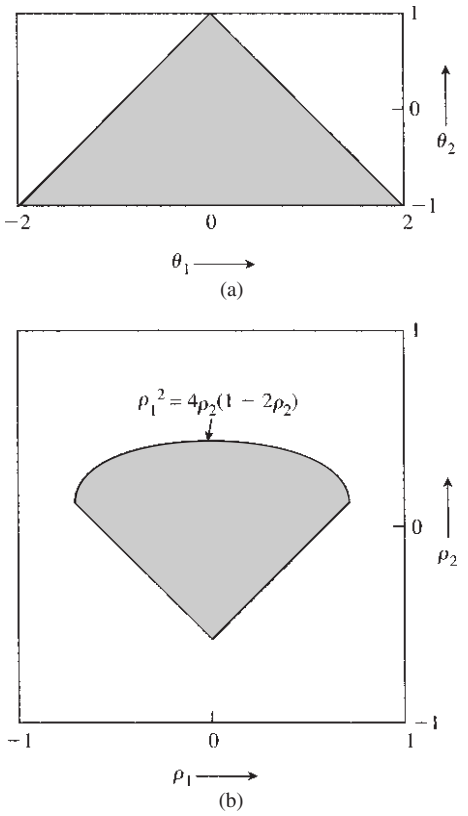

FIGURE 3.6 Admissible regions for (a) θ_1, θ_2 and (b) ρ_1, ρ_2 for an invertible MA(2) process.

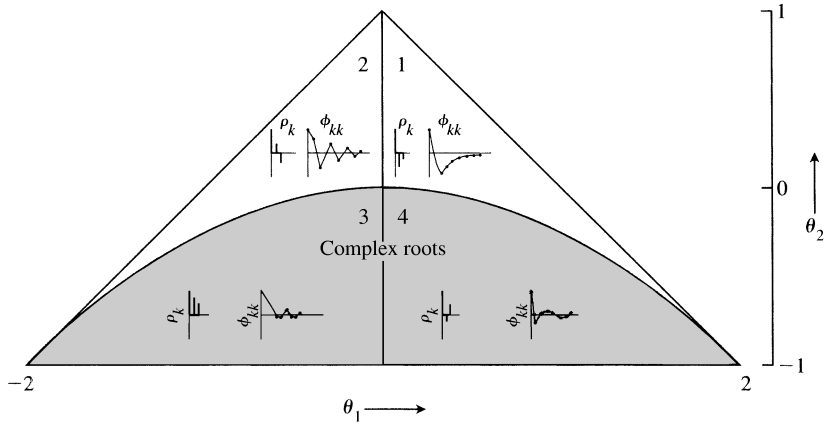

FIGURE 3.7 Autocorrelation and partial autocorrelation functions ρ_k and ϕ_{kk} for various MA(2) models.

autocorrelations and partial autocorrelations for an AR(2) process, illustrates the duality between the MA(2) and the AR(2) processes.

Example. For illustration, consider the second-order moving average model

$$\tilde{z}_t = a_t - 0.8a_{t-1} + 0.5a_{t-2}$$

The variance of the process is $\gamma_0 = \sigma_a^2(1 + (0.8)^2 + (-0.5)^2) = 1.89\sigma_a^2$, and from (3.3.11) the theoretical autocorrelations are

$$\rho_1 = \frac{-0.8(1-(-0.5))}{1+(0.8)^2+(-0.5)^2} = \frac{-1.20}{1.89} = -0.635 \qquad \rho_2 = \frac{-(-0.5)}{1.89} = 0.265$$

and $\rho_k = 0$, for $k > 2$. The theoretical partial autocorrelations are obtained by solving (3.2.31) successively; the first several values are $\phi_{11} = \rho_1 = -0.635$, $\phi_{22} = (\rho_2 - \rho_1^2)/(1 - \rho_1^2) = -0.232$, $\phi_{33} = 0.105$, $\phi_{44} = 0.191$, and $\phi_{55} = 0.102$.

Figure 3.8 shows the autocorrelation and partial autocorrelation functions up to 15 lags for this example. Note the partial autocorrelations ϕ_{kk} display an approximate damped sinusoidal behavior with moderate rate of damping, similar to the behavior depicted for region 4 in Figure 3.7. This is consistent with the fact that the roots of $\theta(B) = 0$ are complex with modulus (damping factor) $D = \sqrt{0.5} \simeq 0.71$ and frequency $f_0 = \cos^{-1}(0.5657)/(2\pi) = 1/6.48$ in this example.

The autocorrelation and partial autocorrelation functions shown in Figure 3.8 were generated using the function ARMAacf() in the R stats package. The commands needed to reproduce the graph are shown below. Note that the moving average parameters in the ARMAacf() function are again entered with their signs reversed since R uses positive signs in defining the moving average operator, rather than the negative signs used here.

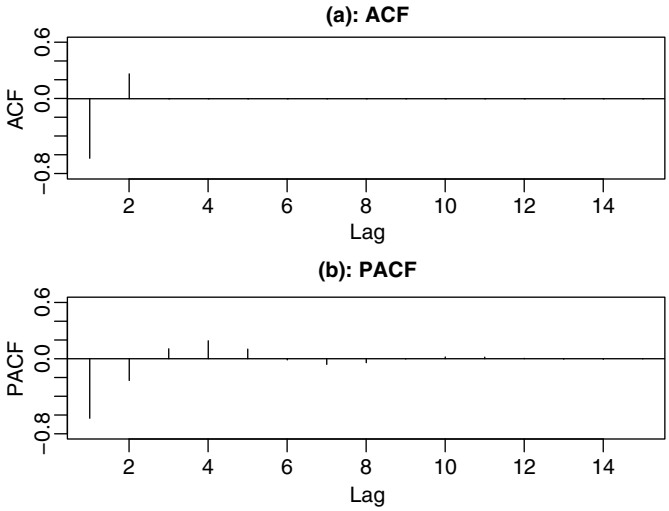

FIGURE 3.8 (a) Autocorrelation function and (b) partial autocorrelation function for the MA(2) model $\tilde{z}_t = a_t - 0.8a_{t-1} + 0.5a_{t-2}$.

```
> ACF=ARMAacf(ar=0,ma=c(-0.8,+0.5),lag.max=15,pacf=FALSE)[-1]
> PACF=ARMAacf(ar=0,ma=c(-0.8,+0.5),lag.max=15,pacf=TRUE)
> par(mfrow=c(2,1))
> plot(ACF,type='h',ylim=c(-0.8,0.6),xlab='lag',main='(a): ACF')
> abline(h=0)
> plot(PACF,type='h',ylim=c(-0.8,0.6),xlab='lag',main='(b):PACF')
> abline(h=0)
> ACF    % Retrieves the autocorrelation coefficients
> PACF   % Retrieves the partial autocorrelation coefficients
```

3.3.5 Duality Between Autoregressive and Moving Average Processes

The previous sections have examined the properties of autoregressive and moving average processes and discussed the *duality* between these processes. As illustrated in Table 3.2 at the end of this chapter, this duality has the following consequences:

1. In a stationary autoregressive process of order p, a_t can be represented as a *finite* weighted sum of previous $\tilde{z}$'s, or $\tilde{z}_t$ as an infinite weighted sum
$$\tilde{z}_t = \phi^{-1}(B)a_t$$
of previous a's. Conversely, an invertible moving average process of order q, $\tilde{z}_t$, can be represented as a finite weighted sum of previous a's, or a_t as an infinite weighted sum
$$\theta^{-1}(B)\tilde{z}_t = a_t$$
of previous $\tilde{z}$'s.
2. The finite MA process has an autocorrelation function that is zero beyond a certain point, but since it is equivalent to an infinite AR process, its partial autocorrelation function is infinite in extent and is dominated by damped exponentials and/or damped sine waves. Conversely, the AR process has a partial autocorrelation function that is zero beyond a certain point, but its autocorrelation function is infinite in extent and consists of a mixture of damped exponentials and/or damped sine waves.
3. For an autoregressive process of finite order p, the parameters are not required to satisfy any conditions to ensure invertibility. However, for stationarity, the roots of $\phi(B) = 0$ must lie outside the unit circle. Conversely, the parameters of the MA process are not required to satisfy any conditions to ensure stationarity. However, for invertibility, the roots of $\theta(B) = 0$ must lie outside the unit circle.
4. The spectrum of a moving average process has an inverse relationship to the spectrum of the corresponding autoregressive process.

3.4 MIXED AUTOREGRESSIVE–MOVING AVERAGE PROCESSES

3.4.1 Stationarity and Invertibility Properties

We have noted earlier that to achieve parsimony it may be necessary to include both autoregressive and moving average terms. Thus, we may need to employ the mixed ARMA model

$$\tilde{z}_t = \phi_1\tilde{z}_{t-1} + \cdots + \phi_p\tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{3.4.1}$$

that is,

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\tilde{z}_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)a_t$$

or

$$\phi(B)\tilde{z}_t = \theta(B)a_t$$

where $\phi(B)$ and $\theta(B)$ are polynomial operators in B of degrees p and q.

We subsequently refer to this process as an ARMA(p, q) process. It may be thought of in two ways:

1. As a pth-order autoregressive process

$$\phi(B)\tilde{z}_t = e_t$$

with e_t following the qth-order moving average process $e_t = \theta(B)a_t$.

2. As a qth-order moving average process

$$\tilde{z}_t = \theta(B)b_t$$

with b_t following the pth-order autoregressive process $\phi(B)b_t = a_t$ so that

$$\phi(B)\tilde{z}_t = \theta(B)\phi(B)b_t = \theta(B)a_t$$

It is obvious that moving average terms on the right of (3.4.1) will not affect the earlier arguments, which establish conditions for stationarity of an autoregressive process. Thus, $\phi(B)\tilde{z}_t = \theta(B)a_t$ will define a stationary process provided that the characteristic equation $\phi(B) = 0$ has all its roots outside the unit circle. Similarly, the roots of $\theta(B) = 0$ must lie outside the unit circle if the process is to be invertible.

Thus, the stationary and invertible ARMA(p, q) process (3.4.1) has both the infinite moving average representation

$$\tilde{z}_t = \psi(B)a_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

where $\psi(B) = \phi^{-1}(B)\theta(B)$, and the infinite autoregressive representation

$$\pi(B)\tilde{z}_t = \tilde{z}_t - \sum_{j=1}^{\infty} \pi_j \tilde{z}_{t-j} = a_t$$

where $\pi(B) = \theta^{-1}(B)\phi(B)$, with both the ψ_j weights and the π_j weights being absolutely summable. The weights ψ_j are determined from the relation $\phi(B)\psi(B) = \theta(B)$ to satisfy

$$\psi_j = \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + \cdots + \phi_p\psi_{j-p} - \theta_j \qquad j > 0$$

with $\psi_0 = 1$, $\psi_j = 0$ for $j < 0$, and $\theta_j = 0$ for $j > q$, while from the relation $\theta(B)\pi(B) = \phi(B)$ the π_j are determined to satisfy

$$\pi_j = \theta_1\pi_{j-1} + \theta_2\pi_{j-2} + \cdots + \theta_q\pi_{j-q} + \phi_j \qquad j > 0$$

with the $\pi_0 = -1$, $\pi_j = 0$ for $j < 0$, and $\phi_j = 0$ for $j > p$. From these relations, the ψ_j and π_j weights can readily be computed recursively in terms of the ϕ_i and θ_i coefficients.

3.4.2 Autocorrelation Function and Spectrum of Mixed Processes

Autocorrelation Function. The autocorrelation function of the mixed process may be derived by a method similar to that used for autoregressive processes in Section 3.2.2. On multiplying throughout in (3.4.1) by $\tilde{z}_{t-k}$ and taking expectations, we see that the autocovariance function satisfies the difference equation

$$\gamma_k = \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p} + \gamma_{za}(k) - \theta_1\gamma_{za}(k-1) - \cdots - \theta_q\gamma_{za}(k-q)$$

where $\gamma_{za}(k)$ is the cross-covariance function between z and a and is defined by $\gamma_{za}(k) = E[\tilde{z}_{t-k}a_t]$. Since z_{t-k} depends only on shocks that have occurred up to time $t-k$ through the infinite moving average representation $\tilde{z}_{t-k} = \psi(B)a_{t-k} = \sum_{j=0}^{\infty}\psi_j a_{t-k-j}$, it follows that

$$\gamma_{za}(k) = \begin{cases} 0 & k > 0 \\ \psi_{-k}\sigma_a^2 & k \le 0 \end{cases}$$

Hence, the preceding equation for γ_k may be expressed as

$$\gamma_k = \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p} - \sigma_a^2(\theta_k\psi_0 + \theta_{k+1}\psi_1 + \cdots + \theta_q\psi_{q-k}) \tag{3.4.2}$$

with the convention that $\theta_0 = -1$. We see that this implies

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

and hence

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p} \qquad k \ge q+1 \tag{3.4.3}$$

or

$$\phi(B)\rho_k = 0 \qquad k \ge q+1$$

Thus, for the ARMA(p, q) process, there will be q autocorrelations $\rho_1, \ldots, \rho_q$ whose values depend directly on the choice of the q moving average parameters θ_i, as well as on the p autoregressive parameters ϕ_j. Also, the p values $\rho_{q-p+1}, \ldots, \rho_q$ provide the necessary starting values for the difference equation $\phi(B)\rho_k = 0$, where $k \ge q+1$, which then entirely determines the autocorrelations at higher lags. If $q - p < 0$, the whole autocorrelation function ρ_j, for $j = 0, 1, 2, \ldots$, will consist of a mixture of damped exponentials and/or damped sine waves, whose nature is dictated by (the roots of) the polynomial $\phi(B)$ and the starting values. If, however, $q - p \ge 0$, there will be $q - p + 1$ initial values $\rho_0, \rho_1, \ldots, \rho_{q-p}$, which do not follow this general pattern. These facts are useful in identifying mixed series.

Variance. When $k = 0$, we have

$$\gamma_0 = \phi_1\gamma_1 + \cdots + \phi_p\gamma_p + \sigma_a^2(1 - \theta_1\psi_1 - \cdots - \theta_q\psi_q) \tag{3.4.4}$$

which has to be solved along with the p equations (3.4.2) for $k = 1, 2, \ldots p$ to obtain $\gamma_0, \gamma_1, \ldots, \gamma_p$.

Spectrum. Using (3.1.12), the spectrum of the mixed ARMA(p, q) process is

$$\begin{aligned} p(f) &= 2\sigma_a^2 \frac{|\theta(e^{-i2\pi f})|^2}{|\phi(e^{-i2\pi f})|^2} \\ &= 2\sigma_a^2 \frac{|1 - \theta_1 e^{-i2\pi f} - \cdots - \theta_q e^{-i2q\pi f}|^2}{|1 - \phi_1 e^{-i2\pi f} - \cdots - \phi_p e^{-i2p\pi f}|^2} \qquad 0 \le f \le \frac{1}{2} \end{aligned} \tag{3.4.5}$$

Partial Autocorrelation Function. The mixed process $\phi(B)\tilde{z}_t = \theta(B)a_t$ can be written as

$$a_t = \theta^{-1}(B)\phi(B)\tilde{z}_t$$

where $\theta^{-1}(B)$ is an infinite series in B. Hence, the partial autocorrelation function of a mixed process is infinite in extent. It behaves eventually like the partial autocorrelation function of a pure moving average process, being dominated by a mixture of damped exponentials and/or damped sine waves, depending on the order of the moving average and the values of the parameters it contains.

3.4.3 First Order Autoregressive First-Order Moving Average Process

A mixed ARMA process of considerable practical importance is the ARMA(1, 1) process

$$\tilde{z}_t - \phi_1 \tilde{z}_{t-1} = a_t - \theta_1 a_{t-1} \tag{3.4.6}$$

that is,

$$(1 - \phi_1 B)\tilde{z}_t = (1 - \theta_1 B)a_t$$

We now derive some of its more important properties.

Stationarity and Invertibility Conditions. First, we note that the process is stationary if $-1 < \phi_1 < 1$, and invertible if $-1 < \theta_1 < 1$. Hence, the admissible parameter space is the square shown in Figure 3.9(a). In addition, from the relations $\psi_1 = \phi_1\psi_0 - \theta_1 = \phi_1 - \theta_1$ and $\psi_j = \phi_1\psi_{j-1}$ for $j > 1$, we find that the ψ_j weights are given by $\psi_j = (\phi_1 - \theta_1)\phi_1^{j-1}$, $j \ge 1$, and similarly it is easily seen that $\pi_j = (\phi_1 - \theta_1)\theta_1^{j-1}$, $j \ge 1$, for the stationary and invertible ARMA(1, 1) process.

Autocorrelation Function. From (3.4.2) and (3.4.4) we obtain

$$\begin{aligned} \gamma_0 &= \phi_1\gamma_1 + \sigma_a^2(1 - \theta_1\psi_1) \\ \gamma_1 &= \phi_1\gamma_0 - \theta_1\sigma_a^2 \\ \gamma_k &= \phi_1\gamma_{k-1} \qquad k \ge 2 \end{aligned}$$

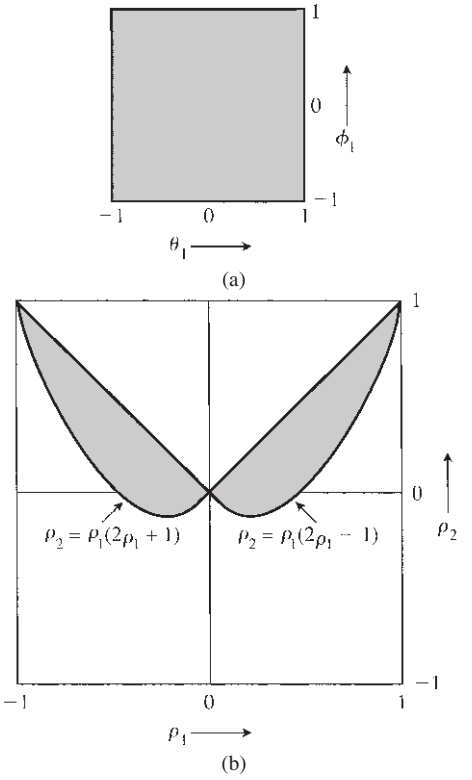

FIGURE 3.9 Admissible regions for (a) ϕ_1, θ_1 and (b) ρ_1, ρ_2 for a stationary and invertible ARMA(1, 1) process.

with $\psi_1 = \phi_1 - \theta_1$. Hence, solving the first two equations for γ_0 and γ_1, the autocovariance function of the process is

$$\begin{aligned}\gamma_0 &= \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2}\sigma_a^2 \\ \gamma_1 &= \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2}\sigma_a^2 \\ \gamma_k &= \phi_1\gamma_{k-1} \qquad k \geq 2\end{aligned} \tag{3.4.7}$$

The last equation gives $\rho_k = \phi_1\rho_{k-1}$, $k \geq 2$, so that $\rho_k = \rho_1\phi_1^{k-1}$, $k > 1$. Thus, the autocorrelation function decays exponentially from the starting value ρ_1, which depends on θ_1 and ϕ_1.[2] This exponential decay is smooth if ϕ_1 is positive and alternates if ϕ_1 is negative. Furthermore, the sign of ρ_1 is determined by the sign of $(\phi_1 - \theta_1)$ and dictates from which side of zero the exponential decay takes place.

[2]By contrast, the autocorrelation function for the AR(1) process decays exponentially from the starting value $\rho_0 = 1$.

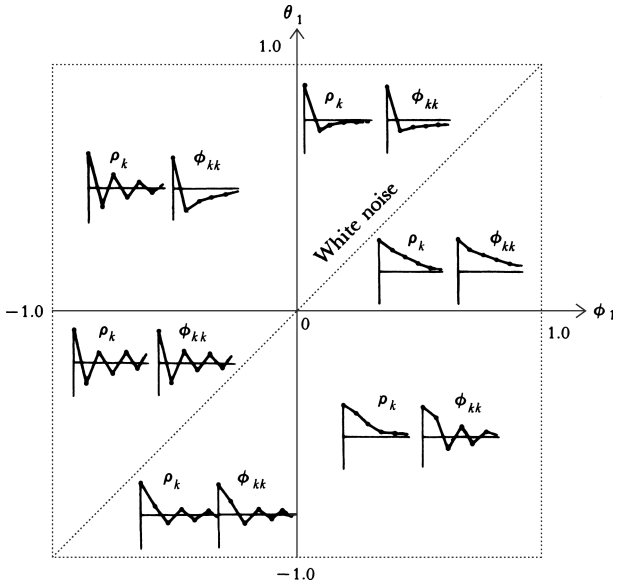

FIGURE 3.10 Autocorrelation and partial autocorrelation functions ρ_k and ϕ_{kk} for various ARMA(1, 1) models.

The first two autocorrelations may be expressed in terms of the parameters of the ARMA(1,1) process, as follows:

$$\rho_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1} \qquad (3.4.8)$$
$$\rho_2 = \phi_1\rho_1$$

Using these expressions and the stationarity and invertibility conditions, it may be shown that ρ_1 and ρ_2 must lie in the region

$$|\rho_2| < |\rho_1|$$
$$\rho_2 > \rho_1(2\rho_1 + 1) \qquad \rho < 0 \qquad (3.4.9)$$
$$\rho_2 > \rho_1(2\rho_1 - 1) \qquad \rho_1 > 0$$

Figure 3.9(b) shows the admissible space for ρ_1 and ρ_2; that is, it indicates which combinations of ρ_1 and ρ_2 are possible for a mixed (1, 1) stationary, invertible process.

Partial Autocorrelation Function. The partial autocorrelation function of the mixed ARMA(1, 1) process consists of a single initial value $\phi_{11} = \rho_1$. Thereafter, it behaves like the partial autocorrelation function of a pure MA(1) process and is dominated by a damped exponential. Thus, as shown in Figure 3.10, when θ_1 is positive, it is dominated by a smoothly damped exponential that decays from a value of ρ_1, with sign determined by the sign of $(\phi_1 - \theta_1)$. Similarly, when θ_1 is negative, it is dominated by an exponential that oscillates as it decays from a value of ρ_1, with sign determined by the sign of $(\phi_1 - \theta_1)$.

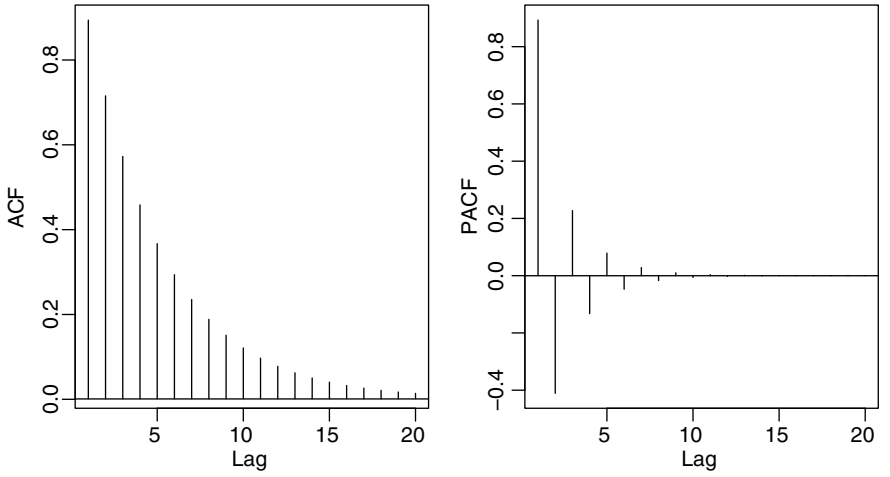

FIGURE 3.11 Theoretical autocorrelation and partial autocorrelation functions of an ARMA(1,1) process with $\phi = 0.8$ and $\theta = -0.6$.

Numerical Example. For numerical illustration, consider the ARMA(1, 1) process,

$$(1 - 0.8B)\tilde{z}_t = (1 + 0.6B)a_t$$

so that $\phi = 0.8$ and $\theta = -0.6$. Further assuming $\sigma_a^2 = 1$, we find from (3.4.7) and (3.4.8) that the variance of $\tilde{z}_t$ is $\gamma_0 = 6.444$, and $\rho_1 = 0.893$. Also, the autocorrelation function satisfies $\rho_j = 0.8\rho_{j-1}$, $j \geq 2$, so that $\rho_j = 0.893(0.8)^{j-1}$, for $j \geq 2$.

The autocorrelation and partial autocorrelation functions are shown in Figure 3.11. The exponential decay in the autocorrelation function is clearly evident from the graph. The partial autocorrelation function also exhibits an exponentially decaying pattern that oscillates in sign due to the negative value of θ. The figure was generated in R using the commands included below. Notice again that the parameter θ, although negative in this example, is entered as + 0.6 since R defines the MA operator $\theta(B)$ as $(1 + \theta B)$ rather that $(1 - \theta B)$ as done in this text.

```
> ACF=ARMAacf(ar=0.8,ma=0.6,20)[-1]
> PACF=ARMAacf(ar=0.8,ma=0.6,20,pacf=TRUE)
> win.graph(width=8,height=4)
> par(mfrow=c(1,2))
> plot(ACF,type="h",xlab="lag");abline(h=0)
> plot(PACF,type="h",xlab="lag");abline(h=0)
```

3.4.4 Summary

Figure 3.12 brings together the admissible regions for the parameters and for the autocorrelations ρ_1, ρ_2 for AR(2), MA(2), and ARMA(1, 1) processes, which are restricted to being both stationary and invertible. Table 3.2 summarizes the properties of mixed ARMA processes and brings together all the important results for autoregressive, moving average, and mixed processes, which will be needed in Chapter 6 to identify models for observed

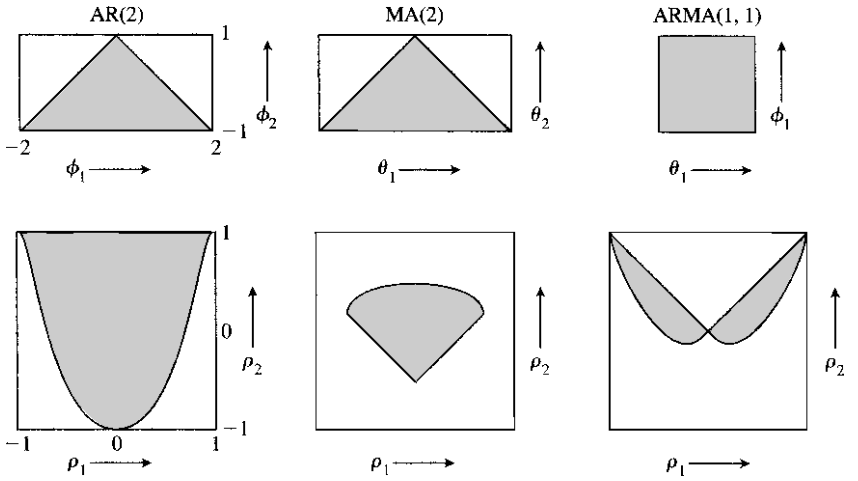

FIGURE 3.12 Admissible regions for the parameters and ρ_1, ρ_2 for AR(2), MA(2), and ARMA(1, 1) processes that are restricted to being both stationary and invertible.

time series. In the next chapter, we extend the mixed ARMA model to produce models that can describe nonstationary behavior of the kind that is frequently met in practice.

APPENDIX A3.1 AUTOCOVARIANCES, AUTOCOVARIANCE GENERATING FUNCTION, AND STATIONARITY CONDITIONS FOR A GENERAL LINEAR PROCESS

Autocovariances. The autocovariance at lag k of the linear process

$$\tilde{z}_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

with $\psi_0 = 1$ is clearly

$$\begin{aligned} \gamma_k &= E[\tilde{z}_t \tilde{z}_{t+k}] \\ &= E\left[\sum_{j=0}^{\infty}\sum_{h=0}^{\infty} \psi_j \psi_h a_{t-j} a_{t+k-h}\right] \\ &= \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \end{aligned} \tag{A3.1.1}$$

using the property (3.1.2) for the autocovariance function of white noise.

Autocovariance Generating Function. The result (A3.1.1) may be substituted in the autocovariance generating function

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k \tag{A3.1.2}$$

TABLE 3.2 Summary of Properties of Autoregressive, Moving Average, and Mixed ARMA Processes

	Autoregressive Process	Moving Average Processes	Mixed Processes
Model in terms of previous $\tilde{z}'s$	$\phi(B)\tilde{z}_t = a_t$	$\theta^{-1}(B)\tilde{z}_t = a_t$	$\theta^{-1}(B)\phi(B)\tilde{z}_t = a_t$
Model in terms of previous a's	$\tilde{z}_t = \phi^{-1}(B)a_t$	$\tilde{z}_t = \theta(B)a_t$	$\tilde{z}_t = \phi^{-1}(B)\theta(B)a_t$
π weights	Finite series	Infinite series	Infinite series
ψ weights	Infinite series	Finite series	Infinite series
Stationarity condition	Roots of $\phi(B) = 0$ lie outside the unit circle	Always stationary	Roots of $\phi(B) = 0$ lie outside the unit circle
Invertibility condition	Always invertible	Roots of $\theta(B) = 0$ lie outside the unit circle	Roots of $\theta(B) = 0$ lie outside the unit circle
Autocorrelation function	Infinite (damped exponentials and/or damped sine waves)	Finite	Infinite (damped exponentials and/or damped sine waves after first $q - p$ lags)
	Tails off	Cuts off after lag q	Tails off
Partial autocorrelation function	Finite	Infinite (dominated by damped exponentials and/or damped sine waves)	Infinite (dominated by damped exponentials and/or damped sine waves after first $p - q$ lags)
	Cuts off after lag p	Tails off	Tails off

to give

$$\gamma(B) = \sigma_a^2 \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} \psi_j \psi_{j+k} B^k$$
$$= \sigma_a^2 \sum_{j=0}^{\infty} \sum_{k=-j}^{\infty} \psi_j \psi_{j+k} B^k$$

since $\psi_h = 0$ for $h < 0$. Writing $j + k = h$, so that $k = h - j$, we have

$$\gamma(B) = \sigma_a^2 \sum_{j=0}^{\infty} \sum_{h=0}^{\infty} \psi_j \psi_h B^{h-j}$$
$$= \sigma_a^2 \sum_{h=0}^{\infty} \psi_h B^h \sum_{j=0}^{\infty} \psi_j B^{-j}$$

that is,

$$\gamma(B) = \sigma_a^2 \psi(B)\psi(B^{-1}) = \sigma_a^2 \psi(B)\psi(F) \tag{A3.1.3}$$

which is the result (3.1.11) quoted in the text.

Stationarity Conditions. If we substitute $B = e^{-i2\pi f}$ and $F = B^{-1} = e^{i2\pi f}$ in the autocovariance generating function (A3.1.2), we obtain half the power spectrum. Hence, the power spectrum of a linear process is

$$p(f) = 2\sigma_a^2 \psi(e^{-i2\pi f})\psi(e^{i2\pi f})$$
$$= 2\sigma_a^2 |\psi(e^{-i2\pi f})|^2 \qquad 0 \le f \le \tfrac{1}{2} \tag{A3.1.4}$$

It follows that the variance of the process is

$$\sigma_z^2 = \int_0^{1/2} p(f)\,df = 2\sigma_a^2 \int_0^{1/2} \psi(e^{-i2\pi f})\psi(e^{i2\pi f})\,df \tag{A3.1.5}$$

Now if the integral (A3.1.5) is to converge, it may be shown (Grenander and Rosenblatt, 1957) that the infinite series $\psi(B)$ must converge for B on or within the unit circle. More directly, for the linear process $\tilde{z}_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$, the condition $\sum_{j=0}^{\infty} |\psi_j| < \infty$ of absolute summability of the coefficients ψ_j implies (see Brockwell and Davis, 1991; Fuller, 1996) that the sum $\sum_{j=0}^{\infty} \psi_j a_{t-j}$ converges with probability 1 and hence represents a valid stationary process.

APPENDIX A3.2 RECURSIVE METHOD FOR CALCULATING ESTIMATES OF AUTOREGRESSIVE PARAMETERS

We now show how Yule–Walker estimates for the parameters of an AR($p + 1$) model may be obtained from the estimates for an AR(p) model fitted to the same time series. This recursive method of calculation, which is due to Levinson (1947) and Durbin (1960), can be used to approximate the partial autocorrelation function, as described in Section 3.2.6.

To illustrate the recursion, consider equations (3.2.35). Yule–Walker estimates are obtained for $k = 2, 3$ from

$$\begin{aligned} r_2 &= \hat{\phi}_{21} r_1 + \hat{\phi}_{22} \\ r_1 &= \hat{\phi}_{21} + \hat{\phi}_{22} r_1 \end{aligned} \tag{A3.2.1}$$

and

$$\begin{aligned} r_3 &= \hat{\phi}_{31} r_2 + \hat{\phi}_{32} r_1 + \hat{\phi}_{33} \\ r_2 &= \hat{\phi}_{31} r_1 + \hat{\phi}_{32} + \hat{\phi}_{33} r_1 \\ r_1 &= \hat{\phi}_{31} + \hat{\phi}_{32} r_1 + \hat{\phi}_{33} r_2 \end{aligned} \tag{A3.2.2}$$

The coefficients $\hat{\phi}_{31}$ and $\hat{\phi}_{32}$ may be expressed in terms of $\hat{\phi}_{33}$ using the last two equations of (A3.2.2). The solution may be written in matrix form as

$$\begin{pmatrix} \hat{\phi}_{31} \\ \hat{\phi}_{32} \end{pmatrix} = \mathbf{R}_2^{-1} \begin{pmatrix} r_2 - \hat{\phi}_{33} r_1 \\ r_1 - \hat{\phi}_{33} r_2 \end{pmatrix} \tag{A3.2.3}$$

where

$$\mathbf{R}_2 = \begin{bmatrix} r_1 & 1 \\ 1 & r_1 \end{bmatrix}$$

Now, (A3.2.3) may be written as

$$\begin{bmatrix} \hat{\phi}_{31} \\ \hat{\phi}_{32} \end{bmatrix} = \mathbf{R}_2^{-1} \begin{bmatrix} r_2 \\ r_1 \end{bmatrix} - \hat{\phi}_{33} \mathbf{R}_2^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \tag{A3.2.4}$$

Using the fact that (A3.2.1) may also be written as

$$\begin{bmatrix} \hat{\phi}_{21} \\ \hat{\phi}_{22} \end{bmatrix} = \mathbf{R}_2^{-1} \begin{bmatrix} r_2 \\ r_1 \end{bmatrix}$$

it follows that (A3.2.4) becomes

$$\begin{bmatrix} \hat{\phi}_{31} \\ \hat{\phi}_{32} \end{bmatrix} = \begin{bmatrix} \hat{\phi}_{21} \\ \hat{\phi}_{22} \end{bmatrix} - \hat{\phi}_{33} \begin{bmatrix} \hat{\phi}_{22} \\ \hat{\phi}_{21} \end{bmatrix}$$

that is,

$$\begin{aligned} \hat{\phi}_{31} &= \hat{\phi}_{21} - \hat{\phi}_{33} \hat{\phi}_{22} \\ \hat{\phi}_{32} &= \hat{\phi}_{22} - \hat{\phi}_{33} \hat{\phi}_{21} \end{aligned} \tag{A3.2.5}$$

To complete the calculation of $\hat{\phi}_{31}$ and $\hat{\phi}_{32}$, we need an expression for $\hat{\phi}_{33}$. On substituting (A3.2.5) in the first of the equations (A3.2.2), we obtain

$$\hat{\phi}_{33} = \frac{r_3 - \hat{\phi}_{21} r_2 - \hat{\phi}_{22} r_1}{1 - \hat{\phi}_{21} r_1 - \hat{\phi}_{22} r_2} \tag{A3.2.6}$$

Thus, the partial autocorrelation $\hat{\phi}_{33}$ is first calculated from $\hat{\phi}_{21}$ and $\hat{\phi}_{22}$, using (A3.2.6), and then the other two coefficients, $\hat{\phi}_{31}$ and $\hat{\phi}_{32}$, may be obtained from (A3.2.5).

