2 Stationary Processes

A key role in time series analysis is played by processes whose properties, or some of them, do not vary with time. If we wish to make predictions, then clearly we must assume that *something* does not vary with time. In extrapolating deterministic functions it is common practice to assume that either the function itself or one of its derivatives is constant. The assumption of a constant first derivative leads to linear extrapolation as a means of prediction. In time series analysis our goal is to predict a series that typically is not deterministic but contains a random component. If this random component is stationary, in the sense of Definition 1.4.2, then we can develop powerful techniques to forecast its future values. These techniques will be developed and discussed in this and subsequent chapters.

2.1 Basic Properties

In Section 1.4 we introduced the concept of stationarity and defined the autocovariance function (ACVF) of a stationary time series $\{X_t\}$ as

$$\gamma(h) = \text{Cov}(X_{t+h},\ X_t), \quad h = 0, \pm 1, \pm 2, \ldots.$$

The autocorrelation function (ACF) of $\{X_t\}$ was defined similarly as the function $\rho(\cdot)$ whose value at lag h is

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

P.J. Brockwell, R.A. Davis, *Introduction to Time Series and Forecasting*,
Springer Texts in Statistics, DOI 10.1007/978-3-319-29854-2_2

The ACVF and ACF provide a useful measure of the degree of dependence among the values of a time series at different times and for this reason play an important role when we consider the prediction of future values of the series in terms of the past and present values. They can be estimated from observations of $X_1, \ldots, X_n$ by computing the *sample* ACVF and ACF as described in Section 1.4.1.

The role of the autocorrelation function in prediction is illustrated by the following simple example. Suppose that $\{X_t\}$ is a stationary Gaussian time series (see Definition A.3.2) and that we have observed X_n. We would like to find the function of X_n that gives us the best predictor of X_{n+h}, the value of the series after another h time units have elapsed. To define the problem we must first say what we mean by "best." A natural and computationally convenient definition is to specify our required predictor to be the function of X_n with minimum mean squared error. In this illustration, and indeed throughout the remainder of this book, we shall use this as our criterion for "best." Now by Proposition A.3.1 the conditional distribution of X_{n+h} given that $X_n = x_n$ is

$$\mathrm{N}\big(\mu + \rho(h)(x_n - \mu), \sigma^2\big(1 - \rho(h)^2\big)\big),$$

where μ and σ^2 are the mean and variance of $\{X_t\}$. It was shown in Problem 1.1 that the value of the constant c that minimizes $E(X_{n+h} - c)^2$ is $c = E(X_{n+h})$ and that the function m of X_n that minimizes $E(X_{n+h} - m(X_n))^2$ is the conditional mean

$$m(X_n) = E(X_{n+h}|X_n) = \mu + \rho(h)(X_n - \mu). \tag{2.1.1}$$

The corresponding mean squared error is

$$E(X_{n+h} - m(X_n))^2 = \sigma^2\big(1 - \rho(h)^2\big). \tag{2.1.2}$$

This calculation shows that at least for stationary Gaussian time series, prediction of X_{n+h} in terms of X_n is more accurate as $|\rho(h)|$ becomes closer to 1, and in the limit as $\rho(h) \to \pm 1$ the best predictor approaches $\mu \pm (X_n - \mu)$ and the corresponding mean squared error approaches 0.

In the preceding calculation the assumption of joint normality of X_{n+h} and X_n played a crucial role. For time series with nonnormal joint distributions the corresponding calculations are in general much more complicated. However, if instead of looking for the best function of X_n for predicting X_{n+h}, we look for the best **linear predictor**, i.e., the best predictor of the form $\ell(X_n) = aX_n + b$, then our problem becomes that of finding a and b to minimize $E(X_{n+h} - aX_n - b)^2$. An elementary calculation (Problem 2.1), shows that the best predictor of this form is

$$\ell(X_n) = \mu + \rho(h)(X_n - \mu) \tag{2.1.3}$$

with corresponding mean squared error

$$E(X_{n+h} - \ell(X_n))^2 = \sigma^2(1 - \rho(h)^2). \tag{2.1.4}$$

Comparison with (2.1.1) and (2.1.3) shows that for Gaussian processes, $\ell(X_n)$ and $m(X_n)$ are the same. In general, of course, $m(X_n)$ will give smaller mean squared error than $\ell(X_n)$, since it is the best of a larger class of predictors (see Problem 1.8). However, the fact that the best linear predictor depends only on the mean and ACF of the series $\{X_t\}$ means that it can be calculated without more detailed knowledge of the joint distributions. This is extremely important in practice because of the difficulty of estimating all of the joint distributions and because of the difficulty of computing the required conditional expectations even if the distributions were known.

As we shall see later in this chapter, similar conclusions apply when we consider the more general problem of predicting X_{n+h} as a function not only of X_n, but also of $X_{n-1}, X_{n-2}, \ldots$. Before pursuing this question we need to examine in more detail the properties of the autocovariance and autocorrelation functions of a stationary time series.

Basic Properties of $\gamma(\cdot)$:

$$\gamma(0) \geq 0,$$

$$|\gamma(h)| \leq \gamma(0) \text{ for all } h,$$

and $\gamma(\cdot)$ is even, i.e.,

$$\gamma(h) = \gamma(-h) \text{ for all } h.$$

Proof The first property is simply the statement that $\text{Var}(X_t) \geq 0$, the second is an immediate consequence of the fact that correlations are less than or equal to 1 in absolute value (or the Cauchy–Schwarz inequality), and the third is established by observing that

$$\gamma(h) = \text{Cov}(X_{t+h}, X_t) = \text{Cov}(X_t, X_{t+h}) = \gamma(-h). \qquad \blacksquare$$

Autocovariance functions have another fundamental property, namely that of nonnegative definiteness.

Definition 2.1.1 A real-valued function κ defined on the integers is **nonnegative definite** if

$$\sum_{i,j=1}^{n} a_i \kappa(i-j) a_j \geq 0 \tag{2.1.5}$$

for all positive integers n and vectors $\mathbf{a} = (a_1, \ldots, a_n)'$ with real-valued components a_i.

Theorem 2.1.1 *A real-valued function defined on the integers is the autocovariance function of a stationary time series if and only if it is even and nonnegative definite.*

Proof To show that the autocovariance function $\gamma(\cdot)$ of any stationary time series $\{X_t\}$ is nonnegative definite, let $\mathbf{a}$ be any $n \times 1$ vector with real components $a_1, \ldots, a_n$ and let $\mathbf{X}_n = (X_n, \ldots, X_1)'$. Then by equation (A.2.5) and the nonnegativity of variances,

$$\text{Var}(\mathbf{a}'\mathbf{X}_n) = \mathbf{a}'\Gamma_n\mathbf{a} = \sum_{i,j=1}^{n} a_i \gamma(i-j) a_j \geq 0,$$

where Γ_n is the covariance matrix of the random vector $\mathbf{X}_n$. The last inequality, however, is precisely the statement that $\gamma(\cdot)$ is nonnegative definite. The converse result, that there exists a stationary time series with autocovariance function κ if κ is even, real-valued, and nonnegative definite, is more difficult to establish (see Brockwell and Davis (1991), Theorem 1.5.1 for a proof). A slightly stronger statement

can be made, namely, that under the specified conditions there exists a stationary *Gaussian* time series $\{X_t\}$ with mean 0 and autocovariance function $\kappa(\cdot)$. ■

Remark 1. An autocorrelation function $\rho(\cdot)$ has all the properties of an autocovariance function and satisfies the additional condition $\rho(0) = 1$. In particular, we can say that $\rho(\cdot)$ is the autocorrelation function of a stationary process if and only if $\rho(\cdot)$ is an ACVF with $\rho(0) = 1$. □

Remark 2. To verify that a given function is nonnegative definite it is often simpler to find a stationary process that has the given function as its ACVF than to verify the conditions (2.1.5) directly. For example, the function $\kappa(h) = \cos(\omega h)$ is nonnegative definite, since (see Problem 2.2) it is the ACVF of the stationary process

$$X_t = A\cos(\omega t) + B\sin(\omega t),$$

where A and B are uncorrelated random variables, both with mean 0 and variance 1. Another illustration is provided by the following example. □

Example 2.1.1 We shall show now that the function defined on the integers by

$$\kappa(h) = \begin{cases} 1, & \text{if } h = 0, \\ \rho, & \text{if } h = \pm 1, \\ 0, & \text{otherwise,} \end{cases}$$

is the ACVF of a stationary time series if and only if $|\rho| \leq \frac{1}{2}$. Inspection of the ACVF of the MA(1) process of Example 1.4.4 shows that κ is the ACVF of such a process if we can find real θ and nonnegative σ^2 such that

$$\sigma^2(1 + \theta^2) = 1$$

and

$$\sigma^2\theta = \rho.$$

If $|\rho| \leq \frac{1}{2}$, these equations give solutions $\theta = (2\rho)^{-1}\left(1 \pm \sqrt{1 - 4\rho^2}\right)$ and $\sigma^2 = \left(1 + \theta^2\right)^{-1}$. However, if $|\rho| > \frac{1}{2}$, there is no real solution for θ and hence no MA(1) process with ACVF κ. To show that there is no *stationary* process with ACVF κ, we need to show that κ is not nonnegative definite. We shall do this directly from the definition (2.1.5). First, if $\rho > \frac{1}{2}$, $K = [\kappa(i - j)]_{i,j=1}^{n}$, and $\mathbf{a}$ is the n-component vector $\mathbf{a} = (1, -1, 1, -1, \ldots)'$, then

$$\mathbf{a}'K\mathbf{a} = n - 2(n - 1)\rho < 0 \text{ for } n > 2\rho/(2\rho - 1),$$

showing that $\kappa(\cdot)$ is not nonnegative definite and therefore, by Theorem 2.1.1, is not an autocovariance function. If $\rho < -\frac{1}{2}$, the same argument with $\mathbf{a} = (1, 1, 1, 1, \ldots)'$ again shows that $\kappa(\cdot)$ is not nonnegative definite.

□

If $\{X_t\}$ is a (weakly) stationary time series, then the vector $(X_1, \ldots, X_n)'$ and the time-shifted vector $(X_{1+h}, \ldots, X_{n+h})'$ have the same mean vectors and covariance matrices for every integer h and positive integer n. A strictly stationary sequence is one in which the joint distributions of these two vectors (and not just the means and covariances) are the same. The precise definition is given below.

Definition 2.1.2

$\{X_t\}$ is a **strictly stationary time series** if

$$(X_1, \ldots, X_n)' \stackrel{d}{=} (X_{1+h}, \ldots, X_{n+h})'$$

for all integers h and $n \geq 1$. (Here $\stackrel{d}{=}$ is used to indicate that the two random vectors have the same joint distribution function.)

For reference, we record some of the elementary properties of strictly stationary time series.

Properties of a Strictly Stationary Time Series $\{X_t\}$:

- **a.** The random variables X_t are identically distributed.
- **b.** $(X_t, X_{t+h})' \stackrel{d}{=} (X_1, X_{1+h})'$ for all integers t and h.
- **c.** $\{X_t\}$ is weakly stationary if $E(X_t^2) < \infty$ for all t.
- **d.** Weak stationarity does not imply strict stationarity.
- **e.** An iid sequence is strictly stationary.

Proof Properties (a) and (b) follow at once from Definition 2.1.2. If $EX_t^2 < \infty$, then by (a) and (b) EX_t is independent of t and $\text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_1, X_{1+h})$, which is also independent of t, proving (c). For (d) see Problem 1.8. If $\{X_t\}$ is an iid sequence of random variables with common distribution function F, then the joint distribution function of $(X_{1+h}, \ldots, X_{n+h})'$ evaluated at $(x_1, \ldots, x_n)'$ is $F(x_1) \cdots F(x_n)$, which is independent of h. ■

One of the simplest ways to construct a time series $\{X_t\}$ that is strictly stationary (and hence stationary if $EX_t^2 < \infty$) is to "filter" an iid sequence of random variables. Let $\{Z_t\}$ be an iid sequence, which by (e) is strictly stationary, and define

$$X_t = g(Z_t, Z_{t-1}, \ldots, Z_{t-q}) \tag{2.1.6}$$

for some real-valued function $g(\cdot, \ldots, \cdot)$. Then $\{X_t\}$ is strictly stationary, since $(Z_{t+h}, \ldots, Z_{t+h-q})' \stackrel{d}{=} (Z_t, \ldots, Z_{t-q})'$ for all integers h. It follows also from the defining equation (2.1.6) that $\{X_t\}$ is ***q*-dependent**, i.e., that X_s and X_t are independent whenever $|t - s| > q$. (An iid sequence is 0-dependent.) In the same way, adopting a second-order viewpoint, we say that a stationary time series is ***q*-correlated** if $\gamma(h) = 0$ whenever $|h| > q$. A white noise sequence is then 0-correlated, while the MA(1) process of Example 1.4.4 is 1-correlated. The moving-average process of order q defined below is q-correlated, and perhaps surprisingly, the converse is also true (Proposition 2.1.1).

