Arima Model - 5

강의 : 김성범 교수님

정리 : 김선중

December 13, 2022

이것은 ARIMA 모델 개요 - Part 5 강의에 대한 노트이다. 이전과는 다르게 영어로 작성해보았다.

Contents

아리마 모델 (ARIMA Model) – Part 5

ARMA

Backward shift operator ← function

$B \cdot X_t = X_{t-1}$

$B^2 \cdot X_t = B \cdot (B \cdot X_t) = B \cdot X_{t-1} = X_{t-2}$

$B^3 \cdot X_t = X_{t-3} \quad \cdots \quad B^m \cdot X_t = X_{t-m}$

AR(1) $\quad X_t = \phi X_{t-1} + a_t$

$= \phi B \cdot X_t + a_t$

$(1-\phi B) X_t = a_t$

$$X_t = \frac{a_t}{1-\phi B}$$

$= (1 + \phi B + \phi^2 B^2 + \phi^3 B^3 + \cdots) a_t$

$= a_t + \phi B a_t + \phi^2 B^2 a_t + \phi^3 B^3 a_t + \cdots$

$= a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \phi^3 a_{t-3} + \cdots$

AR(2) $\quad X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t$

$= \phi_1 B X_t + \phi_2 B^2 X_t + a_t$

$(1 - \phi_1 B - \phi_2 B^2) X_t = a_t$

$$X_t = \frac{1}{1-\phi_1 B - \phi_2 B^2} \cdot a_t$$

$$= \frac{1}{(1-\alpha_1 B)(1-\alpha_2 B)} \cdot a_t$$

$$= \frac{1}{1-\alpha_1 B} \cdot \frac{1}{1-\alpha_2 B} \cdot a_t$$

$= (1 + \alpha_1 B + \alpha_1^2 B^2 + \cdots)(1 + \alpha_2 B + \alpha_2^2 B^2 + \cdots) a_t$

$= a_t + (\alpha_1 + \alpha_2) B a_t + (\alpha_1^2 + \alpha_2^2 + \alpha_1\alpha_2) B^2 a_t + \cdots$

$= a_t + (\alpha_1 + \alpha_2) a_{t-1} + (\alpha_1^2 + \alpha_2^2 + \alpha_1\alpha_2) a_{t-2} + \cdots$

* Forward shift operator

$F \cdot X_t = X_{t+1}$

$F^2 \cdot X_t = X_{t+2} \quad \cdots \quad F^K \cdot X_t = X_{t+K}$

ARMA(1,1)

$X_t = \phi X_{t-1} + a_t + \theta a_{t-1} \rightarrow X_t - \phi B X_t = a_t + \theta a_{t-1}$

$(1-\phi B) X_t = a_t + \theta a_{t-1}$

$X_t = (1-\phi B)^{-1}(a_t + \theta a_{t-1})$

$= (1 + \phi B + \phi^2 B^2 + \cdots)(a_t + \theta a_{t-1})$

$= a_t + \theta a_{t-1} + \phi a_{t-1} + \phi a_{t-2} + \cdots$

$= a_t + (\phi + \theta) a_{t-1} + (\phi^2 + \phi\theta) a_{t-2} + \cdots$

$\boxed{= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots}$ (psi)

$X_t \cdot X_{t-h} = \phi X_{t-1} \cdot X_{t-h} + a_t \cdot X_{t-h} + \theta \cdot a_{t-1} \cdot X_{t-h}$

$E(X_t \cdot X_{t-h}) = \phi E(X_{t-1} \cdot X_{t-h}) + E(a_t \cdot X_{t-h}) + \theta E(a_{t-1} \cdot X_{t-h})$

$Cov(X_t, X_{t-h}) = \phi\, Cov(X_{t-1}, X_{t-h}) + E(a_t \cdot X_{t-h}) + \theta E(a_{t-1} X_{t-h})$

$h=0, \quad \gamma_X(0) = \phi \gamma_X(1) + \underbrace{E(a_t \cdot X_t)}_{①} + \theta \underbrace{E(a_{t-1}, X_t)}_{②}$

① $E(a_t \cdot X_t)$

$= E(a_t \cdot (a_t + \psi_1 a_{t-1} + \psi a_{t-2} + \cdots))$

$= E(a_t a_t) + \psi_1 E(a_t \cdot a_{t-1})^{\to 0} + \cdots$

$= Cov(a_t, a_t)$

$= \sigma_a^2$

② $E(a_{t-1}(a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots))$

$= E(a_{t-1} \cdot a_t)^{\to 0} + \boxed{\psi_1 E(a_{t-1} \cdot a_{t-1})} + \psi_2 E(a_{t-1} a_{t-2})^{\to 0} + \cdots)$

$= \psi_1 \sigma_a^2$

$$\boxed{\therefore \gamma_X(0) = \sigma_X^2 = \phi \gamma_X(1) + \sigma_a^2 + \theta \underbrace{\psi_1}_{=(\phi+\theta)} \sigma_a^2}$$

Autovariance function of ARMA(1,1) $h=0$

$h=1, \quad Cov(X_t, X_{t-1}) = \phi\, Cov(X_{t-1}, X_{t-1}) + \underbrace{E(a_t \cdot X_{t-1})}_{=0} + \theta E(a_{t-1} \cdot X_{t-1})$

$\gamma_X(1) = \phi \gamma_X(0) + \theta \cdot \sigma_a^2 \rightarrow ①$

$\boxed{\gamma_X(0) = \phi \gamma_X(1) + \sigma_a^2 + \theta(\phi + \theta)\sigma_a^2} \rightarrow ②$

①, ②

$$\boxed{\gamma_X(0) = \frac{1 + 2\theta\phi + \theta^2}{1-\phi^2} \cdot \sigma_a^2} \longrightarrow h=0$$

$$\boxed{\gamma_X(1) = \frac{(1+\theta\phi)(\phi+\theta)}{1-\phi^2} \sigma_a^2} \longrightarrow h=1$$

$h \geq 2. \quad Cov(X_t, X_{t-h}) = \phi\, Cov(X_{t-1}, X_{t-h}) + E(a_t \cdot X_{t-h})^{\to 0} + \theta E(a_{t-1} \cdot X_{t-h})^{\to 0}$

$\gamma_X(h) = \phi \gamma_X(h-1)$

$\gamma_X(h) = \phi^{h-1} \gamma_X(1)$ leave

$$\boxed{= \phi^{h-1} \cdot \frac{(1+\theta\phi)(\phi+\theta)}{1-\phi^2} \cdot \sigma_a^2}$$

AR process $\quad$ AR(p) $\qquad$ (WN) $\rightarrow \boxed{\phi^{-1}(B)} \rightarrow$ AR(p)

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t$

$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = a_t$

$X_t - \phi_1 B X_t - \phi_2 B^2 X_t - \cdots - \phi_p B^p X_t = a_t$

$\underbrace{(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)}_{\phi(B)} X_t = a_t$

$\phi(B) X_t = a_t$

$\boxed{X_t = \phi^{-1}(B) \cdot a_t}$ WN

⇒ AR process can be thought of as the output X_t from a linear filter with transfer function $\phi^{-1}(B)$ when the input is white noise a_t.