In general, the recursive formulas are

$$\hat{\phi}_{p+1,j} = \hat{\phi}_{pj} - \hat{\phi}_{p+1,p+1}\hat{\phi}_{p,p+1-j} \qquad j = 1, 2, \ldots, p \tag{A3.2.7}$$

$$\hat{\phi}_{p+1,p+1} = \frac{r_{p+1} - \sum_{j=1}^{p} \hat{\phi}_{pj} r_{p+1-j}}{1 - \sum_{j=1}^{p} \hat{\phi}_{pj} r_j} \tag{A3.2.8}$$

EXERCISES

3.1 Write the following models in B notation:

(1) $\tilde{z}_t - 0.5\tilde{z}_{t-1} = a_t$

(2) $\tilde{z}_t = a_t - 1.3a_{t-1} + 0.4a_{t-2}$

(3) $\tilde{z}_t - 0.5\tilde{z}_{t-1} = a_t - 1.3a_{t-1} + 0.4a_{t-2}$

3.2 For each of the models of Exercise 3.1 and also for the following models, state whether it is (**a**) stationary or (**b**) invertible.

(4) $\tilde{z}_t - 1.5\tilde{z}_{t-1} + 0.6\tilde{z}_{t-2} = a_t$

(5) $\tilde{z}_t - \tilde{z}_{t-1} = a_t - 0.5a_{t-1}$

(6) $\tilde{z}_t - \tilde{z}_{t-1} = a_t - 1.3a_{t-1} + 0.3a_{t-2}$

3.3. For each of the models in Exercise 3.1, obtain:

(a) The first four ψ_j weights

(b) The first four π_j weights

(c) The autocovariance generating function

(d) The first four autocorrelations ρ_j

(e) The variance of $\tilde{z}_t$ assuming that $\sigma_a^2 = 1.0$

3.4. Calculate the first fifteen ψ_j weights for each of the three models in Exercise 3.2 using the function ARMAtoMA in R. See help(ARMAtoMA) for details.

3.5. Classify each of the models (1) to (4) in Exercises 3.1 and 3.2 as a member of the class of ARMA(p, q) processes.

3.6. **(a)** Write down the Yule–Walker equations for models (1) and (4) considered in Exercises 3.1 and 3.2.

(b) Solve these equations to obtain ρ_1 and ρ_2 for the two models.

(c) Obtain the partial autocorrelation function for the two models.

3.7. Consider the first-order autoregressive model $z_t = \theta_0 + \phi z_{t-1} + a_t$, where the constant θ_0 is a function of the mean of the series.

(a) Derive the autocovariances $\gamma_k = E([z_t - \mu][z_{t-k} - \mu])$ for this series.

(b) Calculate and plot the autocorrelation function for $\phi = 0.8$ using the R command ARMAacf(); see help(ARMAacf) for details.

(c) Calculate and plot the partial autocorrelation function for the same process.

3.8. Consider the mixed ARMA(1,1) model $z_t - \phi z_{t-1} = a_t - \theta a_{t-1}$, where $-1 < \phi < 1$ and $E(z_t)$ is assumed to be zero for convenience.

(a) Derive the autocovariances $\gamma_k = E([z_t - \mu][z_{t-k} - \mu])$ for this series.

(b) Calculate and plot the autocorrelation function for $\phi = 0.9$ and $\theta = -0.3$ using R (see Exercise 3.7).

(c) Calculate and plot the partial autocorrelation function for the same process.

3.9. For the AR(2) process $\tilde{z}_t - 1.0\tilde{z}_{t-1} + 0.5\tilde{z}_{t-2} = a_t$:

(a) Calculate ρ_1.

(b) Using ρ_0 and ρ_1 as starting values and the difference equation form for the autocorrelation function, calculate the values of ρ_k for $k = 2, \ldots, 15$.

(c) Use the plotted autocorrelation function to estimate the period and damping factor of the autocorrelation function.

(d) Check the values in (c) by direct calculation using the parameter values and the related roots G_1^{-1} and G_2^{-1} of $\phi(B) = 1 - 1.0B + 0.5B^2$.

3.10. **(a)** Plot the power spectrum $g(f)$ of the autoregressive process of Exercise 3.9, and show that it has a peak at a period that is close to the period in the autocorrelation function.

(b) Graphically, or otherwise, estimate the proportion of the variance of the series in the frequency band between $f = 0.0$ and $f = 0.2$ cycle per data interval.

3.11. **(a)** Why is it important to factorize the autoregressive and moving average operators after fitting a model to an observed series?

(b) It was shown by Jenkins (1975) that the number of mink skins z_t traded annually between 1848 and 1909 in North Canada is adequately represented by the AR(4) model

$$(1 - 0.82B + 0.22B^2 + 0.28B^4)[\ln(z_t) - \mu] = a_t$$

Factorize the autoregressive operator and explain what the factors reveal about the autocorrelation function and the underlying nature of the mink series. The data for the period 1850–1911 are listed as Series N in Part Five of this book. Note that the roots of $\phi(B) = 0$ can be calculated using the R commond polyroot(), where the autoregressive parameters are entered with their signs reversed; see help(polyroot) for details.

3.12. Calculate and plot the theoretical autocorrelation function and partial autocorrelation function for the AR(4) model specified in Exercise 3.11(b).

4

LINEAR NONSTATIONARY MODELS

Many empirical time series (e.g., stock price series) behave as though they had no fixed mean. Even so, they exhibit homogeneity in the sense that apart from local level, or perhaps local level and trend, one part of the series behaves much like any other part. Models that describe such homogeneous nonstationary behavior can be obtained by assuming that some suitable *difference* of the process is stationary. In this chapter, we examine the properties of the important class of models for which the dth difference of the series is a stationary mixed autoregressive–moving average process. These models are called autoregressive integrated moving average (ARIMA) processes.

4.1 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE PROCESSES

4.1.1 Nonstationary First-Order Autoregressive Process

Figure 4.1 shows four time series that have arisen in forecasting and control problems. All of them exhibit behavior suggestive of nonstationarity. Series A, C, and D represent ''uncontrolled'' outputs (concentration, temperature, and viscosity, respectively) from three different chemical processes. These series were collected to show the effect on these outputs of uncontrolled and unmeasured disturbances such as variations in feedstock and ambient temperature. The temperature Series C was obtained by temporarily disconnecting the controllers on the pilot plant involved and recording the subsequent temperature fluctuations. Both A and D were collected on full-scale processes, where it was necessary to maintain some output quality characteristic as close as possible to a fixed level. To achieve this control, another variable had been manipulated to approximately cancel out variations in the output. However, the effect of these manipulations on the output was accurately

Time Series Analysis: Forecasting and Control, Fifth Edition. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung

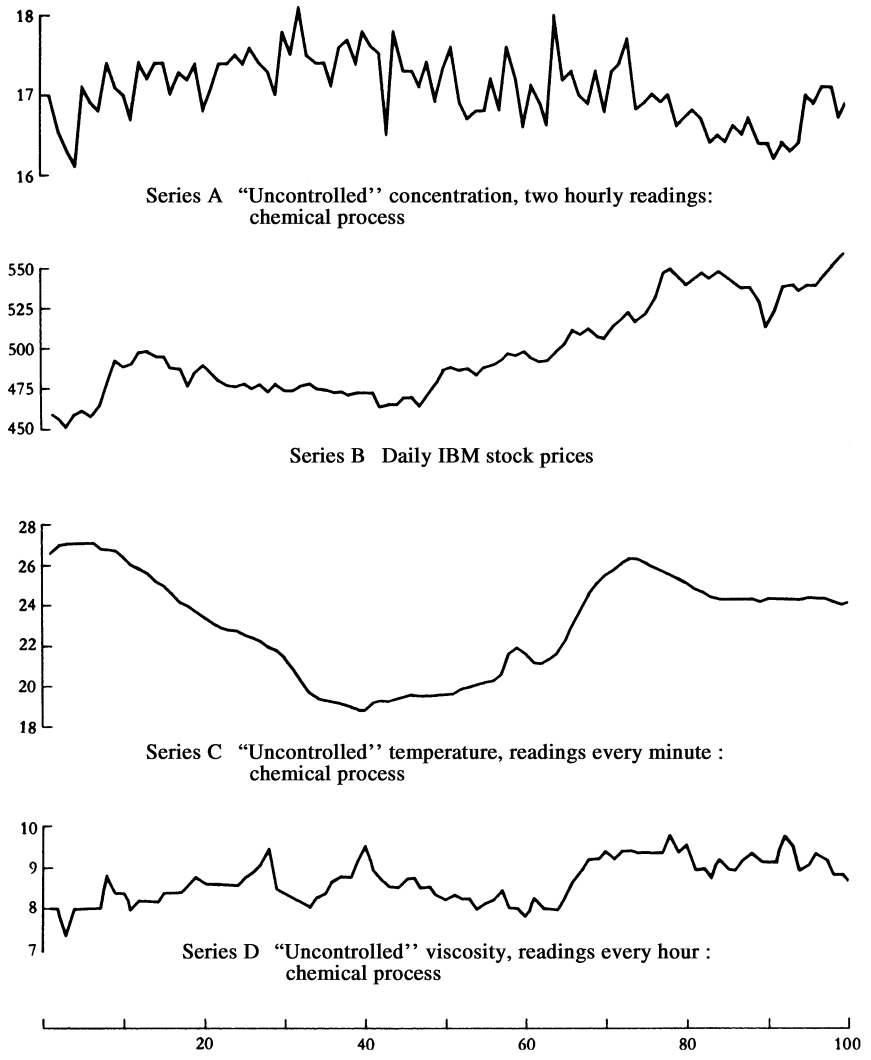

FIGURE 4.1 Typical time series arising in forecasting and control problems.

known in each case, so that it was possible to compensate numerically for the control action. That is, it was possible to calculate very nearly the values of the series that would have been obtained if no corrective action had been taken. It is these compensated values that are recorded here and referred to as the "uncontrolled" series. Series B consists of the daily IBM stock prices during a period beginning in May 1961. A complete list of all the series is given in the collection of time series at the end of this book. In Figure 4.1, 100 successive observations have been plotted from each series and the points joined by straight lines.

There are an unlimited number of ways in which a process can be nonstationary. However, the types of economic and industrial series that we wish to analyze frequently exhibit a particular kind of homogeneous nonstationary behavior that can be represented by a stochastic model, which is a modified form of the autoregressive–moving average

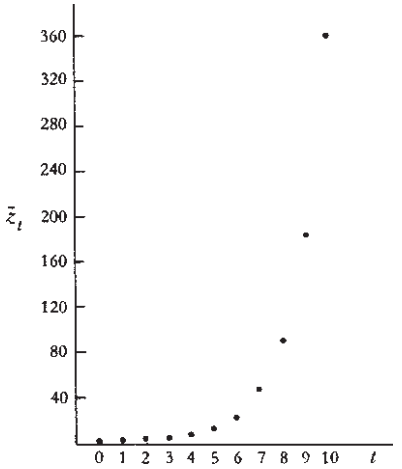

FIGURE 4.2 Realization of the nonstationary first-order autoregressive process $\tilde{z}_t = 2\tilde{z}_{t-1} + a_t$ with $\sigma_a^2 = 1$.

(ARMA) model. In Chapter 3, we considered the mixed ARMA model

$$\phi(B)\tilde{z}_t = \theta(B)a_t \tag{4.1.1}$$

with $\phi(B)$ and $\theta(B)$ polynomial operators in B, of degree p and q, respectively. To ensure stationarity, the roots of $\phi(B) = 0$ must lie outside the unit circle. A natural way of obtaining nonstationary processes is to relax this restriction.

To gain some insight into the possibilities, consider the first-order autoregressive model,

$$(1 - \phi B)\tilde{z}_t = a_t \tag{4.1.2}$$

which is stationary for $|\phi| < 1$. Let us study the behavior of this process for $\phi = 2$, a value outside the stationary range. Figure 4.2 shows a series $\tilde{z}_t$ generated by the model $\tilde{z}_t = 2\tilde{z}_{t-1} + a_t$ using a set of unit random normal deviates a_t and setting $\tilde{z}_0 = 0.7$. It is seen that after a short induction period, the series ''breaks loose'' and essentially follows an exponential curve, with the generating a_t's playing almost no further part. The behavior of series generated by processes of higher order, which violate the stationarity condition, is similar. Furthermore, this behavior is essentially the same whether or not moving average terms are introduced on the right of the model.

4.1.2 General Model for a Nonstationary Process Exhibiting Homogeneity

Autoregressive Integrated Moving Average Model. Although nonstationary models of the kind described above are of value to represent explosive or evolutionary behavior (such as bacterial growth), the applications that we describe in this book are not of this type. So far, we have seen that an ARMA process is stationary if the roots of $\phi(B) = 0$ lie *outside* the unit circle, and exhibits explosive nonstationary behavior if the roots lie *inside* the unit circle. The only case remaining is that the roots of $\phi(B) = 0$ lie *on* the unit circle. It turns out that the resulting models are of great value in representing homogeneous nonstationary

time series. In particular, nonseasonal series are often well represented by models in which one or more of these roots are *unity* and these are considered in the present chapter[1].

Let us consider the model

$$\varphi(B)\tilde{z}_t = \theta(B)a_t \tag{4.1.3}$$

where $\varphi(B)$ is a nonstationary autoregressive operator such that d of the roots of $\varphi(B) = 0$ are unity and the remainder lie outside the unit circle. Then the model can be written as

$$\varphi(B)\tilde{z}_t = \phi(B)(1-B)^d\tilde{z}_t = \theta(B)a_t \tag{4.1.4}$$

where $\phi(B)$ is a *stationary* autoregressive operator. Since $\nabla^d\tilde{z}_t = \nabla^d z_t$, for $d \geq 1$, where $\nabla = 1 - B$ is the differencing operator, we can write the model as

$$\phi(B)\nabla^d z_t = \theta(B)a_t \tag{4.1.5}$$

Equivalently, the process is defined by the two equations

$$\phi(B)w_t = \theta(B)a_t \tag{4.1.6}$$

and

$$w_t = \nabla^d z_t \tag{4.1.7}$$

Thus, we see that the model corresponds to assuming that the dth difference of the series can be represented by a stationary, invertible ARMA process. An alternative way of looking at the process for $d \geq 1$ results from inverting (4.1.7) to give

$$z_t = S^d w_t \tag{4.1.8}$$

where S is the infinite summation operator defined by

$$\begin{aligned} Sx_t = \sum_{h=-\infty}^{t} x_h &= (1 + B + B^2 + \cdots)x_t \\ &= (1-B)^{-1}x_t = \nabla^{-1}x_t \end{aligned}$$

Thus,

$$S = (1-B)^{-1} = \nabla^{-1}$$

The operator S^2 is similarly defined as

$$\begin{aligned} S^2x_t &= Sx_t + Sx_{t-1} + Sx_{t-2} + \cdots \\ &= \sum_{i=-\infty}^{t}\sum_{h=-\infty}^{i} x_h = (1 + 2B + 3B^2 + \cdots)x_t \end{aligned}$$

and so on for higher order d. Equation (4.1.8) implies that the process (4.1.5) can be obtained by summing (or ''integrating'') the stationary process (4.1.6) d times. Therefore, we call the process (4.1.5) an *autoregressive integrated moving average (ARIMA) process*.

[1]In Chapter 9, we consider models, capable of representing seasonality of period s, for which the characteristic equation has roots lying on the unit circle that are the sth roots of unity.

The ARIMA models for nonstationary time series, which were also considered earlier by Yaglom (1955), are of fundamental importance for forecasting and control as discussed by Box and Jenkins (1962, 1963, 1965, 1968a, 1968b, 1969) and Box et al. (1967a). Nonstationary processes were also discussed by Zadeh and Ragazzini (1950), Kalman (1960), and Kalman and Bucy (1961). An earlier procedure for time series analysis that employed differencing was the *variate difference method* (see Tintner (1940) and Rao and Tintner (1963)). However, the motivation, methods, and objectives of this procedure were quite different from those discussed here.

Technically, the *infinite* summation operator $S = (1 - B)^{-1}$ in (4.1.8) cannot actually be used in defining the nonstationary ARIMA processes, since the infinite sums involved will not be convergent. Instead, we can consider the finite summation operator S_m for any positive integer m, given by

$$S_m = (1 + B + B^2 + \cdots + B^{m-1}) \equiv \frac{1 - B^m}{1 - B}$$

Similarly, the finite double summation operator can be defined as

$$\begin{aligned} S_m^{(2)} &= \sum_{j=0}^{m-1} \sum_{i=j}^{m-1} B^i = (1 + 2B + 3B^2 + \cdots + mB^{m-1}) \\ &\equiv \frac{1 - B^m - mB^m(1 - B)}{(1 - B)^2} \end{aligned}$$

since $(1 - B)S_m^{(2)} = S_m - mB^m$, and so on. Then the relation between an integrated ARMA process z_t with $d = 1$, for example, and the corresponding stationary ARMA process $w_t = (1 - B)z_t$, in terms of values back to some earlier time origin $k < t$, can be expressed as

$$z_t = \frac{S_{t-k}}{1 - B^{t-k}} w_t = \frac{1}{1 - B^{t-k}} (w_t + w_{t-1} + \cdots + w_{k+1})$$

so that $z_t = w_t + w_{t-1} + \cdots + w_{k+1} + z_k$ can be thought of as the sum of a finite number of terms from the stationary process w plus an initializing value of the process z at time k. Hence, in the formal definition of the stochastic properties of a nonstationary ARIMA process as generated in (4.1.3), it would typically be necessary to specify initializing conditions for the process at some time point k in the finite (but possibly remote) past. However, these initial condition specifications will have little effect on most of the important characteristics of the process, and such specifications will for the most part not be emphasized in this book.

As mentioned in Chapter 1, the model (4.1.5) is equivalent to representing the process z_t as the output from a linear filter (unless $d = 0$, this is an *unstable* linear filter), whose input is white noise a_t. Alternatively, we can regard it as a device *for transforming the highly dependent, and possibly nonstationary process z_t, to a sequence of uncorrelated random variables a_t*, that is, for transforming the process to white noise.

If in (4.1.5), the autoregressive operator $\phi(B)$ is of order p, the dth difference is taken, and the moving average operator $\theta(B)$ is of order q, we say that we have an ARIMA model of order (p, d, q), or simply an ARIMA(p, d, q) process.

Two Interpretations of the ARIMA Model. We now show that the ARIMA model is an intuitively reasonable model for many time series that occur in practice. First, we note that the local behavior of a stationary time series is heavily dependent on the *level* of $\tilde{z}_t$. This is to be contrasted with the behavior of series such as those in Figure 4.1, where the local behavior of the series appears to be independent of its level.

If we are to use models for which the behavior of the process is independent of its level, we must choose the autoregressive operator $\varphi(B)$ such that

$$\varphi(B)(\tilde{z}_t + c) = \varphi(B)\tilde{z}_t$$

where c is any constant. Thus $\varphi(B)$ must be of the form

$$\varphi(B) = \phi_1(B)(1 - B) = \phi_1(B)\nabla$$

Therefore, a class of processes having the desired property will be of the form

$$\phi_1(B)w_t = \theta(B)a_t$$

where $w_t = \nabla\tilde{z}_t = \nabla z_t$. Required homogeneity excludes the possibility that w_t should increase explosively. This means that either $\phi_1(B)$ is a stationary autoregressive operator or $\phi_1(B) = \phi_2(B)(1 - B)$, so that $\phi_2(B)w_t = \theta(B)a_t$, where now $w_t = \nabla^2 z_t$. In the latter case, the same argument can be applied to the second difference, and so on.

Eventually, we arrive at the conclusion that for the representation of time series that are nonstationary but nevertheless exhibit homogeneity, the operator on the left of (4.1.3) should be of the form $\phi(B)\nabla^d$, where $\phi(B)$ is a *stationary* autoregressive operator. Thus, we are led back to the model (4.1.5).

To approach the model from a somewhat different viewpoint, consider the situation where $d = 0$ in (4.1.4), so that $\phi(B)\tilde{z}_t = \theta(B)a_t$. The requirement that the zeros of $\phi(B)$ lie outside the unit circle would ensure not only that the process $\tilde{z}_t$ was stationary with mean zero, but also that $\nabla z_t, \nabla^2 z_t, \nabla^3 z_t, \ldots$ were each stationary with mean zero. Figure 4.3(a) shows one kind of nonstationary series we would like to represent. This series is homogeneous except in level, in that except for a vertical translation, one part of it looks much the same as another. We can represent such behavior by retaining the requirement that each of the differences be stationary with zero mean, but letting the level ''go free.'' We do this by using the model

$$\phi(B)\nabla z_t = \theta(B)a_t$$

Figure 4.3(b) shows a second kind of nonstationarity or fairly common occurrence. The series has neither a fixed level nor a fixed slope, but its behavior is homogeneous if we allow for differences in these characteristics. We can represent such behavior by the model

$$\phi(B)\nabla^2 z_t = \theta(B)a_t$$

which ensures stationarity and zero mean for all differences after the first and second but allows the level and the slope to ''go free.''

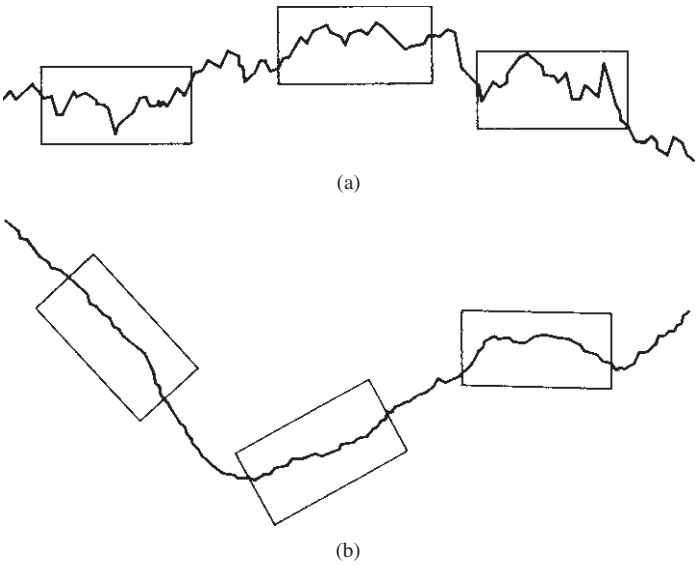

FIGURE 4.3 Two kinds of homogeneous nonstationary behavior. (a) A series showing nonstationarity in level such as can be represented by the model $\phi(B)\nabla z_t = \theta(B)a_t$. (b) A series showing nonstationarity in level and in slope such as can be represented by the model $\phi(B)\nabla^2 z_t = \theta(B)a_t$.

4.1.3 General Form of the ARIMA Model

For reasons to be given below, it is sometimes useful to consider a slight extension of the ARIMA model in (4.1.5), by adding a constant term θ_0, yielding the more general form

$$\varphi(B)z_t = \phi(B)\nabla^d z_t = \theta_0 + \theta(B)a_t \tag{4.1.9}$$

where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$
$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

In what follows:

1. $\phi(B)$ will be called the *autoregressive operator*; it is assumed to be stationary, that is, the roots of $\phi(B) = 0$ lie outside the unit circle.
2. $\varphi(B) = \phi(B)\nabla^d$ will be called the *generalized autoregressive* operator; it is a nonstationary operator with d of the roots of $\varphi(B) = 0$ equal to unity, that is, d unit roots.
3. $\theta(B)$ will be called the *moving average operator*; it is assumed to be invertible, that is, the roots of $\theta(B) = 0$ lie outside the unit circle.

When $d = 0$, this model represents a stationary process. The requirements of stationarity and invertibility apply independently, and, in general, the operators $\phi(B)$ and $\theta(B)$ will not be of the same order. Examples of the stationarity regions for the simple cases of $p = 1, 2$ and the identical invertibility regions for $q = 1, 2$ were given in Chapter 3.

Stochastic and Deterministic Trends. When the constant term θ_0 is omitted, the model (4.1.9) is capable of representing series that have *stochastic* trends, as typified, for example, by random changes in the level and slope of the series. In general, however, we may wish to include a *deterministic* function of time $f(t)$ in the model. In particular, automatic allowance for a deterministic polynomial trend, of degree d, can be made by permitting θ_0 to be nonzero. For example, when $d = 1$, we may use the model with $\theta_0 \neq 0$ to represent a possible deterministic linear trend in the presence of nonstationary noise. Since, from (3.1.22), to allow θ_0 to be nonzero is equivalent to permitting

$$E[w_t] = E[\nabla^d z_t] = \mu_w = \frac{\theta_0}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}$$

to be nonzero, an alternative way of expressing this more general model (4.1.9) is in the form of a stationary invertible ARMA process in $\tilde{w}_t = w_t - \mu_w$. That is,

$$\phi(B)\tilde{w}_t = \theta(B)a_t \tag{4.1.10}$$

Notice, when $d = 1$, for example, $\nabla z_t = w_t = \tilde{w}_t + \mu_w$ implies that $z_t = \tilde{z}_t + \mu_w t + \alpha$, where α is an intercept constant and the process $\tilde{z}_t$ is such that $\nabla \tilde{z}_t = \tilde{w}_t$, which has zero mean. Thus, $\theta_0 \neq 0$ allows for a deterministic linear trend component in z_t with slope $\mu_w = \theta_0/(1 - \phi_1 - \cdots - \phi_p)$.

In many applications, where no physical reason for a deterministic component exists, the mean of w can be assumed to be zero unless such an assumption is inconsistent with the data. In many cases, the assumption of a stochastic trend is more realistic than the assumption of a deterministic trend. This is of special importance in forecasting, since a stochastic trend does not require the series to follow the trend pattern seen in the past. In what follows, when $d > 0$, we will often assume that $\mu_w = 0$, or equivalently, that $\theta_0 = 0$, unless it is clear from the data or from the nature of the problem that a nonzero mean, or more generally a deterministic component of known form, is needed.

Some Important Special Cases of the ARIMA Model. In Chapter 3, we examined some important special cases of the model (4.1.9), corresponding to the stationary situation, $d = 0$. The following models represent some special cases of the nonstationary model ($d \geq 1$), which seem to be common in practice.

1. *The (0, 1, 1) process:*

$$\begin{aligned} \nabla z_t &= a_t - \theta_1 a_{t-1} \\ &= (1 - \theta_1 B)a_t \end{aligned}$$

corresponding to $p = 0$, $d = 1$, $q = 1$, $\phi(B) = 1$, $\theta(B) = 1 - \theta_1 B$.

2. *The (0, 2, 2) process:*

$$\begin{aligned} \nabla^2 z_t &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \\ &= (1 - \theta_1 B - \theta_2 B^2)a_t \end{aligned}$$

corresponding to $p = 0$, $d = 2$, $q = 2$, $\phi(B) = 1$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2$.

3. *The (1, 1, 1) process:*

$$\nabla z_t - \phi_1 \nabla z_{t-1} = a_t - \theta_1 a_{t-1}$$

TABLE 4.1 Summary of Simple Nonstationary Models Fitted to Time Series of Figure 4.1

Series	Model	Order of Model
A	$\nabla z_t = (1 - 0.7B)a_t$	(0, 1, 1)
B	$\nabla z_t = (1 + 0.1B)a_t$	(0, 1, 1)
C	$(1 - 0.8B)\nabla z_t = a_t$	(1, 1, 0)
D	$\nabla z_t = (1 - 0.1B)a_t$	(0, 1, 1)

or

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

corresponding to $p = 1$, $d = 1$, $q = 1$, $\phi(B) = 1 - \phi_1 B$, $\theta(B) = 1 - \theta_1 B$.

For the representation of nonseasonal time series (seasonal models are considered in Chapter 9), we rarely seem to meet situations for which either p, d, or q need to be greater than 2. Frequently, values of zero or unity will be appropriate for one or more of these orders. For example, we show later that Series A, B, C, and D given in Figure 4.1 are reasonably well represented[2] by the simple models shown in Table 4.1.

Nonlinear Transformation of z. The range of useful applications of the model (4.1.9) widens considerably if we allow the possibility of transformation. Thus, we may substitute $z_t^{(\lambda)}$ for z_t, in (4.1.9), where $z_t^{(\lambda)}$ is some nonlinear transformation of z_t, involving one or more parameters λ. A suitable transformation may be suggested by the application, or in some cases it can be estimated from the data. For example, if we were interested in the sales of a recently introduced commodity, we might find that the sales volume was increasing at a rapid rate and that it was the *percentage* fluctuation that showed nonstationary stability (homogeneity) rather than the absolute fluctuation. This would support the analysis of the logarithm of sales since

$$\nabla \log(z_t) = \log\left(\frac{z_t}{z_{t-1}}\right) = \log\left(1 + \frac{\nabla z_t}{z_{t-1}}\right) \simeq \frac{\nabla z_t}{z_{t-1}}$$

where $\nabla z_t / z_{t-1}$ are the relative or percentage changes, the approximation holding if the relative changes are not excessively large. When the data cover a wide range and especially for seasonal data, estimation of the transformation using the approach of Box and Cox (1964) may be helpful (for an example, see Section 9.3.5). This approach considers the family of power transformations of the form $z_t^{(\lambda)} = (z_t^{\lambda} - 1)/\lambda$ for $\lambda \neq 0$ and $z_t^{(0)} = \log(z_t)$ for $\lambda = 0$.

Software to estimate the parameter λ in the Box–Cox power transformation is available in the TSA and MASS libraries of R. For example, the function BoxCox.ar() in the TSA package finds a power transformation so that the transformed series is approximately a Gaussian AR process.

[2]As is discussed more fully later, there are certain advantages in using a nonstationary rather than a stationary model in cases of doubt. In particular, none of the fitted models above assume that z_t has a fixed mean. However, we show in Chapter 7 that it is possible in certain cases to obtain stationary models of slightly better fit.

4.2 THREE EXPLICIT FORMS FOR THE ARIMA MODEL

We now consider three different ''explicit'' forms for the general model (4.1.9). Each of these allows some special aspect to be appreciated. Thus, the current value z_t of the process can be expressed

1. In terms of previous values of the z's and current and previous values of the a's, by direct use of the *difference equation*,
2. In terms of *current and previous shocks* a_{t-j} only, and
3. In terms of a weighted sum of *previous values* z_{t-j} of the process and the current shock a_t.

In this chapter, we are concerned primarily with *nonstationary* models in which $\nabla^d z_t$ is a stationary process and d is greater than zero. For such models, we can, without loss of generality, omit μ from the specification or equivalently replace $\tilde{z}_t$ by z_t. The results of this chapter and the next will, however, apply to stationary models for which $d = 0$, provided that z_t is then interpreted as the *deviation* from the mean μ.

4.2.1 Difference Equation Form of the Model

Direct use of the difference equation permits us to express the current value z_t of the process in terms of previous values of the z's and of the current and previous values of the a's. Thus, if

$$\varphi(B) = \phi(B)(1 - B)^d = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_{p+d} B^{p+d}$$

the general model (4.1.9), with $\theta_0 = 0$, may be written as

$$z_t = \varphi_1 z_{t-1} + \cdots + \varphi_{p+d} z_{t-p-d} - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} + a_t \tag{4.2.1}$$

For example, consider the process represented by the model of order (1, 1, 1)

$$(1 - \phi B)(1 - B)z_t = (1 - \theta B)a_t$$

where, for convenience, we drop the subscript 1 on ϕ_1 and θ_1. Then this process may be written as

$$[1 - (1 + \phi)B + \phi B^2]z_t = (1 - \theta B)a_t$$

that is,

$$z_t = (1 + \phi)z_{t-1} - \phi z_{t-2} + a_t - \theta a_{t-1} \tag{4.2.2}$$

with $\varphi_1 = 1 + \phi$ and $\varphi_2 = -\phi$ in the notation introduced above. For many purposes, and, in particular, for calculating forecasts, the difference equation (4.2.1) is the most convenient form to use.

4.2.2 Random Shock Form of the Model

Model in Terms of Current and Previous Shocks. As discussed in Chapter 3, a linear model can be written as the output z_t from the linear filter

$$\begin{aligned} z_t &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \\ &= a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} \\ &= \psi(B) a_t \end{aligned} \tag{4.2.3}$$

whose input is white noise, or a sequence of uncorrelated shocks a_t with mean 0 and common variance σ_a^2. It is sometimes useful to express the ARIMA model in this form, and, in particular, the ψ weights will be needed in Chapter 5 to calculate the variance of the forecast errors. However, since the nonstationary ARIMA processes are not in statistical equilibrium over time, they cannot be assumed to extend infinitely into the past, and hence an infinite representation as in (4.2.3) will not be possible. But a related finite truncated form, which will be discussed subsequently, always exists. We now show that the ψ weights for an ARIMA process may be obtained directly from the difference equation form of the model.

General Expression for the ψ Weights. If we operate on both sides of (4.2.3) with the generalized autoregressive operator $\varphi(B)$, we obtain

$$\varphi(B) z_t = \varphi(B)\psi(B) a_t$$

However, since $\varphi(B) z_t = \theta(B) a_t$, it follows that

$$\varphi(B)\psi(B) = \theta(B) \tag{4.2.4}$$

Therefore, the ψ weights may be obtained by equating coefficients of B in the expansion

$$\begin{aligned} (1 - \varphi_1 B - \cdots - \varphi_{p+d} B^{p+d})(1 + \psi_1 B + \psi_2 B^2 + \cdots) \\ = (1 - \theta_1 B - \cdots - \theta_q B^q) \end{aligned} \tag{4.2.5}$$

Thus, we find that the ψ_j weights of the ARIMA process can be determined recursively through the equations

$$\psi_j = \varphi_1 \psi_{j-1} + \varphi_2 \psi_{j-2} + \cdots + \varphi_{p+d} \psi_{j-p-d} - \theta_j \quad j > 0$$

with $\psi_0 = 1$, $\psi_j = 0$ for $j < 0$, and $\theta_j = 0$ for $j > q$. We note that for j greater than the larger of $p + d - 1$ and q, the ψ weights satisfy the homogeneous difference equation defined by the generalized autoregressive operator, that is,

$$\varphi(B)\psi_j = \phi(B)(1 - B)^d \psi_j = 0 \tag{4.2.6}$$

where B now operates on the subscript j. Thus, for sufficiently large j, the weights ψ_j are represented by a mixture of polynomials, damped exponentials, and damped sinusoids in the argument j.

Example. For illustration, consider the (1, 1, 1) process (4.2.2), for which

$$\begin{aligned}\varphi(B) &= (1 - \phi B)(1 - B) \\ &= 1 - (1 + \phi)B + \phi B^2\end{aligned}$$

and

$$\theta(B) = 1 - \theta B$$

Substituting in (4.2.5) gives

$$[1 - (1 + \phi)B + \phi B^2](1 + \psi_1 B + \psi_2 B^2 + \cdots) = 1 - \theta B$$

and hence the ψ_j satisfy the recursion $\psi_j = (1 + \phi)\psi_{j-1} - \phi\psi_{j-2}$, $j \geq 2$ with $\psi_0 = 1$ and $\psi_1 = (1 + \phi) - \theta$. Thus, since the roots of $\varphi(B) = (1 - \phi B)(1 - B) = 0$ are $G_1^{-1} = 1$ and $G_2^{-1} = \phi^{-1}$, we have, in general,

$$\psi_j = A_0 + A_1\phi^j \tag{4.2.7}$$

where the constants A_0 and A_1 are determined from the initial values $\psi_0 = A_0 + A_1 = 1$ and $\psi_1 = A_0 + A_1\phi = 1 + \phi - \theta$ as

$$A_0 = \frac{1 - \theta}{1 - \phi} \quad A_1 = \frac{\theta - \phi}{1 - \phi}$$

Thus, informally, we may wish to express model (4.2.2) in the equivalent form

$$z_t = \sum_{j=0}^{\infty}(A_0 + A_1\phi^j)a_{t-j} \tag{4.2.8}$$

Since $|\phi| < 1$, the weights ψ_j tend to A_0 for large j, so that shocks a_{t-j}, which entered in the remote past, receive a constant weight A_0. However, the representation in (4.2.8) is strictly not valid because the infinite sum on the right does not converge in any sense; that is, the weights ψ_j are not absolutely summable as in the case of a stationary process. A related truncated version of the random shock form of the model is always valid, as we discuss in detail shortly. Nevertheless, for notational convenience, we will often refer to the infinite random shock form (4.2.3) of an ARIMA process, even though this form is strictly not convergent, as a simple notational device to represent the valid truncated form in (4.2.14), in situations where the distinction between the two forms is not important.

Truncated Form of the Random Shock Model. For technical purposes, it is necessary and in some cases convenient to consider the model in a form slightly different from (4.2.3). Suppose that we wish to express the current value z_t of the process in terms of the $t - k$ shocks $a_t, a_{t-1}, \ldots, a_{k+1}$, which have entered the system since some time origin $k < t$. This time origin k might, for example, be the time at which the process was first observed.

