CHAPTER 5

Autoregressive Integrated Moving Average (ARIMA) Models

All models are wrong, some are useful.

GEORGE E. P. BOX, *British statistician*

5.1 INTRODUCTION

In the previous chapter, we discussed forecasting techniques that, in general, were based on some variant of exponential smoothing. The general assumption for these models was that any time series data can be represented as the sum of two distinct components: deterministic and stochastic (random). The former is modeled as a function of time whereas for the latter we assumed that some random noise that is added on the deterministic signal generates the stochastic behavior of the time series. One very important assumption is that the random noise is generated through independent shocks to the process. In practice, however, this assumption is often violated. That is, usually successive observations show serial dependence. Under these circumstances, forecasting methods based on exponential smoothing may be inefficient and sometimes inappropriate because they do not take advantage of the serial dependence in the observations in the most effective way. To formally incorporate this dependent structure, in this chapter we will explore a general class of models called autoregressive integrated moving average models or ARIMA models (also known as Box–Jenkins models).

5.2 LINEAR MODELS FOR STATIONARY TIME SERIES

In statistical modeling, we are often engaged in an endless pursuit of finding the ever elusive true relationship between certain inputs and the output. As cleverly put

Introduction to Time Series Analysis and Forecasting
By Douglas C. Montgomery, Cheryl L. Jennings, and Murat Kulahci

by the quote of this chapter, these efforts usually result in models that are nothing but approximations of the "true" relationship. This is generally due to the choices the analyst makes along the way to ease the modeling efforts. A major assumption that often provides relief in modeling efforts is the linearity assumption. A **linear filter**, for example, is a linear operation from one time series x_t to another time series y_t,

$$y_t = L(x_t) = \sum_{i=-\infty}^{+\infty} \psi_i x_{t-i} \tag{5.1}$$

with $t = \ldots, -1, 0, 1, \ldots$. In that regard the linear filter can be seen as a "process" that converts the input, x_t, into an output, y_t, and that conversion is not instantaneous but involves all (present, past, and future) values of the input in the form of a summation with different "weights", $\{\psi_i\}$, on each x_t. Furthermore, the linear filter in Eq. (5.1) is said to have the following properties:

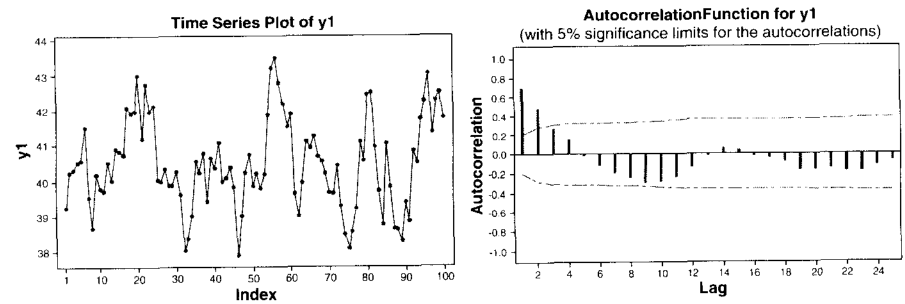
1. **Time-invariant** as the coefficients $\{\psi_i\}$ do not depend on time.
2. **Physically realizable** if $\psi_i = 0$ for $i < 0$; that is, the output y_t is a linear function of the current and past values of the input: $y_t = \psi_0 x_t + \psi_1 x_{t-1} + \cdots$.
3. **Stable** if $\sum_{i=-\infty}^{+\infty} |\psi_i| < \infty$.

In linear filters, under certain conditions, some properties such as **stationarity** of the input time series are also reflected in the output. We discussed stationarity previously in Chapter 2. We will now give a more formal description of it before proceeding further with linear models for time series.

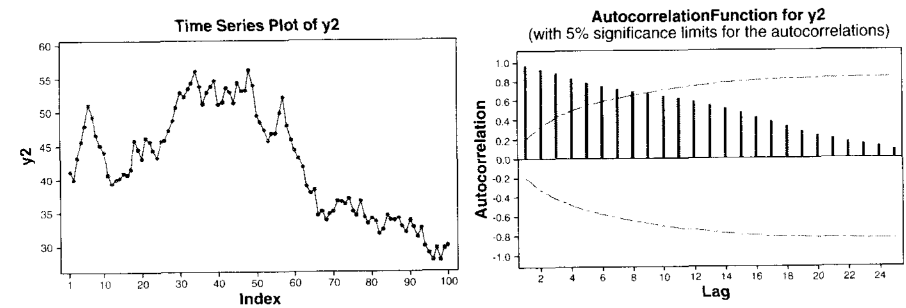
5.2.1 Stationarity

The **stationarity** of a time series is related to its statistical properties in time. That is, in the more strict sense, a stationary time series exhibits similar "statistical behavior" in time and this is often characterized as a constant probability distribution in time. However, it is usually satisfactory to consider the first two moments of the time series and define stationarity (or **weak stationarity**) as follows: (1) the expected value of the time series does not depend on time and (2) the autocovariance function defined as $\mathrm{Cov}(y_t, y_{t+k})$ for any lag k is only a function of k and not time; that is, $\gamma_y(k) = \mathrm{Cov}(y_t, y_{t+k})$.

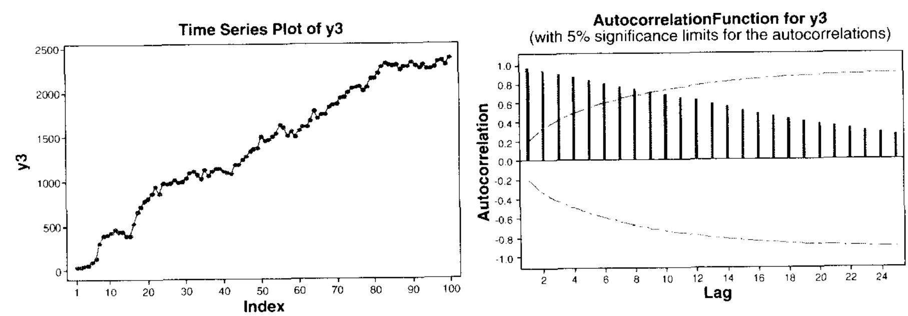
In a crude way, the stationarity of a time series can be determined by taking arbitrary "snapshots" of the process at different points in time and observing the general behavior of the time series. If it exhibits "similar" behavior, one can then proceed with the modeling efforts under the assumption of stationarity. Further preliminary tests also involve observing the behavior of the autocorrelation function. A strong and slowly dying ACF will also suggest deviations from stationarity. Better and more methodological tests of stationarity also exist and we will discuss some of them later in this chapter. Figure 5.1 shows examples of stationary and nonstationary time series data.



(a) $y_{1,t} = 10 + 0.75y_{1,t-1} + \varepsilon_t$



(b) $y_{2,t} = 2 + 0.95y_{2,t-1} + \varepsilon_t$



(c) $y_{3,t} = 20 + y_{3,t-1} + \varepsilon_t$

FIGURE 5.1 Realizations of (a) stationary, (b) near nonstationary, and (c) nonstationary processes.

5.2.2 Stationary Time Series

For a time-invariant and stable linear filter and a stationary input time series x_t with $\mu_x = E(x_t)$ and $\gamma_x(k) = \mathrm{Cov}(x_t, x_{t+k})$, the output time series y_t given in Eq. (5.1) is also a stationary time series with

$$E(y_t) = \mu_y = \sum_{-\infty}^{\infty} \psi_i \mu_x$$

and

$$\mathrm{Cov}(y_t, y_{t+k}) = \gamma_y(k) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i \psi_j \gamma_x(i - j + k)$$

It is then easy to show that the following stable linear process with white noise time series, ε_t, is also stationary:

$$y_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \tag{5.2}$$

where ε_t represents the independent random shocks with $E(\varepsilon_t) = 0$, and

$$\gamma_\varepsilon(h) = \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0 \end{cases}$$

So for the autocovariance function of y_t, we have

$$\begin{aligned} \gamma_y(k) &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \gamma_\varepsilon(i - j + k) \\ &= \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} \end{aligned} \tag{5.3}$$

We can rewrite the linear process in Eq. (5.2) in terms of the **backshift operator**, B, as

$$\begin{aligned} y_t &= \mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots \\ &= \mu + \sum_{i=0}^{\infty} \psi_i B^i \varepsilon_t \\ &= \mu + \underbrace{\left(\sum_{i=0}^{\infty} \psi_i B^i \right)}_{=\Psi(B)} \varepsilon_t \\ &= \mu + \Psi(B) \varepsilon_t \end{aligned} \tag{5.4}$$

This is called the **infinite moving average** and serves as a general class of models for any stationary time series. This is due to a theorem by Wold [1938] and basically states that **any** nondeterministic weakly stationary time series y_t can be represented as in Eq. (5.2), where $\{\psi_i\}$ satisfy $\sum_{i=0}^{\infty} \psi_i^2 < \infty$. A more intuitive interpretation of this theorem is that a stationary time series can be seen as the weighted sum of the present and past random "disturbances." For further explanations see Yule [1927] and Bisgaard and Kulahci [2005]. It can also be seen from Eq. (5.3) that there is a direct relation between the weights $\{\psi_i\}$ and the autocovariance function. In modeling a stationary time series as in Eq. (5.4), it is obviously impractical to attempt to estimate the infinitely many weights given in $\{\psi_i\}$. Although very powerful in providing a general representation of any stationary time series, the infinite moving average model given in Eq. (5.2) is useless in practice except for certain special cases:

1. Finite order moving average (MA) models where, except for a finite number of the weights in $\{\psi_i\}$, they are set to 0.
2. Finite order autoregressive (AR) models, where the weights in $\{\psi_i\}$ are generated using only a finite number of parameters.
3. A mixture of finite order autoregressive and moving average models (ARMA).

We shall now discuss each of these classes of models in great detail.

5.3 FINITE ORDER MOVING AVERAGE (MA) PROCESSES

In finite order moving average or MA models, conventionally ψ_0 is set to 1 and the weights that are not set to 0 are represented by the Greek letter θ with a minus sign in front. Hence a moving average process of order q(MA(q)) is given as

$$y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q} \tag{5.5}$$

where $\{\varepsilon_t\}$ is white noise. Since Eq. (5.5) is a special case of Eq. (5.4) with only finite weights, a MA(q) process is **always** stationary regardless of values of the weights. In terms of the backward shift operator, the MA(q) process is

$$\begin{aligned} y_t &= \mu + \left(1 - \theta_1 B - \cdots - \theta_q B^q\right)\varepsilon_t \\ &= \mu + \left(1 - \sum_{i=1}^{q} \theta_i B^i\right)\varepsilon_t \\ &= \mu + \Theta(B)\,\varepsilon_t \end{aligned} \tag{5.6}$$

where $\Theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i$.

Furthermore, since $\{\varepsilon_t\}$ is white noise, the expected value of the MA(q) process is simply

$$\begin{aligned} E(y_t) &= E\left(\mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}\right) \\ &= \mu \end{aligned} \tag{5.7}$$

and its variance is

$$\begin{aligned} \mathrm{Var}(y_t) = \gamma_y(0) &= \mathrm{Var}\left(\mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}\right) \\ &= \sigma^2\left(1 + \theta_1^2 + \cdots + \theta_q^2\right) \end{aligned} \tag{5.8}$$

Similarly, the autocovariance at lag k can be calculated from

$$\begin{aligned}\gamma_y(k) &= \mathrm{Cov}(y_t, y_{t+k})\\ &= E[(\varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q})(\varepsilon_{t+k} - \theta_1\varepsilon_{t+k-1} - \cdots - \theta_q\varepsilon_{t+k-q})] \quad (5.9)\\ &= \begin{cases}\sigma^2(-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q), & k = 1, 2, \ldots, q\\ 0, \quad k > q\end{cases}\end{aligned}$$

From Eqs. (5.8) and (5.9), the autocovariance function of the MA(q) process is

$$\rho_y(k) = \frac{\gamma_y(k)}{\gamma_y(0)} = \begin{cases}\dfrac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, & k = 1, 2, \ldots, q\\ 0, \quad k > q\end{cases} \tag{5.10}$$

This feature of the ACF is very helpful in identifying the MA model and its appropriate order as it "cuts off" after lag q. In real life applications, however, the sample ACF, $r(k)$, will not necessarily be equal to zero after lag q. It is expected to become very small in absolute value after lag q. For a data set of N observations, this is often tested against $\pm 2/\sqrt{N}$ limits, where $1/\sqrt{N}$ is the approximate value for the standard deviation of the ACF for any lag under the assumption of independence as discussed in Chapter 2.

Note that a more accurate formula for the standard error of the kth sample autocorrelation coefficient is provided by Bartlett [1946] as

$$S(r_k) = N^{-1/2}\left(1 + 2\sum_{j=1}^{k-1} r_j^{*2}\right)^{1/2}$$

where

$$r_j^* = \begin{cases} r_j & \text{for } \rho_j \neq 0\\ 0 & \text{for } \rho_j = 0\end{cases}$$

A special case would be white noise data for which $\rho_j = 0$ for all j's. Hence for a white noise process (i.e., no autocorrelation), a reasonable interval for the sample autocorrelation coefficients to fall in would be $\pm 2/\sqrt{N}$ and any indication otherwise may be considered as evidence for serial dependence in the process.

5.3.1 The First-Order Moving Average Process, MA(1)

The simplest finite order MA model is obtained when $q = 1$ in Eq. (5.5):

$$y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} \tag{5.11}$$

For the first-order moving average or MA(1) model, we have the autocovariance function as

$$\begin{aligned}\gamma_y(0) &= \sigma^2\left(1+\theta_1^2\right)\\ \gamma_y(1) &= -\theta_1\sigma^2\\ \gamma_y(k) &= 0, \quad k>1\end{aligned} \tag{5.12}$$

Similarly, we have the autocorrelation function as

$$\begin{aligned}\rho_y(1) &= \frac{-\theta_1}{1+\theta_1^2}\\ \rho_y(k) &= 0, \quad k>1\end{aligned} \tag{5.13}$$

From Eq. (5.13), we can see that the first lag autocorrelation in MA(1) is bounded as

$$\left|\rho_y(1)\right| = \frac{|\theta_1|}{1+\theta_1^2} \leq \frac{1}{2} \tag{5.14}$$

and the autocorrelation function cuts off after lag 1.

Consider, for example, the following MA(1) model:

$$y_t = 40 + \varepsilon_t + 0.8\varepsilon_{t-1}$$

A realization of this model with its sample ACF is given in Figure 5.2. A visual inspection reveals that the mean and variance remain stable while there are some short runs where successive observations tend to follow each other for very brief durations, suggesting that there is indeed some positive autocorrelation in the data as revealed in the sample ACF plot.

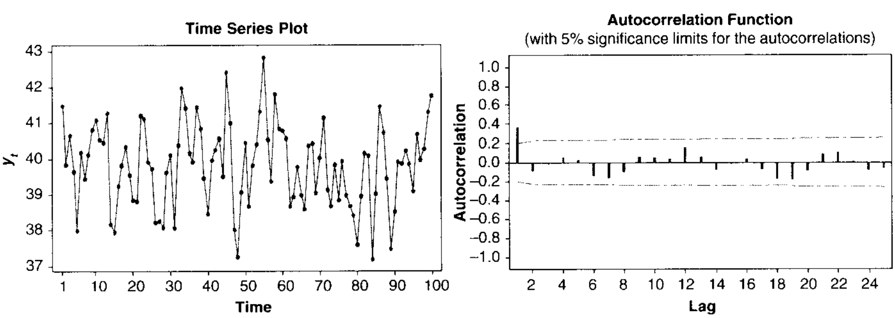


FIGURE 5.2 A realization of the MA(1) process, $y_t = 40 + \varepsilon_t + 0.8\varepsilon_{t-1}$.

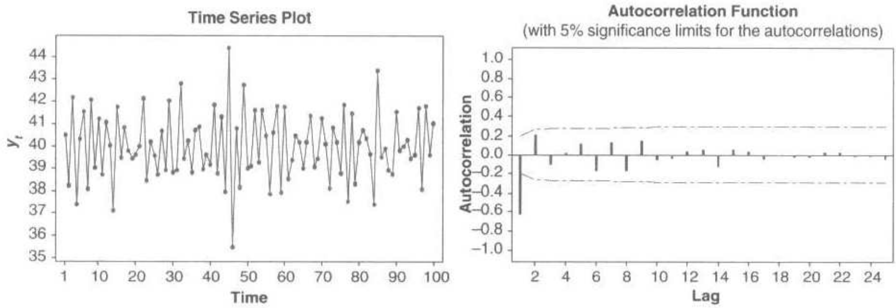


FIGURE 5.3 A realization of the MA(1) process, $y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$.

We can also consider the following model:

$$y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$$

A realization of this model is given in Figure 5.3. We can see that observations tend to oscillate successively. This suggests a negative autocorrelation as confirmed by the sample ACF plot.

5.3.2 The Second-Order Moving Average Process, MA(2)

Another useful finite order moving average process is MA(2), given as

$$\begin{aligned} y_t &= \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} \\ &= \mu + \left(1 - \theta_1 B - \theta_2 B^2\right)\varepsilon_t \end{aligned} \tag{5.15}$$

The autocovariance and autocorrelation functions for the MA(2) model are given as

$$\begin{aligned} \gamma_y(0) &= \sigma^2\left(1 + \theta_1^2 + \theta_2^2\right) \\ \gamma_y(1) &= \sigma^2(-\theta_1 + \theta_1\theta_2) \\ \gamma_y(2) &= \sigma^2(-\theta_2) \\ \gamma_y(k) &= 0, \quad k > 2 \end{aligned} \tag{5.16}$$

and

$$\begin{aligned} \rho_y(1) &= \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2} \\ \rho_y(2) &= \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} \\ \rho_y(k) &= 0, \quad k > 2 \end{aligned} \tag{5.17}$$

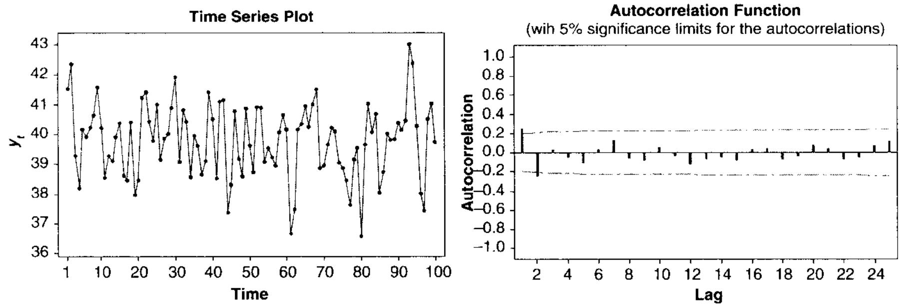


FIGURE 5.4 A realization of the MA(2) process, $y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$.

Figure 5.4 shows the time series plot and the autocorrelation function for a realization of the MA(2) model:

$$y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$$

Note that the sample ACF cuts off after lag 2.

5.4 FINITE ORDER AUTOREGRESSIVE PROCESSES

As mentioned in Section 5.1, while it is quite powerful and important, Wold's decomposition theorem does not help us much in our modeling and forecasting efforts as it implicitly requires the estimation of the infinitely many weights,$\{\psi_i\}$. In Section 5.2 we discussed a special case of this decomposition of the time series by assuming that it can be adequately modeled by only estimating a finite number of weights and setting the rest equal to 0. Another interpretation of the finite order MA processes is that at any given time, of the infinitely many past disturbances, only a finite number of those disturbances "contribute" to the current value of the time series and that the time window of the contributors "moves" in time, making the "oldest" disturbance obsolete for the next observation. It is indeed not too far fetched to think that some processes might have these intrinsic dynamics. However, for some others, we may be required to consider the "lingering" contributions of the disturbances that happened back in the past. This will of course bring us back to square one in terms of our efforts in estimating infinitely many weights. Another solution to this problem is through the autoregressive models in which the infinitely many weights are assumed to follow a distinct pattern and can be successfully represented with only a handful of parameters. We shall now consider some special cases of autoregressive processes.

