Wei: "Time Series"

5 FORECASTING

Uncertainty is a fact of life for both individuals and organizations. Forecasting is essential for planning and operation control in a variety of areas such as production management, inventory systems, quality control, financial planning, and investment analysis. In this chapter we develop the minimum mean square error forecasts for the stationary and nonstationary time series models introduced in Chapters 3 and 4. These models can also be used to update forecasts when new information becomes available. We also discuss the implication of the constructed time series model in terms of its eventual forecast function.

5.1 INTRODUCTION

One of the most important objectives in the analysis of a time series is to forecast its future values. Even if the final purpose of time series modeling is for the control of a system, its operation is usually based on forecasting. The term *forecasting* is used more frequently in recent time series literature than the term *prediction*. However, most forecasting results are derived from a general theory of linear prediction developed by Kolmogorov (1939, 1941), Wiener (1949), Kalman (1960), Yaglom (1962), and Whittle (1983), among others.

Consider the general ARIMA(p, d, q) model

$$\phi(B)(1-B)^d Z_t = \theta(B)a_t \tag{5.1.1}$$

where $\phi(B) = (1-\phi_1 B - \cdots - \phi_p B^p)$, $\theta(B) = (1-\theta_1 B - \cdots - \theta_q B^q)$, and the series a_t is a Gaussian $N(0, \sigma_a^2)$ white noise process. The deterministic trend parameter θ_0 is omitted for simplicity but no loss of generality. Equation (5.1.1) is one of the most commonly used models in forecasting applications. We discuss the minimum mean square error forecast of this model for both cases when $d = 0$ and $d \neq 0$.

5.2 MINIMUM MEAN SQUARE ERROR FORECASTS

5.2.1 Minimum Mean Square Error Forecasts for ARMA Models

To derive the minimum mean square error forecasts, we first consider the case when $d = 0$, i.e., the stationary ARMA model

$$\phi(B)Z_t = \theta(B)a_t. \tag{5.2.1}$$

Because the model is stationary, we can rewrite it in a moving average representation,

$$\begin{aligned} Z_t &= \psi(B)a_t \\ &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \end{aligned} \tag{5.2.2}$$

where

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j = \frac{\theta(B)}{\phi(B)} \tag{5.2.3}$$

and $\psi_0 = 1$. For $t = n + l$, we have

$$Z_{n+l} = \sum_{j=0}^{\infty} \psi_j a_{n+l-j}. \tag{5.2.4}$$

Suppose at time $t = n$ we have the observations $Z_n, Z_{n-1}, Z_{n-2}, \ldots$ and wish to forecast l-step ahead of future value Z_{n+l} as a linear combination of the observations $Z_n, Z_{n-1}, Z_{n-2}, \ldots$. Since Z_t for $t = n, n-1, n-2, \ldots$ can all be written in the form of (5.2.2), we can let the minimum mean square error forecast $\hat{Z}_n(l)$ of Z_{n+l} be

$$\hat{Z}_n(l) = \psi_l^* a_n + \psi_{l+1}^* a_{n-1} + \psi_{l+2}^* a_{n-2} + \cdots \tag{5.2.5}$$

where the ψ_j^* are to be determined. The mean square error of the forecast is

$$E(Z_{n+l} - \hat{Z}_n(l))^2 = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2 + \sigma_a^2 \sum_{j=0}^{\infty} \left[\psi_{l+j} - \psi_{l+j}^*\right]^2,$$

which is easily seen to be minimized when $\psi_{l+j}^* = \psi_{l+j}$. Hence,

$$\hat{Z}_n(l) = \psi_l a_n + \psi_{l+1} a_{n-1} + \psi_{l+2} a_{n-2} + \cdots. \tag{5.2.6}$$

But using (5.2.4) and the fact that

$$E(a_{n+j} \mid Z_n, Z_{n-1}, \ldots) = \begin{cases} 0, & j > 0, \\ a_{n+j}, & j \leq 0, \end{cases}$$

we have

$$E(Z_{n+l} \mid Z_n, Z_{n-1}, \ldots) = \psi_l a_n + \psi_{l+1} a_{n-1} + \psi_{l+2} a_{n-2} + \cdots.$$

Thus, the minimum mean square error forecast of Z_{n+l} is given by its conditional expectation. That is,

$$\hat{Z}_n(l) = E(Z_{n+l} \mid Z_n, Z_{n-1}, \ldots). \tag{5.2.7}$$

$\hat{Z}_n(l)$ is usually read as the l-step ahead of the forecast of Z_{n+l} at the forecast origin n.

The forecast error is

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = \sum_{j=0}^{l-1} \psi_j a_{n+l-j}. \tag{5.2.8}$$

Because $E(e_n(l) \mid Z_t, t \le n) = 0$, the forecast is unbiased with variance

$$\mathrm{Var}(\hat{Z}_n(l)) = \mathrm{Var}(e_n(l)) = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2. \tag{5.2.9}$$

For a normal process, the $(1-\alpha)100\%$ forecast limits are

$$\hat{Z}_n(l) \pm N_{\frac{\alpha}{2}} \left[1 + \sum_{j=1}^{l-1} \psi_j^2\right]^{1/2} \sigma_a \tag{5.2.10}$$

where $N_{\alpha/2}$ is the standard normal deviate such that $P(N > N_{\alpha/2}) = \alpha/2$.

The forecast error $e_n(l)$ as shown in (5.2.8) is a linear combination of the future random shocks entering the system after time n. Specifically, the one-step ahead forecast error is

$$e_n(1) = Z_{n+1} - \hat{Z}_n(1) = a_{n+1}. \tag{5.2.11}$$

Thus, the one-step ahead forecast errors are independent. This implies that $\hat{Z}_n(1)$ is indeed the best forecast of Z_{n+1}. Otherwise, if one-step ahead forecast errors are correlated, then one can calculate the forecast $\hat{a}_{n+1}$ of a_{n+1} from the available errors $a_n, a_{n-1}, a_{n-2} \ldots$ and hence improve the forecast of Z_{n+1} by simply using $\hat{Z}_n(1) + \hat{a}_{n+1}$ as the forecast. However, the forecast errors for longer lead times are correlated. This is true for the forecast errors

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = a_{n+l} + \psi_1 a_{n+l-1} + \cdots + \psi_{l-1} a_{n+1} \tag{5.2.12}$$

and

$$e_{n-j}(l) = Z_{n+l-j} - \hat{Z}_{n-j}(l) = a_{n+l-j} + \psi_1 a_{n+l-j-1} + \cdots + \psi_{l-1} a_{n-j+1}, \tag{5.2.13}$$

which are made at the same lead time l but different origins n and $n-j$ for $j < l$. It is also true for the forecast errors for different lead times made from the same time origin. For example,

$$\mathrm{Cov}[e_n(2), e_n(1)] = E[(a_{n+2} + \psi_1 a_{n+1})(a_{n+1})] = \psi_1 \sigma_a^2. \tag{5.2.14}$$

5.2.2 Minimum Mean Square Error Forecasts for ARIMA Models

We now consider the general nonstationary ARIMA(p, d, q) model with $d \neq 0$, i.e.,

$$\phi(B)(1-B)^d Z_t = \theta(B) a_t, \tag{5.2.15}$$

where $\phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ is a stationary AR operator and $\theta(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$ is an invertible MA operator, respectively. It is interesting to note that although for this process the mean and the second order moments such as the variance and the autocovariance functions vary over time, as shown in Chapter 4, the complete evolution of the process is completely determined by a finite number of fixed parameters. Hence, we can view the forecast of the process as the estimation of a function of these parameters and obtain the minimum mean square error forecast using a Bayesian argument. It is well known that using this approach with respect to the mean square error criterion, which corresponds to a squared loss function, when the series is known up to time n, the optimal forecast of Z_{n+l} is given by its conditional expectation $E(Z_{n+l} \mid Z_n, Z_{n-1}, \ldots)$. The minimum mean square error forecast for the stationary ARMA model discussed earlier is, of course, a special case of the forecast for the ARIMA(p, d, q) model with $d = 0$.

To derive the variance of the forecast for the general ARIMA model, we rewrite the model at time $t+l$ in an AR representation that exists because the model is invertible. Thus,

$$\pi(B) Z_{t+l} = a_{t+l}, \tag{5.2.16}$$

where

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j = \frac{\phi(B)(1-B)^d}{\theta(B)}, \tag{5.2.17}$$

or equivalently

$$Z_{t+l} = \sum_{j=1}^{\infty} \pi_j Z_{t+l-j} + a_{t+l}. \tag{5.2.18}$$

Following Wegman (1986), we apply the operator

$$1 + \psi_1 B + \cdots + \psi_{l-1} B^{l-1}$$

to (5.2.18) and obtain

$$\sum_{j=0}^{\infty} \sum_{k=0}^{l-1} \pi_j \psi_k Z_{t+l-j-k} + \sum_{k=0}^{l-1} \psi_k a_{t+l-k} = 0 \tag{5.2.19}$$

where $\pi_0 = -1$ and $\psi_0 = 1$. It can be easily shown that

$$\sum_{j=0}^{\infty}\sum_{k=0}^{l-1}\pi_j\psi_k Z_{t+l-j-k} = \pi_0 Z_{t+l} + \sum_{m=1}^{l-1}\sum_{i=0}^{m}\pi_{m-i}\psi_i Z_{t+l-m} + \sum_{j=1}^{\infty}\sum_{i=0}^{l-1}\pi_{l-1+j-i}\psi_i Z_{t-j+1}. \tag{5.2.20}$$

Choosing ψ weights so that

$$\sum_{i=0}^{m}\pi_{m-i}\psi_i = 0, \qquad \text{for } m = 1, 2, \ldots, l-1, \tag{5.2.21}$$

we have

$$Z_{t+l} = \sum_{j=1}^{\infty}\pi_j^{(l)} Z_{t-j+1} + \sum_{i=0}^{l-1}\psi_i a_{t+l-i} \tag{5.2.22}$$

where

$$\pi_j^{(l)} = \sum_{i=0}^{l-1}\pi_{l-1+j-i}\psi_i. \tag{5.2.23}$$

Thus, given Z_t, for $t \le n$, we have

$$\begin{aligned}\hat{Z}_n(l) &= E(Z_{n+l} \mid Z_t, t \le n)\\ &= \sum_{j=1}^{\infty}\pi_j^{(l)} Z_{t-j+1},\end{aligned} \tag{5.2.24}$$

because $E(a_{n+j} \mid Z_t, t \le n) = 0$, for $j > 0$. The forecast error is

$$\begin{aligned}e_n(l) &= Z_{n+l} - \hat{Z}_n(l)\\ &= \sum_{j=0}^{l-1}\psi_j a_{n+l-j}\end{aligned} \tag{5.2.25}$$

where the ψ_j weights, by (5.2.21), can be calculated recursively from the π_j weights as follows:

$$\psi_j = \sum_{i=0}^{j-1}\pi_{j-i}\psi_i, \qquad j = 1, \ldots, l-1. \tag{5.2.26}$$

Note that (5.2.25) is exactly the same as (5.2.8). Hence, the results given in (5.2.7) through (5.2.14) hold for both stationary and nonstationary ARMA models.

For a stationary process, $\lim_{l\to\infty}\sum_{j=0}^{l-1}\psi_j^2$ exists. Hence, from (5.2.10), the eventual forecast limits approach two horizontally parallel lines as shown in Figure 5.1(a). For a nonstationary process, $\lim_{l\to\infty}\sum_{j=0}^{l-1}\psi_j^2$ does not exist. In fact, $\sum_{j=0}^{l-1}\psi_j^2$ increases and becomes unbounded as $l \to \infty$. Thus, the forecast limits in this case become wider and wider as the forecast lead l becomes larger and larger, as shown in Figure 5.1(b). The practical implication of the latter case is that the forecaster becomes less certain about the result as the forecast lead time gets larger. For more discussion on the properties of the mean square error forecasts, see Shaman (1983).

5.3 COMPUTATION OF FORECASTS

From the result that the minimum mean square error forecasts $\hat{Z}_n(l)$ of Z_{n+l} at the forecast origin n is given by the conditional expectation

$$\hat{Z}_n(l) = E(Z_{n+l} \mid Z_n, Z_{n-1}, \ldots),$$

we can easily obtain the actual forecasts by directly using the difference equation form of the model. Let

$$\Psi(B) = \phi(B)(1-B)^d = (1 - \Psi_1 B - \cdots - \Psi_{p+d}B^{p+d}).$$

The general ARIMA(p, d, q) model (5.2.15) can be written as the following difference equation form:

$$(1 - \Psi_1 B - \cdots - \Psi_{p+d}B^{p+d})Z_t = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t. \tag{5.3.1}$$

For $t = n + l$, we have

$$\begin{aligned}Z_{n+l} = {} & \Psi_1 Z_{n+l-1} + \Psi_2 Z_{n+l-2} + \cdots + \Psi_{p+d} Z_{n+l-p-d}\\ & + a_{n+l} - \theta_1 a_{n+l-1} - \cdots - \theta_q a_{n+l-q}.\end{aligned} \tag{5.3.2}$$

Taking the conditional expectation at time origin n, we obtain

$$\begin{aligned}\hat{Z}_n(l) = {} & \Psi_1\hat{Z}_n(l-1) + \cdots + \Psi_{p+d}\hat{Z}_n(l-p-d)\\ & + \hat{a}_n(l) - \theta_1\hat{a}_n(l-1) - \cdots - \theta_q\hat{a}_n(l-q)\end{aligned} \tag{5.3.3}$$

where

$$\begin{aligned}&\hat{Z}_n(j) = E(Z_{n+j} \mid Z_n, Z_{n-1}, \ldots), && j \ge 1,\\ &\hat{Z}_n(j) = Z_{n+j}, && j \le 0,\\ &\hat{a}_n(j) = 0, && j \ge 1,\end{aligned}$$

and

$$\hat{a}_n(j) = Z_{n+j} - \hat{Z}_{n+j-1}(1) = a_{n+j}, j \le 0.$$

Example 5.1 To illustrate the above results, we consider the l-step ahead forecast $\hat{Z}_n(l)$ of Z_{n+l} for the following ARIMA(1, 0, 1) or ARMA(1, 1) model:

$$(1 - \phi B)(Z_t - \mu) = (1 - \theta B)a_t. \tag{5.3.4}$$

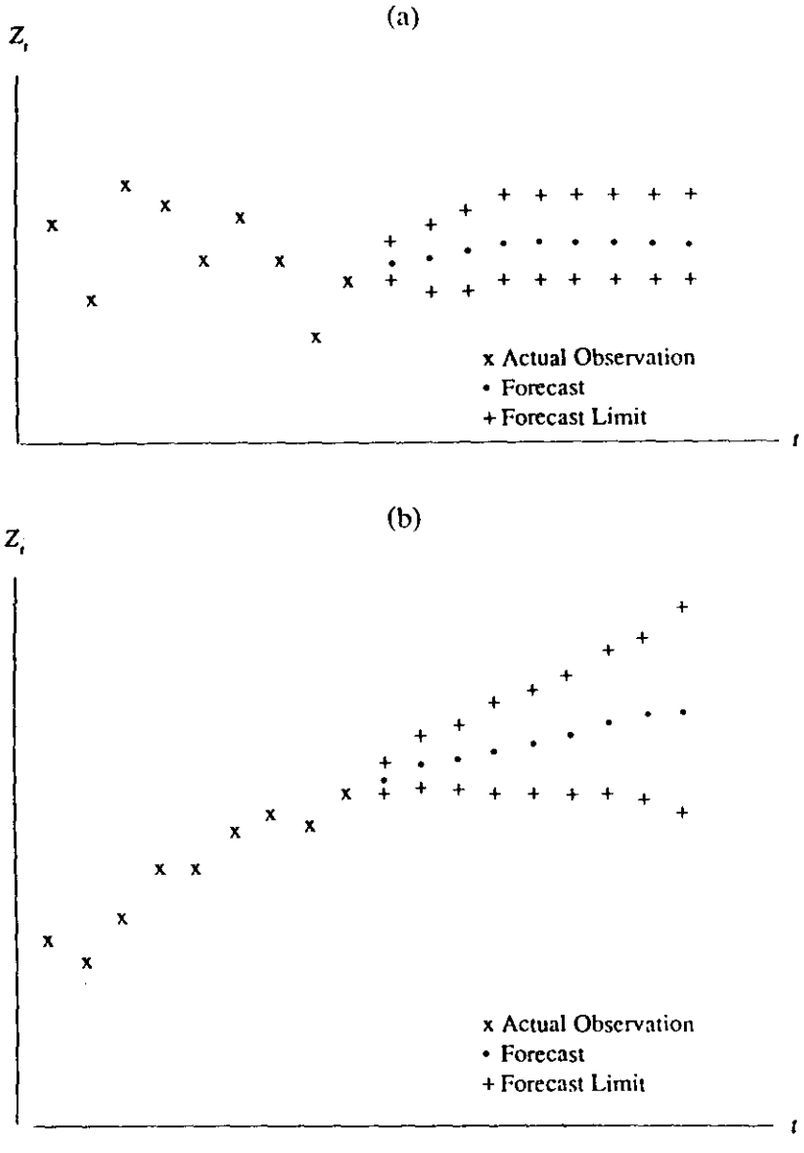


Fig. 5.1 (a) Forecasts for stationary processes. (b) Forecasts for nonstationary processes

1. Calculate the forecast $\hat{Z}_n(l)$ as the conditional expectation from the difference equation form.

For $t = n + l$, we can rewrite the above model in (5.3.4) as

$$Z_{n+l} = \mu + \phi(Z_{n+l-1} - \mu) + a_{n+l} - \theta a_{n+l-1}. \tag{5.3.5}$$

Hence

$$\hat{Z}_n(1) = \mu + \phi(Z_n - \mu) - \theta a_n \tag{5.3.6a}$$

and

$$\begin{aligned}\hat{Z}_n(l) &= \mu + \phi[\hat{Z}_n(l-1) - \mu] \\ &= \mu + \phi^l(Z_n - \mu) - \phi^{l-1}\theta a_n, \qquad l \geq 2.\end{aligned} \tag{5.3.6b}$$

2. Calculate the forecast variance $\text{Var}(\hat{Z}_n(l)) = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2$.

When $|\phi| < 1$, we can calculate the ψ weights from the moving average representation (5.2.2) with $\phi(B) = (1 - \phi B)$ and $\theta(B) = (1 - \theta B)$. That is,

$$(1 - \phi B)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = (1 - \theta B). \tag{5.3.7}$$

Equating the coefficients of B^j on both sides gives

$$\psi_j = \phi^{j-1}(\phi - \theta), \qquad j \geq 1. \tag{5.3.8}$$

Hence, the forecast variance becomes

$$\text{Var}(\hat{Z}_n(l)) = \sigma_a^2 \left\{ 1 + \sum_{j=1}^{l-1} [\phi^{j-1}(\phi - \theta)]^2 \right\}, \tag{5.3.9}$$

which converges to $\sigma_a^2[1 + (\phi - \theta)^2/(1 - \phi^2)]$.

When $\phi = 1$, which corresponds to taking the first difference, the model in (5.3.4) becomes an IMA(1, 1) process

$$(1 - B)Z_t = (1 - \theta B)a_t, \tag{5.3.10}$$

where we note that $(1 - B)\mu = 0$. To calculate the ψ weights needed for the forecast variance, since the MA representation does not exist, we first rewrite (5.3.10) in an AR form when time equals $t + l$, i.e.,

$$\pi(B)Z_{t+l} = a_{t+l}$$

where

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \cdots - = \frac{(1 - B)}{(1 - \theta B)}$$

or

$$(1 - B) = 1 - (\pi_1 + \theta)B - (\pi_2 - \pi_1\theta)B^2 - (\pi_3 - \pi_2\theta)B^3 - \cdots$$

Equating the coefficients of B^j on both sides gives

$$\pi_j = (1 - \theta)\theta^{j-1}, \qquad j \geq 1. \tag{5.3.11}$$

Now, applying (5.2.26) we obtain

$$\psi_1 = \pi_1 = (1-\theta),$$
$$\psi_2 = \pi_2 + \pi_1\psi_1 = (1-\theta)\theta + (1-\theta)^2 = (1-\theta),$$
$$\vdots$$

That is, we have $\psi_j = (1-\theta)$, $1 \le j \le l-1$. Thus, the variance of $\hat{Z}_n(l)$, from (5.2.9), becomes

$$\mathrm{Var}(\hat{Z}_n(l)) = \sigma_a^2[1 + (l-1)(1-\theta)^2], \tag{5.3.12}$$

which approaches $+\infty$ as $l \to \infty$.

As expected, (5.3.12) is the limiting case of (5.3.9) when $\phi \to 1$. Thus, when ϕ is close to 1, the choice between a stationary ARMA(1, 1) model and a nonstationary IMA(1, 1) model has very different implications for forecasting. This can be seen even more clearly from the limiting behavior of the l-step ahead forecast $\hat{Z}_n(l)$ in (5.3.6b). For $|\phi| < 1$, $\hat{Z}_n(l)$ approaches the mean, μ, of the process as $l \to \infty$. When $\phi \to 1$, the first equation of (5.3.6b) implies that $\hat{Z}_n(l) = \hat{Z}_n(l-1)$ for all l. That is, the forecast is free to wander, with no tendency for the values to remain clustered around a fixed level.

5.4 THE ARIMA FORECAST AS A WEIGHTED AVERAGE OF PREVIOUS OBSERVATIONS

Recall that we can always represent an invertible ARIMA model in an autoregressive representation. In this representation, Z_t is expressed as an infinite weighted sum of previous observations plus a random shock, i.e.,

$$Z_{n+l} = \sum_{j=1}^{\infty} \pi_j Z_{n+l-j} + a_{n+l}, \tag{5.4.1}$$

or equivalently

$$\pi(B)Z_{n+l} = a_{n+l},$$

where

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j = \frac{\phi(B)(1-B)^d}{\theta(B)}. \tag{5.4.2}$$

Thus,

$$\hat{Z}_n(l) = \sum_{j=1}^{\infty} \pi_j \hat{Z}_n(l-j), \qquad l \ge 1. \tag{5.4.3}$$

By repeated substitutions, we see that $\hat{Z}_n(l)$ can be expressed as a weighted sum of current and past observations Z_t, for $t \le n$. For example,

$$\begin{aligned}\hat{Z}_n(1) &= \pi_1 Z_n + \pi_2 Z_{n-1} + \pi_3 Z_{n-2} + \cdots \\ &= \sum_{j=1}^{\infty} \pi_j Z_{n+1-j} \\ \hat{Z}_n(2) &= \pi_1 \hat{Z}_n(1) + \pi_2 Z_n + \pi_3 Z_{n-1} + \cdots \\ &= \pi_1 \sum_{j=1}^{\infty} \pi_j Z_{n+1-j} + \sum_{j=1}^{\infty} \pi_{j+1} Z_{n+1-j} \\ &= \sum_{j=1}^{\infty} \pi_j^{(2)} Z_{n+1-j}\end{aligned}$$

where

$$\pi_j^{(2)} = \pi_1\pi_j + \pi_{j+1}, \qquad j \ge 1.$$

More generally, it can be shown by successive substitutions that

$$\hat{Z}_n(l) = \sum_{j=1}^{\infty} \pi_j^{(l)} Z_{n+1-j}, \tag{5.4.4}$$

where

$$\pi_j^{(l)} = \pi_{j+l-1} + \sum_{i=1}^{l-1} \pi_i \pi_j^{(l-i)}, \qquad l > 1, \tag{5.4.5}$$

and

$$\pi_j^{(1)} = \pi_j.$$

Thus, many smoothing results, such as moving averages and exponential smoothing, are special cases of ARIMA forecasting. ARIMA models provide a very natural and optimal way to obtain the required weights for forecasting. The user does not have to specify either the number or the form of weights as required in the moving average method and the exponential smoothing method. It should also be noted that the ARIMA forecasts are minimum mean square error forecasts. This optimal property is not shared in general by the moving average and the exponential smoothing methods.