The general model

$$\varphi(B)z_t = \theta(B)a_t \tag{4.2.9}$$

is a difference equation with the solution

$$z_t = C_k(t - k) + I_k(t - k) \tag{4.2.10}$$

A short discussion of linear difference equations is given in Appendix A4.1. We remind the reader that the solution of such equations closely parallels the solution of linear differential equations. The *complimentary function* $C_k(t-k)$ is the general solution of the homogeneous difference equation

$$\varphi(B)C_k(t-k) = 0 \tag{4.2.11}$$

In general, this solution will consist of a *linear* combination of certain functions of time. These functions are powers t^j, real geometric (exponential) terms G^t, and complex geometric (exponential) terms $D^t \sin(2\pi f_0 t + F)$, where the constants G, f_0, and F are functions of the parameters $(\boldsymbol{\phi}, \boldsymbol{\theta})$ of the model. The coefficients that form the linear combinations of these terms can be determined so as to satisfy a set of initial conditions defined by the values of the process before time $k+1$. The *particular integral* $I_k(t-k)$ is *any* function that satisfies

$$\varphi(B)I_k(t-k) = \theta(B)a_t \tag{4.2.12}$$

It should be carefully noted that in this expression B operates on t and *not on* k. It is shown in Appendix A4.1 that this equation is satisfied for $t-k > q$ by

$$I_k(t-k) = \sum_{j=0}^{t-k-1} \psi_j a_{t-j} = a_t + \psi_1 a_{t-1} + \cdots + \psi_{t-k-1} a_{k+1} \quad t > k \tag{4.2.13}$$

with $I_k(t-k) = 0, t \le k$. This particular integral $I_k(t-k)$, thus, represents the finite truncated form of the infinite random shock form (4.2.3), while the complementary function $C_k(t-k)$ embodies the ''initializing'' features of the process z in the sense that $C_k(t-k)$ is already determined or specified by the time $k+1$. Hence, the truncated form of the random shock model for the ARIMA process (4.1.3) is given by

$$z_t = \sum_{j=0}^{t-k-1} \psi_j a_{t-j} + C_k(t-k) \tag{4.2.14}$$

For illustration, consider Figure 4.4. The above discussion implies that any observation z_t can be considered in relation to any previous time k and can be divided up into two additive parts. The first part $C_k(t-k)$ is the component of z_t, *already determined at time* k, and indicates what the observations prior to time $k+1$ had to tell us about the value of the series at time t. It represents the course that the process would take if at time k, the source of shocks a_t had been ''switched off.'' The second part, $I_k(t-k)$, represents an additional component, *unpredictable at time* k, which embodies the entire effect of shocks entering the system at time k. Hence, to specify an ARIMA process, one must specify the initializing component $C_k(t-k)$ in (4.2.14) for some time origin k in the finite (but possibly remote) past, with the remaining course of the process being determined through the truncated random shock terms in (4.2.14).

Example. For illustration, consider again the example

$$(1-\phi B)(1-B)z_t = (1-\theta B)a_t$$

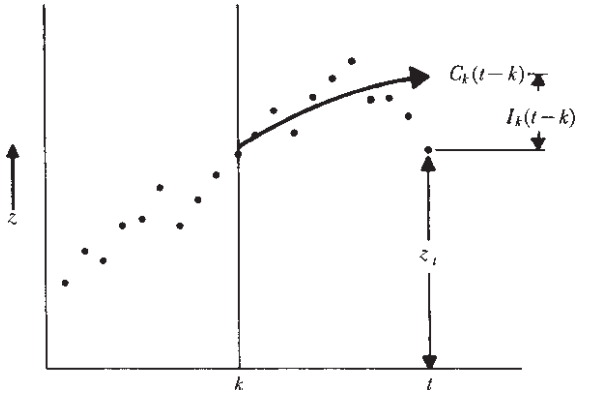

FIGURE 4.4 Role of the complementary function $C_k(t-k)$ and of the particular integral $I_k(t-k)$ in describing the behavior of a time series.

The complementary function is the solution of the difference equation

$$(1-\phi B)(1-B)C_k(t-k)=0$$

that is,

$$C_k(t-k)=b_0^{(k)}+b_1^{(k)}\phi^{t-k}$$

where $b_0^{(k)}$, $b_1^{(k)}$ are coefficients that depend on the past history of the process and, it will be noted, change with the origin k.

Making use of the ψ weights (4.2.7), a particular integral (4.2.13) is

$$I_k(t-k)=\sum_{j=0}^{t-k-1}(A_0+A_1\phi^j)a_{t-j}$$

so that, finally, we can write the model (4.2.8) in the equivalent form

$$z_t=b_0^{(k)}+b_1^{(k)}\phi^{t-k}+\sum_{j=0}^{t-k-1}(A_0+A_1\phi^j)a_{t-j} \tag{4.2.15}$$

Note that since $|\phi|<1$, if $t-k$ is chosen sufficiently large, the term involving ϕ^{t-k} in this expression is negligible and may be ignored.

Link Between the Truncated and Nontruncated Forms of the Random Shock Model. Returning to the general case, we can always think of the process with reference to some (possibly remote) finite origin k, with the process having the truncated random shock form as in (4.2.14). By comparison with the nontruncated form in (4.2.3), one can see that we might, informally, make the correspondence of representing the complementary function $C_k(t-k)$ in terms of the ψ weights as

$$C_k(t-k)=\sum_{j=t-k}^{\infty}\psi_j a_{t-j} \tag{4.2.16}$$

even though, formally, the infinite sum on the right of (4.2.16) does not converge. As mentioned earlier, for notational simplicity, we will often use this correspondence.

In summary, then, for the general model (4.2.9),

1. We can express the value z_t of the process, informally, as an infinite weighted sum of current and previous shocks a_{t-j} , according to

$$z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} = \psi(B)a_t$$

2. The value of z_t can be expressed, more formally, as a weighted finite sum of the $t-k$ current and previous shocks occurring after some origin k, plus a complementary function $C_k(t-k)$. This finite sum consists of the first $t-k$ terms of the infinite sum, so that

$$z_t = C_k(t-k) + \sum_{j=0}^{t-k-1} \psi_j a_{t-j} \tag{4.2.17}$$

Finally, the complementary function $C_k(t-k)$ can be taken, for notational convenience, to be represented as the truncated infinite sum, so that

$$C_k(t-k) = \sum_{j=t-k}^{\infty} \psi_j a_{t-j} \tag{4.2.18}$$

For illustration, consider once more the model

$$(1-\phi B)(1-B)z_t = (1-\theta B)a_t$$

We can write z_t either, informally, as an infinite sum of the a_{t-j}'s

$$z_t = \sum_{j=0}^{\infty} (A_0 + A_1\phi^j)a_{t-j}$$

or, more formally, in terms of the weighted finite sum as

$$z_t = C_k(t-k) + \sum_{j=0}^{t-k-1} (A_0 + A_1\phi^j)a_{t-j}$$

Furthermore, the complementary function can be written as

$$C_k(t-k) = b_0^{(k)} + b_1^{(k)}\phi^{t-k}$$

where $b_0^{(k)}$ and $b_1^{(k)}$, which satisfy the initial conditions through time k, are

$$b_0^{(k)} = \frac{z_k - \phi z_{k-1} - \theta a_k}{1-\phi} \quad b_1^{(k)} = \frac{-\phi(z_k - z_{k-1}) + \theta a_k}{1-\phi}$$

The complementary function can also be represented, informally, as the truncated infinite sum

$$C_k(t-k) = \sum_{j=t-k}^{\infty} (A_0 + A_1\phi^j)a_{t-j}$$

from which it can be seen that $b_0^{(k)}$ and $b_1^{(k)}$ may be represented as

$$b_0^{(k)} = A_0 \sum_{j=t-k}^{\infty} a_{t-j} = \frac{1-\theta}{1-\phi} \sum_{j=t-k}^{\infty} a_{t-j}$$

$$b_1^{(k)} = A_1 \sum_{j=t-k}^{\infty} \phi^{j-(t-k)} a_{t-j} = \frac{\theta-\phi}{1-\phi} \sum_{j=t-k}^{\infty} \phi^{j-(t-k)} a_{t-j}$$

Complementary Function as a Conditional Expectation. One consequence of the truncated form (4.2.14) is that for $m > 0$,

$$\begin{aligned} C_k(t-k) = C_{k-m}(t-k+m) + \psi_{t-k}a_k + \psi_{t-k+1}a_{k-1} + \cdots \\ + \psi_{t-k+m-1}a_{k-m+1} \end{aligned} \tag{4.2.19}$$

which shows how the complementary function changes as the origin k is changed. Now denote by $E_k[z_t]$ the *conditional expectation of* z_t, *at time* k. That is the expectation given complete historical knowledge of the process up to, but not beyond time k. To calculate this expectation, note that

$$E_k[a_j] = \begin{cases} 0 & j > k \\ a_j & j \le k \end{cases}$$

That is, *standing at time* k, the expected values of the future a's are zero and of those that have happened already are their actually realized values.

By taking conditional expectations at time k on both sides of (4.2.17), we obtain $E_k[z_t] = C_k(t-k)$. Thus, for $(t-k) > q$, the complementary function provides the expected value of the future value z_t of the process, *viewed from time* k and based on knowledge of the past. The particular integral shows how that expectation is modified by *subsequent* events represented by the shocks $a_{k+1}, a_{k+2}, \ldots, a_t$. In the problem of forecasting, which we discuss in Chapter 5, it will turn out that $C_k(t-k)$ is the minimum mean square error forecast of z_t made at time k. Equation (4.2.19) may be used in ''updating'' this forecast.

4.2.3 Inverted Form of the Model

Model in Terms of Previous z's and the Current Shock a_t. We have seen in Section 3.1.1 that the model

$$z_t = \psi(B)a_t$$

may also be written in the inverted form

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

or

$$\pi(B)z_t = \left(1 - \sum_{j=1}^{\infty} \pi_j B^j\right) z_t = a_t \tag{4.2.20}$$

Thus, z_t is an infinite weighted sum of previous values of z, plus a random shock:

$$z_t = \pi_1 z_{t-1} + \pi_2 z_{t-2} + \cdots + a_t$$

Because of the invertibility condition, the π weights must form a convergent series; that is, $\pi(B)$ must converge on or within the unit circle.

General Expression for the π Weights. To derive the π weights for the general ARIMA model, we can substitute (4.2.20) in

$$\varphi(B)z_t = \theta(B)a_t$$

to obtain

$$\varphi(B)z_t = \theta(B)\pi(B)z_t$$

Hence, the π weights can be obtained explicitly by equating coefficients of B in

$$\varphi(B) = \theta(B)\pi(B) \tag{4.2.21}$$

that is,

$$\begin{aligned}(1 - \varphi_1 B - \cdots - \varphi_{p+d} B^{p+d}) = (1 - \theta_1 B - \cdots - \theta_q B^q) \\ \times (1 - \pi_1 B - \pi_2 B^2 - \cdots)\end{aligned} \tag{4.2.22}$$

Thus, we find that the π_j weights of the ARIMA process can be determined recursively through

$$\pi_j = \theta_1 \pi_{j-1} + \theta_2 \pi_{j-2} + \cdots + \theta_q \pi_{j-q} + \varphi_j \quad j > 0$$

with the convention $\pi_0 = -1$, $\pi_j = 0$ for $j < 0$, and $\varphi_j = 0$ for $j > p + d$. It will be noted that for j greater than the larger of $p + d$ and q, the π weights satisfy the homogeneous difference equation defined by the *moving average operator*

$$\theta(B)\pi_j = 0$$

where B now operates on j. Hence, for sufficiently large j, the π weights will exhibit similar behavior as the autocorrelation function (3.2.5) of an autoregressive process; that is, they follow a mixture of damped exponentials and damped sine waves.

Another interesting fact is that if $d \geq 1$, the π weights in (4.2.20) sum to unity. This may be verified by substituting $B = 1$ in (4.2.21). Thus, $\varphi(B) = \phi(B)(1 - B)^d$ is zero when $B = 1$ and $\theta(1) \neq 0$, because the roots of $\theta(B) = 0$ lie outside the unit circle. Hence, it follows from (4.2.21) that $\pi(1) = 0$, that is,

$$\sum_{j=1}^{\infty} \pi_j = 1 \tag{4.2.23}$$

Therefore, if $d \geq 1$, the process may be written in the form

$$z_t = \bar{z}_{t-1}(\pi) + a_t \tag{4.2.24}$$

where

$$\bar{z}_{t-1}(\pi) = \sum_{j=1}^{\infty} \pi_j z_{t-j}$$

is a *weighted average* of previous values of the process.

Example. We again consider, for illustration, the ARIMA(1, 1, 1) process:

$$(1 - \phi B)(1 - B)z_t = (1 - \theta B)a_t$$

Then, using (4.2.21),

$$\pi(B) = \varphi(B)\theta^{-1}(B) = [1 - (1 + \phi)B + \phi B^2](1 + \theta B + \theta^2 B^2 + \cdots)$$

so that

$$\pi_1 = \phi + (1 - \theta) \quad \pi_2 = (\theta - \phi)(1 - \theta) \quad \pi_j = (\theta - \phi)(1 - \theta)\theta^{j-2}, \quad j \geq 3.$$

The first seven π weights corresponding to $\phi = -0.3$ and $\theta = 0.5$ are given in Table 4.2. Thus, z_t would be generated by a weighted average of previous values, plus an additional shock, according to

$$z_t = (0.2z_{t-1} + 0.4z_{t-2} + 0.2z_{t-3} + 0.1z_{t-4} + \cdots) + a_t$$

We notice, in particular, that the π weights die out as more and more remote values of z_{t-j} are involved. This happens when $-1 < \theta < 1$, so that the series is invertible.

We mention in passing that, for models fitted to actual time series, the convergent π weights usually die out rather quickly. Thus, although z_t may be theoretically dependent on the remote past, the representation

$$z_t = \sum_{j=1}^{\infty} \pi_j z_{t-j} + a_t$$

will usually show that z_t is dependent *to an important extent* only on recent past values z_{t-j} of the time series. This is still true even though for nonstationary models with $d > 0$, the ψ weights in the ''weighted shock'' representation

$$z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

do not die out to zero. What happens, of course, is that all the information that remote values of the shocks a_{t-j} supply about z_t is contained in recent values $z_{t-1}, z_{t-2}, \cdots$ of the series. In particular, the expectation $E_k[z_t]$, which in theory is conditional on complete history of the process up to time k, can usually be computed to sufficient accuracy from *recent* values of the time series. This fact is particularly important in forecasting applications.

TABLE 4.2 First Seven π Weights for an ARIMA(1, 1, 1) Process with $\phi = -0.3$, $\theta = 0.5$

j	1	2	3	4	5	6	7
π_j	0.2	0.4	0.2	0.1	0.05	0.025	0.0125

4.3 INTEGRATED MOVING AVERAGE PROCESSES

A nonstationary model that is useful in representing some commonly occurring series is the (0, 1, 1) process:

$$\nabla z_t = a_t - \theta a_{t-1}$$

The model contains only two parameters, θ and σ_a^2. Figure 4.5 shows two time series generated by this model from the same sequence of random normal deviates a_t. For the first series, $\theta = 0.6$, and for the second, $\theta = 0$. Models of this kind have often been found useful in inventory control problems, in representing certain kinds of disturbances occurring in industrial processes, and in econometrics. We will show in Chapter 7 that this simple process can, with suitable parameter values, supply useful representations of Series A, B, and D shown in Figure 4.1. Another valuable model is the (0, 2, 2) process

$$\nabla^2 z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

which contains three parameters, θ_1, θ_2, and σ_a^2. Figure 4.6 shows two series generated from this model using the same set of normal deviates. For the first series, the parameters $(\theta_1, \theta_2) = (0, 0)$ and for the second $(\theta_1, \theta_2) = (1.5, -0.8)$. The series tend to be much smoother than those generated by the (0, 1, 1) process. The (0, 2, 2) models are useful in representing disturbances (such as Series C) in systems with a large degree of inertia. Both the (0, 1, 1) and the (0, 2, 2) models are special cases of the class

$$\nabla^d z_t = \theta(B) a_t \tag{4.3.1}$$

We call these models *integrated moving average* (IMA) processes, of order $(0, d, q)$, and consider their properties in the following section.

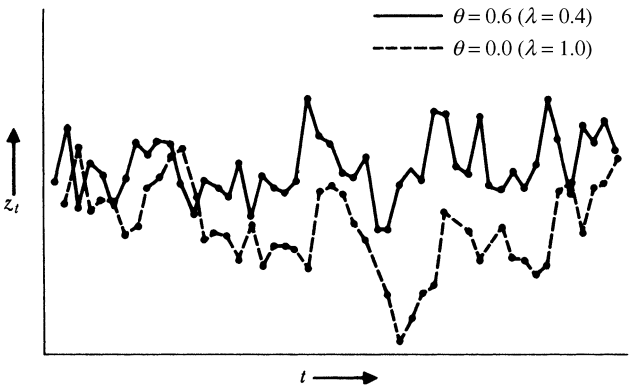

FIGURE 4.5 Two time series generated from IMA(0, 1, 1) models.

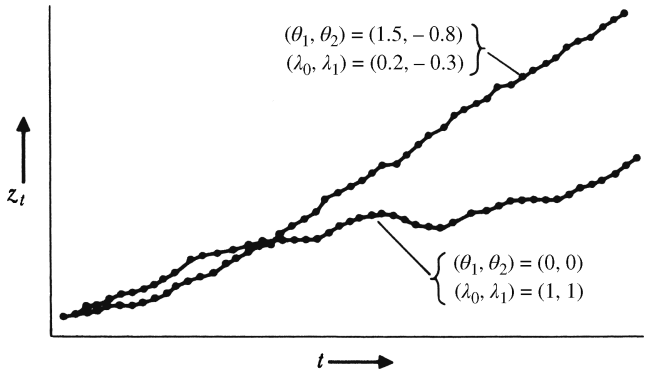

FIGURE 4.6 Two time series generated from IMA(0, 2, 2) models.

4.3.1 Integrated Moving Average Process of Order (0, 1, 1)

Difference Equation Form. The IMA(0, 1, 1) process

$$\nabla z_t = (1 - \theta B)a_t \quad -1 < \theta < 1$$

possesses useful representational capability, and we now study its properties in more detail. The model can be written in terms of the z's and the a's in the form

$$z_t = z_{t-1} + a_t - \theta a_{t-1} \tag{4.3.2}$$

Random Shock Form of Model. Alternatively, we can obtain z_t in terms of the a's alone by summing on both sides of (4.3.2). Before doing this, there is some advantage in expressing the right-hand operator in terms of ∇ rather than B. Thus, we can write

$$1 - \theta B = (1 - \theta)B + (1 - B) = (1 - \theta)B + \nabla = \lambda B + \nabla$$

where $\lambda = 1 - \theta$, and the invertibility region in terms of λ is defined by $0 < \lambda < 2$. Hence

$$\nabla z_t = \lambda a_{t-1} + \nabla a_t$$

Relative to some time origin $k < t$, applying the finite summation operator $S_{t-k} = 1 + B + \cdots + B^{t-k-1} = (1 - B^{t-k})/(1 - B)$, we obtain

$$(1 - B^{t-k})z_t = \lambda S_{t-k} a_{t-1} + (1 - B^{t-k})a_t \tag{4.3.3}$$

so that

$$z_t = a_t + \lambda(a_{t-1} + a_{t-2} + \cdots + a_{k+1}) + (z_k - \theta a_k) \tag{4.3.4}$$

In comparison to $z_t = \sum_{j=0}^{t-k-1} \psi_j a_{t-j} + C_k(t-k)$, the weights are $\psi_0 = 1, \psi_j = \lambda$ for $j \geq 1$. Also, the complementary function is $C_k(t-k) = z_k - \theta a_k = b_0^{(k)}$ (a ''constant'' b_0 for each k), which is the solution of the difference equation (1 - B)$C_k(t-k) = 0$. Moreover, in the infinite form $z_t = a_t + \lambda \sum_{j=1}^{\infty} a_{t-j}$, we may identify $b_0^{(k)}$ with $\lambda \sum_{j=t-k}^{\infty} a_{t-j}$. For this model, then, the complementary function is simply a constant (i.e., a polynomial in t of degree zero) representing the current ''level'' of the process and associated with the particular origin of

reference k. If the origin is changed from $k-1$ to k, then b_0 is ''updated'' according to

$$b_0^{(k)} = b_0^{(k-1)} + \lambda a_k$$

since using (4.3.2), $b_0^{(k)} = z_k + (\lambda - 1)a_k = z_{k-1} - \theta a_{k-1} + \lambda a_k$.

Inverted Form of Model. Finally, we can consider the model in the form

$$\pi(B)z_t = a_t$$

or equivalently, in the form

$$z_t = \sum_{j=1}^{\infty} \pi_j z_{t-j} + a_t = \bar{z}_{t-1}(\pi) + a_t$$

where $\bar{z}_{t-1}(\pi)$ is a weighted moving average of previous values of the process.

Using (4.2.21), the π weights for the IMA(0, 1, 1) process are given by

$$(1 - \theta B)\pi(B) = 1 - B$$

that is,

$$\begin{aligned}\pi(B) = \frac{1-B}{1-\theta B} &= \frac{1 - \theta B - (1-\theta)B}{1-\theta B} \\ &= 1 - (1-\theta)(B + \theta B^2 + \theta^2 B^3 + \cdots)\end{aligned}$$

so that

$$\pi_j = (1-\theta)\theta^{j-1} = \lambda(1-\lambda)^{j-1} \quad j \geq 1$$

Thus, the process may be written as

$$z_t = \bar{z}_{t-1}(\lambda) + a_t \tag{4.3.5}$$

The weighted moving average of previous values of the process

$$\bar{z}_{t-1}(\lambda) = \lambda \sum_{j=1}^{\infty} (1-\lambda)^{j-1} z_{t-j} \tag{4.3.6}$$

is, in this case, an *exponentially weighted moving average* (EWMA). This term reflects the fact that the weights

$$\lambda \quad \lambda(1-\lambda) \quad \lambda(1-\lambda)^2 \quad \lambda(1-\lambda)^3 \cdots$$

fall off exponentially (i.e., as a geometric progression) as j increases. The weight function for an IMA(0, 1, 1) process, with $\lambda = 0.4$ (or $\theta = 0.6$), is shown in Figure 4.7.

Although the invertibility condition is satisfied for $0 < \lambda < 2$, in practice, we are most often concerned with values of λ between zero and 1 (i.e., $0 < \theta < 1$). We note that if λ had a value equal to 1, the weight function would consist of a single spike ($\pi_1 = 1, \pi_j = 0$ for $j > 1$). As the value λ approaches zero, the exponential weights die out more and more slowly and the EWMA stretches back further into past values of the process. Finally, with

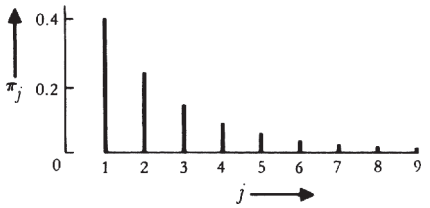

FIGURE 4.7 The π weights for an IMA process of order (0, 1, 1) with $\lambda = 1 - \theta = 0.4$.

$\lambda = 0$ and $\theta = 1$, the model $(1 - B)z_t = (1 - B)a_t$ is equivalent to $z_t = \theta_0 + a_t$, with θ_0 being given by the mean of all past values.

Since $b_0^{(k)} = z_k - \theta a_k = z_{k+1} - a_{k+1}$, or $z_{k+1} = b_0^{(k)} + a_{k+1}$, on comparison with (4.3.5) it follows that for this process, the complementary function $b_0^{(k)} = C_k(t - k)$ in (4.3.4) is

$$b_0^{(k)} = \bar{z}_k(\lambda) \tag{4.3.7}$$

an exponentially weighted average of values up to the origin k. In fact, (4.3.4) may be written as

$$z_t = \bar{z}_k(\lambda) + \lambda \sum_{j=1}^{t-k-1} a_{t-j} + a_t$$

We have seen that the complementary function $C_k(t - k)$ can be thought of as telling us what is known about the future value of the process at time t, based on knowledge of the past when *we are standing at time k*. For the IMA(0, 1, 1) process, this takes the form of information about the ''level'' or location of the process $b_0^{(k)} = \bar{z}_k(\lambda)$. At time k, our knowledge of the future behavior of the process is that it will diverge from this level in accordance with the ''random walk'' represented by $\lambda \sum_{j=1}^{t-k-1} a_{t-j} + a_t$, whose expectation is zero and whose behavior we cannot predict. As soon as a new observation is available, that is, as soon as we move our origin to time $k + 1$, the level will be updated to $b_0^{(k+1)} = \bar{z}_{k+1}(\lambda)$.

Important Properties of the IMA(0, 1, 1) Process. Since the process is nonstationary, it does not vary in a stable manner about a fixed mean. However, the exponentially weighted moving average $\bar{z}_t(\lambda)$ can be regarded as measuring the local level of the process at time t. From its definition (4.3.6), we obtain the well-known recursion formula for the EWMA:

$$\bar{z}_t(\lambda) = \lambda z_t + (1 - \lambda)\bar{z}_{t-1}(\lambda) \tag{4.3.8}$$

This expression shows that for the IMA(0, 1, 1) model, each new level is arrived at by interpolating between the new observation and the previous level. If λ is equal to unity, $\bar{z}_t(\lambda) = z_t$ which would ignore all evidence concerning location coming from previous observations. On the other hand, if λ had some value close to zero, $\bar{z}_1(\lambda)$ would rely heavily on the previous value $\bar{z}_{t-1}(\lambda)$, which would have weight $1 - \lambda$. Only the small weight λ would be given to the new observation.

Now consider the two equations

$$\begin{aligned} z_t &= \bar{z}_{t-1}(\lambda) + a_t \\ \bar{z}_t(\lambda) &= \bar{z}_{t-1}(\lambda) + \lambda a_t \end{aligned} \tag{4.3.9}$$

the latter being obtained by substituting (4.3.5) in (4.3.8) and is also directly derivable from (4.3.7).

It was pointed out by Muth (1960) that the two equations (4.3.9) provide a useful way of thinking about the generation of the process. The first equation shows how, with the level of the system at $\bar{z}_{t-1}(\lambda)$, a shock a_t is added at time t and produces the value z_t. However, the second equation shows that only a proportion λ of the shock is actually absorbed into the level and has a lasting influence, the remaining proportion $\theta = 1 - \lambda$ of the shock being dissipated. Now a new level $\bar{z}_t(\lambda)$ having been established by the absorption of a_t, a new shock a_{t+1} enters the system at time $t + 1$. Equations (4.3.9), with subscripts increased by unity, will then show how this shock produces z_{t+1} and how a proportion λ of it is absorbed into the system to produce the new level $\bar{z}_{t+1}(\lambda)$, and so on.

Equation (4.3.4) can be used to obtain variance and correlation features of the IMA(0, 1, 1) process directly. For example, with reference to the origin k and treating the initializing function $b_0^{(k)}$ as constant, we find that

$$\text{var}[z_t] = \sigma_a^2[1 + (t - k - 1)\lambda^2] \tag{4.3.10}$$

which does not converge as t increases. We might also view this variance as, essentially, the variance of the difference $z_t - z_k$, treating $a_k = 0$ in (4.3.4). In particular, in the case of a random walk process, $z_t = z_{t-1} + a_t$, we have $\lambda = 1$, and this variance function grows proportionally with $t - k$, whereas for more common situations with $0 < \lambda < 1$ (i.e., $0 < \theta < 1$) and especially for λ close to zero, the variance function of $z_t - z_k$ grows much more slowly with $t - k$. In addition, for $s > 0$, $\text{cov}[z_t, z_{t+s}] = \sigma_a^2[\lambda + (t - k - 1)\lambda^2]$, which implies that $\text{corr}[z_t, z_{t+s}]$ will be close to 1 for $t - k$ large relative to s (and λ not close to zero). Hence, it follows that adjacent values of the process will be highly positively correlated, so the process will tend to exhibit rather smooth behavior (unless λ is close to zero).

The properties of the IMA(0, 1, 1) process with deterministic drift

$$\nabla z_t = \theta_0 + (1 - \theta_1 B)a_t$$

are discussed in Appendix A4.2.

4.3.2 Integrated Moving Average Process of Order (0, 2, 2)

Difference Equation Form. The IMA(0, 2, 2) process

$$\nabla^2 z_t = (1 - \theta_1 B - \theta_2 B^2)a_t \tag{4.3.11}$$

can be used to represent series exhibiting stochastic trends (e.g., see Fig. 4.6), and we now study its general properties within the invertibility region:

$$-1 < \theta_2 < 1 \quad \theta_2 + \theta_1 < 1 \quad \theta_2 - \theta_1 < 1$$

Proceeding as before, z_t can be written explicitly in terms of z's and a's as

$$z_t = 2z_{t-1} - z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

Alternatively, we can rewrite the right-hand operator in terms of differences:

$$1 - \theta_1 B - \theta_2 B^2 = (\lambda_0 \nabla + \lambda_1)B + \nabla^2$$

and on equating coefficients, we find expressions for the θ's in terms of the λ's, and vice versa, as follows:

$$\begin{aligned} \theta_1 &= 2 - \lambda_0 - \lambda_1 & \lambda_0 &= 1 + \theta_2 \\ \theta_2 &= \lambda_0 - 1 & \lambda_1 &= 1 - \theta_1 - \theta_2 \end{aligned} \tag{4.3.12}$$

The IMA(0, 2, 2) model may then be rewritten as

$$\nabla^2 z_t = (\lambda_0 \nabla + \lambda_1) a_{t-1} + \nabla^2 a_t \tag{4.3.13}$$

There is an important advantage in using this form of the model, as compared with (4.3.11). This stems from the fact that if we set $\lambda_1 = 0$ in (4.3.13), we obtain

$$\nabla z_t = [1 - (1 - \lambda_0)B]a_t$$

which corresponds to a (0, 1, 1) process, with $\theta = 1 - \lambda_0$. However, if we set $\theta_2 = 0$ in (4.3.11), we obtain

$$\nabla^2 z_t = (1 - \theta_1 B)a_t$$

As will be shown in Chapter 5, for a series generated by the (0, 2, 2) model, the optimal forecasts lie along a straight line, the level and *slope* of which are continually updated as new data become available. By contrast, a series generated by a (0, 1, 1) model can supply no information about slope but only about a continually updated level. It can be an important question whether a linear trend, as well as the level, can be forecasted and updated. When the choice is between these two models, this question turns on whether or not λ_1 in (4.3.13) is zero.

The invertibility region for an IMA(0, 2, 2) process is the same as that given for an MA(2) process in Chapter 3. It may be written in terms of the θ's and λ's as follows:

$$\begin{aligned} \theta_2 + \theta_1 &< 1 & 0 < 2\lambda_0 + \lambda_1 &< 4 \\ \theta_2 - \theta_1 &< 1 & \lambda_1 &> 0 \\ -1 < \theta_2 &< 1 & \lambda_0 &> 0 \end{aligned} \tag{4.3.14}$$

The triangular region for the θ's was shown in Figure 3.6 and the corresponding region for the λ's is shown in Figure 4.8.

Truncated and Infinite Random Shock Forms of Model. On applying the finite double summation operator $S^{(2)}_{t-k}$, relative to a time origin k, to (4.3.13), we find that

$$\begin{aligned} [1 - B^{t-k} - (t-k)B^{t-k}(1-B)]z_t = {} & [\lambda_0(S_{t-k} - (t-k)B^{t-k}) + \lambda_1 S^{(2)}_{t-k}]a_{t-1} \\ & + [1 - B^{t-k} - (t-k)B^{t-k}(1-B)]a_t \end{aligned}$$

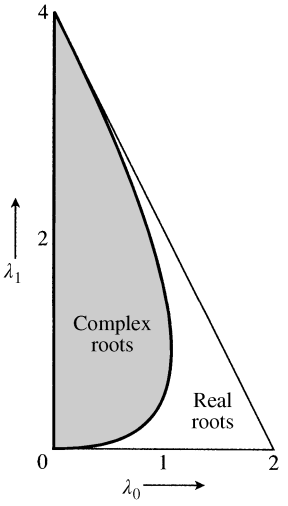

FIGURE 4.8 Invertibility region for parameters λ_0 and λ_1 of an IMA(0, 2, 2) process.

Hence, we obtain the truncated form of the random shock model as

$$\begin{aligned} z_t &= \lambda_0 S_{t-k-1} a_{t-1} + \lambda_1 S^{(2)}_{t-k-1} a_{t-1} + a_t + b_0^{(k)} + b_1^{(k)}(t-k) \\ &= \lambda_0 \sum_{j=1}^{t-k-1} a_{t-j} + \lambda_1 \sum_{j=1}^{t-k-1} j a_{t-j} + a_t + C_k(t-k) \end{aligned} \tag{4.3.15}$$

So, for this process, the ψ weights are

$$\psi_0 = 1 \quad \psi_1 = (\lambda_0 + \lambda_1) \cdots \quad \psi_j = (\lambda_0 + j\lambda_1) \cdots$$

The complementary function is the solution of

$$(1 - B)^2 C_k(t-k) = 0$$

that is,

$$C_k(t-k) = b_0^{(k)} + b_1^{(k)}(t-k) \tag{4.3.16}$$

which is a polynomial in $(t-k)$ of degree 1 whose coefficients depend on the location of the origin k. From (4.3.15), we find that these coefficients are given explicitly as

$$\begin{aligned} b_0^{(k)} &= z_k - (1 - \lambda_0) a_k \\ b_1^{(k)} &= z_k - z_{k-1} - (1 - \lambda_1) a_k + (1 - \lambda_0) a_{k-1} \end{aligned}$$

Also, by considering the differences $b_0^{(k)} - b_0^{(k-1)}$ and $b_1^{(k)} - b_1^{(k-1)}$, it follows that if the origin is updated from $k-1$ to k, then b_0 and b_1 are updated according to

$$\begin{aligned} b_0^{(k)} &= b_0^{(k-1)} + b_1^{(k-1)} + \lambda_0 a_k \\ b_1^{(k)} &= b_1^{(k-1)} + \lambda_1 a_k \end{aligned} \tag{4.3.17}$$

We see that when this model is appropriate, our expectation of the future behavior of the series, judged from origin k, would be represented by the straight line (4.3.16), having location $b_0^{(k)}$ and slope $b_1^{(k)}$. In practice, the process will, by time t, have diverged from this line because of the influence of the random component

$$\lambda_0 \sum_{j=1}^{t-k-1} a_{t-j} + \lambda_1 \sum_{j=1}^{t-k-1} j a_{t-j} + a_t$$

which at time k is unpredictable. Moreover, on moving from origin $k-1$ to origin k, the intercept and slope are updated according to (4.3.17).

Informally, through (4.3.15) we may also obtain the infinite random shock form as

$$z_t = \lambda_0 \sum_{j=1}^{\infty} a_{t-j} + \lambda_1 \sum_{j=1}^{\infty} j a_{t-j} + a_t = \lambda_0 S a_{t-1} + \lambda_1 S^2 a_{t-1} + a_t \tag{4.3.18}$$

So by comparison with (4.3.15), the complementary function can be represented informally as

$$C_k(t-k) = \lambda_0 \sum_{j=t-k}^{\infty} a_{t-j} + \lambda_1 \sum_{j=t-k}^{\infty} j a_{t-j} = b_0^{(k)} + b_1^{(k)}(t-k)$$

By writing the second infinite sum above in the form

$$\sum_{j=t-k}^{\infty} j a_{t-j} = (t-k) \sum_{j=t-k}^{\infty} a_{t-j} + \sum_{j=t-k}^{\infty} [j-(t-k)] a_{t-j}$$

we see that the coefficients $b_0^{(k)}$ and $b_1^{(k)}$ can be associated with

$$b_0^{(k)} = \lambda_0 S a_k + \lambda_1 S^2 a_{k-1} = (\lambda_0 - \lambda_1) S a_k + \lambda_1 S^2 a_k$$
$$b_1^{(k)} = \lambda_1 S a_k$$

Inverted Form of Model. Finally, we consider the model in the inverted form:

$$z_t = \sum_{j=1}^{\infty} \pi_j z_{t-j} + a_t = \bar{z}_{t-1}(\pi) + a_t$$

Using (4.2.22), we find on equating coefficients in

$$1 - 2B + B^2 = (1 - \theta_1 B - \theta_2 B^2)(1 - \pi_1 B - \pi_2 B^2 - \cdots)$$

that the π weights of the IMA(0, 2, 2) process are

$$\begin{aligned} \pi_1 &= 2 - \theta_1 = \lambda_0 + \lambda_1 \\ \pi_2 &= \theta_1(2 - \theta_1) - (1 + \theta_2) = \lambda_0 + 2\lambda_1 - (\lambda_0 + \lambda_1)^2 \\ (1 - \theta_1 B - \theta_2 B^2)\pi_j &= 0 \qquad j \geq 3 \end{aligned} \tag{4.3.19}$$

where B now operates on j.