MA(q)

$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$

$= a_t - \theta_1 B \cdot a_t - \theta_2 B^2 a_t - \cdots - \theta_q B^q \cdot a_t$

$= (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) a_t$

$= \theta(B) \cdot a_t.$ $\qquad$ $a_t \rightarrow \boxed{\theta(B)} \rightarrow$ MA(q)

ARMA(p,q)

$X_t = \underbrace{\phi_1 X_{t-1} + \cdots + \phi_p X_{t-p}}_{AR} + \underbrace{a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}}_{MA}$

$\phi(B) X_t = \theta(B) \cdot a_t$

$X_t = \boxed{\phi^{-1}(B) \cdot \theta(B)}\, a_t$

$a_t \rightarrow \boxed{\quad} \rightarrow X_t \rightarrow$ ARMA(p,q)

1 Backward shift operator B

An operator B is called the *backward shift operator*. It is a function that maps a random variable which depends on time (i.e. X_t or a_t), backward by one step.

$$B(X_t) = X_{t-1} \tag{1}$$

We occasionally write BX_t instead of $B(X_t)$ to stand for the same thing. Denoting $B^m = B \circ B \circ \cdots \circ B$ by the m times composition of B, we have

$$\begin{aligned} BX_t &= X_{t-1} \\ B^2 X_t &= X_{t-2} \\ &\vdots \\ B^m X_t &= X_{t-m}. \end{aligned}$$

The professor tries to explain AR(1) model and AR(2) model in terms of this operator B, which is illustrated as follows.

1.1 AR(1) model described by B

Recall that AR(1) is called the *autoregressive model* of order 1, which predict X_t by means of X_{t-1} and a_t.

$$X_t = \phi X_{t-1} + a_t. \tag{2}$$

Here, a_t is a white noise so that the expectation of a_t is 0 ;

$$\mathbb{E}[a_t] = 0, \tag{3}$$

the variance of a_t is independent of t ;

$$\mathbb{V}[a_t] = {\sigma_a}^2, \tag{4}$$

and a_t's in distinct timsteps are independent one another ;

$$\mathrm{Cov}(a_t, a_{t+h}) = 0 \quad \text{for} \quad h \neq 0. \tag{5}$$

Recall that $\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ for any random variable X. By (3) and (4),

$$\mathbb{E}[{a_t}^2] = \mathbb{V}[a_t] + (\mathbb{E}[a_t])^2 = {\sigma_a}^2 + 0^2 = {\sigma_a}^2. \tag{6}$$

Moreover, by (3) and (5),

$$\mathbb{E}[a_t a_s] = \mathbb{E}[(a_t - \mathbb{E}[a_t])(a_s - \mathbb{E}[a_s])] = \mathrm{Cov}(a_t, a_s) = 0 \tag{7}$$

if $t \neq s$. Summarizing the above two equations ((6) and (7)) yields

$$\mathbb{E}[a_t a_s] = \begin{cases} {\sigma_a}^2 & (t = s) \\ 0 & (t \neq s) \end{cases} \tag{8}$$

We can rewrite the equation (2) as

$$X_t = \phi B X_t + a_t$$
$$(1 - \phi B) X_t = a_t$$

where $1 - \phi B$ can be thought of as an oprator, which maps X_t to $X_t - \phi B X_t$. Here, 1 should be viewed as a identity operator. But, I think the professor regard $1 - \phi B$ as a '1 minus a real number ϕB. Moreover, he assumes that $|\phi B| < 1$ in that

$$\begin{aligned} X_t &= \frac{1}{1 - \phi B} a_t \\ &= (1 + \phi B + \phi^2 B^2 + \cdots) a_t \\ &= a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \cdots . \end{aligned} \tag{9}$$

where he use the formula for geometric series.[1] As a result, X_t is expressed as a (infinite) linear combination of a_{t-h} for $h = 0, 1, 2, \cdots$.

1.2 AR(2) model described by B

Recall that AR(2) is an autoregressive model of order 2, which predict X_t by means of X_{t-1}, X_{t-2} and a_t.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t \tag{10}$$

Using the backward shift operator B, we get

$$X_t = \phi_1 B X_t + \phi_2 B^2 X_t + a_t$$
$$(1 - \phi_1 B - \phi_2 B^2) X_t = a_t$$

Again, regarding B as a real number, $1 - \phi_1 B - \phi_2 B^2$ can be thought of as a quadratic polynomial of B. So there exists numbers[2] α_1 and α_2 such that ($\alpha_1 + \alpha_2 = \phi_1$, $\alpha_1 \alpha_2 = -\phi_2$)

$$1 - \phi_1 B - \phi_2 B^2 = (1 - \alpha_1 B)(1 - \alpha_2 B).$$

Thus,

$$\begin{aligned} X_t &= \frac{1}{1 - \phi_1 B - \phi_2 B^2} a_t \\ &= \frac{1}{(1 - \alpha_1 B)(1 - \alpha_2 B)} a_t \\ &= (1 + \alpha_1 B + {\alpha_1}^2 B^2 + \cdots)(1 + \alpha_2 B + {\alpha_2}^2 B^2 + \cdots) a_t \\ &= a_t + (\alpha_1 + \alpha_2) a_{t-1} + ({\alpha_1}^2 + \alpha_1 \alpha_2 + {\alpha_2}^2) a_{t-2} + ({\alpha_1}^3 + {\alpha_1}^2 \alpha_2 + \alpha_1 {\alpha_2}^2 + {\alpha_2}^3) a_{t-3} + \cdots . \end{aligned}$$

[1] An alternative way to equate $(1 - \phi B)^{-1}$ and $1 + \phi B + \phi^2 B^2 + \cdots$ is illustrated in Shumway's book. In this book, he considers the inverse operator $\phi^{-1}(B)$ of $\phi(B) = 1 - \phi B$. See equations (3.6), (3.12), (3.13), (3.14).