For an invertible process, the π weights in (5.4.3) or (5.4.4) form a convergent series. This implies that for a given degree of accuracy, $\hat{Z}_n(l)$ depends only on a finite number of recent observations. The associated π weights provide very useful information for many important managerial decisions.

Example 5.2 For the ARMA(1, 1) model in (5.3.4) with $|\theta| < 1$, we have from (5.4.2) with $d = 0$

$$(1-\phi B) = (1-\pi_1 B - \pi_2 B^2 - \cdots)(1-\theta B) \tag{5.4.6}$$

or

$$(1-\phi B) = 1-(\pi_1+\theta)B-(\pi_2-\pi_1\theta)B^2-(\pi_3-\pi_2\theta)B^3-\cdots.$$

Equating the coefficients or B^j on both sides gives

$$\pi_j = (\phi-\theta)\theta^{j-1}, \qquad j \geq 1. \tag{5.4.7}$$

Assuming that $\mu = 0$, from (5.4.3), we have

$$\hat{Z}_n(l) = \sum_{j=1}^{\infty}(\phi-\theta)\theta^{j-1}\hat{Z}_n(l-j). \tag{5.4.8}$$

When $l = 1$, we get

$$\hat{Z}_n(1) = (\phi-\theta)\sum_{j=1}^{\infty}\theta^{j-1}Z_{n+1-j}. \tag{5.4.9}$$

For $l = 2$, from (5.4.4) and (5.4.5),

$$\begin{aligned}\hat{Z}_n(2) &= \sum_{j=1}^{\infty}\pi_j^{(2)}Z_{n+1-j}\\ &= \sum_{j=1}^{\infty}[\pi_{j+1}+\pi_1\pi_j]Z_{n+1-j}\\ &= \sum_{j=1}^{\infty}[(\phi-\theta)\theta^j+(\phi-\theta)^2\theta^{j-1}]Z_{n+1-j}\\ &= \phi(\phi-\theta)\sum_{j=1}^{\infty}\theta^{j-1}Z_{n+1-j}.\end{aligned} \tag{5.4.10}$$

Again $\hat{Z}_n(2)$ is a weighted average of the previous observations with the $\pi_j^{(2)}$ weights given by $\pi_j^{(2)} = \phi(\phi-\theta)\theta^{j-1}$ for $j \geq 1$. Note that comparing (5.4.9) and (5.4.10), we see that

$$\hat{Z}_n(2) = \phi\hat{Z}_n(1), \tag{5.4.11}$$

which as expected, agrees with (5.3.6b) with $l = 2$ and $\mu = 0$.

To see the implications of these weights, let us examine (5.4.9) more carefully. For $|\theta| < 1$, $\pi_j = \theta^{j-1}(\phi-\theta)$ converges to 0 as $j \to 0$. This implies that the more recent observations have a greater influence on the forecast. For $|\theta| \geq 1$, although the model is still stationary, its AR representation does not exist. To see the trouble, note that if $|\theta| > 1$, the π weights rapidly approach $-\infty$ or $+\infty$ as j increases. This implies that the more remote past observations have a much greater influence on the forecast. When $|\theta| = 1$, the π weights become $\pi_j = (\phi-1)$ for $\theta = 1$, and $\pi_j = (-1)^j(1+\phi)$ for $\theta = -1$, which have the same absolute value for all j. This means that all past and present observations are equally important in their effect on the forecast. Thus, a meaningful forecast can be derived only from an invertible process. The models corresponding to $|\theta| > 1$ and $|\theta| = 1$ are both noninvertible.

5.5 UPDATING FORECASTS

Recall that when a time series Z_t is available for $t \leq n$, the l-step ahead minimum mean square forecast error for the forecast origin n using the general ARIMA model is obtained in (5.2.25), which, for convenience, is listed again in the following:

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = \sum_{j=0}^{l-1}\psi_j a_{n+l-j}. \tag{5.5.1}$$

In particular, the one-step ahead forecast error is

$$e_n(1) = Z_{n+1} - \hat{Z}_n(1) = a_{n+1}. \tag{5.5.2}$$

Clearly, the result can be reversed, giving

$$Z_n - \hat{Z}_{n-1}(1) = a_n. \tag{5.5.3}$$

From (5.5.1), it is clear that

$$e_{n-1}(l+1) = e_n(l) + \psi_l a_n, \tag{5.5.4}$$

where

$$e_{n-1}(l+1) = Z_{n+l} - \hat{Z}_{n-1}(l+1),$$

and

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l).$$

Hence, after substituting, rearranging, and using (5.5.3), we have the following updating equation:

$$\hat{Z}_n(l) = \hat{Z}_{n-1}(l+1) + \psi_l[Z_n - \hat{Z}_{n-1}(1)], \tag{5.5.5}$$

or equivalently

$$\hat{Z}_{n+1}(l) = \hat{Z}_n(l+1) + \psi_l[Z_{n+1} - \hat{Z}_n(1)]. \tag{5.5.6}$$

The updated forecast is obtained by adding to the previous forecast a constant multiple ψ_l of the one-step ahead forecast error $a_{n+1} = Z_{n+1} - \hat{Z}_n(1)$. This is certainly sensible. For example, when the value Z_{n+1} becomes available and

is found to be higher than the previous forecast, (resulting in a positive forecast error $a_{n+1} = Z_{n+1} - \hat{Z}_n(1)$), we will naturally modify the forecast $\hat{Z}_n(l+1)$ made earlier by proportionally adding a constant multiple of this error.

5.6 EVENTUAL FORECAST FUNCTIONS

Let the ARIMA model be

$$\Psi(B)Z_t = \theta(B)a_t$$

where $\Psi(B) = \phi(B)(1-B)^d$. Recall from Equation (5.3.3) that

$$\hat{Z}_n(l) = \Psi_1\hat{Z}_n(l-1) + \Psi_2\hat{Z}_n(l-2) + \cdots + \Psi_{p+d}\hat{Z}_n(l-p-d) + \hat{a}_n(l) - \theta_1\hat{a}_n(l-1) - \cdots - \theta_q\hat{a}_n(l-q).$$

When $l > q$, $\hat{Z}_n(l)$ becomes

$$\hat{Z}_n(l) = \Psi_1\hat{Z}_n(l-2) + \cdots + \Psi_{p+d}\hat{Z}_n(l-p-d)$$

or

$$\Psi(B)\hat{Z}_n(l) = 0. \tag{5.6.1}$$

That is, $\hat{Z}_n(l)$ for $l > q$ satisfies the homogeneous difference equation of order $(p+d)$. Let $\Psi(B) = \prod_{i=1}^{N}(1-R_iB)^{m_i}$ with $\sum_{i=1}^{N} m_i = (p+d)$. Then from Theorem 2.7.1, the general solution is given by

$$\hat{Z}_n(l) = \sum_{i=1}^{N}\left(\sum_{j=0}^{m_i-1} C_{ij}^{(n)} l^j\right) R_i^l \tag{5.6.2}$$

for $l > (q-p-d)$, where the $C_{ij}^{(n)}$ are constants that depend on time origin n. For a given forecast origin n, they are fixed constants for all forecast lead times l. The constants change when the forecast origin is changed.

The solution in (5.6.2) is called the eventual forecast function because it holds only for $l > (q-p-d)$. When $q < (p+d)$, the eventual forecast function actually holds for all lead times $l > 0$. In general, the function is the unique curve that passes through the $(p+d)$ values given by $\hat{Z}_n(q)$, $\hat{Z}_n(q-1)$, ..., $\hat{Z}_n(q-p-d+1)$, where $\hat{Z}_n(-j) = Z_{n-j}$ for $j \geq 0$. For the ARIMA(p, d, q) model with $q = 0$, i.e., the ARIMA$(p, d, 0)$ model, the function passes through the points $Z_n, Z_{n-1}, \ldots$, and $Z_{n-p-d+1}$.

Example 5.3 For the ARIMA(1, 0, 1) model given in (5.3.4),

$$(1-\phi B)(Z_t - \mu) = (1-\theta B)a_t,$$

the forecast $\hat{Z}_n(l)$ satisfies the difference equation $(1-\phi B)(\hat{Z}_t(l) - \mu) = 0$ for $l > 1$. The eventual forecast function is given by

$$\hat{Z}_n(l) - \mu = C_1^{(n)}\phi^l$$

or

$$\hat{Z}_n(l) = \mu + C_1^{(n)}\phi^l \tag{5.6.3}$$

for $l > (q-p-d) = 0$ and constant $C_1^{(n)}$. The eventual forecast function passes through $\hat{Z}_n(1)$. As $l \to \infty$, $\hat{Z}_n(l)$ approaches the mean μ of the stationary process as expected.

Example 5.4 Consider the ARIMA(1, 1, 1) model,

$$(1-\phi B)(1-B)Z_t = (1-\theta B)a_t.$$

The eventual forecast function is the solution of $(1-\phi B)(1-B)\hat{Z}_n(l) = 0$ for $l > 1$, and is given by

$$\hat{Z}_n(l) = C_1^{(n)} + C_2^{(n)}\phi^l \tag{5.6.4}$$

for $l > (q-p-d) = -1$. The function passes through the first forecast $\hat{Z}_n(1)$ and the last observation Z_n.

Example 5.5 Consider the ARIMA(0, 2, 1) model,

$$(1-B)^2 Z_t = (1-\theta B)a_t.$$

The eventual forecast function is the solution of $(1-B)^2\hat{Z}_n(l) = 0$ for $l > 1$, and from (5.6.2) it is given by

$$\hat{Z}_n(l) = C_1^{(n)} + C_2^{(n)}l \tag{5.6.5}$$

for $l > (q-p-d) = -1$. The function is a straight line passing through $\hat{Z}_n(1)$ and Z_n.

5.7 A NUMERICAL EXAMPLE

As a numerical example, consider the AR(1) model,

$$(1-\phi B)(Z_t - \mu) = a_t$$

with $\phi = .6$, $\mu = 9$, and $\sigma_a^2 = .1$. Suppose that we have the observations $Z_{100} = 8.9$, $Z_{99} = 9$, $Z_{98} = 9$, $Z_{97} = 9.6$, and want to forecast Z_{101}, Z_{102}, Z_{103}, and Z_{104} with their associated 95% forecast limits. We proceed as follows:

1. The AR(1) model can be written as

$$Z_t - \mu = \phi(Z_{t-1} - \mu) + a_t,$$

and the general form of the forecast equation is

$$\begin{aligned}\hat{Z}_t(l) &= \mu + \phi(\hat{Z}_t(l-1) - \mu) \\ &= \mu + \phi^l(Z_t - \mu), \qquad l \geq 1.\end{aligned} \tag{5.7.1}$$

Thus,

$$\hat{Z}_{100}(1) = 9 + .6(8.9 - 9) = 8.94,$$
$$\hat{Z}_{100}(2) = 9 + (.6)^2(8.9 - 9) = 8.964,$$
$$\hat{Z}_{100}(3) = 9 + (.6)^3(8.9 - 9) = 8.9784,$$
$$\hat{Z}_{100}(4) = 9 + (.6)^4(8.9 - 9) = 8.98704.$$

2. To obtain the forecast limits, we obtain the ψ weights from the relationship

$$(1 - \phi B)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = 1. \tag{5.7.2}$$

That is,

$$\psi_j = \phi^j, \qquad j \geq 0. \tag{5.7.3}$$

The 95% forecast limits for Z_{101} from (5.2.10) are

$$8.94 \pm 1.96\sqrt{.1} \qquad \text{or} \qquad 8.320 < Z_{101} < 9.560.$$

The 95% forecast limits for Z_{102} are

$$8.964 \pm 1.96\sqrt{1 + (.6)^2}\sqrt{.1} \qquad \text{or} \qquad 8.241 < Z_{102} < 9.687.$$

The 95% forecast limits for Z_{103} and Z_{104} can be obtained similarly. The results are shown in Figure 5.2.

3. Suppose now that the observation at $t = 101$ turns out to be $Z_{101} = 8.8$. Because $\psi_l = \phi^l = (.6)^l$, we can update the forecasts for Z_{102}, Z_{103}, and Z_{104} by using (5.5.5) as follows.

$$\hat{Z}_{101}(1) = \hat{Z}_{100}(2) + \psi_1[Z_{101} - \hat{Z}_{100}(1)]$$
$$= 8.964 + .6(8.8 - 8.94) = 8.88$$
$$\hat{Z}_{101}(2) = \hat{Z}_{100}(3) + \psi_2[Z_{101} - \hat{Z}_{100}(1)]$$
$$= 8.9784 + (.6)^2(8.8 - 8.94) = 8.928$$
$$\hat{Z}_{101}(3) = \hat{Z}_{100}(4) + \psi_3[Z_{101} - \hat{Z}_{100}(1)]$$
$$= 8.98704 + (.6)^3(8.8 - 8.94) = 8.9568.$$

The earlier forecasts for Z_{102}, Z_{103}, and Z_{104} made at $t = 100$ are adjusted downward due to the negative forecast error made for Z_{101}.

Forecasting discussed above is based on the assumption that the parameters are known in the model. In practice, the parameters are, of course, unknown and have to be estimated from the given observations $\{Z_1, Z_2, \ldots, Z_n\}$. However, with respect to the forecasting origin $t = n$, the estimates are known constants and hence the results remain the same under this conditional sense. Estimation of the model parameters is discussed in Chapter 7.

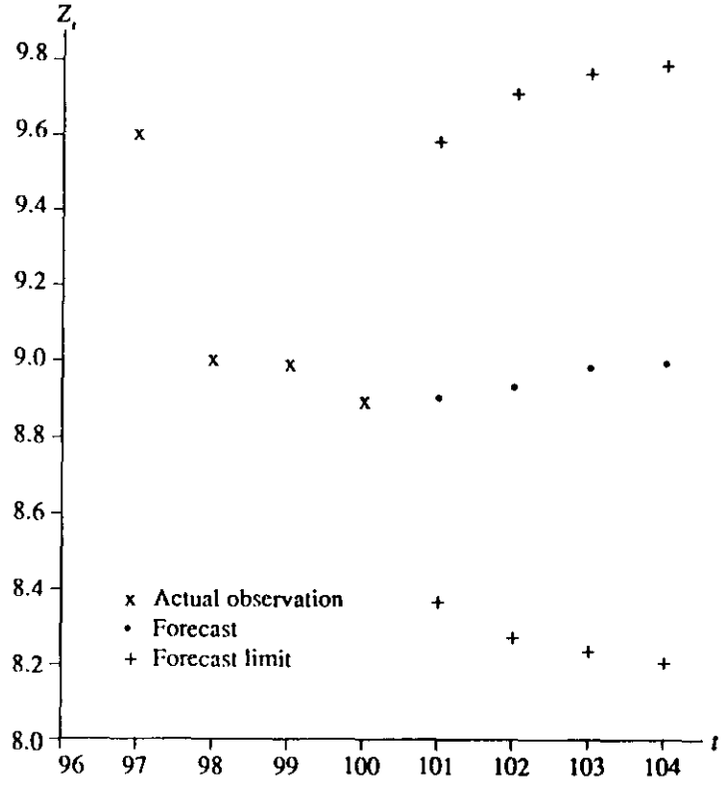


Fig. 5.2 Forecasts with 95% forecast limits for an AR(1) process.

Exercises

5.1 For each of the following models:
(i) $(1 - \phi_1 B)(Z_t - \mu) = a_t$,
(ii) $(1 - \phi_1 B - \phi_2 B^2)(Z_t - \mu) = a_t$,
(iii) $(1 - \phi_1 B)(1 - B)Z_t = (1 - \theta_1 B)a_t$.
(a) Find the l-step ahead forecast $\hat{Z}_n(l)$ of Z_{n+l}.
(b) Find the variance of the l-step ahead forecast error for $l = 1$, 2, and 3.

5.2 (a) Show that the covariance between forecast errors from different origins is given by

$$\text{Cov}\,[e_n(l), e_{n-j}(l)] = \sigma_a^2 \sum_{i=j}^{l-1} \psi_i \psi_{i-j}, \qquad l > j.$$

(b) Show that the covariance between forecast errors from the same origin but with different lead times is given by

$$\text{Cov}[e_n(l), e_n(l+j)] = \sigma_a^2 \sum_{i=0}^{l-1} \psi_i \psi_{i+j}.$$

5.3 Consider the model

$$(1-.68B)(1-B)^2 Z_t = (1-.75B+.34B^2)a_t.$$

(a) Compute and plot the correlations between the error of the forecast $\hat{Z}_t(5)$ and those of the forecasts $\hat{Z}_{t-j}(5)$ for $j = 1, 2, \ldots, 10$.

(b) Compute and plot the correlations between the error of the forecast $\hat{Z}_t(3)$ and those of $\hat{Z}_t(l)$ for $l = 1, 2, \ldots, 10$.

5.4 A sales series was fitted by the ARIMA(2, 1, 0) model

$$(1-1.4B+.7B^2)(1-B)Z_t = a_t$$

where $\hat{\sigma}_a^2 = 58000$ and the last 5 observations are 560, 580, 640, 770, and 800.

(a) Calculate the forecasts of the next 3 observations.

(b) Find the 95% forecast limits for the forecasts in (a).

(c) Find the eventual forecast function.

5.5 Consider the IMA(1, 1) model

$$(1-B)Z_t = (1-\theta B)a_t.$$

(a) Write down the forecast equation that generates the forecasts.

(b) Find the 95% forecast limits produced by this model.

(c) Express the forecasts as a weighted average of previous observations.

(d) Discuss the connection of this model with the simple exponential smoothing method.

5.6 (a) Show that (5.2.23) and (5.4.5) are equivalent.

(b) Illustrate the equivalence of (5.2.23) and (5.4.5) using the model in Exercise 5.5

5.7 Consider an AR(2) model $(1-\phi_1 B-\phi_2 B^2)(Z_t-\mu) = a_t$, where $\phi_1 = 1.2$, $\phi_2 = -.6$, $\mu = 65$, and $\sigma_a^2 = 1$. Suppose we have the observations $Z_{76} = 60.4$, $Z_{77} = 58.9$, $Z_{78} = 64.7$, $Z_{79} = 70.4$, and $Z_{80} = 62.6$.

(a) Forecast Z_{81}, Z_{82}, Z_{83}, and Z_{84}.

(b) Find the 95% forecast limits for the forecasts in (a).

(c) Suppose that the observation at $t = 81$ turns out to be $Z_{81} = 62.2$. Find the updated forecasts for Z_{82}, Z_{83}, and Z_{84}.

5.8 Consider the model

$$(1-.43B)(1-B)Z_t = a_t$$

and the observations $Z_{49} = 33.4$ and $Z_{50} = 33.9$.

(a) Compute the forecast $\hat{Z}_{50}(l)$, for $l = 1, 2, \ldots, 10$, and their 90% forecast limits.

(b) What is the eventual forecast function for the forecasts made at $t = 50$?

(c) At time $t = 51$, Z_{51} became known and equaled 34.1. Update the forecasts obtained in (a).

5.9 Consider the ARIMA(0, 1, 1) model

$$(1-B)Z_t = (1-.8B)a_t.$$

(a) Find the π weights for the following AR representation:

$$Z_t = \hat{Z}_{t-1} + a_t$$

where $\hat{Z}_t = \sum_{j=1}^{\infty} \pi_j Z_{t-j}$, and show that $\sum_{j=1}^{\infty} \pi_j = 1$.

(b) Let $\hat{Z}_t(2) = \sum_{j=1}^{\infty} \pi_j^{(2)} Z_{t-j+1}$ be the two-step ahead forecasts of Z_{t+2} at time t. Express $\pi_j^{(2)}$ in terms of the weights π_j.

5.10 Obtain the eventual forecast function for the following models:

(a) $(1-.6B)Z_t = (1-.8B+.3B^2)a_t$,

(b) $(1-.3B)(1-B) = .4 + a_t$,

(c) $(1-1.2B+.6B^2)(Z_t-65) = a_t$,

(d) $(1-B)^2 Z_t = (1-.2B-.3B^2)a_t$,

(e) $(1-.6B)(1-B)^2 Z_t = (1-.75B+.34B^2)a_t$.

6 MODEL IDENTIFICATION

In time series analysis, the most crucial steps are to identify and build a model based on the available data. This requires a good understanding of the processes discussed in Chapters 3 and 4, particularly the characteristics of these processes in terms of their ACF, ρ_k, and PACF, ϕ_{kk}. In practice, these ACF and PACF are unknown, and for a given observed time series $Z_1, Z_2, \ldots$, and Z_n, they have to be estimated by the sample ACF, $\hat{\rho}_k$, and PACF, $\hat{\phi}_{kk}$ discussed in Section 2.5. Thus, in model identification, our goal is to match patterns in the sample ACF, $\hat{\rho}_k$, and sample PACF, $\hat{\phi}_{kk}$, with the known patterns of the ACF, ρ_k, and PACF, ϕ_{kk}, for the ARMA models. For example, since we know that the ACF cuts off at lag 1 for an MA(1) model, a large single significant spike at lag 1 for $\hat{\rho}_k$ will indicate an MA(1) model as a possible underlying process.

After introducing systematic and useful steps for model identification, we give illustrative examples of identifying models for a wide variety of actual time series data. We also discuss some recently introduced identification tools such as the inverse autocorrelation and extended sample autocorrelation functions.

6.1 STEPS FOR MODEL IDENTIFICATION

To illustrate the model identification, we consider the general ARIMA(p, d, q) model

$$(1-\phi_1 B-\cdots-\phi_p B^p)(1-B)^d Z_t=\theta_0+(1-\theta_1 B-\cdots-\theta_q B^q)a_t. \qquad (6.1.1)$$

...del identification refers to the methodology in identifying the required t... ...ormations, such as variance stabilizing transformations and differencing t... ...ormations, the decision to include the deterministic parameter θ_0 when ... ...nd the proper orders of p and q for the model. Given a time series, we ...e... ...following useful steps to identify a tentative model.

Step 1. Plot the time series data and choose proper transformations.

In any time series analysis, the first step is to plot the data. Through careful examination of the plot, we usually get a good idea about whether the series contains a trend, seasonality, outliers, nonconstant variances, and other nonnormal and nonstationary phenomena. This understanding often provides a basis for postulating a possible data transformation.

In time series analysis, the most commonly used transformations are variance-stabilizing transformations and differencing. Since differencing may create some negative values, we should always apply variance-stabilizing transformations before taking differences. A series with nonconstant variance often needs a logarithmic transformation. More generally, to stabilize the variance, we can apply Box–Cox's power transformation discussed in Section 4.3.2. Since a variance-stabilizing transformation, if necessary, is often chosen before we do any further analysis, we refer to this transformed data as the original series in the following discussion unless mentioned otherwise.