The MA(q) Process:

$\{X_t\}$ is a **moving-average process of order *q*** if

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{2.1.7}$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ and $\theta_1, \ldots, \theta_q$ are constants.

It is a simple matter to check that (2.1.7) defines a stationary time series that is strictly stationary if $\{Z_t\}$ is iid noise. In the latter case, (2.1.7) is a special case of (2.1.6) with g a linear function.

The importance of MA(q) processes derives from the fact that *every* q-correlated process is an MA(q) process. This is the content of the following proposition, whose proof can be found in Brockwell and Davis (1991), Section 3.2. The extension of this result to the case $q = \infty$ is essentially Wold's decomposition (see Section 2.6).

Proposition 2.1.1 *If $\{X_t\}$ is a stationary q-correlated time series with mean 0, then it can be represented as the* MA(q) *process in (2.1.7).*

2.2 Linear Processes

The class of linear time series models, which includes the class of autoregressive moving-average (ARMA) models, provides a general framework for studying stationary processes. In fact, every second-order stationary process is either a linear process or can be transformed to a linear process by subtracting a *deterministic* component. This result is known as Wold's decomposition and is discussed in Section 2.6.

Definition 2.2.1

The time series $\{X_t\}$ is a **linear process** if it has the representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \tag{2.2.1}$$

for all t, where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ and $\{\psi_j\}$ is a sequence of constants with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

In terms of the backward shift operator B, (2.2.1) can be written more compactly as

$$X_t = \psi(B)Z_t, \tag{2.2.2}$$

where $\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$. A linear process is called a **moving average** or **MA(∞)** if $\psi_j = 0$ for all $j < 0$, i.e., if

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

Remark 1. The condition $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ ensures that the infinite sum in (2.2.1) converges (with probability one), since $E|Z_t| \le \sigma$ and

$$E|X_t| \le \sum_{j=-\infty}^{\infty} \left(|\psi_j| E|Z_{t-j}|\right) \le \left(\sum_{j=-\infty}^{\infty} |\psi_j|\right) \sigma < \infty.$$

It also ensures that $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ and hence (see Appendix C, Example C.1.1) that the series in (2.2.1) converges in mean square, i.e., that X_t is the mean square limit of the partial sums $\sum_{j=-n}^{n} \psi_j Z_{t-j}$. The condition $\sum_{j=-n}^{n} |\psi_j| < \infty$ also ensures convergence in both senses of the more general series (2.2.3) considered in Proposition 2.2.1 below. In Section 11.4 we consider a more general class of linear

processes, the fractionally integrated ARMA processes, for which the coefficients are not absolutely summable but only square summable. □

The operator $\psi(B)$ can be thought of as a linear filter, which when applied to the white noise "input" series $\{Z_t\}$ produces the "output" $\{X_t\}$ (see Section 4.3). As established in the following proposition, a linear filter, when applied to any stationary input series, produces a stationary output series.

Proposition 2.2.1 *Let $\{Y_t\}$ be a stationary time series with mean 0 and covariance function γ_Y. If $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, then the time series*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} = \psi(B) Y_t \tag{2.2.3}$$

is stationary with mean 0 and autocovariance function

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h + k - j). \tag{2.2.4}$$

In the special case where $\{X_t\}$ is the linear process (2.2.1),

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2. \tag{2.2.5}$$

Proof The argument used in Remark 1, with σ replaced by $\sqrt{\gamma_Y(0)}$, shows that the series in (2.2.3) is convergent. Since $EY_t = 0$, we have

$$E(X_t) = E\left(\sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}\right) = \sum_{j=-\infty}^{\infty} \psi_j E(Y_{t-j}) = 0$$

and

$$\begin{aligned} E(X_{t+h} X_t) &= E\left[\left(\sum_{j=-\infty}^{\infty} \psi_j Y_{t+h-j}\right)\left(\sum_{k=-\infty}^{\infty} \psi_k Y_{t-k}\right)\right] \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k E(Y_{t+h-j} Y_{t-k}) \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h - j + k), \end{aligned}$$

which shows that $\{X_t\}$ is stationary with covariance function (2.2.4). (The interchange of summation and expectation operations in the above calculations can be justified by the absolute summability of ψ_j.) Finally, if $\{Y_t\}$ is the white noise sequence $\{Z_t\}$ in (2.2.1), then $\gamma_Y(h - j + k) = \sigma^2$ if $k = j - h$ and 0 otherwise, from which (2.2.5) follows. ■

Remark 2. The absolute convergence of (2.2.3) implies (Problem 2.6) that filters of the form $\alpha(B) = \sum_{j=-\infty}^{\infty} \alpha_j B^j$ and $\beta(B) = \sum_{j=-\infty}^{\infty} \beta_j B^j$ with absolutely summable coefficients can be applied successively to a stationary series $\{Y_t\}$ to generate a new stationary series

$$W_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j},$$

where

$$\psi_j = \sum_{k=-\infty}^{\infty} \alpha_k \beta_{j-k} = \sum_{k=-\infty}^{\infty} \beta_k \alpha_{j-k}. \tag{2.2.6}$$

These relations can be expressed in the equivalent form

$$W_t = \psi(B) Y_t,$$

where

$$\psi(B) = \alpha(B)\beta(B) = \beta(B)\alpha(B), \tag{2.2.7}$$

and the products are defined by (2.2.6) or equivalently by multiplying the series $\sum_{j=-\infty}^{\infty} \alpha_j B^j$ and $\sum_{j=-\infty}^{\infty} \beta_j B^j$ term by term and collecting powers of B. It is clear from (2.2.6) and (2.2.7) that the order of application of the filters $\alpha(B)$ and $\beta(B)$ is immaterial. □

Example 2.2.1 An AR(1) Process

In Example 1.4.5, an AR(1) process was defined as a stationary solution $\{X_t\}$ of the equations

$$X_t - \phi X_{t-1} = Z_t, \tag{2.2.8}$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, $|\phi| < 1$, and Z_t is uncorrelated with X_s for each $s < t$. To show that such a solution exists and is the unique stationary solution of (2.2.8), we consider the linear process defined by

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}. \tag{2.2.9}$$

(The coefficients ϕ^j for $j \geq 0$ are absolutely summable, since $|\phi| < 1$.) It is easy to verify directly that the process (2.2.9) is a solution of (2.2.8), and by Proposition 2.2.1 it is also stationary with mean 0 and ACVF

$$\gamma_X(h) = \sum_{j=0}^{\infty} \phi^j \phi^{j+h} \sigma^2 = \frac{\sigma^2 \phi^h}{1 - \phi^2},$$

for $h \geq 0$.

To show that (2.2.9) is the *only* stationary solution of (2.2.8) let $\{Y_t\}$ be *any* stationary solution. Then, iterating (2.2.8), we obtain

$$\begin{aligned} Y_t &= \phi Y_{t-1} + Z_t \\ &= Z_t + \phi Z_{t-1} + \phi^2 Y_{t-2} \\ &= \cdots \\ &= Z_t + \phi Z_{t-1} + \cdots + \phi^k Z_{t-k} + \phi^{k+1} Y_{t-k-1}. \end{aligned}$$

If $\{Y_t\}$ is stationary, then EY_t^2 is finite and independent of t, so that

$$\begin{aligned} E(Y_t - \sum_{j=0}^{k} \phi^j Z_{t-j})^2 &= \phi^{2k+2} E(Y_{t-k-1})^2 \\ &\to 0 \text{ as } k \to \infty. \end{aligned}$$

This implies that Y_t is equal to the mean square limit $\sum_{j=0}^{\infty} \phi^j Z_{t-j}$ and hence that the process defined by (2.2.9) is the unique stationary solution of equation (2.2.8).

It the case $|\phi| > 1$, the series in (2.2.9) does not converge. However, we can rewrite (2.2.8) in the form

$$X_t = -\phi^{-1} Z_{t+1} + \phi^{-1} X_{t+1}. \tag{2.2.10}$$

Iterating (2.2.10) gives

$$\begin{aligned} X_t &= -\phi^{-1} Z_{t+1} - \phi^{-2} Z_{t+2} + \phi^{-2} X_{t+2} \\ &= \cdots \\ &= -\phi^{-1} Z_{t+1} - \cdots - \phi^{-k-1} Z_{t+k+1} + \phi^{-k-1} X_{t+k+1}, \end{aligned}$$

which shows, by the same arguments used above, that

$$X_t = -\sum_{j=1}^{\infty} \phi^{-j} Z_{t+j} \tag{2.2.11}$$

is the unique stationary solution of (2.2.8). This solution should not be confused with the nonstationary solution $\{X_t\}$ of (2.2.8) obtained when X_0 is any specified random variable that is uncorrelated with $\{Z_t\}$.

The solution (2.2.11) is frequently regarded as unnatural, since X_t as defined by (2.2.11) is correlated with *future* values of Z_s, contrasting with the solution (2.2.9), which has the property that X_t is uncorrelated with Z_s for all $s > t$. It is customary therefore in modeling stationary time series to restrict attention to AR(1) processes with $|\phi| < 1$. Then X_t has the representation (2.2.8) in terms of $\{Z_s, s \le t\}$, and we say that $\{X_t\}$ is a **causal** or **future-independent** function of $\{Z_t\}$, or more concisely that $\{X_t\}$ is a causal autoregressive process. It should be noted that every AR(1) process with $|\phi| > 1$ can be reexpressed as an AR(1) process with $|\phi| < 1$ and a new white noise sequence (Problem 3.8). From a second-order point of view, therefore, nothing is lost by eliminating AR(1) processes with $|\phi| > 1$ from consideration.

If $\phi = \pm 1$, there is no stationary solution of (2.2.8) (see Problem 2.8).

□

Remark 3. It is worth remarking that when $|\phi| < 1$ the unique stationary solution (2.2.9) can be found immediately with the aid of (2.2.7). To do this let $\phi(B) = 1 - \phi B$ and $\pi(B) = \sum_{j=0}^{\infty} \phi^j B^j$. Then

$$\psi(B) := \phi(B)\pi(B) = 1.$$

Applying the operator $\pi(B)$ to both sides of (2.2.8), we obtain

$$X_t = \pi(B) Z_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

as claimed. □

2.3 Introduction to ARMA Processes

In this section we introduce, through an example, some of the key properties of an important class of linear processes known as ARMA (autoregressive moving average) processes. These are defined by linear difference equations with constant coefficients.

As our example we shall consider the ARMA(1,1) process. Higher-order ARMA processes will be discussed in Chapter 3.

Definition 2.3.1

The time series $\{X_t\}$ is an **ARMA(1, 1) process** if it is stationary and satisfies (for every t)

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \tag{2.3.1}$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ and $\phi + \theta \neq 0$.

Using the backward shift operator B, (2.3.1) can be written more concisely as

$$\phi(B)X_t = \theta(B)Z_t, \tag{2.3.2}$$

where $\phi(B)$ and $\theta(B)$ are the linear filters

$$\phi(B) = 1 - \phi B \text{ and } \theta(B) = 1 + \theta B,$$

respectively.

