CHAPTER 6

MODEL SPECIFICATION

We have developed a large class of parametric models for both stationary and nonstationary time series—the ARIMA models. We now begin our study and implementation of *statistical inference* for such models. The subjects of the next three chapters, respectively, are:

1. how to choose appropriate values for p, d, and q for a given series;
2. how to estimate the parameters of a specific ARIMA(p,d,q) model;
3. how to check on the appropriateness of the fitted model and improve it if needed.

Our overall strategy will first be to decide on reasonable—but tentative—values for p, d, and q. Having done so, we shall estimate the ϕ's, θ's, and σ_e for that model in the most efficient way. Finally, we shall look critically at the fitted model thus obtained to check its adequacy, in much the same way that we did in Section 3.6 on page 42. If the model appears inadequate in some way, we consider the nature of the inadequacy to help us select another model. We proceed to estimate that new model and check it for adequacy.

With a few iterations of this model-building strategy, we hope to arrive at the best possible model for a given series. The book by George E. P. Box and G. M. Jenkins (1976) so popularized this technique that many authors call the procedure the "Box-Jenkins method." We begin by continuing our investigation of the properties of the sample autocorrelation function.

6.1 Properties of the Sample Autocorrelation Function

Recall from page 46 the definition of the sample or estimated autocorrelation function. For the observed series $Y_1, Y_2, \ldots, Y_n$, we have

$$r_k = \frac{\sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2} \qquad \text{for } k = 1, 2, \ldots \tag{6.1.1}$$

Our goal is to recognize, to the extent possible, patterns in r_k that are characteristic of the known patterns in ρ_k for common ARMA models. For example, we know that $\rho_k = 0$ for $k > q$ in an MA(q) model. However, as the r_k are only estimates of the ρ_k, we

need to investigate their sampling properties to facilitate the comparison of estimated correlations with theoretical correlations.

From the definition of r_k, a ratio of quadratic functions of possibly dependent variables, it should be apparent that the sampling properties of r_k will *not* be obtained easily. Even the expected value of r_k is difficult to determine—recall that the expected value of a ratio is *not* the ratio of the respective expected values. We shall be content to accept a general large-sample result and consider its implications in special cases. Bartlett (1946) carried out the original work. We shall take a more general result from Anderson (1971). A recent discussion of these results may be found in Shumway and Stoffer (2006, p. 519).

We suppose that

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j e_{t-j}$$

where the e_t are independent and identically distributed with zero means and finite, nonzero, common variances. We assume further that

$$\sum_{j=0}^{\infty} |\psi_j| < \infty \qquad \text{and} \qquad \sum_{j=0}^{\infty} j\psi_j^2 < \infty$$

(These will be satisfied by any stationary ARMA model.)

Then, for any fixed m, the joint distribution of

$$\sqrt{n}(r_1 - \rho_1),\ \sqrt{n}(r_2 - \rho_2),\ \ldots,\ \sqrt{n}(r_m - \rho_m)$$

approaches, as $n \to \infty$, a joint normal distribution with zero means, variances c_{jj}, and covariances c_{ij},where

$$c_{ij} = \sum_{k=-\infty}^{\infty} (\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i} + 2\rho_i\rho_j\rho_k^2) \qquad (6.1.2)$$

For large n, we would say that r_k is approximately normally distributed with mean ρ_k and variance c_{kk}/n. Furthermore, $Corr(r_k, r_j) \approx c_{kj}/\sqrt{c_{kk}c_{jj}}$. Notice that the approximate variance of r_k is inversely proportional to the sample size, but $Corr(r_k, r_j)$ is approximately *constant* for large n.