[2] real numbers if $D = b^2 - 4ac = (-\phi_1)^2 - 4 \cdot (-\phi_2) \cdot 1 = {\phi_1}^2 + 4\phi_2 > 0$. but in general, they are complex numbers

1.3 Forward shift operator F

In contrast to the backward shift operator B, the *forward shift operator* F shift a random variable forward.

$$
\begin{aligned}
FX_t &= X_{t+1} \\
F^2 X_t &= X_{t+2} \\
&\vdots \\
F^k X_t &= X_{t+k}
\end{aligned} \tag{11}
$$

for any positive integer k.

2 ARMA(1,1) model

ARMA(1,1) model is a combination of AR(1) model and MA(1) model. We are assuming that X_t is obtained from X_{t-1}, a_t and a_{t-1}.

$$
X_t = \phi X_{t-1} + a_t + \theta a_{t-1}. \tag{12}
$$

Here, we compute the covariance $\gamma(h) = \mathrm{Cov}(X_t, X_{t-h})$ of this model for $h = 0, 1, 2, \cdots$. To this aim, multiply (12) by X_{t-h} on both sides

$$
X_t X_{t-h} = \phi X_{t-1} X_{t-h} + a_t X_{t-h} + \theta a_{t-1} X_{t-h}
$$

and take the expectation on both sides to get[3];

$$
\begin{aligned}
\gamma(h) &= \mathbb{E}[X_t X_{t-h}] \\
&= \phi \mathbb{E}[X_{t-1} X_{t-h}] + \mathbb{E}[a_t X_{t-h}] + \theta \mathbb{E}[a_{t-1} X_{t-h}] \\
&= \phi \mathrm{Cov}(X_{t-1}, X_{t-h}) + \mathbb{E}[a_t X_{t-h}] + \theta \mathbb{E}[a_{t-1} X_{t-h}]
\end{aligned} \tag{13}
$$

2.1 autocovariance for $h = 0$: $\gamma(0)$

Suppose $h = 0$. Before computation, we can convert the equation (12) using the operator B ;

$$
\begin{aligned}
(1 - \phi B) X_t &= a_t + \theta a_{t-1} \\
X_t &= \frac{1}{1 - \phi B}(a_t + \theta a_{t-1}) \\
&= (1 + \phi B + \phi^2 B^2 + \cdots)(a_t + \theta a_{t-1}) \\
&= a_t + \theta a_{t-1} + \phi a_{t-1} + \phi\theta a_{t-2} + \phi^2 a_{t-2} + \phi^2 \theta a_{t-3} + \cdots \\
&= a_t + (\theta + \phi) a_{t-1} + \phi(\theta + \phi) a_{t-2} + \cdots \\
&= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots
\end{aligned} \tag{14}
$$

[3] 여기에서 $\mathbb{E}[X_t] = \mathbb{E}[X_{t-h}] = 0$임을 가정하고 있는 것일까? Covariance의 정의는 $\mathrm{Cov}(X, Y) = \mathbb{E}[(X - [X])(Y - [Y])]$인데 마치 $\mathrm{Cov}(X, Y) = \mathbb{E}[XY]$인 것처럼 쓰고 있다. Shumway의 책에서는 x_t의 평균이 0임을 가정하고 있다. x_t의 평균이 0이 아닐 경우 평균 μ를 뺀 $x_t - \mu$를 고려하라고 되어 있다. See p. 86, Definition 3.1.

By (14), the second term and the third term of (13) are evalauted as

$$
\begin{aligned}
\mathbb{E}[a_t X_t] &= \mathbb{E}[a_t(a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots)] \\
&= \mathbb{E}[{a_t}^2] + \psi_1 \mathbb{E}[a_t a_{t-1}] + \psi_2 \mathbb{E}[a_t a_{t-2}] + \cdots \\
&= {\sigma_a}^2 + 0 + 0 + \cdots = {\sigma_a}^2,
\end{aligned}
$$

and

$$
\begin{aligned}
\theta \mathbb{E}[a_{t-1} X_t] &= \theta \mathbb{E}[a_{t-1}(a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots)] \\
&= \theta \left(\mathbb{E}[a_{t-1} a_t] + \psi_1 \mathbb{E}[{a_{t-1}}^2] + \psi_2 \mathbb{E}[a_{t-1} a_{t-2}] + \cdots \right) \\
&= \theta \left(0 + \psi_1 {\sigma_a}^2 + 0 + \cdots \right) \\
&= \theta \psi_1 {\sigma_a}^2,
\end{aligned}
$$

where we used (8) in the computations. Therefore, the autocovariance function of ARMA(1,1) when $h = 0$ is

$$
\begin{aligned}
\gamma(0) &= \phi \mathrm{Cov}(X_{t-1}, X_t) + {\sigma_a}^2 + \theta \psi_1 {\sigma_a}^2 \\
&= \phi \gamma(1) + (1 + \theta^2 + \theta \phi) {\sigma_a}^2,
\end{aligned}
\tag{15}
$$

which is also the variance $\mathbb{V}(X_t)$ of the ARMA(1,1) model.

2.2 autocovariance for $h = 1$: $\gamma(1)$

Suppose $h = 1$. The equation (12) applied to $t - 1$ yields

$$
\begin{aligned}
X_{t-1} &= \phi X_{t-2} + a_{t-1} + \theta a_{t-2} \\
X_{t-1} &= \phi B X_{t-1} + a_{t-1} + \theta a_{t-2} \\
(1 - \phi B) X_{t-1} &= a_{t-1} + \theta a_{t-2} \\
X_{t-1} &= (1 + \phi B + \phi^2 B^2 + \cdots)(a_{t-1} + \theta a_{t-2}) \\
&= a_{t-1} + \psi_1 a_{t-2} + \psi_2 a_{t-3} + \cdots
\end{aligned}
$$

for the same ψ_{t-h} $(h = 1, 2, \cdots)$. Plugging $h = 1$ and the above equation into (13),

$$
\begin{aligned}
\gamma(1) &= \phi \mathrm{Cov}(X_{t-1}, X_{t-1}) + \mathbb{E}[a_t X_{t-1}] + \theta \mathbb{E}[a_{t-1} X_{t-1}] \\
&= \phi \gamma(0) + \mathbb{E}[a_t(a_{t-1} + \psi_1 a_{t-2} + \psi_2 a_{t-3} + \cdots)] + \theta \mathbb{E}[a_{t-1}(a_{t-1} + \psi_1 a_{t-2} + \psi_2 a_{t-3} + \cdots)] \\
&= \phi \gamma(0) + 0 + \theta \times {\sigma_a}^2 \\
&= \phi \gamma(0) + \theta {\sigma_a}^2
\end{aligned}
\tag{16}
$$

by repeated uses of (8).

