Imagine a battery of I such computers generating sequences $\{y_t^{(1)}\}_{t=-\infty}^{\infty}$, $\{y_t^{(2)}\}_{t=-\infty}^{\infty}, \ldots, \{y_t^{(I)}\}_{t=-\infty}^{\infty}$, and consider selecting the observation associated with date t from each sequence:

$$\{y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(I)}\}.$$

This would be described as a sample of I realizations of the random variable Y_t. This random variable has some density, denoted $f_{Y_t}(y_t)$, which is called the *unconditional density* of Y_t. For example, for the Gaussian white noise process, this density is given by

$$f_{Y_t}(y_t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-y_t^2}{2\sigma^2}\right].$$

The *expectation* of the tth observation of a time series refers to the mean of this probability distribution, provided it exists:

$$E(Y_t) \equiv \int_{-\infty}^{\infty} y_t f_{Y_t}(y_t)\, dy_t. \tag{3.1.3}$$

We might view this as the probability limit of the ensemble average:

$$E(Y_t) = \operatorname*{plim}_{I \to \infty} (1/I) \sum_{i=1}^{I} Y_t^{(i)}. \tag{3.1.4}$$

For example, if $\{Y_t\}_{t=-\infty}^{\infty}$ represents the sum of a constant μ plus a Gaussian white noise process $\{\varepsilon_t\}_{t=-\infty}^{\infty}$,

$$Y_t = \mu + \varepsilon_t, \tag{3.1.5}$$

then its mean is

$$E(Y_t) = \mu + E(\varepsilon_t) = \mu. \tag{3.1.6}$$

If Y_t is a time trend plus Gaussian white noise,

$$Y_t = \beta t + \varepsilon_t, \tag{3.1.7}$$

then its mean is

$$E(Y_t) = \beta t. \tag{3.1.8}$$

Sometimes for emphasis the expectation $E(Y_t)$ is called the *unconditional mean* of Y_t. The unconditional mean is denoted μ_t:

$$E(Y_t) = \mu_t.$$

Note that this notation allows the general possibility that the mean can be a function of the date of the observation t. For the process [3.1.7] involving the time trend, the mean [3.1.8] is a function of time, whereas for the constant plus Gaussian white noise, the mean [3.1.6] is not a function of time.

The *variance* of the random variable Y_t (denoted γ_{0t}) is similarly defined as

$$\gamma_{0t} \equiv E(Y_t - \mu_t)^2 = \int_{-\infty}^{\infty} (y_t - \mu_t)^2 f_{Y_t}(y_t)\, dy_t. \tag{3.1.9}$$

For example, for the process [3.1.7], the variance is

$$\gamma_{0t} = E(Y_t - \beta t)^2 = E(\varepsilon_t^2) = \sigma^2.$$

Autocovariance

Given a particular realization such as $\{y_t^{(1)}\}_{t=-\infty}^{\infty}$ on a time series process, consider constructing a vector $\mathbf{x}_t^{(1)}$ associated with date t. This vector consists of the $[j + 1]$ most recent observations on y as of date t for that realization:

$$\mathbf{x}_t^{(1)} \equiv \begin{bmatrix} y_t^{(1)} \\ y_{t-1}^{(1)} \\ \vdots \\ y_{t-j}^{(1)} \end{bmatrix}.$$

We think of each realization $\{y_t\}_{t=-\infty}^{\infty}$ as generating one particular value of the vector $\mathbf{x}_t$ and want to calculate the probability distribution of this vector $\mathbf{x}_t^{(i)}$ across realizations i. This distribution is called the *joint distribution* of $(Y_t, Y_{t-1}, \ldots, Y_{t-j})$. From this distribution we can calculate the jth *autocovariance* of Y_t (denoted γ_{jt}):

$$\begin{aligned}\gamma_{jt} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(y_t - \mu_t)(y_{t-j} - \mu_{t-j}) \\ &\quad \times f_{Y_t,Y_{t-1},\ldots,Y_{t-j}}(y_t, y_{t-1}, \ldots, y_{t-j})\, dy_t\, dy_{t-1} \cdots dy_{t-j} \qquad [3.1.10] \\ &= E(Y_t - \mu_t)(Y_{t-j} - \mu_{t-j}).\end{aligned}$$

Note that [3.1.10] has the form of a covariance between two variables X and Y:

$$\mathrm{Cov}(X, Y) = E(X - \mu_X)(Y - \mu_Y).$$

Thus [3.1.10] could be described as the covariance of Y_t with its own lagged value; hence, the term "autocovariance." Notice further from [3.1.10] that the 0th autocovariance is just the variance of Y_t, as anticipated by the notation γ_{0t} in [3.1.9].

The autocovariance γ_{jt} can be viewed as the $(1, j + 1)$ element of the variance-covariance matrix of the vector $\mathbf{x}_t$. For this reason, the autocovariances are described as the second moments of the process for Y_t.

Again it may be helpful to think of the jth autocovariance as the probability limit of an ensemble average:

$$\gamma_{jt} = \operatorname*{plim}_{I\to\infty} (1/I) \sum_{i=1}^{I} [Y_t^{(i)} - \mu_t] \cdot [Y_{t-j}^{(i)} - \mu_{t-j}]. \qquad [3.1.11]$$

As an example of calculating autocovariances, note that for the process in [3.1.5] the autocovariances are all zero for $j \neq 0$:

$$\gamma_{jt} = E(Y_t - \mu)(Y_{t-j} - \mu) = E(\varepsilon_t\varepsilon_{t-j}) = 0 \qquad \text{for } j \neq 0.$$

Stationarity

If neither the mean μ_t nor the autocovariances γ_{jt} depend on the date t, then the process for Y_t is said to be *covariance-stationary* or *weakly stationary*:

$$\begin{aligned} E(Y_t) &= \mu && \text{for all } t \\ E(Y_t - \mu)(Y_{t-j} - \mu) &= \gamma_j && \text{for all } t \text{ and any } j. \end{aligned}$$

For example, the process in [3.1.5] is covariance-stationary:

$$E(Y_t) = \mu$$

$$E(Y_t - \mu)(Y_{t-j} - \mu) = \begin{cases} \sigma^2 & \text{for } j = 0 \\ 0 & \text{for } j \neq 0. \end{cases}$$

By contrast, the process of [3.1.7] is not covariance-stationary, because its mean, βt, is a function of time.

Notice that if a process is covariance-stationary, the covariance between Y_t and Y_{t-j} depends only on j, the length of time separating the observations, and not on t, the date of the observation. It follows that for a covariance-stationary process, γ_j and γ_{-j} would represent the same magnitude. To see this, recall the definition

$$\gamma_j = E(Y_t - \mu)(Y_{t-j} - \mu). \qquad [3.1.12]$$

If the process is covariance-stationary, then this magnitude is the same for any value of t we might have chosen; for example, we can replace t with $t + j$:

$$\gamma_j = E(Y_{t+j} - \mu)(Y_{[t+j]-j} - \mu) = E(Y_{t+j} - \mu)(Y_t - \mu) = E(Y_t - \mu)(Y_{t+j} - \mu).$$

But referring again to the definition [3.1.12], this last expression is just the definition of γ_{-j}. Thus, for any covariance-stationary process,

$$\gamma_j = \gamma_{-j} \qquad \text{for all integers } j. \qquad [3.1.13]$$

A different concept is that of *strict stationarity*. A process is said to be strictly stationary if, for any values of $j_1, j_2, \ldots, j_n$, the joint distribution of $(Y_t, Y_{t+j_1}, Y_{t+j_2}, \ldots, Y_{t+j_n})$ depends only on the intervals separating the dates $(j_1, j_2, \ldots, j_n)$ and not on the date itself (t). Notice that if a process is strictly stationary with finite second moments, then it must be covariance-stationary—if the densities over which we are integrating in [3.1.3] and [3.1.10] do not depend on time, then the moments μ_t and γ_{jt} will not depend on time. However, it is possible to imagine a process that is covariance-stationary but not strictly stationary; the mean and autocovariances could not be functions of time, but perhaps higher moments such as $E(Y_t^3)$ are.

In this text the term "stationary" by itself is taken to mean "covariance-stationary."

A process $\{Y_t\}$ is said to be *Gaussian* if the joint density

$$f_{Y_t, Y_{t+j_1}, \ldots, Y_{t+j_n}}(y_t, y_{t+j_1}, \ldots, y_{t+j_n})$$

is Gaussian for any $j_1, j_2, \ldots, j_n$. Since the mean and variance are all that are needed to parameterize a multivariate Gaussian distribution completely, a covariance-stationary Gaussian process is strictly stationary.

Ergodicity

We have viewed expectations of a time series in terms of ensemble averages such as [3.1.4] and [3.1.11]. These definitions may seem a bit contrived, since usually all one has available is a single realization of size T from the process, which we earlier denoted $\{y_1^{(1)}, y_2^{(1)}, \ldots, y_T^{(1)}\}$. From these observations we would calculate the sample mean $\bar{y}$. This, of course, is not an ensemble average but rather a time average:

$$\bar{y} \equiv (1/T) \sum_{t=1}^{T} y_t^{(1)}. \qquad [3.1.14]$$

Whether time averages such as [3.1.14] eventually converge to the ensemble concept $E(Y_t)$ for a stationary process has to do with *ergodicity*. A covariance-stationary process is said to be *ergodic for the mean* if [3.1.14] converges in probability to $E(Y_t)$ as $T \to \infty$.[1] A process will be ergodic for the mean provided that the autocovariance γ_j goes to zero sufficiently quickly as j becomes large. In Chapter 7 we will see that if the autocovariances for a covariance-stationary process satisfy

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty, \qquad [3.1.15]$$

then $\{Y_t\}$ is ergodic for the mean.

Similarly, a covariance-stationary process is said to be ergodic for second moments if

$$[1/(T-j)] \sum_{t=j+1}^{T} (Y_t - \mu)(Y_{t-j} - \mu) \xrightarrow{p} \gamma_j$$

for all j. Sufficient conditions for second-moment ergodicity will be presented in Chapter 7. In the special case where $\{Y_t\}$ is a stationary Gaussian process, condition [3.1.15] is sufficient to ensure ergodicity for all moments.

For many applications, stationarity and ergodicity turn out to amount to the same requirements. For purposes of clarifying the concepts of stationarity and ergodicity, however, it may be helpful to consider an example of a process that is stationary but not ergodic. Suppose the mean $\mu^{(i)}$ for the ith realization $\{y_t^{(i)}\}_{t=-\infty}^{\infty}$ is generated from a $N(0, \lambda^2)$ distribution, say

$$Y_t^{(i)} = \mu^{(i)} + \varepsilon_t. \qquad [3.1.16]$$

Here $\{\varepsilon_t\}$ is a Gaussian white noise process with mean zero and variance σ^2 that is independent of $\mu^{(i)}$. Notice that

$$\mu_t = E(\mu^{(i)}) + E(\varepsilon_t) = 0.$$

Also,

$$\gamma_{0t} = E(\mu^{(i)} + \varepsilon_t)^2 = \lambda^2 + \sigma^2$$

and

$$\gamma_{jt} = E(\mu^{(i)} + \varepsilon_t)(\mu^{(i)} + \varepsilon_{t-j}) = \lambda^2 \qquad \text{for } j \neq 0.$$

Thus the process of [3.1.16] is covariance-stationary. It does not satisfy the sufficient condition [3.1.15] for ergodicity for the mean, however, and indeed, the time average

$$(1/T) \sum_{t=1}^{T} Y_t^{(i)} = (1/T) \sum_{t=1}^{T} (\mu^{(i)} + \varepsilon_t) = \mu^{(i)} + (1/T) \sum_{t=1}^{T} \varepsilon_t$$

converges to $\mu^{(i)}$ rather than to zero, the mean of Y_t.

