1.32 If the sequence of L^2 random variables x_n is such that $E(x_n) \to \mu$ and

$$E[(x_n - Ex_n)^2] \to 0,$$

$x_n \xrightarrow{p} \mu$.

1.33 Show that, if $x_n \xrightarrow{d} x, y_n \xrightarrow{d} y$, with x_n and y_n independent, then $x_n + y_n \xrightarrow{d} x + y$, with x and y independent.

Section T1.12

1.34 Suppose

$$x_t = \beta_0 + \beta_1 t,$$

where β_0 and β_1 are constants. Prove as $n \to \infty$,

$$\widehat{\rho}_x(h) \to 1$$

for fixed h, where $\widehat{\rho}_x(h)$ is the ACF (1.32).

1.35 Suppose x_t is a weakly stationary time series with mean zero and with absolutely summable autocovariance function, $\gamma(h)$, such that

$$\sum_{h=-\infty}^{\infty} \gamma(h) = 0.$$

Prove that $\sqrt{n}\, \bar{x} \xrightarrow{p} 0$, where $\bar{x}$ is the sample mean (1.30).

1.36 Let x_t be a linear process of the form (1.26), satisfying (1.27). If we define

$$\tilde{\gamma}(h) = n^{-1} \sum_{t=1}^{n} (x_{t+h} - \mu_x)(x_t - \mu_x),$$

show that

$$n^{1/2}\big(\tilde{\gamma}(h) - \widehat{\gamma}(h)\big) = o_p(1).$$

Hint: The Markov Inequality

$$P\{|x| \geq \epsilon\} < \frac{E|x|}{\epsilon}$$

can be helpful for the cross-product terms.

1.37 For a linear process of the form

$$x_t = \sum_{j=0}^{\infty} \phi_1^j w_{t-j},$$

with $|\phi_1| < 1$, show that

$$\sqrt{n} \frac{(\widehat{\rho}_x(1) - \rho_x(1))}{\sqrt{1 - \rho_x^2(1)}} \xrightarrow{d} N(0, 1),$$

and construct a 95% confidence interval for ϕ_1 when $\widehat{\rho}_x(1) = .64$.

CHAPTER 2

Time Series Regression and ARIMA Models

2.1 Introduction

In Chapter 1, we introduced autocorrelation and cross-correlation functions (ACF's and CCF's) as tools for clarifying relations that may occur within and between time series at various lags. In addition, we have explained how to build linear models based on classical regression theory for exploiting the associations indicated by large values of the ACF or CCF. The time domain methods of this chapter, contrasted with the frequency domain methods introduced in later chapters, are appropriate when we are dealing with possibly nonstationary, shorter time series; these series are the rule rather than the exception in applications arising in economics and the social sciences. In addition, the emphasis in these fields is usually on forecasting future values, which is easily treated as a regression problem. This chapter develops a number of regression techniques for time series that are all related to classical ordinary and weighted or correlated least squares.

Classical regression is often insufficient for explaining all of the interesting dynamics of a time series. For example, the ACF of the residuals of the simple linear regression fit to the global temperature data (see Example 1.22 of Chapter 1) reveals additional structure in the data that the regression did not capture. Instead, the introduction of correlation as a phenomenon that may be generated through lagged linear relations leads to proposing the ***autoregressive (AR)*** and ***autoregressive moving average (ARMA)*** models. Adding nonstationary models to the mix leads to the ***autoregressive integrated moving average (ARIMA)*** model popularized in the landmark work by Box and Jenkins (1970). The ***Box–Jenkins method*** for identifying a plausible ARIMA model is given in this chapter along with techniques for ***parameter estimation*** and ***forecasting*** for these models. In the later sections, we present ***long memory ARMA***, ***threshold autoregressive models***, ***regression with ARMA errors***, and an extension of the Box–Jenkins method for predicting a single output from a collection of possible input series is considered where the inputs themselves may follow ARIMA models, commonly referred to as ***transfer function models***. Finally, we present ***ARCH models*** and the analysis of volatility.