5.4.1 First-Order Autoregressive Process, AR(1)

Let us first consider again the time series given in Eq. (5.2):

$$\begin{aligned} y_t &= \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \\ &= \mu + \sum_{i=0}^{\infty} \psi_i B^i \varepsilon_t \\ &= \mu + \Psi(B)\varepsilon_t \end{aligned}$$

where $\Psi(B) = \sum_{i=0}^{\infty} \psi_i B^i$. As in the finite order MA processes, one approach to modeling this time series is to assume that the contributions of the disturbances that are way in the past should be small compared to the more recent disturbances that the process has experienced. Since the disturbances are independently and identically distributed random variables, we can simply assume a set of infinitely many weights in descending magnitudes reflecting the diminishing magnitudes of contributions of the disturbances in the past. A simple and yet intuitive set of such weights can be created following an exponential decay pattern. For that we will set $\psi_i = \phi^i$, where $|\phi| < 1$ to guarantee the exponential "decay." In this notation, the weights on the disturbances starting from the current disturbance and going back in past will be $1, \phi, \phi^2, \phi^3, \ldots$ Hence Eq. (5.2) can be written as

$$\begin{aligned} y_t &= \mu + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots \\ &= \mu + \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} \end{aligned} \tag{5.18}$$

From Eq. (5.18), we also have

$$y_{t-1} = \mu + \varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + \cdots \tag{5.19}$$

We can then combine Eqs. (5.18) and (5.19) as

$$\begin{aligned} y_t &= \mu + \varepsilon_t + \underbrace{\phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots}_{=\phi y_{t-1} - \phi\mu} \\ &= \underbrace{\mu - \phi\mu}_{=\delta} + \phi y_{t-1} + \varepsilon_t \\ &= \delta + \phi y_{t-1} + \varepsilon_t \end{aligned} \tag{5.20}$$

where $\delta = (1 - \phi)\mu$. The process in Eq. (5.20) is called a **first-order autoregressive process**, AR(1), because Eq. (5.20) can be seen as a regression of y_t on y_{t-1} and hence the term **auto**regressive process.

The assumption of $|\phi| < 1$ that is made to make the weights decay exponentially in time also guarantees that $\sum_{i=0}^{\infty} |\psi_i| < \infty$. Hence an AR(1) process is **stationary** if $|\phi| < 1$. The mean of a stationary AR(1) process is

$$E(y_t) = \mu = \frac{\delta}{1-\phi} \tag{5.21}$$

The autocovariance function of a stationary AR(1) can be calculated from Eq. (5.18) as

$$\gamma(k) = \sigma^2 \phi^k \frac{1}{1-\phi^2} \quad \text{for } k = 0,\ 1,\ 2, \ldots \tag{5.22}$$

The variance is then given as

$$\gamma(0) = \sigma^2 \frac{1}{1-\phi^2} \tag{5.23}$$

Correspondingly, the autocorrelation function for a stationary AR(1) process is given as

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi^k \quad \text{for } k = 0,\ 1,\ 2,\ \ldots \tag{5.24}$$

Hence the ACF for a stationary AR(1) process has an exponential decay form.

A realization of the following AR(1) model,

$$y_t = 8 + 0.8 y_{t-1} + \varepsilon_t$$

is shown in Figure 5.5. As in the MA(1) model with $\theta = -0.8$, we can observe some short runs during which observations tend to move in the upward or downward

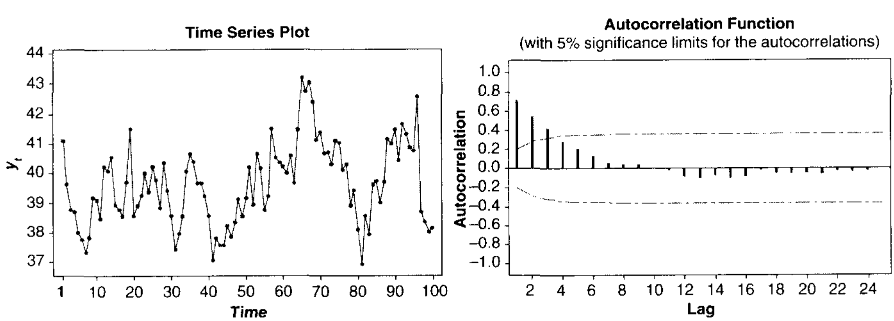


FIGURE 5.5 A realization of the AR(1) process, $y_t = 8 + 0.8y_{t-1} + \varepsilon_t$.

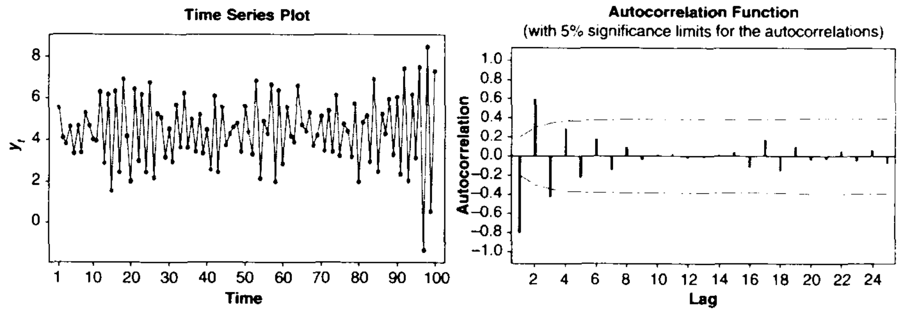


FIGURE 5.6 A realization of the AR(1) process, $y_t = 8 - 0.8y_{t-1} + \varepsilon_t$.

direction. As opposed to the MA(1) model, however, the duration of these runs tends to be longer and the trend tends to linger. This can also be observed in the sample ACF plot.

Figure 5.6 shows a realization of the AR(1) model $y_t = 8 - 0.8y_{t-1} + \varepsilon_t$. We observe that instead of lingering runs, the observations exhibit jittery up/down movements because of the negative ϕ value.

5.4.2 Second-Order Autoregressive Process, AR(2)

In this section, we will first start with the obvious extension of Eq. (5.20) to include the observation y_{t-2} as

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \tag{5.25}$$

We will then show that Eq. (5.25) can be represented in the infinite MA form and provide the conditions of stationarity for y_t in terms of ϕ_1 and ϕ_2. For that we will rewrite Eq. (5.25) as

$$(1 - \phi_1 B - \phi_2 B^2)y_t = \delta + \varepsilon_t \tag{5.26}$$

or

$$\Phi(B)y_t = \delta + \varepsilon_t \tag{5.27}$$

Furthermore, applying $\Phi(B)^{-1}$ to both sides, we obtain

$$\begin{aligned} y_t &= \underbrace{\Phi(B)^{-1}\,\delta}_{=\mu} + \underbrace{\Phi(B)^{-1}}_{=\Psi(B)}\varepsilon_t \\ &= \mu + \Psi(B)\,\varepsilon_t \end{aligned} \tag{5.28}$$

$$= \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

$$= \mu + \sum_{i=0}^{\infty} \psi_i B^i \varepsilon_t$$

where

$$\mu = \Phi(B)^{-1} \delta \tag{5.29}$$

and

$$\Phi(B)^{-1} = \sum_{i=0}^{\infty} \psi_i B^i = \Psi(B) \tag{5.30}$$

We can use Eq. (5.30) to obtain the weights in Eq. (5.28) in terms of ϕ_1 and ϕ_2. For that, we will use

$$\Phi(B)\,\Psi(B) = 1 \tag{5.31}$$

That is,

$$(1 - \phi_1 B - \phi_2 B^2)(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots) = 1$$

or

$$\begin{aligned} \psi_0 + (\psi_1 - \phi_1 \psi_0)\, B + (\psi_2 - \phi_1 \psi_1 - \phi_2 \psi_0)\, B^2 \\ + \cdots + \left(\psi_j - \phi_1 \psi_{j-1} - \phi_2 \psi_{j-2}\right) B^j + \cdots = 1 \end{aligned} \tag{5.32}$$

Since on the right-hand side of the Eq. (5.32) there are no backshift operators, for $\Phi(B)\,\Psi(B) = 1$, we need

$$\begin{aligned} \psi_0 &= 1 \\ (\psi_1 - \phi_1 \psi_0) &= 0 \\ (\psi_j - \phi_1 \psi_{j-1} - \phi_2 \psi_{j-2}) &= 0 \quad \text{for all } j = 2, 3, \ldots \end{aligned} \tag{5.33}$$

The equations in (5.33) can indeed be solved for each ψ_j in a futile attempt to estimate infinitely many parameters. However, it should be noted that the ψ_j in Eq. (5.33) satisfy the second-order linear difference equation and that they can be expressed as the solution to this equation in terms of the two roots m_1 and m_2 of the associated polynomial

$$m^2 - \phi_1 m - \phi_2 = 0 \tag{5.34}$$

If the roots obtained by

$$m_1, m_2 = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

satisfy $|m_1|, |m_2| < 1$, then we have $\sum_{i=0}^{\infty} |\psi_i| < \infty$. Hence if the roots m_1 and m_2 are both less than 1 in absolute value, then the AR(2) model is stationary. Note that if the roots of Eq. (5.34) are complex conjugates of the form $a \pm ib$, the condition for stationarity is that $\sqrt{a^2 + b^2} < 1$.

Furthermore, under the condition that $|m_1|, |m_2| < 1$, the AR(2) time series, $\{y_t\}$, has an infinite MA representation as in Eq. (5.28).

Now that we have established the conditions for the stationarity of an AR(2) time series, let us now consider its mean, autocovariance, and autocorrelation functions. From Eq. (5.25), we have

$$\begin{aligned} E(y_t) &= \delta + \phi_1 E(y_{t-1}) + \phi_2 E(y_{t-2}) + 0 \\ \mu &= \delta + \phi_1 \mu + \phi_2 \mu \\ \Rightarrow \mu &= \frac{\delta}{1 - \phi_1 - \phi_2} \end{aligned} \tag{5.35}$$

Note that for $1 - \phi_1 - \phi_2 = 0, m = 1$ is one of the roots for the associated polynomial in Eq. (5.34) and hence the time series is deemed nonstationary. The autocovariance function is

$$\begin{aligned} \gamma(k) &= \mathrm{Cov}(y_t, y_{t-k}) \\ &= \mathrm{Cov}(\delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, y_{t-k}) \\ &= \phi_1 \mathrm{Cov}(y_{t-1}, y_{t-k}) + \phi_2 \mathrm{Cov}(y_{t-2}, y_{t-k}) + \mathrm{Cov}(\varepsilon_t, y_{t-k}) \\ &= \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + \begin{cases} \sigma^2 & \text{if } k = 0 \\ 0 & \text{if } k > 0 \end{cases} \end{aligned} \tag{5.36}$$

Thus $\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma^2$ and

$$\gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2), \quad k = 1, 2, \ldots \tag{5.37}$$

The equations in (5.37) are called the **Yule–Walker** equations for $\gamma(k)$. Similarly, we can obtain the autocorrelation function by dividing Eq. (5.37) by $\gamma(0)$:

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2), \quad k = 1, 2, \ldots \tag{5.38}$$

The Yule–Walker equations for $\rho(k)$ in Eq. (5.38) can be solved recursively as

$$\begin{aligned}
\rho(1) &= \phi_1 \underbrace{\rho(0)}_{=1} + \phi_2 \underbrace{\rho(-1)}_{=\rho(1)} \\
&= \frac{\phi_1}{1-\phi_2} \\
\rho(2) &= \phi_1 \rho(1) + \phi_2 \\
\rho(3) &= \phi_1 \rho(2) + \phi_2 \rho(1) \\
&\vdots
\end{aligned}$$

A general solution can be obtained through the roots m_1 and m_2 of the associated polynomial $m^2 - \phi_1 m - \phi_2 = 0$. There are three cases.

Case 1. If m_1 and m_2 are distinct, real roots, we then have

$$\rho(k) = c_1 m_1^k + c_2 m_2^k, \quad k = 0, 1, 2, \ldots \tag{5.39}$$

where c_1 and c_2 are particular constants and can, for example, be obtained from $\rho(0)$ and $\rho(1)$. Moreover, since for stationarity we have $|m_1|, |m_2| < 1$, in this case, the autocorrelation function is a **mixture of two exponential decay terms**.

Case 2. If m_1 and m_2 are complex conjugates in the form of $a \pm ib$, we then have

$$\rho(k) = R^k \left[c_1 \cos(\lambda k) + c_2 \sin(\lambda k)\right], \quad k = 0, 1, 2, \ldots \tag{5.40}$$

where $R = |m_i| = \sqrt{a^2 + b^2}$ and λ is determined by $\cos(\lambda) = a/R$, $\sin(\lambda) = b/R$. Hence we have $a \pm ib = R\left[\cos(\lambda) \pm i \sin(\lambda)\right]$. Once again c_1 and c_2 are particular constants. The ACF in this case has the form of a **damped sinusoid**, with damping factor R and frequency λ; that is, the period is $2\pi/\lambda$.

Case 3. If there is one real root m_0, $m_1 = m_2 = m_0$, we then have

$$\rho(k) = (c_1 + c_2 k) m_0^k \quad k = 0, 1, 2, \ldots \tag{5.41}$$

In this case, the ACF will exhibit an exponential decay pattern.

In case 1, for example, an AR(2) model can be seen as an "adjusted" AR(1) model for which a single exponential decay expression as in the AR(1) model is not enough to describe the pattern in the ACF, and hence an additional exponential decay expression is "added" by introducing the second lag term, y_{t-2}.

Figure 5.7 shows a realization of the AR(2) process

$$y_t = 4 + 0.4 y_{t-1} + 0.5 y_{t-2} + \varepsilon_t$$

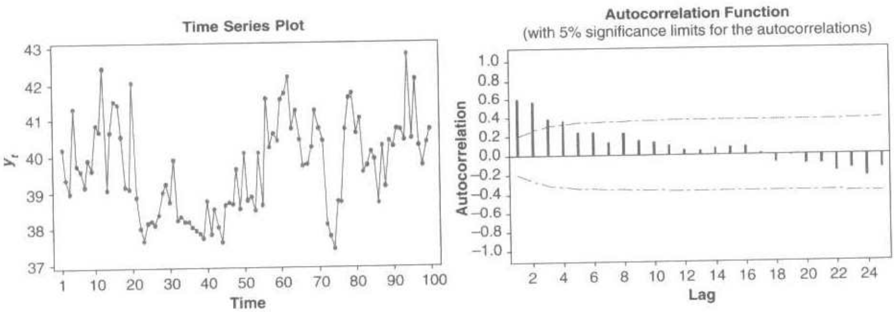


FIGURE 5.7 A realization of the AR(2) process, $y_t = 4 + 0.4y_{t-1} + 0.5y_{t-2} + \varepsilon_t$.

Note that the roots of the associated polynomial of this model are real. Hence the ACF is a mixture of two exponential decay terms.

Similarly, Figure 5.8 shows a realization of the following AR(2) process

$$y_t = 4 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_t.$$

For this process, the roots of the associated polynomial are complex conjugates. Therefore the ACF plot exhibits a damped sinusoid behavior.

5.4.3 General Autoregressive Process, AR(p)

From the previous two sections, a general, pth-order AR model is given as

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{5.42}$$

where ε_t is white noise. Another representation of Eq. (5.42) can be given as

$$\Phi(B) y_t = \delta + \varepsilon_t \tag{5.43}$$

where $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$.

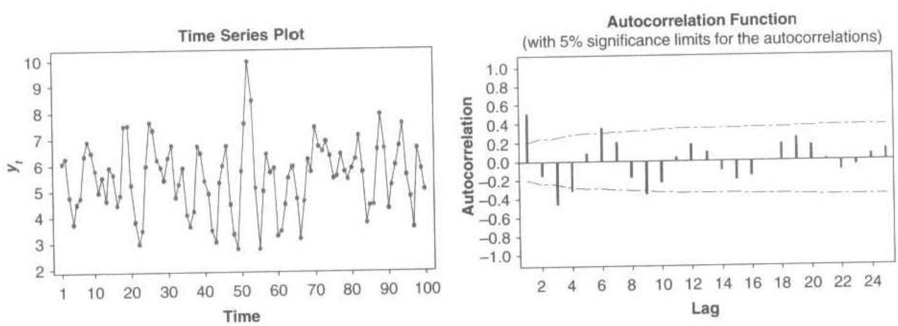


FIGURE 5.8 A realization of the AR(2) process, $y_t = 4 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_t$.

The AR(p) time series $\{y_t\}$ in Eq. (5.42) is stationary if the roots of the associated polynomial

$$m^p - \phi_1 m^{p-1} - \phi_2 m^{p-2} - \cdots - \phi_p = 0 \tag{5.44}$$

are less than one in absolute value. Furthermore, under this condition, the AR(p) time series $\{y_t\}$ is also said to have an **absolutely summable** infinite MA representation

$$y_t = \mu + \Psi(B)\,\varepsilon_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \tag{5.45}$$

where $\Psi(B) = \Phi(B)^{-1}$ with $\sum_{i=0}^{\infty} |\psi_i| < \infty$.

As in AR(2), the weights of the random shocks in Eq. (5.45) can be obtained from $\Phi(B)\,\Psi(B) = 1$ as

$$\begin{aligned} \psi_j &= 0, \quad j < 0 \\ \psi_0 &= 1 \\ \psi_j - \phi_1\psi_{j-1} - \phi_2\psi_{j-2} - \cdots - \phi_p\psi_{j-p} &= 0 \quad \text{for all } j = 1, 2, \ldots \end{aligned} \tag{5.46}$$

We can easily show that, for stationary AR(p),

$$E(y_t) = \mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}$$

and

$$\begin{aligned} \gamma(k) &= \operatorname{Cov}(y_t, y_{t-k}) \\ &= \operatorname{Cov}(\delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, y_{t-k}) \\ &= \sum_{i=1}^{p} \phi_i \operatorname{Cov}(y_{t-i}, y_{t-k}) + \operatorname{Cov}(\varepsilon_t, y_{t-k}) \\ &= \sum_{i=1}^{p} \phi_i \gamma(k-i) + \begin{cases} \sigma^2 & \text{if } k = 0 \\ 0 & \text{if } k > 0 \end{cases} \end{aligned} \tag{5.47}$$

Thus we have

$$\gamma(0) = \sum_{i=1}^{p} \phi_i \gamma(i) + \sigma^2 \tag{5.48}$$

$$\Rightarrow \gamma(0)\left[1 - \sum_{i=1}^{p} \phi_i \rho(i)\right] = \sigma^2 \tag{5.49}$$

By dividing Eq. (5.47) by $\gamma(0)$ for $k > 0$, it can be observed that the ACF of an AR(p) process satisfies the Yule–Walker equations

$$\rho(k) = \sum_{i=1}^{p} \phi_i \rho(k-i), \quad k = 1, 2, \ldots \tag{5.50}$$

The equations in (5.50) are pth-order **linear difference equations**, implying that the ACF for an AR(p) model can be found through the p roots of the associated polynomial in Eq. (5.44). For example, if the roots are all distinct and real, we have

$$\rho(k) = c_1 m_1^k + c_2 m_2^k + \cdots + c_p m_p^k, \quad k = 1, 2, \ldots \tag{5.51}$$

where $c_1, c_2, \ldots, c_p$ are particular constants. However, in general, the roots may not all be distinct or real. Thus the ACF of an AR(p) process can be a **mixture of exponential decay and damped sinusoid** expressions depending on the roots of Eq. (5.44).

5.4.4 Partial Autocorrelation Function, PACF

In Section 5.2, we saw that the ACF is an excellent tool in identifying the order of an MA(q) process, because it is expected to "cut off" after lag q. However, in the previous section, we pointed out that the ACF is not as useful in the identification of the order of an AR(p) process for which it will most likely have a mixture of exponential decay and damped sinusoid expressions. Hence such behavior, while indicating that the process might have an AR structure, fails to provide further information about the order of such structure. For that, we will define and employ the **partial autocorrelation function** (PACF) of the time series. But before that, we discuss the concept of partial correlation to make the interpretation of the PACF easier.

Partial Correlation

Consider three random variables X, Y, and Z. Then consider simple linear regression of X on Z and Y on Z as

$$\hat{X} = a_1 + b_1 Z \quad \text{where } b_1 = \frac{\text{Cov}(Z, X)}{\text{Var}(Z)}$$

and

$$\hat{Y} = a_2 + b_2 Z \quad \text{where } b_2 = \frac{\text{Cov}(Z, Y)}{\text{Var}(Z)}$$

Then the errors can be obtained from

$$X^* = X - \hat{X} = X - (a_1 + b_1 Z)$$

and

$$Y^* = Y - \hat{Y} = Y - (a_2 + b_2 Z)$$

Then the **partial correlation** between X and Y after adjusting for Z is defined as the correlation between X^* and Y^*; $\text{corr}(X^*, Y^*) = \text{corr}(X - \hat{X}, Y - \hat{Y})$. That is, partial correlation can be seen as the correlation between two variables after being adjusted for a common factor that may be affecting them. The generalization is of course possible by allowing for adjustment for more than just one factor.