3.2. *White Noise*

The basic building block for all the processes considered in this chapter is a sequence $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ whose elements have mean zero and variance σ^2,

$$E(\varepsilon_t) = 0 \qquad [3.2.1]$$

$$E(\varepsilon_t^2) = \sigma^2, \qquad [3.2.2]$$

and for which the ε's are uncorrelated across time:

[1]Often "ergodicity" is used in a more general sense; see Anderson and Moore (1979, p. 319) or Hannan (1970, pp. 201–20).

$$E(\varepsilon_t \varepsilon_\tau) = 0 \qquad \text{for } t \neq \tau. \tag{3.2.3}$$

A process satisfying [3.2.1] through [3.2.3] is described as a *white noise process*.

We shall on occasion wish to replace [3.2.3] with the slightly stronger condition that the ε's are independent across time:

$$\varepsilon_t,\ \varepsilon_\tau \text{ independent for } t \neq \tau. \tag{3.2.4}$$

Notice that [3.2.4] implies [3.2.3] but [3.2.3] does not imply [3.2.4]. A process satisfying [3.2.1] through [3.2.4] is called an *independent white noise process*.

Finally, if [3.2.1] through [3.2.4] hold along with

$$\varepsilon_t \sim N(0, \sigma^2), \tag{3.2.5}$$

then we have the *Gaussian white noise process*.

3.3. *Moving Average Processes*

The First-Order Moving Average Process

Let $\{\varepsilon_t\}$ be white noise as in [3.2.1] through [3.2.3], and consider the process

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \tag{3.3.1}$$

where μ and θ could be any constants. This time series is called a *first-order moving average process*, denoted $MA(1)$. The term "moving average" comes from the fact that Y_t is constructed from a weighted sum, akin to an average, of the two most recent values of ε.

The expectation of Y_t is given by

$$E(Y_t) = E(\mu + \varepsilon_t + \theta\varepsilon_{t-1}) = \mu + E(\varepsilon_t) + \theta \cdot E(\varepsilon_{t-1}) = \mu. \tag{3.3.2}$$

We used the symbol μ for the constant term in [3.3.1] in anticipation of the result that this constant term turns out to be the mean of the process.

The variance of Y_t is

$$\begin{aligned} E(Y_t - \mu)^2 &= E(\varepsilon_t + \theta\varepsilon_{t-1})^2 \\ &= E(\varepsilon_t^2 + 2\theta\varepsilon_t\varepsilon_{t-1} + \theta^2\varepsilon_{t-1}^2) \\ &= \sigma^2 + 0 + \theta^2\sigma^2 \\ &= (1 + \theta^2)\sigma^2. \end{aligned} \tag{3.3.3}$$

The first autocovariance is

$$\begin{aligned} E(Y_t - \mu)(Y_{t-1} - \mu) &= E(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2}) \\ &= E(\varepsilon_t\varepsilon_{t-1} + \theta\varepsilon_{t-1}^2 + \theta\varepsilon_t\varepsilon_{t-2} + \theta^2\varepsilon_{t-1}\varepsilon_{t-2}) \\ &= 0 + \theta\sigma^2 + 0 + 0. \end{aligned} \tag{3.3.4}$$

Higher autocovariances are all zero:

$$E(Y_t - \mu)(Y_{t-j} - \mu) = E(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-j} + \theta\varepsilon_{t-j-1}) = 0 \qquad \text{for } j > 1. \tag{3.3.5}$$

Since the mean and autocovariances are not functions of time, an $MA(1)$ process is covariance-stationary regardless of the value of θ. Furthermore, [3.1.15] is clearly satisfied:

$$\sum_{j=0}^{\infty} |\gamma_j| = (1 + \theta^2)\sigma^2 + |\theta\sigma^2|.$$

Thus, if $\{\varepsilon_t\}$ is Gaussian white noise, then the $MA(1)$ process [3.3.1] is ergodic for all moments.

The jth *autocorrelation* of a covariance-stationary process (denoted ρ_j) is defined as its jth autocovariance divided by the variance:

$$\rho_j \equiv \gamma_j/\gamma_0. \qquad [3.3.6]$$

Again the terminology arises from the fact that ρ_j is the correlation between Y_t and Y_{t-j}:

$$\text{Corr}(Y_t, Y_{t-j}) = \frac{\text{Cov}(Y_t, Y_{t-j})}{\sqrt{\text{Var}(Y_t)}\sqrt{\text{Var}(Y_{t-j})}} = \frac{\gamma_j}{\sqrt{\gamma_0}\sqrt{\gamma_0}} = \rho_j.$$

Since ρ_j is a correlation, $|\rho_j| \leq 1$ for all j, by the Cauchy-Schwarz inequality. Notice also that the 0th autocorrelation ρ_0 is equal to unity for any covariance-stationary process by definition.

From [3.3.3] and [3.3.4], the first autocorrelation for an $MA(1)$ process is given by

$$\rho_1 = \frac{\theta\sigma^2}{(1 + \theta^2)\sigma^2} = \frac{\theta}{(1 + \theta^2)}. \qquad [3.3.7]$$

Higher autocorrelations are all zero.

The autocorrelation ρ_j can be plotted as a function of j as in Figure 3.1. Panel (a) shows the autocorrelation function for white noise, while panel (b) gives the autocorrelation function for the $MA(1)$ process:

$$Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}.$$

For different specifications of θ we would obtain different values for the first autocorrelation ρ_1 in [3.3.7]. Positive values of θ induce positive autocorrelation in the series. In this case, an unusually large value of Y_t is likely to be followed by a larger-than-average value for Y_{t+1}, just as a smaller-than-average Y_t may well be followed by a smaller-than-average Y_{t+1}. By contrast, negative values of θ imply negative autocorrelation—a large Y_t might be expected to be followed by a small value for Y_{t+1}.

The values for ρ_1 implied by different specifications of θ are plotted in Figure 3.2. Notice that the largest possible value for ρ_1 is 0.5; this occurs if $\theta = 1$. The smallest value for ρ_1 is -0.5, which occurs if $\theta = -1$. For any value of ρ_1 between -0.5 and 0.5, there are two different values of θ that could produce that autocorrelation. This is because the value of $\theta/(1 + \theta^2)$ is unchanged if θ is replaced by $1/\theta$:

$$\rho_1 = \frac{(1/\theta)}{1 + (1/\theta)^2} = \frac{\theta^2 \cdot (1/\theta)}{\theta^2[1 + (1/\theta)^2]} = \frac{\theta}{\theta^2 + 1}.$$

For example, the processes

$$Y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$$

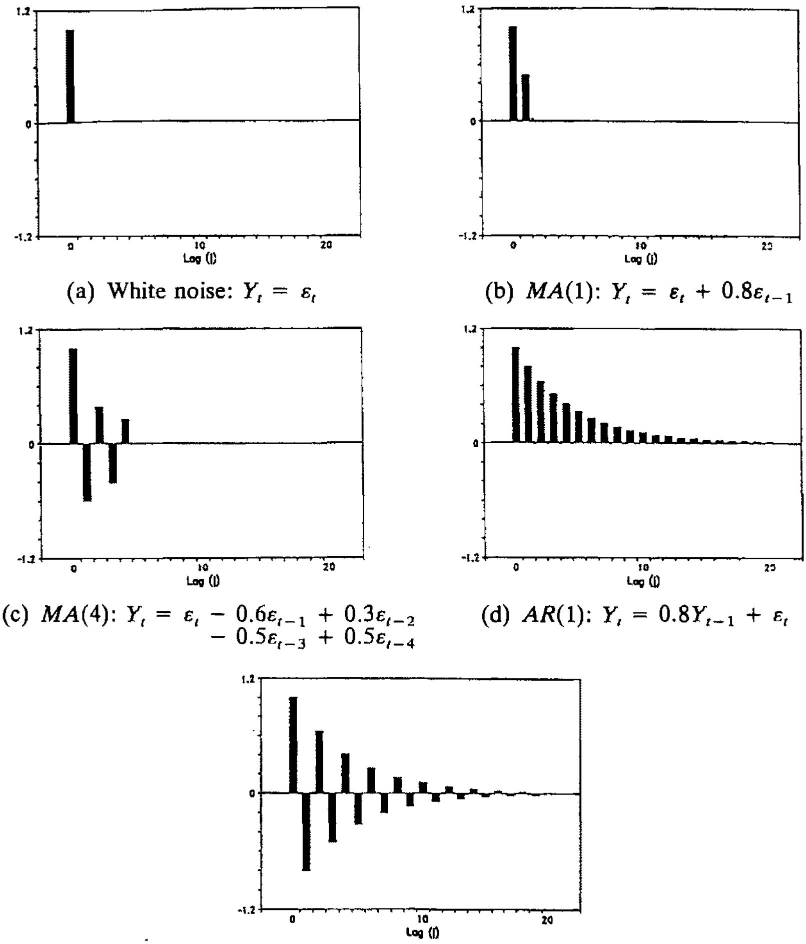
and

$$Y_t = \varepsilon_t + 2\varepsilon_{t-1}$$

would have the same autocorrelation function:

$$\rho_1 = \frac{2}{(1 + 2^2)} = \frac{0.5}{(1 + 0.5^2)} = 0.4.$$

We will have more to say about the relation between two $MA(1)$ processes that share the same autocorrelation function in Section 3.7.



(a) White noise: $Y_t = \varepsilon_t$

(b) $MA(1)$: $Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}$

(c) $MA(4)$: $Y_t = \varepsilon_t - 0.6\varepsilon_{t-1} + 0.3\varepsilon_{t-2} - 0.5\varepsilon_{t-3} + 0.5\varepsilon_{t-4}$

(d) $AR(1)$: $Y_t = 0.8Y_{t-1} + \varepsilon_t$

(e) $AR(1)$: $Y_t = -0.8Y_{t-1} + \varepsilon_t$

FIGURE 3.1 Autocorrelation functions for assorted *ARMA* processes.

The q*th-Order Moving Average Process*

A *qth-order moving average process*, denoted $MA(q)$, is characterized by

$$Y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}, \qquad [3.3.8]$$

where $\{\varepsilon_t\}$ satisfies [3.2.1] through [3.2.3] and $(\theta_1, \theta_2, \ldots, \theta_q)$ could be any real numbers. The mean of [3.3.8] is again given by μ:

$$E(Y_t) = \mu + E(\varepsilon_t) + \theta_1 \cdot E(\varepsilon_{t-1}) + \theta_2 \cdot E(\varepsilon_{t-2}) + \cdots + \theta_q \cdot E(\varepsilon_{t-q}) = \mu.$$

The variance of an $MA(q)$ process is

$$\gamma_0 = E(Y_t - \mu)^2 = E(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q})^2. \qquad [3.3.9]$$

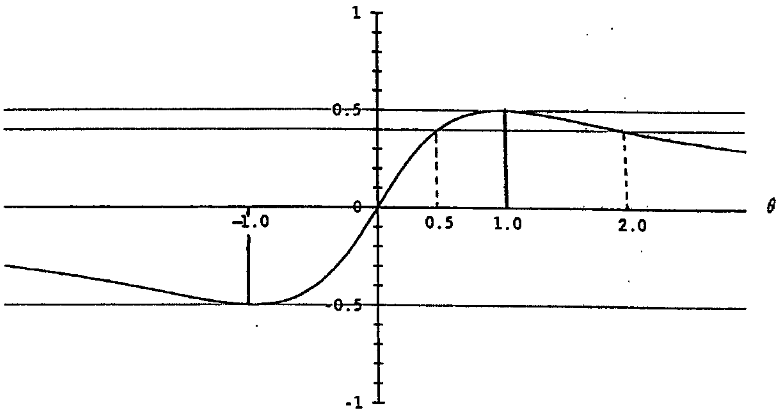


FIGURE 3.2 The first autocorrelation (ρ_1) for an $MA(1)$ process possible for different values of θ.

