3 ARMA Models

In this chapter we introduce an important parametric family of stationary time series, the autoregressive moving-average, or ARMA, processes. For a large class of autocovariance functions $\gamma(\cdot)$ it is possible to find an ARMA process $\{X_t\}$ with ACVF $\gamma_X(\cdot)$ such that $\gamma(\cdot)$ is well approximated by $\gamma_X(\cdot)$. In particular, for any positive integer K, there exists an ARMA process $\{X_t\}$ such that $\gamma_X(h) = \gamma(h)$ for $h = 0, 1, \ldots, K$. For this (and other) reasons, the family of ARMA processes plays a key role in the modeling of time series data. The linear structure of ARMA processes also leads to a substantial simplification of the general methods for linear prediction discussed earlier in Section 2.5.

3.1 ARMA(p, q) Processes

In Section 2.3 we introduced an ARMA(1,1) process and discussed some of its key properties. These included existence and uniqueness of stationary solutions of the defining equations and the concepts of causality and invertibility. In this section we extend these notions to the general ARMA(p, q) process.

P.J. Brockwell, R.A. Davis, *Introduction to Time Series and Forecasting*,
Springer Texts in Statistics, DOI 10.1007/978-3-319-29854-2_3

Definition 3.1.1

$\{X_t\}$ is an **ARMA(p, q) process** if $\{X_t\}$ is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{3.1.1}$$

where $\{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right)$ and the polynomials $\left(1 - \phi_1 z - \ldots - \phi_p z^p\right)$ and $\left(1 + \theta_1 z + \ldots + \theta_q z^q\right)$ have no common factors.

The process $\{X_t\}$ is said to be an **ARMA(p, q) process with mean** μ if $\{X_t - \mu\}$ is an ARMA(p, q) process.

It is convenient to use the more concise form of (3.1.1)

$$\phi(B) X_t = \theta(B) Z_t, \tag{3.1.2}$$

where $\phi(\cdot)$ and $\theta(\cdot)$ are the pth and qth-degree polynomials

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$

and B is the backward shift operator ($B^j X_t = X_{t-j}$, $B^j Z_t = Z_{t-j}$, $j = 0, \pm 1, \ldots$). The time series $\{X_t\}$ is said to be an **autoregressive process of order p** (or AR(p)) if $\theta(z) \equiv 1$, and a **moving-average process of order q** (or MA(q)) if $\phi(z) \equiv 1$.

An important part of Definition 3.1.1 is the requirement that $\{X_t\}$ be stationary. In Section 2.3 we showed, for the ARMA(1,1) equations (2.3.1), that a stationary solution exists (and is unique) if and only if $\phi_1 \neq \pm 1$. The latter is equivalent to the condition that the autoregressive polynomial $\phi(z) = 1 - \phi_1 z \neq 0$ for $z = \pm 1$. The analogous condition for the general ARMA(p, q) process is $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for all complex z with $|z| = 1$. (Complex z is used here, since the zeros of a polynomial of degree $p > 1$ may be either real or complex. The region defined by the set of complex z such that $|z| = 1$ is referred to as the unit circle.) If $\phi(z) \neq 0$ for all z on the unit circle, then there exists $\delta > 0$ such that

$$\frac{1}{\phi(z)} = \sum_{j=-\infty}^{\infty} \chi_j z^j \quad \text{for } 1 - \delta < |z| < 1 + \delta,$$

and $\sum_{j=-\infty}^{\infty} |\chi_j| < \infty$. We can then define $1/\phi(B)$ as the linear filter with absolutely summable coefficients

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

Applying the operator $\chi(B) := 1/\phi(B)$ to both sides of (3.1.2), we obtain

$$X_t = \chi(B)\phi(B) X_t = \chi(B)\theta(B) Z_t = \psi(B) Z_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \tag{3.1.3}$$

where $\psi(z) = \chi(z)\theta(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$. Using the argument given in Section 2.3 for the ARMA(1,1) process, it follows that $\psi(B) Z_t$ is the unique stationary solution of (3.1.1).

Existence and Uniqueness:

A stationary solution $\{X_t\}$ of equation (3.1.1) exists (and is also the unique stationary solution) if and only if

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \quad \text{for all } |z| = 1. \tag{3.1.4}$$

In Section 2.3 we saw that the ARMA(1,1) process is causal, i.e., that X_t can be expressed in terms of $Z_s,\ s \le t$, if and only if $|\phi_1| < 1$. For a general ARMA(p, q) process the analogous condition is that $\phi(z) \neq 0$ for $|z| \le 1$, i.e., the zeros of the autoregressive polynomial must all be greater than 1 in absolute value.

Causality:

An ARMA(p, q) process $\{X_t\}$ is **causal**, or a **causal function of** $\{Z_t\}$, if there exist constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \text{ for all } t. \tag{3.1.5}$$

Causality is equivalent to the condition

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \text{ for all } |z| \le 1. \tag{3.1.6}$$

The proof of the equivalence between causality and (3.1.6) follows from elementary properties of power series. From (3.1.3) we see that $\{X_t\}$ is causal if and only if $\chi(z) := 1/\phi(z) = \sum_{j=0}^{\infty} \chi_j z^j$ (assuming that $\phi(z)$ and $\theta(z)$ have no common factors). But this, in turn, is equivalent to (3.1.6).

The sequence $\{\psi_j\}$ in (3.1.5) is determined by the relation $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z)$, or equivalently by the identity

$$\left(1 - \phi_1 z - \cdots - \phi_p z^p\right)(\psi_0 + \psi_1 z + \cdots) = 1 + \theta_1 z + \cdots + \theta_q z^q.$$

Equating coefficients of $z^j, j = 0, 1, \ldots$, we find that

$$\begin{aligned}
1 &= \psi_0, \\
\theta_1 &= \psi_1 - \psi_0 \phi_1, \\
\theta_2 &= \psi_2 - \psi_1 \phi_1 - \psi_0 \phi_2, \\
&\vdots
\end{aligned}$$

or equivalently,

$$\psi_j - \sum_{k=1}^{p} \phi_k \psi_{j-k} = \theta_j, \quad j = 0, 1, \ldots, \tag{3.1.7}$$

where $\theta_0 := 1$, $\theta_j := 0$ for $j > q$, and $\psi_j := 0$ for $j < 0$.

Invertibility, which allows Z_t to be expressed in terms of $X_s,\ s \le t$, has a similar characterization in terms of the moving-average polynomial.

Invertibility:

An ARMA(p, q) process $\{X_t\}$ is **invertible** if there exist constants $\{\pi_j\}$ such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \text{ for all } t.$$

Invertibility is equivalent to the condition

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \neq 0 \text{ for all } |z| \leq 1.$$

Interchanging the roles of the AR and MA polynomials, we find from (3.1.7) that the sequence $\{\pi_j\}$ is determined by the equations

$$\pi_j + \sum_{k=1}^{q} \theta_k \pi_{j-k} = -\phi_j, \quad j = 0, 1, \ldots, \tag{3.1.8}$$

where $\phi_0 := -1$, $\phi_j := 0$ for $j > p$, and $\pi_j := 0$ for $j < 0$.

Example 3.1.1 An ARMA(1,1) Process Consider the ARMA(1,1) process $\{X_t\}$ satisfying the equations

$$X_t - 0.5X_{t-1} = Z_t + 0.4Z_{t-1}, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right). \tag{3.1.9}$$

Since the autoregressive polynomial $\phi(z) = 1 - 0.5z$ has a zero at $z = 2$, which is located outside the unit circle, we conclude from (3.1.4) and (3.1.6) that there exists a unique ARMA process satisfying (3.1.9) that is also causal. The coefficients $\{\psi_j\}$ in the MA(∞) representation of $\{X_t\}$ are found directly from (3.1.7):

$$\begin{aligned}
\psi_0 &= 1, \\
\psi_1 &= 0.4 + 0.5, \\
\psi_2 &= 0.5(0.4 + 0.5), \\
\psi_j &= 0.5^{j-1}(0.4 + 0.5), \quad j = 1, 2, \ldots.
\end{aligned}$$

The MA polynomial $\theta(z) = 1 + 0.4z$ has a zero at $z = -1/0.4 = -2.5$, which is also located outside the unit circle. This implies that $\{X_t\}$ is invertible with coefficients $\{\pi_j\}$ given by [see (3.1.8)]

$$\begin{aligned}
\pi_0 &= 1, \\
\pi_1 &= -(0.4 + 0.5), \\
\pi_2 &= -(0.4 + 0.5)(-0.4), \\
\pi_j &= -(0.4 + 0.5)(-0.4)^{j-1}, \quad j = 1, 2, \ldots.
\end{aligned}$$

(A direct derivation of these formulas for $\{\psi_j\}$ and $\{\pi_j\}$ was given in Section 2.3 without appealing to the recursions (3.1.7) and (3.1.8).)