Step 2. Compute and examine the sample ACF and the sample PACF of the original series to further confirm a necessary degree of differencing. Some general rules are:

1. If the sample ACF decays very slowly (the individual ACF may not be large) and the sample PACF cuts off after lag 1, it indicates that differencing is needed. Try taking the first differencing $(1-B)Z_t$. One can also use the test proposed by Dickey and Fuller (1979). In a borderline case, differencing is generally recommended (see Dickey, Bell, and Miller [1986]).
2. More generally, to remove nonstationarity we may need to consider a higher order differencing $(1-B)^d Z_t$ for $d>1$. In most cases, d is either 0, 1, or 2. Some authors argue that the consequences of unnecessary differencing are much less serious than those of underdifferencing. But do beware of the artifacts created by overdifferencing, so that unnecessary overparameterization can be avoided.

Step 3. Compute and examine the sample ACF and PACF of the properly transformed and differenced series to identify the orders of p and q (where we recall that p is the highest order in the autoregressive polynomial $(1-\phi_1 B-\cdots-\phi_p B^p)$, and q is the highest order in the moving average polynomial $(1-\theta_1 B-\cdots-\theta_q B^q)$. Usually, the needed orders of these p and q are less than or equal to 3. Table 6.1 summarizes the important results for selecting p and q.

It is useful and interesting to note that a strong duality exists between the AR and the MA models in terms of their ACFs and PACFs. To build a reasonable ARIMA model, ideally, we need a minimum of $n=50$ observations, and the number of sample ACF and PACF to be calculated should be about $n/4$, although occasionally for data of good quality one may be able to identify an adequate model with a smaller sample size.

We identify the orders p and q by matching the patterns in the sample ACF and PACF with the theoretical patterns of known models. The art of a

Table 6.1 Characteristics of theoretical ACF and PACF for stationary processes.

Process	ACF	PACF
AR(p)	Tails off as exponential decay or damped sine wave	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off as exponential decay or damped sine wave
ARMA(p, q)	Tails off after lag $(q-p)$	Tails off after lag $(p-q)$

time series analyst's model identification is very much like the method of an FBI agent's criminal search. Most criminals disguise themselves to avoid being recognized. This is also true of ACF and PACF. The sampling variation and the correlation among the sample ACF and PACF as shown in Section 2.5 often disguise the theoretical ACF and PACF patterns. Hence, in the initial model identification we always concentrate on the general broad features of these sample ACF and PACF without focusing on the fine details. Model improvement can be easily achieved at a later stage of diagnostic checking.

The estimated variances of both the sample ACF and PACF given in (2.5.21) and (2.5.27) are very rough approximations. Some authors recommend that a conservative threshold of 1.5 standard deviations be used in checking the significance of the short-term lags of these ACF and PACF at the initial model identification phase. This is especially true for a relatively short series.

Step 4. Test the deterministic trend term θ_0 when $d > 0$.

As discussed in Section 4.2, for a nonstationary model, $\phi(B)(1-B)^d Z_t = \theta_0 + \theta(B)a_t$, the parameter θ_0 is usually omitted so that it is capable of representing series with random changes in the level, slope, or trend. However, if there is reason to believe that the differenced series contains a deterministic trend mean, we can test for its inclusion by comparing the sample mean $\overline{W}$ of the differenced series $W_t = (1-B)^d Z_t$ with its approximate standard error $S_{\overline{W}}$. To derive $S_{\overline{W}}$, we note from Section 2.5.1 that $\lim_{n\to\infty} n\,\mathrm{Var}(\overline{W}) = \sum_{j=-\infty}^{\infty} \gamma_j$, and hence

$$\sigma_{\overline{W}}^2 = \frac{\gamma_0}{n}\sum_{j=-\infty}^{\infty}\rho_j = \frac{1}{n}\sum_{j=-\infty}^{\infty}\gamma_j = \frac{1}{n}\gamma(1), \tag{6.1.2}$$

where $\gamma(B)$ is the autocovariance generating function defined in (2.6.8) and $\gamma(1)$ is its value at $B = 1$. Thus, the variance and hence the standard error for $\overline{W}$ is model dependent. For example, for the ARIMA(1, d, 0) model, $(1-\phi B)W_t = a_t$, we have, from (2.6.9),

$$\gamma(B) = \frac{\sigma_a^2}{(1-\phi B)(1-\phi B^{-1})}$$

so that

$$\sigma_{\overline{W}}^2 = \frac{\sigma_a^2}{n}\frac{1}{(1-\phi)^2} = \frac{\sigma_W^2}{n}\frac{1-\phi^2}{(1-\phi)^2} = \frac{\sigma_W^2}{n}\left(\frac{1+\phi}{1-\phi}\right) = \frac{\sigma_W^2}{n}\left(\frac{1+\rho_1}{1-\rho_1}\right), \tag{6.1.3}$$

where we note that $\sigma_W^2 = \sigma_a^2/(1-\phi^2)$. The required standard error is

$$S_{\overline{W}} = \sqrt{\frac{\hat{\gamma}_0}{n}\left(\frac{1+\hat{\rho}_1}{1-\hat{\rho}_1}\right)}. \tag{6.1.4}$$

Expressions of $S_{\overline{W}}$ for other models can be derived similarly. However, at the model identification phase, since the underlying model is unknown, most available software use the approximation

$$S_{\overline{W}} = \left[\frac{\hat{\gamma}_0}{n}(1 + 2\hat{\rho}_1 + 2\hat{\rho}_2 + \cdots + 2\hat{\rho}_k)\right]^{1/2} \tag{6.1.5}$$

where $\hat{\gamma}_0$ is the sample variance and $\hat{\rho}_1, \ldots, \hat{\rho}_k$ are the first k significant sample ACFs of $\{W_t\}$. Under the null hypothesis $\rho_k = 0$ for $k \geq 1$, Equation (6.1.5) reduces to

$$S_{\overline{W}} = \sqrt{\hat{\gamma}_0/n}. \tag{6.1.6}$$

Alternatively, one can include θ_0 initially and discard it at the final model estimation if the preliminary estimation result is not significant. Parameter estimation is discussed in the next chapter.

6.2 EMPIRICAL EXAMPLES

In this section we present a variety of real-world examples to illustrate the method of model identification. Several mainframe computer programs such as BMDP, MINITAB, SAS, SCA, and SPSS, and PC software such as AUTOBOX and RATS are available to facilitate the methods. Some programs are available for both the mainframe and personal computers.

Example 6.1 Figure 6.1 shows Series W1, which is the daily average number of defects per truck found in the final inspection at the end of the assembly line of a truck manufacturing plant. The data consist of 45 daily observations of consecutive business days between November 4 to January 10, as reported in Bun (1976, p. 134). The figure suggests a stationary process with constant mean and variance. The sample ACF and sample PACF are calculated in Table 6.2 and plotted in Figure 6.2. The fact that the sample ACF decays exponentially and the sample PACF has a single spike at lag 1 indicates that the series is likely to be generated by an AR(1) process,

$$(1-\phi B)(Z_t - \mu) = a_t. \tag{6.2.1}$$

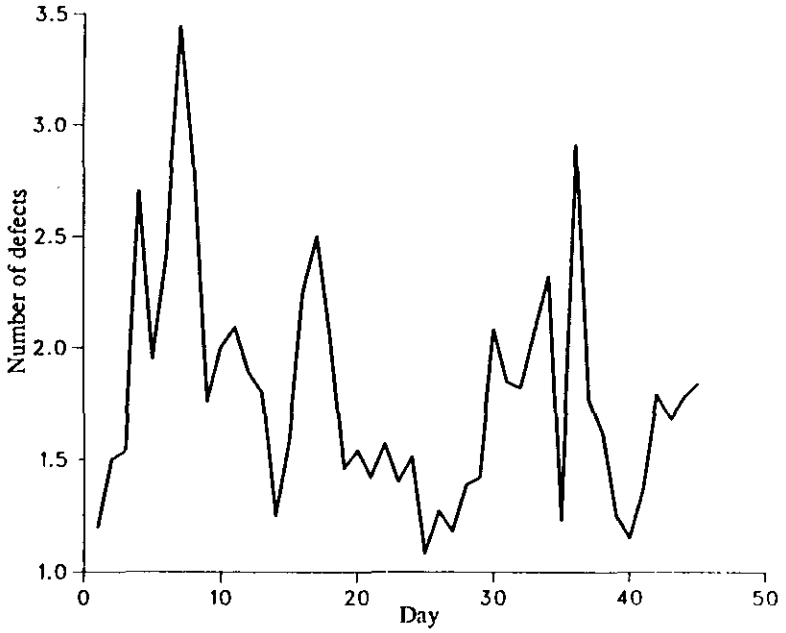


Fig. 6.1 Daily average number of truck manufacturing defects (Series W1).

Example 6.2 Series W2 is the classic series of the Wolf yearly sunspot numbers between 1700 and 1984, giving a total of $n = 285$ observations. Scientists believe that the sunspot numbers affect the weather of the earth and hence human activities such as agriculture, telecommunications, and others. The Wolf sunspot numbers have been discussed extensively in time series literature, e.g., Yule (1927), Bartlett (1950), Whittle (1954), Brillinger and Rosenblatt (1967), and others. This series is also known as the Wolfer sunspot numbers, named after Wolfer, who was a student of Wolf's. For an interesting account of the history of the series, see Izenman (1985). The series between 1700 and 1955 is

Table 6.2 Sample ACF and sample PACF for the daily average number of truck manufacturing defects (Series W1).

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.43	.26	.14	.08	−.09	−.07	−.21	−.11	−.05	−.01
St.E.	.15	.15	.17	.18	.19	.19	.19	.19	.19	.19
$\hat{\phi}_{kk}$	.43	.09	.00	.00	−.16	.00	−.18	.07	.05	.01
St.E.	.15	.15	.15	.15	.15	.15	.15	.15	.15	.15

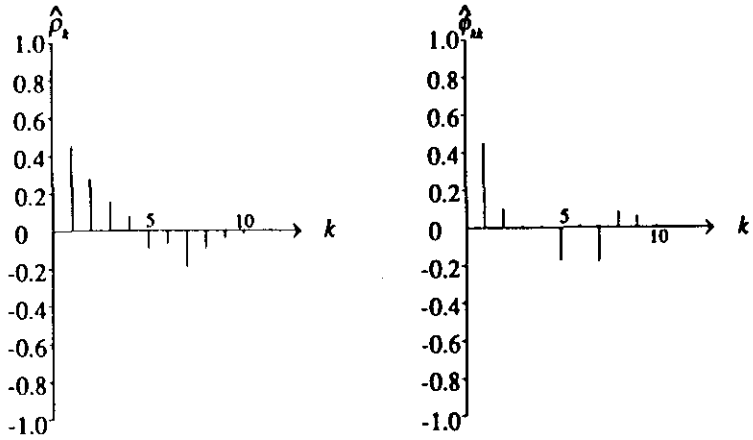


Fig. 6.2 Sample ACF and PACF of the daily average number of truck manufacturing defects (Series W1).

from Waldmeirer (1961), and the remaining observations are calculated from the monthly sunspot numbers in Andrews and Herzberg (1985).

The plot of the data is given in Figure 6.3. It indicates that the series is stationary in the mean. To investigate whether the series is also stationary in the variance, we calculated the following preliminary residual sum of squares:

$$S(\lambda) = \sum_{t=1}^{n} (Z_t(\lambda) - \hat{\mu})^2 \tag{6.2.2}$$

for various values of the power transformation parameter as shown in Table 6.3, where $\hat{\mu}$ is the corresponding sample mean of the transformed series. These calculations suggest that a square root transformation be applied to the data.

The sample ACF and sample PACF of the transformed data are computed as shown in Table 6.4 and Figure 6.4. The sample ACF shows a damping sine-cosine wave, and the sample PACF has relatively large spikes at lags 1, 2, and 9, suggesting that a tentative model may be an AR(2).

$$(1 - \phi_1 B - \phi_2 B^2)(\sqrt{Z_t} - \mu) = a_t, \tag{6.2.3}$$

or an AR(9)

$$(1 - \phi_1 B - \cdots - \phi_9 B^9)(\sqrt{Z_t} - \mu) = a_t. \tag{6.2.4}$$

By ignoring the values of $\hat{\phi}_{kk}$ beyond lag 3, Box and Jenkins (1976) suggest that an AR(3) model, $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(\sqrt{Z_t} - \mu) = a_t$, is also possible, even though their analysis was based on the nontransformed data between 1770

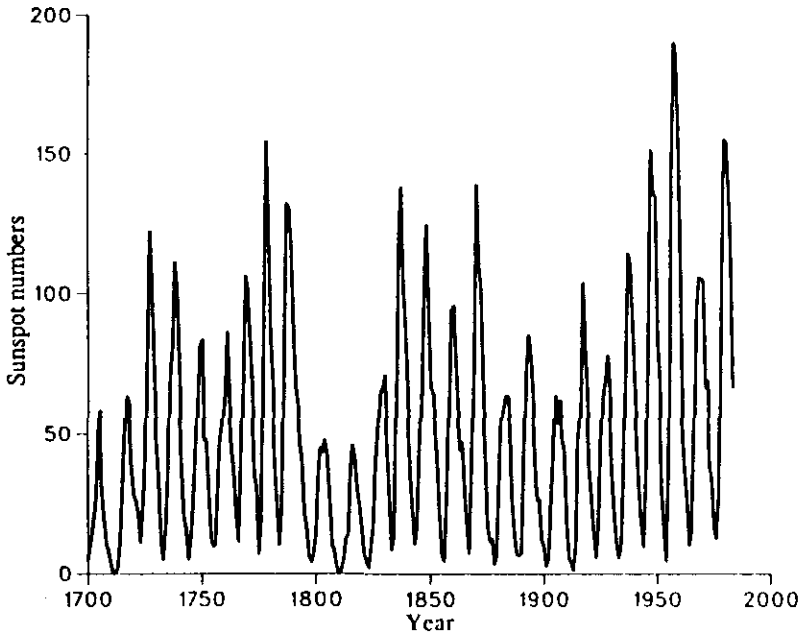


Fig. 6.3 Wolf yearly sunspot numbers, 1700–1983 (Series W2).

Table 6.3 Residual sum of squares in the power transformation.

λ	Residual sum of squares
1.0	13.82
0.5	9.77
0.0	11.06
−0.5	25.86
−1.0	147.68

and 1869. Because of the large autocorrelation .65 at lag 11, many scientists believe that the series contains a cycle of eleven years. We examine this series carefully in later chapters.

Example 6.3 Series W3 is a laboratory series of blowfly data taken from Nicholson (1950). A fixed number of adult blowflies with balanced sex ratios were put inside a cage and given a fixed amount of food daily. The blowfly population was then enumerated every other day for approximately two years,

Table 6.4 Sample ACF and sample PACF for the square root transformed sunspot numbers (Series W2).

k					$\hat{\rho}_k$					
1–10	.81	.45	.06	−.25	−.41	−.40	−.21	.08	.40	.61
St.E.	.06	.09	.10	.10	.10	.11	.11	.11	.11	.11
11–20	.65	.50	.22	−.06	−.28	−.37	−.34	−.21	−.00	.20
St.E.	.12	.14	.15	.15	.15	.15	.15	.16	.16	.16
21–30	.35	.37	.26	.05	−.15	−.30	−.37	−.32	−.20	−.03
St.E.	.16	.16	.16	.17	.17	.17	.17	.17	.17	.17
k					$\hat{\phi}_{kk}$					
1–10	.81	−.62	−.16	−.02	−.05	.13	.24	.16	.28	.03
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06
11–20	−.01	−.05	−.06	.09	−.01	−.08	−.09	−.11	−.01	−.05
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06
21–30	.05	−.03	−.10	−.08	.02	−.03	.01	.05	−.05	.01
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

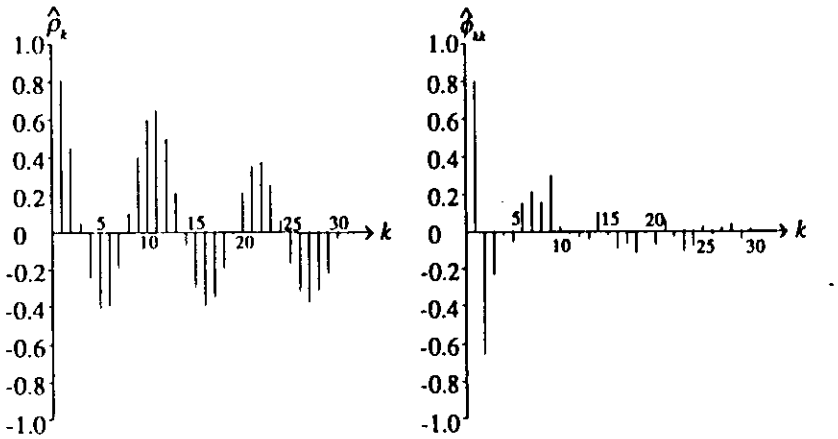


Fig. 6.4 Sample ACF and sample PACF for the square root transformed sunspot numbers (Series W2).

giving a total of $n = 364$ observations. Brillinger, Guckenheimer, Guttorp, and Oster (1980) first applied time series analysis on the data set. Later Tong (1983) considered the following two subseries:

Blowfly A: for Z_t between $20 \le t \le 145$,

Blowfly B: for Z_t between $218 \le t \le 299$,

and argued that the series Blowfly A is nonlinear. Series W3 used in our analysis is the series Blowfly B of 82 observations as shown in Figure 6.5.

The data plot suggests that the series is stationary in the mean. However, the power transformation analysis indicates that a square root or a logarithmic transformation is needed as shown in Table 6.5.

The sample ACF and PACF are calculated for the square root transformed data as shown in Table 6.6 and Figure 6.6. The sample $\hat{\rho}_k$ tails off exponentially and $\hat{\phi}_{kk}$ cuts off after lag 1. Thus, the following AR(1) model is entertained:

$$(1 - \phi B)(\sqrt{Z_t} - \mu) = a_t \qquad (6.2.5)$$

Example 6.4 Recall the monthly series of 300 unemployed females between ages 16 and 19 in the United States from January 1961 to December 1985 as shown in Figure 4.1. The series is labeled as Series W4. It is clearly nonsta-

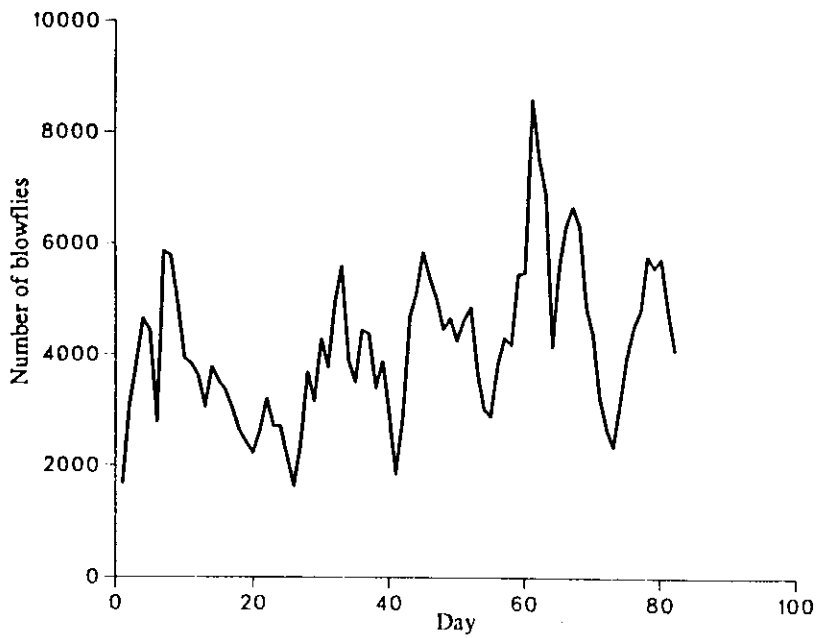


Fig. 6.5 Blowfly data (Series W3).

Table 6.5 Results of the power transformation on blowfly data.

λ	Residual sum of squares
1.0	55.31
0.5	50.09
0.0	50.38
−0.5	56.67
−1.0	71.50

tionary in the mean, suggesting differencing is needed. This is further confirmed by the sustained large spikes of the sample ACF shown in Table 6.7 and Figure 6.7. The differenced series is now stationary and is shown in Figure 6.8. The sample ACF and sample PACF are computed for this differenced series in Table 6.8 and also plotted in Figure 6.9. The sample ACF now cuts off af-

Table 6.6 Sample ACF and sample PACF for the blowfly data (Series W3).

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.73	.49	.30	.20	.12	.02	−.01	−.04	−.01	−.03
St.E.	.11	.16	.18	.18	.19	.19	.19	.19	.19	.19
$\hat{\phi}_{kk}$	.73	−.09	−.04	.04	−.03	−.12	.07	−.05	.07	−.08
St.E.	.11	.11	.11	.11	.11	.11	.11	.11	.11	.11

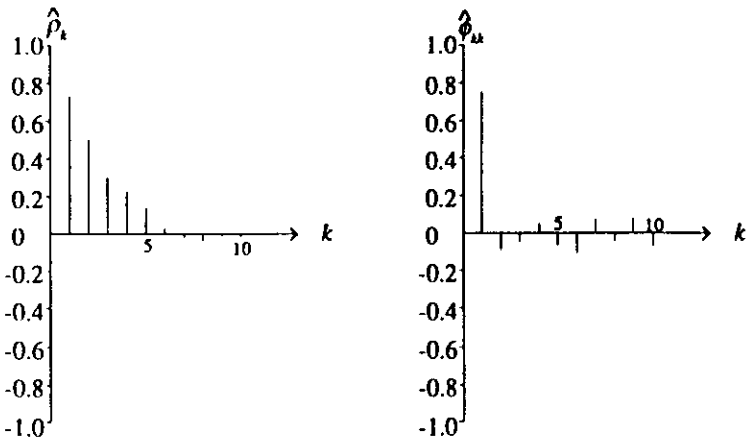


Fig. 6.6 Sample ACF and sample PACF for the blowfly data (Series W3).