We first investigate the range of values of ϕ and θ for which a stationary solution of (2.3.1) exists. If $|\phi| < 1$, let $\chi(z)$ denote the power series expansion of $1/\phi(z)$, i.e., $\sum_{j=0}^{\infty} \phi^j z^j$, which has absolutely summable coefficients. Then from (2.2.7) we conclude that $\chi(B)\phi(B) = 1$. Applying $\chi(B)$ to each side of (2.3.2) therefore gives

$$X_t = \chi(B)\theta(B)Z_t = \psi(B)Z_t,$$

where

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j = \left(1 + \phi B + \phi^2 B^2 + \cdots\right)(1 + \theta B).$$

By multiplying out the right-hand side or using (2.2.6), we find that

$$\psi_0 = 1 \text{ and } \psi_j = (\phi + \theta)\phi^{j-1} \text{ for } j \geq 1.$$

As in Example 2.2.1, we conclude that the MA(∞) process

$$X_t = Z_t + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j} \tag{2.3.3}$$

is the unique stationary solution of (2.3.1).

Now suppose that $|\phi| > 1$. We first represent $1/\phi(z)$ as a series of powers of z with absolutely summable coefficients by expanding in powers of z^{-1}, giving (Problem 2.7)

$$\frac{1}{\phi(z)} = -\sum_{j=1}^{\infty} \phi^{-j} z^{-j}.$$

Then we can apply the same argument as in the case where $|\phi| < 1$ to obtain the unique stationary solution of (2.3.1). We let $\chi(B) = -\sum_{j=1}^{\infty} \phi^{-j} B^{-j}$ and apply $\chi(B)$ to each side of (2.3.2) to obtain

$$X_t = \chi(B)\theta(B)Z_t = -\theta\phi^{-1}Z_t - (\theta + \phi) \sum_{j=1}^{\infty} \phi^{-j-1} Z_{t+j}. \tag{2.3.4}$$

If $\phi = \pm 1$, there is no stationary solution of (2.3.1). Consequently, there is no such thing as an ARMA(1,1) process with $\phi = \pm 1$ according to our definition.

We can now summarize our findings about the existence and nature of the stationary solutions of the ARMA(1,1) recursions (2.3.2) as follows:

- A stationary solution of the ARMA(1,1) equations exists if and only if $\phi \neq \pm 1$.
- If $|\phi| < 1$, then the unique stationary solution is given by (2.3.3). In this case we say that $\{X_t\}$ is **causal** or a causal function of $\{Z_t\}$, since X_t can be expressed in terms of the current and past values Z_s, $s \leq t$.
- If $|\phi| > 1$, then the unique stationary solution is given by (2.3.4). The solution is **noncausal**, since X_t is then a function of Z_s, $s \geq t$.

Just as causality means that X_t is expressible in terms of Z_s, $s \leq t$, the dual concept of invertibility means that Z_t is expressible in terms of X_s, $s \leq t$. We show now that the ARMA(1,1) process defined by (2.3.1) is invertible if $|\theta| < 1$. To demonstrate this, let $\xi(z)$ denote the power series expansion of $1/\theta(z)$, i.e., $\sum_{j=0}^{\infty}(-\theta)^j z^j$, which has absolutely summable coefficients. From (2.2.6) it therefore follows that $\xi(B)\theta(B) = 1$, and applying $\xi(B)$ to each side of (2.3.2) gives

$$Z_t = \xi(B)\phi(B)X_t = \pi(B)X_t,$$

where

$$\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j = \left(1 - \theta B + (-\theta)^2 B^2 + \cdots\right)(1 - \phi B).$$

By multiplying out the right-hand side or using (2.2.6), we find that

$$Z_t = X_t - (\phi + \theta)\sum_{j=1}^{\infty}(-\theta)^{j-1}X_{t-j}. \tag{2.3.5}$$

Thus the ARMA(1,1) process is **invertible**, since Z_t can be expressed in terms of the present and past values of the process X_s, $s \leq t$. An argument like the one used to show noncausality when $|\phi| > 1$ shows that the ARMA(1,1) process is **noninvertible** when $|\theta| > 1$, since then

$$Z_t = -\phi\theta^{-1}X_t + (\theta + \phi)\sum_{j=1}^{\infty}(-\theta)^{-j-1}X_{t+j}. \tag{2.3.6}$$

We summarize these results as follows:

- If $|\theta| < 1$, then the ARMA(1,1) process is **invertible**, and Z_t is expressed in terms of X_s, $s \leq t$, by (2.3.5).
- If $|\theta| > 1$, then the ARMA(1,1) process is **noninvertible**, and Z_t is expressed in terms of X_s, $s \geq t$, by (2.3.6).

Remark 1. In the cases $\theta = \pm 1$, the ARMA(1,1) process is invertible in the more general sense that Z_t is a mean square limit of finite linear combinations of X_s, $s \leq t$, although it cannot be expressed explicitly as an infinite linear combination of X_s, $s \leq t$ (see Section 4.4 of Brockwell and Davis (1991)). In this book the term *invertible* will always be used in the more restricted sense that $Z_t = \sum_{j=0}^{\infty}\pi_j X_{t-j}$, where $\sum_{j=0}^{\infty}|\pi_j| < \infty$. □

Remark 2. If the ARMA(1,1) process $\{X_t\}$ is noncausal or noninvertible with $|\theta| > 1$, then it is possible to find a new white noise sequence $\{W_t\}$ such that $\{X_t\}$ is a causal and noninvertible ARMA(1,1) process relative to $\{W_t\}$ (Problem 4.10). Therefore, from a second-order point of view, nothing is lost by restricting attention to causal and invertible ARMA(1,1) models. This last sentence is also valid for higher-order ARMA models. □

2.4 Properties of the Sample Mean and Autocorrelation Function

A stationary process $\{X_t\}$ is characterized, at least from a second-order point of view, by its mean μ and its autocovariance function $\gamma(\cdot)$. The estimation of μ, $\gamma(\cdot)$, and the autocorrelation function $\rho(\cdot) = \gamma(\cdot)/\gamma(0)$ from observations $X_1, \ldots, X_n$ therefore plays a crucial role in problems of inference and in particular in the problem of constructing an appropriate model for the data. In this section we examine some of the properties of the sample estimates $\bar{x}$ and $\hat{\rho}(\cdot)$ of μ and $\rho(\cdot)$, respectively.

2.4.1 Estimation of μ

The moment estimator of the mean μ of a stationary process is the sample mean

$$\bar{X}_n = n^{-1}(X_1 + X_2 + \cdots + X_n). \tag{2.4.1}$$

It is an unbiased estimator of μ, since

$$E(\bar{X}_n) = n^{-1}(EX_1 + \cdots + EX_n) = \mu.$$

The mean squared error of $\bar{X}_n$ is

$$\begin{aligned} E(\bar{X}_n - \mu)^2 &= \operatorname{Var}(\bar{X}_n) \\ &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{Cov}(X_i, X_j) \\ &= n^{-2}\sum_{i-j=-n}^{n}(n - |i-j|)\gamma(i-j) \\ &= n^{-1}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma(h). \end{aligned} \tag{2.4.2}$$

Now if $\gamma(h) \to 0$ as $h \to \infty$, the right-hand side of (2.4.2) converges to zero, so that $\bar{X}_n$ converges in mean square to μ. If $\sum_{h=-\infty}^{\infty}|\gamma(h)| < \infty$, then (2.4.2) gives $\lim_{n\to\infty} n\operatorname{Var}(\bar{X}_n) = \sum_{|h|<\infty}\gamma(h)$. We record these results in the following proposition.

Proposition 2.4.1 *If $\{X_t\}$ is a stationary time series with mean μ and autocovariance function $\gamma(\cdot)$, then as $n \to \infty$,*

$$\operatorname{Var}(\bar{X}_n) = E(\bar{X}_n - \mu)^2 \to 0 \quad \text{if } \gamma(n) \to 0,$$

$$nE(\bar{X}_n - \mu)^2 \to \sum_{|h|<\infty}\gamma(h) \quad \text{if } \sum_{h=-\infty}^{\infty}|\gamma(h)| < \infty.$$

To make inferences about μ using the sample mean $\bar{X}_n$, it is necessary to know the distribution or an approximation to the distribution of $\bar{X}_n$. If the time series is Gaussian (see Definition A.3.2), then by Remark 2 of Section A.3 and (2.4.2),

$$n^{1/2}(\bar{X}_n - \mu) \sim \mathrm{N}\left(0, \sum_{|h|<n}\left(1 - \frac{|h|}{n}\right)\gamma(h)\right).$$

It is easy to construct exact confidence bounds for μ using this result if $\gamma(\cdot)$ is known, and approximate confidence bounds if it is necessary to estimate $\gamma(\cdot)$ from the observations.

For many time series, in particular for linear and ARMA models, $\bar{X}_n$ is approximately normal with mean μ and variance $n^{-1}\sum_{|h|<\infty}\gamma(h)$ for large n (see Brockwell and Davis (1991), p. 219). An approximate 95 % confidence interval for μ is then

$$\left(\bar{X}_n - 1.96v^{1/2}/\sqrt{n},\ \bar{X}_n + 1.96v^{1/2}/\sqrt{n}\right), \tag{2.4.3}$$

where $v = \sum_{|h|<\infty}\gamma(h)$. Of course, v is not generally known, so it must be estimated from the data. The estimator computed in the program ITSM is $\hat{v} = \sum_{|h|<\sqrt{n}}\left(1 - |h|/\sqrt{n}\right)\hat{\gamma}(h)$. For ARMA processes this is a good approximation to v for large n.

Example 2.4.1 An AR(1) Model

Let $\{X_t\}$ be an AR(1) process with mean μ, defined by the equations

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t,$$

where $|\phi| < 1$ and $\{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right)$. From Example 2.2.1 we have $\gamma(h) = \phi^{|h|}\sigma^2/(1-\phi^2)$ and hence $v = \left(1 + 2\sum_{h=1}^{\infty}\phi^h\right)\sigma^2/\left(1-\phi^2\right) = \sigma^2/(1-\phi)^2$. Approximate 95 % confidence bounds for μ are therefore given by $\bar{x}_n \pm 1.96\sigma n^{-1/2}/(1-\phi)$. Since ϕ and σ are unknown in practice, they must be replaced in these bounds by estimated values.

□

2.4.2 Estimation of $\gamma(\cdot)$ and $\rho(\cdot)$

Recall from Section 1.4.1 that the sample autocovariance and autocorrelation functions are defined by

$$\hat{\gamma}(h) = n^{-1}\sum_{t=1}^{n-|h|}\left(X_{t+|h|} - \bar{X}_n\right)\left(X_t - \bar{X}_n\right) \tag{2.4.4}$$

and

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}. \tag{2.4.5}$$

Both the estimators $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ are biased even if the factor n^{-1} in (2.4.4) is replaced by $(n-h)^{-1}$. Nevertheless, under general assumptions they are nearly unbiased for large sample sizes. The sample ACVF has the desirable property that for each $k \geq 1$ the k-dimensional sample covariance matrix

$$\hat{\Gamma}_k = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(k-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(k-2) \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\gamma}(k-1) & \hat{\gamma}(k-2) & \cdots & \hat{\gamma}(0) \end{bmatrix} \tag{2.4.6}$$

is nonnegative definite. To see this, first note that if $\hat{\Gamma}_m$ is nonnegative definite, then $\hat{\Gamma}_k$ is nonnegative definite for all $k < m$. So assume $k \geq n$ and write

$$\hat{\Gamma}_k = n^{-1}TT',$$

where T is the $k \times 2k$ matrix

$$T = \begin{bmatrix} 0 & \cdots & 0 & 0 & Y_1 & Y_2 & \cdots & Y_k \\ 0 & \cdots & 0 & Y_1 & Y_2 & \cdots & Y_k & 0 \\ \vdots & & & & & & & \vdots \\ 0 & Y_1 & Y_2 & \cdots & Y_k & 0 \cdots & 0 & \end{bmatrix},$$

$Y_i = X_i - \bar{X}_n$, $i = 1, \ldots, n$, and $Y_i = 0$ for $i = n+1, \ldots, k$. Then for any real $k \times 1$ vector $\mathbf{a}$ we have

$$\mathbf{a}'\hat{\Gamma}_k\mathbf{a} = n^{-1}(\mathbf{a}'T)(T'\mathbf{a}) \geq 0, \tag{2.4.7}$$

and consequently the sample autocovariance matrix $\hat{\Gamma}_k$ and sample autocorrelation matrix

$$\hat{R}_k = \hat{\Gamma}_k/\gamma(0) \tag{2.4.8}$$

are nonnegative definite. Sometimes the factor n^{-1} is replaced by $(n-h)^{-1}$ in the definition of $\hat{\gamma}(h)$, but the resulting covariance and correlation matrices $\hat{\Gamma}_n$ and $\hat{R}_n$ may not then be nonnegative definite. We shall therefore use the definitions (2.4.4) and (2.4.5) of $\hat{\gamma}(h)$ and $\hat{\rho}(h)$.