Partial Autocorrelation Function

Following the above definition, the **partial autocorrelation function** between y_t and y_{t-k} is the autocorrelation between y_t and y_{t-k} after adjusting for y_{t-1}, $y_{t-2}, \ldots, y_{t-k+1}$. Hence for an AR($p$) model the partial autocorrelation function between y_t and y_{t-k} for $k > p$ should be equal to zero. A more formal definition can be found below.

Consider a stationary time series model $\{y_t\}$ that is not necessarily an AR process. Further consider, for any fixed value of k, the Yule–Walker equations for the ACF of an AR(p) process given in Eq. (5.50) as

$$\rho(j) = \sum_{i=1}^{k} \phi_{ik}\rho(j-i), \quad j = 1, 2, \ldots, k \tag{5.52}$$

or

$$\begin{aligned}
\rho(1) &= \phi_{1k} + \phi_{2k}\rho(1) + \cdots + \phi_{kk}\rho(k-1) \\
\rho(2) &= \phi_{1k}\rho(1) + \phi_{2k} + \cdots + \phi_{kk}\rho(k-2) \\
&\vdots \\
\rho(k) &= \phi_{1k}\rho(k-) + \phi_{2k}\rho(k-2) + \cdots + \phi_{kk}
\end{aligned}$$

Hence we can write the equations in (5.52) in matrix notation as

$$\begin{bmatrix} 1 & \rho(1) & \rho(2) & \ldots & \rho(k-1) \\ \rho(1) & 1 & \rho(3) & \ldots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \ldots & \rho(k-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \ldots & 1 \end{bmatrix} \begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \phi_{3k} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(k) \end{bmatrix} \tag{5.53}$$

or

$$\mathbf{P}_k \phi_k = \boldsymbol{\rho}_k \tag{5.54}$$

where

$$\mathbf{P}_k = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(k-1) \\ \rho(1) & 1 & \rho(3) & \dots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(k-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \dots & 1 \end{bmatrix},$$

$$\phi_k = \begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \phi_{3k} \\ \vdots \\ \phi_{kk} \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\rho}_k = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(k) \end{bmatrix}$$

Thus to solve for ϕ_k, we have

$$\phi_k = \mathbf{P}_k^{-1}\boldsymbol{\rho}_k \tag{5.55}$$

For any given k, $k = 1, 2, \dots$, the last coefficient ϕ_{kk} is called the partial autocorrelation of the process at lag k. Note that for an AR(p) process $\phi_{kk} = 0$ for $k > p$. Hence we say that the PACF cuts off after lag p for an AR(p). This suggests that the PACF can be used in identifying the order of an AR process similar to how the ACF can be used for an MA process.

For sample calculations, $\hat{\phi}_{kk}$, the sample estimate of ϕ_{kk}, is obtained by using the sample ACF, $r(k)$. Furthermore, in a sample of N observations from an AR(p) process, $\hat{\phi}_{kk}$ for $k > p$ is approximately normally distributed with

$$E(\hat{\phi}_{kk}) \approx 0 \quad \text{and} \quad \text{Var}(\hat{\phi}_{kk}) \approx \frac{1}{N} \tag{5.56}$$

Hence the 95% limits to judge whether any $\hat{\phi}_{kk}$ is statistically significantly different from zero are given by $\pm 2/\sqrt{N}$. For further detail see Quenouille [1949], Jenkins [1954, 1956], and Daniels [1956].

Figure 5.9 shows the sample PACFs of the models we have considered so far. In Figure 5.9a we have the sample PACF of the realization of the MA(1) model with $\theta = 0.8$ given in Figure 5.3. It exhibits an exponential decay pattern. Figure 9b shows the sample PACF of the realization of the MA(2) model in Figure 5.4 and it also has an exponential decay pattern in absolute value since for this model the roots of the associated polynomial are real. Figures 5.9c and 5.9d show the sample PACFs of the realization of the AR(1) model with $\phi = 0.8$ and $\phi = -0.8$, respectively. In both cases the PACF "cuts off" after the first lag. That is, the only significant sample PACF value is at lag 1, suggesting that the AR(1) model is indeed appropriate to fit the data. Similarly, in Figures 5.9e and 5.9f, we have the sample PACFs of the realizations of the AR(2) model. Note that the sample PACF cuts off after lag 2.

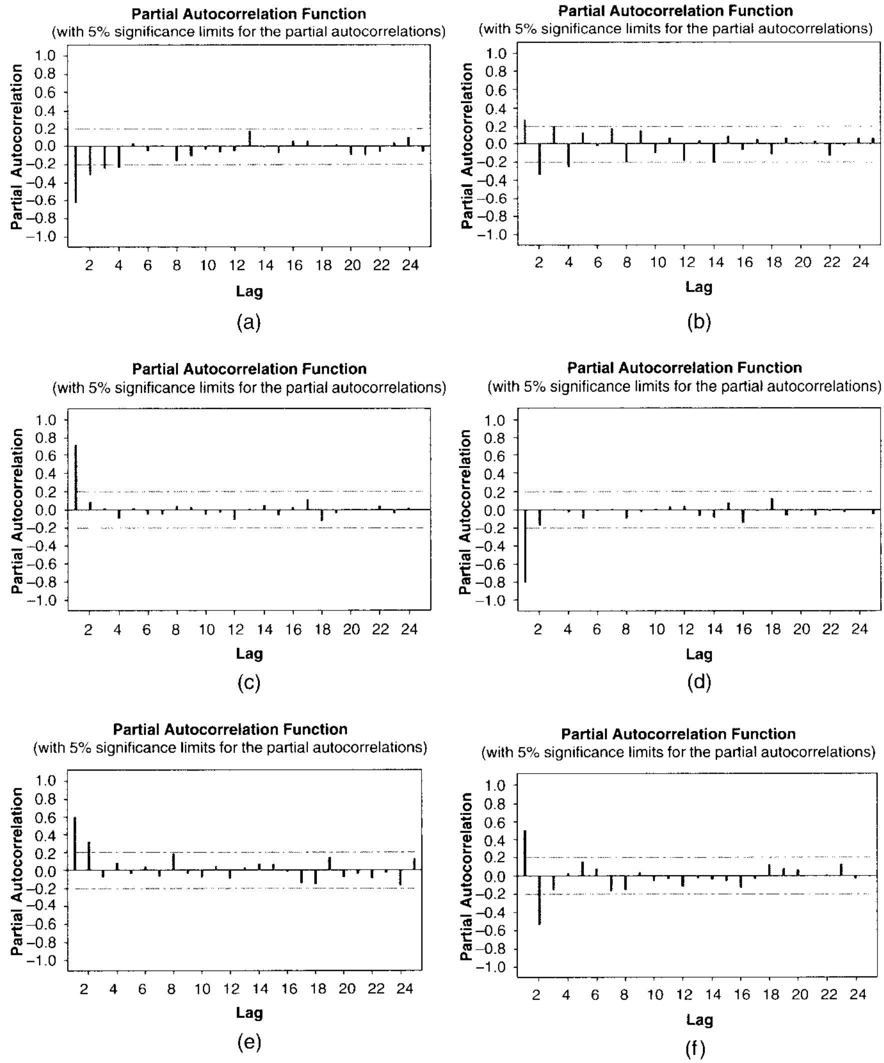


FIGURE 5.9 Partial autocorrelation functions for the realizations of (a) MA(1) process, $y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$; (b) MA(2) process, $y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$; (c) AR(1) process, $y_t = 8 + 0.8y_{t-1} + \varepsilon_t$; (d) AR(1) process, $y_t = 8 - 0.8y_{t-1} + \varepsilon_t$; (e) AR(2) process, $y_t = 4 + 0.4y_{t-1} + 0.5y_{t-2} + \varepsilon_t$; and (f) AR(2) process, $y_t = 4 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_t$.

Invertibility of MA Models

In the previous section we showed that the PACF "cuts off" after lag p for an AR(p). The PACF of an MA(q) model, however, exhibits a more complicated pattern. For that we define an **invertible** moving average process as the following: the MA(q) process in Eq. (5.5) is said to be invertible if it has an absolutely summable infinite AR representation.

Consider the MA(q) process

$$y_t = \mu + \left(1 - \sum_{i=1}^{q} \theta_i B^i\right) \varepsilon_t$$
$$= \mu + \Theta(B)\varepsilon_t$$

After multiplying both sides with $\Theta(B)^{-1}$, we have

$$\begin{aligned} \Theta(B)^{-1} y_t &= \Theta(B)^{-1} \mu + \varepsilon_t \\ \prod(B) y_t &= \delta + \varepsilon_t \end{aligned} \tag{5.57}$$

where $\Pi(B) = 1 - \sum_{i=1}^{\infty} \pi_i B^i = \Theta(B)^{-1}$ and $\Theta(B)^{-1} \mu = \delta$. Hence the infinite AR representation of an MA(q) process is given as

$$y_t - \sum_{i=1}^{\infty} \pi_i y_{t-i} = \delta + \varepsilon_t \tag{5.58}$$

with $\sum_{i=1}^{\infty} |\pi_i| < \infty$. The π_i can be determined from

$$(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)(1 - \pi_1 B - \pi_2 B^2 + \cdots) = 1 \tag{5.59}$$

which in turn yields

$$\begin{aligned} \pi_1 + \theta_1 &= 0 \\ \pi_2 - \theta_1 \pi_1 + \theta_2 &= 0 \\ &\vdots \\ \pi_j - \theta_1 \pi_{j-1} - \cdots - \theta_q \pi_{j-q} &= 0 \end{aligned} \tag{5.60}$$

with $\pi_0 = -1$ and $\pi_j = 0$ for $j < 0$. Hence as in the previous arguments for the stationarity of AR(p) models, the π_i are the solutions to the qth-order linear difference equations and therefore the condition for the invertibility of an MA(q) process turns out to be very similar to the stationarity condition of an AR(p) process: the roots of the associated polynomial given in Eq. (5.60) should be less than 1 in absolute value,

$$m^q - \theta_1 m^{q-1} - \theta_2 m^{q-2} - \cdots - \theta_q = 0 \tag{5.61}$$

An invertible MA(q) process can then be written as an infinite AR process.

Correspondingly, for such a process, adjusting for $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$ does not necessarily eliminate the correlation between y_t and y_{t-k} and therefore its PACF will never "cut off." In general, the PACF of an MA(q) process is a **mixture of exponential decay and damped sinusoid** expressions.

The ACF and the PACF do have very distinct and indicative properties for MA and AR models, respectively. Therefore, in model identification, we strongly recommend the use of both the sample ACF and the sample PACF **simultaneously**.

5.5 MIXED AUTOREGRESSIVE–MOVING AVERAGE (ARMA) PROCESSES

In the previous sections we have considered special cases of Wold's decomposition of a stationary time series represented as a weighted sum of infinite random shocks. In an AR(1) process, for example, the weights in the infinite sum are forced to follow an exponential decay form with ϕ as the rate of decay. Since there are no restrictions apart from $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ on the weights (ψ_i), it may not be possible to approximate them by an exponential decay pattern. For that, we will need to increase the order of the AR model to approximate any pattern that these weights may in fact be exhibiting. On some occasions, however, it is possible to make simple adjustments to the exponential decay pattern by adding only a few terms and hence to have a more parsimonious model. Consider, for example, that the weights ψ_i do indeed exhibit an exponential decay pattern with a constant rate except for the fact that ψ_1 is not equal to this rate of decay as it would be in the case of an AR(1) process. Hence instead of increasing the order of the AR model to accommodate for this "anomaly," we can add an MA(1) term that will simply adjust ψ_1 while having no effect on the rate of exponential decay pattern of the rest of the weights. This results in a mixed **autoregressive moving average** or ARMA(1,1) model. In general, an ARMA(p, q) model is given as

$$\begin{aligned} y_t &= \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \\ &= \delta + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \end{aligned} \tag{5.62}$$

or

$$\Phi(B)\, y_t = \delta + \Theta(B)\, \varepsilon_t \tag{5.63}$$

where ε_t is a white noise process.

Stationarity of ARMA (p, q) Process

The **stationarity** of an ARMA process is related to the AR component in the model and can be checked through the roots of the associated polynomial

$$m^p - \phi_1 m^{p-1} - \phi_2 m^{p-2} - \cdots - \phi_p = 0. \tag{5.64}$$

If all the roots of Eq. (5.64) are less than one in absolute value, then ARMA(p, q) is stationary. This also implies that, under this condition, ARMA(p, q) has an infinite

MA representation as

$$y_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \mu + \Psi(B)\varepsilon_t \tag{5.65}$$

with $\Psi(B) = \Phi(B)^{-1}\,\Theta(B)$. The coefficients in $\Psi(B)$ can be found from

$$\psi_i - \phi_1\psi_{i-1} - \phi_2\psi_{i-2} - \cdots - \phi_p\psi_{i-p} = \begin{cases} -\theta_i, & i = 1, \ldots, q \\ 0, & i > q \end{cases} \tag{5.66}$$

and $\psi_0 = 1$.

Invertibility of ARMA (p, q) Process

Similar to the stationarity condition, the **invertibility** of an ARMA process is related to the MA component and can be checked through the roots of the associated polynomial

$$m^q - \theta_1 m^{q-1} - \theta_2 m^{q-2} - \cdots - \theta_q = 0 \tag{5.67}$$

If all the roots of Eq. (5.65) are less than one in absolute value, then ARMA(p, q) is said to be invertible and has an infinite AR representation,

$$\Pi(B)y_t = \alpha + \varepsilon_t \tag{5.68}$$

where $\alpha = \Theta(B)^{-1}\,\delta$ and $\Pi(B) = \Theta(B)^{-1}\,\Phi(B)$. The coefficients in $\Pi(B)$ can be found from

$$\pi_i - \theta_1\pi_{i-1} - \theta_2\pi_{i-2} - \cdots - \theta_q\pi_{i-q} = \begin{cases} \phi_i, & i = 1, \ldots, p \\ 0, & i > p \end{cases} \tag{5.69}$$

and $\pi_0 = -1$.

In Figure 5.10 we provide realizations of two ARMA(1,1) models:

$$y_t = 16 + 0.6y_{t-1} + \varepsilon_t + 0.8\varepsilon_{t-1} \quad \text{and} \quad y_t = 16 - 0.7y_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1}.$$

Note that the sample ACFs and PACFs exhibit exponential decay behavior (sometimes in absolute value depending on the signs of the AR and MA coefficients).

ACF and PACF of ARMA(p,q) Process

As in the stationarity and invertibility conditions, the ACF and PACF of an ARMA process are determined by the AR and MA components, respectively. It can therefore be shown that the ACF and PACF of an ARMA(p, q) both exhibit exponential decay and/or damped sinusoid patterns, which makes the identification of the order of the ARMA(p, q) model relatively more difficult. For that, additional sample functions such as the Extended Sample ACF (ESACF), the Generalized Sample PACF

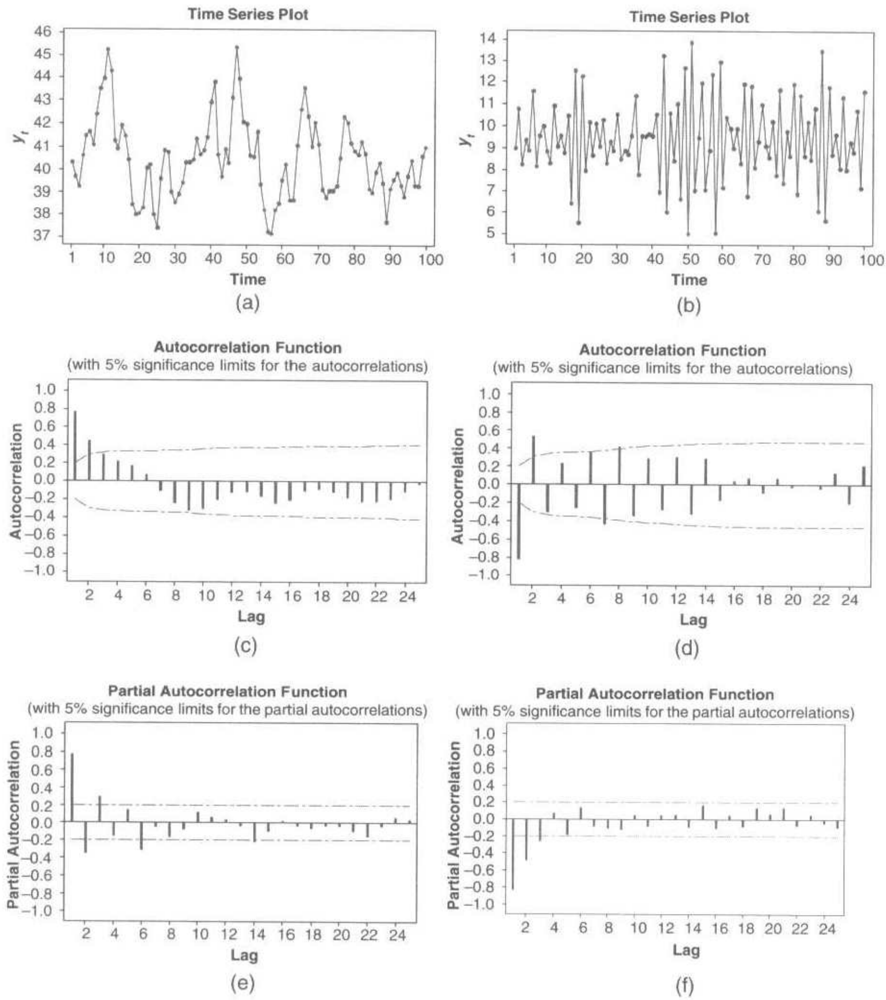


FIGURE 5.10 Two realizations of the ARMA(1,1) model: (a) $y_t = 16 + 0.6y_{t-1} + \varepsilon_t + 0.8\varepsilon_{t-1}$ and (b) $y_t = 16 - 0.7y_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1}$. (c) The ACF of (a), (d) the ACF of (b), (e) the PACF of (a), and (f) the PACF of (b).

(GPACF), the Inverse ACF (IACF), and canonical correlations can be used. For further information see Box, Jenkins, and Reinsel [1994], Wei [2006], Tiao and Box [1981], Tsay and Tiao [1984], and Abraham and Ledolter [1984]. However, the availability of sophisticated statistical software packages such as Minitab, JMP, and SAS makes it possible for the practitioner to consider several different models with various orders and compare them based on the model selection criteria such as AIC, AICC, and SIC as described in Chapter 2 and residual analysis.

The theoretical values of the ACF and PACF for stationary time series are summarized in Table 5.1. The summary of the sample ACFs and PACFs of the realizations

TABLE 5.1 Behavior of Theoretical ACF and PACF for Stationary Processes

Model	ACF	PACF
MA(q)	Cuts off after lag q	Exponential decay and/or damped sinusoid
AR(p)	Exponential decay and/or damped sinusoid	Cuts off after lag p
ARMA(p,q)	Exponential decay and/or damped sinusoid	Exponential decay and/or damped sinusoid

of some of the models we have covered in this chapter are given in Table 5.2, Table 5.3, and Table 5.4 for MA, AR, and ARMA models, respectively.