□

Example 3.1.2 An AR(2) Process

Let $\{X_t\}$ be the AR(2) process

$$X_t = 0.7X_{t-1} - 0.1X_{t-2} + Z_t, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right).$$

The autoregressive polynomial for this process has the factorization $\phi(z) = 1 - 0.7z + 0.1z^2 = (1 - 0.5z)(1 - 0.2z)$, and is therefore zero at $z = 2$ and $z = 5$. Since these zeros lie outside the unit circle, we conclude that $\{X_t\}$ is a causal AR(2) process with coefficients $\{\psi_j\}$ given by

$$\begin{aligned}\psi_0 &= 1,\\ \psi_1 &= 0.7,\\ \psi_2 &= 0.7^2 - 0.1,\\ \psi_j &= 0.7\psi_{j-1} - 0.1\psi_{j-2}, \quad j = 2, 3, \ldots.\end{aligned}$$

While it is a simple matter to calculate ψ_j numerically for any j, it is possible also to give an explicit solution of these difference equations using the theory of linear difference equations (see Brockwell and Davis (1991), Section 3.6).

□

The option `Model>Specify` of the program ITSM allows the entry of any causal ARMA(p, q) model with $p < 28$ and $q < 28$. This option contains a causality check and will immediately let you know if the entered model is noncausal. (A causal model can be obtained by setting all the AR coefficients equal to 0.001.) Once a causal model has been entered, the coefficients ψ_j in the MA(∞) representation of the process can be computed by selecting `Model>AR/MA Infinity`. This option will also compute the AR(∞) coefficients π_j, provided that the model is invertible.

Example 3.1.3 An ARMA(2,1) Process

Consider the ARMA(2,1) process defined by the equations

$$X_t - 0.75X_{t-1} + 0.5625X_{t-2} = Z_t + 1.25Z_{t-1}, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right).$$

The AR polynomial $\phi(z) = 1 - 0.75z + 0.5625z^2$ has zeros at $z = 2\left(1 \pm i\sqrt{3}\right)/3$, which lie outside the unit circle. The process is therefore causal. On the other hand, the MA polynomial $\theta(z) = 1 + 1.25z$ has a zero at $z = -0.8$, and hence $\{X_t\}$ is not invertible.

□

Remark 1. It should be noted that causality and invertibility are properties not of $\{X_t\}$ alone, but rather of the relationship between the two processes $\{X_t\}$ and $\{Z_t\}$ appearing in the defining ARMA equations (3.1.1). □

Remark 2. If $\{X_t\}$ is an ARMA process defined by $\phi(B)X_t = \theta(B)Z_t$, where $\theta(z) \neq 0$ if $|z| = 1$, then it is always possible (see Brockwell and Davis (1991), p. 127) to find polynomials $\tilde{\phi}(z)$ and $\tilde{\theta}(z)$ and a white noise sequence $\{W_t\}$ such that $\tilde{\phi}(B)X_t = \tilde{\theta}(B)W_t$ and $\tilde{\theta}(z)$ and $\tilde{\phi}(z)$ are nonzero for $|z| \leq 1$. However, if the original white noise sequence $\{Z_t\}$ is iid, then the new white noise sequence will not be iid unless $\{Z_t\}$ is Gaussian.

□

In view of Remark 2, we will focus our attention principally on causal and invertible ARMA processes.

3.2 The ACF and PACF of an ARMA(p, q) Process

In this section we discuss three methods for computing the autocovariance function $\gamma(\cdot)$ of a causal ARMA process $\{X_t\}$. The autocorrelation function is readily found from the ACVF on dividing by $\gamma(0)$. The partial autocorrelation function (PACF) is also found from the function $\gamma(\cdot)$.

3.2.1 Calculation of the ACVF

First we determine the ACVF $\gamma(\cdot)$ of the causal ARMA(p, q) process defined by

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right), \tag{3.2.1}$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$. The causality assumption implies that

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \tag{3.2.2}$$

where $\sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z)$, $|z| \le 1$. The calculation of the sequence $\{\psi_j\}$ was discussed in Section 3.1.

First Method. From Proposition 2.2.1 and the representation (3.2.2), we obtain

$$\gamma(h) = E(X_{t+h}X_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}. \tag{3.2.3}$$

Example 3.2.1 The ARMA(1,1) Process

Substituting from (2.3.3) into (3.2.3), we find that the ACVF of the process defined by

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right), \tag{3.2.4}$$

with $|\phi| < 1$ is given by

$$\begin{aligned}
\gamma(0) &= \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 \\
&= \sigma^2 \left[1 + (\theta + \phi)^2 \sum_{j=0}^{\infty} \phi^{2j} \right] \\
&= \sigma^2 \left[1 + \frac{(\theta + \phi)^2}{1 - \phi^2} \right], \\
\gamma(1) &= \sigma^2 \sum_{j=0}^{\infty} \psi_{j+1} \psi_j \\
&= \sigma^2 \left[\theta + \phi + (\theta + \phi)^2 \phi \sum_{j=0}^{\infty} \phi^{2j} \right] \\
&= \sigma^2 \left[\theta + \phi + \frac{(\theta + \phi)^2 \phi}{1 - \phi^2} \right],
\end{aligned}$$

and

$$\gamma(h) = \phi^{h-1} \gamma(1), \quad h \ge 2.$$

□

Example 3.2.2 The MA(q) Process

For the process

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right),$$

Equation (3.2.3) immediately gives the result

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, & \text{if } |h| \le q, \\ 0, & \text{if } |h| > q, \end{cases}$$

where θ_0 is defined to be 1. The ACVF of the MA(q) process thus has the distinctive feature of vanishing at lags greater than q. Data for which the sample ACVF is small for lags greater than q therefore suggest that an appropriate model might be a moving average of order q (or less). Recall from Proposition 2.1.1 that *every* zero-mean stationary process with correlations vanishing at lags greater than q can be represented as a moving-average process of order q or less.

□

Second Method. If we multiply each side of the equations

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

by X_{t-k}, $k = 0, 1, 2, \ldots$, and take expectations on each side, we find that

$$\gamma(k) - \phi_1 \gamma(k-1) - \cdots - \phi_p \gamma(k-p) = \sigma^2 \sum_{j=0}^{\infty} \theta_{k+j} \psi_j, \quad 0 \le k < m, \tag{3.2.5}$$

and

$$\gamma(k) - \phi_1 \gamma(k-1) - \cdots - \phi_p \gamma(k-p) = 0, \quad k \ge m, \tag{3.2.6}$$

where $m = \max(p, q+1)$, $\psi_j := 0$ for $j < 0$, $\theta_0 := 1$, and $\theta_j := 0$ for $j \notin \{0, \ldots, q\}$. In calculating the right-hand side of (3.2.5) we have made use of the expansion (3.2.2). Equations (3.2.6) are a set of homogeneous linear difference equations with constant coefficients, for which the solution is well known (see, e.g., Brockwell and Davis (1991), Section 3.6) to be of the form

$$\gamma(h) = \alpha_1 \xi_1^{-h} + \alpha_2 \xi_2^{-h} + \cdots + \alpha_p \xi_p^{-h}, \quad h \ge m - p, \tag{3.2.7}$$

where $\xi_1, \ldots, \xi_p$ are the roots (assumed to be distinct) of the equation $\phi(z) = 0$, and $\alpha_1, \ldots, \alpha_p$ are arbitrary constants. (For further details, and for the treatment of the case where the roots are not distinct, see Brockwell and Davis (1991), Section 3.6.) Of course, we are looking for the solution of (3.2.6) that also satisfies (3.2.5). We therefore substitute the solution (3.2.7) into (3.2.5) to obtain a set of m linear equations that then uniquely determine the constants $\alpha_1, \ldots, \alpha_p$ and the $m - p$ autocovariances $\gamma(h)$, $0 \le h < m - p$.

Example 3.2.3 The ARMA(1,1) Process

For the causal ARMA(1,1) process defined in Example 3.2.1, equations (3.2.5) are

$$\gamma(0) - \phi\gamma(1) = \sigma^2(1 + \theta(\theta + \phi)) \tag{3.2.8}$$

and

$$\gamma(1) - \phi\gamma(0) = \sigma^2\theta. \tag{3.2.9}$$

Equation (3.2.6) takes the form

$$\gamma(k) - \phi\gamma(k-1) = 0, \quad k \geq 2. \tag{3.2.10}$$

The solution of (3.2.10) is

$$\gamma(h) = \alpha\phi^h, \quad h \geq 1.$$

Substituting this expression for $\gamma(h)$ into the two preceding equations (3.2.8) and (3.2.9) gives two linear equations for α and the unknown autocovariance $\gamma(0)$. These equations are easily solved, giving the autocovariances already found for this process in Example 3.2.1.