Since Equation (6.1.2) is clearly difficult to interpret in its present generality, we shall consider some important special cases and simplifications. Suppose first that $\{Y_t\}$ is white noise. Then Equation (6.1.2) reduces considerably, and we obtain

$$Var(r_k) \approx \frac{1}{n} \text{ and } Corr(r_k, r_j) \approx 0 \text{ for } k \neq j \qquad (6.1.3)$$

Next suppose that $\{Y_t\}$ is generated by an AR(1) process with $\rho_k = \phi^k$ for $k > 0$. Then, after considerable algebra and summing several geometric series, Equation (6.1.2) with $i = j$ yields

$$Var(r_k) \approx \frac{1}{n}\left[\frac{(1+\phi^2)(1-\phi^{2k})}{1-\phi^2} - 2k\phi^{2k}\right] \qquad (6.1.4)$$

In particular,

$$Var(r_1) \approx \frac{1-\phi^2}{n} \tag{6.1.5}$$

Notice that the closer ϕ is to ± 1, the more precise our estimate of ρ_1 $(= \phi)$ becomes.

For large lags, the terms in Equation (6.1.4) involving ϕ^k may be ignored, and we have

$$Var(r_k) \approx \frac{1}{n}\left[\frac{1+\phi^2}{1-\phi^2}\right] \text{ for large } k \tag{6.1.6}$$

Notice that here, in contrast to Equation (6.1.5), values of ϕ close to ± 1 imply large variances for r_k. Thus we should not expect nearly as precise estimates of $\rho_k = \phi^k \approx 0$ for large k as we do of $\rho_k = \phi^k$ for small k.

For the AR(1) model, Equation (6.1.2) can also be simplified (after much algebra) for general $0 < i < j$ as

$$c_{ij} = \frac{(\phi^{j-i} - \phi^{j+i})(1+\phi^2)}{1-\phi^2} + (j-i)\phi^{j-i} - (j+i)\phi^{j+i} \tag{6.1.7}$$

In particular, we find

$$Corr(r_1, r_2) \approx 2\phi\sqrt{\frac{1-\phi^2}{1+2\phi^2-3\phi^4}} \tag{6.1.8}$$

Based on Equations (6.1.4) through (6.1.8), Exhibit 6.1 gives approximate standard deviations and correlations for several lags and a few values of ϕ in AR(1) models.

Exhibit 6.1 Large Sample Results for Selected r_k from an AR(1) Model

ϕ	$\sqrt{Var(r_1)}$	$\sqrt{Var(r_2)}$	$Corr(r_1, r_2)$	$\sqrt{Var(r_{10})}$
± 0.9	$0.44/\sqrt{n}$	$0.807/\sqrt{n}$	± 0.97	$2.44/\sqrt{n}$
± 0.7	$0.71/\sqrt{n}$	$1.12/\sqrt{n}$	± 0.89	$1.70/\sqrt{n}$
± 0.4	$0.92/\sqrt{n}$	$1.11/\sqrt{n}$	± 0.66	$1.18/\sqrt{n}$
± 0.2	$0.98/\sqrt{n}$	$1.04/\sqrt{n}$	± 0.38	$1.04/\sqrt{n}$

For the MA(1) case, Equation (6.1.2) simplifies as follows:

$$c_{11} = 1 - 3\rho_1^2 + 4\rho_1^4 \quad \text{and} \quad c_{kk} = 1 + 2\rho_1^2 \text{ for } k > 1 \tag{6.1.9}$$

Furthermore,

$$c_{12} = 2\rho_1(1-\rho_1^2) \tag{6.1.10}$$

Based on these expressions, Exhibit 6.2 lists large-sample standard deviations and correlations for the sample autocorrelations for several lags and several θ-values. Notice again that the sample autocorrelations can be highly correlated and that the standard deviation of r_k is larger for $k > 1$ than for $k = 1$.