Remark 1. The matrices $\hat{\Gamma}_k$ and $\hat{R}_k$ are in fact nonsingular if there is at least one nonzero Y_i, or equivalently if $\hat{\gamma}(0) > 0$. To establish this result, suppose that $\hat{\gamma}(0) > 0$ and $\hat{\Gamma}_k$ is singular. Then there is equality in (2.4.7) for some nonzero vector $\mathbf{a}$, implying that $\mathbf{a}'T = 0$ and hence that the rank of T is less than k. Let Y_i be the first nonzero value of $Y_1, Y_2, \ldots, Y_k$, and consider the $k \times k$ submatrix of T consisting of columns $(i+1)$ through $(i+k)$. Since this matrix is lower right triangular with each diagonal element equal to Y_i, its determinant has absolute value $|Y_i|^k \neq 0$. Consequently, the submatrix is nonsingular, and T must have rank k, a contradiction. □

Without further information beyond the observed data $X_1, \ldots, X_n$, it is impossible to give reasonable estimates of $\gamma(h)$ and $\rho(h)$ for $h \geq n$. Even for h slightly smaller than n, the estimates $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ are unreliable, since there are so few pairs (X_{t+h}, X_t) available (only one if $h = n-1$). A useful guide is provided by Jenkins (1976), p. 33 who suggest that n should be at least about 50 and $h \leq n/4$.

The sample ACF plays an important role in the selection of suitable models for the data. We have already seen in Example 1.4.6 and Section 1.6 how the sample ACF can be used to test for iid noise. For systematic inference concerning $\rho(h)$, we need the sampling distribution of the estimator $\hat{\rho}(h)$. Although the distribution of $\hat{\rho}(h)$ is intractable for samples from even the simplest time series models, it can usually be well approximated by a normal distribution for large sample sizes. For linear models and in particular for ARMA models (see Theorem 7.2.2 of Brockwell and Davis (1991) for exact conditions) $\hat{\boldsymbol{\rho}}_k = (\hat{\rho}(1), \ldots, \hat{\rho}(k))'$ is approximately distributed for large n as $N(\boldsymbol{\rho}_k, n^{-1}W)$, i.e.,

$$\hat{\boldsymbol{\rho}} \approx N\big(\boldsymbol{\rho}, n^{-1}W\big), \tag{2.4.9}$$

where $\boldsymbol{\rho} = (\rho(1), \ldots, \rho(k))'$, and W is the covariance matrix whose (i, j) element is given by **Bartlett's formula**

$$w_{ij} = \sum_{k=-\infty}^{\infty} \big\{\rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i)\big\}.$$

Simple algebra shows that

$$w_{ij} = \sum_{k=1}^{\infty}\{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\} \times \{\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\}, \tag{2.4.10}$$

which is a more convenient form of w_{ij} for computational purposes.

Example 2.4.2 iid Noise

If $\{X_t\} \sim \text{IID}\big(0, \sigma^2\big)$, then $\rho(h) = 0$ for $|h| > 0$, so from (2.4.10) we obtain

$$w_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

For large n, therefore, $\hat{\rho}(1), \ldots, \hat{\rho}(h)$ are approximately independent and identically distributed normal random variables with mean 0 and variance n^{-1}. This result is the basis for the test that data are generated from iid noise using the sample ACF described in Section 1.6. (See also Example 1.4.6.)

□

Example 2.4.3 An MA(1) Process

If $\{X_t\}$ is the MA(1) process of Example 1.4.4, i.e., if

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

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, then from (2.4.10)

$$w_{ii} = \begin{cases} 1 - 3\rho^2(1) + 4\rho^4(1), & \text{if } i = 1, \\ 1 + 2\rho^2(1), & \text{if } i > 1, \end{cases}$$

is the approximate variance of $n^{-1/2}(\hat{\rho}(i) - \rho(i))$ for large n. In Figure 2-1 we have plotted the sample autocorrelation function $\hat{\rho}(k)$, $k = 0, \ldots, 40$, for 200 observations from the MA(1) model

$$X_t = Z_t - .8Z_{t-1}, \tag{2.4.11}$$

where $\{Z_t\}$ is a sequence of iid N(0, 1) random variables. Here $\rho(1) = -0.8/1.64 = -0.4878$ and $\rho(h) = 0$ for $h > 1$. The lag-one sample ACF is found to be $\hat{\rho}(1) = -0.4333 = -6.128n^{-1/2}$, which would cause us (in the absence of our prior knowledge of $\{X_t\}$) to reject the hypothesis that the data are a sample from an iid noise sequence. The fact that $|\hat{\rho}(h)| \leq 1.96n^{-1/2}$ for $h=2, \ldots, 40$ strongly suggests that the data are from a model in which observations are uncorrelated past lag 1. Figure 2-1 shows the bounds $\pm 1.96n^{-1/2}(1 + 2\rho^2(1))^{1/2}$, indicating the compatibility of the data with the model (2.4.11). Since, however, $\rho(1)$ is not normally known in advance, the autocorrelations $\hat{\rho}(2), \ldots, \hat{\rho}(40)$ would in practice have been compared with the more

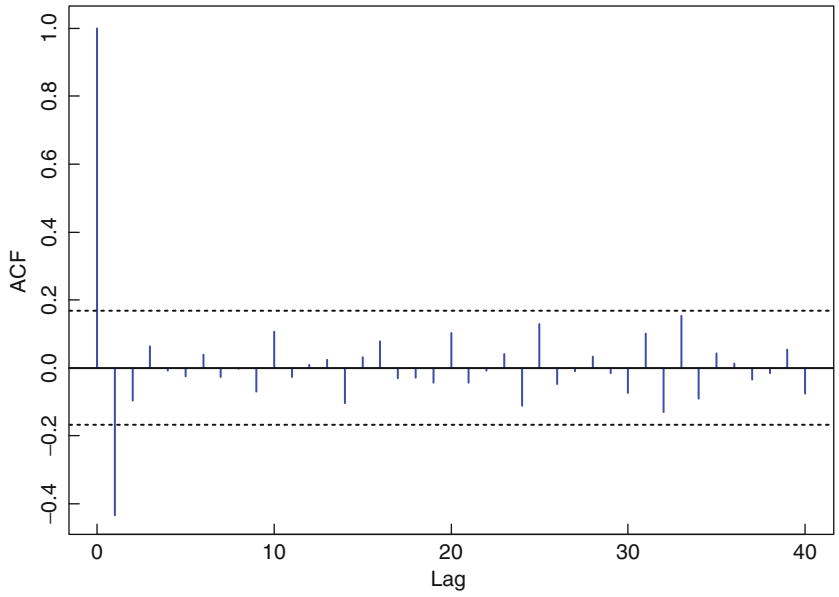

Figure 2-1
The sample autocorrelation function of $n = 200$ observations of the MA(1) process in Example 2.4.3, showing the bounds $\pm 1.96 n^{-1/2}(1+2\hat{\rho}^2(1))^{1/2}$

stringent bounds $\pm 1.96 n^{-1/2}$ or with the bounds $\pm 1.96 n^{-1/2}(1+2\hat{\rho}^2(1))^{1/2}$ in order to check the hypothesis that the data are generated by a moving-average process of order 1. Finally, it is worth noting that the lag-one correlation -0.4878 is well inside the 95 % confidence bounds for $\rho(1)$ given by $\hat{\rho}(1) \pm 1.96 n^{-1/2}(1 - 3\hat{\rho}^2(1) + 4\hat{\rho}^4(1))^{1/2} = -0.4333 \pm 0.1053$. This further supports the compatibility of the data with the model $X_t = Z_t - 0.8 Z_{t-1}$.

□

Example 2.4.4 An AR(1) Process

For the AR(1) process of Example 2.2.1,

$$X_t = \phi X_{t-1} + Z_t,$$

where $\{Z_t\}$ is iid noise and $|\phi| < 1$, we have, from (2.4.10) with $\rho(h) = \phi^{|h|}$,

$$\begin{aligned} w_{ii} &= \sum_{k=1}^{i} \phi^{2i}\left(\phi^{-k} - \phi^{k}\right)^2 + \sum_{k=i+1}^{\infty} \phi^{2k}\left(\phi^{-i} - \phi^{i}\right)^2 \\ &= \left(1 - \phi^{2i}\right)\left(1 + \phi^2\right)\left(1 - \phi^2\right)^{-1} - 2i\phi^{2i}, \end{aligned} \tag{2.4.12}$$

$i = 1, 2, \ldots$. In Figure 2-2 we have plotted the sample ACF of the Lake Huron residuals $y_1, \ldots, y_{98}$ from Figure 1-10 together with 95 % confidence bounds for $\rho(i)$, $i = 1, \ldots, 40$, assuming that data are generated from the AR(1) model

$$Y_t = 0.791 Y_{t-1} + Z_t \tag{2.4.13}$$

[see equation (1.4.3)]. The confidence bounds are computed from $\hat{\rho}(i) \pm 1.96 n^{-1/2} w_{ii}^{1/2}$, where w_{ii} is given in (2.4.12) with $\phi = 0.791$. The model ACF, $\rho(i) = (0.791)^i$, is also plotted in Figure 2-2. Notice that the model ACF just touches the confidence bounds at lags 2–4. This suggests some incompatibility of the data with the model (2.4.13). A much better fit to the residuals is provided by the second-order autoregression defined by (1.4.4).

□

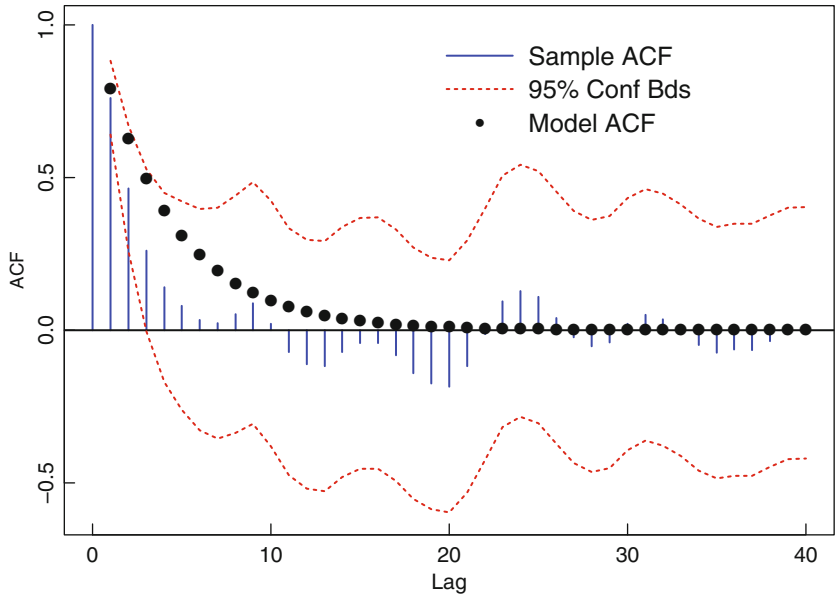

Figure 2-2
The sample autocorrelation function of the Lake Huron residuals of Figure 1-10 showing the bounds $\hat{\rho}(i) \pm 1.96 n^{-1/2} w_{ii}^{1/2}$ and the model ACF $\rho(i) = (0.791)^i$