5.6 NONSTATIONARY PROCESSES

It is often the case that while the processes may not have a constant level, they exhibit homogeneous behavior over time. Consider, for example, the linear trend process given in Figure 5.1c. It can be seen that different snapshots taken in time do exhibit similar behavior except for the main level of the process. Similarly, processes may show nonstationarity in the slope as well. We will call a time series, y_t, homogeneous, nonstationary if it is not stationary but its first difference, that is, $w_t = y_t - y_{t-1} = (1 - B)\, y_t$, or higher-order differences, $w_t = (1 - B)^d\, y_t$, produce a stationary time series. We will further call y_t an **autoregressive integrated moving average** (ARIMA) process of orders p, d, and q—that is, ARIMA(p, d, q)—if its dth difference, denoted by $w_t = (1 - B)^d\, y_t$, produces a stationary ARMA(p, q) process. The term integrated is used since, for $d = 1$, for example, we can write y_t as the sum (or "integral") of the w_t process as

$$
\begin{aligned}
y_t &= w_t + y_{t-1} \\
&= w_t + w_{t-1} + y_{t-2} \\
&= w_t + w_{t-1} + \cdots + w_1 + y_0
\end{aligned}
\tag{5.70}
$$

Hence an ARIMA(p, d, q) can be written as

$$\Phi(B)(1 - B)^d\, y_t = \delta + \Theta(B)\, \varepsilon_t \tag{5.71}$$

Thus once the differencing is performed and a stationary time series $w_t = (1 - B)^d\, y_t$ is obtained, the methods provided in the previous sections can be used to obtain the full model. In most applications first differencing ($d = 1$) and occasionally second differencing ($d = 2$) would be enough to achieve stationarity. However, sometimes transformations other than differencing are useful in reducing a nonstationary time series to a stationary one. For example, in many economic time series the variability of the observations increases as the average level of the process increases; however, the percentage of change in the observations is relatively independent of level. Therefore taking the logarithm of the original series will be useful in achieving stationarity.

Some Examples of ARIMA(p, d, q) Processes

The **random walk process, ARIMA(0, 1, 0)** is the simplest nonstationary model. It is given by

$$(1 - B)y_t = \delta + \varepsilon_t \tag{5.72}$$

suggesting that first differencing eliminates all serial dependence and yields a white noise process.

Consider the process $y_t = 20 + y_{t-1} + \varepsilon_t$. A realization of this process together with its sample ACF and PACF are given in Figure 5.11a–c. We can see that the sample ACF dies out very slowly, while the sample PACF is only significant at the first lag. Also note that the PACF value at the first lag is very close to one. All this evidence suggests that the process is not stationary. The first difference, $w_t = y_t - y_{t-1}$, and its sample ACF and PACF are shown in Figure 5.11d–f. The time series plot of w_t implies that the first difference is stationary. In fact, the sample ACF and PACF do not show any significant values. This further suggests that differencing the original data once "clears out" the autocorrelation. Hence the data can be modeled using the random walk model given in Eq. (5.72).

The **ARIMA(0, 1, 1) process** is given by

$$(1 - B)y_t = \delta + (1 - \theta B)\varepsilon_t \tag{5.73}$$

The infinite AR representation of Eq. (5.73) can be obtained from Eq. (5.69)

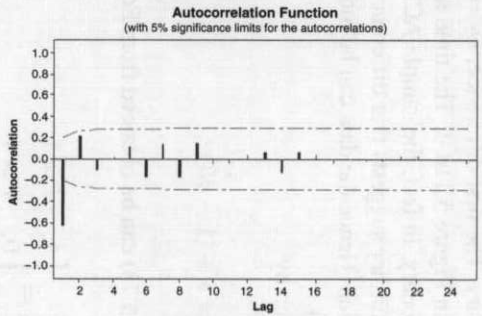
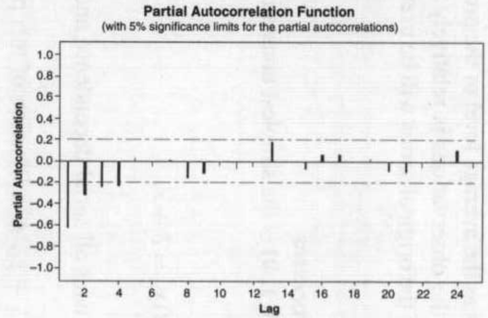
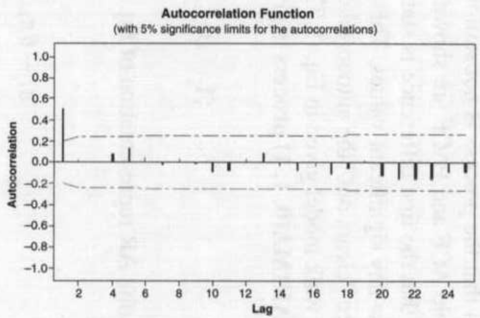
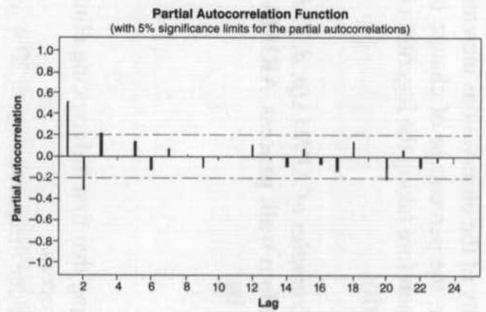
$$\pi_i - \theta\pi_{i-1} = \begin{cases} 1, & i = 1 \\ 0, & i > 1 \end{cases} \tag{5.74}$$

with $\pi_0 = -1$. Thus we have

$$\begin{aligned} y_t &= \alpha + \sum_{i=1}^{\infty} \pi_i y_{t-i} + \varepsilon_t \\ &= \alpha + (1 - \theta)(y_{t-1} + \theta y_{t-2} + \cdots) + \varepsilon_t \end{aligned} \tag{5.75}$$

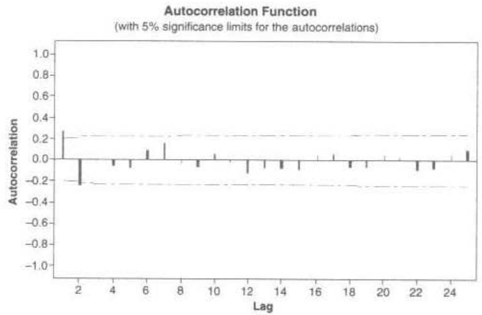
This suggests that an ARIMA(0, 1, 1) (a.k.a. IMA(1, 1)) can be written as an exponentially weighted moving average (EWMA) of all past values.

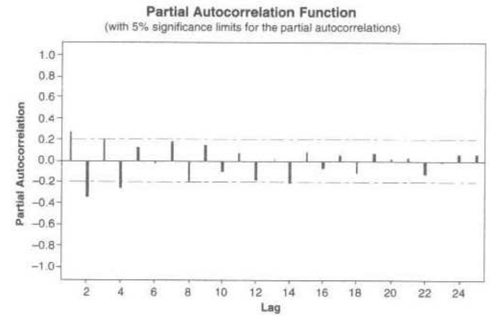
TABLE 5.2 Sample ACFs and PACFs for Some Realizations of MA(1) and MA(2) Models

<table>
<tr><th>Model</th><th>Sample ACF</th><th>Sample PACF</th></tr>
<tr><td>MA(1)
$y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$</td><td></td><td></td></tr>
<tr><td>$y_t = 40 + \varepsilon_t + 0.8\varepsilon_{t-1}$</td><td></td><td></td></tr>
</table>

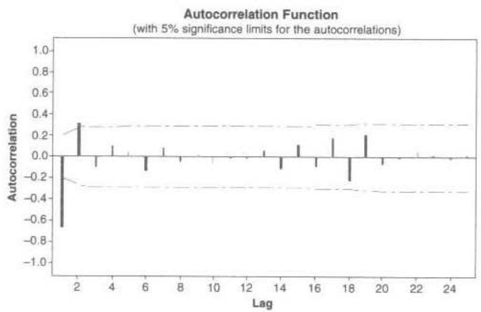
MA(2)

$$y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$$





$$y_t = 40 + \varepsilon_t - 1.1\varepsilon_{t-1} + 0.8\varepsilon_{t-2}$$



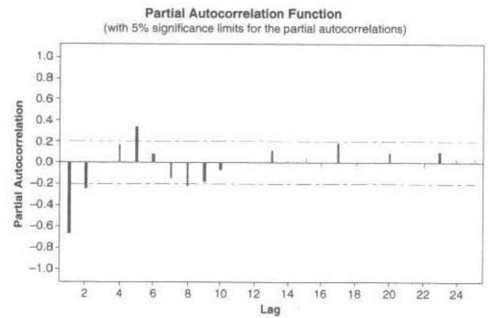
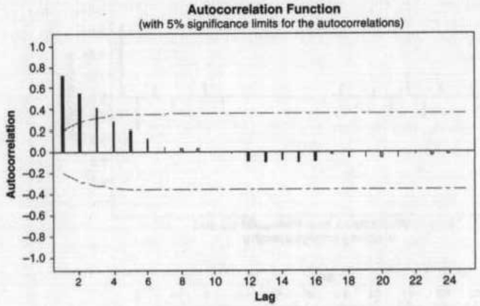
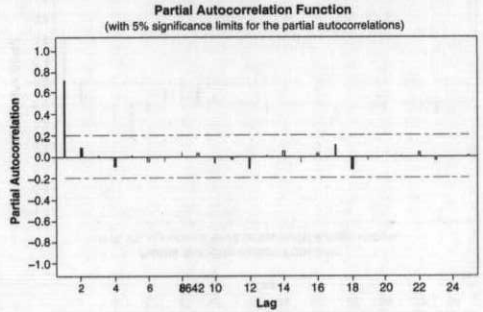
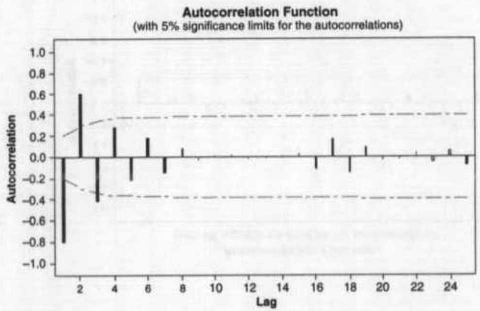
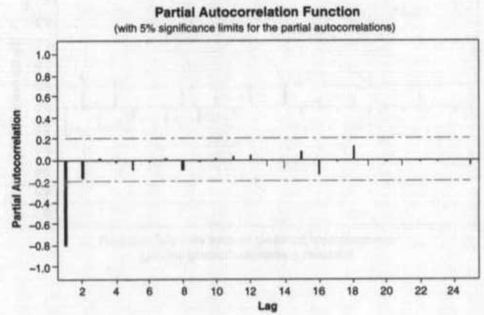
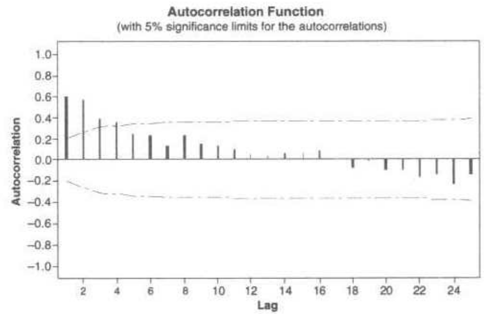


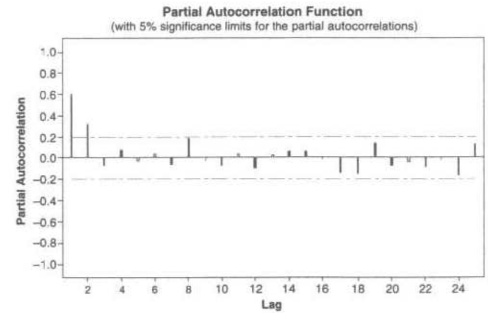
TABLE 5.3 Sample ACFs and PACFs for Some Realizations of AR(1) and AR(2) Models

Model	Sample ACF	Sample PACF
AR(1) $y_t = 8 + 0.8y_{t-1} + \varepsilon_t$		
$y_t = 8 - 0.8y_{t-1} + \varepsilon_t$		

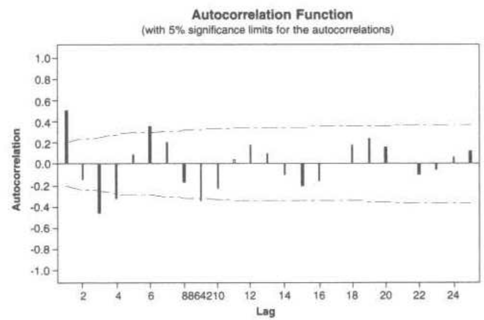
AR(2)

$$y_t = 4 + 0.4y_{t-1} + 0.5y_{t-2} + \varepsilon_t$$





$$y_t = 4 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_t$$



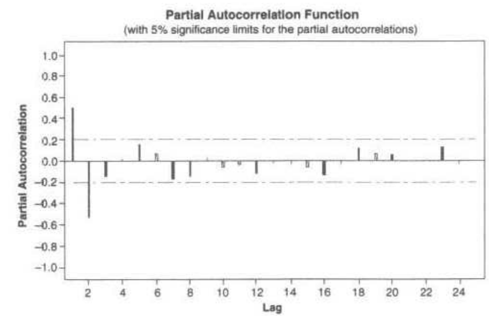
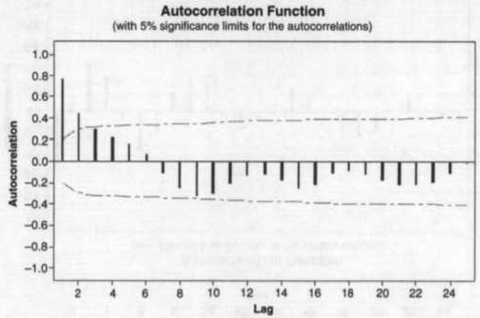
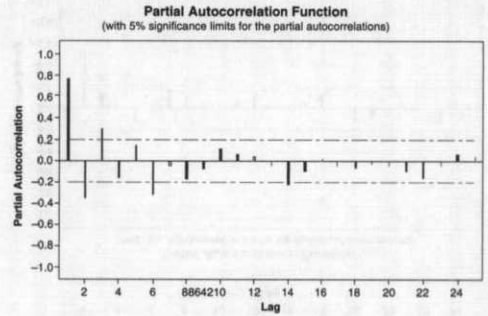
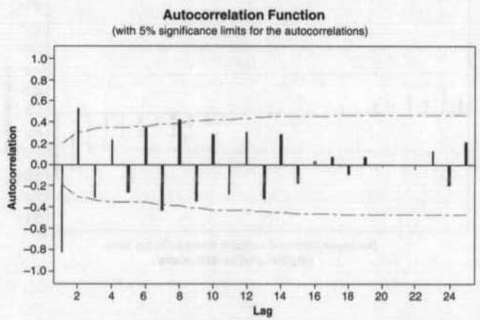
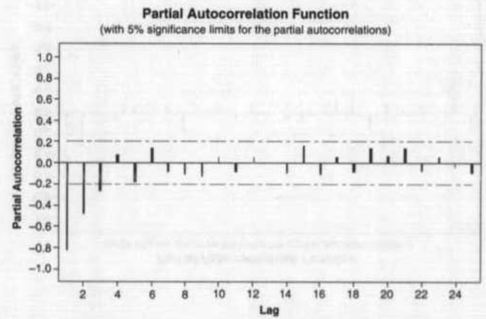


TABLE 5.4 Sample ACFs and PACFs for Some Realizations of ARMA(1,1) Models

Model	Sample ACF	Sample PACF
ARMA(1.1) $y_t = 16 + 0.6y_{t-1} + \varepsilon_t + 0.8\varepsilon_{t-1}$		
$y_t = 16 - 0.7y_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1}$		

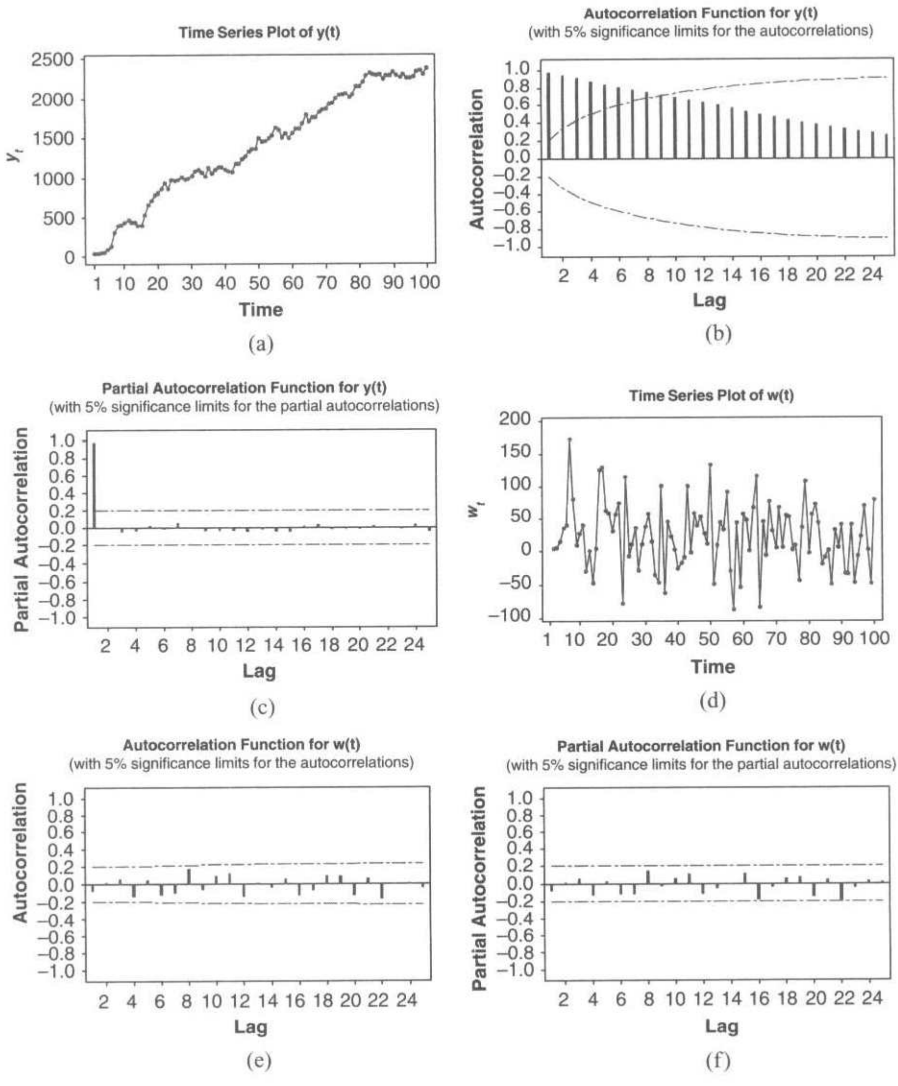


FIGURE 5.11 A realization of the ARIMA(0, 1, 0) model, y_t, its first difference, w_t, and their sample ACFs and PACFs.