□

Example 3.2.4 The General AR(2) Process

For the causal AR(2) process defined by

$$\left(1 - \xi_1^{-1}B\right)\left(1 - \xi_2^{-1}B\right)X_t = Z_t, \quad |\xi_1|, |\xi_2| > 1, \xi_1 \neq \xi_2,$$

we easily find from (3.2.7) and (3.2.5) using the relations

$$\phi_1 = \xi_1^{-1} + \xi_2^{-1}$$

and

$$\phi_2 = -\xi_1^{-1}\xi_2^{-1}$$

that

$$\gamma(h) = \frac{\sigma^2\xi_1^2\xi_2^2}{(\xi_1\xi_2 - 1)(\xi_2 - \xi_1)}\left[(\xi_1^2 - 1)^{-1}\xi_1^{1-h} - (\xi_2^2 - 1)^{-1}\xi_2^{1-h}\right]. \tag{3.2.11}$$

Figures 3-1, 3-2, 3-3, and 3-4 illustrate some of the possible forms of $\gamma(\cdot)$ for different values of ξ_1 and ξ_2. Notice that in the case of complex conjugate roots $\xi_1 = re^{i\theta}$ and $\xi_2 = re^{-i\theta}$, $0 < \theta < \pi$, we can write (3.2.11) in the more illuminating form

$$\gamma(h) = \frac{\sigma^2 r^4 \cdot r^{-h}\sin(h\theta + \psi)}{(r^2 - 1)(r^4 - 2r^2\cos 2\theta + 1)\sin\theta}, \tag{3.2.12}$$

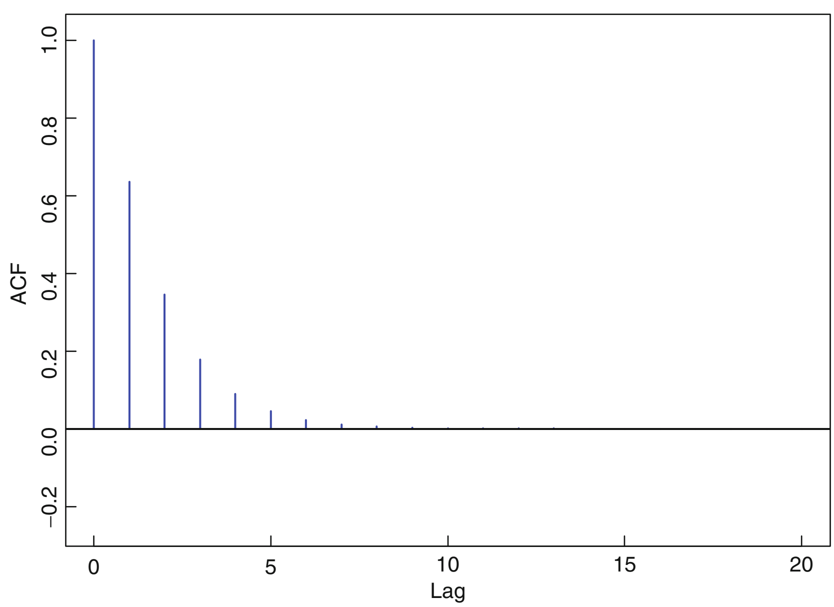

Figure 3-1 The model ACF of the AR(2) series of Example 3.2.4 with $\xi_1 = 2$ and $\xi_2 = 5$

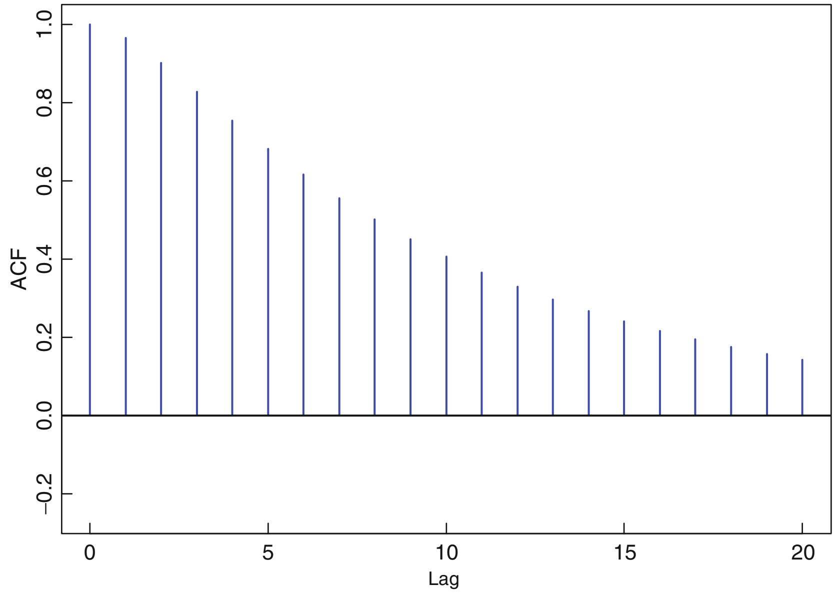

Figure 3-2
The model ACF of the AR(2) series of Example 3.2.4 with ξ_1=10/9 and ξ_2=2

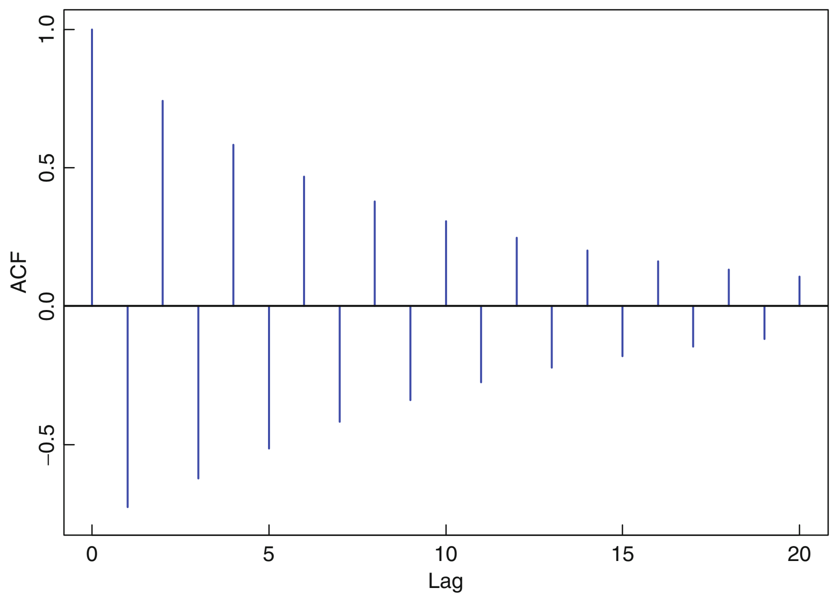

Figure 3-3
The model ACF of the AR(2) series of Example 3.2.4 with $\xi_1 = -10/9$ and $\xi_2 = 2$

where

$$\tan\psi = \frac{r^2+1}{r^2-1}\tan\theta \tag{3.2.13}$$

and $\cos\psi$ has the same sign as $\cos\theta$. Thus in this case $\gamma(\cdot)$ has the form of a damped sinusoidal function with damping factor r^{-1} and period $2\pi/\theta$. If the roots are close to the unit circle, then r is close to 1, the damping is slow, and we obtain a nearly sinusoidal autocovariance function.

□

Third Method. The autocovariances can also be found by solving the first $p+1$ equations of (3.2.5) and (3.2.6) for $\gamma(0)\ldots,\gamma(p)$ and then using the subsequent equations to solve successively for $\gamma(p+1), \gamma(p+2), \ldots$. This is an especially convenient method for numerical determination of the autocovariances $\gamma(h)$ and is used in the option `Model>ACF/PACF>Model` of the program ITSM.

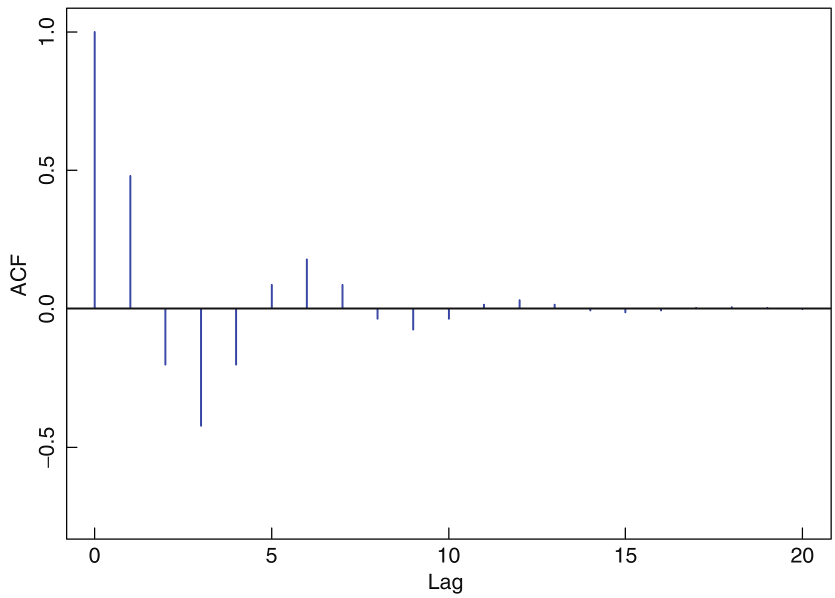

Figure 3-4
The model ACF of the AR(2) series of Example 3.2.4 with $\xi_1 = 2(1 + i\sqrt{3})/3$ and $\xi_2 = 2(1 - i\sqrt{3})/3$

Example 3.2.5 Consider again the causal ARMA(1,1) process of Example 3.2.1. To apply the third method we simply solve (3.2.8) and (3.2.9) for $\gamma(0)$ and $\gamma(1)$. Then $\gamma(2), \gamma(3), \ldots$ can be found successively from (3.2.10). It is easy to check that this procedure gives the same results as those obtained in Examples 3.2.1 and 3.2.3.