As a result, we get a system of linear equations (15) and (16) with unknowns $\gamma(0)$ and $\gamma(1)$. Solving the system by substitution yields

$$
\begin{aligned}
\gamma(0) &= \phi \gamma(1) + (1 + \theta^2 + \theta \phi) {\sigma_a}^2 \\
&= \phi^2 \gamma(0) + (1 + 2\theta\phi + \theta^2) {\sigma_a}^2
\end{aligned}
$$

$$\gamma(0) = \frac{1 + 2\theta\phi + \theta^2}{1 - \phi^2}{\sigma_a}^2 \tag{17}$$

and

$$\begin{aligned}\gamma(1) &= \phi\gamma(0) + \theta{\sigma_a}^2 \\ &= \phi^2\gamma(1) + (\phi + \phi\theta^2 + \phi^2\theta + \theta){\sigma_a}^2\end{aligned}$$

$$\gamma(1) = \frac{\phi + \theta\phi^2 + \theta^2\phi + \theta}{1 - \phi^2}{\sigma_a}^2 = \frac{(\phi + \theta)(1 + \phi\theta)}{1 - \phi^2}{\sigma_a}^2 \tag{18}$$

2.3 autocovariance for $h \geq 2$: $\gamma(h)$

By the similar reasoning as in 2.1 and 2.2, we have

$$\begin{aligned}X_{t-h} &= (1 + \phi B + \phi^2 B^2 + \cdots)(a_{t-h} + \theta a_{t-h-1}) \\ &= a_{t-h} + \psi_1 a_{t-h-1} + \psi_2 a_{t-h-2} + \cdots\end{aligned}$$

Substituting the above equation to (13) yields

$$\begin{aligned}\gamma(h) &= \phi\mathrm{Cov}(X_{t-1}, X_{t-h}) + \mathbb{E}[a_t X_{t-h}] + \theta\mathbb{E}[a_{t-1}X_{t-h}] \\ &= \phi\gamma(h-1) + \mathbb{E}[a_t(a_{t-h} + \psi_1 a_{t-h-1} + \psi_2 a_{t-h-2} + \cdots)] + \theta\mathbb{E}[a_{t-1}(a_{t-h} + \psi_1 a_{t-h-1} + \psi_2 a_{t-h-2} + \cdots)] \\ &= \phi\gamma(h-1).\end{aligned} \tag{19}$$

Here, all the terms of the form $\mathbb{E}[a_t a_{t-l}]$ or $\mathbb{E}[a_{t-1}a_{t-l}]$ has vanished since $h \geq 2$. Using induction on the above reccurence formula (19), we have $\gamma(h) = \phi^{h-1}\gamma(1)$ for $h \geq 2$. Combined with (17) and (18), we get

$$\gamma(h) = \begin{cases} \frac{1+2\theta\phi+\theta^2}{1-\phi^2}{\sigma_a}^2 & (h = 0) \\ \frac{(\phi+\theta)(1+\phi\theta)\phi^{h-1}}{1-\phi^2}{\sigma_a}^2 & (h \geq 1) \end{cases} \tag{20}$$

Since $\gamma(-h) = \gamma(h)$, equation (20) is the explicit formula of $\gamma(h)$ for all integers h.

3 AR, MA and ARMA described by operators

The lecture ends by giving a summary for AR model, MA model and ARMA model. The moral is that those models can be expressed in terms of *polynomials* of B.

3.1 AR model

Consider the AR(p) model, characterized by the equation

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t \tag{21}$$

The above equation can be converted into

$$(1 - \phi B - \phi^2 B^2 - \cdots - \phi^p B^p)X_t = a_t \tag{22}$$

We denote

$$\phi(B) = 1 - \phi B - \phi^2 B^2 - \cdots - \phi^p B^p, \tag{23}$$

so that $\phi(B)$ is an operator (a polynomial of operator B) acting on a random variable. The equation (22) can be expressed as

$$\phi(B) X_t = a_t \tag{24}$$

If the inverse operator $\phi^{-1}(B)$ of $\phi(B)$ exists[4], the above equation goes

$$X_t = \phi^{-1}(B) a_t.$$

Here is a comment by the professor, summarizing the above equation :

> AR process can be thought of as making the output X_t from a linear filter with transfer function $\phi^{-1}(B)$ when the input is a white noise a_t.

3.2 MA model

The MA(q) model is governed by the equation[5]

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}. \tag{25}$$

We have

$$X_t = (1 - \theta B - \theta^2 B^2 - \cdots - \theta^q B^q) a_t \tag{26}$$

$$\theta(B) = 1 - \theta B - \theta^2 B^2 - \cdots - \theta^q B^q \tag{27}$$

$$X_t = \theta(B) a_t \tag{28}$$

3.3 ARMA model

The ARMA(p, q) model has the equation

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}. \tag{29}$$

By the similar procedures that we did earlier,

$$\phi(B) X_t = \theta(B) a_t \tag{30}$$

holds for the same $\phi(B)$ and $\theta(B)$ defined in (23) and (27), repsectively. Taking the inverse $\phi^{-1}(B)$ of $\phi(B)$ on both sides yields

$$X_t = \phi^{-1}(B)\theta(B) a_t. \tag{31}$$

[4]actually, it exists if $|\phi| < 1$ whose explicit form is written in (9). See equations (3.6), (3.12), (3.13), (3.14) in Shumway's book.

[5]Note that the professor wrote minus sign instead of plus sign. The choice of signs are irrelavant of the context. Confer to the second footnote at page 90, in Shumway's book.

related pages in Shumway's book

differencing may be more appropriate. Differencing is also a viable tool if the trend is fixed, as in Example 2.3. That is, e.g., if $\mu_t = \beta_1 + \beta_2 t$ in the model (2.25), differencing the data produces stationarity (see Problem 2.6):

$$x_t - x_{t-1} = (\mu_t + y_t) - (\mu_{t-1} + y_{t-1}) = \beta_2 + y_t - y_{t-1}.$$

Because differencing plays a central role in time series analysis, it receives its own notation. The first difference is denoted as

$$\nabla x_t = x_t - x_{t-1}. \tag{2.29}$$

As we have seen, the first difference eliminates a linear trend. A second difference, that is, the difference of (2.29), can eliminate a quadratic trend, and so on. In order to define higher differences, we need a variation in notation that we use, for the first time here, and often in our discussion of ARIMA models in Chapter 3.

Definition 2.4 *We define the* **backshift operator** *by*

$$Bx_t = x_{t-1}$$

and extend it to powers $B^2 x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}$*, and so on. Thus,*

$$B^k x_t = x_{t-k}. \tag{2.30}$$

It is clear that we may then rewrite (2.29) as

$$\nabla x_t = (1 - B)x_t, \tag{2.31}$$

and we may extend the notion further. For example, the second difference becomes

$$\begin{aligned} \nabla^2 x_t &= (1 - B)^2 x_t = (1 - 2B + B^2)x_t \\ &= x_t - 2x_{t-1} + x_{t-2} \end{aligned}$$

by the linearity of the operator. To check, just take the difference of the first difference $\nabla(\nabla x_t) = \nabla(x_t - x_{t-1}) = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2})$.