2.2 Autoregressive Moving Average Models

The classical regression model in Section 1.8 of Chapter 1 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 created with the idea that the present 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 1.9 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 1.8 gives an indication that forecasting for such a model might be a distinct possibility, say, through some version such as

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

where the quantity on the left-hand side denotes the forecast at the next period $t+1$ based on current and past observed values $x_1, x_2, \ldots, x_t$. We will make this notion more precise in our discussion of forecasting (Section 2.5).

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 1. For example, the lagged scatterplot matrix for the Southern Oscillation Index (SOI), shown in Figure 1.18, gives a distinct indication that lags 1 and 2, for example, are linearly associated with the present value. The ACF shown in Figure 1.13 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 1.19. 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 definition of an ***autoregressive model of order p***, abbreviated as ***AR(p)***, of the form

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + w_t, \tag{2.1}$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are constants and w_t is a white noise series with mean zero and variance σ_w^2. We assume for simplicity in notation that the mean of x_t is zero. If the mean, μ, of x_t is not zero, we can replace x_t by $x_t - \mu$ in (2.1), or write

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + w_t, \tag{2.2}$$

where $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$. We note several different ways (2.1) can be written that will be used in the sequel. First, define $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)'$ and $\boldsymbol{x}_{t-1} = (x_{t-1}, x_{t-2}, \ldots, x_{t-p})'$ so that

$$x_t = \boldsymbol{\phi}' \boldsymbol{x}_{t-1} + w_t \tag{2.3}$$

and the AR(p) model becomes the regression model of Section 1.8. Some technical difficulties, however, develop from applying that model because $\boldsymbol{x}_{t-1}$ has random components, whereas $\boldsymbol{z}_t$ was assumed to be fixed. A second useful form follows by using the backshift operator (1.41) to write the AR(p) model, (2.1), as

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)x_t = w_t, \tag{2.4}$$

or even more concisely as

$$\phi(B)x_t = w_t, \tag{2.5}$$

where the ***autoregressive operator***

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p \tag{2.6}$$

is an operator whose properties are important in solving (2.5) for x_t.

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 the variance of x_t is bounded, we can represent an AR(1) model by

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j}, \tag{2.7}$$

in the mean square sense (see Section T1.10). This conclusion follows from the fact that

$$\lim_{k\to\infty} E\left(x_t - \sum_{j=0}^{k-1}\phi^j w_{t-j}\right)^2 = \lim_{k\to\infty} \phi^{2k} E\left(x_{t-k}^2\right) = 0.$$

Alternately, we could simply have defined an AR(1) model to be the stationary process given in equation (2.7), because, with $|\phi| < 1$,

$$\begin{aligned} x_t &= \sum_{j=0}^{\infty}\phi^j w_{t-j} = (\sum_{j=1}^{\infty}\phi^j w_{t-j}) + w_t \\ &= \phi(\sum_{j=0}^{\infty}\phi^j w_{t-1-j}) + w_t = \phi x_{t-1} + w_t. \end{aligned} \tag{2.8}$$

The AR(1) process defined by (2.7) is stationary with mean

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

and autocovariance function,

$$\begin{aligned} \gamma(h) &= \operatorname{cov}(x_{t+h}, x_t) = E\left[\left(\sum_{j=0}^{\infty}\phi^j w_{t+h-j}\right)\left(\sum_{k=0}^{\infty}\phi^j w_{t-k}\right)\right] \\ &= \sigma_w^2\sum_{j=0}^{\infty}\phi^j\phi^{j+h} = \sigma_w^2\phi^h\sum_{j=0}^{\infty}\phi^{2j} = \frac{\sigma_w^2\phi^h}{1-\phi^2}, \quad h \ge 0. \end{aligned} \tag{2.9}$$

Recall that $\gamma(h) = \gamma(-h)$, so we will only exhibit the autocovariance function for $h \ge 0$. From (2.9), the ACF of an AR(1) is

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h, \quad h \ge 0, \tag{2.10}$$

and $\rho(h)$ satisfies the recursion

$$\rho(h) = \phi\rho(h-1), \quad h \ge 1. \tag{2.11}$$

Well will discuss the ACF of a general AR(p) model in Section 2.4.