2.5 Forecasting Stationary Time Series

We now consider the problem of predicting the values X_{n+h}, $h > 0$, of a stationary time series with known mean μ and autocovariance function γ in terms of the values $\{X_n, \ldots, X_1\}$, up to time n. Our goal is to find the *linear combination* of $1, X_n, X_{n-1}, \ldots, X_1$, that forecasts X_{n+h} with minimum mean squared error. The best linear predictor in terms of $1, X_n, \ldots, X_1$ will be denoted by $P_n X_{n+h}$ and clearly has the form

$$P_n X_{n+h} = a_0 + a_1 X_n + \cdots + a_n X_1. \tag{2.5.1}$$

It remains only to determine the coefficients $a_0, a_1, \ldots, a_n$, by finding the values that minimize

$$S(a_0, \ldots, a_n) = E(X_{n+h} - a_0 - a_1 X_n - \cdots - a_n X_1)^2. \tag{2.5.2}$$

(We already know from Problem 1.1 that $P_0 Y = E(Y)$.) Since S is a quadratic function of $a_0, \ldots, a_n$ and is bounded below by zero, it is clear that there is at least one value of $(a_0, \ldots, a_n)$ that minimizes S and that the minimum $(a_0, \ldots, a_n)$ satisfies the equations

$$\frac{\partial S(a_0, \ldots, a_n)}{\partial a_j} = 0, \quad j = 0, \ldots, n. \tag{2.5.3}$$

Evaluation of the derivatives in equation (2.5.3) gives the equivalent equations

$$E\left[X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n+1-i}\right] = 0, \tag{2.5.4}$$

$$E\left[(X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n+1-i}) X_{n+1-j}\right] = 0, \quad j = 1, \ldots, n. \tag{2.5.5}$$

These equations can be written more neatly in vector notation as

$$a_0 = \mu \left(1 - \sum_{i=1}^{n} a_i\right) \tag{2.5.6}$$

and

$$\Gamma_n \mathbf{a}_n = \boldsymbol{\gamma}_n(h), \tag{2.5.7}$$

where

$$\mathbf{a}_n = (a_1, \ldots, a_n)', \qquad \Gamma_n = \left[\gamma(i-j)\right]_{i,\,j=1}^{n},$$

and

$$\boldsymbol{\gamma}_n(h) = (\gamma(h), \gamma(h+1), \ldots, \gamma(h+n-1))'.$$

Hence,

$$P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu), \tag{2.5.8}$$

where $\mathbf{a}_n$ satisfies (2.5.7). From (2.5.8) the expected value of the prediction error $X_{n+h} - P_n X_{n+h}$ is zero, and the mean square prediction error is therefore

$$\begin{aligned} E(X_{n+h} - P_n X_{n+h})^2 &= \gamma(0) - 2\sum_{i=1}^{n} a_i \gamma(h+i-1) + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \gamma(i-j) a_j \\ &= \gamma(0) - \mathbf{a}_n' \boldsymbol{\gamma}_n(h), \end{aligned} \tag{2.5.9}$$

where the last line follows from (2.5.7).

Remark 1. To show that equations (2.5.4) and (2.5.5) determine $P_n X_{n+h}$ uniquely, let $\left\{a_j^{(1)}, j = 0, \ldots, n\right\}$ and $\left\{a_j^{(2)}, j = 0, \ldots, n\right\}$ be two solutions and let Z be the difference between the corresponding predictors, i.e.,

$$Z = a_0^{(1)} - a_0^{(2)} + \sum_{j=1}^{n} \left(a_j^{(1)} - a_j^{(2)}\right) X_{n+1-j}.$$

Then

$$Z^2 = Z\left(a_0^{(1)} - a_0^{(2)} + \sum_{j=1}^{n} \left(a_j^{(1)} - a_j^{(2)}\right) X_{n+1-j}\right).$$

But from (2.5.4) and (2.5.5) we have $EZ = 0$ and $E(Z X_{n+1-j}) = 0$ for $j = 1, \ldots, n$. Consequently, $E(Z^2) = 0$ and hence $Z = 0$. □

Properties of $P_n X_{n+h}$:

1. $P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu)$, where $\mathbf{a}_n = (a_1, \ldots, a_n)'$ satisfies (2.5.7).
2. $E(X_{n+h} - P_n X_{n+h})^2 = \gamma(0) - \mathbf{a}_n' \boldsymbol{\gamma}_n(h)$, where $\boldsymbol{\gamma}_n(h) = (\gamma(h), \ldots, \gamma(h+n-1))'$.
3. $E(X_{n+h} - P_n X_{n+h}) = 0$.
4. $E[(X_{n+h} - P_n X_{n+h}) X_j] = 0$, $j = 1, \ldots, n$.

Remark 2. Notice that properties 3 and 4 are exactly equivalent to (2.5.4) and (2.5.5). They can be written more succinctly in the form

$$E[(\text{Error}) \times (\text{PredictorVariable})] = 0. \tag{2.5.10}$$

The equations (2.5.10), one for each predictor variable, therefore uniquely determine $P_n X_{n+h}$. □

Example 2.5.1 One-Step Prediction of an AR(1) Series

Consider now the stationary time series defined in Example 2.2.1 by the equations

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

where $|\phi| < 1$ and $\{Z_t\} \sim \text{WN}(0, \sigma^2)$. From (2.5.7) and (2.5.8), the best linear predictor of X_{n+1} in terms of $\{1, X_n, \ldots, X_1\}$ is (for $n \geq 1$)

$$P_n X_{n+1} = \mathbf{a}_n' \mathbf{X}_n,$$

where $\mathbf{X}_n = (X_n, \ldots, X_1)'$ and

$$\begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{n-1} \\ \phi & 1 & \phi & \cdots & \phi^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{bmatrix}. \tag{2.5.11}$$

A solution of (2.5.11) is clearly

$$\mathbf{a}_n = (\phi, 0, \ldots, 0)',$$

and hence the best linear predictor of X_{n+1} in terms of $\{X_1, \ldots, X_n\}$ is

$$P_n X_{n+1} = \mathbf{a}_n' \mathbf{X}_n = \phi X_n,$$

with mean squared error

$$E(X_{n+1} - P_n X_{n+1})^2 = \gamma(0) - \mathbf{a}_n' \boldsymbol{\gamma}_n(1) = \frac{\sigma^2}{1 - \phi^2} - \phi\gamma(1) = \sigma^2.$$

A simpler approach to this problem is to guess, by inspection of the equation defining X_{n+1}, that the best predictor is ϕX_n. Then to verify this conjecture, it suffices to check (2.5.10) for each of the predictor variables $1, X_n, \ldots, X_1$. The prediction error of the predictor ϕX_n is clearly $X_{n+1} - \phi X_n = Z_{n+1}$. But $E(Z_{n+1}Y) = 0$ for $Y = 1$ and for $Y = X_j$, $j = 1, \ldots, n$. Hence, by (2.5.10), ϕX_n is the required best linear predictor in terms of $1, X_1, \ldots, X_n$.

□

2.5.1 Prediction of Second-Order Random Variables

Suppose now that Y and W_n, …, W_1 are *any* random variables with finite second moments and that the means $\mu = EY$, $\mu_i = EW_i$ and covariances $\text{Cov}(Y, Y)$, $\text{Cov}(Y, W_i)$, and $\text{Cov}(W_i, W_j)$ are all known. It is convenient to introduce the random vector $\mathbf{W} = (W_n, \ldots, W_1)'$, the corresponding vector of means $\boldsymbol{\mu}_W = (\mu_n, \ldots, \mu_1)'$, the vector of covariances

$$\boldsymbol{\gamma} = \text{Cov}(Y, \mathbf{W}) = (\text{Cov}(Y, W_n), \text{Cov}(Y, W_{n-1}), \ldots, \text{Cov}(Y, W_1))',$$

and the covariance matrix

$$\Gamma = \text{Cov}(\mathbf{W}, \mathbf{W}) = \left[\text{Cov}(W_{n+1-i}, W_{n+1-j})\right]_{i,j=1}^{n}.$$

Then by the same arguments used in the calculation of $P_n X_{n+h}$, the best linear predictor of Y in terms of $\{1, W_n, \ldots, W_1\}$ is found to be

$$P(Y|\mathbf{W}) = \mu_Y + \mathbf{a}'(\mathbf{W} - \boldsymbol{\mu}_W), \tag{2.5.12}$$

where $\mathbf{a} = (a_1, \ldots, a_n)'$ is any solution of

$$\Gamma \mathbf{a} = \boldsymbol{\gamma}. \tag{2.5.13}$$

The mean squared error of the predictor is

$$E\left[(Y - P(Y|\mathbf{W}))^2\right] = \mathrm{Var}(Y) - \mathbf{a}'\boldsymbol{\gamma}. \tag{2.5.14}$$

Example 2.5.2 Estimation of a Missing Value

Consider again the stationary series defined in Example 2.2.1 by the equations

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

where $|\phi| < 1$ and $\{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right)$. Suppose that we observe the series at times 1 and 3 and wish to use these observations to find the linear combination of $1, X_1$, and X_3 that estimates X_2 with minimum mean squared error. The solution to this problem can be obtained directly from (2.5.12) and (2.5.13) by setting $Y = X_2$ and $\mathbf{W} = (X_1, X_3)'$. This gives the equations

$$\begin{bmatrix} 1 & \phi^2 \\ \phi^2 & 1 \end{bmatrix} \mathbf{a} = \begin{bmatrix} \phi \\ \phi \end{bmatrix},$$

with solution

$$\mathbf{a} = \frac{1}{1+\phi^2} \begin{bmatrix} \phi \\ \phi \end{bmatrix}.$$

The best estimator of X_2 is thus

$$P(X_2|\mathbf{W}) = \frac{\phi}{1+\phi^2}(X_1 + X_3),$$

with mean squared error

$$E[(X_2 - P(X_2|\mathbf{W}))^2] = \frac{\sigma^2}{1-\phi^2} - \mathbf{a}' \begin{bmatrix} \frac{\phi\sigma^2}{1-\phi^2} \\ \frac{\phi\sigma^2}{1-\phi^2} \end{bmatrix} = \frac{\sigma^2}{1+\phi^2}.$$

□

2.5.2 The Prediction Operator $P(\cdot|\mathbf{W})$

For any given $\mathbf{W} = (W_n, \ldots, W_1)'$ and Y with finite second moments, we have seen how to compute the best linear predictor $P(Y|\mathbf{W})$ of Y in terms of $1, W_n, \ldots, W_1$ from (2.5.12) and (2.5.13). The function $P(\cdot|\mathbf{W})$, which converts Y into $P(Y|\mathbf{W})$, is called a **prediction operator**. (The operator P_n defined by equations (2.5.7) and (2.5.8) is an example with $\mathbf{W} = (X_n, X_{n-1}, \ldots, X_1)'$.) Prediction operators have a number of useful properties that can sometimes be used to simplify the calculation of best linear predictors. We list some of these below.

Properties of the Prediction Operator $P(\cdot|\,\mathbf{W})$:
Suppose that $EU^2 < \infty$, $EV^2 < \infty$, $\Gamma = \text{Cov}(\mathbf{W}, \mathbf{W})$, and $\beta, \alpha_1, \ldots, \alpha_n$ are constants.

1. $P(U|\mathbf{W}) = EU + \mathbf{a}'(\mathbf{W} - E\mathbf{W})$, where $\Gamma\mathbf{a} = \text{Cov}(U, \mathbf{W})$.
2. $E[(U - P(U|\mathbf{W}))\mathbf{W}] = \mathbf{0}$ and $E[U - P(U|\mathbf{W})] = 0$.
3. $E[(U - P(U|\mathbf{W}))^2] = \text{Var}(U) - \mathbf{a}'\text{Cov}(U, \mathbf{W})$.
4. $P(\alpha_1 U + \alpha_2 V + \beta|\mathbf{W}) = \alpha_1 P(U|\mathbf{W}) + \alpha_2 P(V|\mathbf{W}) + \beta$.
5. $P\left(\sum_{i=1}^n \alpha_i W_i + \beta|\mathbf{W}\right) = \sum_{i=1}^n \alpha_i W_i + \beta$.
6. $P(U|\mathbf{W}) = EU$ if $\text{Cov}(U, \mathbf{W}) = \mathbf{0}$.
7. $P(U|\mathbf{W}) = P(P(U|\mathbf{W}, \mathbf{V})|\mathbf{W})$ if $\mathbf{V}$ is a random vector such that the components of $E(\mathbf{V}\mathbf{V}')$ are all finite.