Table 6.7 Sample ACF and sample PACF of Series W4 — the U.S. monthly series of unemployed females between ages 16 and 19 from January 1961 to December 1985.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.97	.96	.95	.94	.93	.93	.92	.91	.90	.90
St.E.	.06	.10	.13	.15	.17	.18	.20	.21	.22	.24
$\hat{\phi}_{kk}$	.97	.37	.10	.08	.02	.08	.02	.02	−.04	.00
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

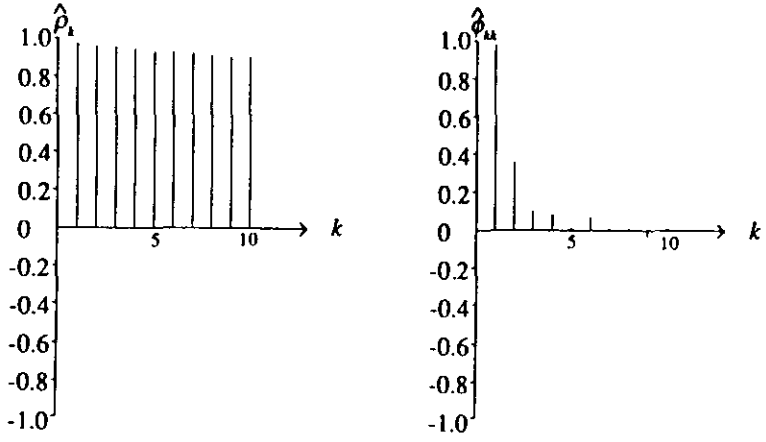


Fig. 6.7 Sample ACF and sample PACF of Series W4 — the U.S. monthly series of unemployed females between ages 16 and 19 from January 1961 to December 1985.

ter lag 1 while the sample PACF tails off. This pattern is very similar to the ACF and PACF for MA(1) with positive θ_1 as shown in Figure 3.10, suggesting an MA(1) model for the differenced series and hence an IMA(1, 1) model for the original series. To test whether a deterministic trend parameter θ_0 is needed, we calculate the t-ratio $\overline{W}/S_{\overline{W}} = 1.0502/2.4223 = .4335$, which is not significant. Thus, our proposed model is

$$(1-B)Z_t = (1-\theta_1 B)a_t. \tag{6.2.6}$$

ample 6.5 The accidental death rate is a vital statistic for many state l federal governments. Figure 6.10 shows Series W5, which is the yearly lental death rate (per 100,000 population) of Pennsylvania between 1950 1984 published in the 1984 Pennsylvania Vital Statistics Annual Report by t nsylvania State Health Data Center. The series is clearly nonstation-

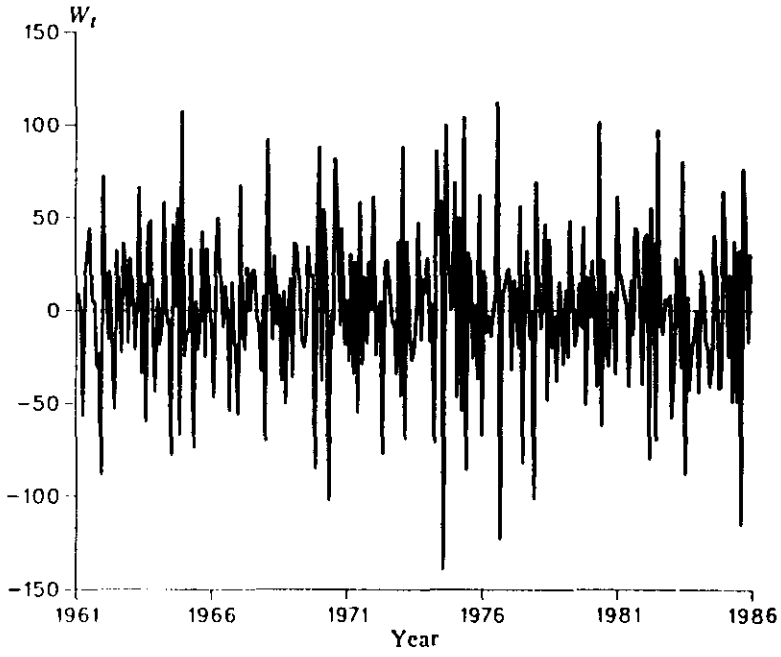


Fig. 6.8 The differenced monthly series, $W_t = (1-B)Z_t$, of the U.S. monthly series of unemployed females between ages 16 and 19 from January 1961 to December 1985 (Series W4).

ary with a decreasing trend. This nonstationarity is also shown by the slowly decaying ACF in Table 6.9 and Figure 6.11. Both Figure 6.11 and the evaluation of power transformations indicate no transformations other than differencing are needed. The sample ACF and PACF of the differenced data shown in Table 6.10 and Figure 6.12 suggest a white noise phenomenon. The t-ratio, $\overline{W}/S_{\overline{W}} = -.5618/.2507 = -2.24$, implies that a deterministic trend term is

Table 6.8 Sample ACF and sample PACF of the differenced U.S. monthly Series $W_t = (1-B)Z_t$, of unemployed females between ages 16 and 19 (Series W4) $\overline{W} = 1.0502$, $S_{\overline{W}} = 2.4223$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.41	.06	−.08	.06	−.09	.07	−.03	.07	−.06	−.08
St.E.	.06	.07	.07	.07	.07	.07	.07	.07	.07	.07
$\hat{\phi}_{kk}$	−.41	−.14	.14	−.04	−.11	−.02	−.02	.05	−.01	.06
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

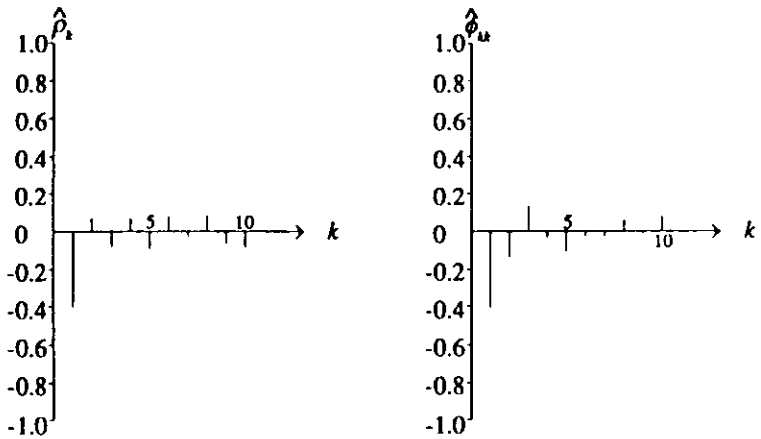
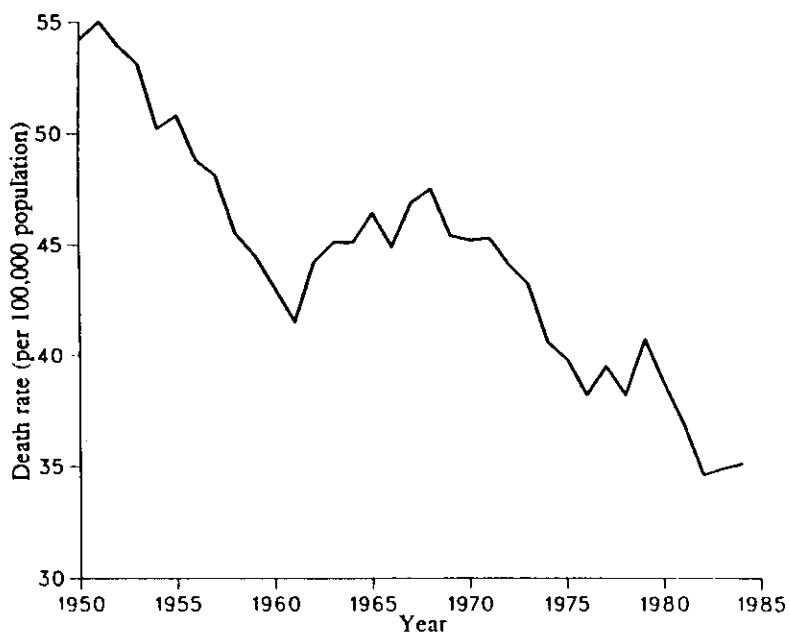


Fig. 6.9 Sample ACF and sample PACF of the differenced U.S. monthly series, $W_t = (1-B)Z_t$, of unemployed females between the ages of 16 and 19 (Series W4).



Fi[illegible]10 Series W5—The yearly Pennsylvania accidental death rate between 1950 [illegible]1.

Table 6.9 Sample ACF and sample PACF of the Pennsylvania accidental death rate between 1950 and 1984 (Series W5).

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.87	.73	.57	.44	.34	.25	.15	.08	.01	−.02
St.E.	.17	.27	.32	.35	.36	.37	.38	.38	.38	.38
$\hat{\phi}_{kk}$	.87	−.13	−.15	.03	.06	−.08	−.14	.04	−.02	.03
St.E.	.17	.17	.17	.17	.17	.17	.17	.17	.17	.17

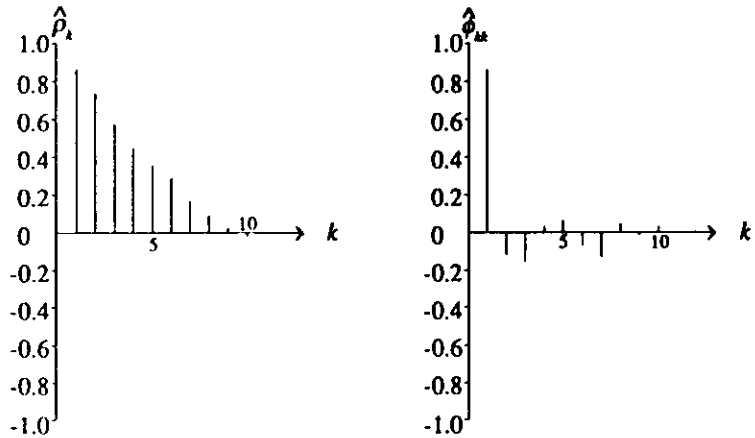


Fig. 6.11 Sample ACF and sample PACF of Series W5—the yearly Pennsylvania accidental death rate between 1950 and 1984.

recommended. Hence, the following random walk model with drift is entertained:

$$(1-B)Z_t = \theta_0 + a_t. \tag{6.2.7}$$

Based on the ACF and PACF of the original nondifferenced data in Table 6.9, one may also suggest an alternative AR(1) model

$$(1-\phi_1 B)(Z_t - \mu) = a_t. \tag{6.2.8}$$

However, the clear downward trend should give an estimate of ϕ_1 close to 1. We investigate both models in Chapter 7 when we discuss parameter estimation.

Example 6.6 We now examine Series W6, which is the yearly U.S. tobacco production from 1871 to 1984 published in the 1985 Agricultural Statistics by the United States Department of Revenue and shown in Figure 4.2. The plot indicates that the series is nonstationary both in the mean and the variance. In fact, the standard deviation of the series is roughly proportional to the level

Table 6.10 Sample ACF and sample PACF for the differenced series of the Pennsylvania accidental death rate from 1950 to 1984 (Series W5) $\overline{W} = -.5618$, $S_{\overline{W}} = .2507$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.10	.27	−.04	.09	.02	.05	−.08	.01	−.05	−.12
St.E.	.17	.17	.18	.18	.18	.18	.18	.19	.19	.19
$\hat{\phi}_{kk}$	−.10	.27	.01	.02	.05	.03	−.10	−.02	−.01	−.14
St.E.	.17	.17	.17	.17	.17	.17	.17	.17	.17	.17

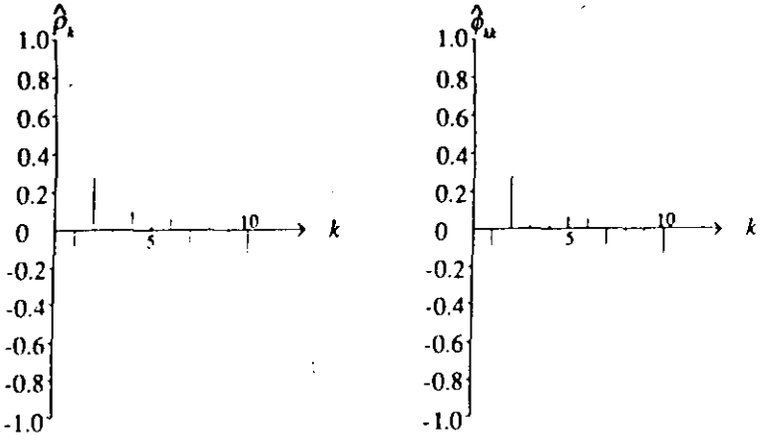


Fig. 6.12 Sample ACF and sample PACF for the differenced series of the Pennsylvania accidental death rate (Series W5).

of the series. Hence, from the results of Section 4.3.2, a logarithmic transformation is suggested, which is also confirmed by the λ value of the power transformation calculated in Table 6.11. These transformed data are plotted in Figure 6.13 and show an upward trend with a constant variance.

The very slowly decaying ACF as shown in Table 6.12 and Figure 6.14 further supports the need for differencing. Hence, the sample ACF and PACF for the differenced data, $W_t = (1-B)\ln Z_t$, are calculated in Table 6.13 with their plots in Figure 6.15. The ACF cuts off after lag 1, and the PACF tails off exponentially, which looks very similar to Figure 3.10 with $\theta_1 > 0$. It suggests that IMA(1, 1) is a possible model. To determine whether a deterministic trend

Table 6.11 Result of the power transformation on the tobacco production data.

λ	Residual sum of squares
1.0	7.88
0.5	5.95
0.0	5.11
−0.5	5.55
−1.0	7.92

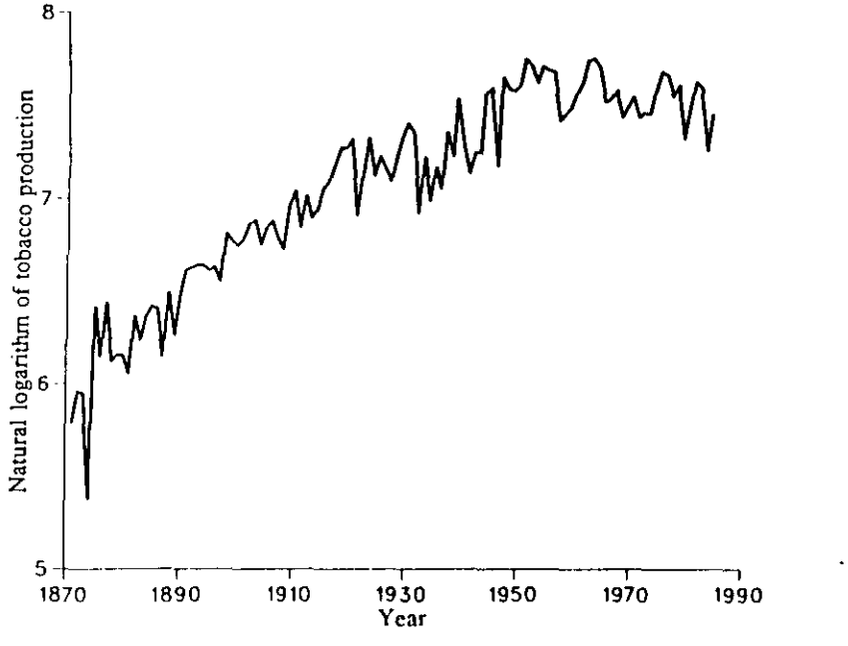


Fig. 6.13 Natural logarithms of the U.S. yearly tobacco production in million pounds (Series W6).

term θ_0 is needed, we examine the t-ratio, $t = \overline{W}/S_{\overline{W}} = .0147/.0186 = .7903$, which is not significant. Hence, we entertain the following IMA(1, 1) model as our tentative model:

$$(1-B)\ln Z_t = (1-\theta_1 B)a_t. \tag{6.2.9}$$

Table 6.12 Sample ACF and sample PACF for natural logarithms of the U.S. yearly tobacco production (Series W6).

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.90	.88	.84	.79	.78	.76	.75	.72	.69	.66
St.E.	.15	.19	.22	.24	.27	.28	.30	.32	.33	.34
$\hat{\phi}_{kk}$	.90	.37	.05	−.11	.15	.14	.08	−.11	−.12	.00
St.E.	.09	.09	.09	.09	.09	.09	.09	.09	.09	.09

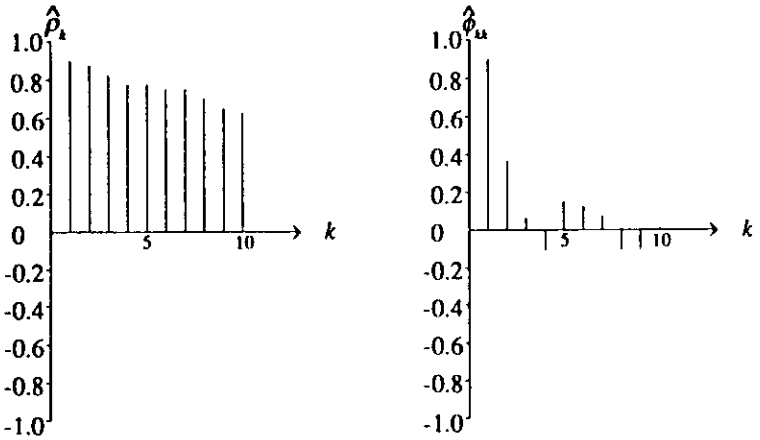


Fig. 6.14 Sample ACF and sample PACF for natural logarithms of the U.S. yearly tobacco production (Series W6).

Table 6.13 Sample ACF and sample PACF for the differenced series of natural logarithms of the U.S. yearly tobacco production (Series W6).

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.51	.11	−.09	.02	−.03	.00	.04	.04	−.05	−.01
St.E.	.12	.12	.12	.12	.12	.12	.12	.12	.12	.12
$\hat{\phi}_{kk}$	−.51	−.20	−.17	−.14	−.13	−.12	−.04	.06	.02	−.03
St.E.	.09	.09	.09	.09	.09	.09	.09	.09	.09	.09

$\overline{W} = .0147$, $S_{\overline{W}} = .0186$

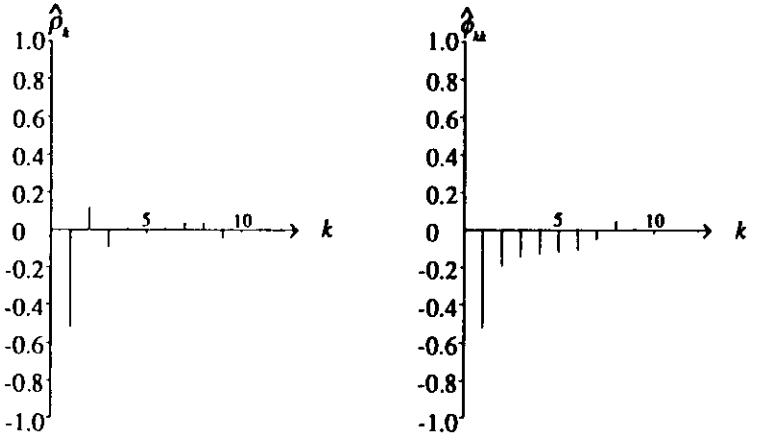


Fig. 6.15 Sample ACF and sample PACF for the differenced natural logarithms of the tobacco data (Series W6).

Example 6.7 Figure 6.16(a) shows Series W7—the yearly number of lynx pelts sold by the Hudson's Bay Company in Canada between 1857 and 1911 as reported in Andrews and Herzberg (1985). The result of the power transformation in Table 6.14 indicates that a logarithmic transformation is required. The natural logarithm of the series is stationary and is plotted in Figure 6.16(b).

The sample ACF in Table 6.15 and Figure 6.17 show a clear sine-cosine phenomenon indicating an AR(p) model with $p \geq 2$. The three significant PACF strongly suggest $p = 3$. Thus, our entertained model is

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(\ln Z_t - \mu) = a_t. \tag{6.2.10}$$

A related series that was studied by many time series analysts is the number of Canadian lynx trapped for the years from 1821 to 1934. References include Campbell and Walker (1977), Tong (1977), Priestly (1981, Section 5.5), and Lin (1987). The series of lynx pelt sales that we analyze here is much shorter and has received much less attention in the literature. We use this series extensively for various illustrations in this book.

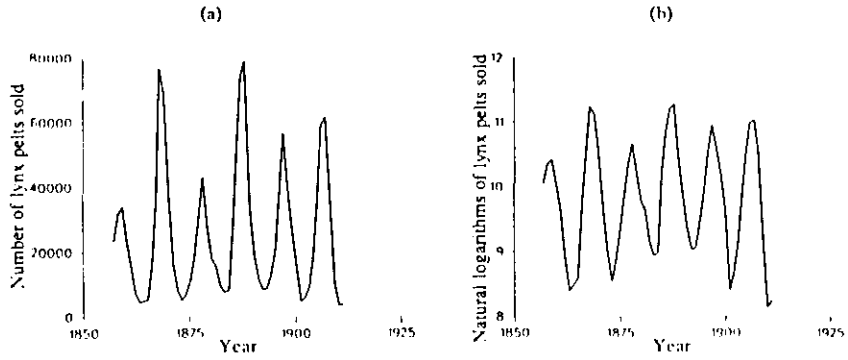


Fig. 6.16 The Canadian lynx pelt sales data (Series W7). (a) The yearly number of lynx pelts sold in Canada between 1857 and 1911. (b) The natural logarithms of yearly numbers of lynx pelts sold in Canada between 1857 and 1911.

Table 6.14 Result of the power transformation on the lynx pelt sales.

λ	Residual sum of squares
1.0	107.1642
0.5	61.1585
0.0	51.6603
−0.5	68.1996
−1.0	129.2240

Table 6.15 Sample ACF and PACF for natural logarithms of the yearly number of lynx pelts sold (Series W7).

	(a) ACF, $\hat{\rho}_k$									
1–10	.73	.22	−.32	−.69	−.76	−.53	−.08	.35	.61	.59
St.E.	.13	.19	.20	.21	.25	.29	.30	.30	.31	.33
11–20	.31	−.06	−.41	−.58	−.49	−.21	.16	.44	.54	.40
St.E.	.35	.36	.37	.37	.39	.40	.40	.41	.41	.43
	(b) PACF, $\hat{\phi}_{kk}$									
1–10	.73	−.68	−.36	−.20	−.09	−.08	.13	−.08	.06	−.07
St.E.	.13	.13	.13	.13	.13	.13	.13	.13	.13	.13
11–20	−.13	.05	−.19	.07	−.02	−.04	.14	−.04	.10	−.09
St.E.	.13	.13	.13	.13	.13	.13	.13	.13	.13	.13

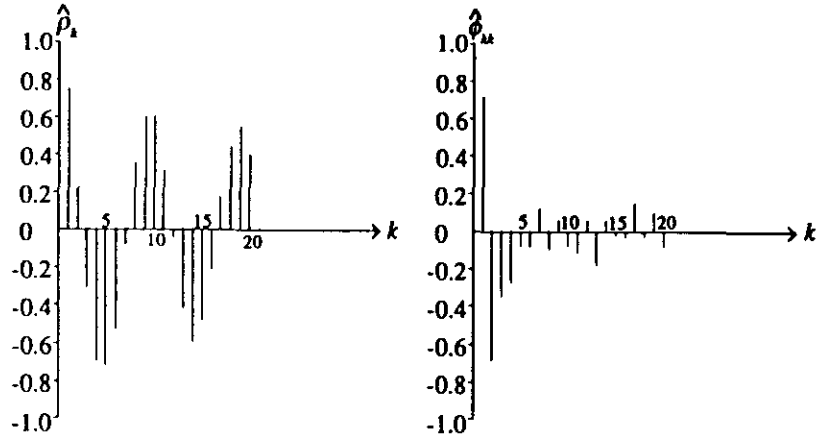


Fig. 6.17 Sample ACF and PACF for natural logarithms of the yearly number of lynx pelts sold (Series W7).

