbf{A}^{-1} \begin{pmatrix} b_0 \\ b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_r \end{pmatrix} \sigma_a^2$$

where $\mathbf{A}$ is the $(r+1)$ by $(r+1)$ matrix with (i,j) entry:

$$A_{ij} = \begin{cases} \phi_{i-j},\ j = 1,\ i = 1,2,\ldots,r+1 \\ \phi_{i-j} + \phi_{i+j-2},\ j = 2,3,\ldots,r+1,\ i = 1,2,\ldots,r+1 \end{cases}$$

$$\phi_k = \begin{cases} \phi_k, k = 0,1,\ldots,p \\ 0, \ otherwise \end{cases}$$

5. For $k > r = max(p,q)$ calculate γ_k recursively from

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p}$$

6. Divide γ_k by γ_0 to obtain ρ_k for $k = 1,2,\ldots$.

PROBLEMS

3.1 Two stochastic processes, z_{1t} and z_{2t}, have theoretical autocovariance functions at lag k of γ_{1k} and γ_{2k}, respectively, where $\gamma_{1k} = 0$ for $k > 2$ and $\gamma_{2k} = 0$ for $k > 3$. Derive the theoretical autocorrelation function (ACF) for the process $z_{3t} = z_{1t} + bz_{2t}$ in terms of the theoretical autocovariance functions for z_{1t} and z_{2t} where b is a constant. Assume that z_{1t} and z_{2t} are independent of one another.

3.2 Using equations, clearly explain how Cholesky decomposition works.

3.3 For an AR(2) process given by

$$(1 - 1.1B + 0.24B^2)z_t = a_t$$

a) calculate ρ_1,

b) using ρ_1 as starting values for the difference equation in [3.2.9], determine ρ_k, $k = 2,3,\ldots,12$,

c) plot the ACF for this model and comment about its behaviour.

3.4 A constrained AR(3) model without the second AR parameter, ϕ_2, is written as

$$(1 - \phi_1 B - \phi_3 B^3)(z_t - \mu) = a_t$$

From basic principles, derive the Yule-Walker equations for this specific AR model.

3.5 Compare the advantages and disadvantages of using the following three methods for estimating the PACF. Briefly explain how each method works.

a) Cramer's rule,

b) Durbin's method [see Box and Jenkins (1976) and also Durbin (1960)],

c) Pagano's (1972) technique.

3.6 Using equations, explain how the Burg algorithm works for estimating the parameters of an AR(p) model. As an example, show how the Burg algorithm is employed for estimating the parameters of an AR(2) model.

3.7 From first principles, derive the theoretical ACF for a MA(2) process. Using the Yule-Walker equations, determine the theoretical PACF for this process.

3.8 For the ARMA(1,1) process in [3.4.2], derive the two main equations that are required to solve for γ_k, the theoretical autocovariance function of this process. Use these equations to solve for γ_k, $k = 0,1,2,\ldots$.

3.9 An ARMA model is written as

$$(1 - 0.8B + 0.12B^2)z_t = (1 - 0.2B)a_t$$

Prove whether or not this model is stationary.

3.10 Using the hints given with [3.4.20], prove that for stationarity, the roots of $\phi(B) = 0$ must lie outside the unit circle.

3.11 An ARMA(p,q) model is given as

$$(1 - 0.7B)z_t = (1 - 0.4B - 0.21B^2)a_t$$

Prove whether or not this model is invertible.

3.12 The constrained AR(3) model for the annual flows of the St. Lawrence River at Ogdensburg, New York, is given in [3.2.19] as

$$(1 - 0.619B - 0.177B^3)(z_t - 6818.63) = a_t$$

Write this model in inverted form.

3.13 For the ARMA(1,1) model in [3.4.2], determine

a) the parameters ψ, ψ_2 and ψ_3 in the random shock operator, and

b) the parameters π_1, π_2 and π_3 in the inverted operator.

3.14 Express the model given by

$$(1 - 0.6B)(z_t - 15) = (1 - 0.8B)a_t$$

in

a) random shock form, and

b) inverted form.

3.15 Prove that the Box-Cox power transformation in [3.4.30] is continuous at $\lambda = 0$.

3.16 One method for causing non-normal data to become normal is to invoke the Box-Cox transformation in [3.4.30]. Subsequent to this, an ARMA(p,q) model can be fitted to the data that now approximately follow a normal distribution. Other approaches are also available for modelling non-normal data. Describe other transformations suggested by Jain and Singh (1986) for applying to non-normal data sets. Briefly explain how Lewis (1985) and other authors cited in his paper handle the problem of modelling data that do not follow a normal distribution.

3.17 By employing [3.5.7] and [3.5.10], obtain the autocovariance function and normalized spectrum for an

a) AR(1) model, and

b) ARMA(1,1) model.

3.18 An environmental systems model of a watershed is depicted in Figure 3.6.1. Suppose that the precipitation input to this system is ARMA(2,1). Derive the types of models that this precipitation causes for groundwater storage and streamflow runoff. Write down the theoretical ACF's for the precipitation, groundwater storage and runoff models.

3.19 Section 3.6 explains how ARMA models can realistically describe the watershed system displayed in Figure 3.6.1. Investigate the validity of ARMA models for describing another environmental system such as a system of reservoirs or a sewage treatment facility.