Example 2.1 The Sample Path of an AR(1) Process

Figure 2.1 shows a time plot of two AR(1) processes, one with $\phi = 0.9$ and one with $\phi = -0.9$; in both cases, $\sigma_w^2 = 1$. In the first case, $\rho(h) = .9^h$, for $h \ge 0$, so the observations close together in time are positively correlated with each other. This result means that the observations at

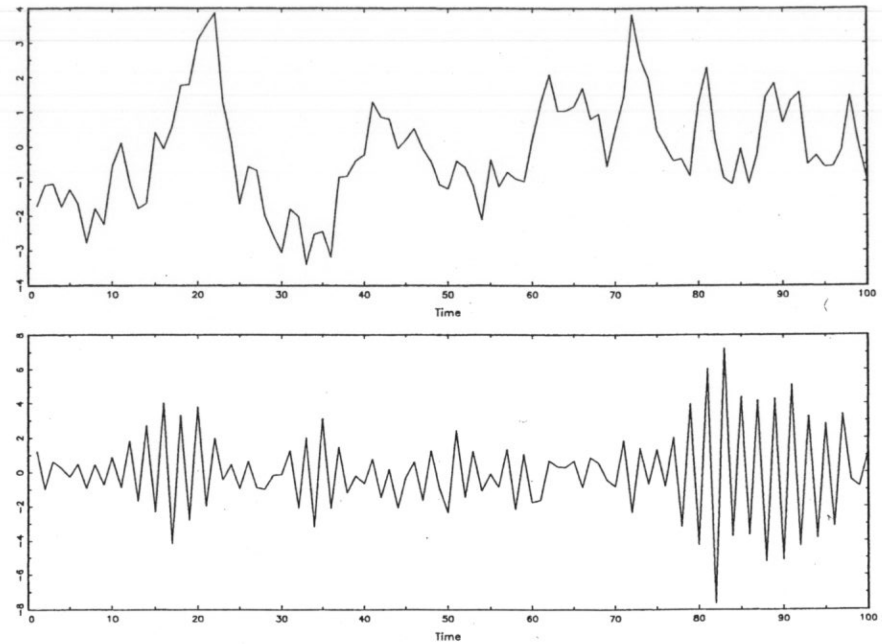

Figure 2.1: Simulated AR(1) models: $\phi = 0.9$ (top); $\phi = -0.9$ (bottom).

contiguous time points will tend be close in value to each other; this fact shows up in top of Figure 2.1 as a very smooth sample path for x_t. Now, contrast this to the case in which $\phi = -0.9$, so that $\rho(h) = (-.9)^h$, for $h \ge 0$. 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 2.1, where if an observation, x_t, is positive [negative], the next observation, x_{t+1}, is typically negative [positive], and the next observation, x_{t+2} is typically positive [negative]. Thus, in this case, the sample path is very choppy.

Example 2.2 Explosive AR Models and Causality

In Chapter 1, Problem 1.8, 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, $\sum_{j=0}^{k-1} \phi^j w_{t-j}$ will not converge in mean square as $k \to \infty$, so the intuition used to get (2.7) 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,

$$\begin{aligned} x_t &= \phi^{-1}x_{t+1} - \phi^{-1}w_{t+1} = \phi^{-1}\left(\phi^{-1}x_{t+2} - \phi^{-1}w_{t+2}\right) - \phi^{-1}w_{t+1} \\ &\vdots \\ &= \phi^{-k}x_{t+k} - \sum_{j=1}^{k-1}\phi^{-j}w_{t+j}, \qquad (2.12) \end{aligned}$$