If the roots of the characteristic equation $1 - \theta_1 B - \theta_2 B^2 = 0$ are real, the π weights are a mixture of two damped exponentials. If the roots are complex, the weights follow a

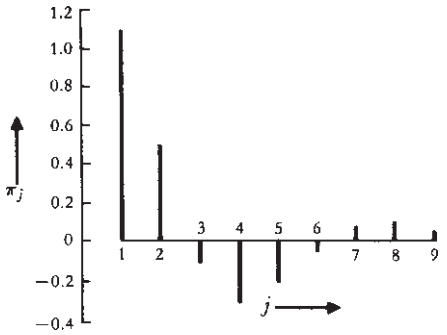

FIGURE 4.9 The π weights for an IMA(0, 2, 2) process with $\lambda_0 = 0.5$, $\lambda_1 = 0.6$.

damped sine wave. Figure 4.9 shows the weights for a process with $\theta_1 = 0.9$ and $\theta_2 = -0.5$, that is, $\lambda_0 = 0.5$ and $\lambda_1 = 0.6$. For these parameter values, the characteristic equation has complex roots (the discriminant $\theta_1^2 + 4\theta_2 = -1.19$ is less than zero). Hence, the weights in Figure 4.9 follow a damped sine wave, as expected.

4.3.3 General Integrated Moving Average Process of Order $(0, d, q)$

Difference Equation Form. The general integrated moving average process of order $(0, d, q)$ is

$$\nabla^d z_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) a_t = \theta(B) a_t \tag{4.3.20}$$

where the zeros of $\theta(B)$ must lie outside the unit circle for the process to be invertible. This model may be written explicitly in terms of past z's and a's in the form

$$z_t = d z_{t-1} - \frac{1}{2} d(d-1) z_{t-2} + \cdots + (-1)^{d+1} z_{t-d} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

Random Shock Form of Model. To obtain z_t in terms of the a_t's, we write the right-hand operator in (4.3.20) in terms of $\nabla = 1 - B$. In this way, we obtain

$$(1 - \theta_1 B - \cdots - \theta_q B^q) = (\lambda_{d-q} \nabla^{q-1} + \cdots + \lambda_0 \nabla^{d-1} + \cdots + \lambda_{d-1}) B + \nabla^d \tag{4.3.21}$$

where, as before, the λ's may be written explicitly in terms of the θ's, by equating coefficients of B.

On substituting (4.3.21) in (4.3.20) and summing d times, informally, we obtain

$$z_t = (\lambda_{d-q} \nabla^{q-d-1} + \cdots + \lambda_0 S + \cdots + \lambda_{d-1} S^d) a_{t-1} + a_t \tag{4.3.22}$$

Thus, for $q > d$, we notice that in addition to the d sums, we pick up $q - d$ additional terms $\nabla^{q-d-1} a_{t-1}, \ldots$ involving $a_{t-1}, a_{t-2}, \ldots, a_{t+d-q}$.

If we write this solution in terms of finite sums of a's entering the system after some origin k, we obtain the same form of equation, but with an added complementary function, which is the solution of

$$\nabla^d C_k(t-k) = 0$$

that is, the polynomial

$$C_k(t-k) = b_0^{(k)} + b_1^{(k)}(t-k) + b_2^{(k)}(t-k)^2 + \cdots + b_{d-1}^{(k)}(t-k)^{d-1}$$

As before, the complementary function $C_k(t-k)$ represents the finite behavior of the process, which is predictable at time k. Similarly, the coefficients $b_j^{(k)}$ may be expressed, informally, in terms of the infinite sums up to origin k, that is, $Sa_k, S^2a_k, \ldots, S^da_k$. Accordingly, we can discover how the coefficients $b_j^{(k)}$ change as the origin is changed, from $k-1$ to k.

Inverted Form of Model. Finally, the model can be expressed in the inverted form

$$\pi(B)z_t = a_t$$

or

$$z_t = \bar{z}_{t-1}(\pi) + a(t)$$

The π weights may be obtained by equating coefficients in (4.2.22), that is,

$$(1-B)^d = (1-\theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)(1-\pi_1 B - \pi_2 B^2 - \cdots) \tag{4.3.23}$$

This expression implies that for j greater than the larger of d and q, the π weights satisfy the homogeneous difference equation

$$\theta(B)\pi_j = 0$$

defined by the moving average operator. Hence, for sufficiently large j, the weights π_j follow a mixture of damped exponentials and sine waves.

IMA Process of Order (0, 2, 3). One final special case of sufficient interest to merit comment is the IMA process of order (0, 2, 3):

$$\nabla^2 z_t = (1-\theta_1 B - \theta_2 B^2 - \theta_3 B^3)a_t$$

Proceeding as before, if we apply the finite double summation operator, this model can be written in truncated random shock form as

$$z_t = \lambda_{-1}a_{t-1} + \lambda_0 \sum_{j=1}^{t-k-1} a_{t-j} + \lambda_1 \sum_{j=1}^{t-k-1} j a_{t-j} + a_t + b_0^{(k)} + b_1^{(k)}(t-k)$$

where the relations between the λ's and θ's are

$$\begin{aligned} \theta_1 &= 2 - \lambda_{-1} - \lambda_0 - \lambda_1 & \lambda_{-1} &= -\theta_3 \\ \theta_2 &= \lambda_0 - 1 + 2\lambda_{-1} & \lambda_0 &= 1 + \theta_2 + 2\theta_3 \\ \theta_3 &= -\lambda_{-1} & \lambda_1 &= 1 - \theta_1 - \theta_2 - \theta_3 \end{aligned}$$

Alternatively, it can be written, informally, in the infinite integrated form as

$$z_t = \lambda_{-1}a_{t-1} + \lambda_0 S a_{t-1} + \lambda_1 S^2 a_{t-1} + a_t$$

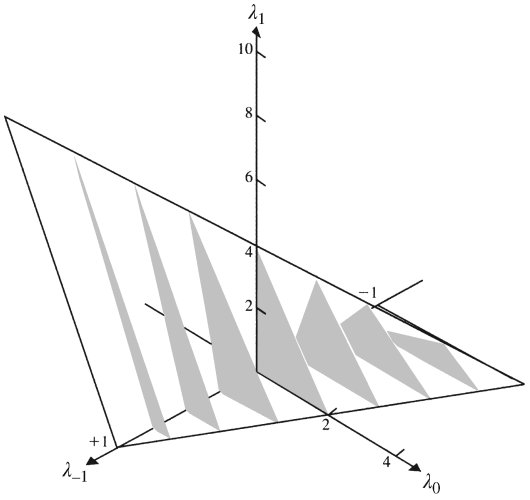

FIGURE 4.10 Invertibility region for parameters λ_{-1}, λ_0, and λ_1 and of an IMA(0, 2, 3) process.

Finally, the invertibility region is defined by

$$\begin{aligned} \theta_1 + \theta_2 + \theta_3 &< 1 & \lambda_1 &> 0 \\ -\theta_1 + \theta_2 - \theta_3 &< 1 & 2\lambda_0 + \lambda_1 &< 4(1 - \lambda_{-1}) \\ \theta_3(\theta_3 - \theta_1) - \theta_2 &< 1 & \lambda_0(1 + \lambda_{-1}) &> -\lambda_1\lambda_{-1} \\ -1 < \theta_3 &< 1 & -1 < \lambda_{-1} &< 1 \end{aligned}$$

as is shown in Figure 4.10.

In Chapter 5, we show how forecasts of future values of a time series can be generated in an optimal manner when the model is an ARIMA process. In studying these forecasts, we make considerable use of the various model forms discussed in this chapter.

APPENDIX A4.1 LINEAR DIFFERENCE EQUATIONS

In this book, we are often concerned with linear difference equations. In particular, the ARIMA model relates an output z_t to an input a_t in terms of the difference equation

$$\begin{aligned} &z_t - \varphi_1 z_{t-1} - \varphi_2 z_{t-2} - \cdots - \varphi_{p'} z_{t-p'} \\ &\quad = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \end{aligned} \tag{A4.1.1}$$

where $p' = p + d$.

Alternatively, we may write (A4.1.1) as

$$\varphi(B) z_t = \theta(B) a_t$$

where

$$\begin{aligned} \varphi(B) &= 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_{p'} B^{p'} \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \end{aligned}$$

We now derive an expression for the general solution of the difference equation (A4.1.1) relative to an origin $k < t$.

1. We show that the general solution may be written as

$$z_t = C_k(t-k) + I_k(t-k)$$

   where $C_k(t-k)$ is the complementary function and $I_k(t-k)$ is a ''particular integral.''
2. We then derive a general expression for the complementary function $C_k(t-k)$.
3. Finally, we derive a general expression for a particular integral $I_k(t-k)$.

General Solution. The argument is identical to that for the solution of linear differential or linear algebraic equations. Suppose that z_t' is any particular solution of

$$\varphi(B)z_t = \theta(B)a_t \tag{A4.1.2}$$

that is, it satisfies

$$\varphi(B)z_t' = \theta(B)a_t \tag{A4.1.3}$$

On subtracting (A4.1.3) from (A4.1.2), we obtain

$$\varphi(B)(z_t - z_t') = 0$$

Thus $z_t'' = z_t - z_t'$ satisfies

$$\varphi(B)z_t'' = 0 \tag{A4.1.4}$$

Now

$$z_t = z_t' + z_t''$$

and hence the general solution of (A4.1.2) is the sum of the complementary function z_t'', which is the general solution of the homogeneous difference equation (A4.1.4), and a particular integral z_t', which is any particular solution of (A4.1.2). Relative to any origin $k < t$, we denote the complementary function z_t'' by $C_k(t-k)$ and the particular integral z_t' by $I_k(t-k)$.

Evaluation of the Complementary Function.

Distinct Roots. Consider the homogeneous difference equation

$$\varphi(B)z_t = 0 \tag{A4.1.5}$$

where

$$\varphi(B) = (1 - G_1B)(1 - G_2B)\cdots(1 - G_{p'}B) \tag{A4.1.6}$$

and where we assume in the first instance that $G_1, G_2, \ldots, G_{p'}$ are *distinct*. Then, it is shown below that the general solution of (A4.1.5) at time t, when the series is referred to an origin

at time k, is

$$z_t = A_1 G_1^{t-k} + A_2 G_2^{t-k} + \cdots + A_{p'} G_{p'}^{t-k} \tag{A4.1.7}$$

where the A_i's are constants. Thus, a real root G of $\varphi(B) = 0$ contributes a damped exponential term G^{t-k} to the complementary function. A pair of complex roots contributes a damped sine wave term $D^{t-k}\sin(2\pi f_0 t + F)$.

To see that the expression given in (A4.1.7) does satisfy (A4.1.5), we can substitute (A4.1.7) in (A4.1.5) to give

$$\varphi(B)(A_1 G_1^{t-k} + A_2 G_2^{t-k} + \cdots + A_{p'} G_{p'}^{t-k}) = 0 \tag{A4.1.8}$$

Now consider

$$\begin{aligned}\varphi(B) G_i^{t-k} &= (1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_{p'} B^{p'}) G_i^{t-k} \\ &= G_i^{t-k-p'}(G_i^{p'} - \varphi_1 G_i^{p'-1} - \cdots - \varphi_{p'})\end{aligned}$$

We see that $\varphi(B) G_i^{t-k}$ vanishes for each value of i if

$$G_i^{p'} - \varphi_1 G_i^{p'-1} - \cdots - \varphi_{p'} = 0$$

that is, if $B_i = 1/G_i$ is a root of $\varphi(B) = 0$. Now, since (A4.1.6) implies that the roots of $\varphi(B) = 0$ are $B_i = 1/G_i$, it follows that $\varphi(B) G_i^{t-k}$ is zero for all i and hence (A4.1.8) holds, confirming that (A4.1.7) is a general solution of (A4.1.5).

To prove (A4.1.7) directly, consider the special case of the second-order equation:

$$(1 - G_1 B)(1 - G_2 B) z_t = 0$$

which we can write as

$$(1 - G_1 B) y_t = 0 \tag{A4.1.9}$$

where

$$y_t = (1 - G_2 B) z_t \tag{A4.1.10}$$

Now (A4.1.9) implies that

$$y_t = G_1 y_{t-1} = G_1^2 y_{t-2} = \cdots = G_1^{t-k} y_k$$

and hence

$$y_t = D_1 G_1^{t-k}$$

where $D_1 = y_k$ is a constant determined by the starting value y_k. Hence (A4.1.10) may be written as

$$\begin{aligned} z_t &= G_2 z_{t-1} + D_1 G_1^{t-k} \\ &= G_2(G_2 z_{t-2} + D_1 G_1^{t-k-1}) + D_1 G_1^{t-k} \\ &\quad \vdots \\ &= G_2^{t-k} z_k + D_1(G_1^{t-k} + G_2 G_1^{t-k-1} + \cdots + G_2^{t-k-1} G_1) \\ &= G_2^{t-k} z_k + \frac{D_1}{1 - G_2/G_1}(G_1^{t-k} - G_2^{t-k}) \\ &= A_1 G_1^{t-k} + A_2 G_2^{t-k} \end{aligned} \tag{A4.1.11}$$

where A_1, A_2 are constants determined by the starting values of the series. By an extension of the argument above, it may be shown that the general solution of (A4.1.5), when the roots of $\varphi(B) = 0$ are distinct, is given by (A4.1.7).

Equal Roots. Suppose that $\varphi(B) = 0$ has d equal roots G_0^{-1}, so that $\varphi(B)$ contains a factor $(1 - G_0 B)^d$. In particular, consider the solution (A4.1.11) for the second-order equation when both G_1 and G_2 are equal to G_0. Then, (A4.1.11) reduces to

$$z_t = G_0^{t-k} z_k + D_1 G_0^{t-k}(t-k)$$

or

$$z_t = [A_0 + A_1(t-k)] G_0^{t-k}$$

In general, if there are d equal roots G_0, it may be verified by direct substitution in (A4.1.5) that the general solution is

$$\begin{aligned} z_t = [&A_0 + A_1(t-k) + A_2(t-k)^2 + \cdots \\ &+ A_{d-1}(t-k)^{d-1}] G_0^{t-k} \end{aligned} \tag{A4.1.12}$$

In particular, when the equal roots G_0 are all equal to unity as in the IMA $(0, d, q)$ process, the solution is

$$z_t = A_0 + A_1(t-k) + A_2(t-k)^2 + \cdots + A_{d-1}(t-k)^{d-1} \tag{A4.1.13}$$

that is, a polynomial in $t - k$ of degree $d - 1$.

In general, when $\varphi(B)$ factors according to

$$(1 - G_1 B)(1 - G_2 B) \cdots (1 - G_p B)(1 - G_0 B)^d$$

the complementary function is

$$C_k(t-k) = G_0^{t-k} \sum_{j=0}^{d-1} A_j (t-k)^j + \sum_{i=1}^{p} D_i G_i^{t-k} \tag{A4.1.14}$$

Thus, in general, the complementary function consists of a mixture of damped exponential terms G^{t-k}, polynomial terms $(t-k)^j$, damped sine wave terms of the form $D^{t-k} \sin(2\pi f_0 t + F)$, and combinations of these functions.

Evaluation of the "Particular Integral". We now show that a particular integral $I_k(s-k)$, satisfying

$$\varphi(B)I_k(t-k) = \theta(B)a_t \quad t-k > q \tag{A4.1.15}$$

is a function defined as follows:

$$\begin{aligned} I_k(s-k) &= 0 \quad s \le k \\ I_k(1) &= a_{k+1} \\ I_k(2) &= a_{k+2} + \psi_1 a_{k+1} \\ &\vdots \\ I_k(t-k) &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots + \psi_{t-k-1} a_{k+1} \quad t > k \end{aligned} \tag{A4.1.16}$$

where the ψ weights are those appearing in the form (4.2.3) of the model. Thus, the ψ weights satisfy

$$\varphi(B)\psi(B)a_t = \theta(B)a_t \tag{A4.1.17}$$

Now the terms on the left-hand side of (A4.1.17) may be set out as follows:

$$\begin{array}{r|l} a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots + \psi_{t-k-1} a_{k+1} & +\psi_{t-k} a_k + \cdots \\ -\varphi_1(a_{t-1} + \psi_1 a_{t-2} + \cdots + \psi_{t-k-2} a_{k+1} & +\psi_{t-k-1} a_k + \cdots) \\ -\varphi_2(\quad \cdots & \cdots) \\ \vdots & \vdots \\ -\varphi_{p'}(a_{t-p'} + \cdots + \psi_{t-k-p'-1} a_{k+1} & +\psi_{t-k-p'} a_k + \cdots \end{array} \tag{A4.1.18}$$

Since the right-hand side of (A4.1.17) is

$$a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

it follows that the first $q+1$ columns in this array sum to $a_t, -\theta_1 a_{t-1}, \ldots, -\theta_q a_{t-q}$. Now the left-hand term in (A4.1.15), where $I_k(s-k)$ is given by (A4.1.16), is equal to the sum of the terms in the first $(t-k)$ columns of the array, that is, those to the left of the vertical line. Therefore, if $t-k < q$, that is, the vertical line is drawn after $q+1$ columns, the sum of all terms up to the vertical line is equal to $\theta(B)a_t$. This shows that (A4.1.16) is a particular integral of the difference equation.

Example. Consider the IMA(0, 1, 1) process

$$z_t - z_{t-1} = a_t - \theta a_{t-1} \tag{A4.1.19}$$

for which $\psi_j = 1-\theta$ for $j \ge 1$. Then

$$\begin{aligned} I_k(0) &= 0 \\ I_k(1) &= a_{k+1} \\ &\vdots \end{aligned} \tag{A4.1.20}$$

$$I_k(t-k) = a_t + (1-\theta)\sum_{j=1}^{t-k-1} a_{t-j} \quad t-k>1$$

Now if $z_t = I_k(t-k)$ is a solution of (A4.1.19), then

$$I_k(t-k) - I_k(t-k-1) = a_t - \theta a_{t-1}$$

and as is easily verified, while this is not satisfied by (A4.1.20) for $t-k=1$, it is satisfied by (A4.1.20) for $t-k>1$, that is, for $t-k>q$.

APPENDIX A4.2 IMA(0, 1, 1) PROCESS WITH DETERMINISTIC DRIFT

The general model $\phi(B)\nabla^d z_t = \theta_0 + \theta(B)a_t$ can also be written as

$$\phi(B)\nabla^d z_t = \theta(B)\varepsilon_t$$

with the shocks ε_t having a nonzero mean $\xi = \theta_0/(1-\theta_1-\cdots-\theta_q)$. For example, the IMA(0, 1, 1) model is then

$$\nabla z_t = (1-\theta B)\varepsilon_t$$

with $E[\varepsilon_t] = \xi = \theta_0/(1-\theta)$. In this form, z_t could represent, for example, the outlet temperature from a reactor when heat was being supplied from a heating element at a fixed rate. Now if

$$\varepsilon_t = \xi + a_t \tag{A4.2.1}$$

where a_t is white noise with zero mean, then with reference to a time origin k, the integrated form of the model is

$$z_t = b_0^{(k)} + \lambda\sum_{j=1}^{t-k-1}\varepsilon_{t-j} + \varepsilon_t \tag{A4.2.2}$$

with $\lambda = 1-\theta$. Substituting for (A4.2.1) in (A4.2.2), the model written in terms of the a's is

$$z_t = b_0^{(k)} + \lambda\xi(t-k-1) + \xi + \lambda\sum_{j=1}^{t-k-1} a_{t-j} + a_t \tag{A4.2.3}$$

Thus, we see that z_t contains a deterministic slope or drift due to the term $\lambda\xi(t-k-1)$, with the slope of the deterministic linear trend equal to $\lambda\xi = \theta_0$. Moreover, if we denote the ''level'' of the process at time $t-1$ by l_{t-1}, where

$$z_t = l_{t-1} + a_t$$

we see that the level is changed from time $t-1$ to time t, according to

$$l_t = l_{t-1} + \lambda\xi + \lambda a_t$$

The change in the level, thus, contains a deterministic component $\lambda\xi = \theta_0$, as well as a stochastic component λa_t.

APPENDIX A4.3 ARIMA PROCESSES WITH ADDED NOISE

In this appendix, we consider the effect of added noise (e.g., measurement error) to a general ARIMA(p, d, q) process. The results are also relevant to determine the nature of the reduced form ARIMA model of an observed process in structural component models (see Section 9.4), in which an observed series Z_t is presumed to be represented as the sum of two unobservable component processes that follow specified ARIMA models.

A4.3.1 Sum of Two Independent Moving Average Processes

As a necessary preliminary to what follows, consider a stochastic process w_t, which is the sum of two *independent* moving average processes of orders q_1 and q_2, respectively. That is,

$$w_t = w_{1t} + w_{2t} = \theta_1(B)a_t + \theta_2(B)b_t \tag{A4.3.1}$$

where $\theta_1(B)$ and $\theta_2(B)$ are polynomials in B, of orders q_1 and q_2, and the white noise processes a_t and b_t have zero means, variances σ_a^2 and σ_b^2, and are mutually independent. Suppose that $q = \max(q_1, q_2)$; then since

$$\gamma_j(w) = \gamma_j(w_1) + \gamma_j(w_2)$$

it is clear that the autocovariance function $\gamma_j(w)$ for w_t must be zero for $j > q$. It follows that there exists a representation of w_t as a single MA(q) process:

$$w_t = \theta(B)u_t \tag{A4.3.2}$$

where u_t is a white noise process with mean zero and variance σ_u^2. Thus, the sum of two independent moving average processes is another moving average process, whose order is the same as that of the component process of higher order.

The parameters in the MA(q) model can be deduced by equating the autocovariances of w_t, as determined from the representation in (A4.3.1), with the autocovariances of the basic MA(q) model (A4.3.2), as given in Section 3.3.2. For an example, suppose that $w_{1t} = \theta_1(B)a_t = (1 - \theta_{1,1}B)a_t$ is MA(1) and $w_{2t} = \theta_2(B)b_t = (1 - \theta_{1,2}B - \theta_{2,2}B^2)b_t$ is MA(2), so that $w_t = \theta(B)u_t$ is MA(2) with

$$\begin{aligned} w_t &= (1 - \theta_{1,1}B)a_t + (1 - \theta_{1,2}B - \theta_{2,2}B^2)b_t \\ &= (1 - \theta_1 B - \theta_2 B^2)u_t \end{aligned}$$

The parameters of the MA(2) model for w_t can be determined by considering

$$\begin{aligned} \gamma_0(w) &= (1 + \theta_{1,1}^2)\sigma_a^2 + (1 + \theta_{1,2}^2 + \theta_{2,2}^2)\sigma_b^2 \equiv (1 + \theta_1^2 + \theta_2^2)\sigma_u^2 \\ \gamma_1(w) &= -\theta_{1,1}\sigma_a^2 + (-\theta_{1,2} + \theta_{1,2}\theta_{2,2})\sigma_b^2 \equiv (-\theta_1 + \theta_1\theta_2)\sigma_u^2 \\ \gamma_2(w) &= -\theta_{2,2}\sigma_b^2 \equiv -\theta_2\sigma_u^2 \end{aligned}$$

and solving for θ_1, θ_2, and σ_u^2 in terms of given values for the autocovariances $\gamma_0(w)$, $\gamma_1(w)$, $\gamma_2(w)$ as determined from the left-hand-side expressions for these.

A4.3.2 Effect of Added Noise on the General Model

Correlated Noise. Consider the general nonstationary model for the process z_t of order (p, d, q):

$$\phi(B)\nabla^d z_t = \theta(B)a_t \tag{A4.3.3}$$

Suppose that we cannot observe z_t itself, but only $Z_t = z_t + b_t$, where b_t represents some extraneous noise (e.g., measurement error) or simply some additional unobserved component that together with z_t forms the observed process Z_t, and b_t may be autocorrelated. We wish to determine the nature of the model for the observed process Z_t. In general, applying $\phi(B)\nabla^d$ to both sides of $Z_t = z_t + b_t$, we have

$$\phi(B)\nabla^d Z_t = \theta(B)a_t + \phi(B)\nabla^d b_t$$

If the noise b_t follows a stationary ARMA process of order $(p_1, 0, q_1)$,

$$\phi_1(B)b_t = \theta_1(B)\alpha_t \tag{A4.3.4}$$

where α_t is a white noise process independent of the a_t process, then

$$\underbrace{\phi_1(B)\phi(B)\nabla^d}_{p_1+p+d} Z_t = \underbrace{\phi_1(B)\theta(B)}_{p_1+q} a_t + \underbrace{\phi(B)\theta_1(B)\nabla^d}_{p+q_1+d} \alpha_t \tag{A4.3.5}$$

where the values below the braces indicate the degrees of the various polynomials in B. Now the right-hand side of (A4.3.5) is of the form (A4.3.1). Let $P = p_1 + p$ and Q be equal to whichever of $(p_1 + q)$ and $(p + q_1 + d)$ is larger. Then we can write

$$\phi_2(B)\nabla^d Z_t = \theta_2(B)u_t$$

with u_t a white noise process, and the Z_t process is seen to be an ARIMA of order (P, d, Q). The stationary AR operator in the ARIMA model for Z_t is determined as $\phi_2(B) = \phi_1(B)\phi(B)$, and the parameters of the MA operator $\theta_2(B)$ and σ_u^2 are determined in the same manner as described in Section A4.3.1, that is, by equating the nonzero autocovariances from the representations:

$$\phi_1(B)\theta(B)a_t + \phi(B)\theta_1(B)\nabla^d\alpha_t = \theta_2(B)u_t$$

Added White Noise. If, as might be true in some applications, the added noise is white, then $\phi_1(B) = \theta_1 B = 1$ in (A4.3.4), and we obtain

$$\phi(B)\nabla^d Z_t = \theta_2(B)u_t \tag{A4.3.6}$$

with

$$\theta_2(B)u_t = \theta(B)a_t + \phi(B)\nabla^d b_t$$

which is of order (p, d, Q) where Q is the larger of q and $(p + d)$. If $p + d \leq q$, the order of the process with error is the same as that of the original process. The only effect of the added white noise is to change the values of the θ's (but not the ϕ's).

Effect of Added White Noise on an Integrated Moving Average Process. In particular, an IMA process of order $(0, d, q)$, with white noise added, remains an IMA of order $(0, d, q)$ if $d \leq q$; otherwise, it becomes an IMA of order $(0, d, d)$. In either case, the parameters of the process are changed by the addition of noise, with the representation $\nabla^d Z_t = \theta_2(B)u_t$ as in (A4.3.6). The nature of these changes can be determined by equating the autocovariances of the dth differences of the process, with added noise, to those of the dth differences of a simple IMA process, that is, as a special case of the above, by equating the nonzero autocovariances in the representation

$$\theta(B)a_t + \nabla^d b_t = \theta_2(B)u_t$$

The procedure will now be illustrated with an example.

A4.3.3 Example for an IMA(0, 1, 1) Process with Added White Noise

Consider the properties of the process $Z_t = z_t + b_t$ when

$$z_t = z_{t-1} - (1 - \lambda)a_{t-1} + a_t \tag{A4.3.7}$$

and the b_t and a_t are mutually independent white noise processes. The Z_t process has first difference $W_t = Z_t - Z_{t-1}$ given by

$$W_t = [1 - (1 - \lambda)B]a_t + (1 - B)b_t \tag{A4.3.8}$$

The autocovariances for the first differences W_t are

$$\begin{aligned} \gamma_0 &= \sigma_a^2[1 + (1 - \lambda)^2] + 2\sigma_b^2 \\ \gamma_1 &= -\sigma_a^2(1 - \lambda) - \sigma_b^2 \\ \gamma_j &= 0 \qquad j \geq 2 \end{aligned} \tag{A4.3.9}$$

The fact that the γ_j are zero beyond the first lag confirms that the process with added noise is, as expected, an IMA process of order (0, 1, 1). To obtain explicitly the parameters of the IMA that represents the noisy process, we suppose that it can be written as

$$Z_t = Z_{t-1} - (1 - \Lambda)u_{t-1} + u_t \tag{A4.3.10}$$

where u_t is a white noise process. The process (A4.3.10) has first differences $W_t = Z_t - Z_{t-1}$ with autocovariances

$$\begin{aligned} \gamma_0 &= \sigma_u^2[1 + (1 - \Lambda)^2] \\ \gamma_1 &= -\sigma_u^2(1 - \Lambda) \\ \gamma_j &= 0 \qquad j \geq 2 \end{aligned} \tag{A4.3.11}$$

Equating (A4.3.9) and (A4.3.11), we can solve for Λ and σ_u^2 explicitly. Thus

$$\frac{\Lambda^2}{1-\Lambda^2} = \frac{\lambda^2}{1-\lambda+\sigma_b^2/\sigma_a^2}$$

$$\sigma_u^2 = \sigma_a^2 \frac{\lambda^2}{\Lambda^2} \tag{A4.3.12}$$

Suppose, for example, that the original series has $\lambda = 0.5$ and $\sigma_b^2 = \sigma_a^2$; then, $\Lambda = 0.333$ and $\sigma_u^2 = 2.25\sigma_a^2$.

A4.3.4 Relation between the IMA(0, 1, 1) Process and a Random Walk

The process

$$z_t = z_{t-1} + a_t \tag{A4.3.13}$$

which is an IMA(0, 1, 1) process, with $\lambda = 1(\theta = 0)$, is called a *random walk*. If the a_t are steps taken forward or backward at time t, then z_t will represent the position of the walker at time t.

Any IMA(0, 1, 1) process can be thought of as a random walk buried in white noise b_t, uncorrelated with the shocks a_t associated with the random walk process. If the noisy process is $Z_t = z_t + b_t$, where z_t is defined by (A4.3.13), then using (A4.3.12), we have

$$Z_t = Z_{t-1} - (1-\Lambda)u_{t-1} + u_t$$

with

$$\frac{\Lambda^2}{1-\Lambda^2} = \frac{\sigma_a^2}{\sigma_b^2} \qquad \sigma_u^2 = \frac{\sigma_a^2}{\Lambda^2} \tag{A4.3.14}$$

A4.3.5 Autocovariance Function of the General Model with Added Correlated Noise

Suppose that the basic process is an ARIMA process of order (p, d, q):

$$\phi(B)\nabla^d z_t = \theta(B)a_t$$

and that $Z_t = z_t + b_t$ is observed, where the stationary process b_t, which has autocovariance function $\gamma_j(b)$, is independent of the process a_t, and hence of z_t. Suppose that $\gamma_j(w)$ is the autocovariance function for $w_t = \nabla^d z_t = \phi^{-1}(B)\theta(B)a_t$ and that $W_t = \nabla^d Z_t$. We require the autocovariance function for W_t. Now

$$\nabla^d(Z_t - b_t) = \phi^{-1}(B)\theta(B)a_t$$

$$W_t = w_t + v_t$$

where

$$v_t = \nabla^d b_t = (1-B)^d b_t$$

Hence

$$\begin{aligned}\gamma_j(W) &= \gamma_j(w) + \gamma_j(v)\\ \gamma_j(v) &= (1-B)^d(1-F)^d\gamma_j(b)\\ &= (-1)^d(1-B)^{2d}\gamma_{j+d}(b)\end{aligned}$$

and

$$\gamma_j(W) = \gamma_j(w) + (-1)^d(1-B)^{2d}\gamma_{j+d}(b) \tag{A4.3.15}$$

For example, suppose that correlated noise b_t is added to an IMA(0, 1, 1) process defined by $w_t = \nabla z_t = (1-\theta B)a_t$. Then the autocovariances of the first difference W_t of the ''noisy'' process will be

$$\begin{aligned}\gamma_0(W) &= \sigma_a^2(1+\theta^2) + 2[\gamma_0(b) - \gamma_1(b)]\\ \gamma_1(W) &= -\sigma_a^2\theta + [2\gamma_1(b) - \gamma_0(b) - \gamma_2(b)]\\ \gamma_j(W) &= [2\gamma_j(b) - \gamma_{j-1}(b) - \gamma_{j+1}(b)] \qquad j \geq 2\end{aligned}$$

In particular, if b_t was first-order autoregressive, so that $b_t = \phi b_{t-1} + \alpha_t$,

$$\begin{aligned}\gamma_0(W) &= \sigma_a^2(1+\theta^2) + 2\gamma_0(b)(1-\phi)\\ \gamma_1(W) &= -\sigma_a^2\theta - \gamma_0(b)(1-\phi)^2\\ \gamma_j(W) &= -\gamma_0(b)\phi^{j-1}(1-\phi)^2 \qquad j \geq 2\end{aligned}$$

where $\gamma_0(b) = \sigma_\alpha^2/(1-\phi^2)$. In fact, from (A4.3.5), the resulting noisy process $Z_t = z_t + b_t$ is in this case defined by

$$(1-\phi B)\nabla Z_t = (1-\phi B)(1-\theta B)a_t + (1-B)\alpha_t$$

which will be of order (1, 1, 2), and for the associated ARMA(1, 2) process $W_t = \nabla Z_t$, we know that the autocovariances satisfy $\gamma_j(W) = \phi\gamma_{j-1}(W)$ for $j \geq 3$ [e.g., see (3.4.3)] as is shown explicitly above.

EXERCISES

4.1. For each of the models

(1) $(1-B)z_t = (1-0.5B)a_t$

(2) $(1-B)z_t = (1-0.2B)a_t$

(3) $(1-0.5B)(1-B)z_t = a_t$

(4) $(1-0.2B)(1-B)z_t = a_t$

(5) $(1-0.2B)(1-B)z_t = (1-0.5B)a_t$

(a) Obtain the first seven ψ_j weights.

(b) Obtain the first seven π_j weights.

(c) Classify as a member of the class of ARIMA(*p, d, q*) processes.

4.2. For the five models of Exercise 4.1, and using where appropriate the results there obtained,

(a) Write each model in random shock form.

(b) Write each model as a complementary function plus a particular integral in relation to an origin $k = t - 3$.

(c) Write each model in inverted form.

4.3. Consider the IMA(0, 2, 2) process with parameters $\theta_1 = 0.8$ and $\theta_2 = -0.4$.

(a) Is the process invertible? If so, what is the expected pattern of the π weights?

(b) Calculate and plot the first ten π weights for the original series z_t and comment.

(c) Calculate and plot the first ten π weights for the differenced series $w_t = (1 - B)^2 z_t$.

4.4 Given the following series of random shocks a_t, and given that $z_0 = 20, z_{-1} = 19$,

t	a_t	t	a_t	t	a_t
0	−0.3	5	−0.6	10	−0.4
1	0.6	6	1.7	11	0.9
2	0.9	7	−0.9	12	0.0
3	0.2	8	−1.3	13	−1.4
4	0.1	9	−0.6	14	−0.6

(a) Use the difference equation form of the model to obtain $z_1, z_2, \ldots, z_{14}$ for each of the five models in Exercise 4.1.

(b) Plot the resulting series.

4.5. Using the inverted forms of each of the models in Exercise 4.1, obtain z_{12}, z_{13}, and z_{14}, using only the values $z_1, z_2, \ldots, z_{11}$ derived in Exercise 4.4 and a_{12}, a_{13}, and a_{14}. Confirm that the values agree with those obtained in Exercise 4.4.

4.6. Consider the IMA(0, 1, 1) model $(1 - B)z_t = (1 - \theta)a_t$, where the a_t are i.i.d. $N(0, \sigma_a^2)$.

(a) Derive the expected value and variance of z_t, $t = 1, 2, \ldots$, assuming that the process starts at time $t = 1$ with $z_0 = 10$.

(b) Derive the correlation coefficient ρ_k between z_t and z_{t-k}, conditioning on $z_0 = 10$. Assume that t is much larger than the lag k.

(c) Provide an approximate value for the autocorrelation coefficient ρ_k derived in part (c).