Example 2.5.3 One-Step Prediction of an AR(p) Series

Suppose now that $\{X_t\}$ is a stationary time series satisfying the equations

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t, \quad t = 0, \pm 1, \ldots,$$

where $\{Z_t\} \sim \text{WN}\left(0, \sigma^2\right)$ and Z_t is uncorrelated with X_s for each $s < t$. Then if $n > p$, we can apply the prediction operator P_n to each side of the defining equations, using properties (4), (5), and (6) to get

$$P_n X_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p}.$$

□

Example 2.5.4 An AR(1) Series with Nonzero Mean

The time series $\{Y_t\}$ is said to be an AR(1) process with mean μ if $\{X_t = Y_t - \mu\}$ is a zero-mean AR(1) process. Defining $\{X_t\}$ as in Example 2.5.1 and letting $Y_t = X_t + \mu$, we see that Y_t satisfies the equation

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + Z_t. \tag{2.5.15}$$

If $P_n Y_{n+h}$ is the best linear predictor of Y_{n+h} in terms of $\{1, Y_n, \ldots, Y_1\}$, then application of P_n to (2.5.15) with $t = n+1, n+2, \ldots$ gives the recursions

$$P_n Y_{n+h} - \mu = \phi(P_n Y_{n+h-1} - \mu), \quad h = 1, 2, \ldots.$$

Noting that $P_n Y_n = Y_n$, we can solve these equations recursively for $P_n Y_{n+h}$, $h = 1, 2, \ldots$, to obtain

$$P_n Y_{n+h} = \mu + \phi^h (Y_n - \mu). \tag{2.5.16}$$

The corresponding mean squared error is [from (2.5.14)]

$$E(Y_{n+h} - P_n Y_{n+h})^2 = \gamma(0)[1 - \mathbf{a}_n' \boldsymbol{\rho}_n(h)]. \tag{2.5.17}$$

From Example 2.2.1 we know that $\gamma(0) = \sigma^2/\left(1-\phi^2\right)$ and $\rho(h) = \phi^h$, $h \geq 0$. Hence, substituting $\mathbf{a}_n = \left(\phi^h, 0, \ldots, 0\right)'$ [from (2.5.16)] into (2.5.17) gives

$$E(Y_{n+h} - P_n Y_{n+h})^2 = \sigma^2\left(1 - \phi^{2h}\right)/\left(1 - \phi^2\right). \tag{2.5.18}$$

□

Remark 3. In general, if $\{Y_t\}$ is a stationary time series with mean μ and if $\{X_t\}$ is the zero-mean series defined by $X_t = Y_t - \mu$, then since the collection of all linear combinations of $1, Y_n, \ldots, Y_1$ is the same as the collection of all linear combinations of $1, X_n, \ldots, X_1$, the linear predictor of any random variable W in terms of $1, Y_n, \ldots, Y_1$ is the same as the linear predictor in terms of $1, X_n, \ldots, X_1$. Denoting this predictor by P_nW and applying P_n to the equation $Y_{n+h} = X_{n+h} + \mu$ gives

$$P_nY_{n+h} = \mu + P_nX_{n+h}. \tag{2.5.19}$$

Thus the best linear predictor of Y_{n+h} can be determined by finding the best linear predictor of X_{n+h} and then adding μ. Note from (2.5.8) that since $E(X_t) = 0$, P_nX_{n+h} is the same as the best linear predictor of X_{n+h} in terms of $X_n, \ldots, X_1$ only. □

2.5.3 The Durbin–Levinson Algorithm

In view of Remark 3 above, we can restrict attention from now on to zero-mean stationary time series, making the necessary adjustments for the mean if we wish to predict a stationary series with nonzero mean. If $\{X_t\}$ is a zero-mean stationary series with autocovariance function $\gamma(\cdot)$, then in principle the equations (2.5.12) and (2.5.13) completely solve the problem of determining the best linear predictor P_nX_{n+h} of X_{n+h} in terms of $\{X_n, \ldots, X_1\}$. However, the direct approach requires the determination of a solution of a system of n linear equations, which for large n may be difficult and time-consuming. In cases where the process is defined by a system of linear equations (as in Examples 2.5.2 and 2.5.3) we have seen how the linearity of P_n can be used to great advantage. For more general stationary processes it would be helpful if the one-step predictor P_nX_{n+1} based on n previous observations could be used to simplify the calculation of $P_{n+1}X_{n+2}$, the one-step predictor based on $n+1$ previous observations. Prediction algorithms that utilize this idea are said to be **recursive**. Two important examples are the Durbin–Levinson algorithm, discussed in this section, and the innovations algorithm, discussed in Section 2.5.4 below.

We know from (2.5.12) and (2.5.13) that if the matrix Γ_n is nonsingular, then

$$P_nX_{n+1} = \boldsymbol{\phi}_n'\mathbf{X}_n = \phi_{n1}X_n + \cdots + \phi_{nn}X_1,$$

where

$$\boldsymbol{\phi}_n = \Gamma_n^{-1}\boldsymbol{\gamma}_n,$$

$\boldsymbol{\gamma}_n = (\gamma(1), \ldots, \gamma(n))'$, and the corresponding mean squared error is

$$v_n := E(X_{n+1} - P_nX_{n+1})^2 = \gamma(0) - \boldsymbol{\phi}_n'\boldsymbol{\gamma}_n.$$

A useful sufficient condition for nonsingularity of *all* the autocovariance matrices $\Gamma_1, \Gamma_2, \ldots$ is $\gamma(0) > 0$ and $\gamma(h) \to 0$ as $h \to \infty$. (For a proof of this result see Brockwell and Davis (1991), Proposition 5.1.1.)

The Durbin–Levinson Algorithm:
The coefficients $\phi_{n1}, \ldots, \phi_{nn}$ can be computed recursively from the equations

$$\phi_{nn} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)\right] v_{n-1}^{-1}, \tag{2.5.20}$$

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix} \tag{2.5.21}$$

and

$$v_n = v_{n-1}\left[1 - \phi_{nn}^2\right], \tag{2.5.22}$$

where $\phi_{11} = \gamma(1)/\gamma(0)$ and $v_0 = \gamma(0)$.

Proofs 1 The definition of ϕ_{11} ensures that the equation

$$R_n\phi_n = \boldsymbol{\rho}_n \tag{2.5.23}$$

(where $\boldsymbol{\rho}_n = (\rho(1), \ldots, \rho(n))'$) is satisfied for $n = 1$. The first step in the proof is to show that ϕ_n, defined recursively by (2.5.20) and (2.5.21), satisfies (2.5.23) for all n. Suppose this is true for $n = k$. Then, partitioning R_{k+1} and defining

$$\boldsymbol{\rho}_k^{(r)} := (\rho(k), \rho(k-1), \ldots, \rho(1))'$$

and

$$\phi_k^{(r)} := (\phi_{kk}, \phi_{k,k-1}, \ldots, \phi_{k1})',$$

we see that the recursions imply

$$\begin{aligned} \mathbf{R}_{k+1}\phi_{k+1} &= \begin{bmatrix} \mathbf{R}_k & \boldsymbol{\rho}_k^{(r)} \\ \boldsymbol{\rho}_k^{(r)\prime} & 1 \end{bmatrix} \begin{bmatrix} \phi_k - \phi_{k+1,k+1}\phi_k^{(r)} \\ \phi_{k+1,k+1} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\rho}_k - \phi_{k+1,k+1}\boldsymbol{\rho}_k^{(r)} + \phi_{k+1,k+1}\boldsymbol{\rho}_k^{(r)} \\ \boldsymbol{\rho}_k^{(r)\prime}\phi_k - \phi_{k+1,k+1}\boldsymbol{\rho}_k^{(r)\prime}\phi_k^{(r)} + \phi_{k+1,k+1} \end{bmatrix} \\ &= \boldsymbol{\rho}_{k+1}, \end{aligned}$$

as required. Here we have used the fact that if $R_k\phi_k = \boldsymbol{\rho}_k$, then $R_k\phi_k^{(r)} = \boldsymbol{\rho}_k^{(r)}$. This is easily checked by writing out the component equations in reverse order. Since (2.5.23) is satisfied for $n = 1$, it follows by induction that the coefficient vectors ϕ_n defined recursively by (2.5.20) and (2.5.21) satisfy (2.5.23) for all n.

It remains only to establish that the mean squared errors

$$v_n := E(X_{n+1} - \phi_n'\mathbf{X}_n)^2$$

satisfy $v_0 = \gamma(0)$ and (2.5.22). The fact that $v_0 = \gamma(0)$ is an immediate consequence of the definition $P_0X_1 := E(X_1) = 0$. Since we have shown that $\phi_n'\mathbf{X}_n$ is the best linear predictor of X_{n+1}, we can write, from (2.5.9) and (2.5.21),

$$v_n = \gamma(0) - \phi_n'\gamma_n = \gamma(0) - \phi_{n-1}'\gamma_{n-1} + \phi_{nn}\phi_{n-1}^{(r)\prime}\gamma_{n-1} - \phi_{nn}\gamma(n).$$

Applying (2.5.9) again gives

$$v_n = v_{n-1} + \phi_{nn}\left(\phi_{n-1}^{(r)\prime}\gamma_{n-1} - \gamma(n)\right),$$

and hence, by (2.5.20),

$$v_n = v_{n-1} - \phi_{nn}^2(\gamma(0) - \phi_{n-1}'\gamma_{n-1}) = v_{n-1}\left(1 - \phi_{nn}^2\right). \quad \blacksquare$$

Remark 4. Under the conditions of the proposition, the function defined by $\alpha(0) = 1$ and $\alpha(n) = \phi_{nn}$, $n = 1, 2, \ldots$, is known as the **partial autocorrelation function** (PACF) of $\{X_t\}$, discussed further in Section 3.2. Equation (2.5.22) shows the relation between $\alpha(n)$ and the reduction in the one-step mean squared error as the number of predictors is increased from $n-1$ to n. □

2.5.4 The Innovations Algorithm

The recursive algorithm to be discussed in this section is applicable to *all* series with finite second moments, regardless of whether they are stationary or not. Its application, however, can be simplified in certain special cases.

Suppose then that $\{X_t\}$ is a zero-mean series with $E|X_t|^2 < \infty$ for each t and

$$E(X_i X_j) = \kappa(i, j). \tag{2.5.24}$$

We denote the best one-step linear predictors and their mean squared errors by

$$\hat{X}_n = \begin{cases} 0, & \text{if } n = 1, \\ P_{n-1}X_n, & \text{if } n = 2, 3, \ldots, \end{cases}$$

and

$$v_n = E(X_{n+1} - P_n X_{n+1})^2.$$

We shall also introduce the **innovations**, or one-step prediction errors,

$$U_n = X_n - \hat{X}_n.$$

In terms of the vectors $\mathbf{U}_n = (U_1, \ldots, U_n)'$ and $\mathbf{X}_n = (X_1, \ldots, X_n)'$ the last equations can be written as

$$\mathbf{U}_n = A_n\mathbf{X}_n, \tag{2.5.25}$$

where A_n has the form

$$A_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{11} & 1 & 0 & \cdots & 0 \\ a_{22} & a_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ a_{n-1,n-1} & a_{n-1,n-2} & a_{n-1,n-3} & \cdots & 1 \end{bmatrix}.$$

(If $\{X_t\}$ is stationary, then $a_{ij} = -a_j$ with a_j as in (2.5.7) with $h = 1$.) This implies that A_n is nonsingular, with inverse C_n of the form

$$C_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 1 & 0 & \cdots & 0 \\ \theta_{22} & \theta_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1 \end{bmatrix}.$$

The vector of one-step predictors $\hat{\mathbf{X}}_n := (X_1, P_1X_2, \ldots, P_{n-1}X_n)'$ can therefore be expressed as

$$\hat{\mathbf{X}}_n = \mathbf{X}_n - \mathbf{U}_n = C_n\mathbf{U}_n - \mathbf{U}_n = \boldsymbol{\Theta}_n\left(\mathbf{X}_n - \hat{\mathbf{X}}_n\right), \tag{2.5.26}$$

where

$$\boldsymbol{\Theta}_n = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 0 & 0 & \cdots & 0 \\ \theta_{22} & \theta_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 0 \end{bmatrix}$$

and $\mathbf{X}_n$ itself satisfies

$$\mathbf{X}_n = C_n\left(\mathbf{X}_n - \hat{\mathbf{X}}_n\right). \tag{2.5.27}$$

Equation (2.5.26) can be rewritten as

$$\hat{X}_{n+1} = \begin{cases} 0, & \text{if } n = 0, \\ \sum_{j=1}^{n} \theta_{nj}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right), & \text{if } n = 1, 2, \ldots, \end{cases} \tag{2.5.28}$$

from which the one-step predictors $\hat{X}_1, \hat{X}_2, \ldots$ can be computed recursively once the coefficients θ_{ij} have been determined. The following algorithm generates these coefficients and the mean squared errors $v_i = E\left(X_{i+1} - \hat{X}_{i+1}\right)^2$, starting from the covariances $\kappa(i, j)$.