Since the ε's are uncorrelated, the variance [3.3.9] is[2]

$$\gamma_0 = \sigma^2 + \theta_1^2\sigma^2 + \theta_2^2\sigma^2 + \cdots + \theta_q^2\sigma^2 = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma^2. \quad [3.3.10]$$

For $j = 1, 2, \ldots, q$,

$$\begin{aligned} \gamma_j &= E[(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}) \\ &\quad \times (\varepsilon_{t-j} + \theta_1\varepsilon_{t-j-1} + \theta_2\varepsilon_{t-j-2} + \cdots + \theta_q\varepsilon_{t-j-q})] \\ &= E[\theta_j\varepsilon_{t-j}^2 + \theta_{j+1}\theta_1\varepsilon_{t-j-1}^2 + \theta_{j+2}\theta_2\varepsilon_{t-j-2}^2 + \cdots + \theta_q\theta_{q-j}\varepsilon_{t-q}^2]. \end{aligned} \quad [3.3.11]$$

Terms involving ε's at different dates have been dropped because their product has expectation zero, and θ_0 is defined to be unity. For $j > q$, there are no ε's with common dates in the definition of γ_j, and so the expectation is zero. Thus,

$$\gamma_j = \begin{cases} [\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \cdots + \theta_q\theta_{q-j}]\cdot\sigma^2 & \text{for } j = 1, 2, \ldots, q \\ 0 & \text{for } j > q. \end{cases} \quad [3.3.12]$$

For example, for an $MA(2)$ process,

$$\begin{aligned} \gamma_0 &= [1 + \theta_1^2 + \theta_2^2]\cdot\sigma^2 \\ \gamma_1 &= [\theta_1 + \theta_2\theta_1]\cdot\sigma^2 \\ \gamma_2 &= [\theta_2]\cdot\sigma^2 \\ \gamma_3 &= \gamma_4 = \cdots = 0. \end{aligned}$$

For any values of $(\theta_1, \theta_2, \ldots, \theta_q)$, the $MA(q)$ process is thus covariance-stationary. Condition [3.1.15] is satisfied, so for Gaussian ε_t the $MA(q)$ process is also ergodic for all moments. The autocorrelation function is zero after q lags, as in panel (c) of Figure 3.1.

The Infinite-Order Moving Average Process

The $MA(q)$ process can be written

$$Y_t = \mu + \sum_{j=0}^{q} \theta_j\varepsilon_{t-j}$$

[2]See equation [A.5.18] in Appendix A at the end of the book.

with $\theta_0 \equiv 1$. Consider the process that results as $q \to \infty$:

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots . \qquad [3.3.13]$$

This could be described as an $MA(\infty)$ process. To preserve notational flexibility later, we will use ψ's for the coefficients of an infinite-order moving average process and θ's for the coefficients of a finite-order moving average process.

Appendix 3.A to this chapter shows that the infinite sequence in [3.3.13] generates a well defined covariance-stationary process provided that

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty. \qquad [3.3.14]$$

It is often convenient to work with a slightly stronger condition than [3.3.14]:

$$\sum_{j=0}^{\infty} |\psi_j| < \infty. \qquad [3.3.15]$$

A sequence of numbers $\{\psi_j\}_{j=0}^{\infty}$ satisfying [3.3.14] is said to be *square summable*, whereas a sequence satisfying [3.3.15] is said to be *absolutely summable*. Absolute summability implies square-summability, but the converse does not hold—there are examples of square-summable sequences that are not absolutely summable (again, see Appendix 3.A).

The mean and autocovariances of an $MA(\infty)$ process with absolutely summable coefficients can be calculated from a simple extrapolation of the results for an $MA(q)$ process:[3]

$$E(Y_t) = \lim_{T\to\infty} E(\mu + \psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots + \psi_T\varepsilon_{t-T}) \qquad [3.3.16]$$
$$= \mu$$

$$\begin{aligned}\gamma_0 &= E(Y_t - \mu)^2 \\ &= \lim_{T\to\infty} E(\psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots + \psi_T\varepsilon_{t-T})^2 \\ &= \lim_{T\to\infty} (\psi_0^2 + \psi_1^2 + \psi_2^2 + \cdots + \psi_T^2)\cdot\sigma^2\end{aligned} \qquad [3.3.17]$$

$$\begin{aligned}\gamma_j &= E(Y_t - \mu)(Y_{t-j} - \mu) \\ &= \sigma^2(\psi_j\psi_0 + \psi_{j+1}\psi_1 + \psi_{j+2}\psi_2 + \psi_{j+3}\psi_3 + \cdots).\end{aligned} \qquad [3.3.18]$$

Moreover, an $MA(\infty)$ process with absolutely summable coefficients has absolutely summable autocovariances:

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty. \qquad [3.3.19]$$

Hence, an $MA(\infty)$ process satisfying [3.3.15] is ergodic for the mean (see Appendix 3.A). If the ε's are Gaussian, then the process is ergodic for all moments.

[3]Absolute summability of $\{\psi_j\}_{j=0}^{\infty}$ and existence of the second moment $E(\varepsilon_t^2)$ are sufficient conditions to permit interchanging the order of integration and summation. Specifically, if $\{X_T\}_{T=1}^{\infty}$ is a sequence of random variables such that

$$\sum_{T=1}^{\infty} E|X_T| < \infty,$$

then

$$E\left\{\sum_{T=1}^{\infty} X_T\right\} = \sum_{T=1}^{\infty} E(X_T).$$

See Rao (1973, p. 111).

3.4. *Autoregressive Processes*

The First-Order Autoregressive Process

A *first-order autoregression*, denoted $AR(1)$, satisfies the following difference equation:

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t. \qquad [3.4.1]$$

Again, $\{\varepsilon_t\}$ is a white noise sequence satisfying [3.2.1] through [3.2.3]. Notice that [3.4.1] takes the form of the first-order difference equation [1.1.1] or [2.2.1] in which the input variable w_t is given by $w_t = c + \varepsilon_t$. We know from the analysis of first-order difference equations that if $|\phi| \geq 1$, the consequences of the ε's for Y accumulate rather than die out over time. It is thus perhaps not surprising that when $|\phi| \geq 1$, there does not exist a covariance-stationary process for Y_t with finite variance that satisfies [3.4.1]. In the case when $|\phi| < 1$, there is a covariance-stationary process for Y_t satisfying [3.4.1]. It is given by the stable solution to [3.4.1] characterized in [2.2.9]:

$$\begin{aligned} Y_t &= (c + \varepsilon_t) + \phi\cdot(c + \varepsilon_{t-1}) + \phi^2\cdot(c + \varepsilon_{t-2}) + \phi^3\cdot(c + \varepsilon_{t-3}) + \cdots \\ &= [c/(1 - \phi)] + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \cdots. \end{aligned} \qquad [3.4.2]$$

This can be viewed as an $MA(\infty)$ process as in [3.3.13] with ψ_j given by ϕ^j. When $|\phi| < 1$, condition [3.3.15] is satisfied:

$$\sum_{j=0}^{\infty} |\psi_j| = \sum_{j=0}^{\infty} |\phi|^j,$$

which equals $1/(1 - |\phi|)$ provided that $|\phi| < 1$. The remainder of this discussion of first-order autoregressive processes assumes that $|\phi| < 1$. This ensures that the $MA(\infty)$ representation exists and can be manipulated in the obvious way, and that the $AR(1)$ process is ergodic for the mean.

Taking expectations of [3.4.2], we see that

$$E(Y_t) = [c/(1 - \phi)] + 0 + 0 + \cdots,$$

so that the mean of a stationary $AR(1)$ process is

$$\mu = c/(1 - \phi). \qquad [3.4.3]$$

The variance is

$$\begin{aligned} \gamma_0 &= E(Y_t - \mu)^2 \\ &= E(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \cdots)^2 \\ &= (1 + \phi^2 + \phi^4 + \phi^6 + \cdots)\cdot\sigma^2 \\ &= \sigma^2/(1 - \phi^2), \end{aligned} \qquad [3.4.4]$$

while the jth autocovariance is

$$\begin{aligned} \gamma_j &= E(Y_t - \mu)(Y_{t-j} - \mu) \\ &= E[\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots + \phi^j\varepsilon_{t-j} + \phi^{j+1}\varepsilon_{t-j-1} \\ &\quad + \phi^{j+2}\varepsilon_{t-j-2} + \cdots] \times [\varepsilon_{t-j} + \phi\varepsilon_{t-j-1} + \phi^2\varepsilon_{t-j-2} + \cdots] \\ &= [\phi^j + \phi^{j+2} + \phi^{j+4} + \cdots]\cdot\sigma^2 \\ &= \phi^j[1 + \phi^2 + \phi^4 + \cdots]\cdot\sigma^2 \\ &= [\phi^j/(1 - \phi^2)]\cdot\sigma^2. \end{aligned} \qquad [3.4.5]$$

It follows from [3.4.4] and [3.4.5] that the autocorrelation function,

$$\rho_j = \gamma_j/\gamma_0 = \phi^j, \tag{3.4.6}$$

follows a pattern of geometric decay as in panel (d) of Figure 3.1. Indeed, the autocorrelation function [3.4.6] for a stationary $AR(1)$ process is identical to the dynamic multiplier or impulse-response function [1.1.10]; the effect of a one-unit increase in ε_t on Y_{t+j} is equal to the correlation between Y_t and Y_{t+j}. A positive value of ϕ, like a positive value of θ for an $MA(1)$ process, implies positive correlation between Y_t and Y_{t+1}. A negative value of ϕ implies negative first-order but positive second-order autocorrelation, as in panel (e) of Figure 3.1.

Figure 3.3 shows the effect on the appearance of the time series $\{y_t\}$ of varying the parameter ϕ. The panels show realizations of the process in [3.4.1] with $c = 0$ and $\varepsilon_t \sim N(0, 1)$ for different values of the autoregressive parameter ϕ. Panel (a) displays white noise ($\phi = 0$). A series with no autocorrelation looks choppy and patternless to the eye; the value of one observation gives no information about the value of the next observation. For $\phi = 0.5$ (panel (b)), the series seems smoother, with observations above or below the mean often appearing in clusters of modest duration. For $\phi = 0.9$ (panel (c)), departures from the mean can be quite prolonged; strong shocks take considerable time to die out.

The moments for a stationary $AR(1)$ were derived above by viewing it as an $MA(\infty)$ process. A second way to arrive at the same results is to assume that the process is covariance-stationary and calculate the moments directly from the difference equation [3.4.1]. Taking expectations of both sides of [3.4.1],

$$E(Y_t) = c + \phi \cdot E(Y_{t-1}) + E(\varepsilon_t). \tag{3.4.7}$$

Assuming that the process is covariance-stationary,

$$E(Y_t) = E(Y_{t-1}) = \mu. \tag{3.4.8}$$

Substituting [3.4.8] into [3.4.7],

$$\mu = c + \phi\mu + 0$$

or

$$\mu = c/(1 - \phi), \tag{3.4.9}$$

reproducing the earlier result [3.4.3].