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

$$\phi(B)x_t = w_t, \qquad (2.13)$$

where $\phi(B) = 1 - \phi B$, and $|\phi| < 1$. Also, write the model in equation (2.7) using operator form as

$$x_t = \sum_{j=0}^{\infty}\psi_j w_{t-j} = \psi(B)w_t, \qquad (2.14)$$

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 $\phi(B)x_t$ from (2.13) for w_t in (2.14) to obtain

$$x_t = \psi(B)w_t = \psi(B)\phi(B)x_t. \qquad (2.15)$$

Equating coefficients on the left- and right-hand sides of (2.15), we get

$$1 = (1 + \psi_1 B + \psi_2 B^2 + \cdots + \psi_j B^j + \cdots)(1 - \phi B). \qquad (2.16)$$

Reorganizing the coefficients in (2.16),

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

we see that for each $j = 1, 2, ...$, the coefficient of B^j on the right, must be zero (because it is zero on the left). The coefficient of B on the right 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, $\psi_j - \psi_{j-1}\phi = 0$, which leads to the general solution $\psi_j = \phi^j$.

This example makes it clear that $\psi(B)$ is also the inverse of the operator $\phi(B)$. In operator form, we took the following steps starting with the AR(1) model, $\phi(B)x_t = w_t$, where $\phi(B) = (1 - \phi B)$.

(i) Multiply each side by the inverse operator (assuming it exists)

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

(ii) Write the result as $x_t = \psi(B)w_t$, where we defined $\psi(B) = \phi^{-1}(B)$.

(iii) Solve for $\phi^{-1}(B)$ by matching the coefficients in $\psi(B)\phi(B) = 1$.

The solution, of course, was $\phi^{-1}(B) = 1 + \phi B + \phi^2 B^2 + \cdots + \phi^j B^j + \cdots$. Notice the operators behave like polynomials. That is, consider the polynomial $\phi(z) = 1 - \phi z$, where z is a complex number and $|\phi| < 1$. Then,

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

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. In such cases, we write

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \ldots + \theta_q w_{t-q} \qquad (2.17)$$

where there are q lags in the moving average and $\theta_1, \theta_2, \ldots, \theta_q$ are parameters that determine the overall pattern of the process. The system is the same as the infinite moving average defined as the linear process (2.14), where $\psi_0 = 1$, $\psi_j = \theta_j, j = 1, \ldots, q$, and $\psi_j = 0$ for other values. We may also write the MA(q) process in the equivalent form

$$x_t = \theta(B)w_t, \qquad (2.18)$$

where the ***moving average operator***

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q \tag{2.19}$$

defines a linear combination of values in the shift operator $B^k w_t = w_{t-k}$ as before.

Unlike the autoregressive process, the moving average process is stationary for any values of the parameters $\theta_1, ..., \theta_q$; details of this result are provided in Section 2.4.

Example 2.3 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

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

The ACF of a general MA(q) model will be presented in Section 2.4.

Note $|\rho(1)| \leq 1/2$ for all values of θ (Problem 2.1). The time series is "one-dependent;" that is, 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 $\theta = 0.5$, for example, x_t and x_{t-1} are positively correlated, and $\rho(1) = 0.4$. When $\theta = -0.5$, x_t and x_{t-1} are negatively correlated, $\rho(1) = -0.4$. Figure 2.2 shows a time plot of these two processes with $\sigma_w^2 = 1$. The series in Figure 2.2, where $\theta = 0.5$, is smoother than the series in Figure 2.2, where $\theta = -0.5$.