Exhibit 6.2 Large-Sample Results for Selected r_k from an MA(1) Model

θ	$\sqrt{Var(r_1)}$	$\sqrt{Var(r_k)}$ for $k > 1$	$Corr(r_1, r_2)$
±0.9	$0.71/\sqrt{n}$	$1.22/\sqrt{n}$	∓0.86
±0.7	$0.73/\sqrt{n}$	$1.20/\sqrt{n}$	∓0.84
±0.5	$0.79/\sqrt{n}$	$1.15/\sqrt{n}$	∓0.74
±0.4	$0.89/\sqrt{n}$	$1.11/\sqrt{n}$	∓0.53

For a general MA(q) process and $i = j = k$, Equation (6.1.2) reduces to

$$c_{kk} = 1 + 2\sum_{j=1}^{q} \rho_j^2 \text{ for } k > q$$

so that

$$Var(r_k) = \frac{1}{n}\left[1 + 2\sum_{j=1}^{q} \rho_j^2\right] \text{ for } k > q \tag{6.1.11}$$

For an observed time series, we can replace ρ's by r's, take the square root, and obtain an estimated standard deviation of r_k, that is, the **standard error** of r_k for large lags. A test of the hypothesis that the series is MA(q) could be carried out by comparing r_k to plus and minus two standard errors. We would reject the null hypothesis if and only if r_k lies outside these bounds. In general, we should not expect the sample autocorrelation to mimic the true autocorrelation in great detail. Thus, we should not be surprised to see ripples or "trends" in r_k that have no counterparts in the ρ_k.

6.2 The Partial and Extended Autocorrelation Functions

Since for MA(q) models the autocorrelation function is zero for lags beyond q, the sample autocorrelation is a good indicator of the order of the process. However, the autocorrelations of an AR(p) model do not become zero after a certain number of lags—they die off rather than cut off. So a different function is needed to help determine the order of autoregressive models. Such a function may be defined as the correlation between Y_t and Y_{t-k} *after removing the effect of the intervening variables* $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_{t-k+1}$. This coefficient is called the partial autocorrelation at lag k and will be denoted by ϕ_{kk}. (The reason for the seemingly redundant double subscript on ϕ_{kk} will become apparent later on in this section.)

There are several ways to make this definition precise. If $\{Y_t\}$ is a normally distributed time series, we can let

$$\phi_{kk} = Corr(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k+1}) \tag{6.2.1}$$

That is, ϕ_{kk} is the correlation in the bivariate distribution of Y_t and Y_{t-k} conditional on $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k+1}$.

An alternative approach, not based on normality, can be developed in the following way. Consider predicting Y_t based on a linear function of the intervening variables Y_{t-1}, $Y_{t-2},\ldots, Y_{t-k+1}$, say, $\beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_{k-1} Y_{t-k+1}$, with the β's chosen to minimize the mean square error of prediction. If we assume that the β's have been so chosen and then think backward in time, it follows from stationarity that the best "predictor" of Y_{t-k} based on the same $Y_{t-1}, Y_{t-2},\ldots, Y_{t-k+1}$ will be $\beta_1 Y_{t-k+1} + \beta_2 Y_{t-k+2} + \cdots + \beta_{k-1} Y_{t-1}$. The **partial autocorrelation function** at lag k is then defined to be the correlation between the prediction errors; that is,

$$\begin{aligned}\phi_{kk} = Corr(&Y_t - \beta_1 Y_{t-1} - \beta_2 Y_{t-2} - \cdots - \beta_{k-1} Y_{t-2},\\ &Y_{t-k} - \beta_1 Y_{t-k+1} - \beta_2 Y_{t-k+2} - \cdots - \beta_{k-1} Y_{t-1})\end{aligned} \tag{6.2.2}$$

(For normally distributed series, it can be shown that the two definitions coincide.) By convention, we take $\phi_{11} = 1$.

As an example, consider ϕ_{22}. It is shown in Appendix F on page 218 that the best linear prediction of Y_t based on Y_{t-1} alone is just $\rho_1 Y_{t-1}$. Thus, according to Equation (6.2.2), we will obtain ϕ_{22} by computing

$$Cov(Y_t - \rho_1 Y_{t-1}, Y_{t-2} - \rho_1 Y_{t-1}) = \gamma_0(\rho_2 - \rho_1^2 - \rho_1^2 + \rho_1^2) = \gamma_0(\rho_2 - \rho_1^2)$$