The Innovations Algorithm:
The coefficients $\theta_{n1}, \ldots, \theta_{nn}$ can be computed recursively from the equations

$$v_0 = \kappa(1, 1),$$

$$\theta_{n,n-k} = v_k^{-1}\left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}v_j\right), \quad 0 \le k < n,$$

and

$$v_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j.$$

(It is a trivial matter to solve first for v_0, then successively for $\theta_{11}, v_1; \theta_{22}, \theta_{21}, v_2; \theta_{33}, \theta_{32}, \theta_{31}, v_3; \ldots$.)

Proof See Brockwell and Davis (1991), Proposition 5.2.2. ■

Remark 5. While the Durbin–Levinson recursion gives the coefficients of $X_n, \ldots, X_1$ in the representation $\hat{X}_{n+1} = \sum_{j=1}^{n} \phi_{nj}X_{n+1-j}$, the innovations algorithm gives the coefficients of $\left(X_n - \hat{X}_n\right), \ldots, \left(X_1 - \hat{X}_1\right)$, in the alternative expansion $\hat{X}_{n+1} = \sum_{j=1}^{n} \theta_{nj}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right)$. The latter expansion has a number of advantages deriving from the fact that the innovations are uncorrelated (see Problem 2.20). It can also be greatly simplified in the case of ARMA(p, q) series, as we shall see in Section 3.3.

An immediate consequence of (2.5.28) is the innovations representation of X_{n+1} itself. Thus (defining $\theta_{n0} := 1$),

$$X_{n+1} = X_{n+1} - \hat{X}_{n+1} + \hat{X}_{n+1} = \sum_{j=0}^{n} \theta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right), \quad n = 0, 1, 2, \ldots.$$

□

Example 2.5.5 Recursive Prediction of an MA(1)

If $\{X_t\}$ is the time series defined by

$$X_t = Z_t + \theta Z_{t-1}, \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right),$$

then $\kappa(i, j) = 0$ for $|i-j| > 1, \kappa(i, i) = \sigma^2\left(1+\theta^2\right)$, and $\kappa(i, i+1) = \theta\sigma^2$. Application of the innovations algorithm leads at once to the recursions

$$\theta_{nj} = 0, 2 \le j \le n,$$
$$\theta_{n1} = v_{n-1}^{-1}\theta\sigma^2,$$
$$v_0 = (1 + \theta^2)\sigma^2,$$

and

$$v_n = \left[1 + \theta^2 - v_{n-1}^{-1}\theta^2\sigma^2\right]\sigma^2.$$

For the particular case

$$X_t = Z_t - 0.9Z_{t-1}, \quad \{Z_t\} \sim \text{WN}(0, 1),$$

the mean squared errors v_n of $\hat{X}_{n+1}$ and coefficients θ_{nj}, $1 \le j \le n$, in the innovations representation

$$\hat{X}_{n+1} = \sum_{j=1}^{n} \theta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right) = \theta_{n1} \left(X_n - \hat{X}_n \right)$$

are found from the recursions to be as follows:

$$\begin{aligned}
&v_0 = 1.8100,\\
&\theta_{11} = -0.4972, \quad v_1 = 1.3625,\\
&\theta_{21} = -0.6606, \quad \theta_{22} = 0, \quad v_2 = 1.2155,\\
&\theta_{31} = -0.7404, \quad \theta_{32} = 0, \quad \theta_{33} = 0, \quad v_3 = 1.1436,\\
&\theta_{41} = -0.7870, \quad \theta_{42} = 0, \quad \theta_{43} = 0, \quad \theta_{44} = 0, \quad v_4 = 1.1017.
\end{aligned}$$

If we apply the Durbin–Levinson algorithm to the same problem, we find that the mean squared errors v_n of $\hat{X}_{n+1}$ and coefficients ϕ_{nj}, $1 \le j \le n$, in the representation

$$\hat{X}_{n+1} = \sum_{j=1}^{n} \phi_{nj} X_{n+1-j}$$

are as follows:

$$\begin{aligned}
&v_0 = 1.8100,\\
&\phi_{11} = -0.4972, \quad v_1 = 1.3625,\\
&\phi_{21} = -0.6606, \quad \phi_{22} = -0.3285, \quad v_2 = 1.2155,\\
&\phi_{31} = -0.7404, \quad \phi_{32} = -0.4892, \quad \phi_{33} = -0.2433, \quad v_3 = 1.1436,\\
&\phi_{41} = -0.7870, \quad \phi_{42} = -0.5828, \quad \phi_{43} = -0.3850, \quad \phi_{44} = -0.1914, \quad v_4 = 1.1017.
\end{aligned}$$

Notice that as n increases, v_n approaches the white noise variance and θ_{n1} approaches θ. These results hold for any MA(1) process with $|\theta| < 1$. The innovations algorithm is particularly well suited to forecasting MA(q) processes, since for them $\theta_{nj} = 0$ for $n - j > q$. For AR(p) processes the Durbin–Levinson algorithm is usually more convenient, since $\phi_{nj} = 0$ for $n - j > p$.

□

2.5.5 Recursive Calculation of the *h*-Step Predictors

For h-step prediction we use the result

$$P_n(X_{n+k} - P_{n+k-1}X_{n+k}) = 0, \quad k \geq 1. \tag{2.5.29}$$

This follows from (2.5.10) and the fact that

$$E[(X_{n+k} - P_{n+k-1}X_{n+k} - 0)X_{n+j-1}] = 0, \quad j = 1, \ldots, n.$$

Hence,

$$\begin{aligned} P_nX_{n+h} &= P_nP_{n+h-1}X_{n+h} \\ &= P_n\hat{X}_{n+h} \\ &= P_n\left(\sum_{j=1}^{n+h-1} \theta_{n+h-1,j}\left(X_{n+h-j} - \hat{X}_{n+h-j}\right)\right). \end{aligned}$$

Applying (2.5.29) again and using the linearity of P_n we find that

$$P_nX_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}\left(X_{n+h-j} - \hat{X}_{n+h-j}\right), \tag{2.5.30}$$

where the coefficients θ_{nj} are determined as before by the innovations algorithm. Moreover, the mean squared error can be expressed as

$$\begin{aligned} E(X_{n+h} - P_nX_{n+h})^2 &= EX_{n+h}^2 - E(P_nX_{n+h})^2 \\ &= \kappa(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 v_{n+h-j-1}. \end{aligned} \tag{2.5.31}$$

2.5.6 Prediction of a Stationary Process in Terms of Infinitely Many Past Values

It is often useful, when many past observations $X_m, \ldots, X_0, X_1, \ldots, X_n$ ($m < 0$) are available, to evaluate the best linear predictor of X_{n+h} in terms of $1, X_m, \ldots, X_0, \ldots, X_n$. This predictor, which we shall denote by $P_{m,n}X_{n+h}$, can easily be evaluated by the methods described above. If $|m|$ is large, this predictor can be approximated by the sometimes more easily calculated mean square limit

$$\tilde{P}_nX_{n+h} = \lim_{m \to -\infty} P_{m,n}X_{n+h}.$$

We shall refer to $\tilde{P}_n$ as the **prediction operator based on the infinite past,** $\{X_t, -\infty < t \leq n\}$. Analogously we shall refer to P_n as the **prediction operator based on the finite past**, $\{X_1, \ldots, X_n\}$. (Mean square convergence of random variables is discussed in Appendix C.)

2.5.7 Determination of $\tilde{P}_n X_{n+h}$

If $\{X_n\}$ is a zero-mean stationary process with autocovariance function $\gamma(\cdot)$ then, just as $P_n X_{n+h}$ is characterized by equation (2.5.10), $\tilde{P}_n X_{n+h}$ is characterized by the equations

$$E\left[\left(X_{n+h} - \tilde{P}_n X_{n+h}\right) X_{n+1-i}\right] = 0, \quad i = 1, 2, \ldots.$$

If we can find a solution to these equations, it will necessarily be the uniquely defined predictor $\tilde{P}_n X_{n+h}$. An approach to this problem that is often effective is to *assume* that $\tilde{P}_n X_{n+h}$ can be expressed in the form

$$\tilde{P}_n X_{n+h} = \sum_{j=1}^{\infty} \alpha_j X_{n+1-j},$$

in which case the preceding equations reduce to

$$E\left[\left(X_{n+h} - \sum_{j=1}^{\infty} \alpha_j X_{n+1-j}\right) X_{n+1-i}\right] = 0, \quad i = 1, 2, \ldots,$$

or equivalently,

$$\sum_{j=1}^{\infty} \gamma(i-j)\alpha_j = \gamma(h+i-1), \quad i = 1, 2, \ldots.$$

This is an infinite set of linear equations for the unknown coefficients α_i that determine $\tilde{P}_n X_{n+h}$, *provided that the resulting series converges*.

Properties of $\tilde{P}_n$:
Suppose that $EU^2 < \infty$, $EV^2 < \infty$, a, b, and c are constants, and $\Gamma = \mathrm{Cov}(\mathbf{W}, \mathbf{W})$.

1. $E[(U - \tilde{P}_n(U))X_j] = 0$, $j \le n$.
2. $\tilde{P}_n(aU + bV + c) = a\tilde{P}_n(U) + b\tilde{P}_n(V) + c$.
3. $\tilde{P}_n(U) = U$ if U is a limit of linear combinations of X_j, $j \le n$.
4. $\tilde{P}_n(U) = EU$ if $\mathrm{Cov}(U, X_j) = 0$ for all $j \le n$.

These properties can sometimes be used to simplify the calculation of $\tilde{P}_n X_{n+h}$, notably when the process $\{X_t\}$ is an ARMA process.

Example 2.5.6 Consider the causal invertible ARMA(1,1) process $\{X_t\}$ defined by

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right).$$

We know from (2.3.3) and (2.3.5) that we have the representations

$$X_{n+1} = Z_{n+1} + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{n+1-j}$$

and

$$Z_{n+1} = X_{n+1} - (\phi + \theta) \sum_{j=1}^{\infty} (-\theta)^{j-1} X_{n+1-j}.$$

Applying the operator $\tilde{P}_n$ to the second equation and using the properties of $\tilde{P}_n$ gives

$$\tilde{P}_n X_{n+1} = (\phi + \theta) \sum_{j=1}^{\infty} (-\theta)^{j-1} X_{n+1-j}.$$

Applying the operator $\tilde{P}_n$ to the first equation and using the properties of $\tilde{P}_n$ gives

$$\tilde{P}_n X_{n+1} = (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{n+1-j}.$$

Hence,

$$X_{n+1} - \tilde{P}_n X_{n+1} = Z_{n+1},$$

and so the mean squared error of the predictor $\tilde{P}_n X_{n+1}$ is $EZ_{n+1}^2 = \sigma^2$.

□

2.6 The Wold Decomposition

Consider the stationary process

$$X_t = A\cos(\omega t) + B\sin(\omega t),$$

where $\omega \in (0, \pi)$ is constant and A, B are uncorrelated random variables with mean 0 and variance σ^2. Notice that

$$X_n = (2\cos\omega)X_{n-1} - X_{n-2} = \tilde{P}_{n-1}X_n, \quad n = 0, \pm 1, \ldots,$$

so that $X_n - \tilde{P}_{n-1}X_n = 0$ for all n. Processes with the latter property are said to be **deterministic**.