Consider the time series data in Figure 5.12a. It looks like the mean of the process is changing (moving upwards) in time. Yet the change in the mean (i.e., nonstationarity) is not as obvious as in the previous example. The sample ACF plot of the data in Figure 5.12b dies relatively slowly and the sample PACF of the data in Figure 5.12c shows two significant values at lags 1 and 2. Hence we might be tempted to model this data using an AR(2) model because of the exponentially decaying ACF and significant PACF at the first two lags. Indeed, we might even have a good fit using an AR(2)

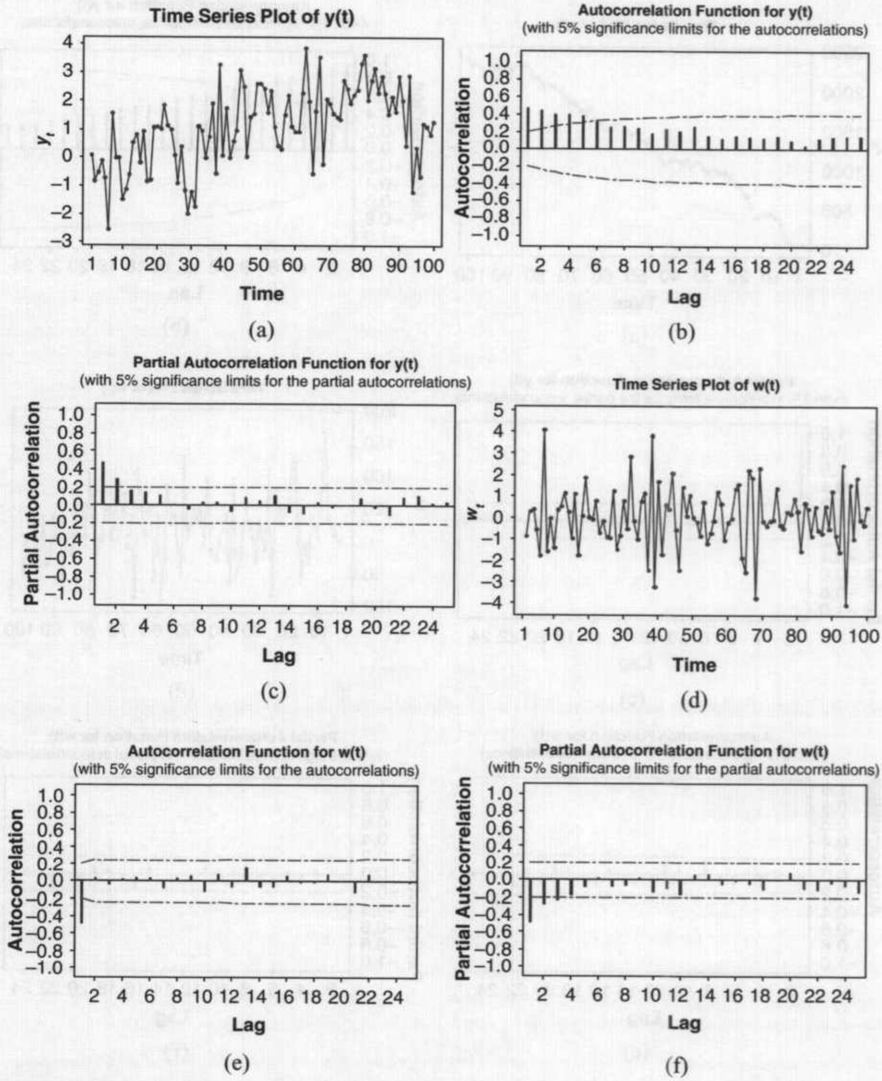


FIGURE 5.12 A realization of the ARIMA(0, 1, 1) model, y_t, its first difference, w_t, and their sample ACFs and PACFs.

model. We should nevertheless check the roots of the associated polynomial given in Eq. (5.34) to make sure that its roots are less than 1 in absolute value. Also note that a technically stationary process will behave more and more nonstationary as the roots of the associated polynomial approach unity. For that, observe the realization of the near nonstationary process, $y_t = 2 + 0.95y_{t-1} + \varepsilon_t$, given in Figure 5.1b. Based on the visual inspection, however, we may deem the process nonstationary and proceed with taking the first difference of the data. This is because the ϕ value of the AR(1) model is close to 1. Under these circumstances, where the nonstationarity of the process is dubious, we strongly recommend that the analyst refer back to basic underlying process knowledge. If, for example, the process mean is expected to wander off as in some financial data, assuming that the process is nonstationary and proceeding with differencing the data would be more appropriate. For the data given in Figure 5.12a, its first difference given in Figure 5.12d looks stationary. Furthermore, its sample ACF and PACF given in Figures 5.12e and 5.12f, respectively, suggest that an MA(1) model would be appropriate for the first difference since its ACF cuts off after the first lag and the PACF exhibits an exponential decay pattern. Hence the ARIMA (0, 1, 1) model given in Eq. (5.73) can be used for this data.

5.7 TIME SERIES MODEL BUILDING

A three-step iterative procedure is used to build an ARIMA model. First, a tentative model of the ARIMA class is identified through analysis of historical data. Second, the unknown parameters of the model are estimated. Third, through residual analysis, diagnostic checks are performed to determine the adequacy of the model, or to indicate potential improvements. We shall now discuss each of these steps in more detail.

5.7.1 Model Identification

Model identification efforts should start with preliminary efforts in understanding the type of process from which the data is coming and how it is collected. The process's perceived characteristics and sampling frequency often provide valuable information in this preliminary stage of model identification. In today's data rich environments, it is often expected that the practitioners would be presented with "enough" data to be able to generate reliable models. It would nevertheless be recommended that 50 or preferably more observations should be initially considered. Before engaging in rigorous statistical model-building efforts, we also strongly recommend the use of "creative" plotting of the data, such as the simple time series plot and scatter plots of the time series data y_t versus y_{t-1}, y_{t-2}, and so on. For the y_t versus y_{t-1} scatter plot, for example, this can be achieved in a data set of N observations by plotting the first $N - 1$ observations versus the last $N - 1$. Simple time series plots should be used as the preliminary assessment tool for stationarity. The visual inspection of these plots should later be confirmed as described earlier in this chapter. If nonstationarity is suspected, the time series plot of the first (or dth) difference should also be considered. The unit root test by Dickey and Fuller [1979] can also be performed to make sure that the differencing is indeed needed. Once the stationarity can be presumed, the sample ACF and PACF of the time series of the original time series (or its dth difference if necessary) should be obtained. Depending on the nature of the autocorrelation, the first 20–25 sample autocorrelations and partial autocorrelations should be sufficient. More care should be taken of course if the process exhibits strong autocorrelation and/or seasonality, as we will discuss in the following sections. Table 5.1 together with the $\pm 2/\sqrt{N}$ limits can be used as a guide for identifying AR or MA models.

As discussed earlier, the identification of ARMA models would require more care, as both the ACF and PACF will exhibit exponential decay and/or damped sinusoid behavior.

We have already discussed that the differenced series $\{w_t\}$ may have a nonzero mean, say, μ_w. At the identification stage we may obtain an indication of whether or not a nonzero value of μ_w is needed by comparing the sample mean of the differenced series, say, $\bar{w} = \sum_{t=1}^{n-d} [w/(n-d)]$, with its approximate standard error. Box, Jenkins, and Reinsel [1994] give the approximate standard error of $\bar{w}$ for several useful ARIMA(p, d, q) models.

Identification of the appropriate ARIMA model requires skills obtained by experience. Several excellent examples of the identification process are given in Box et al. [1994, Chap. 6] and Montgomery et al. [1990].

5.7.2 Parameter Estimation

There are several methods such as methods of moments, maximum likelihood, and least squares that can be employed to estimate the parameters in the tentatively identified model. However, unlike the regression models of Chapter 2, most ARIMA models are **nonlinear** models and require the use of a nonlinear model fitting procedure. However, this is usually automatically performed by sophisticated software packages such as Minitab, JMP, and SAS. In some software packages, the user may have the choice of estimation method and can accordingly choose the most appropriate method based on the problem specifications.

5.7.3 Diagnostic Checking

After a tentative model has been fit to the data, we must examine its adequacy and, if necessary, suggest potential improvements. This is done through residual analysis. The residuals for an ARMA(p, q) process can be obtained from

$$\hat{\varepsilon}_t = y_t - \left(\hat{\delta} + \sum_{i=1}^{p} \hat{\phi}_i y_{t-i} - \sum_{i=1}^{q} \hat{\theta}_i \hat{\varepsilon}_{t-i}\right) \tag{5.76}$$

If the specified model is adequate and hence the appropriate orders p and q are identified, it should transform the observations to a white noise process. Thus the residuals in Eq. (5.76) should behave like white noise.

Let the sample autocorrelation function of the residuals be denoted by $\{r_e(k)\}$. If the model is appropriate, then the residual sample autocorrelation function should have no structure to identify. That is, the autocorrelation should not differ significantly from zero for all lags greater than one. If the form of the model were correct and if we knew the true parameter values, then the standard error of the residual autocorrelations would be $N^{-1/2}$.

Rather than considering the $r_e(k)$ terms individually, we may obtain an indication of whether the first K residual autocorrelations considered together indicate adequacy of the model. This indication may be obtained through an approximate chi-square test of model adequacy. The test statistic is

$$Q = (N - d) \sum_{k=1}^{K} r_e^2(k) \tag{5.77}$$

which is approximately distributed as chi-square with $K - p - q$ degrees of freedom if the model is appropriate. If the model is inadequate, the calculated value of Q will be too large. Thus we should reject the hypothesis of model adequacy if Q exceeds an approximate small upper tail point of the chi-square distribution with $K - p - q$ degrees of freedom. Further details of this test are in Chapter 2 and in the original reference by Box and Pierce [1970]. The modification of this test by Ljung and Box [1978] presented in Chapter 2 is also useful in assessing model adequacy.

5.7.4 Examples of Building ARIMA Models

In this section we shall present two examples of the identification, estimation, and diagnostic checking process. One example presents the analysis for a stationary time series, while the other is an example of modeling a nonstationary series.

Example 5.1

Table 5.5 shows the weekly total number of loan applications in a local branch of a national bank for the last two years. It is suspected that there should be some relationship (i.e., autocorrelation) between the number of applications in the current week and the number of loan applications in the previous weeks. Modeling that relationship will help the management to proactively plan for the coming weeks through reliable forecasts. As always, we start our analysis with the time series plot of the data, shown in Figure 5.13.

Figure 5.13 shows that the weekly data tend to have short runs and that the data seem to be indeed autocorrelated. Next, we visually inspect the stationarity. Although there might be a slight drop in the mean for the second year (weeks 53–104), in general it seems to be safe to assume stationarity.

We now look at the sample ACF and PACF plots in Figure 5.14. Here are possible interpretations of the ACF plot:

1. It cuts off after lag 2 (or maybe even 3), suggesting a MA(2) (or MA(3)) model.
2. It has an (or a mixture of) exponential decay(s) pattern suggesting an AR(p) model.

To resolve the conflict, consider the sample PACF plot. For that, we have only one interpretation; it cuts off after lag 2. Hence we use the second interpretation of the sample ACF plot and assume that the appropriate model to fit is the AR(2) model.

Table 5.6 shows the Minitab output for the AR(2) model. The parameter estimates are $\hat{\phi}_1 = 0.27$ and $\hat{\phi}_2 = 0.42$, and they turn out to be significant (see the P-values).

TABLE 5.5 Weekly Total Number of Loan Applications for the Last Two Years

Week	Applications	Week	Applications	Week	Applications	Week	Applications
1	71	27	62	53	66	79	63
2	57	28	77	54	71	80	61
3	62	29	76	55	59	81	73
4	64	30	88	56	57	82	72
5	65	31	71	57	66	83	65
6	67	32	72	58	51	84	70
7	65	33	66	59	59	85	54
8	82	34	65	60	56	86	63
9	70	35	73	61	57	87	62
10	74	36	76	62	55	88	60
11	75	37	81	63	53	89	67
12	81	38	84	64	74	90	59
13	71	39	68	65	64	91	74
14	75	40	63	66	70	92	61
15	82	41	66	67	74	93	61
16	74	42	71	68	69	94	52
17	78	43	67	69	64	95	55
18	75	44	69	70	68	96	61
19	73	45	63	71	64	97	56
20	76	46	61	72	70	98	61
21	66	47	68	73	73	99	60
22	69	48	75	74	59	100	65
23	63	49	66	75	68	101	55
24	76	50	81	76	59	102	61
25	65	51	72	77	66	103	59
26	73	52	77	78	63	104	63

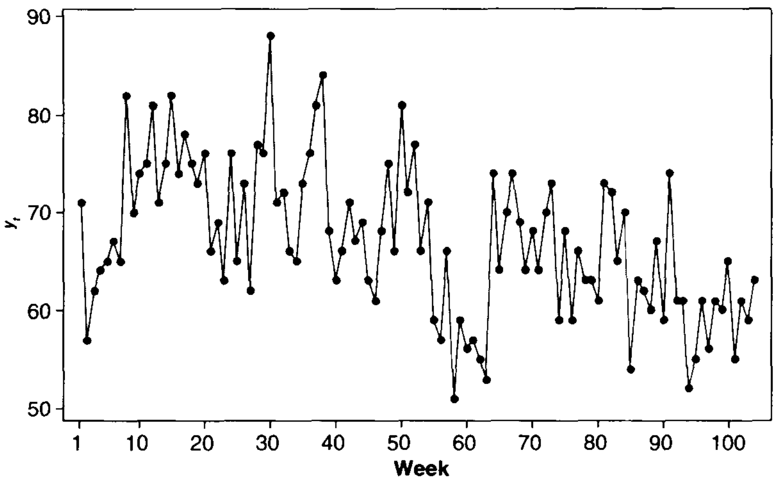
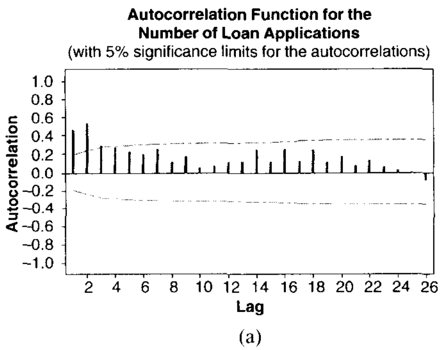


FIGURE 5.13 Time series plot of the weekly total number of loan applications.



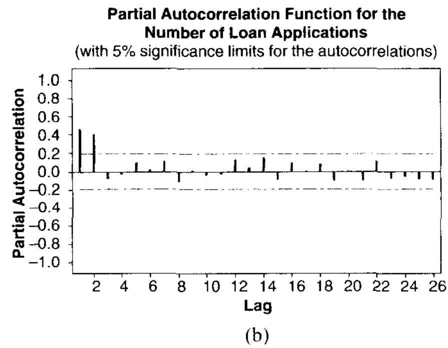


FIGURE 5.14 ACF and PACF for the weekly total number of loan applications.

MSE is calculated to be 39.35. The modified Box–Pierce test suggests that there is no autocorrelation left in the residuals. We can also see this in the ACF and PACF plots of the residuals in Figure 5.15.

As the last diagnostic check, we have the 4-in-1 residual plots in Figure 5.16 provided by Minitab: Normal Probability Plot, Residuals versus Fitted Value, Histogram of the Residuals, and Time Series Plot of the Residuals. They indicate that the fit is indeed acceptable.

TABLE 5.6 Minitab Output for the AR(2) Model for the Loan Application Data

```
Final Estimates of Parameters

Type          Coef   SE Coef      T      P
AR    1     0.2682    0.0903   2.97  0.004
AR    2     0.4212    0.0908   4.64  0.000
Constant   20.7642    0.6157  33.73  0.000
Mean        66.844     1.982


Number of observations:  104
Residuals:     SS =  3974.30 (backforecasts excluded)
               MS =  39.35   DF = 101


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag             12      24      36      48
Chi-Square     6.2    16.0    24.9    32.0
DF               9      21      33      45
P-Value      0.718   0.772   0.843   0.927
```

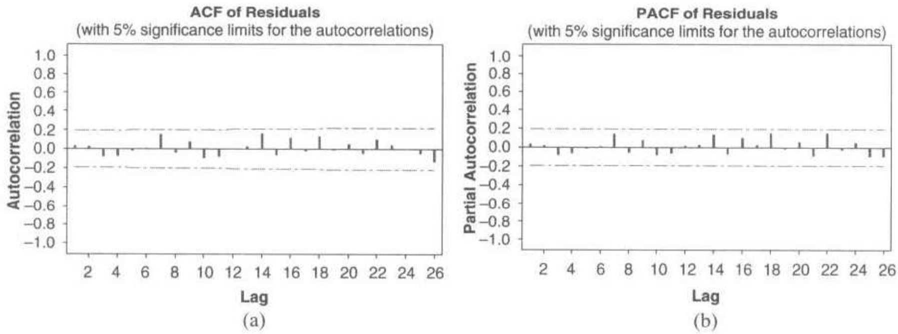


FIGURE 5.15 The sample ACF and PACF of the residuals for the AR(2) model in Table 5.6.

Figure 5.17 shows the actual data and the fitted values. It looks like the fitted values smooth out the highs and lows in the data.

Note that, in this example, we often and deliberately used "vague" words such as "seems" or "looks like." It should be clear by now that the methodology presented in this chapter has a very sound theoretical foundation. However, as in any modeling effort, we should also keep in mind the subjective component of model identification. In fact, as we mentioned earlier, time series model fitting can be seen as a mixture of science and art and can best be learned by practice and experience. The next example will illustrate this point further. ■

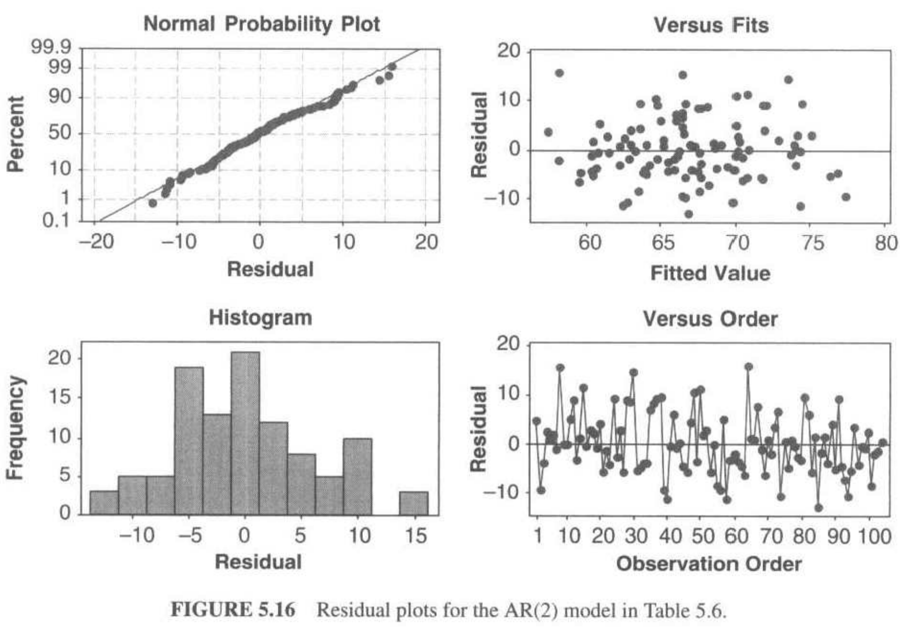


FIGURE 5.16 Residual plots for the AR(2) model in Table 5.6.

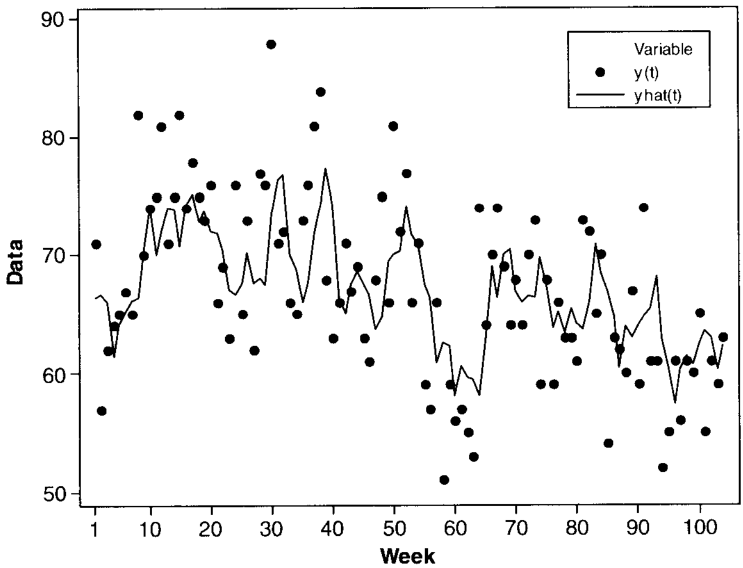


FIGURE 5.17 Time series plot of the actual data and fitted values for the AR(2) model in Table 5.6.

Example 5.2

Consider the Dow Jones Index data from Chapter 4. A time series plot of the data is given in Figure 5.18. The process shows signs of nonstationarity with changing mean and possibly variance.

Similarly, the slowly decreasing sample ACF and sample PACF with significant value at lag 1, which is close to 1 in Figure 5.19, confirm that indeed the process

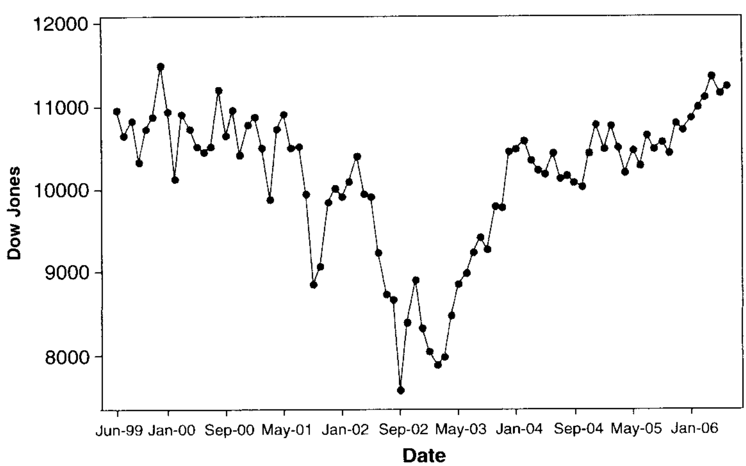


FIGURE 5.18 Time series plot of the Dow Jones Index from June 1999 to June 2006.