6.3 INVERSE AUTOCORRELATION FUNCTION (IACF)

Let

$$\phi_p(B)(Z_t - \mu) = \theta_q(B)a_t \tag{6.3.1}$$

be an ARMA(p, q) model where $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ is a stationary autoregressive operator, $\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$ is an invertible moving average operator, and the a_t is a white noise series with a zero mean and a constant variance σ_a^2. We can rewrite Equation (6.3.1) as the moving average representation

$$(Z_t - \mu) = \frac{\theta_q(B)}{\phi_p(B)}a_t = \psi(B)a_t, \tag{6.3.2}$$

where $\psi(B) = \theta_q(B)/\phi_p(B)$. From (2.6.9), the autocovariance generating function of this model is given by

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k = \sigma_a^2 \psi(B)\psi(B^{-1})$$
$$= \sigma_a^2 \frac{\theta_q(B)\theta_q(B^{-1})}{\phi_p(B)\phi_p(B^{-1})}. \tag{6.3.3}$$

Assume that $|\gamma(B)| > 0$ for all $|B| \leq 1$ and let

$$\gamma^{(I)}(B) = \frac{1}{\gamma(B)} = \sum_{k=-\infty}^{\infty} \gamma_k^{(I)} B^k$$
$$= \frac{1}{\sigma_a^2} \frac{\phi_p(B)\phi_p(B^{-1})}{\theta_q(B)\theta_q(B^{-1})}. \tag{6.3.4}$$

Clearly, from the assumptions on $\phi_p(B)$ and $\theta_q(B)$, the coefficients $\{\gamma_k^{(I)}\}$ in (6.3.4) form a positive definite, square summable sequence that is symmetric about the origin. Hence, they are proper autocovariances for a process. With respect to the model given in (6.3.1), the process having $\gamma^{(I)}(B)$ as its autocovariance generating function is referred to as the inverse process. Hence, $\gamma^{(I)}(B)$ is also referred to as the inverse autocovariance generating function of $\{Z_t\}$. Naturally, the inverse autocorrelation generating function of $\{Z_t\}$, from (2.6.10), is given by

$$\rho^{(I)}(B) = \frac{\gamma^{(I)}(B)}{\gamma_0^{(I)}}. \tag{6.3.5}$$

The kth lag inverse autocorrelation is defined as

$$\rho_k^{(I)} = \frac{\gamma_k^{(I)}}{\gamma_0^{(I)}}, \tag{6.3.6}$$

which, of course, is equal to the coefficient of B^k or B^{-k} in $\rho^{(I)}(B)$. As a function of k, $\rho_k^{(I)}$ is called the inverse autocorrelation function (IACF).

From Equations (6.3.3) and (6.3.4), it is clear that if $\{Z_t\}$ is an ARMA (p, q) process, then the inverse process will be an ARMA(q, p) process. Specifically, if $\{Z_t\}$ is an AR(p) process with autocorrelations tailing off, then the inverse process will be an MA(p) process with its autocorrelations cutting off at lag p. In other words, the inverse autocorrelation function of an AR(p) process will cut off at lag p. Similarly, an MA(q) process with autocorrelations cutting off at lag q will have its inverse autocorrelations tailing off. Hence, the inverse autocorrelation function of a process exhibits characteristics similar to the partial autocorrelation function and can be used as an identification tool in model identification.

For an AR(p) process, it is easily seen that the inverse autocorrelation function is given by

$$\rho_k^{(I)} = \begin{cases} \dfrac{-\phi_k + \phi_1\phi_{k+1} + \cdots + \phi_{p-k}\phi_p}{1 + \phi_1^2 + \cdots + \phi_p^2}, & k = 1, \ldots, p, \\ 0, & k > p. \end{cases} \tag{6.3.7}$$

Since one can approximate a series by an AR(p) model with a sufficiently large p, one method to obtain a sample inverse autocorrelation $\hat{\rho}_k^{(I)}$ is to replace the AR parameters by their estimates, i.e.,

$$\hat{\rho}_k^{(I)} = \begin{cases} \dfrac{-\hat{\phi}_k + \hat{\phi}_1\hat{\phi}_{k+1} + \cdots + \hat{\phi}_{p-k}\hat{\phi}_p}{1 + \hat{\phi}_1^2 + \cdots + \hat{\phi}_p^2}, & k = 1, \ldots, p, \\ 0, & k > p. \end{cases} \tag{6.3.8}$$

The parameter estimation is discussed in Chapter 7. Under the null hypothesis of a white noise process, the standard error of $\hat{\rho}_k^{(I)}$ is given by

$$S_{\hat{\rho}_k^{(I)}} = \sqrt{\frac{1}{n}}. \tag{6.3.9}$$

Thus, one can use the limits $\pm 2/\sqrt{n}$ to assess the significance of the sample inverse autocorrelations.

Example 6.8 As illustrations, Table 6.16 shows sample inverse autocorrelations for two time series examined in Section 6.2. Table 6.16(a) is the sample inverse autocorrelation for the daily average series of truck manufacturing defects that we discussed in Example 6.1 in which we entertained an AR(1) model for the series. Note that although the inverse autocorrelation function seems to cut off at lag 1, it is not statistically significant at $\alpha = .05$. Table 6.16(b) is the sample inverse autocorrelation for the natural logarithms of the Canadian lynx pelt sales that was identified as an AR(3) model based on the three significant PACF examined in Example 6.7. However, in terms of the IACF, the maximum AR order will be 2. In fact, because the standard error of the sample IACF is .14, the model implied by the IACF may be an AR(1), which is clearly not appropriate from the ACF in Table 6.15(a).

It is not unusual that the values of the IACF are smaller than those of PACF at lower lags in the above examples. If the underlying model is an AR(p), we

Table 6.16 Sample inverse autocorrelation functions (SIACF).

(a) The daily average number of truck manufacturing defects (Series W1)

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k^{(I)}$	−.27	.05	−.03	−.07	.10	−.11	.22	.01	−.05	.02
St.E.	.15	.15	.15	.15	.15	.15	.15	.15	.15	.15

(b) The natural logarithms of Canadian lynx pelt sales (Series W7)

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k^{(I)}$	−.67	.23	.01	−.03	.09	−.11	.15	−.11	.07	−.13
St.E.	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14

know that $\phi_{kk} = \rho_k^{(I)} = 0$ for $k > p$. But, Equation (6.3.7) implies that

$$\rho_p^{(I)} = \frac{-\phi_p}{1 + \phi_1^2 + \cdots + \phi_p^2}, \tag{6.3.10}$$

and from the discussion in the closing paragraph of Section 2.3, we have

$$\phi_{pp} = \phi_p. \tag{6.3.11}$$

Hence, $|\phi_{pp}| > |\rho_p^{(I)}|$, and sample IACF in general are smaller than sample PACF, particularly at lower lags. In a recent study, Abraham and Ledolter (1984) conclude that, as an identification aid, the PACF generally outperforms the IACF. Some computer programs such as SAS and SCA provide both PACF and IACF options for analysis.

The inverse autocorrelation function was first introduced by Cleveland (1972) through the inverse of a spectrum and the relationship between the spectrum and the autocorrelation function. This leads to another method to obtain a sample inverse autocorrelation function. We return to this point in Chapter 11.

6.4 EXTENDED SAMPLE AUTOCORRELATION FUNCTION AND OTHER IDENTIFICATION PROCEDURES

6.4.1 Extended Sample Autocorrelation Function (ESACF)

From the previous empirical examples, it seems clear that, due to the cutting off property of the PACF and IACF for AR models and the same cutting off property of the ACF for MA models, the identification of the order p of an AR model and the order q of an MA model through the sample ACF, PACF, and IACF are relatively simple. However, for a mixed ARMA process, the ACF, PACF, and IACF all exhibit tapering off behavior, which makes the identification of the orders p and q much more difficult. One commonly used method is based on the fact that if Z_t follows an ARMA(p, q) model

$$(1 - \phi_1 B - \cdots - \phi_p B^p)Z_t = \theta_0 + (1 - \theta_1 B - \cdots - \theta_q B^q)a_t, \tag{6.4.1a}$$

or equivalently

$$Z_t = \theta_0 + \sum_{i=1}^{p} \phi_i Z_{t-1} - \sum_{i=1}^{q} \theta_i a_{t-1} + a_t, \tag{6.4.1b}$$

then

$$\begin{aligned} Y_t &= (1 - \phi_1 B - \cdots - \phi_p B^p)Z_t \\ &= Z_t - \sum_{i=1}^{p} \phi_i Z_{t-i} \end{aligned} \tag{6.4.2}$$

follows an MA(q) model

$$Y_t = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t \tag{6.4.3}$$

where without loss of generality we assume $\theta_0 = 0$. Thus, some authors such as Tiao and Box (1981) suggest using the sample ACF of the estimated residuals

$$\hat{Y}_t = (1 - \hat{\phi}_1 B - \cdots - \hat{\phi}_p B^p)Z_t \tag{6.4.4}$$

from an ordinary least squares (OLS) AR fitting to identify q and hence obtaining the orders p and q for the ARMA(p, q) model. For example, an MA(2) residual process from an AR(1) fitting implies a mixed ARMA(1, 2) model. But, as we show in Section 7.4, because the OLS estimates of the AR parameter ϕ_is in (6.4.4) are not consistent when the underlying model is a mixed ARMA(p, q) with $q > 0$, the procedure may lead to incorrect identification.

To derive consistent estimates of ϕ_i, suppose that n observations adjusted for mean are available from the ARMA(p, q) process in (6.4.1a). If an AR(p) is fitted to the data, i.e.,

$$Z_t = \sum_{i=1}^{p} \phi_i Z_{t-i} + e_t, \qquad t = p + 1, \ldots, n \tag{6.4.5}$$

where e_t represents the error term, then the OLS estimates $\hat{\phi}_i^{(0)}$ of ϕ_i, $i = 1, \ldots, p$, will be inconsistent and the estimated residuals

$$\hat{e}_t^{(0)} = Z_t - \sum_{i=1}^{p} \hat{\phi}_i^{(0)} Z_{t-i}$$

will not be white noise. In fact, if $q \geq 1$, the lagged values $\hat{e}_{t-i}^{(0)}$, $i = 1, \ldots, q$, will contain some information about the process Z_t. This leads to the following iterated regressions. First, we consider an AR(p) regression plus an added term $\hat{e}_{t-1}^{(0)}$, i.e.,

$$Z_t = \sum_{i=1}^{p} \phi_i^{(1)} Z_{t-i} + \beta_1^{(1)} \hat{e}_{t-1}^{(0)} + e_t^{(1)}, \qquad t = p + 2, \ldots, n \tag{6.4.6}$$

where the superscript (1) refers to the first iterated regression and $e_t^{(1)}$ represents the corresponding error term. The OLS estimates $\hat{\phi}_i^{(1)}$ will be consistent if $q = 1$. However, if $q > 1$, the $\hat{\phi}_i^{(1)}$ will again be inconsistent, the estimated

residuals $\hat{e}_t^{(1)}$ are not white noise, and the lagged values $\hat{e}_{t-i}^{(1)}$ will contain some information about Z_t. We thus consider the second iterated AR(p) regression

$$Z_t = \sum_{i=1}^{p} \phi_i^{(2)} Z_{t-i} + \beta_1^{(2)} \hat{e}_{t-1}^{(1)} + \beta_2^{(2)} \hat{e}_{t-2}^{(0)} + e_t^{(2)}, \qquad t = p+3, \ldots, n. \quad (6.4.7)$$

The OLS estimates $\hat{\phi}_i^{(2)}$ will be consistent if $q = 2$. For $q > 2$, the $\hat{\phi}_i^{(2)}$ will again be inconsistent. However, consistent estimates can be obtained by repeating the above iteration. That is, the OLS estimates $\hat{\phi}_i^{(q)}$ obtained from the following qth iterated AR(p) regression will be consistent:

$$Z_t = \sum_{i=1}^{p} \phi_i^{(q)} Z_{t-i} + \sum_{i=1}^{q} \beta_i^{(q)} \hat{e}_{t-i}^{(q-i)} + e_t^{(q)}, \qquad t = p+q+1, \ldots, n \quad (6.4.8)$$

where $\hat{e}_t^{(j)} = Z_t - \sum_{i=1}^{p} \phi_i^{(j)} Z_{t-i} - \sum_{i=1}^{q} \hat{\beta}_i^{(j)} \hat{e}_{t-i}^{(j-i)}$ is the estimated residual of the jth iterated AR(p) regression and the $\hat{\phi}_i^{(j)}$ and $\hat{\beta}_i^{(j)}$ are the corresponding least square estimates.

In practice, the true order p and q of the ARMA(p, q) model are usually unknown and have to be estimated. However, based on the preceding consideration, Tsay and Tiao (1984) suggest a general set of iterated regressions and introduce the concept of the extended sample autocorrelation function (ESACF) to estimate the orders p and q. Specifically, for $m = 0, 1, 2, \ldots$, let $\hat{\phi}_i^{(j)}, i = 1, \ldots, m$, be the OLS estimates from the jth iterated AR(m) regression of the ARMA process Z_t. They define the mth ESACF $\hat{\rho}_j^{(m)}$ of Z_t as the sample autocorrelation function for the transformed series

$$Y_t^{(j)} = (1 - \hat{\phi}_1^{(j)} B - \cdots - \hat{\phi}_m^{(j)} B^m) Z_t. \quad (6.4.9)$$

It is useful to arrange $\hat{\rho}_j^{(m)}$ in a two-way table as shown in Table 6.17 where the first row corresponding to $\hat{\rho}_j^{(0)}$ gives the standard sample ACF of Z_t, the

Table 6.17 The ESACF Table.

AR	MA					
	0	1	2	3	4	⋯
0	$\hat{\rho}_1^{(0)}$	$\hat{\rho}_2^{(0)}$	$\hat{\rho}_3^{(0)}$	$\hat{\rho}_4^{(0)}$	$\hat{\rho}_5^{(0)}$	⋯
1	$\hat{\rho}_1^{(1)}$	$\hat{\rho}_2^{(1)}$	$\hat{\rho}_3^{(1)}$	$\hat{\rho}_4^{(1)}$	$\hat{\rho}_5^{(1)}$	⋯
2	$\hat{\rho}_1^{(2)}$	$\hat{\rho}_2^{(2)}$	$\hat{\rho}_3^{(2)}$	$\hat{\rho}_4^{(2)}$	$\hat{\rho}_5^{(2)}$	⋯
3	$\hat{\rho}_1^{(3)}$	$\hat{\rho}_2^{(3)}$	$\hat{\rho}_3^{(3)}$	$\hat{\rho}_4^{(3)}$	$\hat{\rho}_5^{(3)}$	⋯
⋮			⋮			

Table 6.18 The asymptotic ESACF for an ARMA(1,1) model.

AR	MA					
	0	1	2	3	4	⋯
0	X	X	X	X	X	⋯
1	X	0	0	0	0	⋯
2	X	X	0	0	0	⋯
3	X	X	X	0	0	⋯
4	X	X	X	X	0	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋱

second row gives the first ESACF $\hat{\rho}_j^{(1)}$, and so on. The rows are numbered 0, 1, ... to specify the AR order, and the columns are numbered in a similar way for the MA order.

Note that the ESACF $\hat{\rho}_j^{(m)}$ is a function of n, the number of observations, even though it is not explicitly shown. In fact, it can be shown (see Tsay and Tiao [1984]) that for an ARMA(p, q) process, we have the following convergence in probability, i.e., for $m = 1, 2, \ldots$, and $j = 1, 2, \ldots$, we have

$$\hat{\rho}_j^{(m)} \xrightarrow{P} \begin{cases} 0, & 0 \le m - p < j - q, \\ X \neq 0, & \text{otherwise.} \end{cases} \quad (6.4.10)$$

Thus, by (6.4.10), the asymptotic ESACF table for an ARMA(1, 1) model becomes the one shown in Table 6.18. The zeros can be seen to form a triangle with the vertex at the (1, 1) position. More generally, for an ARMA(p, q) process, the vertex of the zero triangle in the asymptotic ESACF will be at the (p, q) position. Hence, the ESACF can be a useful tool in model identification, particularly for a mixed ARMA model.

Of course, in practice, we have finite samples, and the limit of $\hat{\rho}_j^{(m)}$ for $0 \le m - p < j - q$ may not be exactly zero. However, the asymptotic variance of $\hat{\rho}_j^{(m)}$ can be approximated using Bartlett's formula or more crudely by $(n - m - j)^{-1}$ on the hypothesis that the transformed series $Y_t^{(j)}$ of (6.4.9) is white noise. The ESACF table can then be constructed using indicator symbols with X referring to values greater than or less than ± 2 standard deviations and 0 for values within ± 2 standard deviations.

Example 6.9 To illustrate the method, we use SCA to compute the ESACF for the natural logarithms of Canadian lynx pelt sales discussed in Example 6.7. Table 16.19(a) shows the ESACF and Table 16.19(b) the corresponding indicator symbols for the ESACF of the series. The vertex of the triangle suggests a mixed ARMA(2, 2) model. This is different from an AR(3) model that we en-

Table 6.19 (a) The ESACF for natural logarithms of Canadian lynx pelt sales.

AR	MA									
	0	1	2	3	4	5	6	7	8	9
					(a) The ESACF					
0	.73	.22	−.32	−.69	−.76	−.53	−.08	.33	.61	.59
1	.68	.22	−.28	−.60	−.65	.51	−.09	.31	.53	.57
2	−.54	.01	.20	−.15	.17	−.26	.25	−.15	−.05	.10
3	−.53	.06	.32	.09	−.02	−.10	.15	−.16	−.20	.06
4	.00	.13	.37	−.12	−.02	−.09	−.01	−.18	−.16	−.03
					(b) The ESACF with indicator symbols					
0	X	0	X	X	X	X	0	X	X	X
1	X	0	0	X	X	X	0	X	X	X
2	X	0	0	0	0	0	0	0	0	0
3	X	0	X	0	0	0	0	0	0	0
4	0	X	X	0	0	0	0	0	0	0

tertained earlier based on the characteristic of the PACF. We further examine the series in Chapter 7.

Example 6.10 As another example, Table 6.20 shows the ESACF for Series C of 226 temperature readings in Box and Jenkins (1976). Because the vertex of the zero triangle occurs at the (2, 0) position, the ESACF approach suggests an AR(2) model for this series.

It should be noted that the OLS estimation is used in the iterated regression. Because the OLS regression estimation can always be applied to a regression model regardless of whether a series is stationary or nonstationary, invertible or noninvertible, Tsay and Tiao (1984) allow the roots of AR and MA polynomials to be on or outside the unit circle in their definition of the ESACF.

Table 6.20 The ESACF of Series C in Box and Jenkins.

AR	MA								
	0	1	2	3	4	5	6	7	8
0	X	X	X	X	X	X	X	X	X
1	X	X	X	X	X	X	X	X	X
2	0	0	0	0	0	0	0	0	0
3	X	0	0	0	0	0	0	0	0
4	X	X	0	0	0	0	0	0	0
5	X	X	X	0	0	0	0	0	0

Hence, the ESACF is computed from the original nondifferenced data. As a result, the ARIMA(p, d, q) process is identified as the ARMA(P, q) process with $P = p + d$. For example, based on the sample ACF and PACF of Series C as given in Table 6.21, Box and Jenkins (1976) suggested an ARIMA(1, 1, 0) model or an ARIMA(0, 2, 0) model. However, both these ARIMA(1, 1, 0) and ARIMA(0, 2, 0) models are identified as an ARMA(2, 0) model in terms of the ESACF.

Because the ESACF proposed by Tsay and Tiao (1984) is defined on the original series, they argue that the use of ESACF eliminates the need for differencing and provides a unified identification procedure to both stationary and nonstationary processes. To see whether a series is nonstationary, they suggest that for given specified values of p and q, the iterated AR estimates can be examined to see whether the AR polynomial contains a nonstationary factor with roots on the unit circle. For example, in the above Series C in Box and Jenkins (1976), the iterated AR estimates are $\hat{\phi}_1^{(0)} = 1.81$ and $\hat{\phi}_2^{(0)} = -.82$. Because $\hat{\phi}^{(0)}(B) \simeq (1-B)(1-.8B)$, it indicates that the AR polynomial contains the factor $(1-B)$. However, other than a few nice exceptions, the task of identifying nonstationarity through this approach is generally difficult. The real advantage of the ESACF is for the identification of p and q for the mixed ARMA mod-

Table 6.21 Sample ACF and PACF of Series C in Box and Jenkins (1976).

k	1	2	3	4	5	6	7	8	9	10
					(a) Z_t					
$\hat{\rho}_k$	.98	.94	.90	.85	.80	.75	.69	.64	.58	.52
St.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07
$\hat{\phi}_{kk}$	.98	−.81	−.03	−.02	−.10	−.07	−.01	−.03	.04	−.04
St.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07
					(b) $(1-B)Z_t$					
k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.80	.65	.53	.44	.38	.32	.26	.19	.14	.14
St.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07
$\hat{\phi}_{kk}$	.80	−.01	−.01	.06	.03	−.03	−.01	−.08	.00	.10
St.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07
					(c) $(1-B)^2Z_t$					
k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.08	−.07	−.12	−.06	.01	−.02	.05	−.05	−.12	.12
St.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07
$\hat{\phi}_{kk}$	−.08	−.08	−.14	−.10	−.03	−.05	.02	−.06	−.16	.09
St.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07

els. This advantage, I think, can be much better utilized if the ESACF is used for a properly transformed stationary series. This is particularly true because a tentatively identified model will be subjected to more efficient estimation procedures (such as the maximum likelihood estimation), which generally require stationarity.

Due to sampling variations and correlations among sample ACF, the pattern in the ESACF table from most time series may not be as clear-cut as those shown in the above examples. But from the author's experience, models can usually be identified without much difficulty through a joint study of ACF, PACF, and ESACF.

Some computer programs such as AUTOBOX and SCA provide the option to compute the ESACF in the model identification phase.

6.4.2 Other Identification Procedures

Other model identification procedures include the information criterion (AIC) proposed by Akaike (1974); the R-and-S-array introduced by Gray, Kelley, and McIntire (1978); and the corner method suggested by Beguin, Gourieroux, and Monfort (1980). The statistical properties of the statistics used in the R-and-S-array approach and the corner method are still largely unknown, and the software needed for these methods is not easily available. Interested readers are referred to their original research papers listed in the reference section of this book. The information criterion is discussed later in Chapter 7.

At this point, it is appropriate to say that model identification is both a science and an art. One should not use one method to the exclusion of others. Through careful examination of the ACF, PACF, IACF, ESACF, and other properties of time series, model identification truly becomes the most interesting aspect of time series analysis.

Exercises

6.1 Identify appropriate time series models from the sample ACF below. Justify your choice using the knowledge of the theoretical ACF for ARIMA models.