Since

$$\begin{aligned}Var(Y_t - \rho_1 Y_{t-1}) &= Var(Y_{t-2} - \rho_1 Y_{t-1})\\ &= \gamma_0(1 + \rho_1^2 - 2\rho_1^2)\\ &= \gamma_0(1 - \rho_1^2)\end{aligned}$$

we have that, for any stationary process, the lag 2 partial autocorrelation can be expressed as

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \tag{6.2.3}$$

Consider now an AR(1) model. Recall that $\rho_k = \phi^k$ so that

$$\phi_{22} = \frac{\phi^2 - \phi^2}{1 - \phi^2} = 0$$

We shall soon see that for the AR(1) case, $\phi_{kk} = 0$ for all $k > 1$. Thus the partial autocorrelation is nonzero for lag 1, the order of the AR(1) process, but is zero for all lags greater than 1. We shall show this to be generally the case for AR(p) models. Sometimes we say that the partial autocorrelation function for an AR(p) process *cuts off* after the lag exceeds the order of the process.

Consider a general AR(p) case. It will be shown in Chapter 9 that the best linear predictor of Y_t based on a linear function of the variables $Y_{t-1}, Y_{t-2},\ldots, Y_p,\ldots, Y_{t-k+1}$ for $k > p$ is $\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p}$. Also, the best linear predictor of Y_{t-k} is some function of $Y_{t-1}, Y_{t-2},\ldots, Y_p,\ldots, Y_{t-k+1}$, call it $h(Y_{t-1}, Y_{t-2},\ldots, Y_p,\ldots, Y_{t-k+1})$. So the covariance between the two prediction errors is

$$
\begin{aligned}
&Cov(Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \cdots - \phi_p Y_{t-p}, \\
&\qquad\qquad Y_{t-k} - h(Y_{t-k+1}, Y_{t-k+2}, \ldots, Y_{t-1})) \\
&= Cov(e_t, Y_{t-k} - h(Y_{t-k+1}, Y_{t-k+2}, \ldots, Y_{t-1})) \\
&= 0 \quad \text{since } e_t \text{ is independent of } Y_{t-k}, Y_{t-k+1}, Y_{t-k+2}, \ldots, Y_{t-1}
\end{aligned}
$$

Thus we have established the key fact that, for an AR(p) model,

$$\phi_{kk} = 0 \quad \text{for} \quad k > p \tag{6.2.4}$$

For an MA(1) model, Equation (6.2.3) quickly yields

$$\phi_{22} = \frac{-\theta^2}{1 + \theta^2 + \theta^4} \tag{6.2.5}$$

Furthermore, for the MA(1) case, it may be shown that

$$\phi_{kk} = -\frac{\theta^k(1 - \theta^2)}{1 - \theta^{2(k+1)}} \quad \text{for} \quad k \geq 1 \tag{6.2.6}$$

Notice that the partial autocorrelation of an MA(1) model never equals zero but essentially decays to zero exponentially fast as the lag increases—rather like the autocorrelation function of the AR(1) process. More generally, it can be shown that the partial autocorrelation of an MA(q) model behaves very much like the *autocorrelation* of an AR(q) model.

A general method for finding the partial autocorrelation function for any stationary process with autocorrelation function ρ_k is as follows (see Anderson 1971, pp. 187–188, for example). For a given lag k, it can be shown that the ϕ_{kk} satisfy the Yule-Walker equations (which first appeared in Chapter 4 on page 79):

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \phi_{k3}\rho_{j-3} + \cdots + \phi_{kk}\rho_{j-k} \quad \text{for } j = 1, 2, \ldots, k \tag{6.2.7}$$

More explicitly, we can write these k linear equations as

$$
\left.
\begin{aligned}
\phi_{k1} + \quad \rho_1\phi_{k2} + \quad \rho_2\phi_{k3} + \cdots + \rho_{k-1}\phi_{kk} &= \rho_1 \\
\rho_1\phi_{k1} + \quad \phi_{k2} + \quad \rho_1\phi_{k3} + \cdots + \rho_{k-2}\phi_{kk} &= \rho_2 \\
\vdots \qquad\qquad & \\
\rho_{k-1}\phi_{k1} + \rho_{k-2}\phi_{k2} + \rho_{k-3}\phi_{k3} + \cdots + \phi_{kk} &= \rho_k
\end{aligned}
\right\} \tag{6.2.8}
$$

Here we are treating $\rho_1, \rho_2, \ldots, \rho_k$ as given and wish to solve for $\phi_{k1}, \phi_{k2}, \ldots, \phi_{kk}$ (discarding all but ϕ_{kk}).