Definition 2.5 Differences of order d *are defined as*

$$\nabla^d = (1 - B)^d, \tag{2.32}$$

where we may expand the operator $(1 - B)^d$ *algebraically to evaluate for higher integer values of* d*. When* $d = 1$*, we drop it from the notation.*

The first difference (2.29) is an example of a linear filter applied to eliminate a trend. Other filters, formed by averaging values near x_t, can produce adjusted series that eliminate other kinds of unwanted fluctuations, as in Chapter 3. The differencing technique is an important component of the ARIMA model of Box and Jenkins (1970) (see also Box et al., 1994), to be discussed in Chapter 3.

3.2 Autoregressive Moving Average Models

The classical regression model of Chapter 2 was developed for the static case, namely, we only allow the dependent variable to be influenced by current values of the independent variables. In the time series case, it is desirable to allow the dependent variable to be influenced by the past values of the independent variables and possibly by its own past values. If the present can be plausibly modeled in terms of only the past values of the independent inputs, we have the enticing prospect that forecasting will be possible.

Introduction to Autoregressive Models

Autoregressive models are based on the idea that the current value of the series, x_t, can be explained as a function of p past values, $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$, where p determines the number of steps into the past needed to forecast the current value. As a typical case, recall Example 1.10 in which data were generated using the model

$$x_t = x_{t-1} - .90x_{t-2} + w_t,$$

where w_t is white Gaussian noise with $\sigma_w^2 = 1$. We have now assumed the current value is a particular *linear* function of past values. The regularity that persists in Figure 1.9 gives an indication that forecasting for such a model might be a distinct possibility, say, through some version such as

$$x_{n+1}^n = x_n - .90x_{n-1},$$

where the quantity on the left-hand side denotes the forecast at the next period $n+1$ based on the observed data, $x_1, x_2, \ldots, x_n$. We will make this notion more precise in our discussion of forecasting (§3.5).

The extent to which it might be possible to forecast a real data series from its own past values can be assessed by looking at the autocorrelation function and the lagged scatterplot matrices discussed in Chapter 2. For example, the lagged scatterplot matrix for the Southern Oscillation Index (SOI), shown in Figure 2.7, gives a distinct indication that lags 1 and 2, for example, are linearly associated with the current value. The ACF shown in Figure 1.14 shows relatively large positive values at lags 1, 2, 12, 24, and 36 and large negative values at 18, 30, and 42. We note also the possible relation between the SOI and Recruitment series indicated in the scatterplot matrix shown in Figure 2.8. We will indicate in later sections on transfer function and vector AR modeling how to handle the dependence on values taken by other series.

The preceding discussion motivates the following definition.

Definition 3.1 *An* **autoregressive model of order p**, *abbreviated* **AR(p)**, *is of the form*

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t, \tag{3.1}$$

where x_t is stationary, $\phi_1, \phi_2, \ldots, \phi_p$ are constants ($\phi_p \neq 0$). Unless otherwise stated, we assume that w_t is a Gaussian white noise series with mean zero and *variance* σ_w^2. *The mean of* x_t *in (3.1) is zero. If the mean,* μ, *of* x_t *is not zero, replace* x_t *by* $x_t - \mu$ *in (3.1), i.e.,*

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \cdots + \phi_p(x_{t-p} - \mu) + w_t,$$

or write

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t, \tag{3.2}$$

where $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$.

We note that (3.2) is similar to the regression model of §2.2, and hence the term auto (or self) regression. Some technical difficulties, however, develop from applying that model because the regressors, $x_{t-1}, \ldots, x_{t-p}$, are random components, whereas z_t was assumed to be fixed. A useful form follows by using the backshift operator (2.30) to write the AR(p) model, (3.1), as

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)x_t = w_t, \tag{3.3}$$

or even more concisely as

$$\phi(B)x_t = w_t. \tag{3.4}$$

The properties of $\phi(B)$ are important in solving (3.4) for x_t. This leads to the following definition.

Definition 3.2 *The* **autoregressive operator** *is defined to be*

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \tag{3.5}$$

We initiate the investigation of AR models by considering the first-order model AR(1) given by $x_t = \phi x_{t-1} + w_t$. Iterating backwards k times, we get

$$\begin{aligned} x_t &= \phi x_{t-1} + w_t = \phi(\phi x_{t-2} + w_{t-1}) + w_t \\ &= \phi^2 x_{t-2} + \phi w_{t-1} + w_t \\ &\ \ \vdots \\ &= \phi^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j w_{t-j}. \end{aligned}$$

This method suggests that, by continuing to iterate backwards, and provided that $|\phi| < 1$ and x_t is stationary, we can represent an AR(1) model as a linear process given by[1]

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j}. \tag{3.6}$$

[1]Note that $\lim_{k\to\infty} E\left(x_t - \sum_{j=0}^{k-1} \phi^j w_{t-j}\right)^2 = \lim_{k\to\infty} \phi^{2k} E\left(x_{t-k}^2\right) = 0$, so (3.6) exists in the mean square sense (see Appendix A for a definition).

The AR(1) process defined by (3.6) is stationary with mean

$$E(x_t) = \sum_{j=0}^{\infty} \phi^j E(w_{t-j}) = 0,$$

and autocovariance function,

$$\begin{aligned}
\gamma(h) &= \operatorname{cov}(x_{t+h}, x_t) = E\left[\left(\sum_{j=0}^{\infty} \phi^j w_{t+h-j}\right)\left(\sum_{k=0}^{\infty} \phi^k w_{t-k}\right)\right] \\
&= \sigma_w^2 \sum_{j=0}^{\infty} \phi^j \phi^{j+h} = \sigma_w^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma_w^2 \phi^h}{1-\phi^2}, \quad h \geq 0. \qquad (3.7)
\end{aligned}$$

Recall that $\gamma(h) = \gamma(-h)$, so we will only exhibit the autocovariance function for $h \geq 0$. From (3.7), the ACF of an AR(1) is

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h, \quad h \geq 0, \qquad (3.8)$$

and $\rho(h)$ satisfies the recursion

$$\rho(h) = \phi\, \rho(h-1), \quad h = 1, 2, \ldots. \qquad (3.9)$$

We will discuss the ACF of a general AR(p) model in §3.4.