Notice that formula [3.4.9] is clearly not generating a sensible statement if $|\phi| \geq 1$. For example, if $c > 0$ and $\phi > 1$, then Y_t in [3.4.1] is equal to a positive constant plus a positive number times its lagged value plus a mean-zero random variable. Yet [3.4.9] seems to assert that Y_t would be negative on average for such a process! The reason that formula [3.4.9] is not valid when $|\phi| \geq 1$ is that we assumed in [3.4.8] that Y_t is covariance-stationary, an assumption which is not correct when $|\phi| \geq 1$.

To find the second moments of Y_t in an analogous manner, use [3.4.3] to rewrite [3.4.1] as

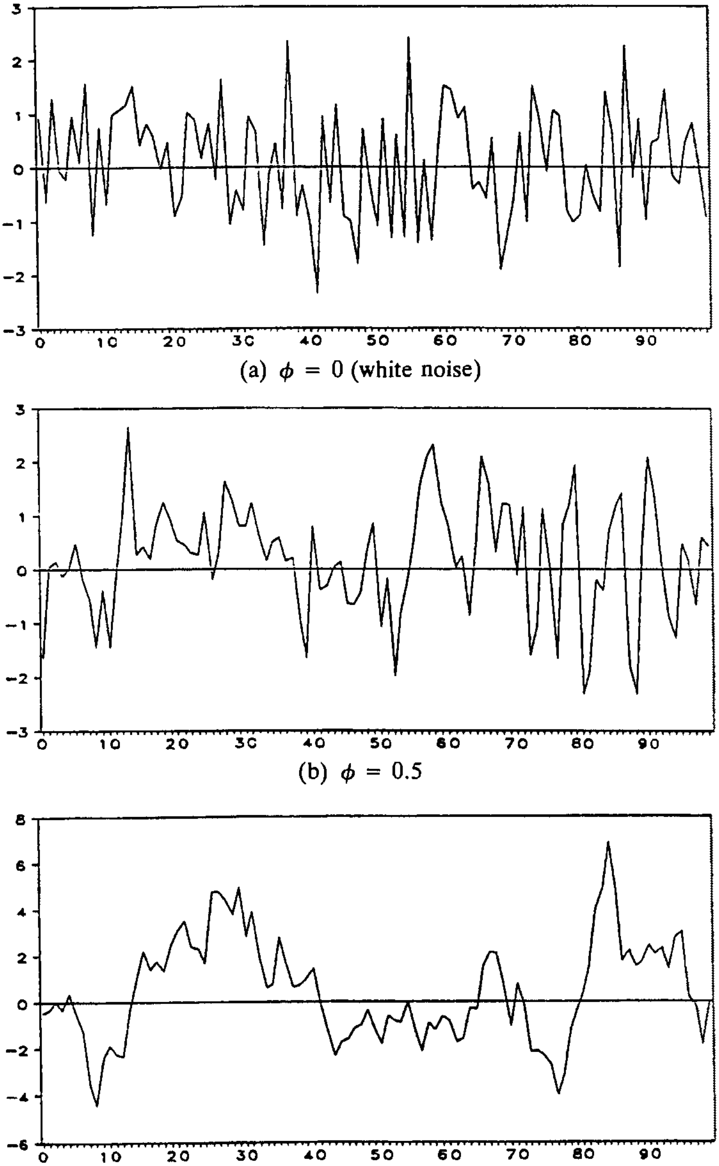
$$Y_t = \mu(1 - \phi) + \phi Y_{t-1} + \varepsilon_t$$

or

$$(Y_t - \mu) = \phi(Y_{t-1} - \mu) + \varepsilon_t. \tag{3.4.10}$$

Now square both sides of [3.4.10] and take expectations:

$$E(Y_t - \mu)^2 = \phi^2 E(Y_{t-1} - \mu)^2 + 2\phi E[(Y_{t-1} - \mu)\varepsilon_t] + E(\varepsilon_t^2). \tag{3.4.11}$$



(a) $\phi = 0$ (white noise)

(b) $\phi = 0.5$

(c) $\phi = 0.9$

FIGURE 3.3 Realizations of an $AR(1)$ process, $Y_t = \phi Y_{t-1} + \varepsilon_t$, for alternative values of ϕ.

Recall from [3.4.2] that $(Y_{t-1} - \mu)$ is a linear function of $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$:

$$(Y_{t-1} - \mu) = \varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + \cdots.$$

But ε_t is uncorrelated with $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$, so ε_t must be uncorrelated with $(Y_{t-1} - \mu)$. Thus the middle term on the right side of [3.4.11] is zero:

$$E[(Y_{t-1} - \mu)\varepsilon_t] = 0. \qquad [3.4.12]$$

Again, assuming covariance-stationarity, we have

$$E(Y_t - \mu)^2 = E(Y_{t-1} - \mu)^2 = \gamma_0. \qquad [3.4.13]$$

Substituting [3.4.13] and [3.4.12] into [3.4.11],

$$\gamma_0 = \phi^2\gamma_0 + 0 + \sigma^2$$

or

$$\gamma_0 = \sigma^2/(1 - \phi^2),$$

reproducing [3.4.4].

Similarly, we could multiply [3.4.10] by $(Y_{t-j} - \mu)$ and take expectations:

$$\begin{aligned} &E[(Y_t - \mu)(Y_{t-j} - \mu)] \\ &\qquad = \phi \cdot E[(Y_{t-1} - \mu)(Y_{t-j} - \mu)] + E[\varepsilon_t(Y_{t-j} - \mu)]. \end{aligned} \qquad [3.4.14]$$

But the term $(Y_{t-j} - \mu)$ will be a linear function of $\varepsilon_{t-j}, \varepsilon_{t-j-1}, \varepsilon_{t-j-2}, \ldots$, which, for $j > 0$, will be uncorrelated with ε_t. Thus, for $j > 0$, the last term on the right side in [3.4.14] is zero. Notice, moreover, that the expression appearing in the first term on the right side of [3.4.14],

$$E[(Y_{t-1} - \mu)(Y_{t-j} - \mu)],$$

is the autocovariance of observations on Y separated by $j - 1$ periods:

$$E[(Y_{t-1} - \mu)(Y_{[t-1]-[j-1]} - \mu)] = \gamma_{j-1}.$$

Thus, for $j > 0$, [3.4.14] becomes

$$\gamma_j = \phi\gamma_{j-1}. \qquad [3.4.15]$$

Equation [3.4.15] takes the form of a first-order difference equation,

$$y_t = \phi y_{t-1} + w_t,$$

in which the autocovariance γ takes the place of the variable y and in which the subscript j (which indexes the order of the autocovariance) replaces t (which indexes time). The input w_t in [3.4.15] is identically equal to zero. It is easy to see that the difference equation [3.4.15] has the solution

$$\gamma_j = \phi^j\gamma_0,$$

which reproduces [3.4.6]. We now see why the impulse-response function and autocorrelation function for an $AR(1)$ process coincide—they both represent the solution to a first-order difference equation with autoregressive parameter ϕ, an initial value of unity, and no subsequent shocks.

The Second-Order Autoregressive Process

A *second-order autoregression*, denoted $AR(2)$, satisfies

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, \qquad [3.4.16]$$

or, in lag operator notation,

$$(1 - \phi_1 L - \phi_2 L^2)Y_t = c + \varepsilon_t. \qquad [3.4.17]$$

The difference equation [3.4.16] is stable provided that the roots of

$$(1 - \phi_1 z - \phi_2 z^2) = 0 \qquad [3.4.18]$$

lie outside the unit circle. When this condition is satisfied, the $AR(2)$ process turns out to be covariance-stationary, and the inverse of the autoregressive operator in [3.4.17] is given by

$$\psi(L) = (1 - \phi_1 L - \phi_2 L^2)^{-1} = \psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \cdots. \qquad [3.4.19]$$

Recalling [1.2.44], the value of ψ_j can be found from the (1, 1) element of the matrix $\mathbf{F}$ raised to the jth power, as in expression [1.2.28]. Where the roots of [3.4.18] are distinct, a closed-form expression for ψ_j is given by [1.2.29] and [1.2.25]. Exercise 3.3 at the end of this chapter discusses alternative algorithms for calculating ψ_j.

Multiplying both sides of [3.4.17] by $\psi(L)$ gives

$$Y_t = \psi(L)c + \psi(L)\varepsilon_t. \qquad [3.4.20]$$

It is straightforward to show that

$$\psi(L)c = c/(1 - \phi_1 - \phi_2) \qquad [3.4.21]$$

and

$$\sum_{j=0}^{\infty} |\psi_j| < \infty; \qquad [3.4.22]$$

the reader is invited to prove these claims in Exercises 3.4 and 3.5. Since [3.4.20] is an absolutely summable $MA(\infty)$ process, its mean is given by the constant term:

$$\mu = c/(1 - \phi_1 - \phi_2). \qquad [3.4.23]$$

An alternative method for calculating the mean is to assume that the process is covariance-stationary and take expectations of [3.4.16] directly:

$$E(Y_t) = c + \phi_1 E(Y_{t-1}) + \phi_2 E(Y_{t-2}) + E(\varepsilon_t),$$

implying

$$\mu = c + \phi_1 \mu + \phi_2 \mu + 0,$$

reproducing [3.4.23].

To find second moments, write [3.4.16] as

$$Y_t = \mu \cdot (1 - \phi_1 - \phi_2) + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

or

$$(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \varepsilon_t. \qquad [3.4.24]$$

Multiplying both sides of [3.4.24] by $(Y_{t-j} - \mu)$ and taking expectations produces

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} \qquad \text{for } j = 1, 2, \ldots. \qquad [3.4.25]$$

Thus, the autocovariances follow the same second-order difference equation as does the process for Y_t, with the difference equation for γ_j indexed by the lag j. The autocovariances therefore behave just as the solutions to the second-order difference equation analyzed in Section 1.2. An $AR(2)$ process is covariance-stationary provided that ϕ_1 and ϕ_2 lie within the triangular region of Figure 1.5.

When ϕ_1 and ϕ_2 lie within the triangular region but above the parabola in that figure, the autocovariance function γ_j is the sum of two decaying exponential functions of j. When ϕ_1 and ϕ_2 fall within the triangular region but below the parabola, γ_j is a damped sinusoidal function.