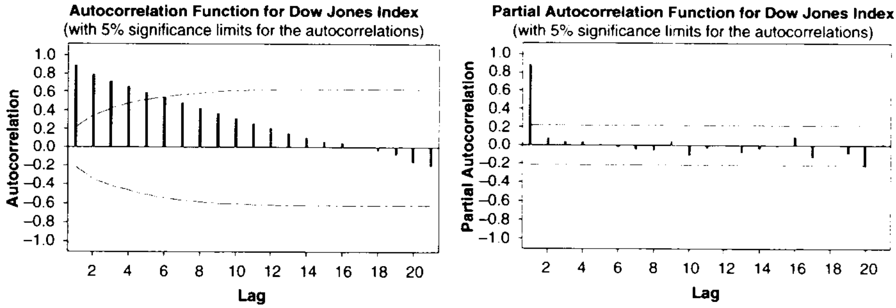


FIGURE 5.19 Sample ACF and PACF of the Dow Jones Index.

can be deemed nonstationary. On the other hand, one might argue that the significant sample PACF value at lag 1 suggests that the AR(1) model might also fit the data well. We will consider this interpretation first and fit an AR(1) model to the Dow Jones Index data.

Table 5.7 shows the Minitab output for the AR(1) model. Although it is close to 1, the AR(1) model coefficient estimate $\hat{\phi} = 0.9045$ turns out to be quite significant and the modified Box–Pierce test suggests that there is no autocorrelation left in the residuals. This is also confirmed by the sample ACF and PACF plots of the residuals given in Figure 5.20.

The only concern in the residual plots in Figure 5.21 is in the changing variance observed in the time series plot of the residuals. This is indeed a very important issue

TABLE 5.7 Minitab Output for the AR(1) Model for the Dow Jones Index

```
Final Estimates of Parameters

Type           Coef   SE Coef        T        P
AR   1       0.9045    0.0500    18.10    0.000
Constant     984.94     44.27    22.25    0.000
Mean        10309.9     463.4


Number of observations:  85
Residuals:      SS =  13246015 (backforecasts excluded)
                MS =  159591   DF = 83


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag              12       24       36       48
Chi-Square      2.5     14.8     21.4     29.0
DF               10       22       34       46
P-Value       0.991    0.872    0.954    0.977
```

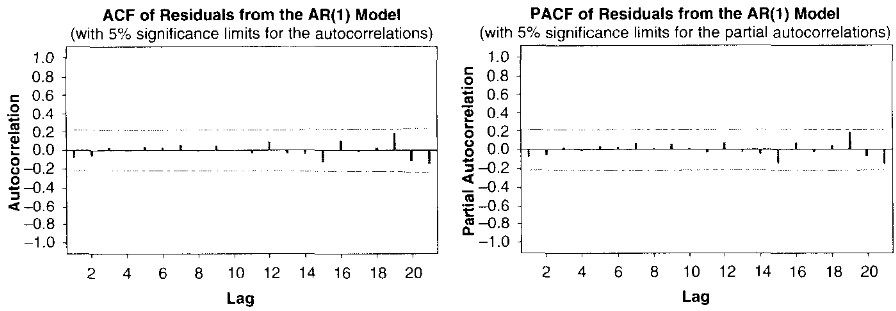


FIGURE 5.20 Sample ACF and PACF of the residuals from the AR(1) model for the Dow Jones Index data.

since it violates the constant variance assumption. We will discuss this issue further in Section 7.3 but for illustration purposes we will ignore it in this example.

Overall it can be argued that an AR(1) model provides a decent fit to the data. However, we will now consider the earlier interpretation and assume that the Dow Jones Index data comes from a nonstationary process. We then take the first difference of the data as shown in Figure 5.22. While there are once again some serious concerns about changing variance, the level of the first difference remains the same. If we ignore the changing variance and look at the sample ACF and PACF plots given in Figure 5.23, we may conclude that the first difference is in fact white noise. That is, since these plots do not show any sign of significant autocorrelation, a model we may consider for the Dow Jones Index data would be the random walk model, ARIMA (0, 1, 0).

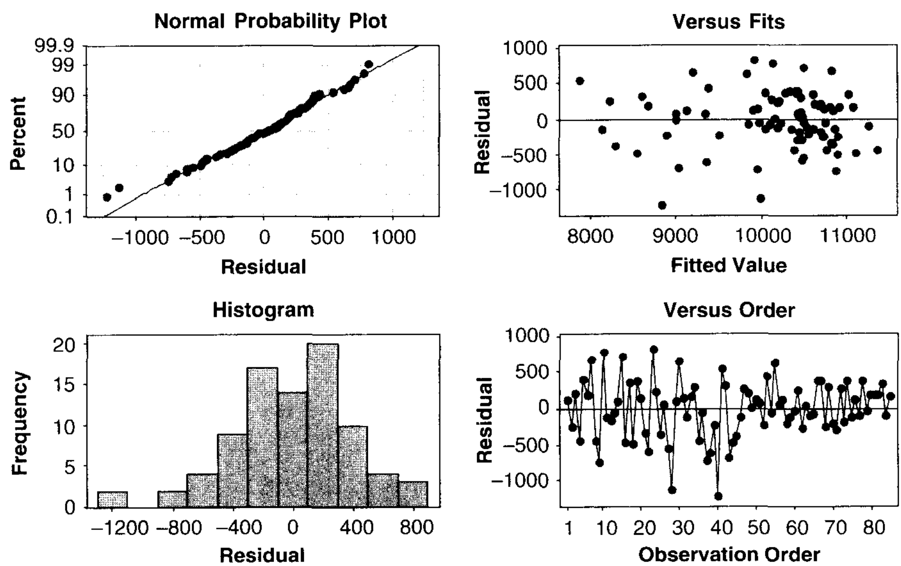


FIGURE 5.21 Residual plots from the AR(1) model for the Dow Jones Index data.

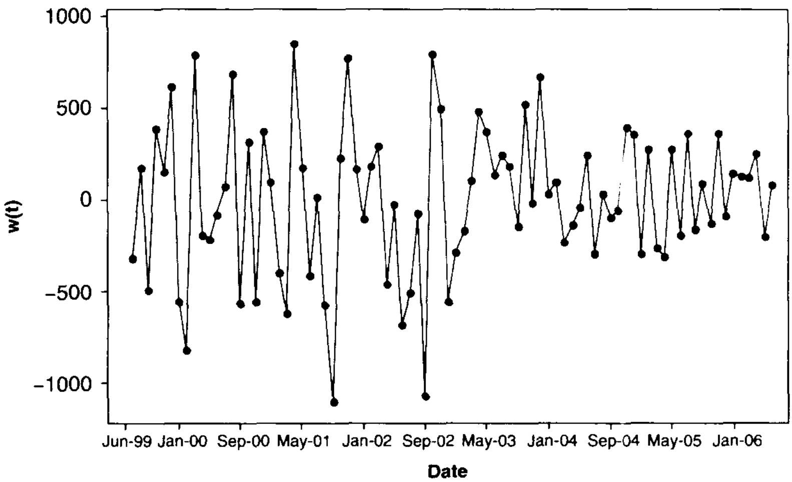


FIGURE 5.22 Time series plot of the first difference $w(t)$ of the Dow Jones Index data.

Now the analyst has to decide between the two models: AR(1) and ARIMA (0, 1, 0). One can certainly use some of the criteria we discussed in Section 2.6.2 to choose one of these models. Since these two models are fundamentally quite different, we strongly recommend that the analyst use the subject matter/process knowledge as much as possible. Do we expect a financial index such as the Dow Jones Index to wander about a fixed mean as implied by the AR(1)? In most cases involving financial data, the answer would be no. Hence a model such as ARIMA(0, 1, 0) that takes into account the inherent nonstationarity of the process should be preferred. However, we do have a problem with the proposed model. A random walk model means that the price changes are random and cannot be predicted. If we have a higher price today compared to yesterday, that would have no bearing on the forecasts tomorrow. That is, tomorrow's price can be higher or lower than today's and we would have no way to forecast it effectively. This further suggests that the best forecast for tomorrow's price is in fact the price we have today. This is obviously not a reliable and effective forecasting model. This very same issue of the random walk models for financial

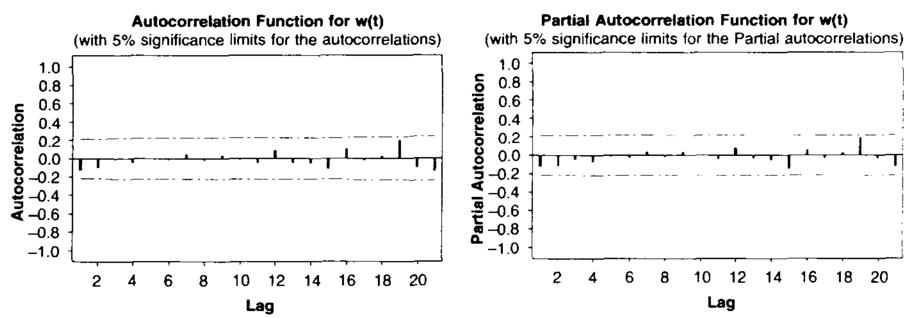


FIGURE 5.23 Sample ACF and PACF plots of the first difference of the Dow Jones Index data.

data has been discussed in great detail in the literature. We simply used this data to illustrate that in time series model fitting we can end up with fundamentally different models that will fit the data equally well. At this point, process knowledge can provide the needed guidance in picking the "right" model.

It should be noted that, in this example, we tried to keep the models simple for illustration purposes. Indeed, a more thorough analysis would (and should) pay close attention to the changing variance issue. In fact, this is a very common concern particularly when dealing with financial data. For that, we once again refer the reader to Section 7.3. ■

5.8 FORECASTING ARIMA PROCESSES

Once an appropriate time series model has been fit, it may be used to generate forecasts of future observations. If we denote the current time by T, the forecast for $y_{T+\tau}$ is called the τ-period-ahead forecast and denoted by $\hat{y}_{T+\tau}(T)$. The standard criterion to use in obtaining the best forecast is the mean squared error for which the expected value of the squared forecast errors, $E[(y_{T+\tau} - \hat{y}_{T+\tau}(T))^2] = E[e_T(\tau)^2]$, is minimized. It can be shown that the best forecast in the mean square sense is the conditional expectation of $y_{T+\tau}$ given current and previous observations, that is, $y_T, y_{T-1}, \ldots$:

$$\hat{y}_{T+\tau}(T) = E[y_{T+\tau} \mid y_T, y_{T-1}, \ldots] \tag{5.78}$$

Consider, for example, an ARIMA (p, d, q) process at time $T + \tau$ (i.e., τ period in the future):

$$y_{T+\tau} = \delta + \sum_{i=1}^{p+d} \phi_i y_{T+\tau-i} + \varepsilon_{T+\tau} - \sum_{i=1}^{q} \theta_i \varepsilon_{T+\tau-i} \tag{5.79}$$

Further consider its infinite MA representation,

$$y_{T+\tau} = \mu + \sum_{i=1}^{\infty} \psi_i \varepsilon_{T+\tau-i} \tag{5.80}$$

We can partition Eq. (5.80) as

$$y_{T+\tau} = \mu + \sum_{i=1}^{\tau-1} \psi_i \varepsilon_{T+\tau-i} + \sum_{i=\tau}^{\infty} \psi_i \varepsilon_{T+\tau-i} \tag{5.81}$$

In this partition we can clearly see that the $\sum_{i=1}^{\tau-1} \psi_i \varepsilon_{T+\tau-i}$ component involves the future errors whereas the $\sum_{i=\tau}^{\infty} \psi_i \varepsilon_{T+\tau-i}$ component involves the present and past errors. From the relationship between the current and past observations and the corresponding random shocks as well as the fact that the random shocks are assumed to have mean zero and to be independent, we can show that the best forecast in the mean square sense is

$$\hat{y}_{T+\tau}(T) = E[y_{T+\tau} | y_T, y_{T-1}, \ldots] = \mu + \sum_{i=\tau}^{\infty} \psi_i \varepsilon_{T+\tau-i} \tag{5.82}$$

since

$$E[\varepsilon_{T+\tau-i} | y_T, y_{T-1}, \ldots] = \begin{cases} 0 & \text{if } i < \tau \\ \varepsilon_{T+\tau-i} & \text{if } i \geq \tau \end{cases}$$

Subsequently, the forecast error is calculated from

$$e_T(\tau) = y_{T+\tau} - \hat{y}_{T+\tau}(T) = \sum_{i=0}^{\tau-1} \psi_i \varepsilon_{T+\tau-i} \tag{5.83}$$

Since the forecast error in Eq. (5.83) is a linear combination of random shocks, we have

$$E[e_T(\tau)] = 0 \tag{5.84}$$

$$\begin{aligned} \text{Var}[e_T(\tau)] &= \text{Var}\left[\sum_{i=0}^{\tau-1} \psi_i \varepsilon_{T+\tau-i}\right] = \sum_{i=0}^{\tau-1} \psi_i^2 \text{Var}(\varepsilon_{T+\tau-i}) \\ &= \sigma^2 \sum_{i=0}^{\tau-1} \psi_i^2 \qquad (5.85) \\ &= \sigma^2(\tau), \quad \tau = 1, 2, \ldots \end{aligned}$$

It should be noted that the variance of the forecast error gets bigger with increasing forecast lead times τ. This intuitively makes sense as we should expect more uncertainty in our forecasts further into the future. Moreover, if the random shocks are assumed to be normally distributed, $N(0, \sigma^2)$, then the forecast errors will also be normally distributed with $N(0, \sigma^2(\tau))$. We can then obtain the $100(1-\alpha)$ percent prediction intervals for the future observations from

$$P\left(\hat{y}_{T+\tau}(T) - z_{\alpha/2}\sigma(\tau) < y_{T+\tau} < \hat{y}_{T+\tau}(T) + z_{\alpha/2}\sigma(\tau)\right) = 1 - \alpha \tag{5.86}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution, $N(0, 1)$. Hence the $100(1-\alpha)$ percent prediction interval for $y_{T+\tau}$ is

$$\hat{y}_{T+\tau}(T) \pm z_{\alpha/2}\sigma(\tau) \tag{5.87}$$

There are two issues with the forecast equation in (5.82). First, it involves infinitely many terms in the past. However, in practice, we will only have a finite amount of data. For a sufficiently large data set, this can be overlooked. Second, Eq. (5.82) requires knowledge of the magnitude of random shocks in the past, which is unrealistic. A solution to this problem is to "estimate" the past random shocks through one-step-ahead forecasts. For the ARIMA model we can calculate

$$\hat{\varepsilon}_t = y_t - \left[\delta + \sum_{i=1}^{p+d} \phi_i y_{t-i} - \sum_{i=1}^{q} \theta_i \hat{\varepsilon}_{t-i}\right] \tag{5.88}$$

recursively by setting the initial values of the random shocks to zero for $t < p + d + 1$. For more accurate results, these initial values together with the y_t for $t \leq 0$ can also be obtained using back-forecasting. For further details, see Box, Jenkins, and Reinsel [1994].

As an illustration consider forecasting the ARIMA(1, 1, 1) process

$$(1 - \phi B)(1 - B)\, y_{T+\tau} = (1 - \theta B)\, \varepsilon_{T+\tau} \tag{5.89}$$

We will consider two of the most commonly used approaches:

1. As discussed earlier, this approach involves the infinite MA representation of the model in Eq. (5.89), also known as the **random shock** form of the model:

$$\begin{aligned} y_{T+\tau} &= \sum_{i=1}^{\infty} \psi_i \varepsilon_{T+\tau-i} \\ &= \psi_1 \varepsilon_{T+\tau-1} + \psi_2 \varepsilon_{T+\tau-2} + \cdots \end{aligned} \tag{5.90}$$

   Hence the τ-step-ahead forecast can be calculated from

$$\hat{y}_{T+\tau}(T) = \psi_\tau \varepsilon_T + \psi_{\tau+1} \varepsilon_{T-1} + \cdots \tag{5.91}$$

   The weights ψ_i can be calculated from

$$(\psi_0 + \psi_1 B + \cdots)\left(1 - \phi_1 B - \cdots - \phi_p B^p\right)(1 - B) = (1 - \theta B) \tag{5.92}$$

   and the random shocks can be estimated using the one-step-ahead forecast error; for example, ε_T can be replaced by $e_T(1) = y_T - \hat{y}_T(T - 1)$.

2. Another approach that is often employed in practice is to use **difference equations** as given by

$$y_{T+\tau} = (1 + \phi)\, y_{T+\tau-1} - \phi y_{T+\tau-2} + \varepsilon_{T+\tau} - \theta \varepsilon_{T+\tau-1} \tag{5.93}$$

   For $\tau = 1$, the best forecast in the mean squared error sense is

$$\hat{y}_{T+1}(T) = E\left[y_{T+1} \mid y_T, y_{T-1}, \ldots\right] = (1 + \phi)\, y_T - \phi y_{T-1} - \theta e_T(1) \tag{5.94}$$

   We can further show that for lead times $\tau > 2$, the forecast is

$$\hat{y}_{T+\tau}(T) = (1 + \phi)\, \hat{y}_T(\tau - 1) - \phi \hat{y}_T(\tau - 2) \tag{5.95}$$

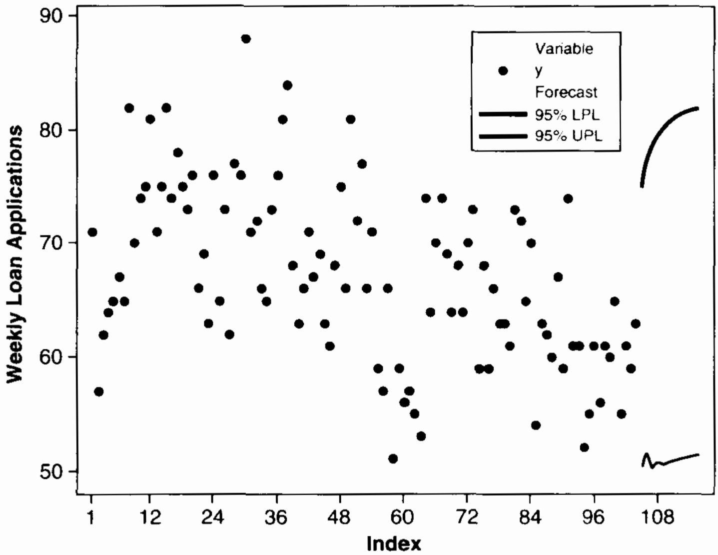


FIGURE 5.24 Forecasts for the weekly loan application data.