(a) $n = 121$, data $= Z_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.15	−.08	.04	.08	.08	.03	.02	.05	.04	−.11

(b) $n = 250$, data $= Z_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.63	.36	−.17	.09	−.07	.06	−.08	.10	−.11	.06

(c) $n = 250$, data $= Z_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	−.35	.17	.09	−.06	.01	−.01	−.04	.07	−.07	.09

(d) $n = 100$, data $= Z_t$, $W_t = (1-B)Z_t$, $\overline{W} = 2.5$, $S_W^2 = 20.$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_Z(k)$	.99	.98	.98	.97	.94	.91	.89	.86	.85	.83
$\hat{\rho}_W(k)$	.45	−.04	.12	.06	−.18	.16	−.07	.05	.10	.09

(e) $n = 100$, data $= Z_t$, $W_t = (1-B)Z_t$, $\overline{W} = 35$, $S_W^2 = 15$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_Z(k)$	.94	.93	.90	.89	.87	.86	.84	.81	.80	.80
$\hat{\rho}_W(k)$	.69	.50	.33	.19	.10	.08	.03	.01	.01	.00

6.2 Identify proper models for the following data sets (read across):

(a)	−2.401	−.574	.382	−.535	−1.639	−.960	−1.118
	−.719	−1.236	.117	−.493	−2.282	−1.823	.645
	−.179	.589	1.413	.370	.082	−.531	−1.891
	−.961	−.865	−.790	−1.476	−2.491	−4.479	−2.809
	−2.154	−1.532	−2.119	−3.349	−1.588	.740	.907
	1.540	.557	2.259	2.622	.701	2.463	2.714
	2.089	3.750	4.322	3.186	3.192	2.939	3.263
	3.279	.295	.227	1.356	1.912	1.060	.370
	−.195	.340	1.084	1.237	.610	2.126	3.960
	3.317	2.167	1.292	.595	.140	−.082	−.769
	.870	1.551	2.610	2.193	1.353	−.600	−.455
	.203	1.472	1.367	1.875	2.082	1.604	2.033
	3.746	2.954	.676	1.163	1.368	.343	−.334
	1.041	1.328	1.325	.968	1.970	2.296	2.896
	1.918	1.569					
(b)	−1.453	.867	.727	−.765	−1.317	.024	−.542
	−.048	−.805	.858	−.563	−1.986	−.454	1.738
	−.566	.697	1.060	−.478	−.140	−.581	−1.572
	.174	−.289	−.270	−1.002	−1.605	−2.984	−.122
	.469	−.239	−1.200	−2.077	.421	1.693	.463
	.996	−.367	1.925	1.267	−.872	2.043	1.236
	.461	2.497	2.072	.593	1.281	1.023	1.500
	1.321	−1.673	.050	1.219	1.098	−.087	−.266
	−.417	.457	.880	.586	−.132	1.760	2.684
	.941	.177	−.008	−.180	−.217	−.165	−.720
	1.332	1.029	1.679	.627	.038	−1.412	−.095
	.476	1.350	.484	1.055	.957	.355	1.071
	2.526	.707	−1.096	.757	.670	−.477	−.540
	1.241	.704	.528	.173	1.389	1.115	1.519
	.180	.419					
(c)	3.485	5.741	5.505	3.991	3.453	4.773	4.142
	4.598	3.796	5.430	3.960	2.541	4.054	6.155
	3.778	5.066	5.422	3.908	4.302	3.876	2.888
	4.613	4.075	4.054	3.288	2.654	1.215	3.979
	3.452	3.569	2.523	1.584	3.998	5.135	3.842

	4.404	3.077	5.432	4.795	2.747	5.767	4.988
	4.311	6.456	6.114	4.785	5.646	5.516	6.121
	6.059	3.196	5.050	6.231	6.119	4.988	4.885
	4.777	5.666	6.081	5.801	5.126	7.067	8.015
	6.358	5.752	5.700	5.614	5.629	5.705	5.155
	7.204	6.871	7.555	6.565	6.081	4.719	6.090
	6.637	7.492	6.635	7.264	7.221	6.694	7.493
	9.012	7.274	5.622	7.593	7.533	6.432	6.424
	8.219	7.668	7.534	7.232	8.501	8.266	8.748
	7.501	7.856					
(d)	.315	−.458	−.488	−.170	.565	−.344	−1.176
	−1.054	−.826	.710	−.341	−1.809	−1.242	−.667
	−.999	2.812	1.286	−1.084	−1.505	−2.556	−.144
	−1.749	−3.032	−2.958	−2.827	−3.392	−2.431	−2.757
	−2.822	−3.314	−2.738	−1.979	−1.671	−2.977	−.709
	.718	.736	.879	1.642	2.180	1.963	.716
	.769	.973	.334	1.309	.878	.062	.169
	.677	1.851	.242	.828	−.317	−1.042	−2.093
	.653	.261	2.020	2.136	1.635	−.141	−1.747
	−2.047	−.752	−.211	−1.062	−1.565	.232	.015
	−.935	−.338	.853	.888	3.069	3.364	3.854
	4.419	2.145	2.291	1.753	1.058	1.048	.200
	1.424	.590	.356	.476	.684	−2.260	−.569
	−1.014	−.207	.638	−.664	−.469	−.215	−.296
	−1.561	.246					

6.3 Consider the yearly data of lumber production (in billions of board feet) in the United States given below.

Year	Production									
1921–30	29.0	35.2	41.0	39.5	41.0	39.8	37.3	36.8	38.7	29.4
1931–40	20.0	13.5	17.2	18.8	22.9	27.6	29.0	24.8	28.8	31.2
1941–50	36.5	36.3	34.3	32.9	28.1	34.1	35.4	37.0	32.2	38.0
1951–60	37.2	37.5	36.7	36.4	37.4	38.2	32.9	33.4	37.2	32.9
1961–70	32.0	33.2	34.7	36.6	36.8	36.6	34.7	36.5	35.8	34.7
1971–80	37.0	37.7	38.6	34.6	32.6	36.3	39.4	40.5	40.6	35.4
1981–82	31.7	30.0								

(a) Plot the data and determine an appropriate model for the series.
(b) Find and plot the forecasts for the next 4 years, and calculate 95% forecast limits.
(c) Update your forecasts when the 1983 observation became available and equaled 34.6.

7 PARAMETER ESTIMATION, DIAGNOSTIC CHECKING, AND MODEL SELECTION

After identifying a tentative model, the next step is to estimate the parameters in the model. With full generality, we consider the general ARMA(p, q) model. That is, we discuss the estimation of parameters $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)'$, $\mu = E(Z_t)$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)$, and $\sigma_a^2 = E(a_t^2)$ in the model

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \cdots + \phi_p \dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

where $\dot{Z}_t = Z_t - \mu$, Z_t ($t = 1, 2, \ldots, n$) are n observed stationary or properly transformed stationary time series, and $\{a_t\}$ are i.i.d. $N(0, \sigma_a^2)$ white noise. Several widely used estimation procedures are discussed.

Once parameters have been estimated, we check on the adequacy of the model for the series. Very often several models can adequately represent a given series. Thus, after introducing diagnostic checking, we also present some criteria that are commonly used for model selection in time series model building.

7.1 THE METHOD OF MOMENTS

This method consists of substituting sample moments such as the sample mean, $\bar{Z}$, sample variance $\hat{\gamma}_0$, and sample ACF $\hat{\rho}_k$ for their theoretical counterparts and solving the resultant equations. For example, in the AR(p) process

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \cdots + \phi_p \dot{Z}_{t-p} + a_t, \tag{7.1.1}$$

the mean $\mu = E(Z_t)$ is estimated by $\bar{Z}$. To estimate $\boldsymbol{\phi}$, we first use the fact that $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}$ for $k > 1$ to obtain the following system of Yule-Walker equations:

$$\rho_1 = \phi_1 + \phi_2 \rho_1 + \phi_3 \rho_2 + \cdots + \phi_p \rho_{p-1}$$
$$\rho_2 = \phi_1 \rho_1 + \phi_2 + \phi_3 \rho_1 + \cdots + \phi_p \rho_{p-2}$$

	4.404	3.077	5.432	4.795	2.747	5.767	4.988
	4.311	6.456	6.114	4.785	5.646	5.516	6.121
	6.059	3.196	5.050	6.231	6.119	4.988	4.885
	4.777	5.666	6.081	5.801	5.126	7.067	8.015
	6.358	5.752	5.700	5.614	5.629	5.705	5.155
	7.204	6.871	7.555	6.565	6.081	4.719	6.090
	6.637	7.492	6.635	7.264	7.221	6.694	7.493
	9.012	7.274	5.622	7.593	7.533	6.432	6.424
	8.219	7.668	7.534	7.232	8.501	8.266	8.748
	7.501	7.856					
(d)	.315	−.458	−.488	−.170	.565	−.344	−1.176
	−1.054	−.826	.710	−.341	−1.809	−1.242	−.667
	−.999	2.812	1.286	−1.084	−1.505	−2.556	−.144
	−1.749	−3.032	−2.958	−2.827	−3.392	−2.431	−2.757
	−2.822	−3.314	−2.738	−1.979	−1.671	−2.977	−.709
	.718	.736	.879	1.642	2.180	1.963	.716
	.769	.973	.334	1.309	.878	.062	.169
	.677	1.851	.242	.828	−.317	−1.042	−2.093
	.653	.261	2.020	2.136	1.635	−.141	−1.747
	−2.047	−.752	−.211	−1.062	−1.565	.232	.015
	−.935	−.338	.853	.888	3.069	3.364	3.854
	4.419	2.145	2.291	1.753	1.058	1.048	.200
	1.424	.590	.356	.476	.684	−2.260	−.569
	−1.014	−.207	.638	−.664	−.469	−.215	−.296
	−1.561	.246					

6.3 Consider the yearly data of lumber production (in billions of board feet) in the United States given below.

Year	Production									
1921–30	29.0	35.2	41.0	39.5	41.0	39.8	37.3	36.8	38.7	29.4
1931–40	20.0	13.5	17.2	18.8	22.9	27.6	29.0	24.8	28.8	31.2
1941–50	36.5	36.3	34.3	32.9	28.1	34.1	35.4	37.0	32.2	38.0
1951–60	37.2	37.5	36.7	36.4	37.4	38.2	32.9	33.4	37.2	32.9
1961–70	32.0	33.2	34.7	36.6	36.8	36.6	34.7	36.5	35.8	34.7
1971–80	37.0	37.7	38.6	34.6	32.6	36.3	39.4	40.5	40.6	35.4
1981–82	31.7	30.0								

(a) Plot the data and determine an appropriate model for the series.
(b) Find and plot the forecasts for the next 4 years, and calculate 95% forecast limits.
(c) Update your forecasts when the 1983 observation became available and equaled 34.6.

7 PARAMETER ESTIMATION, DIAGNOSTIC CHECKING, AND MODEL SELECTION

After identifying a tentative model, the next step is to estimate the parameters in the model. With full generality, we consider the general ARMA(p, q) model. That is, we discuss the estimation of parameters $\phi = (\phi_1, \phi_2, \ldots, \phi_p)'$, $\mu = E(Z_t)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_q)$, and $\sigma_a^2 = E(a_t^2)$ in the model

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \cdots + \phi_p \dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

where $\dot{Z}_t = Z_t - \mu$, Z_t ($t = 1, 2, \ldots, n$) are n observed stationary or properly transformed stationary time series, and $\{a_t\}$ are i.i.d. $N(0, \sigma_a^2)$ white noise. Several widely used estimation procedures are discussed.

Once parameters have been estimated, we check on the adequacy of the model for the series. Very often several models can adequately represent a given series. Thus, after introducing diagnostic checking, we also present some criteria that are commonly used for model selection in time series model building.

7.1 THE METHOD OF MOMENTS

This method consists of substituting sample moments such as the sample mean, $\bar{Z}$, sample variance $\hat{\gamma}_0$, and sample ACF $\hat{\rho}_i$ for their theoretical counterparts and solving the resultant equations. For example, in the AR(p) process

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \cdots + \phi_p \dot{Z}_{t-p} + a_t, \tag{7.1.1}$$

the mean $\mu = E(Z_t)$ is estimated by $\bar{Z}$. To estimate ϕ, we first use the fact that $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}$ for $k > 1$ to obtain the following system of Yule-Walker equations:

$$\rho_1 = \phi_1 + \phi_2 \rho_1 + \phi_3 \rho_2 + \cdots + \phi_p \rho_{p-1}$$
$$\rho_2 = \phi_1 \rho_1 + \phi_2 + \phi_3 \rho_1 + \cdots + \phi_p \rho_{p-2}$$

$$\vdots$$

$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \cdots + \phi_p. \tag{7.1.2}$$

Then, replacing ρ_k by $\hat{\rho}_k$, we obtain the moment estimators $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$ by solving the above linear system of equations. That is,

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{p-2} & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-3} & \hat{\rho}_{p-2} \\ \vdots & & & & & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \hat{\rho}_{p-3} & \cdots & \hat{\rho}_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{bmatrix}. \tag{7.1.3}$$

These estimators are usually called Yule-Walker estimators.

Having obtained $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$, we use the result

$$\begin{aligned} \gamma_0 &= E(\dot{Z}_t\dot{Z}_t) = E[\dot{Z}_t(\phi_1\dot{Z}_{t-1} + \phi_2\dot{Z}_{t-2} + \cdots + \phi_p\dot{Z}_{t-p} + a_t)] \\ &= \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma_a^2 \end{aligned} \tag{7.1.4}$$

and obtain the moment estimator for σ_a^2 as

$$\hat{\sigma}_a^2 = \hat{\gamma}_0(1 - \hat{\phi}_1\hat{\rho}_1 - \hat{\phi}_2\hat{\rho}_2 - \cdots - \hat{\phi}_p\hat{\rho}_p). \tag{7.1.5}$$

Example 7.1 For the AR(1) model,

$$(Z_t - \mu) = \phi_1(Z_{t-1} - \mu) + a_t, \tag{7.1.6}$$

the Yule-Walker estimator for ϕ_1, from (7.1.3), is

$$\hat{\phi}_1 = \hat{\rho}_1. \tag{7.1.7}$$

The moment estimators for μ and σ_a^2 are given by

$$\hat{\mu} = \bar{Z} \tag{7.1.8}$$

and

$$\hat{\sigma}_a^2 = \hat{\gamma}_0(1 - \hat{\phi}_1\hat{\rho}_1) \tag{7.1.9}$$

respectively, where $\hat{\gamma}_0$ is the sample variance of the Z_t series.

Next, let us consider a simple MA(1) model

$$\dot{Z}_t = a_t - \theta_1 a_{t-1}. \tag{7.1.10}$$

Again, μ is estimated by $\bar{Z}$. For θ_1, we use the fact that

$$\rho_1 = \frac{-\theta_1}{1 + \theta_1^2}$$

and solve the above quadratic equation for θ_1 after replacing ρ_1 by $\hat{\rho}_1$. This is to

$$\hat{\theta}_1 = \frac{-1 \pm \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1}. \tag{7.1.11}$$

If $\hat{\rho}_1 = \pm.5$, we have a unique solution $\hat{\theta}_1 = \pm 1$, which gives a noninvertible model. If $|\hat{\rho}_1| > .5$, the real valued moment estimator $\hat{\theta}_1$ does not exist. This is expected, since a real valued MA(1) model always has $|\rho_1| < .5$, as discussed in Section 3.2.1. For $|\hat{\rho}_1| < .5$, there exist two distinct real valued solutions and we always choose the one that satisfies the invertibility condition. After having obtained $\hat{\theta}_1$, we calculate the moment estimator for σ_a^2 as

$$\hat{\sigma}_a^2 = \frac{\hat{\gamma}_0}{1 + \hat{\theta}_1^2}. \tag{7.1.12}$$

The above example of the MA(1) model shows that the moment estimators for MA and mixed ARMA models are complicated. More generally, regardless of AR, MA, or ARMA models, the moment estimators are very sensitive to rounding errors. They are usually used to provide initial estimates needed for a more efficient nonlinear estimation to be discussed later in this chapter. This is particularly true for the MA and mixed ARMA models. The moment estimators are not recommended for final estimation results and should not be used if the process is close to being nonstationary or noninvertible.

7.2 MAXIMUM LIKELIHOOD METHOD

7.2.1 Conditional Maximum Likelihood Estimation

For the general stationary ARMA(p, q) model

$$\dot{Z}_t = \phi_1\dot{Z}_{t-1} + \cdots + \phi_p\dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} \quad \cdots - \theta_q a_{t-q} \tag{7.2.1}$$

where $\dot{Z}_t = Z_t - \mu$ and $\{a_t\}$ are i.i.d. $N(0, \sigma_a^2)$ white noise, the joint probability density of $\mathbf{a} = (a_1, a_2, \ldots, a_n)'$ is given by

$$P(\mathbf{a} \mid \boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \sigma_a^2) = (2\pi\sigma_a^2)^{-n/2}\exp[-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n} a_t^2]. \tag{7.2.2}$$

Rewriting (7.2.1) as

$$a_t = \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + \dot{Z}_t - \phi_1\dot{Z}_{t-1} - \cdots - \phi_p\dot{Z}_{t-p}, \tag{7.2.3}$$

we can write down the likelihood function of the parameters $(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \sigma_a^2)$.

Let $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)'$ and assume the initial conditions $\mathbf{Z}_* = (Z_{1-p}, \ldots, Z_{-1}, Z_0)'$ and $\mathbf{a}_* = (a_{1-q}, \ldots, a_{-1}, a_0)'$. The conditional log-likelihood function

$$\ln L_*(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}, \sigma_a^2) = -\frac{n}{2}\ln 2\pi\sigma_a^2 - \frac{S_*(\boldsymbol{\phi}, \mu, \boldsymbol{\theta})}{2\sigma_a^2} \tag{7.2.4}$$

where

$$S_*(\boldsymbol{\phi}, \mu, \boldsymbol{\theta}) = \sum_{t=1}^{n} a_t^2(\boldsymbol{\phi}, \mu, \boldsymbol{\theta} \mid \mathbf{Z}_*, \mathbf{a}_*, \mathbf{Z}) \tag{7.2.5}$$

is the conditional sum of squares function. The quantities of $\hat{\phi}$, $\hat{\mu}$, and $\hat{\theta}$, which maximize Equation (7.2.4), are called the conditional maximum likelihood estimators. Since $\ln L_*(\phi, \mu, \theta, \sigma_a^2)$ involves the data only through $S_*(\phi, \mu, \theta)$, these estimators are the same as the conditional least squares estimators obtained from minimizing the conditional sum of squares function $S_*(\phi, \mu, \theta)$, which, we note, does not contain the parameter σ_a^2.

There are a few alternatives for specifying the initial conditions $\mathbf{Z}_*$ and $\mathbf{a}_*$. Based on the assumptions that $\{Z_t\}$ is stationary and $\{a_t\}$ is a series of i.i.d. $N(0, \sigma_a^2)$, random variables, we can replace the unknown Z_t by the sample mean $\bar{Z}$ and the unknown a_t by its expected value of 0. For the model in (7.2.1), we may also assume $a_p = a_{p-1} = \cdots = a_{p+1-q} = 0$ and calculate a_t for $t \geq (p+1)$ using (7.2.1). The conditional sum of squares function in (7.2.5) thus becomes

$$S_*(\phi, \mu, \theta) = \sum_{t=p+1}^{n} a_t^2(\phi, \mu, \theta \mid \mathbf{Z}), \tag{7.2.6}$$

which is also the form used by most computer programs.

After obtaining the parameter estimates $\hat{\phi}$, $\hat{\mu}$, and $\hat{\theta}$, the estimate $\hat{\sigma}_a^2$ of σ_a^2 is calculated from

$$\hat{\sigma}_a^2 = \frac{S_*(\hat{\phi}, \hat{\mu}, \hat{\theta})}{\text{d.f.}}, \tag{7.2.7}$$

where the number of degrees of freedom d.f. equals the number of terms used in the sum of $S_*(\hat{\phi}, \hat{\mu}, \hat{\theta})$ minus the number of parameters estimated. If (7.2.6) is used to calculate the sum of squares, d.f. $= (n-p) - (p+q+1) = n - (2p + q + 1)$. For other models, the d.f. should be adjusted accordingly.

7.2.2 Unconditional Maximum Likelihood Estimation and Backcasting Method

As seen from Chapter 5, one of the most important functions of a time series model is to forecast the unknown future values. Naturally, one asks whether we can back-forecast or backcast the unknown values $\mathbf{Z}_* = (Z_{1-p}, \ldots, Z_{-1}, Z_0)'$ and $\mathbf{a}_* = (a_{1-q}, \ldots, a_{-1}, a_0)'$ needed in the computation of the sum of squares and likelihood functions. Indeed, this is possible since any ARMA model can be written in either the forward form

$$(1 - \phi_1 B - \cdots - \phi_p B^p)\dot{Z}_t = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t \tag{7.2.8}$$

or the backward form

$$(1 - \phi_1 F - \cdots - \phi_p F^p)\dot{Z}_t = (1 - \theta_1 F - \cdots - \theta_q F^q)e_t \tag{7.2.9}$$

where $F^j Z_t = Z_{t+j}$. Because of the stationarity, (7.2.8) and (7.2.9) should have exactly the same autocovariance structure. This implies that $\{e_t\}$ is also a white noise series with mean zero and variance σ_e^2. Thus, in the same way as we use the forward form (7.2.8) to forecast the unknown future values Z_{n+j} for $j > 0$ based on the data $(Z_1, Z_2, \ldots, Z_n)$, we can also use the backward form (7.2.9) to backcast the unknown past values Z_j and hence compute a_j for $j \leq 0$ based on the data $\{Z_n, Z_{n-1}, \ldots, Z_1\}$. Therefore, for a further improvement in estimation, Box and Jenkins (1976) suggest the following unconditional log-likelihood function:

$$\ln L(\phi, \mu, \theta, \sigma_a^2) = -\frac{n}{2}\ln 2\pi\sigma_a^2 - \frac{S(\phi, \mu, \theta)}{2\sigma_a^2} \tag{7.2.10}$$

where $S(\phi, \mu, \theta)$ is the unconditional sum of squares function given by

$$S(\phi, \mu, \theta) = \sum_{t=-\infty}^{n} [E(a_t \mid \phi, \mu, \theta, \mathbf{Z})]^2 \tag{7.2.11}$$

and $E(a_t \mid \phi, \mu, \theta, \mathbf{Z})$ is the conditional expectation of a_t given ϕ, μ, θ, and $\mathbf{Z}$. Some of these terms have to be calculated using backcasts illustrated in Example 7.2.

The quantities $\hat{\phi}$, $\hat{\mu}$ and $\hat{\theta}$ that maximize Equation (7.2.10) are called unconditional maximum likelihood estimators. Again, since $\ln L(\phi, \mu, \theta, \sigma_a^2)$ involves the data only through $S(\phi, \mu, \theta)$, these unconditional maximum likelihood estimators are equivalent to the unconditional least squares estimators obtained by minimizing $S(\phi, \mu, \theta)$. In practice, the summation in (7.2.11) is approximated by a finite form

$$S(\phi, \mu, \theta) = \sum_{t=-M}^{n} [E(a_t \mid \phi, \mu, \theta, \mathbf{Z})]^2, \tag{7.2.12}$$

where M is a sufficiently large integer such that the backcast increment $|E(Z_t \mid \phi, \mu, \theta, \mathbf{Z}) - E(Z_{t-1} \mid \phi, \mu, \theta, \mathbf{Z})|$ is less than any arbitrary predetermined small ϵ value for $t \leq -(M+1)$. This implies that $E(Z_t \mid \phi, \mu, \theta, \mathbf{Z}) \simeq \mu$ and hence $E(a_t \mid \phi, \mu, \theta, \mathbf{Z})$ is negligible for $t \leq -(M+1)$.