The autocorrelations are found by dividing both sides of [3.4.25] by γ_0:

$$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} \qquad \text{for } j = 1, 2, \ldots. \tag{3.4.26}$$

In particular, setting $j = 1$ produces

$$\rho_1 = \phi_1 + \phi_2\rho_1$$

or

$$\rho_1 = \phi_1/(1 - \phi_2). \tag{3.4.27}$$

For $j = 2$,

$$\rho_2 = \phi_1\rho_1 + \phi_2. \tag{3.4.28}$$

The variance of a covariance-stationary second-order autoregression can be found by multiplying both sides of [3.4.24] by $(Y_t - \mu)$ and taking expectations:

$$\begin{aligned} E(Y_t - \mu)^2 = {} & \phi_1 \cdot E(Y_{t-1} - \mu)(Y_t - \mu) + \phi_2 \cdot E(Y_{t-2} - \mu)(Y_t - \mu) \\ & + E(\varepsilon_t)(Y_t - \mu), \end{aligned}$$

or

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma^2. \tag{3.4.29}$$

The last term (σ^2) in [3.4.29] comes from noticing that

$$\begin{aligned} E(\varepsilon_t)(Y_t - \mu) &= E(\varepsilon_t)[\phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \varepsilon_t] \\ &= \phi_1 \cdot 0 + \phi_2 \cdot 0 + \sigma^2. \end{aligned}$$

Equation [3.4.29] can be written

$$\gamma_0 = \phi_1\rho_1\gamma_0 + \phi_2\rho_2\gamma_0 + \sigma^2. \tag{3.4.30}$$

Substituting [3.4.27] and [3.4.28] into [3.4.30] gives

$$\gamma_0 = \left[\frac{\phi_1^2}{(1 - \phi_2)} + \frac{\phi_2\phi_1^2}{(1 - \phi_2)} + \phi_2^2\right]\gamma_0 + \sigma^2$$

or

$$\gamma_0 = \frac{(1 - \phi_2)\sigma^2}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]}.$$

The pth-Order Autoregressive Process

A *pth-order autoregression*, denoted *AR*(p), satisfies

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t. \tag{3.4.31}$$

Provided that the roots of

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0 \tag{3.4.32}$$

all lie outside the unit circle, it is straightforward to verify that a covariance-stationary representation of the form

$$Y_t = \mu + \psi(L)\varepsilon_t \tag{3.4.33}$$

exists where

$$\psi(L) = (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)^{-1}$$

and $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Assuming that the stationarity condition is satisfied, one way to find the mean is to take expectations of [3.4.31]:

$$\mu = c + \phi_1\mu + \phi_2\mu + \cdots + \phi_p\mu,$$

or

$$\mu = c/(1 - \phi_1 - \phi_2 - \cdots - \phi_p). \qquad [3.4.34]$$

Using [3.4.34], equation [3.4.31] can be written

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \varepsilon_t. \qquad [3.4.35]$$

Autocovariances are found by multiplying both sides of [3.4.35] by $(Y_{t-j} - \mu)$ and taking expectations:

$$\gamma_j = \begin{cases} \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \cdots + \phi_p\gamma_{j-p} & \text{for } j = 1, 2, \ldots \\ \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma^2 & \text{for } j = 0. \end{cases} \qquad [3.4.36]$$

Using the fact that $\gamma_{-j} = \gamma_j$, the system of equations in [3.4.36] for $j = 0, 1, \ldots, p$ can be solved for $\gamma_0, \gamma_1, \ldots, \gamma_p$ as functions of $\sigma^2, \phi_1, \phi_2, \ldots, \phi_p$. It can be shown[4] that the $(p \times 1)$ vector $(\gamma_0, \gamma_1, \ldots, \gamma_{p-1})'$ is given by the first p elements of the first column of the $(p^2 \times p^2)$ matrix $\sigma^2[\mathbf{I}_{p^2} - (\mathbf{F} \otimes \mathbf{F})]^{-1}$ where $\mathbf{F}$ is the $(p \times p)$ matrix defined in equation [1.2.3] and $\otimes$ indicates the Kronecker product.

Dividing [3.4.36] by γ_0 produces the *Yule-Walker equations*:

$$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \cdots + \phi_p\rho_{j-p} \qquad \text{for } j = 1, 2, \ldots. \qquad [3.4.37]$$

Thus, the autocovariances and autocorrelations follow the same pth-order difference equation as does the process itself [3.4.31]. For distinct roots, their solutions take the form

$$\gamma_j = g_1\lambda_1^j + g_2\lambda_2^j + \cdots + g_p\lambda_p^j, \qquad [3.4.38]$$

where the eigenvalues $(\lambda_1, \ldots, \lambda_p)$ are the solutions to

$$\lambda^p - \phi_1\lambda^{p-1} - \phi_2\lambda^{p-2} - \cdots - \phi_p = 0.$$

3.5. *Mixed Autoregressive Moving Average Processes*

An *ARMA*(p, q) *process* includes both autoregressive and moving average terms:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}, \qquad [3.5.1]$$

or, in lag operator form,

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)Y_t = c + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t. \qquad [3.5.2]$$

Provided that the roots of

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0 \qquad [3.5.3]$$

[4] The reader will be invited to prove this in Exercise 10.1 in Chapter 10.

lie outside the unit circle, both sides of [3.5.2] can be divided by $(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)$ to obtain

$$Y_t = \mu + \psi(L)\varepsilon_t$$

where

$$\psi(L) = \frac{(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)}{(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)}$$

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

$$\mu = c/(1 - \phi_1 - \phi_2 - \cdots - \phi_p).$$

Thus, stationarity of an *ARMA* process depends entirely on the autoregressive parameters $(\phi_1, \phi_2, \ldots, \phi_p)$ and not on the moving average parameters $(\theta_1, \theta_2, \ldots, \theta_q)$.

It is often convenient to write the *ARMA* process [3.5.1] in terms of deviations from the mean:

$$\begin{aligned} Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \cdots \\ + \phi_p(Y_{t-p} - \mu) + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}. \end{aligned} \quad [3.5.4]$$

Autocovariances are found by multiplying both sides of [3.5.4] by $(Y_{t-j} - \mu)$ and taking expectations. For $j > q$, the resulting equations take the form

$$\gamma_j = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \cdots + \phi_p\gamma_{j-p} \qquad \text{for } j = q+1, q+2, \ldots. \quad [3.5.5]$$

Thus, after q lags the autocovariance function γ_j (and the autocorrelation function ρ_j) follow the pth-order difference equation governed by the autoregressive parameters.

Note that [3.5.5] does not hold for $j \leq q$, owing to correlation between $\theta_j\varepsilon_{t-j}$ and Y_{t-j}. Hence, an $ARMA(p, q)$ process will have more complicated autocovariances for lags 1 through q than would the corresponding $AR(p)$ process. For $j > q$ with distinct autoregressive roots, the autocovariances will be given by

$$\gamma_j = h_1\lambda_1^j + h_2\lambda_2^j + \cdots + h_p\lambda_p^j. \quad [3.5.6]$$

This takes the same form as the autocovariances for an $AR(p)$ process [3.4.38], though because the initial conditions $(\gamma_0, \gamma_1, \ldots, \gamma_q)$ differ for the *ARMA* and *AR* processes, the parameters h_k in [3.5.6] will not be the same as the parameters g_k in [3.4.38].

There is a potential for redundant parameterization with *ARMA* processes. Consider, for example, a simple white noise process,

$$Y_t = \varepsilon_t. \quad [3.5.7]$$

Suppose both sides of [3.5.7] are multiplied by $(1 - \rho L)$:

$$(1 - \rho L)Y_t = (1 - \rho L)\varepsilon_t. \quad [3.5.8]$$

Clearly, if [3.5.7] is a valid representation, then so is [3.5.8] for any value of ρ. Thus, [3.5.8] might be described as an $ARMA(1, 1)$ process, with $\phi_1 = \rho$ and $\theta_1 = -\rho$. It is important to avoid such a parameterization. Since any value of ρ in [3.5.8] describes the data equally well, we will obviously get into trouble trying to estimate the parameter ρ in [3.5.8] by maximum likelihood. Moreover, theoretical manipulations based on a representation such as [3.5.8] may overlook key cancellations. If we are using an $ARMA(1, 1)$ model in which θ_1 is close to $-\phi_1$, then the data might better be modeled as simple white noise.

A related overparameterization can arise with an $ARMA(p, q)$ model. Consider factoring the lag polynomial operators in [3.5.2] as in [2.4.3]:

$$(1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L)(Y_t - \mu) \\ = (1 - \eta_1 L)(1 - \eta_2 L) \cdots (1 - \eta_q L)\varepsilon_t. \qquad [3.5.9]$$

We assume that $|\lambda_i| < 1$ for all i, so that the process is covariance-stationary. If the autoregressive operator $(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)$ and the moving average operator $(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)$ have any roots in common, say, $\lambda_i = \eta_j$ for some i and j, then both sides of [3.5.9] can be divided by $(1 - \lambda_i L)$:

$$\prod_{\substack{k=1 \\ k \neq i}}^{p} (1 - \lambda_k L)(Y_t - \mu) = \prod_{\substack{k=1 \\ k \neq j}}^{q} (1 - \eta_k L)\varepsilon_t,$$

or

$$(1 - \phi_1^* L - \phi_2^* L^2 - \cdots - \phi_{p-1}^* L^{p-1})(Y_t - \mu) \\ = (1 + \theta_1^* L + \theta_2^* L^2 + \cdots + \theta_{q-1}^* L^{q-1})\varepsilon_t, \qquad [3.5.10]$$

where

$$(1 - \phi_1^* L - \phi_2^* L^2 - \cdots - \phi_{p-1}^* L^{p-1}) \\ \equiv (1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_{i-1} L)(1 - \lambda_{i+1} L) \cdots (1 - \lambda_p L)$$

$$(1 + \theta_1^* L + \theta_2^* L^2 + \cdots + \theta_{q-1}^* L^{q-1}) \\ \equiv (1 - \eta_1 L)(1 - \eta_2 L) \cdots (1 - \eta_{j-1} L)(1 - \eta_{j+1} L) \cdots (1 - \eta_q L).$$

The stationary $ARMA(p, q)$ process satisfying [3.5.2] is clearly identical to the stationary $ARMA(p - 1, q - 1)$ process satisfying [3.5.10].

3.6. *The Autocovariance-Generating Function*

For each of the covariance-stationary processes for Y_t considered so far, we calculated the sequence of autocovariances $\{\gamma_j\}_{j=-\infty}^{\infty}$. If this sequence is absolutely summable, then one way of summarizing the autocovariances is through a scalar-valued function called the *autocovariance-generating function*:

$$g_Y(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j. \qquad [3.6.1]$$

This function is constructed by taking the jth autocovariance and multiplying it by some number z raised to the jth power, and then summing over all the possible values of j. The argument of this function (z) is taken to be a complex scalar.

Of particular interest as an argument for the autocovariance-generating function is any value of z that lies on the complex unit circle,

$$z = \cos(\omega) - i\sin(\omega) = e^{-i\omega},$$

where $i = \sqrt{-1}$ and ω is the radian angle that z makes with the real axis. If the autocovariance-generating function is evaluated at $z = e^{-i\omega}$ and divided by 2π, the resulting function of ω,

$$s_Y(\omega) = \frac{1}{2\pi} g_Y(e^{-i\omega}) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\omega j},$$

is called the *population spectrum* of Y. The population spectrum will be discussed in detail in Chapter 6. There it will be shown that for a process with absolutely summable autocovariances, the function $s_Y(\omega)$ exists and can be used to calculate all of the autocovariances. This means that if two different processes share the same autocovariance-generating function, then the two processes exhibit the identical sequence of autocovariances.

As an example of calculating an autocovariance-generating function, consider the $MA(1)$ process. From equations [3.3.3] to [3.3.5], its autocovariance-generating function is

$$g_Y(z) = [\theta\sigma^2]z^{-1} + [(1 + \theta^2)\sigma^2]z^0 + [\theta\sigma^2]z^1 = \sigma^2\cdot[\theta z^{-1} + (1 + \theta^2) + \theta z].$$

Notice that this expression could alternatively be written

$$g_Y(z) = \sigma^2(1 + \theta z)(1 + \theta z^{-1}). \qquad [3.6.2]$$

The form of expression [3.6.2] suggests that for the $MA(q)$ process,

$$Y_t = \mu + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t,$$

the autocovariance-generating function might be calculated as

$$\begin{aligned} g_Y(z) = \sigma^2&(1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q) \\ &\times (1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \cdots + \theta_q z^{-q}). \end{aligned} \qquad [3.6.3]$$

This conjecture can be verified by carrying out the multiplication in [3.6.3] and collecting terms by powers of z:

$$\begin{aligned} &(1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q) \times (1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \cdots + \theta_q z^{-q}) \\ &\quad = (\theta_q)z^q + (\theta_{q-1} + \theta_q\theta_1)z^{(q-1)} + (\theta_{q-2} + \theta_{q-1}\theta_1 + \theta_q\theta_2)z^{(q-2)} \\ &\qquad + \cdots + (\theta_1 + \theta_2\theta_1 + \theta_3\theta_2 + \cdots + \theta_q\theta_{q-1})z^1 \\ &\qquad + (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)z^0 \\ &\qquad + (\theta_1 + \theta_2\theta_1 + \theta_3\theta_2 + \cdots + \theta_q\theta_{q-1})z^{-1} + \cdots + (\theta_q)z^{-q}. \end{aligned} \qquad [3.6.4]$$

Comparison of [3.6.4] with [3.3.10] or [3.3.12] confirms that the coefficient on z^j in [3.6.3] is indeed the jth autocovariance.