Example 2.4 Non-uniqueness of MA Models and Invertibility

From Example 2.3, 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,

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

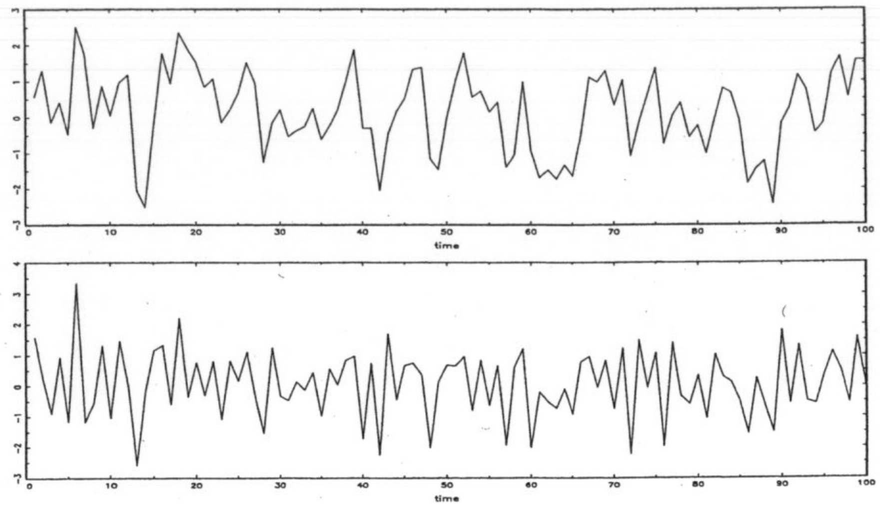

Figure 2.2: Simulated MA(1) models: $\theta = 0.5$ (top); $\theta = -0.5$ (bottom).

Thus, the MA(1) processes

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

and

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

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 (2.40), 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 polynomials, $\theta(z)$, 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 (2.13)-(2.16), 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. A time series x_t, for $t = 0, \pm1, \pm2, ...$, is said to be ARMA(p,q) if x_t is stationary and

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}, \tag{2.20}$$

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. As before, if x_t has a nonzero mean μ, we set $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$ and write the model as

$$x_t = \alpha + \phi_1 x_{t-1} + \cdots \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots \theta_q w_{t-q}. \tag{2.21}$$

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, (2.6), and the MA operator, (2.19). In particular, the ARMA(p,q) model in (2.20) can then can be written in concise form as

$$\phi(B)x_t = \theta(B)w_t. \tag{2.22}$$

Before we discuss the conditions under which (2.20) is causal and invertible, we point out a potential problem with the ARMA model.

Example 2.5 Parameter Redundancy

Consider a white noise process $x_t = w_t$. Equivalently, we can write this as $0.5x_{t-1} = 0.5w_{t-1}$ by shifting back one unit of time and multiplying by 0.5. Now, subtract the two representations to obtain

$$x_t - 0.5x_{t-1} = w_t - 0.5w_{t-1},$$

or

$$x_t = 0.5x_{t-1} - 0.5w_{t-1} + w_t,$$

which looks like an ARMA(1,1) model. Of course, x_t is still white noise; nothing has changed in this regard, but we have hidden the fact that x_t is white noise because of the ***parameter redundancy*** or over-parameterization. Write the parameter redundant model in operator form:

$$(1 - 0.5B)x_t = (1 - 0.5B)w_t.$$

Apply the operator $(1-0.5B)^{-1}$ to both sides to obtain

$$x_t = (1-0.5B)^{-1}(1-0.5B)x_t = (1-0.5B)^{-1}(1-0.5B)w_t = w_t,$$

which is the original model. We can easily detect the problem of over-parameterization with the use of polynomials by writing the AR polynomial $\phi(z) = (1-0.5z)$, the MA polynomial $\theta(z) = (1-0.5z)$, and noting that both polynomials have a ***common factor***, namely $(1-0.5z)$. This common factor immediately identifies the parameter redundancy. Discarding the common factor in each leaves $\phi(z) = 1$ and $\theta(z) = 1$, and we deduce that the model is actually white noise. The consideration of parameter redundancy will be crucial when we discuss estimation for general ARMA models. As this example points out, we might fit an ARMA(1,1) model to white noise data and find that the parameter estimates are significant. If we were unaware of parameter redundancy, we might claim the data are correlated when in fact they are not (Problem 2.19).

Examples 2.2, 2.4, and 2.5 point to a number of problems with the general definition of ARMA(p,q) models, as given by (2.20), or, equivalently, by (2.22). To summarize, we have seen the following problems:

(i) parameter redundant models,

(ii) stationary AR models that depend on the future, and

(iii) MA models that are not unique.