After obtaining the parameter estimates $\hat{\phi}$, $\hat{\mu}$, and $\hat{\theta}$, the estimate $\hat{\sigma}_a^2$ of σ_a^2 can then be calculated as

$$\hat{\sigma}_a^2 = \frac{S(\hat{\phi}, \hat{\mu}, \hat{\theta})}{n}. \tag{7.2.13}$$

For efficiency, the use of backcasts for parameter estimation is important for seasonal models (to be discussed in Chapter 8), for models that are close to being nonstationary, and especially for series that are relatively short. Most computer programs have implemented this option.

Example 7.2 To illustrate the backcasting method, consider the AR(1) model that can be written in the forward form

$$a_t = Z_t - \phi Z_{t-1} \tag{7.2.14}$$

or equivalently in the backward form

$$e_t = Z_t - \phi Z_{t+1} \tag{7.2.15}$$

where, without loss of generality, we assume that $E(Z_t) = 0$. Consider a very simple example with ten observations, $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_{10})$, from the processes that are listed in Table 7.1 under the column $E(Z_t \mid \mathbf{Z})$ for $t = 1, 2, \ldots, 10$. Suppose $\phi = .3$ and we want to calculate the unconditional sum of squares

$$S(\phi = .3) = \sum_{t=-M}^{10} [E(a_t \mid \phi = .3, \mathbf{Z})]^2 \tag{7.2.16}$$

where M is chosen so that $|E(Z_t \mid \phi = .3, \mathbf{Z}) - E(Z_{t-1} \mid \phi = .3, \mathbf{Z})| < .005$ for $t \leq -(M + 1)$. To simplify the notations for this example we write $E(a_t \mid \phi = .3, \mathbf{Z})$ as $E(a_t \mid \mathbf{Z})$ and $E(Z_t \mid \phi = .3, \mathbf{Z})$ as $E(Z_t \mid \mathbf{Z})$.

To obtain $E(a_t \mid \mathbf{Z})$ we use (7.2.14) and compute

$$E(a_t \mid \mathbf{Z}) = E(Z_t \mid \mathbf{Z}) - \phi E(Z_{t-1} \mid \mathbf{Z}). \tag{7.2.17}$$

However, the above computation of $E(a_t \mid \mathbf{Z})$ for $t \leq 1$ involves the unknown Z_t values for $t \leq 0$, which need to be backcasted. To achieve this, we use the backward form in (7.2.15), i.e.,

$$E(Z_t \mid \mathbf{Z}) = E(e_t \mid \mathbf{Z}) + \phi E(Z_{t+1} \mid \mathbf{Z}). \tag{7.2.18}$$

First, we note that in terms of the backward form, e_t for $t \leq 0$ are unknown future random shocks with respect to the observations $Z_n, Z_{n-1}, \ldots, Z_2$, and Z_1. Hence,

$$E(e_t \mid \mathbf{Z}) = 0, \qquad \text{for } t \leq 0. \tag{7.2.19}$$

Therefore, for $\phi = .3$, we have from (7.2.18)

$$\begin{aligned}
E(Z_0 \mid \mathbf{Z}) &= E(e_0 \mid \mathbf{Z}) + .3E(Z_1 \mid \mathbf{Z}) \\
&= 0 + (.3)(-.2) = -.06 \\
E(Z_{-1} \mid \mathbf{Z}) &= E(e_{-1} \mid \mathbf{Z}) + .3E(Z_0 \mid \mathbf{Z}) \\
&= 0 + (.3)(-.06) = -.018 \\
E(Z_{-2} \mid \mathbf{Z}) &= E(e_{-2} \mid \mathbf{Z}) + .3E(Z_{-1} \mid \mathbf{Z}) \\
&= (.3)(-.018) = -.0054 \\
E(Z_{-3} \mid \mathbf{Z}) &= E(e_{-3} \mid \mathbf{Z}) + .3E(Z_{-2} \mid \mathbf{Z}) \\
&= (.3)(-.0054) = -.00162.
\end{aligned}$$

Since $|E(Z_{-3} \mid \mathbf{Z}) - E(Z_{-2} \mid \mathbf{Z})| = .00378 < .005$, the predetermined ϵ value, we choose $M = 2$.

Now, with these backcasted values Z_t for $t \leq 0$, we can return to the forward form in (7.2.17) to compute $E(a_t \mid \mathbf{Z})$ for $\phi = .3$ from $t = -2$ to $t = 10$

as follows:

$$\begin{aligned}
E(a_{-2} \mid \mathbf{Z}) &= E(Z_{-2} \mid \mathbf{Z}) - .3E(Z_{-3} \mid \mathbf{Z}) \\
&= -.0054 - (.3)(-.00162) = -.0049 \\
E(a_{-1} \mid \mathbf{Z}) &= E(Z_{-1} \mid \mathbf{Z}) - .3E(Z_{-2} \mid \mathbf{Z}) \\
&= -.018 - (.3)(-.0054) = -.0164 \\
E(a_0 \mid \mathbf{Z}) &= E(Z_0 \mid \mathbf{Z}) - .3E(Z_{-1} \mid \mathbf{Z}) \\
&= -.06 - (.3)(-.018) = -.0546 \\
E(a_1 \mid \mathbf{Z}) &= E(Z_1 \mid \mathbf{Z}) - .3E(Z_0 \mid \mathbf{Z}) \\
&= -.2 - (.3)(-.06) = -.182 \\
E(a_2 \mid \mathbf{Z}) &= E(Z_2 \mid \mathbf{Z}) - .3E(Z_1 \mid \mathbf{Z}) \\
&= -.4 - (.3)(-.2) = -.34 \\
&\vdots \\
E(a_{10} \mid \mathbf{Z}) &= E(Z_{10} \mid \mathbf{Z}) - .3E(Z_9 \mid \mathbf{Z}) \\
&= -.2 - (.3)(-.1) = -.17.
\end{aligned}$$

All the above computations can be carried out systematically as shown in Table 7.1 and we obtain

$$S(\phi = .3) = \sum_{t=-2}^{10} [E(a_t \mid \phi = .3, \mathbf{Z})]^2 = .8232.$$

Similarly, we can obtain $S(\phi)$ for other values of ϕ and hence find its minimum.

Table 7.1 Calculation of $S(\phi = .3)$ for $(1 - \phi B)Z_t = a_t$ using backcasting method.

t	$E(a_t \mid \mathbf{Z})$	$-.3E(Z_{t-1} \mid \mathbf{Z})$	$E(Z_t \mid \mathbf{Z})$	$.3E(Z_{t+1} \mid \mathbf{Z})$	$E(e_t \mid \mathbf{Z})$
−3			−.0016	−.0016	0
−2	−.0049	.0005	−.0054	−.0054	0
−1	−.0164	.0016	−.018	−.018	0
0	−.0546	.0054	−.06	−.06	0
1	−.182	.018	−.2		
2	−.34	.06	−.4		
3	−.38	.12	−.5		
4	−.35	.15	−.5		
5	−.45	.15	−.6		
6	−.32	.18	−.5		
7	−.25	.15	−.4		
8	−.08	.12	−.2		
9	−.04	.06	−.1		
10	−.17	.03	−.2		

It should be noted that for the AR(1) model we do not need the value $E(e_t \mid \mathbf{Z})$ for $t \geq 1$. For other models, they may be required. However, the procedure is the same. For more detailed examples, we refer readers to Box and Jenkins (1976, p. 212).

7.2.3 Exact Likelihood Functions

Both the conditional and unconditional likelihood functions (7.2.4) and (7.2.10) are approximations. To illustrate the derivation of the exact likelihood function for a time series model, consider the AR(1) process

$$(1-\phi B)\dot{Z}_t = a_t \tag{7.2.20}$$

or

$$\dot{Z}_t = \phi\dot{Z}_{t-1} + a_t$$

where $\dot{Z}_t = (Z_t - \mu)$, $|\phi| < 1$ and the a_t are i.i.d. $N(0, \sigma_a^2)$. Rewriting the process in the moving average representation, we have

$$\dot{Z}_t = \sum_{j=0}^{\infty} \phi^j a_{t-j}. \tag{7.2.21}$$

Clearly, the $\dot{Z}_t$ will be distributed as $N(0, \sigma_a^2/(1-\phi^2))$. However, the $\dot{Z}_t$ are highly correlated. To derive the joint probability density function $P(\dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n)$ of $(\dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n)$ and hence the likelihood function for the parameters, we consider

$$\begin{aligned} e_1 &= \sum_{j=0}^{\infty} \phi^j a_{1-j} = \dot{Z}_1, \\ a_2 &= \dot{Z}_2 - \phi\dot{Z}_1, \\ a_3 &= \dot{Z}_3 - \phi\dot{Z}_2, \\ &\vdots \\ a_n &= \dot{Z}_n - \phi\dot{Z}_{n-1}. \end{aligned} \tag{7.2.22}$$

Note that e_1 follows the normal distribution $N(0, \sigma_a^2/(1-\phi^2))$, a_t, for $2 \leq t \leq n$, follows the normal distribution $N(0, \sigma_a^2)$, and they are all independent of each other. Hence, the joint probability density of $(e_1, a_2, \ldots, a_n)$ is

$$\begin{aligned} &p(e_1, a_2, \ldots, a_n) \\ &= \left[\frac{(1-\phi^2)}{2\pi\sigma_a^2}\right]^{1/2} \exp\left[\frac{-e_1^2(1-\phi^2)}{2\sigma_a^2}\right]\left[\frac{1}{2\pi\sigma_a^2}\right]^{(n-1)/2} \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=2}^{n} a_t^2\right]. \end{aligned} \tag{7.2.23}$$

Now consider the following transformation:

$$\begin{aligned} \dot{Z}_1 &= e_1 \\ \dot{Z}_2 &= \phi\dot{Z}_1 + a_2 \\ \dot{Z}_3 &= \phi\dot{Z}_2 + a_3 \\ &\vdots \\ \dot{Z}_n &= \phi\dot{Z}_{n-1} + a_n. \end{aligned} \tag{7.2.24}$$

The Jacobian for the transformation, from (7.2.22), is

$$J = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ -\phi & 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & -\phi & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \vdots & & & & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & -\phi & 1 \end{bmatrix} = 1,$$

It follows that

$$\begin{aligned} P(\dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n) &= P(e_1, a_2, \ldots, a_n) \\ &= \left[\frac{(1-\phi^2)}{2\pi\sigma_a^2}\right]^{1/2} \exp\left[\frac{-\dot{Z}_1^2(1-\phi^2)}{2\sigma_a^2}\right] \\ &\quad \cdot \left[\frac{1}{2\pi\sigma_a^2}\right]^{(n-1)/2} \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=2}^{n}(\dot{Z}_t - \phi\dot{Z}_{t-1})^2\right]. \end{aligned} \tag{7.2.25}$$

Hence, for a given series $(\dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n)$ we have the following exact log-likelihood function:

$$\ln L(\dot{Z}_1, \ldots, \dot{Z}_n \mid \phi, \mu, \sigma_a^2) = -\frac{n}{2}\ln 2\pi + \frac{1}{2}\ln(1-\phi^2) - \frac{n}{2}\ln\sigma_a^2 - \frac{S(\phi, \mu)}{2\sigma_a^2} \tag{7.2.26}$$

where

$$S(\phi, \mu) = (Z_1 - \mu)^2(1-\phi^2) + \sum_{t=2}^{n}[(Z_t - \mu) - \phi(Z_{t-1} - \mu)]^2 \tag{7.2.27}$$

is the sum of squares term that is a function of only ϕ and μ.

The exact closed form of the likelihood function of a general ARMA model is complicated. Tiao and Ali (1971) derived it for an ARMA(1, 1) model. Newbold (1974) derived it for a general ARMA(p, q) model. Interested readers are advised also to see Ansley (1979), Nicholls and Hall (1979), Ljung and Box (1979), and Hillmer and Tiao (1979) among others for additional references.

7.3 NONLINEAR ESTIMATION

It is clear that the maximum likelihood estimation and the least squares estimation involve minimizing either the conditional sum of squares $S_*(\phi, \mu, \theta)$ or the unconditional sum of squares $S(\phi, \mu, \theta)$. These are the sums of squares of the error terms $a_t's$. For an AR(p) process,

$$a_t = \dot{Z}_t - \phi_1\dot{Z}_{t-1} - \phi_2\dot{Z}_{t-2} - \cdots - \phi_p\dot{Z}_{t-p}, \tag{7.3.1}$$

and a_t is clearly linear in parameters. However, for a model containing an MA factor, the a_t is nonlinear in parameters. To see that, consider a simple ARMA(1, 1) model

$$\dot{Z}_t - \phi_1\dot{Z}_{t-1} = a_t - \theta_1 a_{t-1}. \tag{7.3.2}$$

To calculate a_t, we note that

$$\begin{aligned} a_t &= \dot{Z}_t - \phi_1\dot{Z}_{t-1} + \theta_1 a_{t-1} \\ &= \dot{Z}_t - \phi_1\dot{Z}_{t-1} + \theta_1(\dot{Z}_{t-1} - \phi_1\dot{Z}_{t-2} + \theta_1 a_{t-2}) \\ &= \dot{Z}_t - (\phi_1 - \theta_1)\dot{Z}_{t-1} - \phi_1\theta_1\dot{Z}_{t-2} + \theta_1^2 a_{t-2} \\ &\vdots \end{aligned} \tag{7.3.3}$$

which is clearly nonlinear in the parameters. Hence, for a general ARMA model, a nonlinear least squares estimation procedure must be used to obtain estimates.

The nonlinear least squares procedure involves an iterative search technique. Because a linear model is a special case of the nonlinear model, we can illustrate the main ideas of the nonlinear least squares using the following linear regression model:

$$\begin{aligned} Y_t &= E(Y_t \mid X_{it}'s) + e_t \\ &= \alpha_1 X_{t1} + \alpha_2 X_{t2} + \cdots + \alpha_p X_{tp} + e_t \end{aligned} \tag{7.3.4}$$

for $t = 1, 2, \ldots, n$, where $e_t's$ are i.i.d. $N(0, \sigma_e^2)$ independent of all the X_{it}. Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$ and $\bar{X}$ be the corresponding matrix for the independent variables $X_{it}'s$. From results in linear regression analysis, we know that the least squares estimators are given by

$$\hat{\alpha} = (\bar{X}'\bar{X})^{-1}\bar{X}'\mathbf{Y}, \tag{7.3.5}$$

which follows a multivariate normal distribution MN(α, $V(\hat{\alpha})$) with

$$V(\hat{\alpha}) = \sigma_a^2(\bar{X}'\bar{X})^{-1}. \tag{7.3.6}$$

The minimum residual (error) sum of squares is

$$S(\hat{\alpha}) = \sum_{t=1}^{n}(Y_t - \hat{\alpha}_1 X_{t1} - \hat{\alpha}_2 X_{t2} - \cdots - \hat{\alpha}_p X_{tp})^2. \tag{7.3.7}$$

The least squares estimates in (7.3.5) can also be obtained by the following two-step procedure discussed in Miller and Wichern (1977).

Let $\tilde{\alpha} = (\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_p)'$ be an initial guess value of $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$. We can rewrite the minimum residual sum of squares in (7.3.7) as

$$\begin{aligned} S(\hat{\alpha}) = \sum_{t=1}^{n}[Y_t - \tilde{\alpha}_1 X_{t1} - \cdots - \tilde{\alpha}_p X_{tp} \\ - (\hat{\alpha}_1 - \tilde{\alpha}_1)X_{t1} - \cdots - (\hat{\alpha}_p - \tilde{\alpha}_p)X_{tp}]^2 \end{aligned} \tag{7.3.8}$$

or

$$S(\delta) = S(\hat{\alpha}) = \sum_{t=1}^{n}(\tilde{e}_t - \delta_1 X_{t1} \ldots - \delta_p X_{tp})^2 \tag{7.3.9}$$

where $\tilde{e}_t$'s are estimated residuals based on the initial given values $\tilde{\alpha}$, and $\delta = (\hat{\alpha} - \tilde{\alpha})$. Now, $S(\delta)$ in Equation (7.3.9) and $S(\hat{\alpha})$ in (7.3.7) are in the same form. Hence the least squares value of δ is

$$\delta = (\bar{X}'\bar{X})^{-1}\bar{X}'\tilde{\mathbf{e}} \tag{7.3.10}$$

where $\tilde{\mathbf{e}} = (\tilde{e}_1, \ldots, \tilde{e}_n)'$. Once the values $\delta = (\delta_1, \delta_2, \ldots, \delta_p)$ are calculated, the least squares estimates are given by

$$\hat{\alpha} = \tilde{\alpha} + \delta. \tag{7.3.11}$$

We note that the residual $\tilde{e}_t$ is calculated as $Y_t - \tilde{Y}_t$ where

$$\tilde{Y}_t = \tilde{\alpha}_1 X_{t1} + \cdots + \tilde{\alpha}_p X_{tp}$$

represents a guess of the regression equation obtained by using the original model and the given initial values $\tilde{\alpha}_i's$. Moreover, it is clear from Equation (7.3.4) that

$$\frac{\partial E(Y_t \mid X's)}{\partial \alpha_i} = X_{ti} \tag{7.3.12}$$

for $i = 1, 2, \ldots, p$ and $t = 1, 2, \ldots, n$. Hence, the $\bar{X}$ matrix used in the equation of the least squares estimates in (7.3.5) and (7.3.10) is actually the matrix of the partial derivatives of the regression function with respect to each of the parameters.

Now, consider the following model (linear or nonlinear):

$$Y_t = f(\mathbf{X}_t, \alpha) + e_t, \qquad t = 1, 2, \ldots, n \tag{7.3.13}$$

where $\mathbf{X}_t = (X_{t1}, X_{t2}, \ldots, X_{tp})$ is a set of independent variables corresponding to the observations, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$ is a vector of parameters, and e_t is a white noise series having zero mean and constant variance σ_e^2 independent of $\mathbf{X}$. Let $Y = (Y_1, Y_2, \ldots, Y_n)'$ and $\mathbf{f}(\alpha) = [f(\mathbf{X}_1, \alpha), f(\mathbf{X}_2, \alpha), \ldots, f(\mathbf{X}_n, \alpha)]'$. From the above discussion, the least squares estimators (linear or nonlinear) can always be calculated iteratively as follows:

Step 1. Given any vector of initial guess values $\tilde{\alpha}$, compute the residual $\tilde{\mathbf{e}} = (\mathbf{Y} - \tilde{\mathbf{Y}})$ and the residual sum of squares

$$S(\tilde{\alpha}) = \tilde{\mathbf{e}}'\tilde{\mathbf{e}} = (\mathbf{Y} - \tilde{\mathbf{Y}})'(\mathbf{Y} - \tilde{\mathbf{Y}}), \tag{7.3.14}$$

where $\tilde{\mathbf{Y}} = \mathbf{f}(\tilde{\alpha})$ is a vector of predicted values obtained by replacing the unknown parameters by the initial guess values. Approximate the model $f(X_t, \tilde{\alpha})$ with the first order Taylor series expansion about the initial value $\tilde{\alpha}$. That is,

$$\mathbf{f}(\alpha) = \mathbf{f}(\tilde{\alpha}) + \bar{X}_{\tilde{\alpha}}\delta \tag{7.3.15}$$

where $\delta = (\alpha - \tilde{\alpha})$ and $\bar{X}_{\tilde{\alpha}} = \{X_{ij}\}$ is the $n \times p$ matrix of the partial derivatives at $\tilde{\alpha}$ in the above linear approximation. That is,

$$X_{ij} = \frac{\partial f(\mathbf{X}_t, \alpha)}{\partial \alpha_j}\bigg|_{\alpha=\tilde{\alpha}}, \qquad \begin{matrix} t = 1, 2, \ldots, n, \\ j = 1, 2, \ldots, p. \end{matrix} \tag{7.3.16}$$

Then we calculate

$$\delta = (\bar{X}'_{\tilde{\alpha}}\bar{X}_{\tilde{\alpha}})^{-1}\bar{X}'_{\tilde{\alpha}}\tilde{\mathbf{e}} = (\delta_1, \delta_2, \ldots, \delta_p)'. \tag{7.3.17}$$

Note that for a linear model the $\bar{X}_{\tilde{\alpha}}$ is fixed and equals $\bar{X}$; for a nonlinear model, this $\bar{X}_{\tilde{\alpha}}$ changes from iteration to iteration.

Step 2. Obtain the updated least square estimates

$$\hat{\alpha} = \tilde{\alpha} + \delta \tag{7.3.18}$$

and the corresponding residual sum of squares $S(\hat{\alpha})$. We note that δ_i in δ represents the difference or change in the parameter values. For a linear model, Step 2 gives the final least squares estimates. For a nonlinear model, Step 2 only leads to new initial values for further iterations.

In summary, for a given general ARMA(p, q) model, we can use the nonlinear least squares procedure to find the least squares estimates that minimize the error sum of squares $S_*(\phi, \mu, \theta)$ or $S(\phi, \mu, \theta)$. The nonlinear least squares routine starts with initial guess values of the parameters. It monitors these values in the direction of the smaller sum of squares and updates the initial guess values. The iterations continue until some specified convergence criteria are reached. Some convergence criteria that have been used are the relative reduction in the sum of squares, the maximum change in the parameter values less than a specified level, or the number of iterations greater than a certain number. To achieve a proper and faster convergence, many search algorithms have been developed. One of the algorithms that is commonly used is due to Marquardt (1963). It is a compromise between the Gauss-Newton method and the method of steepest descent. For more discussions on nonlinear estimation, see Draper and Smith (1981), among others.

Properties of the Parameter Estimates Let $\alpha = (\phi, \mu, \theta)$, $\hat{\alpha}$ be the estimate of α, and $X_{\hat{\alpha}}$ be the matrix of the partial derivatives in the final iteration of the nonlinear least squares procedure. We know that $\hat{\alpha}$ is distributed as a multivariate normal distribution MN(α, $V(\hat{\alpha})$). The estimated variance-covariance matrix $\hat{V}(\hat{\alpha})$ of $\hat{\alpha}$ is

$$\begin{aligned} \hat{V}(\hat{\alpha}) &= \hat{\sigma}_a^2(\bar{X}'_{\hat{\alpha}}\bar{X}_{\hat{\alpha}})^{-1} \\ &= (\hat{\sigma}_{\hat{\alpha}_i\hat{\alpha}_j}) \end{aligned} \tag{7.3.19}$$

where $\hat{\sigma}_a^2$ is estimated as in (7.2.7) or (7.2.13) and $\hat{\sigma}_{\hat{\alpha}_i\hat{\alpha}_j}$ is the sample covariance between $\hat{\alpha}_i$ and $\hat{\alpha}_j$. We can test the hypothesis $H_0 : \alpha_i = \alpha_{i0}$ using the following t statistic:

$$t = \frac{\hat{\alpha}_i - \alpha_{i0}}{\sqrt{\hat{\sigma}_{\hat{\alpha}_i\hat{\alpha}_i}}}$$

with the degrees of freedom equaling $n - (p + q + 1)$ for the general ARMA model in (7.2.1). (More generally, the degrees of freedom equals the sample size used in estimation minus the number of parameters estimated in the model.) The estimated correlation matrix of these estimates is

$$\hat{R}(\alpha) = (\hat{\rho}_{\hat{\alpha}_i\hat{\alpha}_j}) \tag{7.3.21}$$

where

$$\hat{\rho}_{\hat{\alpha}_i\hat{\alpha}_j} = \frac{\hat{\sigma}_{\hat{\alpha}_i\hat{\alpha}_j}}{\sqrt{\hat{\sigma}_{\hat{\alpha}_i\hat{\alpha}_i}\hat{\sigma}_{\hat{\alpha}_j\hat{\alpha}_j}}}.$$

A high correlation among estimates indicates overparameterization, which should be avoided as it often causes difficulties in the convergence of the nonlinear least squares.