These equations yield ϕ_{kk} for any stationary process. However, if the process is in fact AR(p), then since for $k = p$ Equations (6.2.8) are just the Yule-Walker equations (page 79), which the AR(p) model is known to satisfy, we must have $\phi_{pp} = \phi_p$. In addition, as we have already seen by an alternative derivation, $\phi_{kk} = 0$ for $k > p$. Thus the partial autocorrelation effectively displays the correct order p of an autoregressive process as the highest lag k before ϕ_{kk} becomes zero.

The Sample Partial Autocorrelation Function

For an observed time series, we need to be able to estimate the partial autocorrelation function at a variety of lags. Given the relationships in Equations (6.2.8), an obvious method is to estimate the ρ's with sample autocorrelations, the corresponding r's, and then solve the resulting linear equations for $k = 1, 2, 3, \ldots$ to get estimates of ϕ_{kk}. We call the estimated function the **sample partial autocorrelation function** (sample PACF) and denote it by $\hat{\phi}_{kk}$.

Levinson (1947) and Durbin (1960) gave an efficient method for obtaining the solutions to Equations (6.2.8) for either theoretical or sample partial autocorrelations. They showed independently that Equations (6.2.8) can be solved recursively as follows:

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\,\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\,\rho_j} \tag{6.2.9}$$

where

$$\phi_{k,j} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j} \quad \text{for } j = 1, 2, \ldots, k-1$$

For example, using $\phi_{11} = \rho_1$ to get started, we have

$$\phi_{22} = \frac{\rho_2 - \phi_{11}\rho_1}{1 - \phi_{11}\rho_1} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

(as before) with $\phi_{21} = \phi_{11} - \phi_{22}\phi_{11}$, which is needed for the next step. Then

$$\phi_{33} = \frac{\rho_3 - \phi_{21}\rho_2 - \phi_{22}\rho_1}{1 - \phi_{21}\rho_1 - \phi_{22}\rho_2}$$

We may thus calculate numerically as many values for ϕ_{kk} as desired. As stated, these recursive equations give us the theoretical partial autocorrelations, but by replacing ρ's with r's, we obtain the estimated or sample partial autocorrelations.

To assess the possible magnitude of the sample partial autocorrelations, Quenoulle (1949) has shown that, under the hypothesis that an AR(p) model is correct, the sample partial autocorrelations at lags greater than p are approximately normally distributed with zero means and variances $1/n$. Thus, for $k > p$, $\pm 2/\sqrt{n}$ can be used as critical limits on $\hat{\phi}_{kk}$ to test the null hypothesis that an AR(p) model is correct.

Mixed Models and the Extended Autocorrelation Function

Exhibit 6.3 summarizes the behavior of the autocorrelation and partial autocorrelation functions that is useful in specifying models.

Exhibit 6.3 General Behavior of the ACF and PACF for ARMA Models

	AR(p)	MA(q)	ARMA(p,q), $p>0$, and $q>0$
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

The Extended Autocorrelation Function

The sample ACF and PACF provide effective tools for identifying pure AR(p) or MA(q) models. However, for a mixed ARMA model, its theoretical ACF and PACF have infinitely many nonzero values, making it difficult to identify mixed models from the sample ACF and PACF. Many graphical tools have been proposed to make it easier to identify the ARMA orders, for example, the corner method (Becuin et al., 1980), the extended autocorrelation (EACF) method (Tsay and Tiao, 1984), and the smallest canonical correlation (SCAN) method (Tsay and Tiao, 1985), among others. We shall outline the EACF method, which seems to have good sampling properties for moderately large sample sizes according to a comparative simulation study done by W. S. Chan (1999).