□

3.2.2 The Autocorrelation Function

Recall that the ACF of an ARMA process $\{X_t\}$ is the function $\rho(\cdot)$ found immediately from the ACVF $\gamma(\cdot)$ as

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

Likewise, for any set of observations $\{x_1, \ldots, x_n\}$, the sample ACF $\hat{\rho}(\cdot)$ is computed as

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

The Sample ACF of an MA(q) Series. Given observations $\{x_1, \ldots, x_n\}$ of a time series, one approach to the fitting of a model to the data is to match the sample ACF of the data with the ACF of the model. In particular, if the sample ACF $\hat{\rho}(h)$ is significantly different from zero for $0 \le h \le q$ and negligible for $h > q$, Example 3.2.2 suggests that an MA(q) model might provide a good representation of the data. In order to apply this criterion we need to take into account the random variation expected in the sample autocorrelation function before we can classify ACF values as "negligible." To resolve this problem we can use Bartlett's formula (Section 2.4), which implies that for a large sample of size n from an MA(q) process, the sample ACF values at lags h greater than q are approximately normally distributed with means 0 and variances $w_{hh}/n = \left(1 + 2\rho^2(1) + \cdots + 2\rho^2(q)\right)/n$. This means that if the sample is from an MA(q) process and if $h > q$, then $\hat{\rho}(h)$ should fall between the bounds $\pm 1.96\sqrt{w_{hh}/n}$ with probability approximately 0.95. In practice we frequently use the more stringent values $\pm 1.96/\sqrt{n}$ as the bounds between which sample autocovariances are considered "negligible." A more effective and systematic approach to the problem of model selection, which also applies to ARMA(p, q) models with $p > 0$ and $q > 0$, will be discussed in Section 5.5.

3.2.3 The Partial Autocorrelation Function

The **partial autocorrelation function (PACF)** of an ARMA process $\{X_t\}$ is the function $\alpha(\cdot)$ defined by the equations

$$\alpha(0) = 1$$

and

$$\alpha(h) = \phi_{hh}, \quad h \geq 1,$$

where ϕ_{hh} is the last component of

$$\boldsymbol{\phi}_h = \Gamma_h^{-1}\boldsymbol{\gamma}_h, \tag{3.2.14}$$

$\Gamma_h = \left[\gamma(i-j)\right]_{i,j=1}^{h}$, and $\boldsymbol{\gamma}_h = \left[\gamma(1), \gamma(2), \ldots, \gamma(h)\right]'$.

For any set of observations $\{x_1, \ldots, x_n\}$ with $x_i \neq x_j$ for some i and j, the **sample PACF** $\hat{\alpha}(h)$ is given by

$$\hat{\alpha}(0) = 1$$

and

$$\hat{\alpha}(h) = \hat{\phi}_{hh}, \quad h \geq 1,$$

where $\hat{\phi}_{hh}$ is the last component of

$$\hat{\boldsymbol{\phi}}_h = \hat{\Gamma}_h^{-1}\hat{\boldsymbol{\gamma}}_h. \tag{3.2.15}$$

We show in the next example that the PACF of a causal AR(p) process is zero for lags greater than p. Both sample and model partial autocorrelation functions can be computed numerically using the program ITSM. Algebraic calculation of the PACF is quite complicated except when q is zero or p and q are both small.

It can be shown (Brockwell and Davis (1991), p. 171) that ϕ_{hh} is the correlation between the prediction errors $X_h - P(X_h|X_1, \ldots, X_{h-1})$ and $X_0 - P(X_0|X_1, \ldots, X_{h-1})$.

Example 3.2.6 The PACF of an AR(p) Process

For the causal AR(p) process defined by

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right),$$

we know (Problem 2.15) that for $h \geq p$ the best linear predictor of X_{h+1} in terms of $1, X_1, \ldots, X_h$ is

$$\hat{X}_{h+1} = \phi_1 X_h + \phi_2 X_{h-1} + \cdots + \phi_p X_{h+1-p}.$$

Since the coefficient ϕ_{hh} of X_1 is ϕ_p if $h = p$ and 0 if $h > p$, we conclude that the PACF $\alpha(\cdot)$ of the process $\{X_t\}$ has the properties

$$\alpha(p) = \phi_p$$

and

$$\alpha(h) = 0 \text{ for } h > p.$$

For $h < p$ the values of $\alpha(h)$ can easily be computed from (3.2.14). For any specified ARMA model the PACF can be evaluated numerically using the option `Model>ACF/PACF>Model` of the program ITSM.

□

Example 3.2.7 The PACF of an MA(1) Process

For the MA(1) process, it can be shown from (3.2.14) (see Problem 3.12) that the PACF at lag h is

$$\alpha(h) = \phi_{hh} = -(-\theta)^h / \left(1 + \theta^2 + \cdots + \theta^{2h}\right).$$

□

The Sample PACF of an AR(p) Series. If $\{X_t\}$ is an AR(p) series, then the sample PACF based on observations $\{x_1, \ldots, x_n\}$ should reflect (with sampling variation) the properties of the PACF itself. In particular, if the sample PACF $\hat{\alpha}(h)$ is significantly different from zero for $0 \le h \le p$ and negligible for $h > p$, Example 3.2.6 suggests that an AR(p) model might provide a good representation of the data. To decide what is meant by "negligible" we can use the result that for an AR(p) process the sample PACF values at lags greater than p are approximately independent $N(0, 1/n)$ random variables. This means that roughly 95 % of the sample PACF values beyond lag p should fall within the bounds $\pm 1.96/\sqrt{n}$. If we observe a sample PACF satisfying $|\hat{\alpha}(h)| > 1.96/\sqrt{n}$ for $0 \le h \le p$ and $|\hat{\alpha}(h)| < 1.96/\sqrt{n}$ for $h > p$, this suggests an AR(p) model for the data. For a more systematic approach to model selection, see Section 5.5.

3.2.4 Examples

Example 3.2.8 The time series plotted in Figure 3-5 consists of 57 consecutive daily *overshorts* from an underground gasoline tank at a filling station in Colorado. If y_t is the measured amount of fuel in the tank at the end of the tth day and a_t is the measured amount sold minus the amount delivered during the course of the tth day, then the overshort at the end of day t is defined as $x_t = y_t - y_{t-1} + a_t$. Due to the error in measuring the current amount of fuel in the tank, the amount sold, and the amount delivered to the station, we view y_t, a_t, and x_t as observed values from some set of random variables Y_t, A_t, and X_t for $t = 1, \ldots, 57$. (In the absence of any measurement error and any leak in the tank, each x_t would be zero.) The data and their ACF are plotted in Figures 3-5 and 3-6. To check the plausibility of an MA(1) model, the bounds $\pm 1.96 \left(1 + 2\hat{\rho}^2(1)\right)^{1/2} / n^{1/2}$ are also plotted in Figure 3-6. Since $\hat{\rho}(h)$ is well within these bounds for $h > 1$, the data

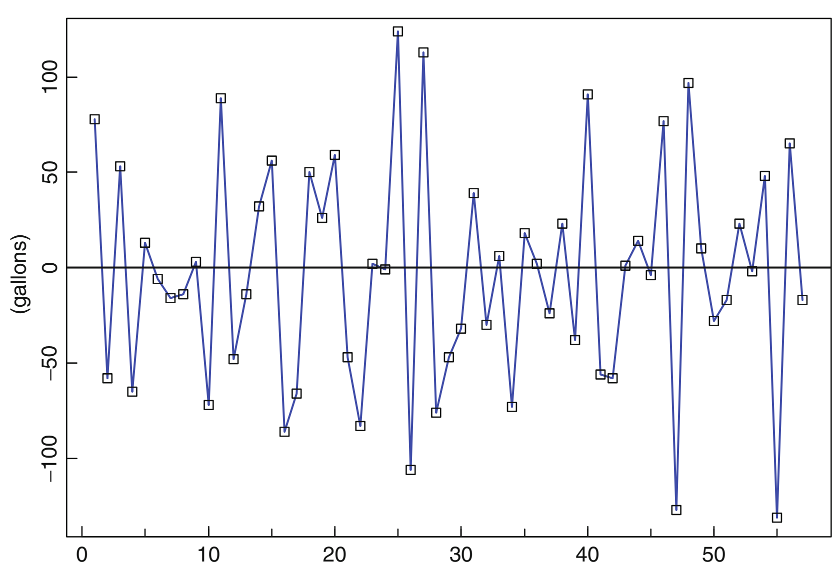

Figure 3-5
Time series of the overshorts in Example 3.2.8

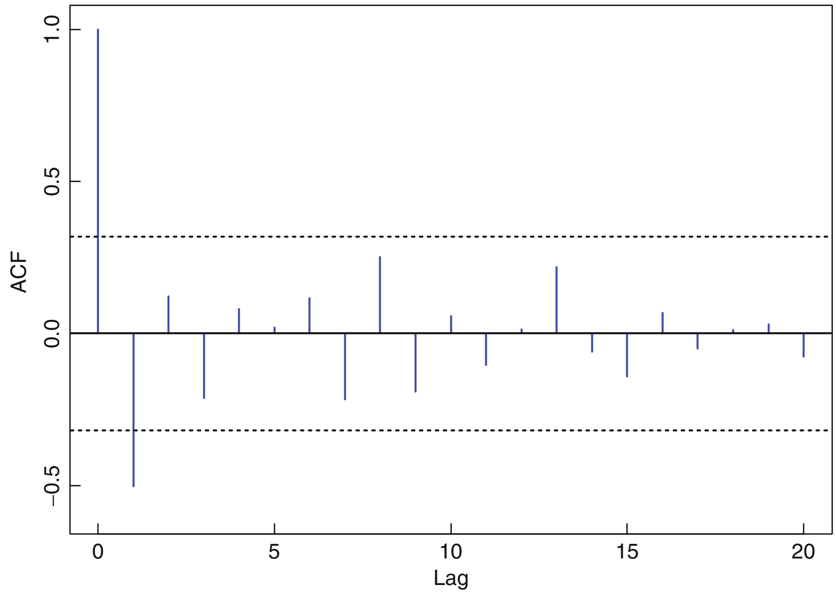

Figure 3-6
The sample ACF of the data in Figure 3-5 showing the bounds $\pm 1.96 n^{-1/2} \left(1 + 2\hat{\rho}^2(1)\right)^{1/2}$ assuming an MA(1) model for the data