The Wold Decomposition:
If $\{X_t\}$ is a nondeterministic stationary time series, then

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t, \tag{2.6.1}$$

where

1. $\psi_0 = 1$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$,
2. $\{Z_t\} \sim \text{WN}\left(0, \sigma^2\right)$,
3. $\text{Cov}(Z_s, V_t) = 0$ for all s and t,
4. $Z_t = \tilde{P}_t Z_t$ for all t,
5. $V_t = \tilde{P}_s V_t$ for all s and t, and
6. $\{V_t\}$ is deterministic.

Here as in Section 2.5, $\tilde{P}_t Y$ denotes the best predictor of Y in terms of linear combinations, or limits of linear combinations of $1, X_s, -\infty < s \leq t$. The sequences $\{Z_t\}$, $\{\psi_j\}$, and $\{V_t\}$ are unique and can be written explicitly as $Z_t = X_t - \tilde{P}_{t-1}X_t$, $\psi_j = E(X_t Z_{t-j})/E\left(Z_t^2\right)$, and $V_t = X_t - \sum_{j=0}^{\infty} \psi_j Z_{t-j}$. (See Brockwell and Davis (1991), p. 188.) For most of the zero-mean stationary time series dealt with in this book (in particular for all ARMA processes) the deterministic component V_t is 0 for all t, and the series is then said to be **purely nondeterministic**.

Example 2.6.1 If $X_t = U_t + Y$, where $\{U_t\} \sim \text{WN}\left(0, \nu^2\right)$, $E(U_t Y) = 0$ for all t, and Y has mean 0 and variance τ^2, then $\tilde{P}_{t-1} X_t = Y$, since Y is the mean square limit as $s \to \infty$ of $[X_{t-1} + \cdots + X_{t-s}]/s$, and $E[(X_t - Y)X_s] = 0$ for all $s \leq t-1$. Hence the sequences in the Wold decomposition of $\{X_t\}$ are given by $Z_t = U_t$, $\psi_0 = 1$, $\psi_j = 0$ for $j > 0$, and $V_t = Y$.

□

Problems

2.1 Suppose that $X_1, X_2, \ldots$, is a stationary time series with mean μ and ACF $\rho(\cdot)$. Show that the best predictor of X_{n+h} of the form $aX_n + b$ is obtained by choosing $a = \rho(h)$ and $b = \mu(1 - \rho(h))$.

2.2 Show that the process

$$X_t = A\cos(\omega t) + B\sin(\omega t), \quad t = 0, \pm 1, \ldots$$

(where A and B are uncorrelated random variables with mean 0 and variance 1 and ω is a fixed frequency in the interval $[0, \pi]$), is stationary and find its mean and autocovariance function. Deduce that the function $\kappa(h) = \cos(\omega h)$, $h = 0, \pm 1, \ldots$, is nonnegative definite.

2.3 a. Find the ACVF of the time series $X_t = Z_t + 0.3Z_{t-1} - 0.4Z_{t-2}$, where $\{Z_t\} \sim \text{WN}(0, 1)$.
b. Find the ACVF of the time series $Y_t = \tilde{Z}_t - 1.2\tilde{Z}_{t-1} - 1.6\tilde{Z}_{t-2}$, where $\{\tilde{Z}_t\} \sim \text{WN}(0, 0.25)$. Compare with the answer found in (a).

2.4 It is clear that the function $\kappa(h) = 1$, $h = 0, \pm 1, \ldots$, is an autocovariance function, since it is the autocovariance function of the process $X_t = Z$, $t = 0, \pm 1, \ldots$, where Z is a random variable with mean 0 and variance 1. By identifying appropriate sequences of random variables, show that the following functions are also autocovariance functions:

a. $\kappa(h) = (-1)^{|h|}$

b. $\kappa(h) = 1 + \cos\left(\frac{\pi h}{2}\right) + \cos\left(\frac{\pi h}{4}\right)$

c. $\kappa(h) = \begin{cases} 1, & \text{if } h = 0, \\ 0.4, & \text{if } h = \pm 1, \\ 0, & \text{otherwise.} \end{cases}$

2.5 Suppose that $\{X_t, t = 0, \pm 1, \ldots\}$ is stationary and that $|\theta| < 1$. Show that for each fixed n the sequence

$$S_m = \sum_{j=1}^{m} \theta^j X_{n-j}$$

is convergent absolutely and in mean square (see Appendix C) as $m \to \infty$.

2.6 Verify the equations (2.2.6).

2.7 Show, using the geometric series $1/(1-x) = \sum_{j=0}^{\infty} x^j$ for $|x| < 1$, that $1/(1-\phi z) = -\sum_{j=1}^{\infty} \phi^{-j} z^{-j}$ for $|\phi| > 1$ and $|z| \geq 1$.

2.8 Show that the autoregressive equations

$$X_t = \phi_1 X_{t-1} + Z_t, \quad t = 0, \pm 1, \ldots,$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ and $|\phi| = 1$, have *no stationary solution*. HINT: Suppose there does exist a stationary solution $\{X_t\}$ and use the autoregressive equation to derive an expression for the variance of $X_t - \phi_1^{n+1} X_{t-n-1}$ that contradicts the stationarity assumption.

2.9 Let $\{Y_t\}$ be the AR(1) plus noise time series defined by

$$Y_t = X_t + W_t,$$

where $\{W_t\} \sim \text{WN}(0, \sigma_w^2)$, $\{X_t\}$ is the AR(1) process of Example 2.2.1, i.e.,

$$X_t - \phi X_{t-1} = Z_t, \{Z_t\} \sim \text{WN}\left(0, \sigma_z^2\right),$$

and $E(W_s Z_t) = 0$ for all s and t.

a. Show that $\{Y_t\}$ is stationary and find its autocovariance function.
b. Show that the time series $U_t := Y_t - \phi Y_{t-1}$ is 1-correlated and hence, by Proposition 2.1.1, is an MA(1) process.
c. Conclude from (b) that $\{Y_t\}$ is an ARMA(1,1) process and express the three parameters of this model in terms of ϕ, σ_w^2, and σ_z^2.

2.10 Use the program ITSM to compute the coefficients ψ_j and π_j, $j = 1, \ldots, 5$, in the expansions

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

and

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

for the ARMA(1,1) process defined by the equations

$$X_t - 0.5 X_{t-1} = Z_t + 0.5 Z_{t-1}, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right).$$

(Select `File>Project>New>Univariate`, then `Model>Specify`. In the resulting dialog box enter 1 for the AR and MA orders, specify $\phi(1) = \theta(1) = 0.5$, and click `OK`. Finally, select `Model>AR/MA Infinity>Default lag` and the values of ψ_j and π_j will appear on the screen.) Check the results with those obtained in Section 2.3.

2.11 Suppose that in a sample of size 100 from an AR(1) process with mean μ, $\phi = .6$, and $\sigma^2 = 2$ we obtain $\bar{x}_{100} = 0.271$. Construct an approximate 95 % confidence interval for μ. Are the data compatible with the hypothesis that $\mu = 0$?

2.12 Suppose that in a sample of size 100 from an MA(1) process with mean μ, $\theta = -0.6$, and $\sigma^2 = 1$ we obtain $\bar{x}_{100} = 0.157$. Construct an approximate 95 % confidence interval for μ. Are the data compatible with the hypothesis that $\mu = 0$?

2.13 Suppose that in a sample of size 100, we obtain $\hat{\rho}(1) = 0.438$ and $\hat{\rho}(2) = 0.145$.

a. Assuming that the data were generated from an AR(1) model, construct approximate 95 % confidence intervals for both $\rho(1)$ and $\rho(2)$. Based on these

two confidence intervals, are the data consistent with an AR(1) model with $\phi = 0.8$?

b. Assuming that the data were generated from an MA(1) model, construct approximate 95% confidence intervals for both $\rho(1)$ and $\rho(2)$. Based on these two confidence intervals, are the data consistent with an MA(1) model with $\theta = 0.6$?

2.14 Let $\{X_t\}$ be the process defined in Problem 2.2.

a. Find P_1X_2 and its mean squared error.

b. Find P_2X_3 and its mean squared error.

c. Find $\tilde{P}_nX_{n+1}$ and its mean squared error.

2.15 Suppose that $\{X_t, t = 0, \pm 1, \ldots\}$ is a stationary process satisfying the equations

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t,$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ and Z_t is uncorrelated with X_s for each $s < t$. Show that the best linear predictor P_nX_{n+1} of X_{n+1} in terms of $1, X_1, \ldots, X_n$, assuming $n > p$, is

$$P_nX_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p}.$$

What is the mean squared error of P_nX_{n+1}?

2.16 Use the program ITSM to plot the sample ACF and PACF up to lag 40 of the sunspot series $D_t, t = 1, 100$, contained in the ITSM file SUNSPOTS.TSM. (Open the project SUNSPOTS.TSM and click on the second yellow button at the top of the screen to see the graphs. Repeated clicking on this button will toggle between graphs of the sample ACF, sample PACF, and both. To see the numerical values, right-click on the graph and select `Info`.) Fit an AR(2) model to the mean-corrected data by selecting `Model>Estimation>Preliminary` and click `Yes` to subtract the sample mean from the data. In the dialog box that follows, enter 2 for the AR order and make sure that the MA order is zero and that the `Yule-Walker` algorithm is selected *without* AICC minimization. Click `OK` and you will obtain a model of the form

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t, \qquad \text{where } \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right),$$

for the mean-corrected series $X_t = D_t - 46.93$. Record the values of the estimated parameters ϕ_1, ϕ_2, and σ^2. Compare the model and sample ACF and PACF by selecting the third yellow button at the top of the screen. Print the graphs by right-clicking and selecting `Print`.

2.17 Without exiting from ITSM, use the model found in the preceding problem to compute forecasts of the next ten values of the sunspot series. (Select `Forecasting>ARMA`, make sure that the number of forecasts is set to 10 and the box `Add the mean to the forecasts` is checked, and then click `OK`. You will see a graph of the original data with the ten forecasts appended. Right-click on the graph and then on `Info` to get the numerical values of the forecasts. Print the graph as described in Problem 2.16.) The details of the calculations will be taken up in Chapter 3 when we discuss ARMA models in detail.

2.18 Let $\{X_t\}$ be the stationary process defined by the equations

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

where $|\theta| < 1$ and $\{Z_t\} \sim \text{WN}(0, \sigma^2)$. Show that the best linear predictor $\tilde{P}_nX_{n+1}$ of X_{n+1} based on $\{X_j, -\infty < j \le n\}$ is

$$\tilde{P}_n X_{n+1} = -\sum_{j=1}^{\infty} \theta^j X_{n+1-j}.$$

What is the mean squared error of the predictor $\tilde{P}_n X_{n+1}$?

2.19 If $\{X_t\}$ is defined as in Problem 2.18 and $\theta = 1$, find the best linear predictor $P_n X_{n+1}$ of X_{n+1} in terms of $X_1, \ldots, X_n$. What is the corresponding mean squared error?

2.20 In the innovations algorithm, show that for each $n \geq 2$, the innovation $X_n - \hat{X}_n$ is uncorrelated with $X_1, \ldots, X_{n-1}$. Conclude that $X_n - \hat{X}_n$ is uncorrelated with the innovations $X_1 - \hat{X}_1, \ldots, X_{n-1} - \hat{X}_{n-1}$.

2.21 Let X_1, X_2, X_4, X_5 be observations from the MA(1) model

$$X_t = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right).$$

a. Find the best linear estimate of the missing value X_3 in terms of X_1 and X_2.
b. Find the best linear estimate of the missing value X_3 in terms of X_4 and X_5.
c. Find the best linear estimate of the missing value X_3 in terms of X_1, X_2, X_4, and X_5.
d. Compute the mean squared errors for each of the estimates in (a)–(c).

2.22 Repeat parts (a)–(d) of Problem 2.21 assuming now that the observations X_1, X_2, X_4, X_5 are from the causal AR(1) model

$$X_t = \phi X_{t-1} + Z_t, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right).$$