4.7. If $\bar{z}_t = \sum_{j=1}^{\infty} \pi_j z_{t+1-j}$, then for models (1) and (2) of Exercise 4.1, which are of the form $(1 - B)z_t = (1 - \theta B)a_t$, $\bar{z}_t$ is an exponentially weighted moving average. For these two models, by actual calculation, confirm that $\bar{z}_{11}, \bar{z}_{12}$, and $\bar{z}_{13}$ satisfy

the relations

$$
\begin{aligned}
z_t &= \bar{z}_{t-1} + a_t \quad \text{(see Exercise 4.5)} \\
\bar{z}_t &= \bar{z}_{t-1} + (1-\theta)a_t \\
&= (1-\theta)z_t + \theta \bar{z}_{t-1}
\end{aligned}
$$

4.8. If $w_{1t} = (1 - \theta_1 B)a_{1t}$ and $w_{2t} = (1 - \theta_2 B)a_{2t}$, show that $w_{3t} = w_{1t} + w_{2t}$ may be written as $w_{3t} = (1 - \theta_3 B)a_{3t}$, and derive an expression for θ_3 and $\sigma^2_{a_3}$ in terms of the parameters of the other two processes. State your assumptions.

4.9. Suppose that $Z_t = z_t + b_t$, where z_t is a first-order autoregressive process $(1 - \phi B)z_t = a_t$ and b_t is a white noise process with variance σ^2_b. What model does the process Z_t follow? State your assumptions.

4.10. **(a)** Simulate a time series of $N = 200$ observations from an IMA(0, 2, 2) model with parameters $\theta_1 = 0.8$ and $\theta_2 = -0.4$ using the arima.sim() function in R; type help(arima.sim) for details. Plot the resulting series and comment on its behavior.

(b) Estimate and plot the autocorrelation function of the simulated time series.

(c) Estimate and plot the autocorrelation functions of the first and second differences of the series.

(d) Comment on the patterns of the autocorrelation functions generated above. Are the results consistent with what you would expect to see for this IMA(0, 2, 2) process?

4.11. Download the daily S& P 500 Index stock price values for the period January 2, 2014 to present from the Internet (e.g., http://research.stlouisfed.org).

(a) Plot the series using R. Calculate and graph the autocorrelation and partial autocorrelation functions for this series. Does the series appear to be stationary?

(b) Repeat the calculations in part **(a)** for the first and second differences of the series. Describe the effects of differencing in this case. Can you suggest a model that might be appropriate for this series?

(c) The return or relative gain on a stock can be calculated as $(z_t - z_{t-1})/z_t$ or $\log(z_t) - \log(z_{t-1})$. Perform this calculation and comment on the stationarity of the resulting series.

4.12. Repeat the analysis in Exercise 11 for the Dow Jones Industrial Average, or for a time series of your own choosing.

5

FORECASTING

In Chapter 4, we discussed the properties of autoregressive integrated moving average (ARIMA) models and examined in detail some special cases that appear to be common in practice. We will now show how these models may be used to forecast future values of an observed time series. In Part Two, we will consider the problem of selecting a suitable model of this form and fitting it to actual data. For the present, however, we proceed as if the model were known *exactly,* bearing in mind that estimation errors in the parameters will not seriously affect the forecasts unless the time series is relatively short.

This chapter will focus on nonseasonal time series. The forecasting, as well as model fitting, of seasonal time series is described in Chapter 9. We show how minimum mean square error (MSE) forecasts may be generated directly from the *difference equation* form of the model. A further recursive calculation yields probability limits for the forecasts. It is emphasized that for practical computation of the forecasts, this approach via the difference equation is the simplest and most elegant. However, to provide insight into the nature of the forecasts, we also consider them from other viewpoints. As a computational tool, we also demonstrate how to generate forecasts and associated probability limits using the R software.

5.1 MINIMUM MEAN SQUARE ERROR FORECASTS AND THEIR PROPERTIES

In Section 4.2, we discussed three explicit forms for the general ARIMA model:

$$\varphi(B)z_t = \theta(B)a_t \tag{5.1.1}$$

Time Series Analysis: Forecasting and Control, Fifth Edition. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung

where $\varphi(B) = \phi(B)\nabla^d$. We begin by recalling these three forms since each one sheds light on a different aspect of the forecasting problem.

We will consider forecasting a value z_{t+l}, $l \geq 1$, when we are currently at time t. This forecast is said to be made at *origin* t for lead *time* l. We now summarize the results of Section 4.2, but writing $t + l$ for t and t for k.

Three Explicit Forms for the Model. An observation z_{t+l} generated by the ARIMA process may be expressed as follows:

1. Directly in terms of the difference equation by

$$z_{t+l} = \varphi_1 z_{t+l-1} + \cdots + \varphi_{p+d} z_{t+l-p-d} - \theta_1 a_{t+l-1} - \cdots - \theta_q a_{t+l-q} + a_{t+l} \tag{5.1.2}$$

2. As an infinite weighted sum of current and previous shocks a_j:

$$z_{t+l} = \sum_{j=0}^{\infty} \psi_j a_{t+l-j} \tag{5.1.3}$$

where $\psi_0 = 1$ and, as in (4.2.5), the ψ weights may be obtained by equating coefficients in

$$\varphi(B)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = \theta(B) \tag{5.1.4}$$

Equivalently, for positive l, with reference to origin $k < t$, the model may be written in the truncated form:

$$\begin{aligned} z_{t+l} &= a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1} \\ &\quad + \psi_l a_t + \cdots + \psi_{t+l-k-1} a_{k+1} + C_k(t + l - k) \\ &= a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1} + C_t(l) \end{aligned} \tag{5.1.5}$$

where $C_k(t + l - k)$ is the complementary function relative to the finite origin k of the process. From (4.2.19), we recall that the complementary function relative to the forecast origin t can be expressed as $C_t(l) = C_k(t + l - k) + \psi_l a_t + \psi_{l+1} a_{t-1} + \cdots + \psi_{t+l-k-1} a_{k+1}$. Informally, $C_t(l)$ is associated with the truncated infinite sum:

$$C_t(l) = \sum_{j=l}^{\infty} \psi_j a_{t+l-j} \tag{5.1.6}$$

3. As an infinite weighted sum of previous observations, plus a random shock,

$$z_{t+l} = \sum_{j=1}^{\infty} \pi_j z_{t+l-j} + a_{t+l} \tag{5.1.7}$$

Also, if $d \geq 1$,

$$\bar{z}_{t+l-1}(\pi) = \sum_{j=1}^{\infty} \pi_j z_{t+l-j} \tag{5.1.8}$$

will be a weighted average, since then $\sum_{j=1}^{\infty} \pi_j = 1$. As in (4.2.22), the π weights may be obtained from

$$\varphi(B) = (1 - \pi_1 B - \pi_2 B^2 - \cdots)\theta(B) \tag{5.1.9}$$

5.1.1 Derivation of the Minimum Mean Square Error Forecasts

Now suppose, at origin t, that we are to make a forecast $\hat{z}_t(l)$ of z_{t+l}, which is to be a linear function of current and previous observations $z_t, z_{t-1}, z_{t-2}, \ldots$. Then, it will also be a linear function of current and previous shocks $a_t, a_{t-1}, a_{t-2}, \ldots$.

Suppose, then, that the best forecast is

$$\hat{z}_t(l) = \psi_l^* a_t + \psi_{l+1}^* a_{t-1} + \psi_{l+2}^* a_{t-2} + \cdots$$

where the weights $\psi_l^*, \psi_{l+1}^*, \ldots$ are to be determined. Then, using (5.1.3), the mean square error of the forecast is

$$\begin{aligned} E[z_{t+l} - \hat{z}_t(l)]^2 =& (1 + \psi_1^2 + \cdots + \psi_{l-1}^2)\sigma_a^2 \\ &+ \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*)^2 \sigma_a^2 \end{aligned} \tag{5.1.10}$$

which is minimized by setting $\psi_{l+j}^* = \psi_{l+j}$. This conclusion is a special case of more general results in prediction theory (Wold, 1938; Kolmogoroff (1939, 1941a, 1941b), Wiener, 1949; Whittle, 1963). We then have

$$\begin{aligned} z_{t+l} &= (a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1}) \\ &\quad + (\psi_l a_t + \psi_{l+1} a_{t-1} + \cdots) \qquad (5.1.11) \\ &= e_t(l) + \hat{z}_t(l) \qquad (5.1.12) \end{aligned}$$

where $e_t(l)$ is the error of the forecast $\hat{z}_t(l)$ at lead time l.

Certain important facts emerge. As before, denote $E[z_{t+l}|z_t, z_{t-1}, \ldots]$, the conditional expectation of z_{t+l} given the knowledge of all the z's up to time t, by $E_t[z_{t+l}]$. We will assume that a_t are a sequence of independent random variables.

1. Then, $E[a_{t+j}|z_t, z_{t-1}, \ldots] = 0$, $j > 0$, and so from (5.1.3),

$$\hat{z}_t(l) = \psi_l a_t + \psi_{l+1} a_{t-1} + \cdots = E_t[z_{t+l}] \tag{5.1.13}$$

Thus, the minimum mean square error forecast at origin t, for lead time l, is the conditional expectation of z_{t+l} at time t. When $\hat{z}_t(l)$ is regarded as a function of l for fixed t, it will be called the *forecast function* for origin t. We note that a minimal requirement on the random shocks a_t in the model (5.1.1) in order for the conditional expectation $E_t[z_{t+l}]$, which always equals the minimum mean square error forecast, to coincide with the minimum mean square error *linear* forecast is that $E_t[a_{t+j}] = 0$, $j > 0$. This property may not hold for certain types of nonlinear processes studied, for example, by Priestley (1988), Tong (1983, 1990), and many subsequent authors. Such processes may, in fact, possess a linear representation as in (5.1.1), but the shocks a_t will not be independent, only uncorrelated, and the best forecast $E_t[z_{t+l}]$ may not coincide with the best linear forecast $\hat{z}_t(l)$ as obtained in (5.1.11).

2. The forecast error for lead time l is

$$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1} \tag{5.1.14}$$

Since

$$E_t[e_t(l)] = 0 \tag{5.1.15}$$

the forecast is unbiased. Also, the variance of the forecast error is

$$V(l) = \text{var}[e_t(l)] = (1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2)\sigma_a^2 \tag{5.1.16}$$

3. It is readily shown that not only is $\hat{z}_t(l)$ the minimum mean square error forecast of z_{t+l}, but that any linear function $\sum_{l=1}^{L} w_l \hat{z}_t(l)$ of the forecasts is also a minimum mean square error forecast of the corresponding linear function $\sum_{l=1}^{L} w_l z_{t+l}$ of the future observations. For example, suppose that using (5.1.13), we have obtained, from monthly data, minimum mean square error forecasts $\hat{z}_t(1)$, $\hat{z}_t(2)$, and $\hat{z}_t(3)$ of the sales of a product 1, 2, and 3 months ahead. Then, it is true that $\hat{z}_t(1) + \hat{z}_t(2) + \hat{z}_t(3)$ is the minimum mean square error forecast of the sales $z_{t+1} + z_{t+2} + z_{t+3}$ during the next quarter.

4. *The Shocks as One-Step-Ahead Forecast Errors*. Using (5.1.14), the one-step-ahead forecast error is

$$e_t(1) = z_{t+1} - \hat{z}_t(1) = a_{t+1} \tag{5.1.17}$$

Hence, the shocks a_t, which generate the process, and which have been introduced so far merely as a set of independent random variables or shocks, turn out to be the *one-step-ahead forecast errors*.

It follows that for a minimum mean square error forecast, the one-step-ahead forecast errors must be uncorrelated. This makes sense, for if the one-step-ahead errors were correlated, the forecast error a_{t+1} could, to some extent, be predicted from available forecast errors $a_t, a_{t-1}, a_{t-2}, \ldots$. If the prediction so obtained was $\hat{a}_{t+1}$, then $\hat{z}_t(1) + \hat{a}_{t+1}$ would be a better forecast of z_{t+1} than was $\hat{z}_t(1)$.

5. *Correlation between the Forecast Errors*. Although the optimal forecast errors at lead time 1 will be uncorrelated, the forecast errors for longer lead times in general will be correlated. In Section A5.1.1, we derive a general expression for the correlation between the forecast errors $e_t(l)$ and $e_{t-j}(l)$, made at the *same* lead time l from *different* origins t and $t - j$.

Now, it is also true that forecast errors $e_t(l)$ and $e_t(l + j)$, made at different lead times from the same origin t, are correlated. One consequence of this is that there will often be a tendency for the forecast function to lie either wholly above or below the values of the series, when they eventually come to hand. In Section A5.1.2, we give a general expression for the correlation between the forecast errors $e_t(l)$ and $e_t(l + j)$, made from the *same* origin.

5.1.2 Three Basic Forms for the Forecast

We have seen that the minimum mean square error forecast $\hat{z}_t(l)$ for lead time l is the conditional expectation $E_t[z_{t+l}]$, of z_{t+l}, at origin t. Using this fact, we can write expressions for the forecast in any one of three different ways, corresponding to the three ways of

expressing the model summarized earlier in this section. To simplify the notation, we will temporarily adopt the convention that square brackets imply that the conditional expectation, at time t, is to be taken. Thus,

$$[a_{t+l}] = E_t[a_{t+l}] \qquad [z_{t+l}] = E_t[z_{t+l}]$$

For $l > 0$, the following are three different ways of expressing the forecasts:

1. *Forecasts from Difference Equation.* Taking conditional expectations at time t in (5.1.2), we obtain

$$\begin{aligned}[z_{t+l}] = \hat{z}_t(l) = \varphi_1[z_{t+l-1}] + \cdots + \varphi_{p+d}[z_{t+l-p-d}] - \theta_1[a_{t+l-1}] \\ - \cdots - \theta_q[a_{t+l-q}] + [a_{t+l}]\end{aligned} \tag{5.1.18}$$

2. *Forecasts in Integrated Form.* Use of (5.1.3) gives

$$\begin{aligned}[z_{t+l}] = \hat{z}_t(l) = [a_{t+l}] + \psi_1[a_{t+l-1}] + \cdots + \psi_{l-1}[a_{t+l}] \\ + \psi_l[a_t] + \psi_{l+1}[a_{t-1}] + \cdots\end{aligned} \tag{5.1.19}$$

yielding the form (5.1.13) discussed above. Alternatively, using the truncated form of the model (5.1.5), we have

$$\begin{aligned}[z_{t+l}] = \hat{z}_t(l) &= [a_{t+l}] + \psi_1[a_{t+l-1}] + \cdots \\ &\quad + \psi_{t+l-k-1}[a_{k+1}] + C_k(t + l - k) \\ &= [a_{t+l}] + \psi_1[a_{t+l-1}] + \cdots + \psi_{l-1}[a_{t+1}] + C_t(l)\end{aligned} \tag{5.1.20}$$

where $C_t(l)$ is the complementary function at origin t.

3. *Forecasts as a Weighted Average of Previous Observations and Forecasts Made at Previous Lead Times from the Same Origin.* Finally, taking conditional expectations in (5.1.7) yields

$$[z_{t+l}] = \hat{z}_t(l) = \sum_{j=1}^{\infty} \pi_j[z_{t+l-j}] + [a_{t+l}] \tag{5.1.21}$$

It is to be noted that the minimum mean square error forecast is defined in terms of the conditional expectation

$$[z_{t+l}] = E_t[z_{t+l}] = E[z_{t+l}|z_t, z_{t-1}, \ldots]$$

which theoretically requires knowledge of the z's stretching back into the infinite past. However, the requirement of invertibility imposed on the ARIMA model ensures that the π weights in (5.1.21) form a convergent series. Hence, for the computation of a forecast, the dependence on z_{t-j} for $j > k$ can typically be ignored. In practice, the π weights usually decay rather quickly, so whatever form of the model is employed, only a moderate length of series $z_t, z_{t-1}, \ldots, z_{t-k}$ is needed to calculate the forecasts to sufficient accuracy. The methods we discuss are easily modified to calculate the exact finite sample forecasts, $E[z_{t+l}|z_t, z_{t-1}, \ldots, z_1]$, based on the finite length of data $z_t, z_{t-1}, \ldots, z_1$.

To calculate the conditional expectations in expressions (5.1.18–5.1.21), we note that if j is a nonnegative integer,

$$\begin{aligned}
[z_{t-j}] &= E_t[z_{t-j}] = z_{t-j} && j = 0, 1, 2, \ldots \\
[z_{t+j}] &= E_t[z_{t+j}] = \hat{z}_t(j) && j = 1, 2, \ldots \\
[a_{t-j}] &= E_t[a_{t-j}] = a_{t-j} = z_{t-j} - \hat{z}_{t-j-1}(1) && j = 0, 1, 2, \ldots \\
[a_{t+j}] &= E_t[a_{t+j}] = 0 && j = 1, 2, \ldots
\end{aligned} \tag{5.1.22}$$

Therefore, to obtain the forecast $\hat{z}_t(l)$, one writes down the model for z_{t+l} in any one of the three explicit forms above and treats the terms on the right according to the following rules:

1. The $z_{t-j}(j = 0, 1, 2, \ldots)$, which have already occurred at origin t, are left unchanged.
2. The $z_{t+j}(j = 1, 2, \ldots)$, which have not yet occurred, are replaced by their forecasts $\hat{z}_t(j)$ at origin t.
3. The $a_{t-j}(j = 0, 1, 2, \ldots)$, which have occurred, are available from $z_{t-j} - \hat{z}_{t-j-1}(1)$.
4. The $a_{t+j}(j = 1, 2, \ldots)$, which have not yet occurred, are replaced by zeros.

For routine calculation, it is easiest to work directly with the difference equation form (5.1.18). Hence, the forecasts for $l = 1, 2, \ldots$ are calculated recursively as

$$\hat{z}_t(l) = \sum_{j=1}^{p+d} \varphi_j \hat{z}_t(l-j) - \sum_{j=l}^{q} \theta_j a_{t+l-j}$$

where $\hat{z}_t(-j) = [z_{t-j}]$ denotes the observed value z_{t-j} for $j \geq 0$, and the moving average terms are not present for lead times $l > q$.

Example: Forecasting Using the Difference Equation Form. We will show in Chapter 7 that the viscosity data in Series C can be represented by the model

$$(1 - 0.8B)(1 - B)z_{t+1} = a_{t+1}$$

that is,

$$(1 - 1.8B + 0.8B^2)z_{t+1} = a_{t+1}$$

or

$$z_{t+l} = 1.8z_{t+l-1} - 0.8z_{t+l-2} + a_{t+l}$$

The forecasts at origin t are given by

$$\begin{aligned}
\hat{z}_t(1) &= 1.8z_t - 0.8z_{t-1} \\
\hat{z}_t(2) &= 1.8\hat{z}_t(1) - 0.8z_t \\
\hat{z}_t(l) &= 1.8\hat{z}_t(l-1) - 0.8\hat{z}_t(l-2) \qquad l = 3, 4, \ldots
\end{aligned} \tag{5.1.23}$$

yielding in a simple recursive calculation.

There are no moving average terms in this model. However, such terms produce no added difficulties. Later in this chapter, we have a series arising in a control problem, for

which the model at time $t + l$ is

$$\nabla^2 z_{t+l} = (1 - 0.9B + 0.5B^2)a_{t+l}$$

or, equivalently, $z_{t+l} = 2z_{t+l-1} - z_{t+l-2} + a_{t+l} - 0.9a_{t+l-1} + 0.5a_{t+l-2}$. Then,

$$\begin{aligned}
\hat{z}_t(1) &= 2z_t - z_{t-1} - 0.9a_t + 0.5a_{t-1} \\
\hat{z}_t(2) &= 2\hat{z}_t(1) - z_t + 0.5a_t \\
\hat{z}_t(l) &= 2\hat{z}_t(l-1) - \hat{z}_t(l-2) \qquad l = 3, 4, \ldots
\end{aligned}$$

In these expressions, we remember that $a_t = z_t - \hat{z}_{t-1}(1)$, $a_{t-1} = z_{t-1} - \hat{z}_{t-2}(1)$, and the forecasting process may be started off initially by setting unknown a values equal to their unconditional expected values of zero. Thus, assuming by convention that data are available starting from time $s = 1$, the necessary a_s's are computed recursively from the difference equation form (5.1.2) of the model:

$$a_s = z_s - \hat{z}_{s-1}(1) = z_s - \left(\sum_{j=1}^{p+d} \varphi_j z_{s-j} - \sum_{j=1}^{q} \theta_j a_{s-j} \right) \qquad s = p + d + 1, \ldots, t$$

setting initial a_s's equal to zero, for $s < p + d + 1$. Alternatively, it is possible to estimate the necessary initial a_s's, as well as the initial z_s's, using *back-forecasting*. This technique, which essentially determines the conditional expectations of the presample a_s's and z_s's, given the available data, is discussed in Chapter 7 with regard to parameter estimation of ARIMA models. However, provided that a sufficient length of data series $z_t, z_{t-1}, \ldots, z_1$ is available, the two different treatments of the initial values will have a negligible effect on the forecasts $\hat{z}_1(l)$.

5.2 CALCULATING FORECASTS AND PROBABILITY LIMITS

5.2.1 Calculation of ψ Weights

It is often the case that forecasts are needed for several lead times $1, 2, \ldots, L$. As already shown, the difference equation form of the model allows the forecasts to be generated recursively in the order $\hat{z}_t(1)$, $\hat{z}_t(2)$, $\hat{z}_t(3)$, and so on. To obtain probability limits for these forecasts, it is necessary to calculate the weights $\psi_1, \psi_2, \ldots, \psi_{L-1}$. This is accomplished using the relation

$$\varphi(B)\psi B = \theta(B) \tag{5.2.1}$$

that is, by equating coefficients of powers of B in

$$\begin{aligned}
(1 - \varphi_1 B - \cdots - \varphi_{p+d} B^{p+d})\ (1 + \psi_1 B + \psi_2 B^2 + \cdots) \\
= (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)
\end{aligned} \tag{5.2.2}$$

Knowing the values of the φ's and the θ's, the values of ψ may be obtained as follows:

$$\begin{aligned}\psi_1 &= \varphi_1 - \theta_1 \\ \psi_2 &= \varphi_1\psi_1 + \varphi_2 - \theta_2 \\ &\vdots \\ \psi_j &= \varphi_j\psi_{j-1} + \cdots + \varphi_{p+d}\psi_{j-p-d} - \theta_j\end{aligned} \tag{5.2.3}$$

where $\psi_0 = 1$, $\psi_j = 0$ for $j < 0$, and $\theta_j = 0$ for $j > q$. If K is the greater of the integers $p + d - 1$ and q, then for $j > K$ the ψ's satisfy the difference equation:

$$\psi_j = \varphi_1\psi_{j-1} + \varphi_2\psi_{j-2} + \cdots + \varphi_{p+d}\psi_{j-p-d} \tag{5.2.4}$$

Thus, the ψ's are easily calculated recursively. For example, for the model $(1 - 1.8B + 0.8B^2)z_t = a_t$, appropriate to Series C, we have

$$(1 - 1.8B + 0.8B^2)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = 1$$

Hence, with $\varphi_1 = 1.8$ and $\varphi_2 = -0.8$, we obtain

$$\begin{aligned}\psi_0 &= 1 \\ \psi_1 &= 1.8 \\ \psi_j &= 1.8\psi_{j-1} - 0.8\psi_{j-2} \qquad j = 2, 3, 4, \ldots\end{aligned}$$

so that $\psi_2 = (1.8 \times 1.8) - (0.8 \times 1.0) = 2.44$ and $\psi_3 = (1.8 \times 2.44) - (0.8 \times 1.8) = 2.95$, and so on.

Before proceeding to discuss the probability limits, we briefly mention the use of the ψ weights for updating of forecasts as new data become available.

5.2.2 Use of the ψ Weights in Updating the Forecasts

Using (5.1.13), we can express the forecasts $\hat{z}_{t+1}(l)$ and $\hat{z}_t(l + 1)$ of the future observation z_{t+l+1} made at origins $t + 1$ and t as

$$\begin{aligned}\hat{z}_{t+1}(l) &= \psi_l a_{t+1} + \psi_{l+1}a_t + \psi_{l+2}a_{t-1} + \cdots \\ \hat{z}_t(l + 1) &= \psi_{l+1}a_t + \psi_{l+2}a_{t-1} + \cdots\end{aligned}$$

On subtraction, it follows that

$$\hat{z}_{t+1}(l) = \hat{z}_t(l + 1) + \psi_l a_{t+1} \tag{5.2.5}$$

Explicitly, the t-origin forecast of z_{t+l+1} can be updated to become the $t + 1$ origin forecast of the same z_{t+l+1}, by adding a constant multiple of the one-step-ahead forecast error $a_{t+l} \equiv z_{t+l} - \hat{z}_t(l)$ with multiplier ψ_l.

This leads to a rather remarkable conclusion. Suppose that we currently have forecasts at origin t for lead times $1, 2, \ldots, L$. Then, as soon as z_{t+l} becomes available, we can calculate $a_{t+1} \equiv z_{t+l} - \hat{z}_t(1)$ and proportionally update to obtain forecasts $\hat{z}_{t+1}(l) = \hat{z}_t(l + 1) + \psi_l a_{t+1}$ at origin $t + 1$, for lead times $1, 2, \ldots, L - 1$. The new forecast $\hat{z}_{t+1}(L)$, for lead time L, cannot be calculated by this means but is easily obtained from the forecasts at shorter lead times, using the difference equation.

TABLE 5.1 Variance Function for Series C

l	1	2	3	4	5	6	7	8	9	10
$V(l)/\sigma_a^2$	1.00	4.24	10.19	18.96	30.24	43.86	59.46	76.79	95.52	115.41

5.2.3 Calculation of the Probability Limits at Different Lead Times

The expression (5.1.16) shows that the variance of the l-steps-ahead forecast error for any origin t is the expected value of

$$e_t^2(l) = [z_{t+l} - \hat{z}_t(l)]^2$$

and is given by

$$V(l) = \left(1 + \sum_{j=1}^{l-1} \psi_j^2\right) \sigma_a^2$$

For example, using the ψ weights calculated above, the function $V(l)/\sigma_a^2$ for Series C is shown in Table 5.1.

Assuming that the a's are normal, it follows that given information up to time t, the conditional probability distribution $p(z_{t+l}|z_t, z_{t-1}, \ldots)$ of a future value z_{t+l} of the process will be normal with mean $\hat{z}_t(l)$ and standard deviation

$$\sigma(l) = \left(1 + \sum_{j=1}^{l-1} \psi_j^2\right)^{1/2} \sigma_a$$

Thus, the variate $(z_{t+l} - \hat{z}_t(l))/\sigma(l)$ will have a unit normal distribution and so $\hat{z}_t(l) \pm u_{\varepsilon/2}\sigma(l)$ provides limits of an interval such that z_{t+l} will lie within the interval with probability $1 - \varepsilon$, where $u_{\varepsilon/2}$ is the deviate exceeded by a proportion $\varepsilon/2$ of the unit normal distribution. Figure 5.1 shows the conditional probability distributions of future values z_{21}, z_{22}, z_{23} for Series C, given information up to origin $t = 20$.

We show in Chapter 7 how an estimate s_a^2, of the variance σ_a^2, may be obtained from time series data. When the number of observations on which this estimate is based is, say, at least 50, s_a may be substituted for σ_a and approximate $1 - \varepsilon$ probability limits $z_{t+l}(-)$ and $z_{t+l}(+)$ for z_{t+l} will be given by

$$z_{t+l}(\pm) = \hat{z}_t(l) \pm u_{\varepsilon/2} \left(1 + \sum_{j=1}^{l-1} \psi_j^2\right)^{1/2} s_a \tag{5.2.6}$$

It follows from Table 7.6 that for Series C, $s_a = 0.134$; hence, the 50 and 95% limits, for z_{t+2}, for example, are given by

$$\text{50\% limits} : \hat{z}_t(2) \pm (0.674)(1 + 1.8^2)^{1/2}(0.134) = \hat{z}_t(2) \pm 0.19$$
$$\text{95\% limits} : \hat{z}_t(2) \pm (1.960)(1 + 1.8^2)^{1/2}(0.134) = \hat{z}_t(2) \pm 0.55$$

Figure 5.2 shows a section of Series C together with the several-steps-ahead forecasts (indicated by crosses) from origins $t = 20$ and $t = 67$. Also shown are the 50 and 95% probability limits for z_{20+l}, for $l = 1$ to 14. The interpretation of the limits $z_{t+l}(-)$ and

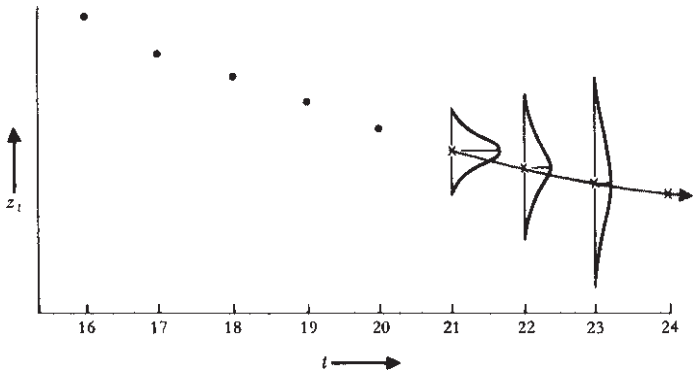

FIGURE 5.1 Conditional probability distributions of future values z_{21}, z_{22}, and z_{23} for Series C, given information up to origin $t = 20$.

$z_{t+l}(+)$ should be noted carefully. These limits are such that *given the information available at origin t*, there is a probability of $1 - \varepsilon$ that the actual value z_{t+l}, when it occurs, will be within them, that is,

$$\Pr\{z_{t+l}(-) < z_{t+l} < z_{t+l}(+)\} = 1 - \varepsilon$$

Also, the probabilities quoted apply to *individual* forecasts and not jointly to the forecasts at different lead times. For example, it is true that with 95% probability, the limits for lead time 10 will include the value z_{t+10} when it occurs. It is not true that the series can be expected to remain within *all* the limits simultaneously with this probability.

5.2.4 Calculation of Forecasts Using R

Forecasts of future values of a time series that follows an ARIMA(p, d, q) can be calculated using R. A convenient option is to use the function sarima.for() in the astsa package. For example, if z represents the observed time series, the command

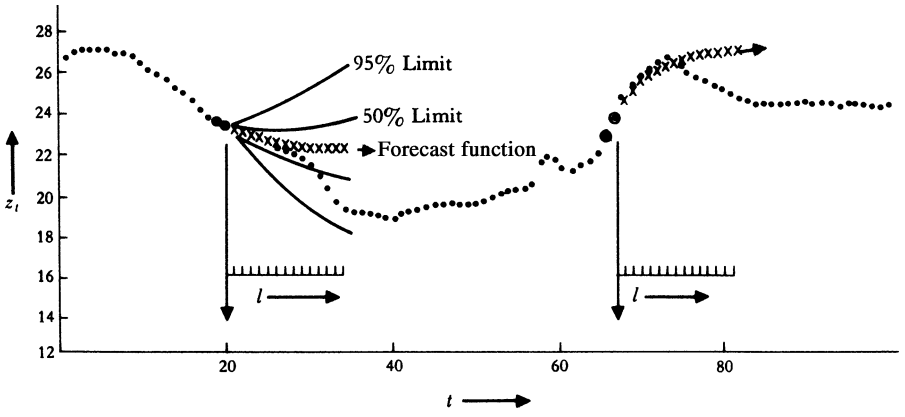

FIGURE 5.2 Forecasts for Series C and probability limits.

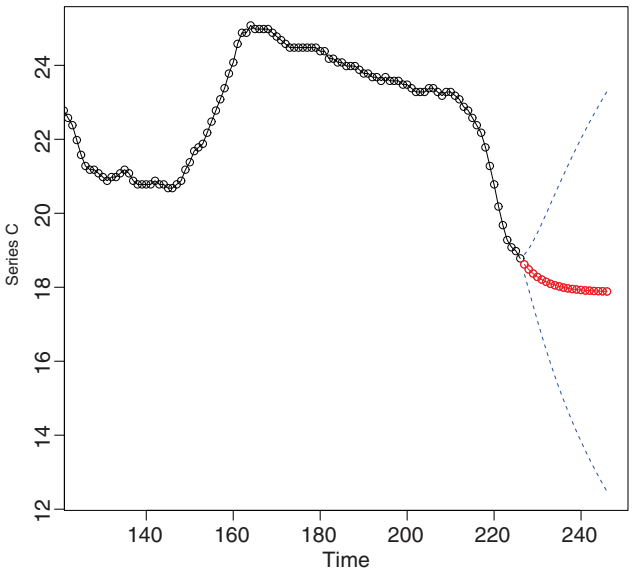

FIGURE 5.3 Forecasts for Series C with ±2 prediction error limits generated using R.

sarima.for(z,n.ahead,p,d,q,no.constant=TRUE) will fit the ARIMA(p, d, q) model without a constant term to the series and generate forecasts from the fitted model. The argument n.ahead specifies the number of forecasts to be generated. The output gives the forecasts and the standard errors of the forecasts, and supplies a graph of the forecasts along with their $+/-$ 2 prediction error limits. Thus, forecasts up to 20 steps ahead for Series C based on the ARIMA(1, 1, 0) model $(1 - \phi B)(1 - B) = a_t$ are generated as follows:

```
> library(astsa)
> seriesC=read.table("SeriesC.txt,header=TRUE)
> m1=sarima.for(seriesC,20,1,1,0,no.constant=FALSE)
> m1   % prints output from file m1
```

This code generates an output file ‘‘m1’’ that includes the forecasts (‘‘pred’’) and the prediction errors (‘‘se’’) of the forecasts. These can be accessed as m1$pred and m1$se, if needed for further analysis. Figure 5.3 shows a graph of the forecasts and their associated ±2 prediction error limits for Series C. We note that the limits become wider as the lead time increases, reflecting the increased uncertainty due to the fact that the series is nonstationary and does not vary around a fixed mean level.

5.3 FORECAST FUNCTION AND FORECAST WEIGHTS

Forecasts are calculated most simply by direct use of the difference equation. From the purely *computational* standpoint, the other model forms are less convenient. However, from the point of view of studying the nature of the forecasts, it is useful to consider in greater detail the alternative forms discussed in Section 5.1.2 and, in particular, to consider the explicit form of the forecast function.

5.3.1 Eventual Forecast Function Determined by the Autoregressive Operator

At time $t + l$, the ARIMA model may be written as

$$z_{t+l} - \varphi_1 z_{t+l-1} - \cdots - \varphi_{p+d} z_{t+l-p-d} = a_{t+l} - \theta_1 a_{t+l-1} - \cdots - \theta_q a_{t+l-q} \tag{5.3.1}$$

Taking the conditional expectations at time t, we have, for $l > q$,

$$\hat{z}_t(l) - \varphi_1 \hat{z}_t(l-1) - \cdots - \varphi_{p+d} \hat{z}_t(l-p-d) = 0 \qquad l > q \tag{5.3.2}$$

where it is understood that $\hat{z}_t(-j) = z_{t-j}$ for $j \geq 0$. This difference equation has the solution

$$\hat{z}_t(l) = b_0^{(t)} f_0(l) + b_1^{(t)} f_1(l) + \cdots + b_{p+d-1}^{(t)} f_{p+d-1}(l) \tag{5.3.3}$$

for $l > q - p - d$. Note that the forecast $\hat{z}_t(l)$ is the complementary function introduced in Chapter 4. In (5.3.3), $f_0(l), f_1(l), \ldots, f_{p+d-1}(l)$ are functions of the lead time l. In general, they could include polynomials, exponentials, sines and cosines, and products of these functions. The functions $f_0(l), f_1(l), \ldots, f_{p+d-1}(l)$ consist of d polynomial terms l^i, $i = 0, \ldots, d-1$, of degree $d-1$, associated with the nonstationary operator $\nabla^d = (1-B)^d$, and p damped exponential and damped sinusoidal terms of the form G^l and $D^l \sin(2\pi f l + F)$, respectively, associated with the roots of $\phi(B) = 0$ for the stationary autoregressive operator. That is, the forecast function has the form

$$\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)} l + \cdots + b_{d-1}^{(t)} l^{d-1} + b_d^{(t)} f_d(l) + b_{d+1}^{(t)} f_{d+1}(l) + \cdots + b_{p+d-1}^{(t)} f_{p+d-1}(l)$$

For instance, if $\phi(B) = 0$ has p distinct real roots $G_1^{-1}, \ldots, G_p^{-1}$, then the last p terms in $\hat{z}_t(l)$ are $b_d^{(t)} G_1^l + b_{d+1}^{(t)} G_2^l + \cdots + b_{p+d-1}^{(t)} G_p^l$. Since the operator $\phi(B)$ is stationary, we have $|G| < 1$ and $D < 1$ and the last p terms in $\hat{z}_t(l)$ are transient and decay to zero as l increases. Hence, the forecast function is dominated by the remaining polynomial terms, $\sum_{i=0}^{d-1} b_i^{(t)} l^i$, as l increases. For a *given origin* t, the coefficients $b_j^{(t)}$ are constants applying to all lead times l, but they change from one origin to the next, *adapting* themselves appropriately to the particular part of the series being considered. From now on we call the function defined by (5.3.3) the *eventual forecast function*; ''eventual'' because when it occasionally happens that $q > p + d$, it supplies the forecasts only for lead times $l > q - p - d$.