Example 3.1 The Sample Path of an AR(1) Process

Figure 3.1 shows a time plot of two AR(1) processes, one with $\phi = .9$ and one with $\phi = -.9$; in both cases, $\sigma_w^2 = 1$. In the first case, $\rho(h) = .9^h$, for $h \geq 0$, so observations close together in time are positively correlated with each other. This result means that observations at contiguous time points will tend to be close in value to each other; this fact shows up in the top of Figure 3.1 as a very smooth sample path for x_t. Now, contrast this to the case in which $\phi = -.9$, so that $\rho(h) = (-.9)^h$, for $h \geq 0$. This result means that observations at contiguous time points are negatively correlated but observations two time points apart are positively correlated. This fact shows up in the bottom of Figure 3.1, where, for example, if an observation, x_t, is positive, the next observation, x_{t+1}, is typically negative, and the next observation, x_{t+2}, is typically positive. Thus, in this case, the sample path is very choppy.

A figure similar to Figure 3.1 can be created in R using the following commands:

```
> par(mfrow=c(2,1))
> plot(arima.sim(list(order=c(1,0,0), ar=.9), n=100),
+  ylab="x",main=(expression("AR(1)  "*phi*" = +.9")))
> plot(arima.sim(list(order=c(1,0,0), ar=-.9), n=100),
+  ylab="x",main=(expression("AR(1)  "*phi*" = -.9")))
```

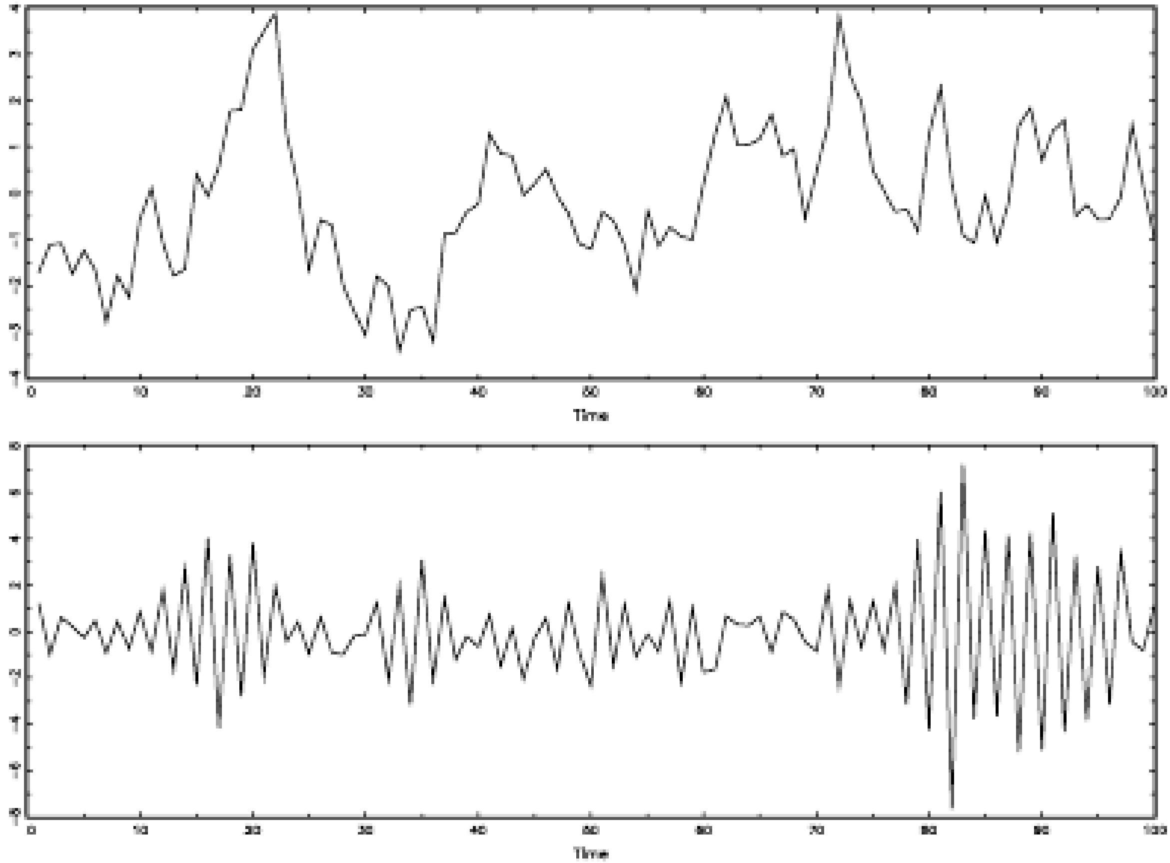

Figure 3.1 Simulated AR(1) models: $\phi = .9$ (top); $\phi = -.9$ (bottom).

Example 3.2 Explosive AR Models and Causality

In Example 1.18, it was discovered that the random walk $x_t = x_{t-1} + w_t$ is not stationary. We might wonder whether there is a stationary AR(1) process with $|\phi| > 1$. Such processes are called explosive because the values of the time series quickly become large in magnitude. Clearly, because $|\phi|^j$ increases without bound as $j \to \infty$, $\sum_{j=0}^{k-1} \phi^j w_{t-j}$ will not converge (in mean square) as $k \to \infty$, so the intuition used to get (3.6) will not work directly. We can, however, modify that argument to obtain a stationary model as follows. Write $x_{t+1} = \phi x_t + w_{t+1}$, in which case,

$$\begin{aligned} x_t &= \phi^{-1} x_{t+1} - \phi^{-1} w_{t+1} = \phi^{-1}\left(\phi^{-1} x_{t+2} - \phi^{-1} w_{t+2}\right) - \phi^{-1} w_{t+1} \\ &\vdots \\ &= \phi^{-k} x_{t+k} - \sum_{j=1}^{k-1} \phi^{-j} w_{t+j}, \end{aligned} \tag{3.10}$$

by iterating forward k steps. Because $|\phi|^{-1} < 1$, this result suggests the stationary future dependent AR(1) model

$$x_t = -\sum_{j=1}^{\infty} \phi^{-j} w_{t+j}.$$

The reader can verify that this is stationary and of the AR(1) form $x_t = \phi x_{t-1} + w_t$. Unfortunately, this model is useless because it requires us to know the future to be able to predict the future. When a process does not depend on the future, such as the AR(1) when $|\phi| < 1$, we will say the process is causal. In the explosive case of this example, the process is stationary, but it is also future dependent, and not causal.