To overcome these problems, we will require some additional restrictions on the model parameters. First, we define the ***AR and MA polynomials*** as

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p, \quad \phi_p \neq 0, \tag{2.23}$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q, \quad \theta_q \neq 0, \tag{2.24}$$

respectively, where z is a complex number.

To address the first problem, we will henceforth refer to an ARMA(p,q) model to mean that it is in its simplest form. That is, in addition to the original definition given in equation (2.20), ***we will also require that*** $\phi(z)$ ***and*** $\theta(z)$ ***have no common factors***. So, the process, $x_t = 0.5x_{t-1} - 0.5w_{t-1} + w_t$, discussed in Example 2.5 is not referred to as an ARMA(1,1) process because, in its reduced form, x_t is white noise.

To address the problem of future-dependent models, we formally introduce the concept of ***causality***. An ARMA(p,q) model, $\phi(B)x_t = \theta(B)w_t$, is said to be causal, if the time series x_t, $t = 0, \pm1, \pm2, ...$, can be written as a one-sided linear process:

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B)w_t, \tag{2.25}$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$, and $\sum_{j=0}^{\infty} |\psi_j| < \infty$; we set $\psi_0 = 1$.

In Example 2.2, the AR(1) process, $x_t = \phi x_{t-1} + w_t$, is causal only when $|\phi| < 1$. Equivalently, the process is causal only when the root of $\phi(z) = 1 - \phi z$ is bigger than one in absolute value. That is, the root, say, z_0, of $\phi(z)$ is $z_0 = 1/\phi$ [because $\phi(z_0) = 0$] and $|z_0| > 1$ because $|\phi| < 1$. In general, we have the following property.

Property P2.1: Causality of an ARMA(p, q) Process
An ARMA(p, q) model is causal only when the roots of $\phi(z)$ lie outside the unit circle; that is, $\phi(z) = 0$ only when $|z| > 1$. The coefficients of the linear process given in (2.25) can be determined by solving

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \le 1.$$

Finally, to address the problem of uniqueness discussed in Example 2.4, we choose the model that allows an infinite autoregressive representation. In particular, an ARMA(p, q) model, $\phi(B)x_t = \theta(B)w_t$, is said to be ***invertible***, if the time series x_t, $t = 0, \pm1, \pm2, \ldots$, can be written as

$$\pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = w_t, \tag{2.26}$$

where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$, and $\sum_{j=0}^{\infty} |\pi_j| < \infty$; we set $\pi_0 = 1$. Analogous to Property P2.1, we have the following property.

Property P2.2: Invertibility of an ARMA(p, q) Process
An ARMA(p, q) model is invertible only when the roots of $\theta(z)$ lie outside the unit circle. The coefficients π_j of $\pi(B)$ given in (2.26) can be determined by solving

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \le 1.$$

The proof of Property P2.1 is given in Section T2.16 (the proof of Property P2.2 is similar and, hence, is not provided). The following examples illustrate these concepts.

Example 2.6 Parameter Redundancy, Causality, and Invertiblity

Consider the process

$$x_t = 0.4x_{t-1} + 0.45x_{t-2} + w_{t-1} + 0.25w_{t-2} + w_t,$$

or, in operator form,

$$(1 - 0.4B - 0.45B^2)x_t = (1 + B + 0.25B^2)w_t.$$

At first, x_t appears to be an ARMA(2,2) process. But, the associated polynomials $\phi(z) = 1 - 0.4z - 0.45z^2 = (1 + 0.5z)(1 - 0.9z)$, and, $\theta(z) = (1 + z + 0.25z^2) = (1 + 0.5z)^2$, have a common factor that can be cancelled. After cancellation, the polynomials become $\phi(z) = (1 - 0.9z)$ and $\theta(z) = (1 + 0.5z)$, so the model is an ARMA(1,1) model, $(1 - 0.9B)x_t = (1 + 0.5B)w_t$, or

$$x_t = 0.9x_{t-1} + 0.5w_{t-1} + w_t. \tag{2.27}$$

The model is causal because $\phi(z) = (1 - 0.9z) = 0$ when $z = 10/9$, which is outside the unit circle. The model is also invertible because the root of $\theta(z) = (1 + 0.5z)$ is $z = -2$, which is outside the unit circle.