We see from (5.3.2) that it is the general autoregressive operator $\varphi(B)$ that determines the mathematical form of the forecast function, that is, the nature of the f's in (5.3.3). Specifically, it determines whether the forecast function is to be a polynomial, a mixture of sines and cosines, a mixture of exponentials, or a combination of these functions.

5.3.2 Role of the Moving Average Operator in Fixing the Initial Values

While the autoregressive operator determines the nature of the eventual forecast function, the moving average operator is influential in determining how that function is to be ''fitted'' to the data and hence how the coefficients $b_0^{(t)}, b_1^{(t)}, \ldots, b_{p+d-1}^{(t)}$ in (5.3.3) are to be calculated and updated.

For example, consider the IMA(0, 2, 3) process:

$$z_{t+l} - 2z_{t+l-1} + z_{t+l-2} = a_{t+l} - \theta_1 a_{t+l-1} - \theta_2 a_{t+l-2} - \theta_3 a_{t+l-3}$$

Taking the conditional expectation, the forecast function becomes

$$\begin{aligned}
\hat{z}_t(1) &= 2z_t - z_{t-1} - \theta_1 a_t - \theta_2 a_{t-1} - \theta_3 a_{t-2} \\
\hat{z}_t(2) &= 2\hat{z}_t(1) - z_t - \theta_2 a_t - \theta_3 a_{t-1} \\
\hat{z}_t(3) &= 2\hat{z}_t(2) - \hat{z}_t(1) - \theta_3 a_t \\
\hat{z}_t(l) &= 2\hat{z}_t(l-1) - \hat{z}_t(l-2) \qquad l > 3
\end{aligned}$$

Therefore, since $\varphi(B) = (1 - B)^2$ in this model, the eventual forecast function is the unique straight line

$$\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)} l \qquad l > 1$$

which passes through $\hat{z}_t(2)$ and $\hat{z}_t(3)$ as shown in Figure 5.4. However, note that if the θ_3 term had not been included in the model, then $q - p - d = 0$, and the forecast would have been given at *all lead times* by the straight line passing through $\hat{z}_t(1)$ and $\hat{z}_t(2)$.

In general, since only one function of the form (5.3.3) can pass through $p + d$ points, the eventual forecast function is that unique curve of the form required by $\varphi(B)$, which passes through the $p + d$ ''pivotal'' values $\hat{z}_t(q), \hat{z}_t(q-1), \ldots, \hat{z}_t(q - p - d + 1)$, where $\hat{z}_t(-j) = z_{t-j}$ $(j = 0, 1, 2, \ldots)$. In the extreme case where $q = 0$, so that the model is of the purely autoregressive form $\varphi(B)z_t = a_t$, the curve passes through the points $z_t, z_{t-1}, \ldots, z_{t-p-d+1}$. Thus, the pivotal values can consist of forecasts or of actual values of the series; they are indicated in the figures by circled points.

The moving average terms help to decide the way in which we ''reach back'' into the series to fit the forecast function determined by the autoregressive operator $\varphi(B)$. Figure 5.5 illustrates the situation for the model of order (1,1,3) given by $(1 - \phi B)\nabla z_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)a_t$. The (hypothetical) weight functions indicate the linear functional dependence of the three forecasts, $\hat{z}_t(1)$, $\hat{z}_t(2)$, and $\hat{z}_t(3)$, on $z_t, z_{t-1}, z_{t-2}, \ldots$. Since the forecast function contains $p + d = 2$ coefficients, it is uniquely determined by the forecasts $\hat{z}_t(3)$ and $\hat{z}_t(2)$, that is, by $\hat{z}_t(q)$ and $\hat{z}_t(q-1)$. We next consider how the forecast weight functions, referred to above, are determined.

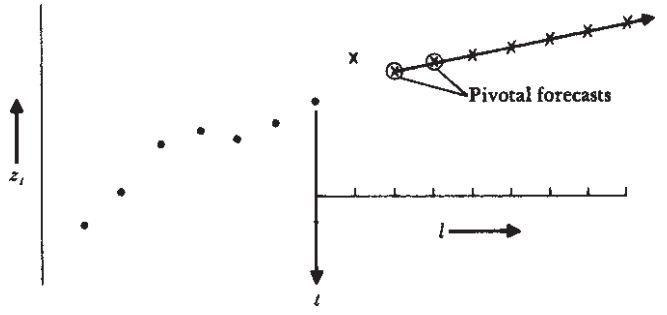

FIGURE 5.4 Eventual forecast function for an IMA(0, 2, 3) process.

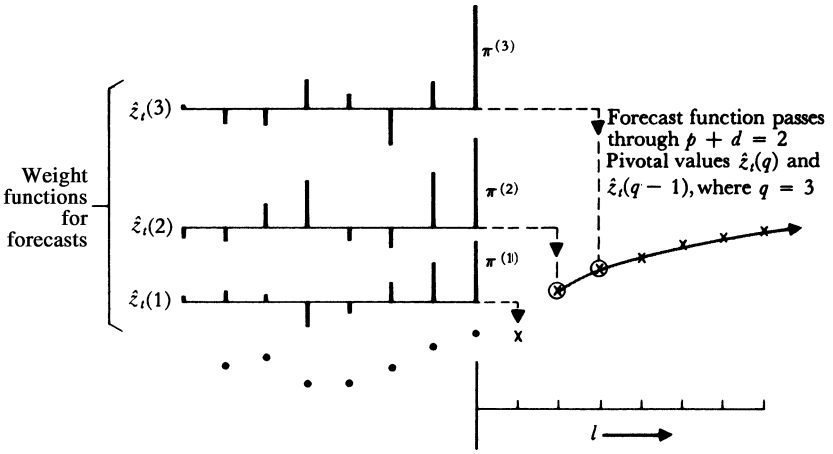

FIGURE 5.5 Dependence of forecast function on observations for a (1, 1, 3) process $(1 - \phi B)\nabla z_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)a_t$.

5.3.3 Lead *l* Forecast Weights

The fact that the general model may also be written in inverted form,

$$a_t = \pi(B)z_t = (1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots)z_t \tag{5.3.4}$$

allows us to write the forecast as in (5.1.21). On substituting for the conditional expectations in (5.1.21), we obtain

$$\hat{z}_t(l) = \sum_{j=1}^{\infty} \pi_j \hat{z}_t(l-j) \tag{5.3.5}$$

where, as before, $\hat{z}_t(-h) = z_{t-h}$ for $h = 0, 1, 2, \ldots$. Thus, in general,

$$\hat{z}_t(l) = \pi_1 \hat{z}_t(l-1) + \cdots + \pi_{l-1}\hat{z}_t(1) + \pi_l z_t + \pi_{l+1} z_{t-1} + \cdots \tag{5.3.6}$$

and, in particular,

$$\hat{z}_t(1) = \pi_1 z_t + \pi_2 z_{t-1} + \pi_3 z_{t-2} + \cdots$$

The forecasts for higher lead times may also be expressed directly as linear functions of the observations $z_t, z_{t-1}, z_{t-2}, \ldots$. For example, the lead 2 forecast at origin t is

$$\begin{aligned}\hat{z}_t(2) &= \pi_1 \hat{z}_t(1) + \pi_2 z_t + \pi_3 z_{t-1} + \cdots \\ &= \pi_1 \sum_{j=1}^{\infty} \pi_j z_{t+1-j} + \sum_{j=1}^{\infty} \pi_{j+1} z_{t+1-j} \\ &= \sum_{j=1}^{\infty} \pi_j^{(2)} z_{t+1-j}\end{aligned}$$

where

$$\pi_j^{(2)} = \pi_1\pi_j + \pi_{j+1} \qquad j = 1, 2, \ldots \tag{5.3.7}$$

Proceeding in this way, it is readily shown that

$$\hat{z}_t(l) = \sum_{j=1}^{\infty} \pi_j^{(l)} z_{t+1-j} \tag{5.3.8}$$

where

$$\pi_j^{(l)} = \pi_{j+l-1} + \sum_{h=1}^{l-1} \pi_h \pi_j^{(l-h)} \qquad j = 1, 2, \ldots \tag{5.3.9}$$

and $\pi_j^{(1)} = \pi_j$. Alternative methods for computing these weights are given in Appendix A5.2.

As seen earlier, the π_j's themselves may be obtained explicitly by equating coefficients in

$$\theta(B)(1 - \pi_1 B - \pi_2 B^2 - \cdots) = \varphi(B)$$

Given these values, the $\pi_j^{(l)}$'s may readily be obtained, if so desired, using (5.3.9) or the results of Appendix A5.2. As an example, consider again the model

$$\nabla^2 z_t = (1 - 0.9B + 0.5B^2)a_t$$

which was fitted to a series, a part of which is shown in Figure 5.6. Equating coefficients in

$$(1 - 0.9B + 0.5B^2)(1 - \pi_1 B - \pi_2 B^2 - \cdots) = 1 - 2B + B^2$$

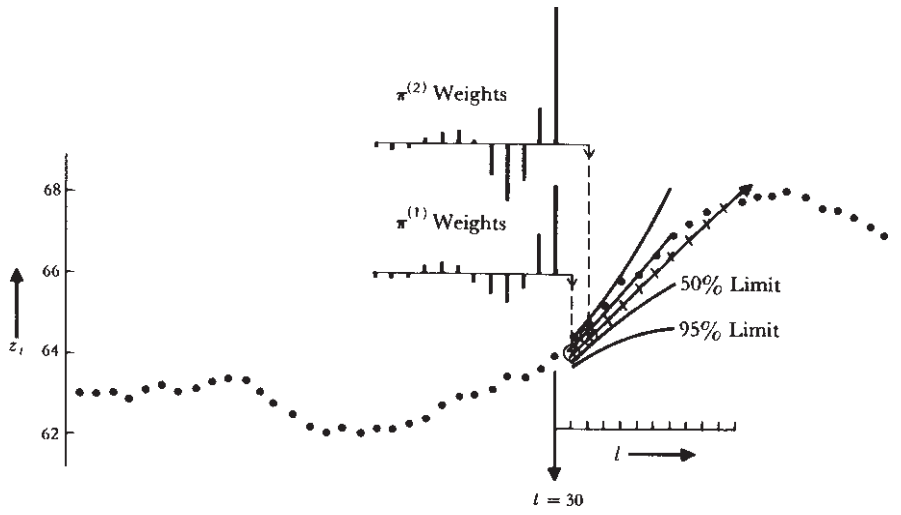

FIGURE 5.6 Part of a series fitted by $\nabla^2 z_t = (1 - 0.9B + 0.5B^2)a_t$ with forecast function for origin $t = 30$, forecast weights, and probability limits.

TABLE 5.2 π Weights for the Model $\nabla^2 z_t = (1 - 0.9B + 0.5B^2)a_t$

j	$\pi_j^{(1)}$	$\pi_j^{(2)}$
1	1.100	1.700
2	0.490	0.430
3	−0.109	−0.463
4	−0.343	−0.632
5	−0.254	−0.336
6	−0.057	0.013
7	0.076	0.181
8	0.097	0.156
9	0.049	0.050
10	−0.004	−0.032
11	−0.028	−0.054
12	−0.023	−0.026

yields the weights $\pi_j = \pi_j^{(1)}$, from which the weights $\pi_j^{(2)}$ may be computed using (5.3.7). The two sets of weights are given for $j = 1, 2, \ldots, 12$ in Table 5.2. In this example, the lead 1 and lead 2 forecasts, expressed in terms of the observations $z_t, z_{t-1}, \ldots$, are

$$\hat{z}_t(1) = 1.10z_t + 0.49z_{t-1} - 0.11z_{t-2} - 0.34z_{t-3} - 0.25z_{t-4} - \cdots$$

and

$$\hat{z}_t(2) = 1.70z_t + 0.43z_{t-1} - 0.46z_{t-2} - 0.63z_{t-3} - 0.34z_{t-4} + \cdots$$

In fact, the weights follow damped sine waves as shown in Figure 5.6.

5.4 EXAMPLES OF FORECAST FUNCTIONS AND THEIR UPDATING

The forecast functions for some special cases of the general ARIMA model will now be considered. We exhibit these in the three forms discussed in Section 5.1.2. While the forecasts are most easily computed from the difference equation itself, the other forms provide insight into the nature of the forecast function in particular cases.

5.4.1 Forecasting an IMA(0, 1, 1) Process

Difference Equation Approach. We first consider the model $\nabla z_t = (1 - \theta B)a_t$. At time $t + l$, the model may be written as

$$z_{t+l} = z_{t+l-1} + a_{t+l} - \theta a_{t+l-1}$$

Taking conditional expectations at origin t yields

$$\begin{aligned} \hat{z}_t(1) &= z_t - \theta a_t \\ \hat{z}_t(l) &= \hat{z}_t(l-1) \qquad l \geq 2 \end{aligned} \tag{5.4.1}$$

Hence, for all lead times, the forecasts at origin t will follow a straight line parallel to the time axis. Using the fact that $z_t = \hat{z}_{t-1}(1) + a_t$, we can write (5.4.1) in either of two useful forms.

The first of these is

$$\hat{z}_t(l) = \hat{z}_{t-l}(l) + \lambda a_t \tag{5.4.2}$$

where $\lambda = 1 - \theta$. This form is identical to the general updating form (5.2.5) for this model, since $\psi_l \equiv \lambda$ and $\hat{z}_{t-1}(l+1) = \hat{z}_{t-1}(l)$ for all $l \geq 1$. This form implies that having seen that our previous forecast $\hat{z}_{t-1}(l)$ falls short of the realized value by a_t, we adjust it by an amount λa_t. It will be recalled from Section 4.3.1 that λ measures the proportion of any given shock a_t, which is permanently absorbed by the ''level'' of the process. Therefore, it is reasonable to increase the forecast by that part λa_t of a_t, which we expect to be absorbed.

The second way of rewriting (5.4.1) is to write $a_t = z_t - \hat{z}_{t-1}(1) = z_t - \hat{z}_{t-1}(l)$ in (5.4.2) to obtain

$$\hat{z}_t(l) = \lambda z_t + (1 - \lambda)\hat{z}_{t-1}(l) \tag{5.4.3}$$

This form implies that the new forecast is a linear interpolation at argument λ between old forecast and new observation. Thus, if λ is very small, we rely principally on a weighted average of past data and heavily discounting the new observation z_t. By contrast, if $\lambda = 1$ ($\theta = 0$), the evidence of past data is completely ignored, $\hat{z}_t(l) = z_t$, and the forecast for all future time is the current value. With $\lambda > 1$, we induce an extrapolation rather than an interpolation between $\hat{z}_{t-1}(l)$ and z_t. The forecast error must now be *magnified* in (5.4.2) to indicate the change in the forecast.

Forecast Function in Integrated Form. The eventual forecast function is the solution of $(1 - B)\hat{z}_t(l) = 0$. Thus, $\hat{z}_t(l) = b_0^{(t)}$, and since $q - p - d = 0$, it provides the forecast for all lead times, that is,

$$\hat{z}_t(l) = b_0^{(t)} \qquad l > 0 \tag{5.4.4}$$

For *any fixed origin*, $b_0^{(t)}$ is a constant, and the forecasts for all lead times will follow a straight line parallel to the time axis. However, the coefficient $b_0^{(t)}$ will be updated as a new observation becomes available and the origin advances. Thus, the forecast function can be thought of as a polynomial of degree zero in the lead time l, with a coefficient that is adaptive with respect to the origin t.

A comparison of (5.4.4) with (5.4.1) shows that

$$b_0^{(t)} = \hat{z}_t(l) = z_t - \theta a_t$$

Equivalently, by referring to (4.3.4), since the truncated integrated form of the model, relative to an initial origin k, is

$$\begin{aligned} z_t &= \lambda S_{t-k-1} a_{t-1} + a_t + (z_k - \theta a_k) \\ &= \lambda(a_{t-1} + \cdots + a_{k+1}) + a_t + (z_k - \theta a_k) \end{aligned}$$

it follows that

$$\hat{z}_t(l) = b_0^{(t)} = \lambda S_{t-k} a_t + (z_k - \theta a_k) = \lambda(a_t + \cdots + a_{k+1}) + (z_k - \theta a_k)$$

Also, $\psi_j = \lambda\,(j = 1, 2, \ldots)$ and hence the adaptive coefficient $b_0^{(t)}$ can be updated from origin t to origin $t + 1$ according to

$$b_0^{(t+1)} = b_0^{(t)} + \lambda a_{t+1} \tag{5.4.5}$$

similar to (5.4.2).

Forecast as a Weighted Average of Previous Observations. Since, for this process, the $\pi_j^{(l)}$ weights of (5.3.8) are also the weights for the one-step-ahead forecast, we can also write, using (4.3.6),

$$\hat{z}_t(l) = b_0^{(t)} = \lambda z_t + \lambda(1-\lambda)z_{t-1} + \lambda(1-\lambda)^2 z_{t-2} + \cdots \tag{5.4.6}$$

Thus, for the IMA(0, 1, 1) model, the forecast for all future time is an *exponentially weighted moving average* of current and past z's.

Example: Forecasting Series A. It will be shown in Chapter 7 that Series A is closely fitted by the model

$$(1 - B)z_t = (1 - 0.7B)a_t$$

In Figure 5.7, the forecasts at origins $t = 39, 40, 41, 42$, and 43 and also at origin $t = 79$ are shown for lead times $1, 2, \ldots, 20$. The weights π_j, which for this model are forecast weights for any lead time, are given in Table 5.3. These weights are shown diagrammatically in their appropriate positions for the forecast $\hat{z}_{39}(l)$ in Figure 5.7.

Variance Functions. Since for this model, $\psi_j = \lambda (j = 1, 2, \ldots)$, the expression (5.1.16) for the variance of the lead l forecast errors is

$$V(l) = \sigma_a^2[1 + (l-1)\lambda^2] \tag{5.4.7}$$

Using the estimate $s_a^2 = 0.101$, appropriate for Series A, in (5.4.7). 50 and 95% probability limits were calculated and are shown in Figure 5.7 for origin $t = 79$.

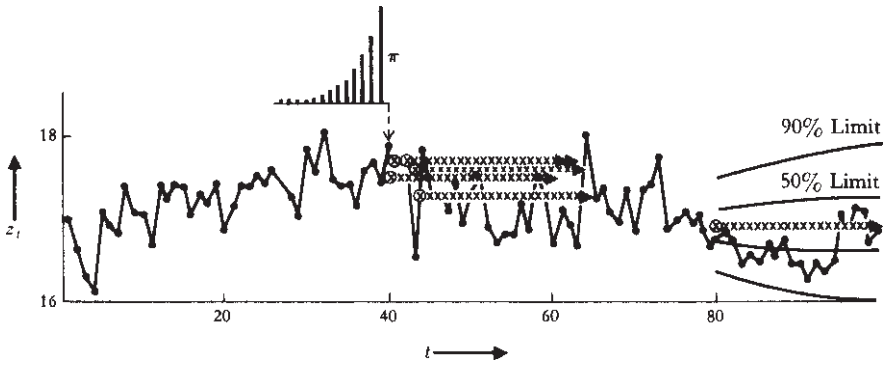

FIGURE 5.7 Part of Series A with forecasts at origins $t = 39, 40, 41, 42, 43$ and at $t = 79$.

TABLE 5.3 Forecast Weights Applied to Previous z's for Any Lead Time Used in Forecasting Series A with Model $\nabla z_t = (1 - 0.7B)a_t$

j	π_j	j	π_j
1	0.300	7	0.035
2	0.210	8	0.025
3	0.147	9	0.017
4	0.103	10	0.012
5	0.072	11	0.008
6	0.050	12	0.006

5.4.2 Forecasting an IMA(0, 2, 2) Process

Difference Equation Approach. We now consider the model $\nabla^2 z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$. At time $t + l$, the model may be written as

$$z_{t+l} = 2z_{t+l-1} - z_{t+l-2} + a_{t+l} - \theta_1 a_{t+l-1} - \theta_2 a_{t+l-2}$$

On taking conditional expectations at time t, we obtain

$$\begin{aligned}
\hat{z}_t(1) &= 2z_t - z_{t-1} - \theta_1 a_t - \theta_2 a_{t-1} \\
\hat{z}_t(2) &= 2\hat{z}_t(1) - z_t - \theta_2 a_t \\
\hat{z}_t(l) &= 2\hat{z}_t(l-1) - \hat{z}_t(l-2) \qquad l \geq 3
\end{aligned}$$

from which the forecasts may be calculated. Forecasting of the series of Figure 5.6 in this way was illustrated in Section (5.1.2). An alternative way of generating the first $L - 1$ of L forecasts is via the updating formula (5.2.5),

$$\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + \psi_l a_{t+1} \tag{5.4.8}$$

The truncated integrated model, as in (4.3.15), is

$$z_t = \lambda_0 S_{t-k-1} a_{t-1} + \lambda_1 S^{(2)}_{t-k-1} a_{t-1} + a_t + b_0^{(k)} + b_1^{(k)}(t-k) \tag{5.4.9}$$

where $\lambda_0 = 1 + \theta_2$ and $\lambda_1 = 1 - \theta_1 - \theta_2$, so that $\psi_j = \lambda_0 + j\lambda_1 (j = 1, 2, \ldots)$. Therefore, the updating function for this model is

$$\hat{z}_{t+1}(l) = \hat{z}_t(l+1) + (\lambda_0 + l\lambda_1)a_{t+1} \tag{5.4.10}$$

Forecast in Integrated Form. The eventual forecast function is the solution of $(1 - B)^2 \hat{z}_t(l) = 0$, that is, $\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)} l$. Since $q - p - d = 0$, the eventual forecast function provides the forecast for all lead times, that is,

$$\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)} l \qquad l > 0 \tag{5.4.11}$$

Thus, the forecast function is a linear function of the lead time l, with coefficients that are adaptive with respect to the origin t. The stochastic model in truncated integrated form is

$$z_{t+l} = \lambda_0 S_{t+l-k-1} a_{t+l-1} + \lambda_1 S^{(2)}_{t+l-k-1} a_{t+l-1} + a_{t+l} + b_0^{(k)} + b_1^{(k)}(t+l-k)$$

and taking expectations at origin t, we obtain

$$\begin{aligned}\hat{z}_t(l) &= \lambda_0 S_{t-k} a_t + \lambda_1 (l a_t + (l+1)a_{t-1} + \cdots + (l+t-k-1)a_{k+1}) \\ &\quad + b_0^{(k)} + b_1^{(k)}(t+l-k) \\ &= [\lambda_0 S_{t-k} a_t + \lambda_1 S^{(2)}_{t-k-1} a_{t-1} + b_0^{(k)} + b_1^{(k)}(t-k)] + (\lambda_1 S_{t-k} a_t + b_1^{(k)})l\end{aligned}$$

The adaptive coefficients may thus be identified as

$$\begin{aligned}b_0^{(t)} &= \lambda_0 S_{t-k} a_t + \lambda_1 S^{(2)}_{t-k-1} a_{t-1} + b_0^{(k)} + b_1^{(k)}(t-k) \\ b_1^{(t)} &= \lambda_1 S_{t-k} a_t + b_1^{(k)}\end{aligned} \tag{5.4.12}$$

or informally based on the infinite integrated form as $b_0^{(t)} = \lambda_0 S a_t + \lambda_1 S^2 a_{t-1}$ and $b_1^{(t)} = \lambda_1 S a_t$. Hence, their updating formulas are

$$\begin{aligned}b_0^{(t)} &= b_0^{(t-1)} + b_1^{(t-1)} + \lambda_0 a_t \\ b_1^{(t)} &= b_1^{(t-1)} + \lambda_1 a_t\end{aligned} \tag{5.4.13}$$

similar to relations (4.3.17). The additional slope term $b_1^{(t-1)}$, which occurs in the updating formula for $b_0^{(t)}$, is an adjustment to change the location parameter b_0 to a value appropriate to the new origin. It will also be noted that λ_0 and λ_1 are the fractions of the shock a_t, which are transmitted to the location parameter and the slope parameter, respectively.

Forecasts as a Weighted Average of Previous Observations. For this model, then, the forecast function is a straight line that passes through the forecasts $\hat{z}_t(1)$ and $\hat{z}_t(2)$. This is illustrated for the series in Figure 5.6, which shows the forecasts made at origin $t = 30$, with appropriate weight functions. It will be seen how dependence of the entire forecast function on previous z's in the series is a reflection of the dependence of $\hat{z}_t(1)$ and $\hat{z}_t(2)$ on these values. The weight functions for $\hat{z}_t(1)$ and $\hat{z}_t(2)$, plotted in the figure, have been given in Table 5.2.

The example illustrates once more that while the AR operator $\varphi(B)$ determines the form of function to be used (a straight line in this case), the MA operator is of importance in determining the way in which that function is ''fitted'' to previous data.

Dependence of the Adaptive Coefficients in the Forecast Function on Previous z's. Since for the general model, the values of the adaptive coefficients in the forecast function are determined by $\hat{z}_t(q), \hat{z}_t(q-1), \ldots, \hat{z}_t(q-p-d+1)$, which can be expressed as functions of the observations, it follows that the same is true for the adaptive coefficients themselves.

For instance, in the case of the model $\nabla^2 z_t = (1 - 0.9B + 0.5B^2)a_t$ of Figure 5.6,

$$\hat{z}_t(1) = b_0^{(t)} + b_1^{(t)} = \sum_{j=1}^{\infty} \pi_j^{(1)} z_{t+1-j}$$

$$\hat{z}_t(2) = b_0^{(t)} + 2b_1^{(t)} = \sum_{j=1}^{\infty} \pi_j^{(2)} z_{t+1-j}$$

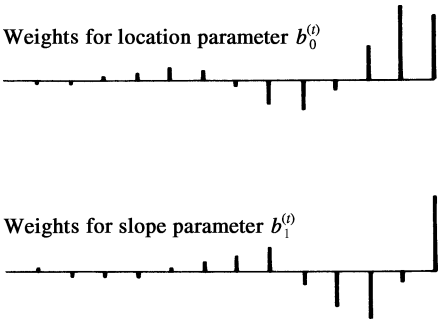

FIGURE 5.8 Weights applied to previous z's determining location and slope for the model $\nabla^2 z_t = (1 - 0.9B + 0.5B^2)a_t$.

so that

$$b_0^{(t)} = 2\hat{z}_t(1) - \hat{z}_t(2) = \sum_{j=1}^{\infty}(2\pi_j^{(1)} - \pi_j^{(2)})z_{t+1-j}$$

and

$$b_1^{(t)} = \hat{z}_t(2) - \hat{z}_t(1) = \sum_{j=1}^{\infty}(\pi_j^{(2)} - \pi_j^{(1)})z_{t+1-j}$$

These weight functions are plotted in Figure 5.8.

Variance of the Forecast Error. Using (5.1.16) and the fact that $\psi_j = \lambda_0 + j\lambda_1$, the variance of the lead l forecast error is

$$V(l) = \sigma_a^2[1 + (l-1)\lambda_0^2 + \frac{1}{6}l(l-1)(2l-1)\lambda_1^2 + \lambda_0\lambda_1 l(l-1)] \tag{5.4.14}$$

Using the estimate $s_a^2 = 0.032$, $\lambda_0 = 0.5$, and $\lambda_1 = 0.6$, the 50 and 95% limits are shown in Figure 5.6 for the forecast at origin $t = 30$.

5.4.3 Forecasting a General IMA(0, d, q) Process

As an example, consider the process of order (0, 1,3):

$$(1 - B)z_{t+l} = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)a_{t+1}$$

Taking conditional expectations at time t, we obtain

$$\begin{aligned}
\hat{z}_t(1) - z_t &= -\theta_1 a_t - \theta_2 a_{t-1} - \theta_3 a_{t-2} \\
\hat{z}_t(2) - \hat{z}_t(1) &= -\theta_2 a_t - \theta_3 a_{t-1} \\
\hat{z}_t(3) - \hat{z}_t(2) &= -\theta_3 a_t \\
\hat{z}_t(l) - \hat{z}_t(l-1) &= 0 \qquad l = 4, 5, 6, \ldots
\end{aligned}$$

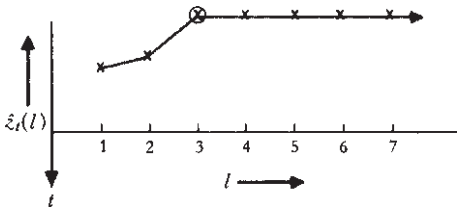

FIGURE 5.9 Forecast function for an IMA(0, 1, 3) process.

Hence, $\hat{z}_t(l) = \hat{z}_t(3) = b_0^{(t)}$ for all $l > 2$, as expected, since $q - p - d = 2$. As shown in Figure 5.9, the forecast function makes two initial ''jumps,'' depending on previous a's, before leveling out to the eventual forecast function.

For the IMA$(0, d, q)$ process, the eventual forecast function satisfies the difference equation $(1 - B)^d \hat{z}_t(l) = 0$, and has for its solution, a polynomial in l of degree $d - 1$:

$$\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)} l + b_2^{(t)} l^2 + \cdots + b_{d-1}^{(t)} l^{d-1}$$

This will provide the forecasts $\hat{z}_t(l)$ for $l - q - d$. The coefficients $b_0^{(t)}, b_1^{(t)}, \ldots, b_{d-1}^{(t)}$ must be updated progressively as the origin advances. The forecast for origin t will make $q - d$ initial ''jumps,'' which depend on $a_t, a_{t-1}, \ldots, a_{t-q+1}$, and after this, will follow the polynomial above.

5.4.4 Forecasting Autoregressive Processes

Consider a process of order $(p, d, 0)$, $\varphi(B)z_t = a_t$. The eventual forecast function is the solution of $\varphi(B)\hat{z}_t(l) = 0$. It applies for all lead times and passes through the last $p + d$ available values of the series. For example, the model for the IBM stock series (Series B) is very nearly

$$(1 - B)z_t = a_t$$

so that

$$\hat{z}_t(l) \approx z_t$$

The best forecast for all future time is very nearly the current value of the stock. The weight function for $\hat{z}_t(l)$ is a spike at time t and there is no averaging over past history.

Stationary Autoregressive Models. The stationary AR(p) process $\phi(B)\tilde{z}_t = a_t$ will in general produce a forecast function that is a mixture of exponentials and damped sines. In particular, for $p = 1$, the model

$$(1 - \phi B)\tilde{z}_t = a_t \qquad -1 < \phi < 1$$

has a forecast function that, for all $l > 0$, is the solution of $(1 - \phi B)\hat{\tilde{z}}_t(l) = 0$. Thus,

$$\hat{\tilde{z}}_t(l) = b_0^{(t)} \phi^l \qquad l > 0 \tag{5.4.15}$$

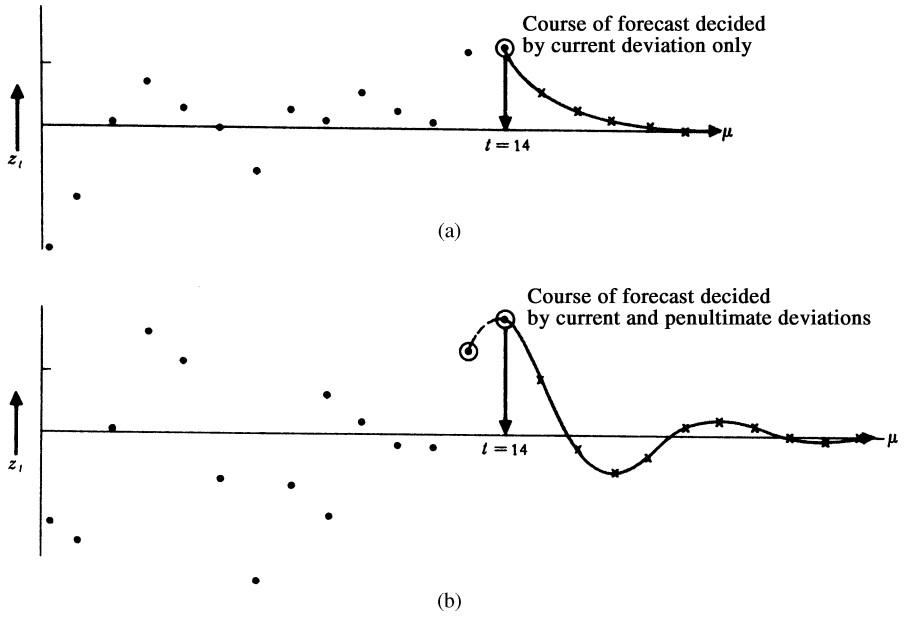

FIGURE 5.10 Forecast functions for (a) the AR(1) process $(1 - 0.5B)\tilde{z}_t = a_t$, and (b) the AR(2) process $(1 - 0.75B + 0.5B^2)\tilde{z}_t = a_t$ from a time origin $t = 14$.

Also, $\hat{\tilde{z}}_t(1) = \phi\tilde{z}_t$, so that $b_0^{(t)} = \tilde{z}_t$ and

$$\hat{\tilde{z}}_t(l) = \tilde{z}_t\phi^l$$

So, the forecasts for the original process z_t are $\hat{z}_t(l) = \mu + \phi^l(z_t - \mu)$.

Hence, the minimum mean square error forecast predicts the current deviation from the mean decaying exponentially to zero. In Figure 5.10(a) a time series is shown that is generated from the process $(1 - 0.5B)\tilde{z}_t = a_t$, with the forecast function at origin $t = 14$. The course of this function is seen to be determined entirely by the single deviation $\tilde{z}_{14}$. Similarly, the minimum mean square error forecast for a second-order autoregressive process is such that the current deviation from the mean is predicted to decay to zero via a damped sine wave or a mixture of two exponentials. Figure 5.10(b) shows a time series generated from the process $(1 - 0.75B + 0.50B^2)\tilde{z}_t = a_t$ and the forecast at origin $t = 14$. Here the course of the forecast function at origin t is determined entirely by the last two deviations, $\tilde{z}_{14}$ and $\tilde{z}_{13}$.

Variance Function for the Forecast from an AR(1) Process. Since the AR(1) process at time $t + l$ may be written as

$$\tilde{z}_{t+l} = a_{t+l} + \phi a_{t+l-1} + \cdots + \phi^{l-1}a_{t+1} + \phi^l\tilde{z}_t$$

it follows from (5.4.15) that

$$e_t(l) = \tilde{z}_{t+l} - \hat{\tilde{z}}_t(l) = a_{t+l} + \phi a_{t+l-1} + \cdots + \phi^{l-1}a_{t+1}$$

Hence,

$$V(l) = \text{var}[e_t(l)] = \sigma_a^2(1 + \phi^2 + \cdots + \phi^{2(l-1)}) = \frac{\sigma_a^2(1 - \phi^{2l})}{1 - \phi^2} \tag{5.4.16}$$

We see that for this stationary process, as l tends to infinity the variance increases to a constant value $\gamma_0 = \sigma_a^2/(1 - \phi^2)$, associated with the variation of the process about the ultimate forecast μ. This is in contrast to the behavior of forecast variance functions for nonstationary models that ''blow up'' for large lead times.