The technique of iterating backwards to get an idea of the stationary solution of AR models works well when $p = 1$, but not for larger orders. A general technique is that of matching coefficients. Consider the AR(1) model in operator form

$$\boxed{\phi(B)x_t = w_t,} \tag{3.11}$$

where $\phi(B) = 1 - \phi B$, and $|\phi| < 1$. Also, write the model in equation (3.6) using operator form as

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B)w_t, \tag{3.12}$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ and $\psi_j = \phi^j$. Suppose we did not know that $\psi_j = \phi^j$. We could substitute $\psi(B)w_t$ from (3.12) for x_t in (3.11) to obtain

$$\boxed{\phi(B)\psi(B)w_t = w_t.} \tag{3.13}$$

The coefficients of B on the left-hand side of (3.13) must be equal to those on right-hand side of (3.13), which means

$$\boxed{(1 - \phi B)(1 + \psi_1 B + \psi_2 B^2 + \cdots + \psi_j B^j + \cdots) = 1.} \tag{3.14}$$

Reorganizing the coefficients in (3.14),

$$1 + (\psi_1 - \phi)B + (\psi_2 - \psi_1\phi)B^2 + \cdots + (\psi_j - \psi_{j-1}\phi)B^j + \cdots = 1,$$

we see that for each $j = 1, 2, \ldots$, the coefficient of B^j on the left must be zero because it is zero on the right. The coefficient of B on the left is $(\psi_1 - \phi)$, and equating this to zero, $\psi_1 - \phi = 0$, leads to $\psi_1 = \phi$. Continuing, the coefficient of B^2 is $(\psi_2 - \psi_1\phi)$, so $\psi_2 = \phi^2$. In general,

$$\psi_j = \psi_{j-1}\phi,$$

with $\psi_0 = 1$, which leads to the general solution $\psi_j = \phi^j$.

Another way to think about the operations we just performed is to consider the AR(1) model in operator form, $\phi(B)x_t = w_t$. Now multiply both sides by $\phi^{-1}(B)$ (assuming the inverse operator exists) to get

$$\phi^{-1}(B)\phi(B)x_t = \phi^{-1}(B)w_t,$$

or

$$x_t = \phi^{-1}(B)w_t.$$

We know already that

$$\boxed{\phi^{-1}(B) = 1 + \phi B + \phi^2 B^2 + \cdots + \phi^j B^j + \cdots,}$$

that is, $\phi^{-1}(B)$ is $\psi(B)$ in (3.12). Thus, we notice that working with operators is like working with polynomials. That is, consider the polynomial $\phi(z) = 1 - \phi z$, where z is a complex number and $|\phi| < 1$. Then,

$$\phi^{-1}(z) = \frac{1}{(1-\phi z)} = 1 + \phi z + \phi^2 z^2 + \cdots + \phi^j z^j + \cdots, \quad |z| \leq 1,$$

and the coefficients of B^j in $\phi^{-1}(B)$ are the same as the coefficients of z^j in $\phi^{-1}(z)$. In other words, we may treat the backshift operator, B, as a complex number, z. These results will be generalized in our discussion of ARMA models. We will find the polynomials corresponding to the operators useful in exploring the general properties of ARMA models.

Introduction to Moving Average Models

As an alternative to the autoregressive representation in which the x_t on the left-hand side of the equation are assumed to be combined linearly, the moving average model of order q, abbreviated as MA(q), assumes the white noise w_t on the right-hand side of the defining equation are combined linearly to form the observed data.

Definition 3.3 *The* **moving average model of order q**, *or* **MA(q)** *model, is defined to be*

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q} \tag{3.15}$$

where there are q lags in the moving average and $\theta_1, \theta_2, \ldots, \theta_q$ ($\theta_q \neq 0$) are parameters.[2] *The noise w_t is assumed to be Gaussian white noise.*

The system is the same as the infinite moving average defined as the linear process (3.12), where $\psi_0 = 1$, $\psi_j = \theta_j$, for $j = 1, \ldots, q$, and $\psi_j = 0$ for other values. We may also write the MA(q) process in the equivalent form

$$\boxed{x_t = \theta(B)w_t,} \tag{3.16}$$

using the following definition.

[2] Some texts and software packages write the MA model with negative coefficients; that is, $x_t = w_t - \theta_1 w_{t-1} - \theta_2 w_{t-2} - \cdots - \theta_q w_{t-q}$.

Definition 3.4 *The* **moving average operator** *is*

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q \tag{3.17}$$

Unlike the autoregressive process, the moving average process is stationary for any values of the parameters $\theta_1, \ldots, \theta_q$; details of this result are provided in §3.4.

Example 3.3 Autocorrelation and Sample Path of an MA(1) Process

Consider the MA(1) model $x_t = w_t + \theta w_{t-1}$. Then,

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma_w^2, & h = 0 \\ \theta\sigma_w^2, & h = 1 \\ 0, & h > 1, \end{cases}$$

and the autocorrelation function is

$$\rho(h) = \begin{cases} \frac{\theta}{(1+\theta^2)}, & h = 1 \\ 0, & h > 1. \end{cases}$$

Note $|\rho(1)| \leq 1/2$ for all values of θ (Problem 3.1). Also, x_t is correlated with x_{t-1}, but not with $x_{t-2}, x_{t-3}, \ldots$. Contrast this with the case of the AR(1) model in which the correlation between x_t and x_{t-k} is never zero. When $\theta = .5$, for example, x_t and x_{t-1} are positively correlated, and $\rho(1) = .4$. When $\theta = -.5$, x_t and x_{t-1} are negatively correlated, $\rho(1) = -.4$. Figure 3.2 shows a time plot of these two processes with $\sigma_w^2 = 1$. The series in Figure 3.2 where $\theta = .5$ is smoother than the series in Figure 3.2, where $\theta = -.5$.

A figure similar to Figure 3.2 can be created in R as follows:

```
> par(mfrow=c(2,1))
> plot(arima.sim(list(order=c(0,0,1), ma=.5), n=100),
+  ylab="x",main=(expression("MA(1)  "*theta*" = +.5")))
> plot(arima.sim(list(order=c(0,0,1), ma=-.5), n=100),
+  ylab="x",main=(expression("MA(1)  "*theta*" = -.5")))
```

Example 3.4 Non-uniqueness of MA Models and Invertibility

Using Example 3.3, we note that for an MA(1) model, $\rho(h)$ is the same for θ and $\frac{1}{\theta}$; try 5 and $\frac{1}{5}$, for example. In addition, the pair $\sigma_w^2 = 1$ and $\theta = 5$ yield the same autocovariance function as the pair $\sigma_w^2 = 25$ and $\theta = 1/5$, namely,

$$\gamma(h) = \begin{cases} 26, & h = 0 \\ 5, & h = 1 \\ 0, & h > 1. \end{cases}$$