appear to be compatible with the model

$$X_t = \mu + Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right). \tag{3.2.16}$$

The mean μ may be estimated by the sample mean $\bar{x}_{57} = -4.035$, and the parameters θ, σ^2 may be estimated by equating the sample ACVF with the model ACVF at lags 0 and 1, and solving the resulting equations for θ and σ^2. This estimation procedure is known as the method of moments, and in this case gives the equations

$$(1 + \theta^2)\sigma^2 = \hat{\gamma}(0) = 3415.72,$$
$$\theta\sigma^2 = \hat{\gamma}(1) = -1719.95.$$

Using the approximate solution $\theta = -1$ and $\sigma^2 = 1708$, we obtain the noninvertible MA(1) model

$$X_t = -4.035 + Z_t - Z_{t-1}, \quad \{Z_t\} \sim \text{WN}(0, 1708).$$

Typically, in time series modeling we have little or no knowledge of the underlying physical mechanism generating the data, and the choice of a suitable class of models is entirely data driven. For the time series of overshorts, the data, through the graph of the ACF, lead us to the MA(1) model. Alternatively, we can attempt to model the mechanism generating the time series of overshorts using a *structural model*. As we will see, the structural model formulation leads us again to the MA(1) model. In the structural model setup, write Y_t, the observed amount of fuel in the tank at time t, as

$$Y_t = y_t^* + U_t, \tag{3.2.17}$$

where y_t^* is the true (or actual) amount of fuel in the tank at time t (not to be confused with y_t above) and U_t is the resulting *measurement error*. The variable y_t^* is an idealized quantity that in principle cannot be observed even with the most sophisticated measurement devices. Similarly, we assume that

$$A_t = a_t^* + V_t, \tag{3.2.18}$$

where a_t^* is the actual amount of fuel sold minus the actual amount delivered during day t, and V_t is the associated measurement error. We further assume that $\{U_t\} \sim$

$\text{WN}\left(0, \sigma_U^2\right)$, $\{V_t\} \sim \text{WN}\left(0, \sigma_V^2\right)$, and that the two sequences $\{U_t\}$ and $\{V_t\}$ are uncorrelated with one another ($E(U_t V_s) = 0$ for all s and t). If the change of level per day due to leakage is μ gallons ($\mu < 0$ indicates leakage), then

$$y_t^* = \mu + y_{t-1}^* - a_t^*. \tag{3.2.19}$$

This equation relates the actual amounts of fuel in the tank at the end of days t and $t-1$, adjusted for the actual amounts that have been sold and delivered during the day. Using (3.2.17)–(3.2.19), the model for the time series of overshorts is given by

$$X_t = Y_t - Y_{t-1} + A_t = \mu + U_t - U_{t-1} + V_t.$$

This model is stationary and 1-correlated, since

$$EX_t = E(\mu + U_t - U_{t-1} + V_t) = \mu$$

and

$$\begin{aligned}\gamma(h) &= E[(X_{t+h} - \mu)(X_t - \mu)] \\ &= E[(U_{t+h} - U_{t+h-1} + V_{t+h})(U_t - U_{t-1} + V_t)] \\ &= \begin{cases} 2\sigma_U^2 + \sigma_V^2, & \text{if } h = 0, \\ -\sigma_U^2, & \text{if } |h| = 1, \\ 0, & \text{otherwise.} \end{cases}\end{aligned}$$

It follows from Proposition 2.1.1 that $\{X_t\}$ is the MA(1) model (3.2.16) with

$$\rho(1) = \frac{\theta_1}{1 + \theta_1^2} = \frac{-\sigma_U^2}{2\sigma_U^2 + \sigma_V^2}.$$

From this equation we see that the measurement error associated with the adjustment $\{A_t\}$ is zero (i.e., $\sigma_V^2 = 0$) if and only if $\rho(1) = -0.5$ or, equivalently, if and only if $\theta_1 = -1$. From the analysis above, the moment estimator of θ_1 for the overshort data is in fact -1, so that we conclude that there is relatively little measurement error associated with the amount of fuel sold and delivered.

We shall return to a more general discussion of structural models in Chapter 8.

□

Example 3.2.9 The Sunspot Numbers

Figure 3-7 shows the sample PACF of the sunspot numbers $S_1, \ldots, S_{100}$ (for the years 1770–1869) as obtained from ITSM by opening the project SUNSPOTS.TSM and clicking on the second yellow button at the top of the screen. The graph also shows the bounds $\pm 1.96/\sqrt{100}$. The fact that all of the PACF values beyond lag 2 fall within the bounds suggests the possible suitability of an AR(2) model for the mean-corrected data set $X_t = S_t - 46.93$. One simple way to estimate the parameters ϕ_1, ϕ_2, and σ^2 of such a model is to require that the ACVF of the model at lags 0, 1, and 2 should match the sample ACVF at those lags. Substituting the sample ACVF values

$$\hat{\gamma}(0) = 1382.2, \quad \hat{\gamma}(1) = 1114.4, \quad \hat{\gamma}(2) = 591.73,$$

for $\gamma(0)$, $\gamma(1)$, and $\gamma(2)$ in the first three equations of (3.2.5) and (3.2.6) and solving for ϕ_1, ϕ_2, and σ^2 gives the fitted model

$$X_t - 1.318X_{t-1} + 0.634X_{t-2} = Z_t, \quad \{Z_t\} \sim \text{WN}(0, 289.2). \tag{3.2.20}$$

(This method of model fitting is called Yule–Walker estimation and will be discussed more fully in Section 5.1.1.)

□

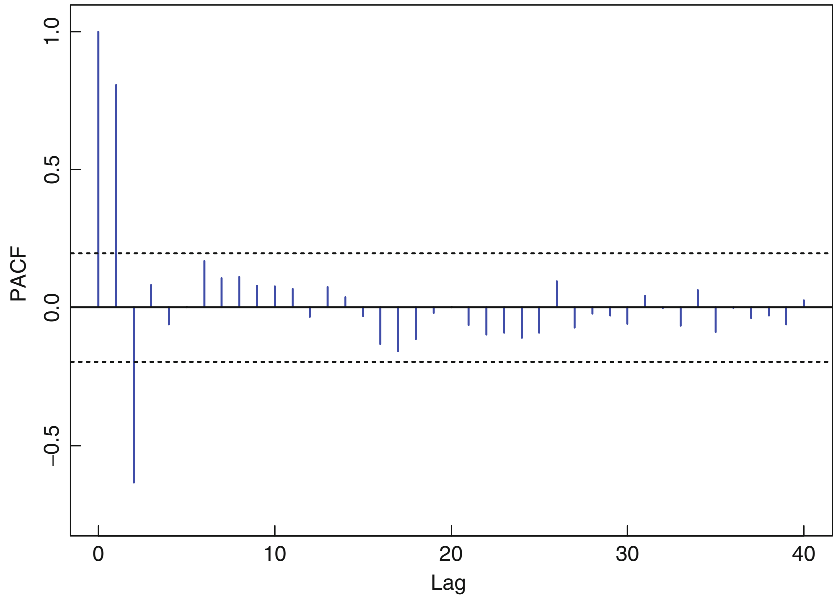

Figure 3-7 The sample PACF of the sunspot numbers with the bounds $\pm 1.96/\sqrt{100}$

3.3 Forecasting ARMA Processes

The innovations algorithm (see Section 2.5.4) provided us with a recursive method for forecasting second-order zero-mean processes that are not necessarily stationary. For the causal ARMA process

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right),$$

it is possible to simplify the application of the algorithm drastically. The idea is to apply it not to the process $\{X_t\}$ itself, but to the transformed process [cf. Ansley (1979)]

$$\begin{cases} W_t = \sigma^{-1}X_t, & t = 1, \ldots, m, \\ W_t = \sigma^{-1}\phi(B)X_t, & t > m, \end{cases} \tag{3.3.1}$$

where

$$m = \max(p, q). \tag{3.3.2}$$

For notational convenience we define $\theta_0 := 1$ and $\theta_j := 0$ for $j > q$. We shall also assume that $p \geq 1$ and $q \geq 1$. (There is no loss of generality in these assumptions, since in the analysis that follows we may take any of the coefficients ϕ_i and θ_i to be zero.)