Nonstationary Autoregressive Models of Order (p, d, 0). For the model

$$\phi(B)\nabla^d z_t = a_t$$

the dth difference of the process decays back to its mean when projected several steps ahead. The mean of $\nabla^d z_t$ will usually be assumed to be zero unless contrary evidence is available. When needed, it is possible to introduce a nonzero mean by replacing $\nabla^d z_t$ by the deviation $(\nabla^d z_t - \mu_w)$ in the model. For example, consider the model

$$(1 - \phi B)(\nabla z_t - \mu_w) = a_t \tag{5.4.17}$$

After substituting $t + j$ for t and taking conditional expectations at origin t, we readily obtain [compare with (5.4.15) et seq.]

$$\hat{z}_t(j) - \hat{z}_t(j-1) - \mu_w = \phi^j(z_t - z_{t-1} - \mu_w)$$

or $\hat{w}_t(j) - \mu_w = \phi^j(w_t - \mu_w)$, where $w_t = \nabla z_t$. This shows how the forecasted *difference* decays exponentially from the initial value $w_t = z_t - z_{t-1}$ to its mean value μ_w. On summing this expression from $j = 1$ to $j = l$, that is, using $\hat{z}_t(l) = \hat{w}_t(l) + \cdots + \hat{w}_t(1) + z_t$, we obtain the forecast function

$$\hat{z}_t(l) = z_t + \mu_w l + (z_t - z_{t-1} - \mu_w)\frac{\phi(1 - \phi^l)}{1 - \phi} \qquad l \geq 1$$

that approaches asymptotically the straight line

$$f(l) = z_t + \mu_w l + (z_t - z_{t-1} - \mu_w)\frac{\phi}{1 - \phi}$$

with deterministic slope μ_w. If the forecasts are generated using the function sarima.for() in the astsa package in R, a deterministic slope can be incorporated into the forecast function by setting the argument no.constant=FALSE. The treatment of the constant term can have a big impact on the forecasts and should be considered carefully when a possible trend might be present.

We now consider the forecasting of some important mixed models.

5.4.5 Forecasting a (1, 0, 1) Process

Difference Equation Approach. Consider the stationary model

$$(1 - \phi B)\tilde{z}_t = (1 - \theta B)a_t$$

The forecasts are readily obtained from

$$\begin{aligned} \hat{\tilde{z}}_t(1) &= \phi\tilde{z}_t - \theta a_t \\ \hat{\tilde{z}}_t(l) &= \phi\hat{\tilde{z}}_t(l-1) \qquad l \geq 2 \end{aligned} \tag{5.4.18}$$

The forecasts decay geometrically to the mean, as in the first-order autoregressive process, but with a lead 1 forecast modified by a factor depending on $a_t = z_t - \hat{z}_{t-1}(1)$. The ψ weights are

$$\psi_j = (\phi - \theta)\phi^{j-1} \qquad j = 1, 2, \ldots$$

and hence, using (5.2.5), the updated forecasts for lead times $1, 2, \ldots, L-1$ could be obtained from previous forecasts for lead times $2, 3, \ldots, L$ according to

$$\hat{\tilde{z}}_{t+1}(l) = \hat{\tilde{z}}_t(l+1) + (\phi - \theta)\phi^{l-1}a_t + 1$$

Integrated Form. The eventual forecast function for all $l > 0$ is the solution of $(1 - \phi B)\hat{\tilde{z}}_t(l) = 0$, that is,

$$\hat{\tilde{z}}_t(l) = b_0^{(t)}\phi^l \qquad l > 0$$

However,

$$\hat{\tilde{z}}_t(l) = b_0^{(t)}\phi = \phi\tilde{z}_t - \theta a_t = \left[\left(1 - \frac{\theta}{\phi}\right)\tilde{z}_t + \frac{\theta}{\phi}\hat{\tilde{z}}_{t-1}(1)\right]\phi$$

Thus,

$$\hat{\tilde{z}}_t(1) = \left[\left(1 - \frac{\theta}{\phi}\right)\tilde{z}_t + \frac{\theta}{\phi}\hat{\tilde{z}}_{t-1}(1)\right]\phi^l \tag{5.4.19}$$

Hence, the forecasted deviation at lead l decays exponentially from an initial value, which is a linear interpolation between the previous lead 1 forecasted deviation and the current deviation. When ϕ is equal to unity, the forecast for all lead times becomes the familiar exponentially weighted moving average and (5.4.19) becomes equal to (5.4.3).

Weights Applied to Previous Observations. The π weights, and hence the weights applied to previous observations to obtain the lead 1 forecasts, as

$$\pi_j = (\phi - \theta)\theta^{j-1} \qquad j = 1, 2, \ldots$$

Note that the weights for this stationary process sum to $(\phi - \theta)/(1 - \theta)$ and not to unity. If ϕ were equal to 1, the process would become a nonstationary IMA(0, 1, 1) process, the weights would then sum to unity, and the behavior of the generated series would be independent of the level of z_t.

For example, Series A is later fitted to a $(1, 0, 1)$ model with $\phi = 0.9$ and $\theta = 0.6$, and hence the weights are $\pi_1 = 0.30$, $\pi_2 = 0.18$, $\pi_3 = 0.11$, $\pi_4 = 0.07$, …, which sum to

0.75. The forecasts (5.4.19) decay very slowly to the mean, and for short lead times are practically indistinguishable from the forecasts obtained from the alternative IMA(0, 1, 1) model $\nabla z_t = a_t - 0.7a_{t-1}$, for which the weights are $\pi_1 = 0.30$, $\pi_2, = 0.21$, $\pi_3 = 0.15$, $\pi_4 = 0.10$, and so on, and sum to unity. The latter model has the advantage that it does not tie the process to a fixed mean.

Variance Function. Since the ψ weights are given by

$$\psi_j = (\phi - \theta)\phi^{j-1} \qquad j = 1, 2, \ldots$$

it follows that the variance function is

$$V(l) = \sigma_a^2 \left[1 + (\phi - \theta)^2 \frac{1 - \phi^{2(l-1)}}{1 - \phi^2}\right] \tag{5.4.20}$$

which increases asymptotically to the value $\sigma_a^2(1 - 2\phi\theta + \theta^2)/(1 - \phi^2)$, the variance γ_0 of the process.

5.4.6 Forecasting a (1, 1, 1) Process

Another important mixed model is the nonstationary (1, 1, 1) process:

$$(1 - \phi B)(1 - B)z_t = (1 - \theta B)a_t$$

Difference Equation Approach. At time $t + 1$, the model may be written

$$z_{t+1} = (1 + \phi)z_{t+l-1} - \phi z_{t+l-2} + a_{t+l} - \theta a_{t+l-1}$$

On taking conditional expectations, we obtain

$$\begin{aligned} \hat{z}_t(1) &= (1 + \phi)z_t - \phi z_{t-1} - \theta a_t \\ \hat{z}_t(l) &= (1 + \phi)\hat{z}_t(l - 1) - \phi\hat{z}_t(l - 2) \qquad l > 1 \end{aligned} \tag{5.4.21}$$

Integrated Form. Since $q < p + d$, the eventual forecast function for all $l > 0$ is the solution of $(1 - \phi B)(1 - B)\hat{z}_t(l) = 0$, which is

$$\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)}\phi^l$$

Substituting for $\hat{z}_t(1)$ and $\hat{z}_t(2)$ in (5.4.21), we find explicitly that

$$\begin{aligned} b_0^{(t)} &= z_t + \frac{\phi}{1 - \phi}(z_t - z_{t-1}) - \frac{\theta}{1 - \theta}a_t \\ b_1^{(t)} &= \frac{\theta a_t - \phi(z_t - z_{t-1})}{1 - \phi} \end{aligned}$$

Thus, finally,

$$\hat{z}_t(l) = z_t + \phi\frac{1-\phi^l}{1-\phi}(z_t - z_{t-1}) - \theta\frac{1-\phi^l}{1-\phi}a_t \tag{5.4.22}$$

It is evident that for large l, the forecast tends to $b_0^{(t)}$.

Weights Applied to Previous Observations. Eliminating a_t from (5.4.22), we obtain the alternative form for the forecast in terms of previous z's:

$$\hat{z}_t(l) = \left[1 - \frac{\theta-\phi}{1-\phi}(1-\phi^l)\right] z_t + \left[\frac{\theta-\phi}{1-\phi}(1-\phi^l)\right] \bar{z}_{t-1}(\theta) \tag{5.4.23}$$

where $\bar{z}_{t-1}(\theta)$ is an exponentially weighted moving average with parameter θ, that is, $\bar{z}_{t-1}(\theta) = (1-\theta)\sum_{j=1}^{\infty}\theta^{j-1}z_{t-j}$. Thus, the π weights for the process consist of a ''spike'' at time t and an EWMA starting at time $t-1$. If we refer to $(1-\alpha)x + \alpha y$ as a linear interpolation between x and y at argument α, the forecast (5.4.23) is a linear interpolation between z and $\bar{z}_{t-1}(\theta)$. The argument for lead time 1 is $\theta - \phi$, but as the lead time is increased, the argument approaches $(\theta-\phi)/(1-\phi)$. For example, when $\theta = 0.9$ and $\phi = 0.5$, the lead 1 forecast is

$$\hat{z}_t(1) = 0.6z_t + 0.4\bar{z}_{t-1}(\theta)$$

and for long lead times, the forecast approaches

$$\hat{z}_t(\infty) = 0.2z_t + 0.8\bar{z}_{t-1}(\theta)$$

5.5 USE OF STATE-SPACE MODEL FORMULATION FOR EXACT FORECASTING

5.5.1 State-Space Model Representation for the ARIMA Process

The use of state-space models for time series analysis began with the work of Kalman (1960) and many of the early developments took place in the field of engineering. These models consist of a state equation that describes the evolution of a dynamic system in time, and a measurement equation that represents the observations as linear combinations of the unobserved state variable corrupted by additive noise. In engineering applications, the state variable generally represents a well-defined set of physical variables, but these variables are not directly observable, and the state equation represents the dynamics that govern the system. In statistical applications, the state-space model is a convenient form to represent many types of models, including autoregressive–moving average (ARMA) models, structural component models of ''signal-plus-noise'' form, or time-varying parameter models. In the literature, state-space models have been used for forecasting, maximum likelihood estimation of parameters, signal extraction, seasonal adjustments, and other applications (see, for example, Durbin and Koopman, 2012). In this section, we introduce the state-space form of an ARIMA model and discuss its use in exact finite sample forecasting. Other applications involving the use of state-space models for likelihood calculations, estimation of structural components, treatment of missing values, and applications related to vector ARMA models will be discussed in Sections 7.4, 9.4, 13.3, and 14.6.

For an ARIMA(p, d, q) process $\varphi(B)z_t = \theta(B)a_t$, define the forecasts $\hat{z}_t(j) = E_t[z_t + j]$ as in Section 5.1, for $j = 0, 1, \ldots, r$, with $r = \max(p + d, q + 1)$, and $\hat{z}_t(0) = z_t$. From the updating equations (5.2.5), we have $\hat{z}_t(j-1) = \hat{z}_{t-1}(j) + \psi_{j-1}a_t$, $j = 1, 2, \ldots, r-1$. Also for $j = r > q$, recall from (5.3.2) that

$$\hat{z}_t(j-1) = \hat{z}_{t-1}(j) + \psi_{j-1}a_t = \sum_{i=1}^{p+d} \varphi_i \hat{z}_{t-1}(j-i) + \psi_{j-1}a_t$$

So we define the ''state'' vector at time t, $\boldsymbol{Y}_t$, with r components as $\boldsymbol{Y}_t = (z_t, \hat{z}_t(1), \ldots, \hat{z}_t(r-1))'$. Then from the relations above, we find that the vector $\boldsymbol{Y}_t$ satisfies the first-order system of equations:

$$\boldsymbol{Y}_t = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & 1 \\ \varphi_r & \varphi_{r-1} & \cdot & \cdot & \cdot & \varphi_1 \end{bmatrix} \boldsymbol{Y}_{t-1} + \begin{bmatrix} 1 \\ \psi_1 \\ \cdot \\ \cdot \\ \cdot \\ \psi_{r-1} \end{bmatrix} a_t \tag{5.5.1}$$

where $\varphi_i = 0$ if $i > p + d$. So we have

$$\boldsymbol{Y}_t = \boldsymbol{\Phi}\boldsymbol{Y}_{t-1} + \boldsymbol{\Psi}a_t \tag{5.5.2}$$

together with the observation equation

$$Z_t = z_t + N_t = [1, 0, \ldots, 0]\boldsymbol{Y}_t + N_t = \mathbf{H}\boldsymbol{Y}_t + N_t \tag{5.5.3}$$

where the additional noise N_t would be present only if the process z_t is observed subject to additional white noise; otherwise, we simply have $z_t = \mathbf{H}\boldsymbol{Y}_t$. The last two equations above constitute what is known as a state-space representation of the model, which consists of a state or transition equation (5.5.2) and an observation equation (5.5.3), and $\boldsymbol{Y}_t$ is known as the state vector. We note that there are many other constructions of the state vector $\boldsymbol{Y}_t$ that will give rise to state-space equations of the general form of (5.5.2) and (5.5.3); that is, the state-space form of an ARIMA model is not unique. The two equations of the form above, in general, represent what is known as a state-space model, with unobservable state vector $\boldsymbol{Y}_t$ and observations Z_t, and can arise in time series settings more general than the context of ARIMA models.

Consider a state-space model of a slightly more general form, with state equation

$$\boldsymbol{Y}_t = \boldsymbol{\Phi}_t\boldsymbol{Y}_{t-1} + \boldsymbol{a}_t \tag{5.5.4}$$

and observation equation

$$Z_t = \mathbf{H}_t\boldsymbol{Y}_t + N_t \tag{5.5.5}$$

where it is assumed that $\boldsymbol{a}_t$ and N_t are independent white noise processes, $\boldsymbol{a}_t$ is a vector white noise process with covariance matrix $\boldsymbol{\Sigma}_a$, and N_t has variance σ_N^2. In this model, the (unobservable) state vector $\boldsymbol{Y}_t$ summarizes the state of the dynamic system through time t, and the state equation (5.5.4) describes the evolution of the dynamic system in time, while the measurement equation (5.5.5) indicates that the observations Z_t consist of linear

combinations of the state variables corrupted by additive white noise. The matrix $\mathbf{\Phi}_t$ in (5.5.4) is an $r \times r$ transition matrix and $\mathbf{H}_t$ in (5.5.5) is a $1 \times \boldsymbol{r}$ vector, which are allowed to vary with time t. Often, in applications these are constant matrices, $\mathbf{\Phi}_t \equiv \mathbf{\Phi}$ and $\mathbf{H}_t \equiv \mathbf{H}$ for all t, that do not depend on t, as in the state-space form (5.5.2) and (5.5.3) of the ARIMA model. In this case, the system or model is said to be *time invariant*. The minimal dimension r of the state vector $\boldsymbol{Y}_t$ in a state-space model needs to be sufficiently large so that the dynamics of the system can be represented by the simple Markovian (first-order) structure as in (5.5.4).

5.5.2 Kalman Filtering Relations for Use in Prediction

For the general state-space model (5.5.4) and (5.5.5), define the finite sample optimal (minimum mean square error matrix) estimate of the state vector $\boldsymbol{Y}_{t+l}$ based on observations $Z_t, \ldots, Z_1$ over the finite past time period, as

$$\hat{\boldsymbol{Y}}_{t+l|t} = E[\boldsymbol{Y}_{t+l} | Z_t, \ldots, Z_1]$$

with

$$\mathbf{V}_{t+1|t} = E[(\boldsymbol{Y}_{t+l} - \hat{\boldsymbol{Y}}_{t+l|t})(\boldsymbol{Y}_{t+l} - \hat{\boldsymbol{Y}}_{t+l|t})']$$

equal to the error covariance matrix. A convenient computational procedure, known as the *Kalman filter* equations, is then available to obtain the current estimate $\hat{\boldsymbol{Y}}_{t|t}$, in particular. It is known that, starting from some appropriate initial values $\boldsymbol{Y}_0 \equiv \hat{\boldsymbol{Y}}_{0|0}$ and $\mathbf{V}_0 \equiv \mathbf{V}_{0|0}$, the optimal filtered estimate, $\hat{\boldsymbol{Y}}_{t|t}$, is given through the following recursive relations:

$$\hat{\boldsymbol{Y}}_{t|t} = \hat{\boldsymbol{Y}}_{t|t-1} + \boldsymbol{K}_t \left(Z_t - \mathbf{H}_t \hat{\boldsymbol{Y}}_{t|t-1} \right) \tag{5.5.6}$$

where

$$\mathbf{K}_t = \mathbf{V}_{t|t-1} \mathbf{H}_t' [\mathbf{H}_t \mathbf{V}_{t|t-1} \mathbf{H}_t' + \sigma_N^2]^{-1} \tag{5.5.7}$$

with

$$\hat{\boldsymbol{Y}}_{t|t-1} = \mathbf{\Phi}_t \hat{\boldsymbol{Y}}_{t-1|t-1} \qquad \mathbf{V}_{t|t-1} = \mathbf{\Phi}_t \mathbf{V}_{t-1|t-1} \mathbf{\Phi}_t' + \mathbf{\Sigma}_a \tag{5.5.8}$$

and

$$\begin{aligned} \mathbf{V}_{t|t} &= [\mathbf{1} - \mathbf{K}_t \mathbf{H}_t] \mathbf{V}_{t|t-1} \\ &= \mathbf{V}_{t|t-1} - \mathbf{V}_{t|t-1} \mathbf{H}_t' [\mathbf{H}_t \mathbf{V}_{t|t-1} \mathbf{H}_t' + \sigma_N^2]^{-1} \mathbf{H}_t \mathbf{V}_{t|t-1} \end{aligned} \tag{5.5.9}$$

for t = 1, 2, ….

In (5.5.6), the quantity $a_{t|t-1} = Z_t - \mathbf{H}_t \hat{\boldsymbol{Y}}_{t|t-1} \equiv Z_t - \hat{Z}_{t|t-1}$ is called the (finite sample) innovation at time t, because it is the new information provided by the measurement Z_t that was not available from the previous observed (finite) history of the system. The factor $\mathbf{K}_t$ is called the *Kalman gain* matrix. The filtering procedure in (5.5.6) has the recursive ''prediction–correction'' or ''updating'' form, and the validity of these equations as representing the minimum mean square error predictor can readily be verified through the principles of updating. For example, verification of (5.5.6) follows from the principle,

for linear prediction, that

$$E[\boldsymbol{Y}_t|Z_t,\ldots,Z_1] = E[\boldsymbol{Y}_t|Z_t - \hat{Z}_{t|t-1}, Z_{t-1},\ldots,Z_1]$$
$$= E[\boldsymbol{Y}_t|Z_{t-1},\ldots,Z_1] + E[\boldsymbol{Y}_t|Z_t - \hat{Z}_{t|t-1}]$$

since $a_{t|t-1} = Z_t - \hat{Z}_{t|t-1}$ is independent of $Z_{t-1},\ldots,Z_1$. From (5.5.6), it is seen that the estimate of $\boldsymbol{Y}_t$ based on observations through time t equals the prediction of $\boldsymbol{Y}_t$ from observations through time $t-1$ updated by the factor $\boldsymbol{K}_t$ times the innovation $a_{t|t-1}$. Equation (5.5.7) indicates that $\boldsymbol{K}_t$ can be interpreted as the regression coefficients of $\boldsymbol{Y}_t$ on the innovation $a_{t|t-1}$, with $\text{var}[a_{t|t-1}] = \mathbf{H}_t\mathbf{V}_{t|t-1}\mathbf{H}_t' + \sigma_N^2$ and $\text{cov}[\boldsymbol{Y}_t, a_{t|t-1}] = \mathbf{V}_{t|t-1}\mathbf{H}_t'$ following directly from (5.5.5) since $a_{t|t-1} = \mathbf{H}_t(Z_t - \hat{Z}_{t|t-1}) + N_t$. Thus, the general *updating relation* is

$$\hat{\boldsymbol{Y}}_{t|t} = \hat{\boldsymbol{Y}}_{t|t-1} + \text{cov}[\boldsymbol{Y}_t, a_{t|t-1}]\{\text{var}[a_{t|t-1}]\}^{-1}a_{t|t-1}$$

where $a_{t|t-1} = Z_t - \hat{Z}_{t|t-1}$, and the relation in (5.5.9) is the usual updating of the error covariance matrix to account for the new information available from the innovation $a_{t|t-1}$, while the *prediction relations* (5.5.8) follow directly from (5.5.4).

In general, forecasts of future state values are available directly as $\hat{\boldsymbol{Y}}_{t+l|t} = \boldsymbol{\Phi}_{t+l}\hat{\boldsymbol{Y}}_{t+l-1|t}$ for $l = 1, 2, \ldots$, with the covariance matrix of the forecast errors generated recursively essentially through (5.5.8) as

$$\mathbf{V}_{t+l|t} = \boldsymbol{\Phi}_{t+l}\mathbf{V}_{t+l-1|t}\boldsymbol{\Phi}_{t+l}' + \boldsymbol{\Sigma}_a$$

Finally, forecasts of future observations, $Z_{t+l} = \mathbf{H}_{t+l}\boldsymbol{Y}_{t+l} + N_{t+l}$, are then available as $\hat{Z}_{t+l|t} = \mathbf{H}_{t+l}\hat{\boldsymbol{Y}}_{t+l|t}$ with forecast error variance

$$v_{t+l|t} = E[(Z_{t+l} - \hat{Z}_{t+l|t})^2] = \mathbf{H}_{t+l}\mathbf{V}_{t+l|t}\mathbf{H}_{t+l}' + \sigma_N^2$$

Use for Exact Forecasting in ARIMA Models. For ARIMA models, with state-space representation (5.5.2) and (5.5.3) and $Z_t = z_t = \mathbf{H}\boldsymbol{Y}_t$ with $\mathbf{H} = [1, 0, \ldots, 0]$, the Kalman filtering procedure constitutes an alternative method to obtain exact finite sample forecasts, based on data $z_t, z_{t-1}, \ldots, z_1$, for future values in the ARIMA process, subject to specification of appropriate initial conditions to use in (5.5.6) to (5.5.9). For stationary zero-mean processes z_t, the appropriate initial values are $\hat{\boldsymbol{Y}}_{0|0} = \mathbf{0}$, a vector of zeros, and $\mathbf{V}_{0|0} = \text{cov}[\boldsymbol{Y}_0] \equiv \mathbf{V}_*$, the covariance matrix of $\boldsymbol{Y}_0$, which can easily be determined under stationarity through the definition of $\boldsymbol{Y}_t$. Specifically, since the state vector $\boldsymbol{Y}_t$ follows the stationary vector AR(1) model $\boldsymbol{Y}_t = \boldsymbol{\Phi}\boldsymbol{Y}_{t-1} + \Psi a_t$, its covariance matrix $\mathbf{V}_* = \text{cov}[\boldsymbol{Y}_t]$ satisfies $\mathbf{V}_* = \boldsymbol{\Phi}\mathbf{V}_*\boldsymbol{\Phi}' + \sigma_a^2\boldsymbol{\Psi}\boldsymbol{\Psi}'$, which can be readily solved for $\mathbf{V}_*$. For nonstationary ARIMA processes, additional assumptions need to be specified (see, for example, Ansley and Kohn (1985) and Bell and Hillmer (1987)).

The forecasts of the ARIMA process z_t are obtained recursively as indicated above, with l-step-ahead forecast $\hat{z}_{t+l|t} = \mathbf{H}\hat{\boldsymbol{Y}}_{t+l|t}$, the first element of the vector $\hat{\boldsymbol{Y}}_{t+l|t}$, where

$$\hat{\boldsymbol{Y}}_{t+l|t} = \boldsymbol{\Phi}\hat{\boldsymbol{Y}}_{t+l-1|t}$$

with forecast error variance $v_{t+l|t} = \mathbf{H}\mathbf{V}_{t+l|t}\mathbf{H}'$. The ''steady-state'' values of the Kalman filtering procedure l-step-ahead forecasts $\hat{z}_{t+l|t}$ and their forecast error variances $v_{t+l|t}$,

which are rapidly approached as t increases, will be identical to the expressions given in Sections 5.1 and 5.2, $\hat{z}_t(l)$ and $V(l) = \sigma_a^2(1 + \sum_{j=1}^{l-1} \psi_j^2)$.

In particular, for the ARIMA process in state-space form, we can obtain the exact (finite sample) one-step-ahead forecasts:

$$\hat{z}_{t|t-1} = E[z_t | z_{t-1}, \ldots, z_1] = \mathbf{H}\hat{\boldsymbol{Y}}_{t|t-1}$$

and their error variances $v_t \equiv \mathbf{H}\mathbf{V}_{t|t-1}\mathbf{H}'$, conveniently through the Kalman filtering equations (5.5.6)–(5.5.9). This can be particularly useful for evaluation of the likelihood function, based on n observations $z_1, \ldots, z_n$ from the ARIMA process, applied to the problem of maximum likelihood estimation of model parameters (see, for example, Jones (1980) and Gardner et al. (1980)). This will be discussed again in Section 7.4.

Innovations Form of State-Space Model and Steady State for Time-Invariant Models. One particular alternative form of the general state variable model, referred to as the innovations or prediction error representation, is worth noting. If we set $\boldsymbol{Y}_t^* = \hat{\boldsymbol{Y}}_{t|t-1}$ and $a_t^* = a_{t|t-1} = Z_t - \mathbf{H}_t\hat{\boldsymbol{Y}}_{t|t-1}$, then from (5.5.6) and (5.5.8) we have

$$\boldsymbol{Y}_{t+1}^* = \boldsymbol{\Phi}_{t+1}\boldsymbol{Y}_t^* + \boldsymbol{\Phi}_{t+l}\mathbf{K}_t a_t^* \equiv \boldsymbol{\Phi}_{t+1}\boldsymbol{Y}_t^* + \boldsymbol{\Psi}_t^* a_t^* \qquad \text{and} \qquad Z_t = \mathbf{H}_t\boldsymbol{Y}_t^* + a_t^*$$

which is also of the general form of a state-space model but with the same white noise process a_t^* (the one-step-ahead prediction errors) involved in both the transition and observation equations.

In the ''stationary case'' (i.e., time-invariant and stable case) of the state-space model, where $\boldsymbol{\Phi}_t \equiv \boldsymbol{\Phi}$ and $\mathbf{H}_t \equiv \mathbf{H}$ in (5.5.4) and (5.5.5) are constant matrices and $\boldsymbol{\Phi}$ has all eigenvalues less than 1 in absolute value, we can obtain the steady-state form of the innovations representation by setting $\boldsymbol{Y}_t^* = E[\boldsymbol{Y}_t | Z_{t-1}, Z_{t-2}, \ldots]$, the projection of $\boldsymbol{Y}_t$ based on the infinite past of $\{Z_t\}$. In this case, in the Kalman filter relations (5.5.7) to (5.5.9), the error covariance matrix $\mathbf{V}_{t+1|t}$ approaches the steady-state matrix $\mathbf{V} = \lim_{t\to\infty} \mathbf{V}_{t+1|t}$ as $t \to \infty$, which satisfies

$$\mathbf{V} = \boldsymbol{\Phi}\mathbf{V}\boldsymbol{\Phi}' - \boldsymbol{\Phi}\mathbf{V}\mathbf{H}'[\mathbf{H}\mathbf{V}\mathbf{H}' + \sigma_N^2]^{-1}\mathbf{H}\mathbf{V}\boldsymbol{\Phi}' + \boldsymbol{\Sigma}_a$$

Then, also, the Kalman gain matrix $\mathbf{K}_t$ in (5.5.7) approaches the steady-state matrix, $\mathbf{K}_t \to \mathbf{K}$, where $\mathbf{K} = \mathbf{V}\mathbf{H}'[\mathbf{H}\mathbf{V}\mathbf{H}' + \sigma_N^2]^{-1}$, $a_t^* = a_{t|t-1}$ tends to $a_t = Z_t - \mathbf{H}\boldsymbol{Y}_t^* \equiv Z_t - E[Z_t | Z_{t-1}, Z_{t-2}, \ldots]$, the one-step-ahead prediction errors, and $\sigma_{t|t-1}^2 = \text{var}[a_{t|t-1}] \to \sigma_a^2 = \text{var}[a_t]$, where $\sigma_a^2 = \mathbf{H}\mathbf{V}\mathbf{H}' + \sigma_N^2$, as $t \to \infty$. These steady-state filtering results for the time-invariant model case also hold under slightly weaker conditions than stability of the transition matrix $\boldsymbol{\Phi}$ (e.g., Harvey (1989), Section 3.3), such as in the nonstationary random walk plus noise model discussed in the example of Section 5.5.3. Hence, in the time-invariant situation, the state variable model can be expressed in the steady-state innovation or prediction error form as

$$\boldsymbol{Y}_{t+l}^* = \boldsymbol{\Phi}\boldsymbol{Y}_t^* + \boldsymbol{\Phi}\mathbf{K}a_t \equiv \boldsymbol{\Phi}\boldsymbol{Y}_t^* + \boldsymbol{\Psi}^* a_t \quad \text{and} \quad Z_t = \mathbf{H}\boldsymbol{Y}_t^* + a_t \tag{5.5.10}$$

In particular, for the ARIMA process $\varphi(B)z_t = \theta(B)a_t$ with no additional observation error so that $Z_t = z_t$, a prediction error form (5.5.10) of the state-space model can be given with state vector $\boldsymbol{Y}_{t+1}^* = (\hat{z}_t(1), \ldots, \hat{z}_t(r^*))'$ of dimension $r^* = \max(p + d, q)$, $\boldsymbol{\Psi}^* = (\psi_1, \ldots, \psi_{r^*})'$, and observation equation $z_t = \hat{z}_{t-1}(1) + a_t$. For example, consider the ARMA(1, 1) process $(1 - \phi B)z_t = (1 - \theta B)a_t$. In addition to the state-space form

with state equation given by (5.5.1) and $\boldsymbol{Y}_t = (z_t, \hat{z}_t(1))'$, we have the innovations form of its state-space representation simply as $\hat{z}_t(1) = \phi\hat{z}_{t-1}(1) + \psi^* a_t$ and $z_t = \hat{z}_{t-1}(1) + a_t$, or $Y^*_{t+1} = \phi Y^*_t + \psi^* a_t$ and $z_t = Y^*_t + a_t$ with the (single) state variable $Y^*_{t+1} = \hat{z}_t(1)$ and $\psi^* = \psi_1 = \phi - \theta$.

5.5.3 Smoothing Relations in the State Variable Model

Another problem of interest within the state variable model framework, particularly in applications to economics and business, is to obtain ''smoothed'' estimates of past values of the state vector $\boldsymbol{Y}_t$ given the observations $Z_1, \ldots, Z_n$ through some fixed time n. One convenient method to obtain the desired estimates, known as the *fixed-interval smoothing* algorithm, makes use of the Kalman filter estimates $\hat{\boldsymbol{Y}}_{t|t}$ obtainable through (5.5.6)–(5.5.9). The smoothing algorithm produces the minimum MSE estimator (predictor) of the state value $\boldsymbol{Y}_t$ given the observations through time n, $\hat{\boldsymbol{Y}}_{t|n} = E[\boldsymbol{Y}_t|Z_1, \ldots, Z_n]$. In general, define $\hat{\boldsymbol{Y}}_{t|T} = E[\boldsymbol{Y}_t|Z_1, \ldots, Z_T]$ and $\mathbf{V}_{t|T} = E[(\boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|T})(\boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|T})']$. We assume that the filtered estimates $\hat{\boldsymbol{Y}}_{t|t}$ and their error covariance matrices $\mathbf{V}_{t|t}$, for $t = 1, \ldots, n$, have already been obtained by the Kalman filter equations. Then, the optimal smoothed estimates are obtained by the (backward) recursive relations, in which the filtered estimate $\hat{\boldsymbol{Y}}_{t|t}$ is updated, as

$$\hat{\boldsymbol{Y}}_{t|n} = \hat{\boldsymbol{Y}}_{t|t} + \mathbf{A}_t(\hat{\boldsymbol{Y}}_{t+1|n} - \hat{\boldsymbol{Y}}_{t+1|t}) \tag{5.5.11}$$

where

$$\mathbf{A}_t = \mathbf{V}_{t|t}\boldsymbol{\Phi}'_{t+l|t}\mathbf{V}^{-1}_{t+1|t} \equiv \text{cov}[\boldsymbol{Y}_t, \boldsymbol{Y}_{t+1} - \hat{\boldsymbol{Y}}_{t+1|t}]\{\text{cov}[\boldsymbol{Y}_{t+1} - \hat{\boldsymbol{Y}}_{t+1|t}]\}^{-1} \tag{5.5.12}$$

and

$$\mathbf{V}_{t|n} = \mathbf{V}_{t|t} - \mathbf{A}_t(\mathbf{V}_{t+1|t} - \mathbf{V}_{t+1|n})\mathbf{A}'_t \tag{5.5.13}$$

The result (5.5.11) is established from the following argument. First, consider $\boldsymbol{u}_t = E[\boldsymbol{Y}_t|Z_1, \ldots, Z_t, \boldsymbol{Y}_{t+1} - \hat{\boldsymbol{Y}}_{t+1|t}, N_{t+1}, \boldsymbol{a}_{t+2}, N_{t+2}, \ldots, \boldsymbol{a}_n, N_n]$. Then, because $\{\boldsymbol{\alpha}_{t+j}, j \geq 2\}$ and $\{N_{t+j},\ j \geq 1\}$ are independent of the other conditioning variables in the definition of $\boldsymbol{u}_t$ and are also independent of $\boldsymbol{Y}_t$, we have $\boldsymbol{u}_t = \hat{\boldsymbol{Y}}_{t|t} + E[\boldsymbol{Y}_t|\boldsymbol{Y}_{t+1} - \hat{\boldsymbol{Y}}_{t+1|t}] = \hat{\boldsymbol{Y}}_{t|t} + \mathbf{A}_t(\boldsymbol{Y}_{t+1} - \hat{\boldsymbol{Y}}_{t+1|t})$, where $\mathbf{A}_t$ is given by (5.5.12). Thus, because the conditioning variables in $\boldsymbol{u}_t$ generate $Z_1, \ldots, Z_n$, it follows that

$$\begin{aligned}\hat{\boldsymbol{Y}}_{t|n} &= \boldsymbol{E}[\boldsymbol{Y}_t|Z_1, \ldots, Z_n] \\ &= E[\boldsymbol{u}_t|Z_1, \ldots, Z_n] = \hat{\boldsymbol{Y}}_{t|t} + \mathbf{A}_t(\hat{\boldsymbol{Y}}_{t+1|n} - \hat{\boldsymbol{Y}}_{t+1|t})\end{aligned}$$

as in (5.5.11). The relation (5.5.13) for the error covariance matrix follows from rather straightforward calculations. This derivation of the fixed-interval smoothing relations is given by Ansley and Kohn (1982).

Thus, it is seen from (5.5.11)–(5.5.13) that the optimal smoothed estimates $\hat{\boldsymbol{Y}}_{t|n}$ are obtained by first obtaining the filtered values $\hat{\boldsymbol{Y}}_{t|t}$ through the forward recursion of the Kalman filter relations, followed by the backward recursions of (5.5.11)–(5.5.13) for $t = n - 1, \ldots, 1$. This type of smoothing procedure has applications for estimation of trend and seasonal components (seasonal adjustment) in economic time series, as will be discussed in Section 9.4. When smoothed estimates $\hat{\boldsymbol{Y}}_{t|n}$ are desired only at a fixed time point (or

only at a few fixed points), for example, in relation to problems that involve the estimation of isolated missing values in a time series, then an alternative ''fixed-point'' smoothing algorithm may be useful (e.g., see Anderson and Moore (1979) or Brockwell and Davis (1991)).