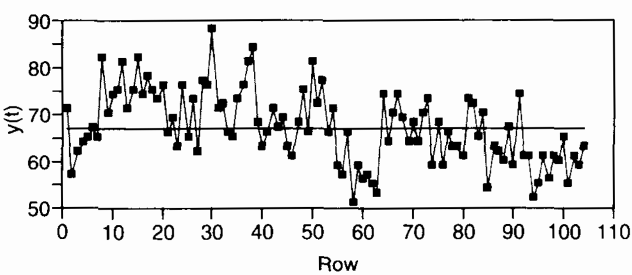
Example 5.3

Consider the loan applications data given in Table 5.5. Now assume that the manager wants to make forecasts for the next 3 months (12 weeks). Hence at the 104th week we need to make 1-step, 2-step, ..., 12-step-ahead predictions, which are obtained and plotted using Minitab in Figure 5.24 together with the 95% prediction interval. ■

Table 5.8 shows the output from JMP for fitting an AR(2) model to the weekly loan application data. In addition to the sample ACF and PACF, JMP provides

TABLE 5.8 JMP AR(2) Output for the Loan Application Data

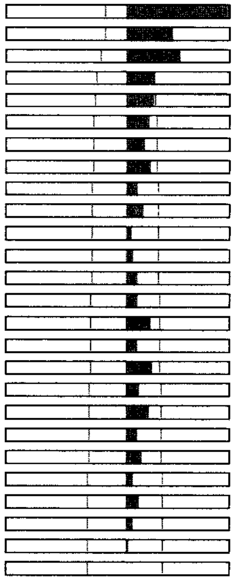
Time Series y(t)



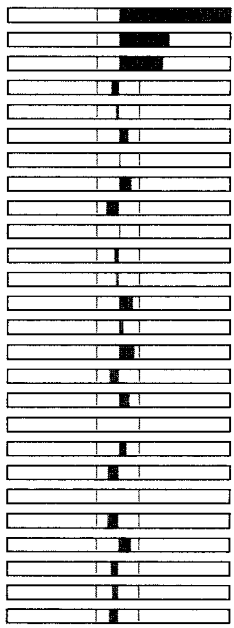
Mean	67.067308
Std	7.663932
N	104
Zero Mean ADF	−0.695158
Single Mean ADF	−6.087814
Trend ADF	−7.396174

TABLE 5.8 (*Continued*)

Time Series Basic Diagnostics



Lag	AutoCorr	Plot Autocorr	Ljung-Box Q	p-Value
0	1.0000		.	.
1	0.4617		22.8186	<.0001
2	0.5314		53.3428	<.0001
3	0.2915		62.6167	<.0001
4	0.2682		70.5487	<.0001
5	0.2297		76.4252	<.0001
6	0.1918		80.5647	<.0001
7	0.2484		87.5762	<.0001
8	0.1162		89.1255	<.0001
9	0.1701		92.4847	<.0001
10	0.0565		92.8587	<.0001
11	0.0716		93.4667	<.0001
12	0.1169		95.1040	<.0001
13	0.1151		96.7080	<.0001
14	0.2411		103.829	<.0001
15	0.1137		105.430	<.0001
16	0.2540		113.515	<.0001
17	0.1279		115.587	<.0001
18	0.2392		122.922	<.0001
19	0.1138		124.603	<.0001
20	0.1657		128.206	<.0001
21	0.0745		128.944	<.0001
22	0.1320		131.286	<.0001
23	0.0708		131.968	<.0001
24	0.0338		132.125	<.0001
25	0.0057		132.130	<.0001



Lag Lag	Partial AutoCorr	Plot Partial Plot Autocorr	Ljung-Box Q	p-Value
0	1.0000			
1	0.4617			
2	0.4045			
3	-0.0629			
4	-0.0220			
5	0.0976			
6	0.0252			
7	0.1155			
8	-0.1017			
9	0.0145			
10	-0.0330			
11	-0.0250			
12	0.1349			
13	0.0488			
14	0.1489			
15	-0.0842			
16	0.1036			
17	0.0105			
18	0.0830			
19	-0.0938			
20	0.0052			
21	-0.0927			
22	0.1149			
23	-0.0645			
24	-0.0473			
25	-0.0742			

TABLE 5.8 JMP AR(2) Output for the Loan Application Data (*Continued*)

Model Comparison

Model	DF	Variance	AIC	SBC	RSquare	-2LogLH
AR(2)	101	39.458251	680.92398	688.85715	0.343	674.92398

Model: AR(2)
Model Summary

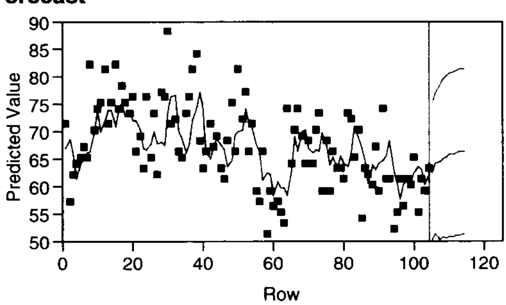
DF	101
Sum of Squared Errors	3985.28336
Variance Estimate	39.4582511
Standard Deviation	6.2815803
Akaike's 'A' Information Criterion	680.923978
Schwarz's Bayesian Criterion	688.857151
RSquare	0.34278547
RSquare Adj	0.32977132
MAPE	7.37857799
MAE	4.91939717
-2LogLikelihood	674.923978

Stable	Yes
Invertible	Yes

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
AR1	1	0.265885	0.089022	2.99	0.0035	21.469383
AR2	2	0.412978	0.090108	4.58	<.0001	
Intercept	0	66.854262	1.833390	36.46	<.0001	

Forecast



Residuals

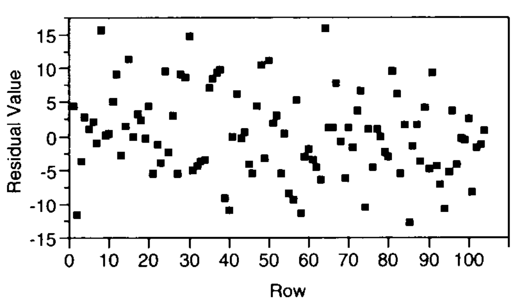


TABLE 5.8 (*Continued*)

Lag	AutoCorr Plot Autocorr	Ljung-Box Q	p-Value
0	1.0000	.	.
1	0.0320	0.1094	0.7408
2	0.0287	0.1986	0.9055
3	-0.0710	0.7489	0.8617
4	-0.0614	1.1647	0.8839
5	-0.0131	1.1839	0.9464
6	0.0047	1.1864	0.9776
7	0.1465	3.6263	0.8217
8	-0.0309	3.7358	0.8801
9	0.0765	4.4158	0.8820
10	-0.0938	5.4479	0.8593
11	-0.0698	6.0251	0.8717
12	0.0019	6.0255	0.9148
13	0.0223	6.0859	0.9430
14	0.1604	9.2379	0.8155
15	-0.0543	9.6028	0.8440
16	0.1181	11.3501	0.7874
17	-0.0157	11.3812	0.8361
18	0.1299	13.5454	0.7582
19	-0.0059	13.5499	0.8093
20	0.0501	13.8788	0.8366
21	-0.0413	14.1056	0.8650
22	0.0937	15.2870	0.8496
23	0.0409	15.5146	0.8752
24	-0.0035	15.5163	0.9047
25	-0.0335	15.6731	0.9242

Lag	Partial Plot Partial
0	1.0000
1	0.0320
2	0.0277
3	-0.0729
4	-0.0580
5	-0.0053
6	0.0038
7	0.1399
8	-0.0454
9	0.0715
10	-0.0803

Lag	AutoCorr Plot Autocorr	Ljung-Box Q	p-Value
11	-0.0586		
12	0.0201		
13	0.0211		
14	0.1306		
15	-0.0669		
16	0.1024		
17	0.0256		
18	0.1477		
19	-0.0027		
20	0.0569		
21	-0.0823		
22	0.1467		
23	-0.0124		
24	0.0448		
25	-0.0869		

the model fitting information including the estimates of the model parameters, the forecasts for 10 periods into the future and the associated prediction intervals, and the residual autocorrelation and partial autocorrelation functions. The AR(2) model is an excellent fit to the data.

5.9 SEASONAL PROCESSES

Time series data may sometimes exhibit strong periodic patterns. This is often referred to as the time series having a seasonal behavior. This mostly occurs when data is taken in specific intervals—monthly, weekly, and so on. One way to represent such data is through an additive model where the process is assumed to be composed of two parts,

$$y_t = S_t + N_t \tag{5.96}$$

where S_t is the deterministic component with periodicity s and N_t is the stochastic component that may be modeled as an ARMA process. In that, y_t can be seen as a process with predictable periodic behavior with some noise sprinkled on top of it. Since the S_t is deterministic and has periodicity s, we have $S_t = S_{t+s}$ or

$$S_t - S_{t-s} = (1 - B^s)S_t = 0 \tag{5.97}$$

Applying the $(1 - B^s)$ operator to Eq. (5.96), we have

$$\begin{aligned} \underbrace{(1 - B^s)y_t}_{\equiv w_t} &= \underbrace{(1 - B^s)S_t}_{=0} + (1 - B^s)N_t \\ w_t &= (1 - B^s)N_t \end{aligned} \tag{5.98}$$

The process w_t can be seen as **seasonally stationary**. Since an ARMA process can be used to model N_t, in general we have

$$\Phi(B)w_t = (1 - B^s)\Theta(B)\varepsilon_t \tag{5.99}$$

where ε_t is white noise.

We can also consider S_t as a stochastic process. We will further assume that after seasonal differencing, $(1 - B^s)$, $(1 - B^s)\,y_t = w_t$ becomes stationary. This, however, may not eliminate all seasonal features in the process. That is, the seasonally differenced data may still show strong autocorrelation at lags s, $2s$, ... So the seasonal ARMA model is

$$\left(1 - \phi_1^* B^s - \phi_2^* B^{2s} - \cdots - \phi_P^* B^{Ps}\right)w_t = \left(1 - \theta_1^* B^s - \theta_2^* B^{2s} - \cdots - \theta_Q^* B^{Qs}\right)\varepsilon_t \tag{5.100}$$

This representation, however, only takes into account the autocorrelation at seasonal lags s, $2s$, ... Hence a more general seasonal ARIMA model of orders $(p, d, q) \times$

(P, D, Q) with period s is

$$\Phi^*(B^s)\Phi(B)(1-B)^d(1-B^s)^D y_t = \delta + \Theta^*(B^s)\Theta(B)\varepsilon_t \tag{5.101}$$

In practice, although it is case specific, it is not expected to have P, D, and Q greater than 1. The results for regular ARIMA processes that we discussed in previous sections apply to the seasonal models given in Eq. (5.101).

As in the nonseasonal ARIMA models, the forecasts for the seasonal ARIMA models can be obtained from the difference equations as illustrated for example in Eq. (5.95) for a nonseasonal ARIMA (1,1,1) process. Similarly the weights in the random shock form given in Eq. (5.90) can be estimated as in Eq. (5.92) to obtain the estimate for the variance of the forecast errors as well as the prediction intervals given in Eqs. (5.85) and (5.86) respectively.

Example 5.4

The ARIMA (0, 1, 1) $\times$ (0, 1, 1) model with $s = 12$ is

$$\underbrace{(1-B)(1-B^{12})y_t}_{w_t} = \left(1 - \theta_1 B - \theta_1^* B^{12} + \theta_1\theta_1^* B^{13}\right)\varepsilon_t$$

For this process, the autocovariances are calculated as

$$\begin{aligned}
\gamma(0) &= \text{Var}(w_t) = \sigma^2\left(1 + \theta_1^2 + \theta_1^{*2} + \left(-\theta_1\theta_1^*\right)^2\right) \\
&= \sigma^2\left(1+\theta_1^2\right)\left(1+\theta_1^{*2}\right) \\
\gamma(1) &= \text{Cov}(w_t, w_{t-1}) = \sigma^2\left(-\theta_1 + \theta_1^*\left(-\theta_1\theta_1^*\right)\right) \\
&= -\theta_1\sigma^2\left(1+\theta_1^*\right) \\
\gamma(2) &= \gamma(3) = \cdots = \gamma(10) = 0 \\
\gamma(11) &= \sigma^2\theta_1\theta_1^* \\
\gamma(12) &= -\sigma^2\theta_1^*\left(1+\theta_1^2\right) \\
\gamma(13) &= \sigma^2\theta_1\theta_1^* \\
\gamma(j) &= 0, \quad j > 13
\end{aligned}$$

■

Example 5.5

Consider the U.S. clothing sales data in Table 4.9. The data obviously exhibit some seasonality and upward linear trend. The sample ACF and PACF plots given in Figure 5.25 indicate a monthly seasonality, $s = 12$, as ACF values at lags 12, 24, 36 are significant and slowly decreasing, and there is a significant PACF value at lag 12

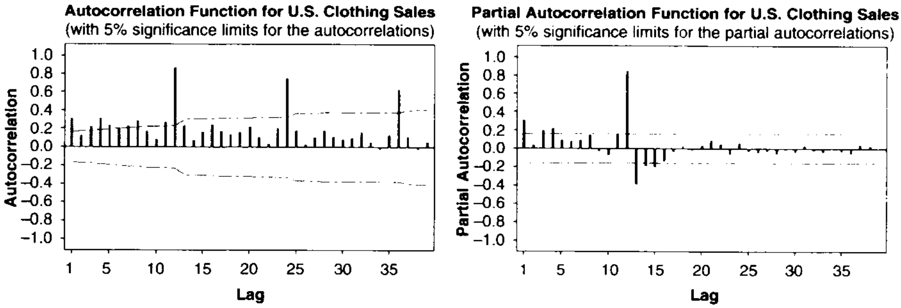


FIGURE 5.25 Sample ACF and PACF plots of the U.S. clothing sales data.

that is close to 1. Moreover, the slowly decreasing ACF in general also indicates a nonstationarity that can be remedied by taking the first difference. Hence we would now consider $w_t = (1 - B)(1 - B^{12})y_t$.

Figure 5.26 shows that first difference together with seasonal differencing—that is, $w_t = (1 - B)(1 - B^{12})y_t$—helps in terms of stationarity and eliminating the seasonality, which is also confirmed by sample ACF and PACF plots given in Figure 5.27. Moreover, the sample ACF with a significant value at lag 1 and the sample PACF with exponentially decaying values at the first 8 lags suggest that a nonseasonal MA(1) model should be used.

The interpretation of the remaining seasonality is a bit more difficult. For that we should focus on the sample ACF and PACF values at lags 12, 24, 36, and so on. The sample ACF at lag 12 seems to be significant and the sample PACF at lags 12,

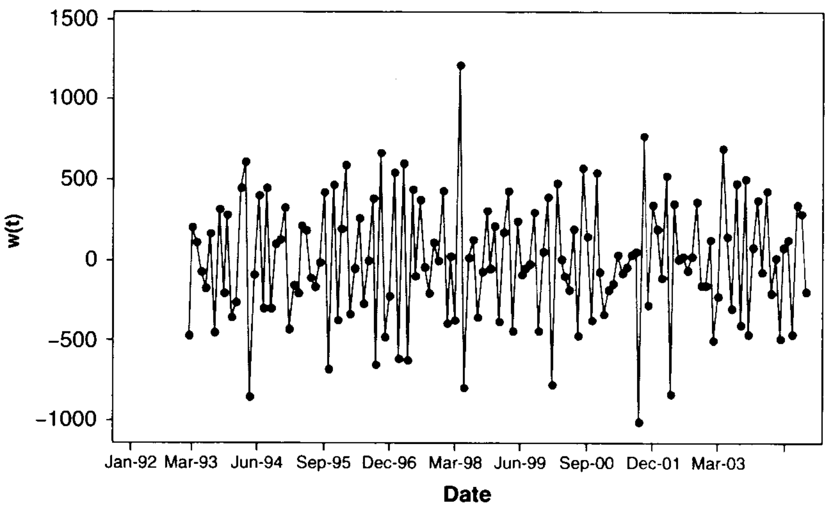


FIGURE 5.26 Time series plot of $w_t = (1 - B)(1 - B^{12})y_t$ for the U.S. clothing sales data.

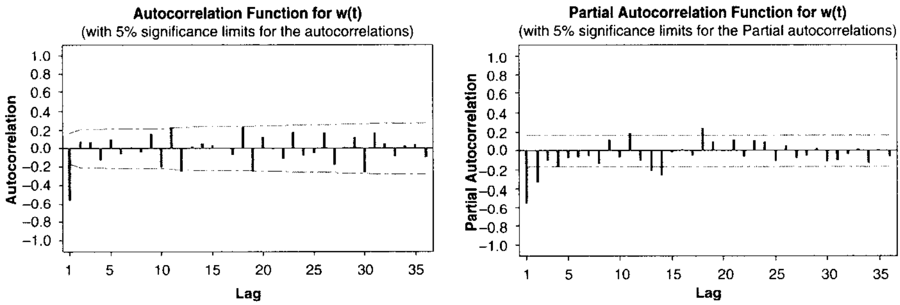


FIGURE 5.27 Sample ACF and PACF plots of $w_t = (1 - B)(1 - B^{12})y_t$.

24, 36 (albeit not significant) seems to be alternating in sign. That suggests that a seasonal MA(1) model can be used as well. Hence an ARIMA $(0, 1, 1) \times (0, 1, 1)_{12}$ model is used to model the data, y_t. The output from Minitab is given in Table 5.9. Both MA(1) and seasonal MA(1) coefficient estimates are significant. As we can see from the sample ACF and PACF plots in Figure 5.28, while there are still some small significant values, as indicated by the modified Box-Pierce statistic, most of the autocorrelation is now modeled out.

The residual plots in Figure 5.29 provided by Minitab seem to be acceptable as well.

Finally, the time series plot of the actual and fitted values in Figure 5.30 suggests that the ARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ model provides a reasonable fit to this highly seasonal and nonstationary time series data. ■

TABLE 5.9 Minitab Output for the ARIMA $(0, 1, 1) \times (0, 1, 1)_{12}$ Model for the U.S. Clothing Sales Data

```
Final Estimates of Parameters

Type        Coef  SE Coef      T      P
MA   1    0.7626   0.0542  14.06  0.000
SMA  12   0.5080   0.0771   6.59  0.000


Differencing: 1 regular, 1 seasonal of order 12
Number of observations:  Original series 155, after
differencing 142
Residuals:     SS =  10033560 (backforecasts excluded)
               MS =  71668   DF = 140


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag              12      24      36      48
Chi-Square     15.8    37.7    68.9    92.6
DF               10      22      34      46
P-Value       0.107   0.020   0.000   0.000
```

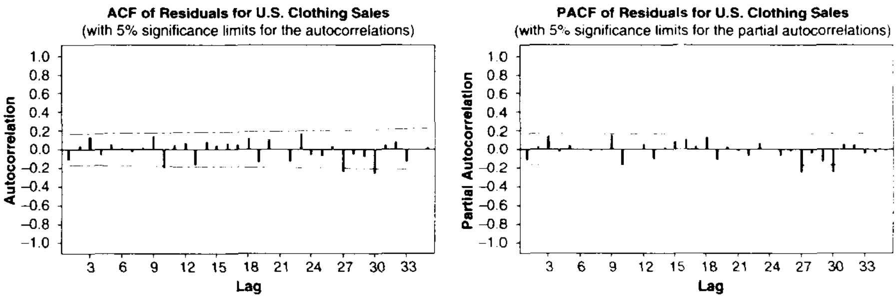


FIGURE 5.28 Sample ACF and PACF plots of residuals from the ARIMA(0, 1, 1) × (0, 1, 1)$_{12}$ model.

5.10 FINAL COMMENTS

ARIMA models (a.k.a. Box–Jenkins models) represent a very powerful and flexible class of models for time series analysis and forecasting. Over the years, they have been very successfully applied to many problems in research and practice. However, there might be certain situations where they may fall short on providing the "right" answers. For example, in ARIMA models, forecasting future observations primarily relies on the past data and implicitly assumes that the conditions at which the data is collected will remain the same in the future as well. In many situations this assumption may (and most likely will) not be appropriate. For those cases, the transfer function–noise

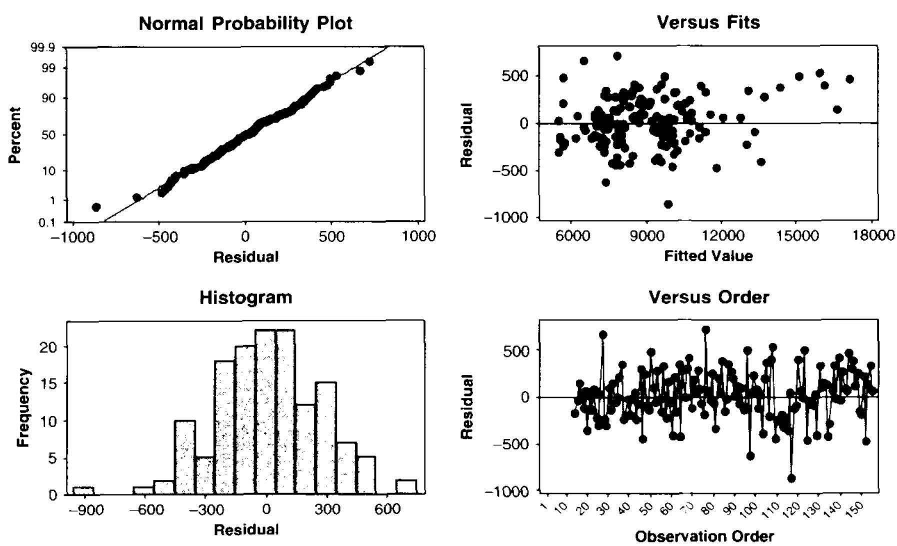


FIGURE 5.29 Residual plots from the ARIMA(0, 1, 1) × (0, 1, 1)$_{12}$ model for the U.S. clothing sales data.