This method for finding $g_Y(z)$ extends to the $MA(\infty)$ case. If

$$Y_t = \mu + \psi(L)\varepsilon_t \qquad [3.6.5]$$

with

$$\psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots \qquad [3.6.6]$$

and

$$\sum_{j=0}^{\infty} |\psi_j| < \infty, \qquad [3.6.7]$$

then

$$g_Y(z) = \sigma^2\psi(z)\psi(z^{-1}). \qquad [3.6.8]$$

For example, the stationary $AR(1)$ process can be written as

$$Y_t - \mu = (1 - \phi L)^{-1}\varepsilon_t,$$

which is in the form of [3.6.5] with $\psi(L) = 1/(1 - \phi L)$. The autocovariance-generating function for an $AR(1)$ process could therefore be calculated from

$$g_Y(z) = \frac{\sigma^2}{(1 - \phi z)(1 - \phi z^{-1})}. \qquad [3.6.9]$$

To verify this claim directly, expand out the terms in [3.6.9]:

$$\frac{\sigma^2}{(1 - \phi z)(1 - \phi z^{-1})} = \sigma^2(1 + \phi z + \phi^2 z^2 + \phi^3 z^3 + \cdots) \times (1 + \phi z^{-1} + \phi^2 z^{-2} + \phi^3 z^{-3} + \cdots),$$

from which the coefficient on z^j is

$$\sigma^2(\phi^j + \phi^{j+1}\phi + \phi^{j+2}\phi^2 + \cdots) = \sigma^2\phi^j/(1 - \phi^2).$$

This indeed yields the jth autocovariance as earlier calculated in equation [3.4.5].

The autocovariance-generating function for a stationary $ARMA(p, q)$ process can be written

$$g_Y(z) = \frac{\sigma^2(1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q)(1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \cdots + \theta_q z^{-q})}{(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p)(1 - \phi_1 z^{-1} - \phi_2 z^{-2} - \cdots - \phi_p z^{-p})}. \tag{3.6.10}$$

Filters

Sometimes the data are *filtered*, or treated in a particular way before they are analyzed, and we would like to summarize the effects of this treatment on the autocovariances. This calculation is particularly simple using the autocovariance-generating function. For example, suppose that the original data Y_t were generated from an $MA(1)$ process,

$$Y_t = (1 + \theta L)\varepsilon_t, \tag{3.6.11}$$

with autocovariance-generating function given by [3.6.2]. Let's say that the data as actually analyzed, X_t, represent the change in Y_t over its value the previous period:

$$X_t = Y_t - Y_{t-1} = (1 - L)Y_t. \tag{3.6.12}$$

Substituting [3.6.11] into [3.6.12], the observed data can be characterized as the following $MA(2)$ process,

$$X_t = (1 - L)(1 + \theta L)\varepsilon_t = [1 + (\theta - 1)L - \theta L^2]\varepsilon_t \equiv [1 + \theta_1 L + \theta_2 L^2]\varepsilon_t, \tag{3.6.13}$$

with $\theta_1 \equiv (\theta - 1)$ and $\theta_2 \equiv -\theta$. The autocovariance-generating function of the observed data X_t can be calculated by direct application of [3.6.3]:

$$g_X(z) = \sigma^2(1 + \theta_1 z + \theta_2 z^2)(1 + \theta_1 z^{-1} + \theta_2 z^{-2}). \tag{3.6.14}$$

It is often instructive, however, to keep the polynomial $(1 + \theta_1 z + \theta_2 z^2)$ in its factored form of the first line of [3.6.13],

$$(1 + \theta_1 z + \theta_2 z^2) = (1 - z)(1 + \theta z),$$

in which case [3.6.14] could be written

$$\begin{aligned} g_X(z) &= \sigma^2(1 - z)(1 + \theta z)(1 - z^{-1})(1 + \theta z^{-1}) \\ &= (1 - z)(1 - z^{-1}) \cdot g_Y(z). \end{aligned} \tag{3.6.15}$$

Of course, [3.6.14] and [3.6.15] represent the identical function of z, and which way we choose to write it is simply a matter of convenience. Applying the filter

$(1 - L)$ to Y_t thus results in multiplying its autocovariance-generating function by $(1 - z)(1 - z^{-1})$.

This principle readily generalizes. Suppose that the original data series $\{Y_t\}$ satisfies [3.6.5] through [3.6.7]. Let's say the data are filtered according to

$$X_t = h(L)Y_t \tag{3.6.16}$$

with

$$h(L) = \sum_{j=-\infty}^{\infty} h_j L^j$$

$$\sum_{j=-\infty}^{\infty} |h_j| < \infty.$$

Substituting [3.6.5] into [3.6.16], the observed data X_t are then generated by

$$X_t = h(1)\mu + h(L)\psi(L)\varepsilon_t \equiv \mu^* + \psi^*(L)\varepsilon_t,$$

where $\mu^* \equiv h(1)\mu$ and $\psi^*(L) \equiv h(L)\psi(L)$. The sequence of coefficients associated with the compound operator $\{\psi_j^*\}_{j=-\infty}^{\infty}$ turns out to be absolutely summable,[5] and the autocovariance-generating function of X_t can accordingly be calculated as

$$g_X(z) = \sigma^2\psi^*(z)\psi^*(z^{-1}) = \sigma^2 h(z)\psi(z)\psi(z^{-1})h(z^{-1}) = h(z)h(z^{-1})g_Y(z). \tag{3.6.17}$$

Applying the filter $h(L)$ to a series thus results in multiplying its autocovariance-generating function by $h(z)h(z^{-1})$.

3.7. *Invertibility*

Invertibility for the MA*(1) Process*

Consider an $MA(1)$ process,

$$Y_t - \mu = (1 + \theta L)\varepsilon_t, \tag{3.7.1}$$

with

$$E(\varepsilon_t\varepsilon_\tau) = \begin{cases} \sigma^2 & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases}$$

[5]Specifically,

$$\begin{aligned}\psi^*(z) &= \left(\sum_{j=-\infty}^{\infty} h_j z^j\right)\left(\sum_{k=0}^{\infty} \psi_k z^k\right) \\ &= (\cdots + h_{-j}z^{-j} + h_{-j+1}z^{-j+1} + \cdots + h_{-1}z^{-1} + h_0 z^0 + h_1 z^1 + \cdots \\ &\quad + h_j z^j + h_{j+1}z^{j+1} + \cdots)(\psi_0 z^0 + \psi_1 z^1 + \psi_2 z^2 + \cdots),\end{aligned}$$

from which the coefficient on z^j is

$$\psi_j^* = h_j\psi_0 + h_{j-1}\psi_1 + h_{j-2}\psi_2 + \cdots = \sum_{v=0}^{\infty} h_{j-v}\psi_v.$$

Then

$$\sum_{j=-\infty}^{\infty} |\psi_j^*| = \sum_{j=-\infty}^{\infty} \left|\sum_{v=0}^{\infty} h_{j-v}\psi_v\right| \le \sum_{j=-\infty}^{\infty}\sum_{v=0}^{\infty} |h_{j-v}\psi_v| = \sum_{v=0}^{\infty} |\psi_v| \sum_{j=-\infty}^{\infty} |h_{j-v}| = \sum_{v=0}^{\infty} |\psi_v| \sum_{j=-\infty}^{\infty} |h_j| < \infty.$$

Provided that $|\theta| < 1$, both sides of [3.7.1] can be multiplied by $(1 + \theta L)^{-1}$ to obtain[6]

$$(1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \cdots)(Y_t - \mu) = \varepsilon_t, \qquad [3.7.2]$$

which could be viewed as an $AR(\infty)$ representation. If a moving average representation such as [3.7.1] can be rewritten as an $AR(\infty)$ representation such as [3.7.2] simply by inverting the moving average operator $(1 + \theta L)$, then the moving average representation is said to be *invertible*. For an $MA(1)$ process, invertibility requires $|\theta| < 1$; if $|\theta| \geq 1$, then the infinite sequence in [3.7.2] would not be well defined.

Let us investigate what invertibility means in terms of the first and second moments of the process. Recall that the $MA(1)$ process [3.7.1] has mean μ and autocovariance-generating function

$$g_Y(z) = \sigma^2(1 + \theta z)(1 + \theta z^{-1}). \qquad [3.7.3]$$

Now consider a seemingly different $MA(1)$ process,

$$\tilde{Y}_t - \mu = (1 + \tilde{\theta} L)\tilde{\varepsilon}_t, \qquad [3.7.4]$$

with

$$E(\tilde{\varepsilon}_t \tilde{\varepsilon}_\tau) = \begin{cases} \tilde{\sigma}^2 & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\tilde{Y}_t$ has the same mean (μ) as Y_t. Its autocovariance-generating function is

$$\begin{aligned} g_{\tilde{Y}}(z) &= \tilde{\sigma}^2(1 + \tilde{\theta} z)(1 + \tilde{\theta} z^{-1}) \\ &= \tilde{\sigma}^2\{(\tilde{\theta}^{-1} z^{-1} + 1)(\tilde{\theta} z)\}\{(\tilde{\theta}^{-1} z + 1)(\tilde{\theta} z^{-1})\} \\ &= (\tilde{\sigma}^2 \tilde{\theta}^2)(1 + \tilde{\theta}^{-1} z)(1 + \tilde{\theta}^{-1} z^{-1}). \end{aligned} \qquad [3.7.5]$$

Suppose that the parameters of [3.7.4], $(\tilde{\theta}, \tilde{\sigma}^2)$, are related to those of [3.7.1] by the following equations:

$$\theta = \tilde{\theta}^{-1} \qquad [3.7.6]$$

$$\sigma^2 = \tilde{\theta}^2 \tilde{\sigma}^2. \qquad [3.7.7]$$

Then the autocovariance-generating functions [3.7.3] and [3.7.5] would be the same, meaning that Y_t and $\tilde{Y}_t$ would have identical first and second moments.

Notice from [3.7.6] that if $|\theta| < 1$, then $|\tilde{\theta}| > 1$. In other words, for any invertible $MA(1)$ representation [3.7.1], we have found a noninvertible $MA(1)$ representation [3.7.4] with the same first and second moments as the invertible representation. Conversely, given any noninvertible representation with $|\tilde{\theta}| > 1$, there exists an invertible representation with $\theta = (1/\tilde{\theta})$ that has the same first and second moments as the noninvertible representation. In the borderline case where $\theta = \pm 1$, there is only one representation of the process, and it is noninvertible.

Not only do the invertible and noninvertible representations share the same moments, either representation [3.7.1] or [3.7.4] could be used as an equally valid description of any given $MA(1)$ process! Suppose a computer generated an infinite sequence of $\tilde{Y}$'s according to [3.7.4] with $\tilde{\theta} > 1$. Thus we know for a fact that the data were generated from an $MA(1)$ process expressed in terms of a noninvertible representation. In what sense could these same data be associated with an invertible $MA(1)$ representation?