REFERENCES

DATA SETS

Stokes, M. A., Drew, L. G. and Stockton, C. W. (1973). Chronology Series 1, Laboratory of Tree Ring Research, University of Arizona, Tucson, Arizona.

Waldmeier, M. (1961). *The Sunspot Activity in the Years 1610-1960.* Schulthas and Company, Zurich, Switzerland.

Yevjevich, V. M. (1963). Fluctuation of wet and dry years, 1, Research data assembly and mathematical models. Hydrology paper No. 1, Colorado State University, Fort Collins, Colorado.

DATA TRANSFORMATIONS

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B,* 26:211-252.

Jain, D. and Singh, V.P. (1986). A comparison of transformation methods for flood frequency analysis. *Water Resources Bulletin* 22(6):903-912.

PHYSICAL BASIS OF TIME SERIES MODELS

Klemes, V. (1978). Physically based stochastic hydrologic analysis. In *Advances in Hydroscience* 11:285-356, Academic Press, New York.

Moss, M. E. and Bryson, M. C. (1974). Autocorrelation structure of monthly streamflows. *Water Resources Research,* 10:737-744.

Parlange, M. B., Katul, G. G., Cuenca, R. H., Kavvas, M. L., Nielsen, D. R. and Mata, M. (1992). Physical basis for a time series model of soil water content. *Water Resources Research,* 28(9):2437-2446.

Salas, J. D. and Smith, R. A. (1981). Physical basis of stochastic models of annual flows. *Water Resources Research,* 17(2):428-430.

STOCHASTIC HYDROLOGY

Fiering, M. B. (1967). *Streamflow Synthesis*. Harvard University Press, Cambridge, Massachusetts.

Hipel, K.W. and McLeod, A.I. (1978). Preservation of the rescaled adjusted range, 2, Simulation studies using Box-Jenkins models. *Water Resources Research* 14(3):509-516.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770-808.

Hurst, H. E. (1956). Methods of using long-term storage in reservoirs. *Proceedings of the Institute of Civil Engineers*, 1:519-543.

McLeod, A. I. and Hipel, K. W. (1978). Preservation of the rescaled adjusted range, 1, A reassessment of the Hurst phenomenon. *Water Resources Research*, 14(3):491-508.

McLeod, A. I., Hipel, K. W. and Lennox, W. C. (1977). Advances in Box-Jenkins modelling, 2, applications. *Water Resources Research*, 13(3):577-586.

TIME SERIES ANALYSIS

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

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. John Wiley, New York.

Barndorff-Nielsen, O. and Schou, G. (1973). On the parameterization of autoregressive models by partial autocorrelations. *Journal of Multivariate Analysis* 3:408-419.

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

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Mass.

Burg, J. P. (1975). *Maximum Entropy Spectral Analysis*. Ph.D. dissertation, Department of Geophysics, Stanford University, Stanford, California.

Durbin, J. (1960). The fitting of time series models. *Revue de L'Institut International de Statistique*, 28(3):233-244.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. John Wiley, New York.

Granger, C. W. J. (1957). A statistical model for sunspot activity. *Astrophysics Journal*, 126:152-158.

Haggan, V. and Oyetunji, O. B. (1984). On the selection of subset autoregressive time series models. *Journal of Time Series Analysis*, 5(2):103-113.

Haykin, S. (1990). *Modern Filters*. MacMillan, New York.

Healy, M. J. R. (1968). Algorithm AS6, triangular decomposition of a symmetric matrix. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 17:195-197.

Hipel, K.W., McLeod, A.I. and Lennox, W.C. (1977). Advances in Box-Jenkins modelling, 1, Model construction. *Water Resources Research* 13(3):567-575.

Kleiner, B., Martin, R. D. and Thomson, D. J. (1979). Robust estimation of power spectrum. *Journal of the Royal Statistical Society*, 41(3):313-351.

Lewis, P.A.W. (1985). Some simple models for continuous variate time series. *Water Resources Bulletin* 21(4):635-644.

McLeod, A. I. (1975). Derivation of the theoretical autocovariance function of autoregressive-moving average time series. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 24(2):255-256.

Otnes, R. K. and Enochson, L. (1972). *Digital Time Series Analysis*. Wiley.

Pagano, M. (1972). An algorithm for fitting autoregressive schemes. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 21:274-281.

Pagano, M. (1973). When is an autoregressive scheme stationary. *Communications in Statistics*, 1(6):533-544.

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

Quenouille, M. H. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, Series B*, 11:68-84.

Robinson, E. A. (1967). *Multichannel Time Series Analysis with Digital Computer Programs*. Holden-Day, San Francisco.

Schur, I. (1917). Uber potenzreihen, die in innern des einheitskreises beschrankt sind. *J. Reine Agnew. Math.*, 147:205-232.

Siddiqui, M. M. (1958). On the inversion of the sample covariance matrix in a stationary autoregressive process. *Annals of Mathematical Statistics*, 29:585-588.

Walker, G. (1931). On periodicity in series of related terms. *Proceedings of the Royal Society, A*, 131:518-532.

Wilkinson, J. H. (1965). *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford.

Yu, G. H. and Lin, Y. C. (1991). A methodology for selecting subset autoregressive time series models. *Journal of Time Series Analysis*, 12(4):363-373.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer sunspot numbers. *Phil. Transactions of the Royal Society, Series A*, 226:267-298.