The autocovariance function $\gamma_X(\cdot)$ of $\{X_t\}$ can easily be computed using any of the methods described in Section 3.2.1. The autocovariances $\kappa(i, j) = E(W_iW_j)$, $i, j \geq 1$, are then found from

$$\kappa(i, j) = \begin{cases} \sigma^{-2}\gamma_X(i-j), & 1 \leq i, j \leq m \\ \sigma^{-2}\left[\gamma_X(i-j) - \sum_{r=1}^{p}\phi_r\gamma_X(r - |i-j|)\right], & \min(i, j) \leq m < \max(i, j) \leq 2m, \\ \sum_{r=0}^{q}\theta_r\theta_{r+|i-j|}, & \min(i, j) > m, \\ 0, & \text{otherwise.} \end{cases} \tag{3.3.3}$$

Applying the innovations algorithm to the process $\{W_t\}$ we obtain

$$\begin{cases} \hat{W}_{n+1} = \sum_{j=1}^{n} \theta_{nj}(W_{n+1-j} - \hat{W}_{n+1-j}), & 1 \le n < m, \\ \hat{W}_{n+1} = \sum_{j=1}^{q} \theta_{nj}(W_{n+1-j} - \hat{W}_{n+1-j}), & n \ge m, \end{cases} \tag{3.3.4}$$

where the coefficients θ_{nj} and the mean squared errors $r_n = E\left(W_{n+1} - \hat{W}_{n+1}\right)^2$ are found recursively from the innovations algorithm with κ defined as in (3.3.3). The notable feature of the predictors (3.3.4) is the vanishing of θ_{nj} when both $n \ge m$ and $j > q$. This is a consequence of the innovations algorithm and the fact that $\kappa(r, s) = 0$ if $r > m$ and $|r - s| > q$.

Observe now that the equations (3.3.1) allow each X_n, $n \ge 1$, to be written as a linear combination of W_j, $1 \le j \le n$, and, conversely, each W_n, $n \ge 1$, to be written as a linear combination of X_j, $1 \le j \le n$. This means that the best linear predictor of any random variable Y in terms of $\{1, X_1, \ldots, X_n\}$ is the same as the best linear predictor of Y in terms of $\{1, W_1, \ldots, W_n\}$. We shall denote this predictor by P_nY. In particular, the one-step predictors of W_{n+1} and X_{n+1} are given by

$$\hat{W}_{n+1} = P_n W_{n+1}$$

and

$$\hat{X}_{n+1} = P_n X_{n+1}.$$

Using the linearity of P_n and the equations (3.3.1) we see that

$$\begin{cases} \hat{W}_t = \sigma^{-1}\hat{X}_t, & t = 1, \ldots, m, \\ \hat{W}_t = \sigma^{-1}\left[\hat{X}_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}\right], & t > m, \end{cases} \tag{3.3.5}$$

which, together with (3.3.1), shows that

$$X_t - \hat{X}_t = \sigma\left[W_t - \hat{W}_t\right] \quad \text{for all } t \ge 1. \tag{3.3.6}$$

Replacing $\left(W_j - \hat{W}_j\right)$ by $\sigma^{-1}\left(X_j - \hat{X}_j\right)$ in (3.3.3) and then substituting into (3.3.4), we finally obtain

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right), & 1 \le n < m, \\ \phi_1 X_n + \cdots + \phi_p X_{n+1-p} + \sum_{j=1}^{q} \theta_{nj}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right), & n \ge m, \end{cases} \tag{3.3.7}$$

and

$$E\left(X_{n+1} - \hat{X}_{n+1}\right)^2 = \sigma^2 E\left(W_{n+1} - \hat{W}_{n+1}\right)^2 = \sigma^2 r_n, \tag{3.3.8}$$

where θ_{nj} and r_n are found from the innovations algorithm with κ as in (3.3.3). Equations (3.3.7) determine the one-step predictors $\hat{X}_2, \hat{X}_3, \ldots$ recursively.

Remark 1. It can be shown (see Brockwell and Davis (1991), Problem 5.6) that if $\{X_t\}$ is invertible, then as $n \to \infty$,

$$E\left(X_n - \hat{X}_n - Z_n\right)^2 \to 0,$$

$$\theta_{nj} \to \theta_j, \; j = 1, \ldots, q,$$

and

$$r_n \to 1.$$

Algebraic calculation of the coefficients θ_{nj} and r_n is not feasible except for very simple models, such as those considered in the following examples. However, numerical implementation of the recursions is quite straightforward and is used to compute predictors in the program ITSM. □

Example 3.3.1 Prediction of an AR(p) Process

Applying (3.3.7) to the ARMA(p, 0) process, we see at once that

$$\hat{X}_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p}, \quad n \geq p.$$

□

Example 3.3.2 Prediction of an MA(q) Process

Applying (3.3.7) to the ARMA(1, q) process with $\phi_1 = 0$ gives

$$\hat{X}_{n+1} = \sum_{j=1}^{\min(n,q)} \theta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right), \quad n \geq 1,$$

where the coefficients θ_{nj} are found by applying the innovations algorithm to the covariances $\kappa(i, j)$ defined in (3.3.3). Since in this case the processes $\{X_t\}$ and $\{\sigma^{-1} W_t\}$ are identical, these covariances are simply

$$\kappa(i, j) = \sigma^{-2} \gamma_X(i-j) = \sum_{r=0}^{q-|i-j|} \theta_r \theta_{r+|i-j|}.$$

□

Example 3.3.3 Prediction of an ARMA(1,1) Process

If

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

and $|\phi| < 1$, then equations (3.3.7) reduce to the single equation

$$\hat{X}_{n+1} = \phi X_n + \theta_{n1}(X_n - \hat{X}_n), \quad n \geq 1.$$

To compute θ_{n1} we use Example 3.2.1 to obtain $\gamma_X(0) = \sigma^2(1 + 2\theta\phi + \theta^2)/\left(1 - \phi^2\right)$. Substituting in (3.3.3) then gives, for $i, j \geq 1$,

$$\kappa(i, j) = \begin{cases} \left(1 + 2\theta\phi + \theta^2\right) / \left(1 - \phi^2\right), & i = j = 1, \\ 1 + \theta^2, & i = j \geq 2, \\ \theta, & |i - j| = 1, i \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

With these values of $\kappa(i, j)$, the recursions of the innovations algorithm reduce to

$$\begin{aligned} r_0 &= \left(1 + 2\theta\phi + \theta^2\right)/\left(1 - \phi^2\right), \\ \theta_{n1} &= \theta / r_{n-1}, \\ r_n &= 1 + \theta^2 - \theta^2 / r_{n-1}, \end{aligned} \tag{3.3.9}$$

which can be solved quite explicitly (see Problem 3.13). □

Example 3.3.4 Numerical Prediction of an ARMA(2,3) Process

In this example we illustrate the steps involved in numerical prediction of an ARMA(2,3) process. Of course, these steps are shown for illustration only. The calculations are all carried out automatically by ITSM in the course of computing predictors for any specified data set and ARMA model. The process we shall consider is the ARMA process defined by the equations

$$X_t - X_{t-1} + 0.24X_{t-2} = Z_t + 0.4Z_{t-1} + 0.2Z_{t-2} + 0.1Z_{t-3}, \tag{3.3.10}$$

where $\{Z_t\} \sim \mathrm{WN}(0, 1)$. Ten values of $X_1, \ldots, X_{10}$ simulated by the program ITSM are shown in Table 3.1. (These were produced using the option `Model>Specify` to specify the order and parameters of the model and then `Model>Simulate` to generate the series from the specified model.)