To write the model as a linear process, we can obtain the ψ-weights using Property P2.1:

$$\begin{aligned}\psi(z) &= \frac{\theta(z)}{\phi(z)} = \frac{(1 + 0.5z)}{(1 - 0.9z)} \\ &= (1 + 0.5z)(1 + 0.9z + 0.9^2z^2 + 0.9^3z^3 + \cdots) \quad |z| \le 1.\end{aligned}$$

The coefficient of z^j in $\psi(z)$ is $\psi_j = (0.5 + 0.9)0.9^{j-1}$, for $j \ge 1$, so (2.27) can be written as

$$x_t = w_t + 1.4\sum_{j=1}^{\infty} 0.9^{j-1} w_{t-j}.$$

Similarly, to find the invertible representation using Property P2.2:

$$\pi(z) = \frac{\phi(z)}{\theta(z)} = (1 - 0.9z)(1 - 0.5z + 0.5^2z^2 - 0.5^3z^3 + \cdots) \quad |z| \le 1.$$

In this case, the π-weights are given by $\pi_j = (-1)^j(0.9 + 0.5)0.5^{j-1}$, for $j \ge 1$, and hence, we can also write (2.27) as

$$x_t = 1.4\sum_{j=1}^{\infty} (-0.5)^{j-1} x_{t-j} + w_t.$$

Example 2.7 Causal Conditions for an AR(2) Process

For an AR(1) model, $(1 - \phi B)x_t = w_t$, to be causal, the root of $\phi(z) = 1 - \phi z$ must lie outside of the unit circle. In this case, the root is $z = 1/\phi$, so that it is easy to go from the causal requirement on the root, that is, $|1/\phi| > 1$, to a requirement on the parameter, that is, $|\phi| < 1$. It is not so easy to establish this relationship for higher order models.

For example, the AR(2) model, $(1-\phi_1 B-\phi_2 B^2)x_t = w_t$, is causal when the two roots of $\phi(z) = 1-\phi_1 z-\phi_2 z^2$ lie outside of the unit circle. Using the quadratic formula, this requirement can be written as

$$\left| \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \right| > 1.$$

The roots of $\phi(z)$ may be real and distinct, real and equal, or a complex conjugate pair. If we denote those roots by z_1 and z_2, we can write $\phi(z) = (1-z_1^{-1}z)(1-z_2^{-1}z)$; note that $\phi(z_1) = \phi(z_2) = 0$. The model can be written in operator form as $(1-z_1^{-1}B)(1-z_2^{-1}B)x_t = w_t$. From this representation, it follows that $\phi_1 = (z_1^{-1}+z_2^{-1})$ and $\phi_2 = -(z_1 z_2)^{-1}$. This relationship can be used to establish the following equivalent condition for causality:

$$\phi_1 + \phi_2 < 1, \quad \phi_2 - \phi_1 < 1, \quad \text{and} \quad |\phi_2| < 1. \tag{2.28}$$

This causality condition specifies a triangular region in the parameter space. We leave the details of the equivalence to the reader (Problem 2.4).

2.3 Homogeneous Difference Equations

The study of the behavior of ARMA processes is greatly enhanced by the use of homogeneous difference equations. This topic is also useful in the study of time domain models and stochastic processes in general. We will give a brief and heuristic account of the topic along with some examples of the usefulness of the theory. For details, the reader is referred to Mickens (1987).