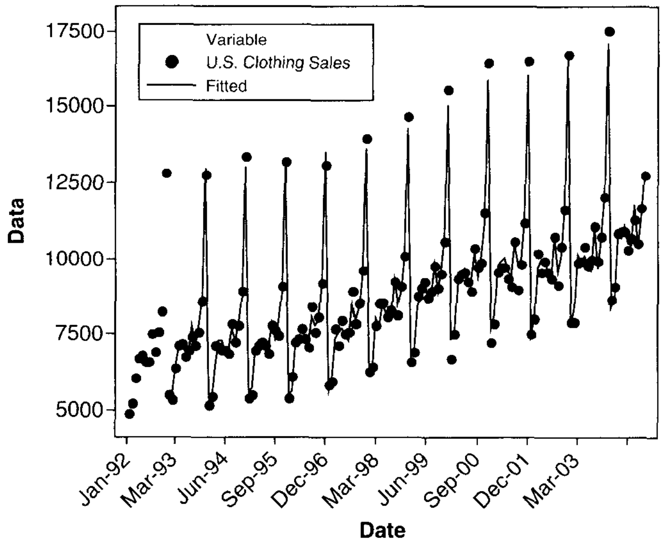


FIGURE 5.30 Time series plot of the actual data and fitted values from the ARIMA(0, 1, 1) × (0, 1, 1)$_{12}$ model for the U.S. clothing sales data.

models, where a set of input variables that may have an effect on the time series are added to the model, provide suitable options. We shall discuss these models in the next chapter. For an excellent discussion of this matter and of time series analysis and forecasting in general, see Jenkins [1979].

EXERCISES

5.1 Consider the time series data shown in Chapter 4, Table E4.2.

a. Fit an appropriate ARIMA model to the first 40 observations of this time series.

b. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors.

c. In Exercise 4.4 you used simple exponential smoothing with $\lambda = 0.2$ to smooth the first 40 time periods of this data and make forecasts of the last 10 observations. Compare the ARIMA forecasts with the exponential smoothing forecasts. How well do both of these techniques work?

5.2 Consider the time series data shown in Table E5.1.

a. Make a time series plot of the data.

b. Calculate and plot the sample autocorrelation and partial autocorrelation functions. Is there significant autocorrelation in this time series?

TABLE E5.1 Data for Exercise 5.2

Period	y_t	Period	y_t	Period	y_t	Period	y_t	Period	y_t
1	29	11	29	21	31	31	28	41	36
2	20	12	28	22	30	32	30	42	35
3	25	13	28	23	37	33	29	43	33
4	29	14	26	24	30	34	34	44	29
5	31	15	27	25	33	35	30	45	25
6	33	16	26	26	31	36	20	46	27
7	34	17	30	27	27	37	17	47	30
8	27	18	28	28	33	38	23	48	29
9	26	19	26	29	37	39	24	49	28
10	30	20	30	30	29	40	34	50	32

c. Identify and fit an appropriate ARIMA model to these data. Check for model adequacy.

d. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors.

5.3 Consider the time series data shown in Table E5.2.

a. Make a time series plot of the data.

b. Calculate and plot the sample autocorrelation and partial autocorrelation functions. Is there significant autocorrelation in this time series?

c. Identify and fit an appropriate ARIMA model to these data. Check for model adequacy.

d. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors.

TABLE E5.2 Data for Exercise 5.3

Period	y_t	Period	y_t	Period	y_t	Period	y_t	Period	y_t
1	500	11	508	21	475	31	639	41	637
2	496	12	510	22	485	32	679	42	606
3	450	13	512	23	495	33	674	43	610
4	448	14	503	24	500	34	677	44	620
5	456	15	505	25	541	35	700	45	613
6	458	16	494	26	555	36	704	46	593
7	472	17	491	27	565	37	727	47	578
8	495	18	487	28	601	38	736	48	581
9	491	19	491	29	610	39	693	49	598
10	488	20	486	30	605	40	65	50	613

5.4 Consider the time series model

$$y_t = 200 + 0.7y_{t-1} + \varepsilon_t$$

a. Is this a stationary time series process?

b. What is the mean of the time series?

c. If the current observation is $y_{100} = 750$, would you expect the next observation to be above or below the mean?

5.5 Consider the time series model

$$y_t = 150 - 0.5y_{t-1} + \varepsilon_t$$

a. Is this a stationary time series process?

b. What is the mean of the time series?

c. If the current observation is $y_{100} = 85$, would you expect the next observation to be above or below the mean?

5.6 Consider the time series model

$$y_t = 50 + 0.8y_{t-1} - 0.15 + \varepsilon_t$$

a. Is this a stationary time series process?

b. What is the mean of the time series?

c. If the current observation is $y_{100} = 160$, would you expect the next observation to be above or below the mean?

5.7 Consider the time series model

$$y_t = 20 + \varepsilon_t + 0.2\varepsilon_{t-1}$$

a. Is this a stationary time series process?

b. Is this an invertible time series?

c. What is the mean of the time series?

d. If the current observation is $y_{100} = 23$, would you expect the next observation to be above or below the mean? Explain your answer.

5.8 Consider the time series model

$$y_t = 50 + 0.8y_{t-1} + \varepsilon_t - 0.2\varepsilon_{t-1}$$

a. Is this a stationary time series process?

b. What is the mean of the time series?

c. If the current observation is $y_{100} = 270$, would you expect the next observation to be above or below the mean?

5.9 The data in Chapter 4, Table E4.4, exhibits a linear trend. Difference the data to remove the trend.

a. Fit an ARIMA model to the first differences.

b. Explain how this model would be used for forecasting.

5.10 Table B.1 in Appendix B contains data on the market yield on U.S. Treasury Securities at 10-year constant maturity.

a. Fit an ARIMA model to this time series, excluding the last 20 observations. Investigate model adequacy. Explain how this model would be used for forecasting.

b. Forecast the last 20 observations.

c. In Exercise 4.10, you were asked to use simple exponential smoothing with $\lambda = 0.2$ to smooth the data, and to forecast the last 20 observations. Compare the ARIMA and exponential smoothing forecasts. Which forecasting method do you prefer?

5.11 Table B.2 contains data on pharmaceutical product sales.

a. Fit an ARIMA model to this time series, excluding the last 10 observations. Investigate model adequacy. Explain how this model would be used for forecasting.

b. Forecast the last 10 observations.

c. In Exercise 4.12, you were asked to use simple exponential smoothing with $\lambda = 0.1$ to smooth the data, and to forecast the last 10 observations. Compare the ARIMA and exponential smoothing forecasts. Which forecasting method do you prefer?

d. How would prediction intervals be obtained for the ARIMA forecasts?

5.12 Table B.3 contains data on chemical process viscosity.

a. Fit an ARIMA model to this time series, excluding the last 20 observations. Investigate model adequacy. Explain how this model would be used for forecasting.

b. Forecast the last 20 observations.

c. Show how to obtain prediction intervals for the forecasts in part b above.

5.13 Table B.4 contains data on the annual U.S. production of blue and gorgonzola cheeses.

a. Fit an ARIMA model to this time series, excluding the last 10 observations. Investigate model adequacy. Explain how this model would be used for forecasting.

b. Forecast the last 10 observations.

c. In Exercise 4.16, you were asked to use exponential smoothing methods to smooth the data, and to forecast the last 10 observations. Compare the

ARIMA and exponential smoothing forecasts. Which forecasting method do you prefer?

d. How would prediction intervals be obtained for the ARIMA forecasts?

5.14 Reconsider the blue and gorgonzola cheese data in Table B.4 and Exercise 5.13. In Exercise 4.17 you were asked to take the first difference of this data and develop a forecasting procedure based on using exponential smoothing on the first differences. Compare this procedure with the ARIMA model of Exercise 5.13.

5.15 Table B.5 shows U.S. beverage manufacturer product shipments. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.16 Table B.6 contains data on the global mean surface air temperature anomaly. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.17 Reconsider the global mean surface air temperature anomaly data shown in Table B.6 and used in Exercise 5.16. In Exercise 4.20 you were asked to use simple exponential smoothing with the optimum value of λ to smooth the data. Compare the results with those obtained with the ARIMA model in Exercise 5.16.

5.18 Table B.7 contains daily closing stock prices for the Whole Foods Market. Develop an appropriate ARIMA model and a procedure for these data. Explain how prediction intervals would be computed.

5.19 Reconsider the Whole Foods Market data shown in Table B.7 and used in Exercise 5.18. In Exercise 4.22 you used simple exponential smoothing with the optimum value of λ to smooth the data. Compare the results with those obtained from the ARIMA model in Exercise 5.18.

5.20 Unemployment rate data is given in Table B.8. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.21 Reconsider the unemployment rate data shown in Table B.8 and used in Exercise 5.21. In Exercise 4.24 you used simple exponential smoothing with the optimum value of λ to smooth the data. Compare the results with those obtained from the ARIMA model in Exercise 5.20.

5.22 Table B.9 contains yearly data on the international sunspot numbers. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.23 Reconsider the sunspot data shown in Table B.9 and used in Exercise 5.22.

a. In Exercise 4.26 you were asked to use simple exponential smoothing with the optimum value of λ to smooth the data, and to use an exponential smoothing procedure for trends. How do these procedures compare to the ARIMA model from Exercise 5.22? Compare the results with those obtained in Exercise 4.26.

b. Do you think that using either exponential smoothing procedure would result in better forecasts than those from the ARIMA model?

5.24 Table B.10 contains seven years of monthly data on the number of airline miles flown in the United Kingdom. This is seasonal data.

a. Using the first six years of data, develop an appropriate ARIMA model and a procedure for these data.

b. Explain how prediction intervals would be computed.

c. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. How well did your procedure work in forecasting the new data?

5.25 Reconsider the airline mileage data in Table B.10 and used in Exercise 5.24.

a. In Exercise 4.27 you used Winters' method to develop a forecasting model using the first six years of data and you made forecasts for the last 12 months. Compare those forecasts with the ones you made using the ARIMA model from Exercise 5.24.

b. Which forecasting method would you prefer and why?

5.26 Table B.11 contains eight years of monthly champagne sales data. This is seasonal data.

a. Using the first seven years of data, develop an appropriate ARIMA model and a procedure for these data.

b. Explain how prediction intervals would be computed.

c. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. How well did your procedure work in forecasting the new data?

5.27 Reconsider the monthly champagne sales data in Table B.11 and used in Exercise 5.26.

a. In Exercise 4.29 you used Winters' method to develop a forecasting model using the first seven years of data and you made forecasts for the last 12 months. Compare those forecasts with the ones you made using the ARIMA model from Exercise 5.26.

b. Which forecasting method would you prefer and why?

5.28 Montgomery et al. [1990] give four years of data on monthly demand for a soft drink. These data are given in Chapter 4, Table E4.5.

a. Using the first three years of data, develop an appropriate ARIMA model and a procedure for these data.

b. Explain how prediction intervals would be computed.

c. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. How well did your procedure work in forecasting the new data?

5.29 Reconsider the soft drink demand data in Table E4.5 and used in Exercise 5.28.

a. In Exercise 4.31 you used Winters' method to develop a forecasting model using the first seven years of data and you made forecasts for the last 12 months. Compare those forecasts with the ones you made using the ARIMA model from the previous exercise.

b. Which forecasting method would you prefer and why?

5.30 Table B.12 presents data on the hourly yield from a chemical process and the operating temperature. Consider only the yield data in this exercise. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.31 Table B.13 presents data on ice cream and frozen yogurt sales. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.32 Table B.14 presents the CO_2 readings from Mauna Loa. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.33 Table B.15 presents data on the occurrence of violent crimes. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.34 Table B.16 presents data on the U.S. gross domestic product (GDP). Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.35 Total annual energy consumption is shown in Table B.17. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.36 Table B.18 contains data on coal production. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.37 Table B.19 contains data on the number of children 0–4 years old who drowned in Arizona. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.38 Data on tax refunds and population are shown in Table B.20. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.39 An ARIMA model has been fit to a time series, resulting in

$$\hat{y}_t = 25 + 0.35y_{t-1} + \varepsilon_t$$

a. Suppose that we are at time period $T = 100$ and $y_{100} = 31$. Determine forecasts for periods 101, 102, 103, ... from this model at origin 100.

b. What is the shape of the forecast function from this model?

c. Suppose that the observation for time period 101 turns out to be $y_{101} = 33$. Revise your forecasts for periods 102, 103, ... using period 101 as the new origin of time.

d. If your estimate $\hat{\sigma}^2 = 2$, find a 95% prediction interval on the forecast of period 101 made at the end of period 100.

5.40 The following ARIMA model has been fit to a time series:

$$\hat{y}_t = 25 + 0.8y_{t-1} - 0.3y_{t-2} + \varepsilon_t$$

a. Suppose that we are at the end of time period $T = 100$ and we know that $y_{100} = 40$ and $y_{99} = 38$. Determine forecasts for periods 101, 102, 103, ... from this model at origin 100.

b. What is the shape of the forecast function from this model?

c. Suppose that the observation for time period 101 turns out to be $y_{101} = 35$. Revise your forecasts for periods 102, 103, ... using period 101 as the new origin of time.

d. If your estimate $\hat{\sigma}^2 = 1$, find a 95% prediction interval on the forecast of period 101 made at the end of period 100.

5.41 The following ARIMA model has been fit to a time series:

$$\hat{y}_t = 25 + 0.8y_{t-1} - 0.2\varepsilon_{t-1} + \varepsilon_t$$

a. Suppose that we are at the end of time period $T = 100$ and we know that the forecast for period 100 was 130 and the actual observed value was $y_{100} = 140$. Determine forecasts for periods 101, 102, 103, ... from this model at origin 100.

b. What is the shape of the forecast function from this model?

c. Suppose that the observation for time period 101 turns out to be $y_{101} = 132$. Revise your forecasts for periods 102, 103, ... using period 101 as the new origin of time.

d. If your estimate $\hat{\sigma}^2 = 1.5$, find a 95% prediction interval on the forecast of period 101 made at the end of period 100.

5.42 The following ARIMA model has been fit to a time series:

$$\hat{y}_t = 20 + \varepsilon_t + 0.45\varepsilon_{t-1} - 0.3\varepsilon_{t-2}$$

a. Suppose that we are at the end of time period $T = 100$ and we know that the observed forecast error for period 100 was 0.5 and for period 99 we know that the observed forecast error was −0.8. Determine forecasts for periods 101, 102, 103, ... from this model at origin 100.

b. What is the shape of the forecast function that evolves from this model?

c. Suppose that the observations for the next four time periods turn out to be 17.5, 21.25, 18.75, and 16.75. Revise your forecasts for periods 102, 103, ... using a rolling horizon approach.

d. If your estimate $\hat{\sigma} = 0.5$, find a 95% prediction interval on the forecast of period 101 made at the end of period 100.

5.43 The following ARIMA model has been fit to a time series:

$$\hat{y}_t = 50 + \varepsilon_t + 0.5\varepsilon_{t-1}$$

a. Suppose that we are at the end of time period $T = 100$ and we know that the observed forecast error for period 100 was 2. Determine forecasts for periods 101, 102, 103, ... from this model at origin 100.

b. What is the shape of the forecast function from this model?

c. Suppose that the observations for the next four time periods turn out to be 53, 55, 46, and 50. Revise your forecasts for periods 102, 103, ... using a rolling horizon approach.

d. If your estimate $\hat{\sigma} = 1$, find a 95% prediction interval on the forecast of period 101 made at the end of period 100.

5.44 For each of the ARIMA models shown below, give the forecasting equation that evolves for lead times $\tau = 1, 2, \ldots, L$. In each case, explain the shape of the resulting forecast function over the forecast lead time.

a. AR(1)

b. AR(2)

c. MA(1)

d. MA(2)

e. ARMA(1, 1)

f. IMA(1, 1)

g. ARIMA(1, 1, 0)

5.45 Use a random number generator and generate 100 observations from the AR(1) model $y_t = 25 + 0.8y_{t-1} + \varepsilon_t$. Assume that the errors are normally and independently distributed with mean zero and variance $\sigma^2 = 1$.

a. Verify that your time series is AR(1).

b. Generate 100 observations for a $N(0, 1)$ process and add these random numbers to the 100 AR(1) observations in part a to create a new time series that is the sum of AR(1) and "white noise."

c. Find the sample autocorrelation and partial autocorrelation functions for the new time series created in part b. Can you identify the new time series?

d. Does this give you any insight about how the new time series might arise in practical settings?

5.46 Assume that you have fit the following model:

$$\hat{y}_t = y_{t-1} + 0.7\varepsilon_{t-1} + \varepsilon_t$$

a. Suppose that we are at the end of time period $T = 100$. What is the equation for forecasting the time series in period 101?

b. What does the forecast equation look like for future periods 102, 103, ...?

c. Suppose that we know that the observed value of y_{100} was 250 and forecast error in period 100 was 12. Determine forecasts for periods 101, 102, 103, ... from this model at origin 100.

d. If your estimate $\hat{\sigma} = 1$, find a 95% prediction interval on the forecast of period 101 made at the end of period 100.

e. Show the behavior of this prediction interval for future lead times beyond period 101. Are you surprised at how wide the interval is? Does this tell you something about the reliability of forecasts from this model at long lead times?

5.47 Consider the AR(1) model $y_t = 25 + 0.75y_{t-1} + \varepsilon_t$. Assume that the variance of the white noise process is $\sigma^2 = 1$.

a. Sketch the theoretical ACF and PACF for this model.

b. Generate 50 realizations of this AR(1) process and compute the sample ACF and PACF. Compare the sample ACF and the sample PACF to the theoretical ACF and PACF. How similar to the theoretical values are the sample values?

c. Repeat part b using 200 realizations. How has increasing the sample size impacted the agreement between the sample and theoretical ACF and PACF? Does this give you any insight about the sample sizes required for model building, or the reliability of models built to short time series?

5.48 Consider the AR(1) model $y_t = 25 + 0.75y_{t-1} + \varepsilon_t$. Assume that the variance of the white noise process is $\sigma^2 = 10$.

a. Sketch the theoretical ACF and PACF for this model.

b. Generate 50 realizations of this AR(1) process and compute the sample ACF and PACF. Compare the sample ACF and the sample PACF to the theoretical ACF and PACF. How similar to the theoretical values are the sample values?

c. Compare the results from part b with the results from part b of Exercise 5.47. How much has changing the variance of the white noise process impacted the results?

d. Repeat part b using 200 realizations. How has increasing the sample size impacted the agreement between the sample and theoretical ACF and PACF? Does this give you any insight about the sample sizes required for model building, or the reliability of models built to short time series?

e. Compare the results from part d with the results from part c of Exercise 5.47. How much has changing the variance of the white noise process impacted the results?

5.49 Consider the AR(2) model $y_t = 25 + 0.6y_{t-1} + 0.25y_{t-2} + \varepsilon_t$. Assume that the variance of the white noise process is $\sigma^2 = 1$.

a. Sketch the theoretical ACF and PACF for this model.

b. Generate 50 realizations of this AR(1) process and compute the sample ACF and PACF. Compare the sample ACF and the sample PACF to the theoretical ACF and PACF. How similar to the theoretical values are the sample values?

c. Repeat part b using 200 realizations. How has increasing the sample size impacted the agreement between the sample and theoretical ACF and PACF? Does this give you any insight about the sample sizes required for model building, or the reliability of models built to short time series?

5.50 Consider the MA(1) model $y_t = 40 + 0.4\varepsilon_{t-1} + \varepsilon_t$. Assume that the variance of the white noise process is $\sigma^2 = 2$.

a. Sketch the theoretical ACF and PACF for this model.

b. Generate 50 realizations of this AR(1) process and compute the sample ACF and PACF. Compare the sample ACF and the sample PACF to the theoretical ACF and PACF. How similar to the theoretical values are the sample values?

c. Repeat part b using 200 realizations. How has increasing the sample size impacted the agreement between the sample and theoretical ACF and PACF? Does this give you any insight about the sample sizes required for model building, or the reliability of models built to short time series?

5.51 Consider the ARMA(1, 1) model $y_t = 50 - 0.7y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t$. Assume that the variance of the white noise process is $\sigma^2 = 2$.

a. Sketch the theoretical ACF and PACF for this model.

b. Generate 50 realizations of this AR(1) process and compute the sample ACF and PACF. Compare the sample ACF and the sample PACF to the theoretical ACF and PACF. How similar to the theoretical values are the sample values?

c. Repeat part b using 200 realizations. How has increasing the sample size impacted the agreement between the sample and theoretical ACF and PACF? Does this give you any insight about the sample sizes required for model building, or the reliability of models built to short time series?