The first step is to compute the covariances $\gamma_X(h)$, $h = 0, 1, 2$, which are easily found from equations (3.2.5) with $k = 0, 1, 2$ to be

$$\gamma_X(0) = 7.17133, \quad \gamma_X(1) = 6.44139, \quad \text{and} \quad \gamma_X(2) = 5.0603.$$

From (3.3.3) we find that the symmetric matrix $K = [\kappa(i, j)]_{i,j=1,2,\ldots}$ is given by

$$K = \begin{bmatrix} 7.1713 & & & & & & & & \\ 6.4414 & 7.1713 & & & & & & & \\ 5.0603 & 6.4414 & 7.1713 & & & & & & \\ 0.10 & 0.34 & 0.816 & 1.21 & & & & & \\ 0 & 0.10 & 0.34 & 0.50 & 1.21 & & & & \\ 0 & 0 & 0.10 & 0.24 & 0.50 & 1.21 & & & \\ \cdot & 0 & 0 & 0.10 & 0.24 & 0.50 & 1.21 & & \\ \cdot & \cdot & 0 & 0 & 0.10 & 0.24 & 0.50 & 1.21 & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

The next step is to solve the recursions of the innovations algorithm for θ_{nj} and r_n using these values for $\kappa(i, j)$. Then

Table 3.1 $\hat{X}_{n+1}$ for the ARMA(2,3) Process of Example 3.3.4

n	X_{n+1}	r_n	θ_{n1}	θ_{n2}	θ_{n3}	$\hat{X}_{n+1}$
0	1.704	7.1713	0			
1	0.527	1.3856	0.8982	1.5305		
2	1.041	1.0057	1.3685	0.7056	−0.1710	
3	0.942	1.0019	0.4008	0.1806	0.0139	1.2428
4	0.555	1.0019	0.3998	0.2020	0.0732	0.7443
5	−1.002	1.0005	0.3992	0.1995	0.0994	0.3138
6	−0.585	1.0000	0.4000	0.1997	0.0998	−1.7293
7	0.010	1.0000	0.4000	0.2000	0.0998	−0.1688
8	−0.638	1.0000	0.4000	0.2000	0.0999	0.3193
9	0.525	1.0000	0.4000	0.2000	0.1000	−0.8731
10		1.0000	0.4000	0.2000	0.1000	1.0638
11		1.0000	0.4000	0.2000	0.1000	
12		1.0000	0.4000	0.2000	0.1000	

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right), & n = 1, 2, \\ X_n - 0.24 X_{n-1} + \sum_{j=1}^{3} \theta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right), & n = 3, 4, \ldots, \end{cases}$$

and

$$E \left(X_{n+1} - \hat{X}_{n+1} \right)^2 = \sigma^2 r_n = r_n.$$

The results are shown in Table 3.1.

□

3.3.1 *h*-Step Prediction of an ARMA(*p*, *q*) Process

As in Section 2.5, we use $P_n Y$ to denote the best linear predictor of Y in terms of $X_1, \ldots, X_n$ (which, as pointed out after (3.3.4), is the same as the best linear predictor of Y in terms of $W_1, \ldots, W_n$). Then from (2.5.30) we have

$$P_n W_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left(W_{n+h-j} - \hat{W}_{n+h-j} \right) = \sigma^2 \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left(X_{n+h-j} - \hat{X}_{n+h-j} \right).$$

Using this result and applying the operator P_n to each side of equation (3.3.1), we conclude that the h-step predictors $P_n X_{n+h}$ satisfy

$$P_n X_{n+h} = \begin{cases} \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left(X_{n+h-j} - \hat{X}_{n+h-j} \right), & 1 \le h \le m - n, \\ \sum_{i=1}^{p} \phi_i P_n X_{n+h-i} + \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left(X_{n+h-j} - \hat{X}_{n+h-j} \right), & h > m - n. \end{cases} \tag{3.3.11}$$

If, as is almost always the case, $n > m = \max(p, q)$, then for all $h \ge 1$,

$$P_n X_{n+h} = \sum_{i=1}^{p} \phi_i P_n X_{n+h-i} + \sum_{j=h}^{q} \theta_{n+h-1,j} \left(X_{n+h-j} - \hat{X}_{n+h-j} \right). \tag{3.3.12}$$

Once the predictors $\hat{X}_1, \ldots \hat{X}_n$ have been computed from (3.3.7), it is a straightforward calculation, with n fixed, to determine the predictors $P_n X_{n+1}, P_n X_{n+2}, P_n X_{n+3}, \ldots$ recursively from (3.3.12) (or (3.3.11) if $n \le m$). The calculations are performed automatically in the `Forecasting>ARMA` option of the program ITSM.

Example 3.3.5 *h*-Step Prediction of an ARMA(2,3) Process

To compute h-step predictors, $h = 1, \ldots, 10$, for the data of Example 3.3.4 and the model (3.3.10), open the project E334.TSM in ITSM and enter the model using the option `Model>Specify`. Then select `Forecasting>ARMA` and specify 10 for the number of forecasts required. You will notice that the white noise variance is automatically set by ITSM to an estimate based on the sample. To retain the model value of 1, you must reset the white noise variance to this value. Then click `OK` and you will see a graph of the original series with the ten predicted values appended. If you right-click on the graph and select `Info`, you will see the numerical results shown in the following table as well as prediction bounds based on the assumption that the series is Gaussian. (Prediction bounds are discussed in the last paragraph of this chapter.) The mean squared errors are calculated as described below. Notice how the predictors converge fairly rapidly to the mean of the process (i.e., zero) as the lead time h increases. Correspondingly, the one-step mean squared error increases from the white noise variance (i.e., 1) at $h = 1$ to the variance of X_t (i.e., 7.1713), which is virtually reached at $h = 10$.

□

The Mean Squared Error of $P_n X_{n+h}$

The mean squared error of $P_n X_{n+h}$ is easily computed by ITSM from the formula

$$\sigma_n^2(h) := E(X_{n+h} - P_n X_{n+h})^2 = \sum_{j=0}^{h-1} \left(\sum_{r=0}^{j} \chi_r \theta_{n+h-r-1,\,j-r} \right)^2 v_{n+h-j-1}, \tag{3.3.13}$$

where the coefficients χ_j are computed recursively from the equations $\chi_0 = 1$ and

$$\chi_j = \sum_{k=1}^{\min(p,j)} \phi_k \chi_{j-k}, \quad j = 1, 2, \ldots. \tag{3.3.14}$$

Example 3.3.6 *h*-Step Prediction of an ARMA(2,3) Process

We now illustrate the use of (3.3.12) and (3.3.13) for the h-step predictors and their mean squared errors by manually reproducing the output of ITSM shown in Table 3.2. From (3.3.12) and Table 3.1 we obtain

$$\begin{aligned} P_{10} X_{12} &= \sum_{i=1}^{2} \phi_i P_{10} X_{12-i} + \sum_{j=2}^{3} \theta_{11,j} \left(X_{12-j} - \hat{X}_{12-j} \right) \\ &= \phi_1 \hat{X}_{11} + \phi_2 X_{10} + 0.2 \left(X_{10} - \hat{X}_{10} \right) + 0.1 \left(X_9 - \hat{X}_9 \right) \\ &= 1.1217 \end{aligned}$$

and

Table 3.2 *h*-step predictors for the ARMA(2,3) Series of Example 3.3.4

h	$P_{10}X_{10+h}$	$\sqrt{MSE}$
1	1.0638	1.0000
2	1.1217	1.7205
3	1.0062	2.1931
4	0.7370	2.4643
5	0.4955	2.5902
6	0.3186	2.6434
7	0.1997	2.6648
8	0.1232	2.6730
9	0.0753	2.6761
10	0.0457	2.6773

$$\begin{aligned} P_{10}X_{13} &= \sum_{i=1}^{2} \phi_i P_{10}X_{13-i} + \sum_{j=3}^{3} \theta_{12,j} \left(X_{13-j} - \hat{X}_{13-j}\right) \\ &= \phi_1 P_{10}X_{12} + \phi_2 \hat{X}_{11} + 0.1 \left(X_{10} - \hat{X}_{10}\right) \\ &= 1.0062. \end{aligned}$$

For $k > 13$, $P_{10}X_k$ is easily found recursively from

$$P_{10}X_k = \phi_1 P_{10}X_{k-1} + \phi_2 P_{10}X_{k-2}.$$

To find the mean squared errors we use (3.3.13) with $\chi_0 = 1$, $\chi_1 = \phi_1 = 1$, and $\chi_2 = \phi_1\chi_1 + \phi_2 = 0.76$. Using the values of θ_{nj} and $v_j(= r_j)$ in Table 3.1, we obtain

$$\sigma_{10}^2(2) = E(X_{12} - P_{10}X_{12})^2 = 2.960$$

and

$$\sigma_{10}^2(3) = E(X_{13} - P_{10}X_{13})^2 = 4.810,$$

in accordance with the results shown in Table 3.2.

□

Large-Sample Approximations

Assuming as usual that the ARMA(p, q) process defined by $\phi(B)X_t = \theta(B)Z_t$, $\{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right)$, is causal and invertible, we have the representations

$$X_{n+h} = \sum_{j=0}^{\infty} \psi_j Z_{n+h-j} \tag{3.3.15}$$

and

$$Z_{n+h} = X_{n+h} + \sum_{j=1}^{\infty} \pi_j X_{n+h-j}, \tag{3.3.16}$$

where $\{\psi_j\}$ and $\{\pi_j\}$ are uniquely determined by equations (3.1.7) and (3.1.8), respectively. Let $\tilde{P}_n Y$ denote the best (i.e., minimum mean squared error) approximation to Y that is a linear combination or limit of linear combinations of X_t, $-\infty < t \le n$, or equivalently [by (3.3.15) and (3.3.16)] of Z_t, $-\infty < t \le n$. The properties of the operator $\tilde{P}_n$ were discussed in Section 2.5.6. Applying $\tilde{P}_n$ to each side of equations (3.3.15) and (3.3.16) gives

$$\tilde{P}_n X_{n+h} = \sum_{j=h}^{\infty} \psi_j Z_{n+h-j} \tag{3.3.17}$$

and

$$\tilde{P}_n X_{n+h} = -\sum_{j=1}^{\infty} \pi_j \tilde{P}_n X_{n+h-j}. \tag{3.3.18}$$