Example. As a simple example of the state-space model and associated Kalman filtering and smoothing, consider a basic structural model in which an observed series Z_t is viewed as the sum of unobserved trend and noise components. To be specific, assume that the observed process can be represented as

$$Z_t = \mu_t + N_t \quad \text{where} \quad \mu_t = \mu_{t-1} + a_t$$

so that μ_t is a random walk process and N_t is an independent (white) noise process. This is a simple example of a time-invariant state-space model with $\mathbf{\Phi} = 1$ and $\mathbf{H} = 1$ in (5.5.4) and (5.5.5) and with the state vector $\boldsymbol{Y}_t = \mu_t$ representing an underlying (unobservable) ''trend or level'' process (or ''permanent'' component). For this model, application of the Kalman filter and associated smoothing algorithm can be viewed as the estimation of the underlying trend process μ_t based on the observed process Z_t. The Kalman filtering relations (5.5.6)–(5.5.9) for this basic model reduce to

$$\hat{\mu}_{t|t} = \hat{\mu}_{t-1|t-1} + K_t(Z_t - \hat{\mu}_{t-1|t-1}) = K_t Z_t + (1 - K_t)\hat{\mu}_{t-1|t-1}$$

where the gain is $K_t = V_{t|t-1}[V_t|t-1 + \sigma_N^2]^{-1}$, with

$$V_{t+1|t} = V_{t|t-1} - V_{t|t-1}[V_{t|t-1} + \sigma_N^2]^{-1}V_{t|t-1} + \sigma_a^2$$

Then $\hat{\mu}_{t|t}$ represents the current estimate of the trend component μ_t given the observations $Z_1, \ldots, Z_t$ through time t. The steady-state solution to the Kalman filter relations is obtained as $t \to \infty$ for V ($V = \lim_{t\to\infty} V_{t+1|t}$), which satisfies $V = V - V[V + \sigma_N^2]^{-1}V + \sigma_a^2$, that is, $V[V + \sigma_N^2]^{-1}V = \sigma_a^2$, and the corresponding steady-state gain is $K = V[V + \sigma_N^2]^{-1}$. In addition, the recursion (5.5.11) for the smoothed estimate of the trend component μ_t becomes

$$\begin{aligned}\hat{\mu}_{t|n} &= \hat{\mu}_{t|t} + A_t(\hat{\mu}_{t+1|n} - \hat{\mu}_{t+1|t}) \\ &= (1 - A_t)\hat{\mu}_{t|t} + A_t\hat{\mu}_{t+1|n} \qquad t = n-1, \ldots, 1\end{aligned}$$

noting that $\hat{\mu}_{t+1|t} = \hat{\mu}_{t|t}$, where $A_t = V_{t|t}V_{t+1|r}^{-1} = V_{t|t}\{V_{t|t} + \sigma_a^2\}^{-1}$ and $V_{t|t} = (1 - K_t)V_{t|t-1}$, with the recursion for the calculation of $V_{t|t-1}$ being as given above. Thus, the smoothed value is a weighted average of the filtered estimate $\hat{\mu}_{t|t}$ at time t and the smoothed estimate $\hat{\mu}_{t+1|n}$ at time $t + 1$. The steady-state form of this smoothing recursion is the same as above with a constant $A = \lim_{t\to\infty} A_t$, which can be found to equal $A = 1 - K$. Hence, the steady-state (backward) smoothing relation (5.5.11) for this example has the same form as the steady-state filter relation already mentioned; that is, they both have the form of an exponential weighted moving average (EWMA) with the same weight.

5.6 SUMMARY

The results of this chapter may be summarized as follows: Let $\bar{z}_t$ be the deviation of an observed time series from any known deterministic function of time $f(t)$. In particular, for a stationary series, $f(t)$ could be equal to μ, the mean of the series, or it could be equal to zero, so that $\bar{z}_t$ was the observed series. Then, consider the general ARIMA model

$$\phi(B)\nabla^d \bar{z}_t = \theta(B)a_t$$

or

$$\varphi(B)\bar{z}_t = \theta(B)a_t$$

Minimum Mean Square Error Forecast. Given the knowledge of the series up to some origin t, the minimum mean square error forecast $\bar{z}_t(l)(l > 0)$ of $\bar{z}_{t+1}$ is the conditional expectation

$$\hat{\bar{z}}_t(l) = [\bar{z}_{t+l}] = E[\bar{z}_{t+l}|\bar{z}_t, \bar{z}_{t-1}, \ldots]$$

Lead 1 Forecast Errors. A necessary consequence is that the lead 1 forecast errors are the generating a_t's in the model and are uncorrelated.

Calculation of the Forecasts. It is usually simplest in practice to compute the forecasts directly from the difference equation to give

$$\begin{aligned}\hat{\bar{z}}_1(l) =& \varphi_1[\bar{z}_{t+l-1}] + \cdots + \varphi_{p+d}[\bar{z}_{t+l-p-d}] + [a_{t+l}] - \theta_1[a_{t+l-1}] \\ & - \cdots - \theta_q[a_t + l - q]\end{aligned} \tag{5.6.1}$$

The conditional expectations in (5.6.1) are evaluated by inserting actual $\bar{z}$'s when these are known, forecasted $\bar{z}$'s for future values, actual a's when these are known, and zeros for future a's. The forecasting process may be initiated by approximating a's by zeros and, in practice, the appropriate form for the model and suitable estimates for the parameters are obtained by methods set out in Chapters 6–8.

Probability Limits for Forecasts. The probability limits may be obtained as follows:

1. By first calculating the ψ weights from

$$\begin{aligned}\psi_0 &= 1 \\ \psi_1 &= \varphi_1 - \theta_1 \\ \psi_2 &= \varphi_1\psi_1 + \varphi_2 - \theta_2 \\ &\vdots \\ \psi_j &= \varphi_1\psi_{j-1} + \cdots + \varphi_{p+d}\psi_{j-p-d} - \theta_j\end{aligned} \tag{5.6.2}$$

where $\theta_j = 0, j > q$.

2. For each desired level of probability ϵ, and for each lead time l, substituting in

$$\tilde{z}_{t+l}(\pm) = \hat{\tilde{z}}_t(l) \pm u_{\epsilon/2}\left(1 + \sum_{j=1}^{l-1} \psi_j^2\right)^{1/2} \sigma_a \tag{5.6.3}$$

where in practice σ_a is replaced by an estimate s_a, of the standard deviation of the white noise process a_t, and $u_{\epsilon/2}$ is the deviate exceeded by a proportion $\epsilon/2$ of the unit normal distribution.

Updating the Forecasts. When a new deviation $\tilde{z}_{t+1}$ comes to hand, the forecasts may be updated to origin $t+1$, by calculating the new forecast error $a_{t+1} = \tilde{z}_{t+1} - \tilde{z}_t(1)$ and using the difference equation (5.6.1) with $t+1$ replacing t. However, an *alternative* method is to use the forecasts $\hat{\tilde{z}}_t(1), \hat{\tilde{z}}_t(2), \ldots, \hat{\tilde{z}}(L)$ at origin t, to obtain the first $L-1$ forecasts $\hat{\tilde{z}}_{t+1}(1), \hat{\tilde{z}}_{t+1}(2), \ldots, \hat{\tilde{z}}_{t+1}(L-1)$ at origin $t+1$, from

$$\hat{\tilde{z}}_{t+1}(l) = \hat{\tilde{z}}_t(l+1) + \psi_l a_{t+1} \tag{5.6.4}$$

and then generate the last forecast $\hat{\tilde{z}}_{t+1}(L)$ using the difference equation (5.6.1).

Other Ways of Expressing the Forecasts. The above is all that is needed for *practical* utilization of the forecasts. However, the following alternative forms provide theoretical insight into the nature of the forecasts generated by different models:

1. *Forecasts in Integrated Form.* For $l > q - p - d$, the forecasts lie on the unique curve

$$\hat{\tilde{z}}_t(l) = b_0^{(t)} f_0(l) + b_1^{(t)} f_1(l) + \cdots + b_{p+d-1}^{(t)} f_{p+d-1}(l) \tag{5.6.5}$$

determined by the ''pivotal'' values $\hat{\tilde{z}}_t(q), \hat{\tilde{z}}_t(q-1), \ldots, \hat{\tilde{z}}_t(q-p-d+1)$, where $\hat{\tilde{z}}_t(-j) = \tilde{z}_{t-j}$ $(j = 0, 1, 2, \ldots)$. If $q > p + d$, the first $q - p - d$ forecasts do not lie on this curve. In general, the stationary autoregressive operator contributes damped exponential and damped sine wave terms to (5.6.5), and the nonstationary operator ∇^d contributes polynomial terms up to degree $d-1$.

The adaptive coefficients $b_j^{(t)}$ in (5.6.5) may be updated from origin t to $t+1$ by amounts depending on the last lead 1 forecast error a_{t+1}, according to the general formula

$$\mathbf{b}^{(t+1)} = \mathbf{L}'\mathbf{b}^{(t)} + \mathbf{g}a_{t+1} \tag{5.6.6}$$

given in Appendix A5.3. Specific examples of the updating are given in (5.4.5) and (5.4.13) for the IMA(0, 1, 1) and IMA(0, 2, 2) processes, respectively.

2. *Forecasts as a Weighted Sum of Past Observations.* It is instructive from a theoretical point of view to express the forecasts as a weighted sum of past observations. Thus, if the model is written in inverted form,

$$a_t = \pi(B)\tilde{z}_t = (1 - \pi_1 B - \pi_2 B^2 - \cdots)\tilde{z}_t$$

the lead 1 forecast is

$$\hat{\tilde{z}}_t(1) = \pi_1 \tilde{z}_t + \pi_2 \tilde{z}_{t-1} + \cdots \tag{5.6.7}$$

and the forecasts for longer lead times may be obtained from

$$\hat{\tilde{z}}_t(l) = \pi_1[\tilde{z}_{t+l-1}] + \pi_2[\tilde{z}_{t+l-2}] + \cdots \tag{5.6.8}$$

where the conditional expectations in (5.6.8) are evaluated by replacing $\tilde{z}$'s by actual values when known, and by forecasted values when unknown.

Alternatively, the forecast for any lead time may be written as a linear function of the available observations. Thus,

$$\hat{\tilde{z}}_t(l) = \sum_{j=1}^{\infty} \pi_j^{(l)} \tilde{z}_{t+l-j}$$

where the $\pi_j^{(l)}$ are functions of the π_j's.

Role of Constant Term in Forecasts. The forecasts will be impacted by the allowance of a nonzero constant term θ_0 in the ARIMA(p, d, q) model, $\varphi(B)z_t = \theta_0 + \theta(B)a_t$, where $\varphi(B) = \phi(B)\nabla^d$. Then, in (5.3.3) and (5.6.5), an additional deterministic polynomial term of degree d, $(\mu_w/d!)l^d$ with $w_t = \nabla^d z_t$ and $\mu_w = E[w_t] = \theta_0/(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$, will be present. This follows because in place of the relation $\varphi(B)\hat{z}_t(l) = \theta_0$ in (5.3.2), the forecasts now satisfy $\varphi(B)\hat{z}_t(l) = \theta_0$, $1 > q$, and the deterministic polynomial term of degree d represents a particular solution to this nonhomogeneous difference equation. Hence, in the instance of a nonzero constant term θ_0, the ARIMA model is also expressible as $\phi(B)(\nabla^d z_t - \mu_w) = \theta(B)a_t$, $\mu_w \neq 0$, and the forecast in the form (5.6.5) may be viewed as representing the forecast value of $\tilde{z}_{t+l} = z_{t+l} - f(t + l)$, where $f(t + l) = (\mu_w/d!)(t + l)^d + g(t + l)$ and $g(t)$ is any fixed deterministic polynomial in t of degree less than or equal to $d - 1$ (including the possibility $g(t) = 0$). For example, in an ARIMA model with $d = 1$ such as the ARIMA(1, 1, 1) model example of Section 5.4.6, but with $\theta_0 \neq 0$, the eventual forecast function of the form $\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)}\phi^l$ will now contain the additional deterministic linear trend term $\mu_w l$, where $\mu_w = \theta_0/(1 - \phi)$, similar to the result in the example for the ARIMA$(1, 1, 0)$ model in (5.4.17). Note that in the special case of a stationary process z_t, with $d = 0$, the additional deterministic term in (5.3.3) reduces to the mean of the process z_t, $\mu = E[z_t]$.

APPENDIX A5.1 CORRELATION BETWEEN FORECAST ERRORS

A5.1.1 Autocorrelation Function of Forecast Errors at Different Origins

Although it is true that for an optimal forecast the forecast errors for lead time 1 will be uncorrelated, this will not generally be true of forecasts at longer lead times. Consider forecasts for lead times l, made at origins t and $t - j$, respectively, where j is a positive integer. Then, if $j = l, l + 1, l + 2, \ldots$, the forecast errors will contain no common component, but for $j = 1, 2, \ldots, l - 1$, certain of the a's will be included in both forecast errors. Specifically,

$$e_t(l) = z_{t+l} - \hat{z}_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1}a_{t+1}$$
$$e_{t-j}(l) = z_{t-j+l} - \hat{z}_{t-j}(l) = a_{t-j+l} + \psi_1 a_{t-j+l-1} + \cdots + \psi_{l-1}a_{t-j+1}$$

TABLE A5.1 Autocorrelations of Forecast Errors at Lead 6 for Series C

j	0	1	2	3	4	5	6
$\rho[e_t(6), e_{t-j(6)}]$	1.00	0.81	0.61	0.41	0.23	0.08	0.00

and for $j < l$, the lag j autocovariance of the forecast errors for lead time l is

$$E[e_t(l)e_{t-j}(l)] = \sigma_a^2 \sum_{i=j}^{l-1} \psi_i \psi_{i-j} \tag{A5.1.1}$$

where $\psi_0 = 1$. The corresponding autocorrelations are

$$\rho[e_t(l), e_{t-j}(l)] = \begin{cases} \dfrac{\sum_{i=j}^{l-1} \psi_i \psi_{i-j}}{\sum_{i=0}^{l-1} \psi_i^2} & 0 \le j \le l \\ 0 & j \ge l \end{cases} \tag{A5.1.2}$$

We show in Chapter 7 that Series C of Figure 4.1 is well fitted by the $(1, 1, 0)$ model $(1 - 0.8B)\nabla z_t = a_t$. To illustrate (A5.1.2), we calculate the autocorrelation function of the forecast errors at lead time 6 for this model. It follows from Section 5.2.1 that the ψ weights $\psi_1, \psi_2, \ldots, \psi_5$ for this model are 1.80, 2.44. 2.95, 3.36, and 3.69, respectively. Thus, for example, the lag 1 autocovariance is

$$\begin{aligned} E[e_t(6)e_{t-1}(6)] &= \sigma_a^2[(1.80 \times 1.00) + (2.44 \times 1.80) + \cdots + (3.69 \times 3.36)] \\ &= 35.70\sigma_a^2 \end{aligned}$$

On dividing by $E[e_t^2(6)] = 43.86\sigma_a^2$, we obtain $\rho[e_t(6), e_{t-1}(6)] = 0.81$. The first six autocorrelations are shown in Table A5.1 and plotted in Figure A5.l(a). As expected, the autocorrelations beyond the fifth are zero.

A5.1.2 Correlation Between Forecast Errors at the Same Origin with Different Lead Times

Suppose that we make a series of forecasts for different lead times from the *same* fixed origin t. Then, the errors for these forecasts will be correlated. We have for $j = 1, 2, 3, \ldots,$

$$\begin{aligned} e_t(l) = z_{t+1} - \hat{z}_t(l) &= a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1} \\ e_t(l+j) = z_{t+l+j} - \hat{z}_t(l+j) &= a_{t+l+j} + \psi_1 a_{t+l+j-1} + \cdots + \psi_{j-1} a_{t+l+1} \\ &\quad + \psi_j a_{t+l} + \psi_{j+1} a_{t+l-1} + \cdots + \psi_{l+j-1} a_{t+1} \end{aligned}$$

so that the covariance between the t-origin forecast errors at lead times l and $l + j$ is $\sigma_a^2 \sum_{i=0}^{l-1} \psi_i \psi_{i+j}$, where $\psi_0 = 1$.

Thus, the correlation coefficient between the t-origin forecast errors at lead times l and $l + j$ is

$$\rho[e_t(l), e_t(l+j)] = \frac{\sum_{i=0}^{l-1} \psi_i \psi_{i+j}}{\left(\sum_{h=0}^{l-1} \psi_h^2 \sum_{g=0}^{l+j-1} \psi_g^2\right)^{1/2}} \tag{A5.1.3}$$

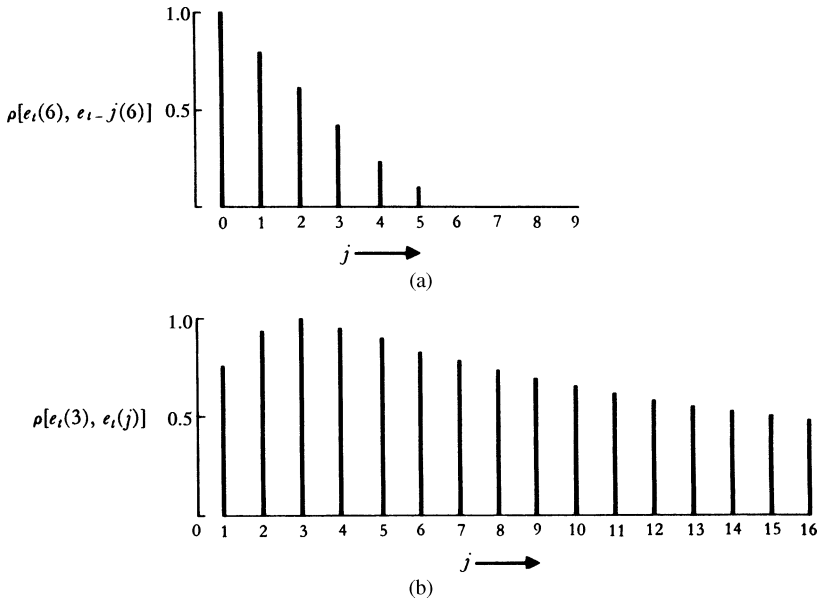

FIGURE A5.1 Correlations between various forecast errors for Series C. (a) Autocorrelations of forecast errors for Series C from different origins at lead time $l = 6$ (b) Correlations between forecast errors for Series C from the same origin at lead time 3 and lead time j.

To illustrate (A5.1.3), we compute, for forecasts made from the same origin, the correlation between the forecast error at lead time 3 and the forecast errors at lead times $j = 1, 2, 3, 4, \ldots, 16$ for Series C. For example, using (A5.1.3) and the ψ weights given in Section 5.2.2,

$$\begin{aligned} E[e_t(3)e_t(5)] &= \sigma_a^2 \sum_{i=0}^{2} \psi_i \psi_{i+2} = \sigma_a^2[\psi_0\psi_2 + \psi_1\psi_3 + \psi_2\psi_4] \\ &= \sigma_a^2[(1.00 \times 2.44) + (1.80 \times 2.95) + (2.44 \times 3.36)] \\ &= 15.94\sigma_a^2 \end{aligned}$$

The correlations for lead times $j = 1, 2, \ldots, 16$ are shown in Table A5.2 and plotted in Figure A5.1(b). As is to be expected, forecasts made from the same origin at different lead times are highly correlated.

APPENDIX A5.2 FORECAST WEIGHTS FOR ANY LEAD TIME

In this appendix we consider an alternative procedure for calculating the forecast weights $\pi_j^{(l)}$ applied to previous z's for any lead time l. To derive this result, we make use of the identity (3.1.7), namely,

$$(1 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - \pi_1 B - \pi_2 B^2 - \cdots) = 1$$

from which the π weights may be obtained in terms of the ψ weights, and vice versa.

TABLE A5.2 Correlation Between Forecast Errors at Lead 3 and at Lead j Made from a Fixed Origin for Series C

j	$\rho[e_t(3), e_t(j)]$	j	$\rho[e_t(3), e_t(j)]$
1	0.76	9	0.71
2	0.94	10	0.67
3	1.00	11	0.63
4	0.96	12	0.60
5	0.91	13	0.57
6	0.85	14	0.54
7	0.80	15	0.52
8	0.75	16	0.50

On equating coefficients, we find, for $j \geq 1$,

$$\psi_j = \sum_{i=1}^{j} \pi_i \psi_{j-i} \qquad (\psi_0 = 1) \tag{A5.2.1}$$

Thus, for example,

$$\begin{aligned} \psi_1 &= \pi_1 & \pi_1 &= \psi_1 \\ \psi_2 &= \pi_1\psi_1 + \pi_2 & \pi_2 &= \psi_2 - \psi_1\pi_1 \\ \psi_3 &= \pi_1\psi_2 + \pi_2\psi_1 + \pi_3 & \pi_3 &= \psi_3 - \psi_1\pi_2 - \psi_2\pi_1 \end{aligned}$$

Now from (5.3.6),

$$\hat{z}_t(l) = \pi_1\hat{z}_t(l-1) + \pi_2\hat{z}_t(l-2) + \cdots + \pi_{l-1}\hat{z}_t(1) + \pi_l z_t + \pi_{l+1}z_{t-1} + \cdots \tag{A5.2.2}$$

Since each of the forecasts in (A5.2.1) is itself a function of the observations $z_t, z_{t-1}, z_{t-2}, \ldots$, we can write

$$\hat{z}_t(l) = \pi_1^{(l)} z_t + \pi_2^{(l)} z_{t-1} + \pi_3^{(l)} z_{t-2} + \cdots$$

where the lead l forecast weights may be calculated from the lead 1 forecast weights $\pi_j^l = \pi_j$. We now show that the weights $\pi_j^{(l)}$ can be obtained using the identity

$$\pi_j^{(l)} = \sum_{i=1}^{l} \psi_{l-i}\pi_{i+j-1} = \pi_{j+l-1} + \psi_1\pi_{j+l-2} + \cdots + \psi_{l-1}\pi_j \tag{A5.2.3}$$

For example, the weights for the forecast at lead time 3 are

$$\begin{aligned} \pi_1^{(3)} &= \pi_3 + \psi_1\pi_2 + \psi_2\pi_1 \\ \pi_2^{(3)} &= \pi_4 + \psi_1\pi_3 + \psi_2\pi_2 \\ \pi_3^{(3)} &= \pi_5 + \psi_1\pi_4 + \psi_2\pi_3 \end{aligned}$$

and so on. To derive (A5.2.3), we write

$$\hat{z}_t(l) = \psi_l a_t + \psi_{l+1} a_{t-1} + \cdots$$
$$\hat{z}_{t+l-1}(l) = \psi_1 a_{t+l-1} + \cdots + \psi_l a_t + \psi_{l+1} a_{t-1} + \cdots$$

On subtraction, we obtain

$$\hat{z}_t(l) = \hat{z}_{t+l-1}(1) - \psi_1 a_{t+l-1} - \psi_2 a_{t+l-2} - \cdots - \psi_{l-1} a_{t+1}$$

Hence,

$$\begin{aligned}\hat{z}_t(l) =& \pi_1 z_{t+l-1} + \pi_2 z_{t+l-2} + \cdots + \pi_{l-1} z_{t+1} + \pi_l z_t + \pi_{l+1} z_{t-1} + \cdots \\ &+ \psi_1(-z_{t+l-1} + \pi_1 z_{t+l-2} + \cdots + \pi_{l-2} z_{t+1} + \pi_{l-1} z_t + \pi_l z_{t-1} + \cdots) \\ &+ \psi_2(-z_{t+l-2} + \pi_1 z_{t+l-3} + \cdots + \pi_{l-3} z_{t+1} + \pi_{l-2} z_t + \pi_{l-1} z_{t-1} + \cdots) \\ &+ \cdots \\ &+ \psi_{l-1}(-z_{t+1} + \pi_1 z_t + \pi_2 z_{t-1} + \cdots)\end{aligned}$$

Using the relation (A5.2.1), each one of the coefficients of $z_{t+l-1}, \ldots, z_{t+1}$ is seen to vanish, as they should, and on collecting terms, we obtain the required result (A5.2.3). Alternatively, we may use the formula in the recursive form

$$\pi_j^{(l)} = \pi_{j+1}^{(l-1)} + \psi_{l-1}\pi_j \tag{A5.2.4}$$

Using the model $\nabla^2 z_t = (1 - 0.9B + 0.5B^2)a_t$ for illustration, we calculate the weights for lead time 2. Equation (A5.2.4) gives

$$\pi_j^{(2)} = \pi_{j+1} + \psi_1 \pi_j$$

and using the weights in Table 5.2, with $\psi_1 = 1.1$ we have, for example,

$$\pi_1^{(2)} = \pi_2 + \psi_1 \pi_1 = 0.490 + (1.1)(1.1) = 1.700$$
$$\pi_2^{(2)} = \pi_3 + \psi_1 \pi_2 = -0.109 + (1.1)(0.49) = 0.430$$

and so on. The first 12 weights have been given in Table 5.2.

APPENDIX A5.3 FORECASTING IN TERMS OF THE GENERAL INTEGRATED FORM

A5.3.1 General Method of Obtaining the Integrated Form

We emphasize once more that for practical computation of the forecasts, the difference equation procedure is by far the simplest. The following general treatment of the integrated form is given only to elaborate further on the forecasts obtained. In this treatment, rather than solving explicitly for the forecast function as we did in the examples given in Section 5.4, it will be appropriate to write down the general form of the eventual forecast function involving $p + d$ adaptive coefficients. We then show how the eventual forecast function needs to be modified to deal with the first $q - p - d$ forecasts if $q > p + d$. Finally, we show how to update the adaptive coefficients from origin t to origin $t + 1$.

If it is understood that $\hat{z}_t(-j) = z_{t-j}$ for $j = 0, 1, 2, \ldots$, then using the conditional expectation argument of Section 5.1.1, the forecasts satisfy the difference equation:

$$\begin{aligned}
&\hat{z}_t(1) - \varphi_1 \hat{z}_t(0) - \cdots - \varphi_{p+d} \hat{z}_t(1-p-d) = -\theta_1 a_t - \cdots - \theta_q a_{t-q+1} \\
&\hat{z}_t(2) - \varphi_1 \hat{z}_t(1) - \cdots - \varphi_{p+d} \hat{z}_t(2-p-d) = -\theta_2 a_t - \cdots - \theta_q a_{t-q+2} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \vdots \qquad\qquad\qquad\qquad\qquad\qquad \text{(A5.3.1)} \\
&\hat{z}_t(q) - \varphi_1 \hat{z}_t(q-1) - \cdots - \varphi_{p+d} \hat{z}_t(q-p-d) = -\theta_q a_t \\
&\hat{z}_t(l) - \varphi_1 \hat{z}_t(l-1) - \cdots - \varphi_{p+d} \hat{z}_t(l-p-d) = 0 \qquad l > q
\end{aligned}$$

The eventual forecast function is the solution of the last equation and may be written as

$$\hat{z}_t(l) = b_0^{(t)} f_0(l) + b_1^{(t)} f_1(l) + \cdots + b_{p+d-1}^{(t)} f_{p+d-1}(l) = \sum_{i=0}^{p+d-1} b_i^{(t)} f_i(l) \qquad l > q - p - d \qquad \text{(A5.3.2)}$$

When q is less than or equal to $p + d$, the eventual forecast function will provide forecasts $\hat{z}_t(1), \hat{z}_t(2), \hat{z}_t(3), \ldots$ for all lead times $l \geq 1$.

As an example of such a model with $q \leq p + d$, suppose that

$$(1 - B)(1 - \sqrt{3}B + B^2)^2 z_t = (1 - 0.5B)a_t$$

so that $p + d = 5$ and $q = 1$. Then,

$$(1 - B)(1 - \sqrt{3}B + B^2)^2 \hat{z}_t(l) = 0 \qquad l = 2, 3, 4, \ldots$$

where B now operates on l and not on t. Solution of this difference equation yields the forecast function

$$\begin{aligned}
\hat{z}_t(l) = b_0^{(t)} + b_1^{(t)} \cos\left(\frac{2\pi l}{12}\right) + b_2^{(t)} l \cos\left(\frac{2\pi l}{12}\right) \\
+ b_3^{(t)} \sin\left(\frac{2\pi l}{12}\right) + b_4^{(t)} l \sin\left(\frac{2\pi l}{12}\right) \qquad l = 1, 2, \ldots
\end{aligned}$$

If q is greater than $p + d$, then for lead times $l \leq q - p - d$, the forecast function will have additional terms containing a_{t-i}'s. Thus,

$$\hat{z}_t(l) = \sum_{i=0}^{p+d-1} b_i^{(t)} f_i(l) + \sum_{i=0}^{j} d_{li} a_{t-i} \qquad l \leq q - p - d \qquad \text{(A5.3.3)}$$

where $j = q - p - d - l$ and the d's may be obtained explicitly by substituting (A5.3.3) in (A5.3.1). For example, consider the stochastic model

$$\nabla^2 z_t = (1 - 0.8B + 0.5B^2 - 0.4B^3 + 0.1B^4)a_t$$

in which $p + d = 2, q = 4, q - p - d = 2$ and $\varphi_1 = 2, \varphi_2 = -1, \theta_1 = 0.8, \theta_2 = -0.5, \theta_3 = 0.4$, and $\theta_4 = -0.1$. Using the recurrence relation (5.2.3), we obtain $\psi_1 = 1.2, \psi_2 = 1.9$,

$\psi_3 = 2.2$, and $\psi_4 = 2.6$. Now, from (A5.3.3),

$$\begin{aligned}\hat{z}_t(1) &= b_0^{(t)} + b_1^{(t)} + d_{10}a_t + d_{11}a_{t-1}\\ \hat{z}_t(2) &= b_0^{(t)} + 2b_1^{(t)} + d_{20}a_t\\ \hat{z}_t(l) &= b_0^{(t)} + b_1^{(t)}l \qquad l > 2\end{aligned} \tag{A5.3.4}$$

Using (A5.3.1) gives

$$\hat{z}_t(4) - 2\hat{z}_t(3) + \hat{z}_t(2) = 0.1a_t$$

so that from (A5.3.4)

$$d_{20}a_t = 0.1a_t$$

and hence $d_{20} = 0.1$. Similarly, from (A5.3.1),

$$\hat{z}_t(3) - 2\hat{z}_t(2) + \hat{z}_t(l) = -0.4a_t + 0.1a_{t-1}$$

and hence using (A5.3.4),

$$-0.2a_t + d_{10}a_t + d_{11}a_{t-1} = -0.4a_t + 0.1a_{t-1}$$

yielding

$$d_{10} = -0.2 \qquad d_{11} = 0.1$$

Hence, the forecast function is

$$\begin{aligned}\hat{z}_t(1) &= b_0^{(t)} + b_1^{(t)} - 0.2a_t + 0.1a_{t-1}\\ \hat{z}_t(2) &= b_0^{(t)} + 2b_1^{(t)} + 0.1a_t\\ \hat{z}_t(l) &= b_0^{(t)} + b_1^{(t)}l \qquad l > 2\end{aligned}$$

A5.3.2 Updating the General Integrated Form

Updating formulas for the coefficients may be obtained using the identity (5.2.5) with $t + 1$ replaced by t:

$$\hat{z}_t(l) = \hat{z}_{t-1}(l + 1) + \psi_l a_t$$

Then, for $l > q - p - d$,

$$\sum_{i=0}^{p+d-1} b_i^{(t)} f_i(l) = \sum_{i=0}^{p+d-1} b_i^{(t-1)} f_i(l + 1) + \psi_l a_t \tag{A5.3.5}$$

By solving $p + d$ such equations for different values of l, we obtain the required updating formula for the individual coefficients, in the form

$$b_i^{(t)} = \sum_{j=0}^{p+d-1} L_{ij} b_j^{(t-1)} + g_i a_t$$

Note that the updating of each of the coefficients of the forecast function depends only on the lead 1 forecast error $a_t = z_t - \hat{z}_{t-1}(1)$.

A5.3.3 Comparison with the Discounted Least-Squares Method

Although to work with the integrated form is an unnecessarily complicated way of computing forecasts, it allows us to compare the present mean square error forecast with another type of forecast that has received considerable attention. Let us write

$$\mathbf{F}_l = \begin{bmatrix} f_0(l) & f_1(l) & \cdots & f_{p+d-1}(l) \\ f_0(l+1) & f_1(l+1) & \cdots & f_{p+d-1}(l+1) \\ \vdots & \vdots & \cdots & \vdots \\ f_0(l+p+d-1) & f_1(l+p+d-1) & \cdots & f_{p+d-1}(l+p+d-1) \end{bmatrix}$$

$$\mathbf{b}^{(t)} = \begin{bmatrix} b_0^{(t)} \\ b_1^{(t)} \\ \vdots \\ b_{p+d-1}^{(t)} \end{bmatrix} \qquad \boldsymbol{\psi}_l = \begin{bmatrix} \psi_l \\ \psi_{l+1} \\ \vdots \\ \psi_{l+p+d-1} \end{bmatrix}$$

Then, using (A5.3.5) for $l, l+1, \ldots, l+p+d+-1$, we obtain for $l > q-p-d$,

$$\mathbf{F}_l \mathbf{b}^{(t)} = \mathbf{F}_{l+1} \mathbf{b}^{t-1} + \boldsymbol{\psi}_l a_t$$

yielding

$$\mathbf{b}^{(t)} = (\mathbf{F}_l^{-1} \mathbf{F}_{l+1}) \mathbf{b}^{(t-1)} + (\mathbf{F}_l^{-1} \boldsymbol{\psi}_l) a_t$$

or

$$\mathbf{b}^{(t)} = \mathbf{L}' \mathbf{b}^{(t-1)} + \mathbf{g} a_t \tag{A5.3.6}$$

Equation (A5.3.6) is of the same algebraic *form* as the updating function given by the ''discounted least-squares'' procedure of Brown (1962) and Brown and Meyer (1961). For comparison, if we denote the forecast error given by that method by e_t, then Brown's updating formula may be written as

$$\boldsymbol{\beta}^{(t)} = \mathbf{L}' \boldsymbol{\beta}^{(t-1)} + \mathbf{h} e_t \tag{A5.3.7}$$

where $\boldsymbol{\beta}^{(t)}$ is his vector of adaptive coefficients. The same matrix $\mathbf{L}$ appears in (A5.3.6) and (A5.3.7). This is inevitable, for this first factor merely allows for changes in the coefficients arising from translation to the new origin and would have to occur in any such formula. For example, consider the straight line forecast function:

$$\hat{z}_{t-1}(l) = b_0^{(t-1)} + b_l^{(t-1)} l$$

where $b_0^{(t-1)}$ is the ordinate at time $t-1$, the origin of the forecast. This can equally well be written as

$$\hat{z}_{t-1}(l) = (b_0^{(t-1)} + b_1^{(t-1)}) + b_1^{(t-1)}(l-1)$$

where now $(b_0^{(t-1)} + b_1^{(t-1)})$ is the ordinate at time t. Obviously, if we update the forecast to origin t, the coefficient b_0 must be suitably adjusted even if the forecast function were to remain unchanged.

In general, the matrix **L** does not change the forecast function, it merely relocates it. The actual updating is done by the vector of coefficients **g** and **h**. We will see that the coefficients **g**, which yield the minimum mean square error forecasts, and the coefficients **h** given by Brown are in general completely different.

Brown's Method of Forecasting.

1. A forecast function is selected from the general class of linear combinations and products of polynomials, exponentials, and sines and cosines.
2. The selected forecast function is fitted to past values by a ''discounted least-squares'' procedure. In this procedure, the coefficients are estimated and updated so that the sum of squares of weighted discrepancies

$$s_w = \sum_{j=0}^{\infty} \omega_j [z_{t-j} - \hat{\hat{z}}_t(-j)]^2 \tag{A5.3.8}$$

between past values of the series and the value given by the forecast function at the corresponding past time are minimized. The weight function ω_j is chosen arbitrarily to fall off geometrically, so that $\omega_j = (1-\alpha)^j$, where the constant α, usually called the *smoothing constant*, is (again arbitrarily) set equal to a value in the range 0.1–0.3.

Difference between the Minimum Mean Square Error Forecasts and those of Brown. To illustrate these comments, consider the forecasting of IBM stock prices, discussed by Brown (1962, p. 141). In this study, he used a quadratic model that would be, in the present notation,

$$\hat{\hat{z}}_t(l) = \beta_0^{(t)} + \beta_1^{(t)} l + \frac{1}{2}\beta_2^{(t)} l^2$$

With this model, he employed his method of discounted least squares to forecast stock prices 3 days ahead. The results obtained from this method are shown for a section of the IBM series in Figure A5.2, where they are compared with the minimum mean square error forecasts.

The discounted least-squares method can be criticized on the following grounds:

1. The nature of the forecast function ought to be decided by the autoregressive operator $\varphi(B)$ in the stochastic model, and not arbitrarily. In particular, it cannot be safely chosen by visual inspection of the time series itself. For example, consider the IBM stock prices plotted in Figure A5.2. It will be seen that a quadratic function might well be used to *fit* short pieces of this series to values already available. If such fitting were relevant to forecasting, we might conclude, as did Brown, that a polynomial forecast function of degree 2 was indicated. The most general linear process for which a quadratic function would produce *minimum mean square error* forecasts at every lead time $l = 1, 2, \ldots$ is defined by the (0, 3, 3) model

$$\nabla^3 z_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3) a_t$$