7.4 ORDINARY LEAST SQUARES (OLS) ESTIMATION IN TIME SERIES ANALYSIS

Regression analysis is possibly the most commonly used statistical method in data analysis. As a result, the ordinary least squares (OLS) estimation developed for standard regression models is perhaps also the most frequently used estimation procedure in statistics. In this section, we discuss some problems of OLS estimation in time series analysis.

Consider the following simple linear regression model:

$$Z_t = \phi X_t + e_t, \qquad t = 1, \ldots, n. \tag{7.4.1}$$

Under the following basic assumptions on the error term e_t:

1. Zero mean: $E(e_t) = 0$
2. Constant Variance: $E(e_t^2) = \sigma_e^2$
3. Nonautocorrelation: $E(e_t e_k) = 0$ for $t \neq k$
4. Uncorrelated with explanatory variable X_t : $E(X_t e_t) = 0$

it is well known that the OLS estimator

$$\hat{\phi} = \frac{\sum_{t=1}^{n} X_t Z_t}{\sum_{t=1}^{n} X_t^2} \tag{7.4.2}$$

is a consistent and the best linear unbiased estimator of ϕ. However, it is important to note that assumption (4) is crucial for this result to hold. Assumption (4) automatically follows if the explanatory variables are nonstochastic. However, in a noncontrollable study, particularly when time series data are involved, the explanatory variables are usually also random variables.

Now, consider the following time series model:

$$Z_t = \phi Z_{t-1} + e_t, \qquad t = 1, \ldots, n. \tag{7.4.3}$$

The OLS estimator of ϕ, based on available data, is

$$\hat{\phi} = \frac{\sum_{t=2}^{n} Z_{t-1} Z_t}{\sum_{t=2}^{n} Z_{t-1}^2}. \tag{7.4.4}$$

We would like to ask whether $\hat{\phi}$ is still unbiased and consistent in this case when the explanatory variable is a lagged dependent variable. The answer depends on the stochastic nature of the error term e_t. To see that, we rewrite $\hat{\phi}$ as

$$\hat{\phi} = \frac{\sum_{t=2}^{n} Z_{t-1} Z_t}{\sum_{t=2}^{n} Z_{t-1}^2} = \frac{\sum_{t=2}^{n} Z_{t-1}(\phi Z_{t-1} + e_t)}{\sum_{t=2}^{n} Z_{t-1}^2}$$

$$= \phi + \frac{\sum_{t=2}^{n} Z_{t-1} e_t}{\sum_{t=2}^{n} Z_{t-1}^2}, \tag{7.4.5}$$

and consider the following two cases:

Case 1: $e_t = a_t$. That is, the e_t is a zero mean white noise series of constant variance σ_a^2. In this case, conditional on the observed Z_t, the expected value of $\hat{\phi}$ is ϕ and hence, $\hat{\phi}$ is still unbiased. Moreover, it is easy to see that $\hat{\phi}$ in (7.4.4) is equivalent to the first lag sample autocorrelation $\hat{\rho}_1$ for the series Z_t. If $|\phi| < 1$, and hence Z_t becomes an AR(1) process with an absolutely summable autocorrelation function, then by Section 2.5, $\hat{\rho}_1$ is a consistent estimator of ρ_1, which is equal to ϕ. Thus, $\hat{\phi}$ in (7.4.4) is a consistent estimator of ϕ.

Case 2: $e_t = (1 - \theta B)a_t$, where the a_t is a zero mean white noise series of constant variance σ_a^2, and hence e_t is an MA(1) process. Under this condition, the series Z_t becomes an ARMA(1, 1) process

$$Z_t = \phi Z_{t-1} + a_t - \theta a_{t-1} \tag{7.4.6}$$

$$E(Z_{t-1} e_t) = E[Z_{t-1}(a_t - \theta a_{t-1})] = -\theta \sigma_a^2. \tag{7.4.7}$$

This shows that autocorrelation in the error term not only violates assumption (3) but also causes a violation of assumption (4) when the explanatory variables contain a lagged dependent variable. Thus, $\hat{\phi}$ in (7.4.4) is no longer unbiased. Even worse, $\hat{\phi}$ is not a consistent estimator of ϕ, because $\hat{\phi} \simeq \hat{\rho}_1$ is a consistent estimator for ρ_1, and for an ARMA(1, 1) process, by (3.4.14),

$$\rho_1 = \frac{(\phi - \theta)(1 - \phi\theta)}{1 + \theta^2 - 2\phi\theta} \neq \phi.$$

In summary, the OLS estimator for the parameter of an explanatory variable in a regression model will be inconsistent unless the error term is uncorrelated with the explanatory variable. For ARMA(p, q) models, this condition usually does not hold except when $q = 0$. Estimation methods discussed in Sections 7.2 and 7.3 are more efficient and commonly used in time series analysis.

7.5 DIAGNOSTIC CHECKING

Time series model building is an iterative procedure. It starts with model identification and parameter estimation. After parameter estimation, we have to assess model adequacy by checking whether the model assumptions are satisfied. The basic assumption is that the $\{a_t\}$ are white noise. That is, the a_t's are uncorrelated random shocks with zero mean and constant variance. For any estimated model, the residuals $\hat{a}_t$'s are estimates of these unobserved white noise a_t's. Hence, model diagnostic checking is accomplished through a careful analysis of the residual series $\{\hat{a}_t\}$. Because this residual series is the product of parameter estimation, the model diagnostic checking is usually contained in the estimation phase of a time series package.

To check whether the errors are normally distributed, one can construct a histogram of the standardized residuals $\hat{a}_t/\hat{\sigma}_a$ and compare it with the standard normal distribution using the chi-square goodness of fit test or even Tukey's simple five-number summary. To check whether the variance is constant, we can examine the plot of residuals or evaluate the effect of different λ values via the Box-Cox method. To check whether the residuals are white noise, we compute the sample ACF and PACF (or IACF) of the residuals to see whether they do not form any pattern and are all statistically insignificant, i.e., within two standard deviations if $\alpha = .05$.

Another useful test is the portmanteau lack of fit test. This test uses all the residual sample ACF's as a unit to check the joint null hypothesis

$$H_0 : \rho_1 = \rho_2 = \cdots = \rho_K = 0,$$

with the test statistic

$$Q = n(n+2)\sum_{k=1}^{K} (n-k)^{-1} \hat{\rho}_k^2. \tag{7.5.1}$$

This test statistic is the modified Q statistic originally proposed by Box and Pierce (1970). Under the null hypothesis of model adequacy, Ljung and Box (1978) and Ansley and Newbold (1979) show that the Q statistic approximately

follows the $\chi^2(K-m)$ distribution where m is the number of parameters estimated in the model.

Based on the results of these residual analyses, if the entertained model is inadequate, a new model can be easily derived. For example, assume the entertained AR(1) model

$$(1-\phi_1 B)(Z_t-\mu)=b_t \tag{7.5.2}$$

produces an MA(1) residual series instead of a white noise series, i.e.,

$$b_t=(1-\theta_1 B)a_t. \tag{7.5.3}$$

Then we should re-identify an ARMA(1, 1) model

$$(1-\phi_1 B)(Z_t-\mu)=(1-\theta_1 B)a_t, \tag{7.5.4}$$

and go through the iterative stages of the model building until a satisfactory model is obtained. As mentioned earlier, if the model should be indeed a mixed model, then the OLS estimates of the AR parameters based on a misidentified model are inconsistent. Although this may sometimes cause problems, the above procedure of using the residuals to modify models usually works fine.

7.6 EMPIRICAL EXAMPLES FOR SERIES W1–W7

For an illustration, we estimated the AR(3) model identified in Example 6.7 for Series W7—the yearly numbers of lynx pelt sales—and obtained the following result:

$$\underset{(.122)\ (.184)\ (.128)\qquad (.038)}{(1-.97B+.12B^2+.50B^3)(\ln Z_t-.58)=a_t} \tag{7.5.5}$$

and $\hat{\sigma}_a^2=.124$ where the values in the parentheses under estimate refer to the standard errors of those estimates. They are all significant except for ϕ_2, and the model can be refitted with ϕ_2 removed if necessary.

To check model adequacy Table 7.2 gives the residual ACF and PACF and the Q statistics. The residual ACF and PACF are all small and exhibit no patterns. For $K=24$, the Q statistic is $Q=26.7$, which is not significant as $\chi^2_{.05}(21)=32.7$, the chi-square value at the significance level $\alpha=.05$ for the degrees of freedom $=K-m=24-3=21$. Thus, we conclude that the AR(3) model fitting is adequate for the data.

Similarly, we use the nonlinear estimation procedure discussed in Section 7.3 to fit the models identified in Section 6.2 for Series W1 to W6. The results are summarized in Table 7.3.

Diagnostic checking similar to the one for Series W7 was performed for each model fitted in Table 7.3. All models except the AR(2) for Series W2 are adequate. Related tables are not be shown here. Instead, we recall that in Example 6.9, Section 6.4.1, Series W7 was alternatively identified as an

Table 7.2 Residual ACF and PACF for the AR(3) model.

(a) ACF $\hat{\rho}_k$

1–12	−.18	−.17	.27	−.00	−.01	−.15	.14	−.09	−.09	.05	.02	.03
St.E.	.14	.14	.15	.16	.16	.16	.16	.16	.16	.16	.16	.16
Q.	1.8	3.5	7.6	7.6	7.6	9.0	10.3	10.8	11.3	11.5	11.5	11.6
13–24	−.25	.18	.02	−.12	.22	−.05	.04	−.03	−.00	.03	−.09	−.15
St.E.	.16	.17	.17	.17	.18	.18	.18	.18	.18	.18	.18	.18
Q.	16.0	18.4	18.5	19.5	23.3	23.5	23.6	23.7	23.7	23.8	24.5	26.7

(b) PACF $\hat{\phi}_{kk}$

1–12	−.18	−.21	.21	.06	.09	−.21	.10	−.15	.02	−.08	.10	.01
St.E.	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14
13–24	−.20	.07	−.04	.07	.17	.02	.04	−.06	−.06	−.04	.01	−.04
St.E.	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14

ARMA(2, 1) model using the ESACF. The estimation of this model gives

$$\underset{(.063)\ (.058)\qquad (.038)\qquad (.121)}{(1-1.55B+.94B^2)(\ln Z_t-.58)=(1-.59B)a_t} \tag{7.6.2}$$

with all parameters being significant and $\hat{\sigma}_a^2=.116$. The result was also presented in Table 7.3. The residual autocorrelations from this ARMA(2, 1) model shown in Table 7.4 also indicate the adequacy of the model. In fact, both AR(3) and ARMA(2, 1) models fit the data almost equally well. This raises the question of model selection to be discussed next.

7.7 MODEL SELECTION CRITERIA

In time series analysis or more generally in any data analysis, there may be several adequate models that can be used to represent a given data set. Sometimes, the best choice is easy; other times the choice can be very difficult. Thus, numerous criteria for model comparison have been introduced in the literature for model selection. They are different from the model identification methods discussed in Chapter 6. Model identification tools such as ACF, PACF, IACF, and ESACF are used only for identifying adequate models. Residuals from all adequate models are white noise and are, in general, indistinguishable in terms of these functions. For a given data set, when there are multiple adequate models, the selection criterion is normally based on summary statistics from residuals computed from a fitted model or on forecast errors calculated from the out-sample forecasts. The latter is often accomplished by using the first portion of the series for model construction and the remaining portion as a holdout period for forecast evaluation. In this section, we introduce some

Table 7.3 Summary of models fitted to Series W1–W7 (the values in the parentheses under each estimate refer to the standard errors of those estimates).

Series	No. of observations	Fitted models	$\hat{\sigma}_a^2$
W1	45	$(1-.43B)(Z_t-1.79)=a_t$ (.134) (.076)	.21
W2	285	$(1-1.33B+.63B^2)(\sqrt{Z_t}-6.3)=a_t$ (.046) (.046) (.169)	1.637
		$(1-1.13B+.42B^2+.10B^3-.16B^4+.14B^5$ (.057) (.088) (.09) (.091) (.098) $+.08B^6-.25B^7+.3B^8-.35B^9)(\sqrt{Z_t}-6.3)=a_t$ (.104) (.104) (.101) (.063) (.169)	1.2986
		$(1-1.17B+.46B^2+.21B^9)(\sqrt{Z_t}-6.3)=a_t$ (.048) (.049) (.028) (.169)	1.362
W3	82	$(1-.73B)(\sqrt{Z_t}-63.47)=a_t$ (.071) (1.157)	45.772
W4	300	$(1-B)Z_t=(1-.51B)a_t$ (.05)	1397.269
W5	35	$(1-B)Z_t=.56+a_t$ (.25)	2.136
		$(1-.96B)(Z_t-44.26)=a_t$ (.051) (.917)	2.415
W6	115	$(1-B)\ln Z_t=(1-.61B)a_t$ (.076)	.028
W7	55	$(1-.97B+.12B^2+.5B^3)(\ln Z_t-.58)=a_t$ (.122) (.184) (.128) (.038)	.124
		$(1-1.55B+.94B^2)(\ln Z_t-.58)=(1-.59B)a_t$ (.063) (.058) (.038) (.121)	.116

Table 7.4 Residual autocorrelations, $\hat{\rho}_k$, from the ARMA(2,1) model.

1–8	−.103	.090	.178	.057	.095	.030	−.003	−.004
St.E.	(.137)	(.139)	(.140)	(.144)	(.145)	(.146)	(.146)	(.146)
Q.	1.	1.	3.	3.	4.	4.	4.	4.
9–16	.064	−.121	.000	.094	−.146	.100	.052	−.141
St.E.	(.146)	(.146)	(.148)	(.148)	(.149)	(.152)	(.153)	(.154)
Q.	4.	5.	5.	6.	7.	8.	8.	10.

model selection criteria based on residuals. Criteria based on out-sample forecast errors are discussed in the next chapter.

1. Akaike's AIC and BIC Criteria. Assume that a statistical model of M parameters is fitted to data. To assess the quality of the model fitting, Akaike (1973, 1974a) introduced an information criterion. The criterion has been called AIC (Akaike's information criterion) in the literature and is defined as

$$\text{AIC}(M) = -2\ln[\text{maximum likelihood}] + 2M \tag{7.7.1}$$

where M is the number of parameters in the model. For the ARMA model and n effective number of observations, recall from (7.2.10) that the log-likelihood function is

$$\ln L = -\frac{n}{2}\ln 2\pi\sigma_a^2 - \frac{1}{2\sigma_a^2}S(\phi, \mu, \theta). \tag{7.7.2}$$

Maximizing (7.7.2) with respect to ϕ, μ, θ, and σ_a^2, we have, from (7.2.13),

$$\ln L = -\frac{n}{2}\ln\hat{\sigma}_a^2 - \frac{n}{2}(1+\ln 2\pi). \tag{7.7.3}$$

Because the second term in (7.7.3) is a constant, the AIC criterion reduces to

$$\text{AIC}(M) = n\ln\hat{\sigma}_a^2 + 2M. \tag{7.7.4}$$

The optimal order of the model is chosen by the value of M, which is a function of p and q, so that AIC(M) is minimum.

Shibata (1976) has shown that the AIC criterion tends to overestimate the order of the autoregression. More recently, Akaike (1978, 1979) has developed a Bayesian extension of the minimum AIC procedure, called BIC, which takes the following form:

$$\text{BIC}(M) = n\ln\hat{\sigma}_a^2 - (n-M)\ln\left(1-\frac{M}{n}\right) + M\ln n + M_z\ln\left[\left(\frac{\hat{\sigma}_z^2}{\hat{\sigma}_a^2}-1\right)/M\right]$$

where $\hat{\sigma}_a^2$ is the maximum likelihood estimate of σ_a^2, M is the number of parameters, and $\hat{\sigma}_z^2$ is the sample variance of the series. Through a simulation study Akaike (1978) has claimed that the BIC is less likely to overestimate the order of the autoregression. For further discussion on the properties of AIC, see Findley (1985).

2. Schwartz's SBC Criterion. Similar to Akaike's BIC, Schwartz (1978) suggested the following Bayesian criterion of model selection, which has been called SBC (Schwartz's Bayesian Criterion):

$$\text{SBC}(M) = n\ln\hat{\sigma}_a^2 + M\ln n. \tag{7.7.6}$$

Table 7.5 AIC values for Series W7.

p \ q	0	1	2	3	4
0	142.1200	93.6185	70.0174	57.8203	55.3785
1	94.7730	69.9294	62.2191	56.4811	58.4985
2	31.9081	*23.7781	41.5118	49.7238	48.1051
3	24.0529	25.2286	43.7349	31.1166	47.2080
4	25.6708	27.4769	88.6398	60.0407	75.5548

Again in (7.7.6), $\hat{\sigma}_a^2$ is the maximum likelihood estimate of σ_a^2, M is the number of parameters in the model, and n is the effective number of observations that is equivalent to the number of residuals that can be calculated from the series.

3. Parzen's CAT Criterion. Parzen (1977) has suggested the following model selection criterion, which he called CAT (criterion for autoregressive transfer functions):

$$\text{CAT}(p) = \begin{cases} -(1+\frac{1}{n}), & p = 0, \\ \frac{1}{n}\sum_{j=1}^{p}\frac{1}{\hat{\sigma}_j^2} - \frac{1}{\hat{\sigma}_p^2}, & p = 1, 2, 3, \ldots \end{cases} \tag{7.7.7}$$

where $\hat{\sigma}_j^2$ is the unbiased estimate of σ_a^2 when an AR(j) model is fitted to the series, and n is the number of observations. The optimal order of p is chosen so that CAT(p) is minimum.

We have introduced only some commonly used model selection criteria. There are many other criteria introduced in the literature. Interested readers are referred to Stone (1979), Hannan and Quinn (1979), Hannan (1980), and others.

Example 7.3 The AIC criterion has become a standard tool in time series model fitting, and its computation is available in many time series programs. In Table 7.5, we use the SAS/ETS software to compute AIC for Series W7—Canadian lynx pelt sales. From the table, it is clear that the minimum AIC occurs for $p = 2$ and $q = 1$. Hence, based on the AIC criterion, an ARMA(2, 1) model should be selected for the data. Note that a competitive AR(3) model that we fitted earlier to the same data set gives the second smallest value of AIC.

Exercises

7.1 Assume that 100 observations from an ARMA(1, 1) model

$$Z_t - \phi_1 Z_{t-1} = a_t - \theta_1 a_{t-1}$$

gave the following estimates: $\hat{\sigma}_z^2 = 10$, $\hat{\rho}_1 = .523$, and $\hat{\rho}_2 = .418$. Find initial estimates for ϕ_1, θ_1, and σ_a^2.

7.2 Assume that 100 observations from an AR(2) model

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$$

gave the following sample ACF: $\hat{\rho}_1 = .8$, $\hat{\rho}_2 = .5$, and $\hat{\rho}_3 = .4$. Estimate ϕ_1 and ϕ_2.

7.3 Given the set of observations 2.2, 4.5, 2.5, 2.3, 1.1, 3.0, 2.1, and 1.0, calculate the conditional sum of squares $S(\theta_1, \theta_2)$ for the MA(2) process with $\theta_1 = -.5$ and $\theta_2 = .2$.

7.4 Given the set of observations 6, 2, 4, 5, 3, 4, 2, and 1, illustrate how to calculate the conditional sum of squares function $S(\phi_1, \theta_1)$ for the ARMA(1, 1) model.

7.5 Consider the following observations from an MA(1) model with $\theta = 4$:

t	Z_t	$W_t = (1-B)Z_t$
0	59	
1	62	3
2	58	−4
3	63	5
4	79	16
5	90	11
6	88	−2

(a) Calculate the conditional sum of squares (with $a_0 = 0$).
(b) Calculate the unconditional sum of squares using the backcasting method as shown in Table 7.1.

7.6 Simulate 100 observations from an ARMA(1, 1) model.
(a) Fit the simulated series with an AR(1) or an MA(1) model. Carry out diagnostic checking, and modify your fitted model from the result of residual analysis.
(b) Estimate the parameters of your modified model, and compare with the true parameter values of the model.

7.7 A summary of models fitted for the series W1 to W7 is given in Table 7.3. Perform residual analysis and model checking for each of the fitted models.

7.8 Use AIC to find a model for each of the series W1 to W7, and compare it with the fitted model given in Table 7.3.

7.9 Suppose $(1-\phi B)Z_t = (1-\theta B)a_t$ is a tentatively entertained model for a process. Given

t	1	2	3	4	5	6	7	8	9	10	11	12
Z_t	−3.1	−.8	1.2	.6	2.8	−.9	.3	−1.4	−2.5	−1.1	.9	1.4

calculate the unconditional sum of squares for $\phi = .4$ and $\theta = .8$.

7.10 Consider the AR(1) model

$$(1-\phi B)(Z_t - \mu) = a_t.$$

(a) For $\mu = 0$, find the maximum likelihood estimator for ϕ and its associated variance.
(b) Find the maximum likelihood estimators for ϕ and μ when $\mu \neq 0$.
(c) Discuss the relationship between the ordinary least square estimator and the maximum likelihood estimator for ϕ in the above model.

8 SEASONAL TIME SERIES MODELS

Because of their common occurrence in our daily activities, we devote a separate chapter to seasonal time series. After a brief introduction of some basic concepts and conventional methods, we extend the autoregressive integrated moving average models to represent seasonal series. Detailed examples are given to illustrate the methods.

8.1 INTRODUCTION

Many business and economic time series contain a seasonal phenomenon that repeats itself after a regular period of time. The smallest time period for this repetitive phenomenon is called the seasonal period. For example, the quarterly series of ice cream sales is high each summer, and the series repeats this phenomenon each year, giving a seasonal period of 4. Similarly, monthly auto sales and earnings tend to decrease during August and September every year because of the changeover to new models, and the monthly sales of toys rise every year in the month of December. The seasonal period in these later cases is 12. Seasonal phenomena may stem from factors such as weather, which affects many business and economic activities like tourism and home building; custom events like Christmas, which is closely related to sales such as jewelry, toys, cards, and stamps; and graduation ceremonies in the summer months, which are directly related to the labor force status in these months.

As an illustration, Figure 8.1 shows the U.S. monthly employment figures (in thousands) for males aged between 16 and 19 years from 1971 to 1981. The seasonal nature of the series is apparent. The numbers increase dramatically in the summer months, with peaks occurring in the month of June when schools are not in session, and decrease in the fall months when schools reopen. The phenomenon repeats itself every 12 months, and thus the seasonal period is 12.

More generally, suppose the series $\{Z_t\}$ is seasonal with seasonal period s. To analyze the data, it is helpful to arrange the series in a two-dimensional