Suppose we have a sequence of numbers $u_0, u_1, u_2, \ldots$ such that

$$u_n - \alpha u_{n-1} = 0, \quad \alpha \neq 0, \quad n = 1, 2, \ldots \tag{2.29}$$

For example, recall (2.11) in which we showed that the ACF of an AR(1) process is a sequence, $\rho(h)$, satisfying $\rho(h) = \phi\rho(h-1)$, for $h = 1, 2, \ldots$ Equation (2.29) represents a ***homogeneous difference equation of order 1***. To solve the equation, we write:

$$\begin{aligned} u_1 &= \alpha u_0 \\ u_2 &= \alpha u_1 = \alpha^2 u_0 \\ &\vdots \\ u_n &= \alpha u_{n-1} = \alpha^n u_0. \end{aligned}$$

Given an initial condition $u_0 = c$, we may solve (2.29), namely, $u_n = \alpha^n c$.

In operator notation, (2.29) can be written as $(1-\alpha B)u_n = 0$. The polynomial associated with (2.29) is $\alpha(z) = 1 - \alpha z$, and the root, say, z_0, of this polynomial is $z_0 = 1/\alpha$; that is $\alpha(z_0) = 0$. We know the solution to (2.29), with initial condition $u_0 = c$, is

$$u_n = \alpha^n c = \left(z_0^{-1}\right)^n c.$$

That is, the solution to the difference equation (2.29) depends only on the initial condition and the inverse of the root to the associated polynomial $\alpha(z)$.

Now suppose that the sequence satisfies

$$u_n - \alpha_1 u_{n-1} - \alpha_2 u_{n-2} = 0, \quad \alpha_2 \neq 0, \quad n = 2, 3, \ldots \tag{2.30}$$

This equation is a ***homogeneous difference equation of order 2***. The corresponding polynomial is

$$\alpha(z) = 1 - \alpha_1 z - \alpha_2 z^2,$$

which has two roots, say, z_1 and z_2; that is, $\alpha(z_1) = \alpha(z_2) = 0$. We will consider two cases. First suppose $z_1 \neq z_2$. Then the general solution to (2.30) is

$$u_n = c_1 z_1^{-n} + c_2 z_2^{-n}, \tag{2.31}$$

where c_1 and c_2 depend on the initial conditions. This claim can be verified by direct substitution of (2.31) into (2.30):

$$\begin{aligned} c_1 z_1^{-n} + c_2 z_2^{-n} & - \alpha_1 \left(c_1 z_1^{-(n-1)} + c_2 z_2^{-(n-1)}\right) - \alpha_2 \left(c_1 z_1^{-(n-2)} + c_2 z_2^{-(n-2)}\right) \\ &= c_1 z_1^{-n}\left(1 - \alpha_1 z_1 - \alpha_2 z_1^2\right) + c_2 z_2^{-n}\left(1 - \alpha_1 z_2 - \alpha_2 z_2^2\right) \\ &= c_1 z_1^{-n}\alpha(z_1) + c_2 z_2^{-n}\alpha(z_2) \\ &= 0. \end{aligned}$$

Given two initial conditions u_0 and u_1, we may solve for c_1 and c_2:

$$\begin{aligned} u_0 &= c_1 + c_2 \\ u_1 &= c_1 z_1^{-1} + c_2 z_2^{-1}, \end{aligned}$$

where z_1 and z_2 can be solved for in terms of α_1 and α_2 using the quadratic formula, for example.

When the roots are equal, $z_1 = z_2$ $(= z_0)$, the general solution to (2.30) is

$$u_n = z_0^{-n}(c_1 + c_2 n). \tag{2.32}$$

This claim can also be verified by direct substitution of (2.32) into (2.30):

$$\begin{aligned} z_0^{-n}(c_1 + c_2 n) & - \alpha_1 \left(z_0^{-(n-1)}[c_1 + c_2(n-1)]\right) - \alpha_2 \left(z_0^{-(n-2)}[c_1 + c_2(n-2)]\right) \\ &= z_0^{-n}(c_1 + c_2 n)\left(1 - \alpha_1 z_0 - \alpha_2 z_0^2\right) + c_2 z_0^{-n+1}\left(\alpha_1 + 2\alpha_2 z_0\right) \\ &= c_2 z_0^{-n+1}\left(\alpha_1 + 2\alpha_2 z_0\right). \end{aligned}$$