For $h = 1$ the jth term on the right of (3.3.18) is just X_{n+1-j}. Once $\tilde{P}_n X_{n+1}$ has been evaluated, $\tilde{P}_n X_{n+2}$ can then be computed from (3.3.18). The predictors $\tilde{P}_n X_{n+3}$, $\tilde{P}_n X_{n+4}, \ldots$ can then be computed successively in the same way. Subtracting (3.3.17) from (3.3.15) gives the h-step prediction error as

$$X_{n+h} - \tilde{P}_n X_{n+h} = \sum_{j=0}^{h-1} \psi_j Z_{n+h-j},$$

from which we see that the mean squared error is

$$\tilde{\sigma}^2(h) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2. \tag{3.3.19}$$

The predictors obtained in this way have the form

$$\tilde{P}_n X_{n+h} = \sum_{j=0}^{\infty} c_j X_{n-j}. \tag{3.3.20}$$

In practice, of course, we have only observations $X_1, \ldots, X_n$ available, so we must truncate the series (3.3.20) after n terms. The resulting predictor is a useful approximation to $P_n X_{n+h}$ if n is large and the coefficients c_j converge to zero rapidly as j increases. It can be shown that the mean squared error (3.3.19) of $\tilde{P}_n X_{n+h}$ can also be obtained by letting $n \to \infty$ in the expression (3.3.13) for the mean squared error of $P_n X_{n+h}$, so that $\tilde{\sigma}^2(h)$ is an easily calculated approximation to $\sigma_n^2(h)$ for large n.

Prediction Bounds for Gaussian Processes

If the ARMA process $\{X_t\}$ is driven by Gaussian white noise (i.e., if $\{Z_t\} \sim$ IID N$\left(0, \sigma^2\right)$), then for each $h \geq 1$ the prediction error $X_{n+h} - P_n X_{n+h}$ is normally distributed with mean 0 and variance $\sigma_n^2(h)$ given by (3.3.19).

Consequently, if $\Phi_{1-\alpha/2}$ denotes the $(1-\alpha/2)$ quantile of the standard normal distribution function, it follows that X_{n+h} lies between the bounds $P_n X_{n+h} \pm \Phi_{1-\alpha/2} \sigma_n(h)$ with probability $(1-\alpha)$. These bounds are therefore called $(1-\alpha)$ prediction bounds for X_{n+h}.

Problems

3.1 Determine which of the following ARMA processes are causal and which of them are invertible. (In each case $\{Z_t\}$ denotes white noise.)

(a) $X_t + 0.2X_{t-1} - 0.48X_{t-2} = Z_t$.

(b) $X_t + 1.9X_{t-1} + 0.88X_{t-2} = Z_t + 0.2Z_{t-1} + 0.7Z_{t-2}$.

(c) $X_t + 0.6X_{t-1} = Z_t + 1.2Z_{t-1}$.

(d) $X_t + 1.8X_{t-1} + 0.81X_{t-2} = Z_t$.

(e) $X_t + 1.6X_{t-1} = Z_t - 0.4Z_{t-1} + 0.04Z_{t-2}$.

3.2 For those processes in Problem 3.1 that are causal, compute and graph their ACF and PACF using the program ITSM.

3.3 For those processes in Problem 3.1 that are causal, compute the first six coefficients $\psi_0, \psi_1, \ldots, \psi_5$ in the causal representation $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ of $\{X_t\}$.

3.4 Compute the ACF and PACF of the AR(2) process

$$X_t = 0.8X_{t-2} + Z_t, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right).$$

3.5 Let $\{Y_t\}$ be the ARMA plus noise time series defined by

$$Y_t = X_t + W_t,$$

where $\{W_t\} \sim \text{WN}\left(0, \sigma_w^2\right)$, $\{X_t\}$ is the ARMA(p, q) process satisfying

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma_z^2\right),$$

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

(a) Show that $\{Y_t\}$ is stationary and find its autocovariance function in terms of σ_W^2 and the ACVF of $\{X_t\}$.

(b) Show that the process $U_t := \phi(B)Y_t$ is r-correlated, where $r = \max(p, q)$ and hence, by Proposition 2.1.1, is an MA(r) process. Conclude that $\{Y_t\}$ is an ARMA(p, r) process.

3.6 Show that the two MA(1) processes

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

$$Y_t = \tilde{Z}_t + \frac{1}{\theta}\tilde{Z}_{t-1}, \quad \{\tilde{Z}_t\} \sim \text{WN}\left(0, \sigma^2\theta^2\right),$$

where $0 < |\theta| < 1$, have the same autocovariance functions.

3.7 Suppose that $\{X_t\}$ is the noninvertible MA(1) process

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

where $|\theta| > 1$. Define a new process $\{W_t\}$ as

$$W_t = \sum_{j=0}^{\infty}(-\theta)^{-j}X_{t-j}$$

and show that $\{W_t\} \sim \text{WN}\left(0, \sigma_W^2\right)$. Express σ_W^2 in terms of θ and σ^2 and show that $\{X_t\}$ has the *invertible* representation (in terms of $\{W_t\}$)

$$X_t = W_t + \frac{1}{\theta}W_{t-1}.$$

3.8 Let $\{X_t\}$ denote the unique stationary solution of the autoregressive equations

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

where $\{Z_t\} \sim \text{WN}\left(0, \sigma^2\right)$ and $|\phi| > 1$. Then X_t is given by the expression (2.2.11). Define the new sequence

$$W_t = X_t - \frac{1}{\phi}X_{t-1},$$

show that $\{W_t\} \sim \text{WN}\left(0, \sigma_W^2\right)$, and express σ_W^2 in terms of σ^2 and ϕ. These calculations show that $\{X_t\}$ is the (unique stationary) solution of the *causal* AR equations

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

3.9 (a) Calculate the autocovariance function $\gamma(\cdot)$ of the stationary time series

$$Y_t = \mu + Z_t + \theta_1 Z_{t-1} + \theta_{12} Z_{t-12}, \quad \{Z_t\} \sim \text{WN}\left(0, \sigma^2\right).$$

(b) Use the program ITSM to compute the sample mean and sample autocovariances $\hat{\gamma}(h)$, $0 \le h \le 20$, of $\{\nabla\nabla_{12}X_t\}$, where $\{X_t, t = 1, \ldots, 72\}$ is the accidental deaths series DEATHS.TSM of Example 1.1.3.

(c) By equating $\hat{\gamma}(1)$, $\hat{\gamma}(11)$, and $\hat{\gamma}(12)$ from part (b) to $\gamma(1)$, $\gamma(11)$, and $\gamma(12)$, respectively, from part (a), find a model of the form defined in (a) to represent $\{\nabla\nabla_{12}X_t\}$.

3.10 By matching the autocovariances and sample autocovariances at lags 0 and 1, fit a model of the form

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t, \quad \{Z_t\} \sim \mathrm{WN}\big(0, \sigma^2\big),$$

to the data STRIKES.TSM of Example 1.1.6. Use the fitted model to compute the best predictor of the number of strikes in 1981. Estimate the mean squared error of your predictor and construct 95 % prediction bounds for the number of strikes in 1981 assuming that $\{Z_t\} \sim$ iid $\mathrm{N}\big(0, \sigma^2\big)$.

3.11 Show that the value at lag 2 of the partial ACF of the MA(1) process

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

where $\{Z_t\} \sim \mathrm{WN}\big(0, \sigma^2\big)$, is

$$\alpha(2) = -\theta^2 / \left(1 + \theta^2 + \theta^4\right).$$

3.12 For the MA(1) process of Problem 3.11, the best linear predictor of X_{n+1} based on $X_1, \ldots, X_n$ is

$$\hat{X}_{n+1} = \phi_{n,1}X_n + \cdots + \phi_{n,n}X_1,$$

where $\phi_n = \big(\phi_{n1}, \ldots, \phi_{nn}\big)'$ satisfies $R_n\phi_n = \rho_n$ [equation (2.5.23)]. By substituting the appropriate correlations into R_n and ρ_n and solving the resulting equations (starting with the last and working up), show that for $1 \le j < n$, $\phi_{n,n-j} = (-\theta)^{-j}\big(1 + \theta^2 + \cdots + \theta^{2j}\big)\phi_{nn}$ and hence that the PACF $\alpha(n) := \phi_{nn} = -(-\theta)^n/\big(1 + \theta^2 + \cdots + \theta^{2n}\big)$.

3.13 The coefficients θ_{nj} and one-step mean squared errors $v_n = r_n\sigma^2$ for the general causal ARMA(1,1) process in Example 3.3.3 can be found as follows:

(a) Show that if $y_n := r_n/(r_n - 1)$, then the last of equation (3.3.9) can be rewritten in the form

$$y_n = \theta^{-2}y_{n-1} + 1, \quad n \ge 1.$$

(b) Deduce that $y_n = \theta^{-2n}y_0 + \sum_{j=1}^{n} \theta^{-2(j-1)}$ and hence determine r_n and θ_{n1}, $n = 1, 2, \ldots$.

(c) Evaluate the limits as $n \to \infty$ of r_n and θ_{n1} in the two cases $|\theta| < 1$ and $|\theta| \ge 1$.