[6]Note from [2.2.8] that

$$(1 + \theta L)^{-1} = [1 - (-\theta)L]^{-1} = 1 + (-\theta)L + (-\theta)^2 L^2 + (-\theta)^3 L^3 + \cdots.$$

Imagine calculating a series $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ defined by

$$\begin{aligned}\varepsilon_t &\equiv (1 + \theta L)^{-1}(\tilde{Y}_t - \mu) \\ &= (\tilde{Y}_t - \mu) - \theta(\tilde{Y}_{t-1} - \mu) + \theta^2(\tilde{Y}_{t-2} - \mu) - \theta^3(\tilde{Y}_{t-3} - \mu) + \cdots,\end{aligned} \qquad [3.7.8]$$

where $\theta = (1/\tilde{\theta})$ is the moving average parameter associated with the invertible $MA(1)$ representation that shares the same moments as [3.7.4]. Note that since $|\theta| < 1$, this produces a well-defined, mean square convergent series $\{\varepsilon_t\}$.

Furthermore, the sequence $\{\varepsilon_t\}$ so generated is white noise. The simplest way to verify this is to calculate the autocovariance-generating function of ε_t and confirm that the coefficient on z^j (the jth autocovariance) is equal to zero for any $j \neq 0$. From [3.7.8] and [3.6.17], the autocovariance-generating function for ε_t is given by

$$g_\varepsilon(z) = (1 + \theta z)^{-1}(1 + \theta z^{-1})^{-1} g_{\tilde{Y}}(z). \qquad [3.7.9]$$

Substituting [3.7.5] into [3.7.9],

$$\begin{aligned}g_\varepsilon(z) &= (1 + \theta z)^{-1}(1 + \theta z^{-1})^{-1}(\tilde{\sigma}^2\tilde{\theta}^2)(1 + \tilde{\theta}^{-1}z)(1 + \tilde{\theta}^{-1}z^{-1}) \\ &= \tilde{\sigma}^2\tilde{\theta}^2,\end{aligned} \qquad [3.7.10]$$

where the last equality follows from the fact that $\tilde{\theta}^{-1} = \theta$. Since the autocovariance-generating function is a constant, it follows that ε_t is a white noise process with variance $\tilde{\theta}^2\tilde{\sigma}^2$.

Multiplying both sides of [3.7.8] by $(1 + \theta L)$,

$$\tilde{Y}_t - \mu = (1 + \theta L)\varepsilon_t$$

is a perfectly valid invertible $MA(1)$ representation of data that were actually generated from the noninvertible representation [3.7.4].

The converse proposition is also true—suppose that the data were really generated from [3.7.1] with $|\theta| < 1$, an invertible representation. Then there exists a noninvertible representation with $\tilde{\theta} = 1/\theta$ that describes these data with equal validity. To characterize this noninvertible representation, consider the operator proposed in [2.5.20] as the appropriate inverse of $(1 + \tilde{\theta}L)$:

$$\begin{aligned}&(\tilde{\theta})^{-1}L^{-1}[1 - (\tilde{\theta}^{-1})L^{-1} + (\tilde{\theta}^{-2})L^{-2} - (\tilde{\theta}^{-3})L^{-3} + \cdots] \\ &\quad = \theta L^{-1}[1 - \theta L^{-1} + \theta^2 L^{-2} - \theta^3 L^{-3} + \cdots].\end{aligned}$$

Define $\tilde{\varepsilon}_t$ to be the series that results from applying this operator to $(Y_t - \mu)$,

$$\tilde{\varepsilon}_t \equiv \theta(Y_{t+1} - \mu) - \theta^2(Y_{t+2} - \mu) + \theta^3(Y_{t+3} - \mu) - \cdots, \qquad [3.7.11]$$

noting that this series converges for $|\theta| < 1$. Again this series is white noise:

$$\begin{aligned}g_{\tilde{\varepsilon}}(z) &= \{\theta z^{-1}[1 - \theta z^{-1} + \theta^2 z^{-2} - \theta^3 z^{-3} + \cdots]\} \\ &\quad \times \{\theta z[1 - \theta z^1 + \theta^2 z^2 - \theta^3 z^3 + \cdots]\}\sigma^2(1 + \theta z)(1 + \theta z^{-1}) \\ &= \theta^2\sigma^2.\end{aligned}$$

The coefficient on z^j is zero for $j \neq 0$, so $\tilde{\varepsilon}_t$ is white noise as claimed. Furthermore, by construction,

$$Y_t - \mu = (1 + \tilde{\theta}L)\tilde{\varepsilon}_t,$$

so that we have found a noninvertible $MA(1)$ representation of data that were actually generated by the invertible $MA(1)$ representation [3.7.1].

Either the invertible or the noninvertible representation could characterize any given data equally well, though there is a practical reason for preferring the invertible representation. To find the value of ε for date t associated with the invertible representation as in [3.7.8], we need to know current and past values of Y. By contrast, to find the value of $\tilde{\varepsilon}$ for date t associated with the noninvertible representation as in [3.7.11], we need to use all of the future values of Y! If the intention is to calculate the current value of ε_t using real-world data, it will be feasible only to work with the invertible representation. Also, as will be noted in Chapters 4 and 5, some convenient algorithms for estimating parameters and forecasting are valid only if the invertible representation is used.

The value of ε_t associated with the invertible representation is sometimes called the *fundamental innovation* for Y_t. For the borderline case when $|\theta| = 1$, the process is noninvertible, but the innovation ε_t for such a process will still be described as the fundamental innovation for Y_t.

Invertibility for the MA(q) *Process*

Consider now the $MA(q)$ process,

$$(Y_t - \mu) = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t \qquad [3.7.12]$$

$$E(\varepsilon_t \varepsilon_\tau) = \begin{cases} \sigma^2 & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases}$$

Provided that the roots of

$$(1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q) = 0 \qquad [3.7.13]$$

lie outside the unit circle, [3.7.12] can be written as an $AR(\infty)$ simply by inverting the MA operator,

$$(1 + \eta_1 L + \eta_2 L^2 + \eta_3 L^3 + \cdots)(Y_t - \mu) = \varepsilon_t,$$

where

$$(1 + \eta_1 L + \eta_2 L^2 + \eta_3 L^3 + \cdots) = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)^{-1}.$$

Where this is the case, the $MA(q)$ representation [3.7.12] is invertible.

Factor the moving average operator as

$$(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q) = (1 - \lambda_1 L)(1 - \lambda_2 L)\cdots(1 - \lambda_q L). \qquad [3.7.14]$$

If $|\lambda_i| < 1$ for all i, then the roots of [3.7.13] are all outside the unit circle and the representation [3.7.12] is invertible. If instead some of the λ_i are outside (but not on) the unit circle, Hansen and Sargent (1981, p. 102) suggested the following procedure for finding an invertible representation. The autocovariance-generating function of Y_t can be written

$$\begin{aligned} g_Y(z) = \sigma^2 \cdot \{&(1 - \lambda_1 z)(1 - \lambda_2 z)\cdots(1 - \lambda_q z)\} \\ \times \{&(1 - \lambda_1 z^{-1})(1 - \lambda_2 z^{-1})\cdots(1 - \lambda_q z^{-1})\}. \end{aligned} \qquad [3.7.15]$$

Order the λ's so that $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ are inside the unit circle and $(\lambda_{n+1}, \lambda_{n+2}, \ldots, \lambda_q)$ are outside the unit circle. Suppose σ^2 in [3.7.15] is replaced by $\sigma^2 \cdot \lambda_{n+1}^2 \cdot \lambda_{n+2}^2 \cdots \lambda_q^2$; since complex λ_i appear as conjugate pairs, this is a positive real number. Suppose further that $(\lambda_{n+1}, \lambda_{n+2}, \ldots, \lambda_q)$ are replaced with their reciprocals, $(\lambda_{n+1}^{-1}, \lambda_{n+2}^{-1}, \ldots, \lambda_q^{-1})$. The resulting function would be

$$\begin{aligned}
&\sigma^2\lambda_{n+1}^2\lambda_{n+2}^2 \cdots \lambda_q^2\left\{\prod_{i=1}^{n}(1-\lambda_i z)\right\}\left\{\prod_{i=n+1}^{q}(1-\lambda_i^{-1}z)\right\}\\
&\quad\times\left\{\prod_{i=1}^{n}(1-\lambda_i z^{-1})\right\}\left\{\prod_{i=n+1}^{q}(1-\lambda_i^{-1}z^{-1})\right\}\\
&=\sigma^2\left\{\prod_{i=1}^{n}(1-\lambda_i z)\right\}\left\{\prod_{i=n+1}^{q}[(\lambda_i z^{-1})(1-\lambda_i^{-1}z)]\right\}\\
&\quad\times\left\{\prod_{i=1}^{n}(1-\lambda_i z^{-1})\right\}\left\{\prod_{i=n+1}^{q}[(\lambda_i z)(1-\lambda_i^{-1}z^{-1})]\right\}\\
&=\sigma^2\left\{\prod_{i=1}^{n}(1-\lambda_i z)\right\}\left\{\prod_{i=n+1}^{q}(\lambda_i z^{-1}-1)\right\}\\
&\quad\times\left\{\prod_{i=1}^{n}(1-\lambda_i z^{-1})\right\}\left\{\prod_{i=n+1}^{q}(\lambda_i z-1)\right\}\\
&=\sigma^2\left\{\prod_{i=1}^{q}(1-\lambda_i z)\right\}\left\{\prod_{i=1}^{q}(1-\lambda_i z^{-1})\right\},
\end{aligned}$$

which is identical to [3.7.15].

The implication is as follows. Suppose a noninvertible representation for an $MA(q)$ process is written in the form

$$Y_t = \mu + \prod_{i=1}^{q}(1-\lambda_i L)\tilde{\varepsilon}_t, \qquad [3.7.16]$$

where

$$\begin{aligned}
|\lambda_i| &< 1 \qquad \text{for } i = 1, 2, \ldots, n\\
|\lambda_i| &> 1 \qquad \text{for } i = n+1, n+2, \ldots, q
\end{aligned}$$

and

$$E(\tilde{\varepsilon}_t\tilde{\varepsilon}_\tau) = \begin{cases}\tilde{\sigma}^2 & \text{for } t = \tau\\ 0 & \text{otherwise.}\end{cases}$$

Then the invertible representation is given by

$$Y_t = \mu + \left\{\prod_{i=1}^{n}(1-\lambda_i L)\right\}\left\{\prod_{i=n+1}^{q}(1-\lambda_i^{-1}L)\right\}\varepsilon_t, \qquad [3.7.17]$$

where

$$E(\varepsilon_t\varepsilon_\tau) = \begin{cases}\tilde{\sigma}^2\lambda_{n+1}^2\lambda_{n+2}^2 \cdots \lambda_q^2 & \text{for } t = \tau\\ 0 & \text{otherwise.}\end{cases}$$

Then [3.7.16] and [3.7.17] have the identical autocovariance-generating function, though only [3.7.17] satisfies the invertibility condition.

From the structure of the preceding argument, it is clear that there are a number of alternative $MA(q)$ representations of the data Y_t associated with all the possible "flips" between λ_i and λ_i^{-1}. Only one of these has all of the λ_i on or inside the unit circle. The innovations associated with this representation are said to be the fundamental innovations for Y_t.

APPENDIX 3.A. *Convergence Results for Infinite-Order Moving Average Processes*

This appendix proves the statements made in the text about convergence for the $MA(\infty)$ process [3.3.13].