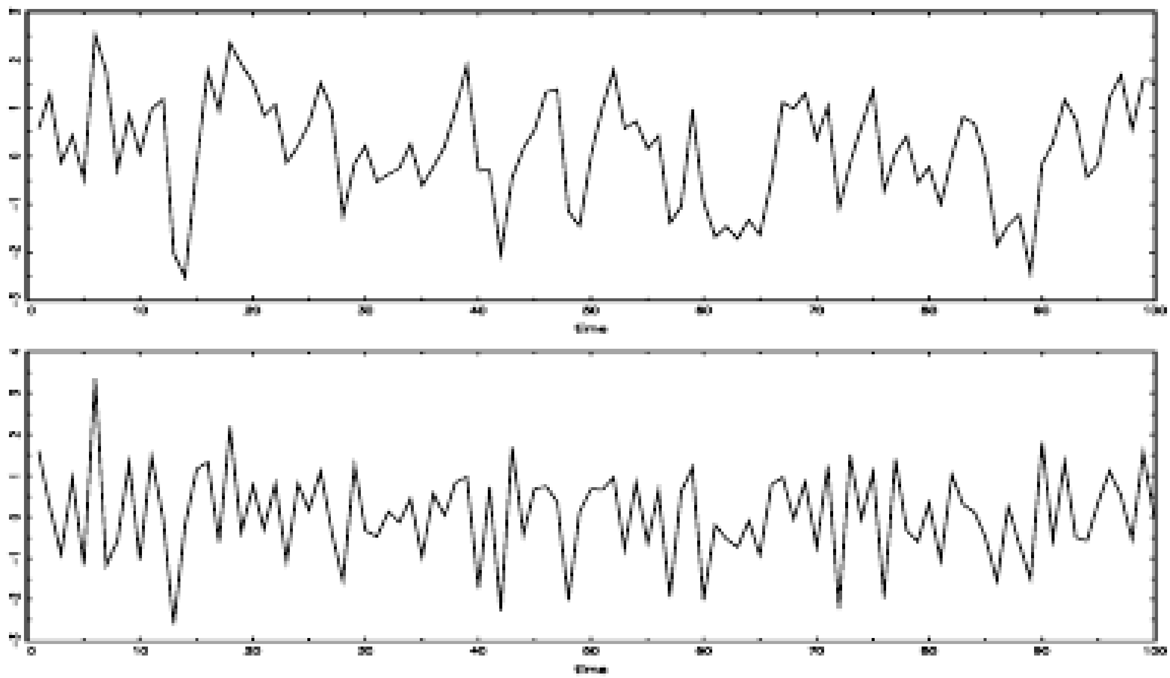

Figure 3.2 Simulated MA(1) models: $\theta = .5$ (top); $\theta = -.5$ (bottom).

Thus, the MA(1) processes

$$x_t = w_t + \frac{1}{5} w_{t-1}, \quad w_t \sim \text{iid} \;\; \text{N}(0, 25)$$

and

$$x_t = v_t + 5 v_{t-1}, \quad v_t \sim \text{iid} \;\; \text{N}(0, 1)$$

are the same because of normality (i.e., all finite distributions are the same). We can only observe the time series x_t and not the noise, w_t or v_t, so we cannot distinguish between the models. Hence, we will have to choose only one of them. For convenience, by mimicking the criterion of causality for AR models, we will choose the model with an infinite AR representation. Such a process is called an invertible process.

To discover which model is the invertible model, we can reverse the roles of x_t and w_t (because we are mimicking the AR case) and write the MA(1) model as $w_t = -\theta w_{t-1} + x_t$. Following the steps that led to (3.6), if $|\theta| < 1$, then $w_t = \sum_{j=0}^{\infty} (-\theta)^j x_{t-j}$, which is the desired infinite AR representation of the model. Hence, given a choice, we will choose the model with $\sigma_w^2 = 25$ and $\theta = 1/5$ because it is invertible.

As in the AR case, the polynomial, $\theta(z)$, corresponding to the moving average operators, $\theta(B)$, will be useful in exploring general properties of MA processes. For example, following the steps of equations (3.11)–(3.14), we can write the MA(1) model as $x_t = \theta(B) w_t$, where $\theta(B) = 1 + \theta B$. If $|\theta| < 1$, then we can write the model as $\pi(B) x_t = w_t$, where $\pi(B) = \theta^{-1}(B)$. Let

$\theta(z) = 1 + \theta z$, for $|z| \le 1$, then $\pi(z) = \theta^{-1}(z) = 1/(1+\theta z) = \sum_{j=0}^{\infty}(-\theta)^j z^j$, and we determine that $\pi(B) = \sum_{j=0}^{\infty}(-\theta)^j B^j$.

Autoregressive Moving Average Models

We now proceed with the general development of autoregressive, moving average, and mixed autoregressive moving average (ARMA), models for stationary time series.

Definition 3.5 *A time series* $\{x_t;\ t = 0, \pm1, \pm2, \dots\}$ *is* **ARMA(p, q)** *if it is stationary and*

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}, \tag{3.18}$$

with $\phi_p \neq 0$, $\theta_q \neq 0$, *and* $\sigma_w^2 > 0$. *The parameters* p *and* q *are called the autoregressive and the moving average orders, respectively. If* x_t *has a nonzero mean* μ, *we set* $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$ *and write the model as*

$$x_t = \alpha + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}. \tag{3.19}$$

Unless stated otherwise, $\{w_t;\ t = 0, \pm1, \pm2, \dots\}$ *is a Gaussian white noise sequence.*

As previously noted, when $q = 0$, the model is called an autoregressive model of order p, AR(p), and when $p = 0$, the model is called a moving average model of order q, MA(q). To aid in the investigation of ARMA models, it will be useful to write them using the AR operator, (3.5), and the MA operator, (3.17). In particular, the ARMA(p, q) model in (3.18) can then be written in concise form as

$$\boxed{\phi(B)x_t = \theta(B)w_t.} \tag{3.20}$$

Before we discuss the conditions under which (3.18) is causal and invertible, we point out a potential problem with the ARMA model.

Example 3.5 Parameter Redundancy

Consider a white noise process $x_t = w_t$. Equivalently, we can write this as $.5x_{t-1} = .5w_{t-1}$ by shifting back one unit of time and multiplying by .5. Now, subtract the two representations to obtain

$$x_t - .5x_{t-1} = w_t - .5w_{t-1},$$

or

$$x_t = .5x_{t-1} - .5w_{t-1} + w_t, \tag{3.21}$$

which looks like an ARMA(1, 1) model. Of course, x_t is still white noise; nothing has changed in this regard [i.e., $x_t = w_t$ is the solution to (3.21)],