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:

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

$$\vdots$$

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

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,

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

These estimators are usually called Yule-Walker estimators.

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

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

and obtain the moment estimator for σ_a^2 as

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

Example 7.1 For the AR(1) model,

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

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

$$\hat{\phi}_1 = \hat{\rho}_1. \qquad (7.1.7)$$

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

$$\hat{\mu} = \bar{Z} \qquad (7.1.8)$$

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Rewriting (7.2.1) as

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

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

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

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

where

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

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

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

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:

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

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

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

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

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

The minimum residual (error) sum of squares is

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

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

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

or

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

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

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

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

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

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

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

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):

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

where $\mathbf{X}_t = (X_{t1}, X_{t2}, \ldots, X_{tp})$ is a set of independent variables corresponding to the observations, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$ 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 $\mathbf{f}(\boldsymbol{\alpha}) = [f(\mathbf{X}_1, \boldsymbol{\alpha}), f(\mathbf{X}_2, \boldsymbol{\alpha}), \ldots, f(\mathbf{X}_n, \boldsymbol{\alpha})]'$. 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 $\bar{X}_{\tilde{\alpha}} = \{X_{tj}\}$ is the $n \times p$ matrix of the partial derivatives at $\tilde{\alpha}$ in the above linear approximation. That is,

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

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 are 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 $\bar{X}_{\hat{\alpha}}$ 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:

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

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

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

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

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

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

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

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:

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

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):

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

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