First we show that absolute summability of the moving average coefficients implies square-summability. Suppose that $\{\psi_j\}_{j=0}^{\infty}$ is absolutely summable. Then there exists an $N < \infty$ such that $|\psi_j| < 1$ for all $j \geq N$, implying $\psi_j^2 < |\psi_j|$ for all $j \geq N$. Then

$$\sum_{j=0}^{\infty} \psi_j^2 = \sum_{j=0}^{N-1} \psi_j^2 + \sum_{j=N}^{\infty} \psi_j^2 < \sum_{j=0}^{N-1} \psi_j^2 + \sum_{j=N}^{\infty} |\psi_j|.$$

But $\sum_{j=0}^{N-1} \psi_j^2$ is finite, since N is finite, and $\sum_{j=N}^{\infty} |\psi_j|$ is finite, since $\{\psi_j\}$ is absolutely summable. Hence $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, establishing that [3.3.15] implies [3.3.14].

Next we show that square-summability does not imply absolute summability. For an example of a series that is square-summable but not absolutely summable, consider $\psi_j = 1/j$ for $j = 1, 2, \ldots$. Notice that $1/j > 1/x$ for all $x > j$, meaning that

$$1/j > \int_j^{j+1} (1/x)\, dx$$

and so

$$\sum_{j=1}^{N} 1/j > \int_1^{N+1} (1/x)\, dx = \log(N + 1) - \log(1) = \log(N + 1),$$

which diverges to ∞ as $N \to \infty$. Hence $\{\psi_j\}_{j=1}^{\infty}$ is not absolutely summable. It is, however, square-summable, since $1/j^2 < 1/x^2$ for all $x < j$, meaning that

$$1/j^2 < \int_{j-1}^{j} (1/x^2)\, dx$$

and so

$$\sum_{j=1}^{N} 1/j^2 < 1 + \int_1^{N} (1/x^2)\, dx = 1 + (-1/x)|_{x=1}^{N} = 2 - (1/N),$$

which converges to 2 as $N \to \infty$. Hence $\{\psi_j\}_{j=1}^{\infty}$ is square-summable.

Next we show that square-summability of the moving average coefficients implies that the $MA(\infty)$ representation in [3.3.13] generates a mean square convergent random variable. First recall what is meant by convergence of a deterministic sum such as $\sum_{j=0}^{\infty} a_j$ where $\{a_j\}$ is just a sequence of numbers. One criterion for determining whether $\sum_{j=0}^{T} a_j$ converges to some finite number as $T \to \infty$ is the *Cauchy criterion*. The Cauchy criterion states that $\sum_{j=0}^{\infty} a_j$ converges if and only if, for any $\varepsilon > 0$, there exists a suitably large integer N such that, for any integer $M > N$,

$$\left| \sum_{j=0}^{M} a_j - \sum_{j=0}^{N} a_j \right| < \varepsilon.$$

In words, once we have summed N terms, calculating the sum out to a larger number M does not change the total by any more than an arbitrarily small number ε.

For a stochastic process such as [3.3.13], the comparable question is whether $\sum_{j=0}^{T} \psi_j \varepsilon_{t-j}$ converges in mean square to some random variable Y_t as $T \to \infty$. In this case the Cauchy criterion states that $\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ converges if and only if, for any $\varepsilon > 0$, there exists a suitably large integer N such that for any integer $M > N$

$$E\left[\sum_{j=0}^{M} \psi_j \varepsilon_{t-j} - \sum_{j=0}^{N} \psi_j \varepsilon_{t-j} \right]^2 < \varepsilon. \qquad [3.A.1]$$

In words, once N terms have been summed, the difference between that sum and the one obtained from summing to M is a random variable whose mean and variance are both arbitrarily close to zero.

Now, the left side of [3.A.1] is simply

$$\begin{aligned} E[\psi_M\varepsilon_{t-M} &+ \psi_{M-1}\varepsilon_{t-M+1} + \cdots + \psi_{N+1}\varepsilon_{t-N-1}]^2 \\ &= (\psi_M^2 + \psi_{M-1}^2 + \cdots + \psi_{N+1}^2)\cdot\sigma^2 \\ &= \left[\sum_{j=0}^{M}\psi_j^2 - \sum_{j=0}^{N}\psi_j^2\right]\cdot\sigma^2. \end{aligned} \qquad [3.A.2]$$

But if $\Sigma_{j=0}^{\infty}\psi_j^2$ converges as required by [3.3.14], then by the Cauchy criterion the right side of [3.A.2] may be made as small as desired by choice of a suitably large N. Thus the infinite series in [3.3.13] converges in mean square provided that [3.3.14] is satisfied.

Finally, we show that absolute summability of the moving average coefficients implies that the process is ergodic for the mean. Write [3.3.18] as

$$\gamma_j = \sigma^2\sum_{k=0}^{\infty}\psi_{j+k}\psi_k.$$

Then

$$|\gamma_j| = \sigma^2\left|\sum_{k=0}^{\infty}\psi_{j+k}\psi_k\right|.$$

A key property of the absolute value operator is that

$$|a+b+c| \le |a| + |b| + |c|.$$

Hence

$$|\gamma_j| \le \sigma^2\sum_{k=0}^{\infty}|\psi_{j+k}\psi_k|$$

and

$$\sum_{j=0}^{\infty}|\gamma_j| \le \sigma^2\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}|\psi_{j+k}\psi_k| = \sigma^2\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}|\psi_{j+k}|\cdot|\psi_k| = \sigma^2\sum_{k=0}^{\infty}|\psi_k|\sum_{j=0}^{\infty}|\psi_{j+k}|.$$

But there exists an $M < \infty$ such that $\Sigma_{j=0}^{\infty}|\psi_j| < M$, and therefore $\Sigma_{j=0}^{\infty}|\psi_{j+k}| < M$ for $k = 0, 1, 2, \ldots$, meaning that

$$\sum_{j=0}^{\infty}|\gamma_j| < \sigma^2\sum_{k=0}^{\infty}|\psi_k|\cdot M < \sigma^2M^2 < \infty.$$

Hence [3.1.15] holds and the process is ergodic for the mean.

Chapter 3 Exercises

3.1. Is the following $MA(2)$ process covariance-stationary?

$$Y_t = (1 + 2.4L + 0.8L^2)\varepsilon_t$$

$$E(\varepsilon_t\varepsilon_\tau) = \begin{cases} 1 & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases}$$

If so, calculate its autocovariances.

3.2. Is the following $AR(2)$ process covariance-stationary?

$$(1 - 1.1L + 0.18L^2)Y_t = \varepsilon_t$$

$$E(\varepsilon_t\varepsilon_\tau) = \begin{cases} 1 & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases}$$

If so, calculate its autocovariances.

3.3. A covariance-stationary $AR(p)$ process,

$$(1 - \phi_1L - \phi_2L^2 - \cdots - \phi_pL^p)(Y_t - \mu) = \varepsilon_t,$$

has an $MA(\infty)$ representation given by

$$(Y_t - \mu) = \psi(L)\varepsilon_t$$

with

$$\psi(L) = 1/[1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p]$$

or

$$[1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p]\,[\psi_0 + \psi_1 L + \psi_2 L^2 + \cdots] = 1.$$

In order for this equation to be true, the implied coefficient on L^0 must be unity and the coefficients on $L^1, L^2, L^3, \ldots$ must be zero. Write out these conditions explicitly and show that they imply a recursive algorithm for generating the $MA(\infty)$ weights $\psi_0, \psi_1, \ldots$. Show that this recursion is algebraically equivalent to setting ψ_j equal to the (1, 1) element of the matrix $\mathbf{F}$ raised to the jth power as in equation [1.2.28].

3.4. Derive [3.4.21].

3.5. Verify [3.4.22].

3.6. Suggest a recursive algorithm for calculating the $AR(\infty)$ weights,

$$(1 + \eta_1 L + \eta_2 L^2 + \cdots)(Y_t - \mu) = \varepsilon_t$$

associated with an invertible $MA(q)$ process,

$$(Y_t - \mu) = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t.$$

Give a closed-form expression for η_j as a function of the roots of

$$(1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q) = 0,$$

assuming that these roots are all distinct.

3.7. Repeat Exercise 3.6 for a noninvertible $MA(q)$ process. (HINT: Recall equation [3.7.17].)

3.8. Show that the $MA(2)$ process in Exercise 3.1 is not invertible. Find the invertible representation for the process. Calculate the autocovariances of the invertible representation using equation [3.3.12] and verify that these are the same as obtained in Exercise 3.1.

Chapter 3 References

Anderson, Brian D. O., and John B. Moore. 1979. *Optimal Filtering*. Englewood Cliffs, N.J.: Prentice-Hall.

Hannan, E. J. 1970. *Multiple Time Series*. New York: Wiley.

Hansen, Lars P., and Thomas J. Sargent. 1981. "Formulating and Estimating Dynamic Linear Rational Expectations Models," in Robert E. Lucas, Jr., and Thomas J. Sargent, eds., *Rational Expectations and Econometric Practice*, Vol. I. Minneapolis: University of Minnesota Press.

Rao, C. Radhakrishna. 1973. *Linear Statistical Inference and Its Applications*, 2d ed. New York: Wiley.

Sargent, Thomas J. 1987. *Macroeconomic Theory*, 2d ed. Boston: Academic Press.

4 Forecasting

This chapter discusses how to forecast time series. Section 4.1 reviews the theory of forecasting and introduces the idea of a linear projection, which is a forecast formed from a linear function of past observations. Section 4.2 describes the forecasts one would use for *ARMA* models if an infinite number of past observations were available. These results are useful in theoretical manipulations and in understanding the formulas in Section 4.3 for approximate optimal forecasts when only a finite number of observations are available.

Section 4.4 describes how to achieve a triangular factorization and Cholesky factorization of a variance-covariance matrix. These results are used in that section to calculate exact optimal forecasts based on a finite number of observations. They will also be used in Chapter 11 to interpret vector autoregressions, in Chapter 13 to derive the Kalman filter, and in a number of other theoretical calculations and numerical methods appearing throughout the text. The triangular factorization is used to derive a formula for updating a forecast in Section 4.5 and to establish in Section 4.6 that for Gaussian processes the linear projection is better than any nonlinear forecast.

Section 4.7 analyzes what kind of process results when two different *ARMA* processes are added together. Section 4.8 states Wold's decomposition, which provides a basis for using an $MA(\infty)$ representation to characterize the linear forecast rule for any covariance-stationary process. The section also describes a popular empirical approach for finding a reasonable approximation to this representation that was developed by Box and Jenkins (1976).

4.1. *Principles of Forecasting*

Forecasts Based on Conditional Expectation

Suppose we are interested in forecasting the value of a variable Y_{t+1} based on a set of variables $\mathbf{X}_t$ observed at date t. For example, we might want to forecast Y_{t+1} based on its m most recent values. In this case, $\mathbf{X}_t$ would consist of a constant plus Y_t, Y_{t-1}, . . . , and Y_{t-m+1}.

Let $Y^*_{t+1|t}$ denote a forecast of Y_{t+1} based on $\mathbf{X}_t$. To evaluate the usefulness of this forecast, we need to specify a *loss function*, or a summary of how concerned we are if our forecast is off by a particular amount. Very convenient results are obtained from assuming a quadratic loss function. A quadratic loss function means choosing the forecast $Y^*_{t+1|t}$ so as to minimize

$$E(Y_{t+1} - Y^*_{t+1|t})^2. \qquad [4.1.1]$$