The EACF method uses the fact that if the AR part of a mixed ARMA model is known, "filtering out" the autoregression from the observed time series results in a pure MA process that enjoys the cutoff property in its ACF. The AR coefficients may be estimated by a finite sequence of regressions. We illustrate the procedure for the case where the true model is an ARMA(1,1) model:

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}$$

In this case, a simple linear regression of Y_t on Y_{t-1} results in an inconsistent estimator of ϕ, even with infinitely many data. Indeed, the theoretical regression coefficient equals $\rho_1 = (\phi - \theta)(1 - \phi\theta)/(1 - 2\phi\theta + \theta^2)$, not ϕ. But the residuals from this regression do contain information about the error process $\{e_t\}$. A second multiple regression is performed that consists of regressing Y_t on Y_{t-1} and on the lag 1 of the residuals from the first regression. The coefficient of Y_{t-1} in the second regression, denoted by $\tilde{\phi}$, turns out to be a consistent estimator of ϕ. Define $W_t = Y_t - \tilde{\phi} Y_{t-1}$, which is then approximately an MA(1) process. For an ARMA(1,2) model, a third regression that regresses Y_t on its lag 1, the lag 1 of the residuals from the second regression, and the lag 2 of the residuals from the first regression leads to the coefficient of Y_{t-1} being a consistent estimator of ϕ. Similarly, the AR coefficients of an ARMA(p,q) model can be consistently estimated via a sequence of q regressions.

As the AR and MA orders are unknown, an iterative procedure is required. Let

$$W_{t,k,j} = Y_t - \tilde{\phi}_1 Y_{t-1} - \cdots - \tilde{\phi}_k Y_{t-k} \tag{6.2.10}$$

be the autoregressive residuals defined with the AR coefficients estimated iteratively assuming the AR order is k and the MA order is j. The sample autocorrelations of $W_{t,k,j}$ are referred to as the extended sample autocorrelations. For $k = p$ and $j \geq q$, $\{W_{t,k,j}\}$ is approximately an MA(q) model, so that its theoretical autocorrelations of lag $q + 1$ or

higher are equal to zero. For $k > p$, an overfitting problem occurs, and this increases the MA order for the W process by the minimum of $k - p$ and $j - q$. Tsay and Tiao (1984) suggested summarizing the information in the sample EACF by a table with the element in the kth row and jth column equal to the symbol X if the lag $j + 1$ sample correlation of $W_{t,k,j}$ is significantly different from 0 (that is, if its magnitude is greater than $1.96/\sqrt{n-j-k}$ since the sample autocorrelation is asymptotically $N(0,1/(n-k-j))$ if the W's are approximately an MA(j) process) and 0 otherwise. In such a table, an MA(p,q) process will have a theoretical pattern of a triangle of zeroes, with the upper left-hand vertex corresponding to the ARMA orders. Exhibit 6.4 displays the schematic pattern for an ARMA(1,1) model. The upper left-hand vertex of the triangle of zeros is marked with the symbol 0^* and is located in the $p = 1$ row and $q = 1$ column—an indication of an ARMA(1,1) model.

Exhibit 6.4 Theoretical Extended ACF (EACF) for an ARMA(1,1) Model

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	x	0*	0	0	0	0	0	0	0	0	0	0	0	0
2	x	x	0	0	0	0	0	0	0	0	0	0	0	0
3	x	x	x	0	0	0	0	0	0	0	0	0	0	0
4	x	x	x	x	0	0	0	0	0	0	0	0	0	0
5	x	x	x	x	x	0	0	0	0	0	0	0	0	0
6	x	x	x	x	x	x	0	0	0	0	0	0	0	0
7	x	x	x	x	x	x	x	0	0	0	0	0	0	0

Of course, the *sample* EACF will never be this clear-cut. Displays like Exhibit 6.4 will contain 8×14 = 112 different estimated correlations, and some will be statistically significantly different from zero by chance (see Exhibit 6.17 on page 124, for an example). We will illustrate the use of the EACF in the next two sections and throughout the remainder of the book.

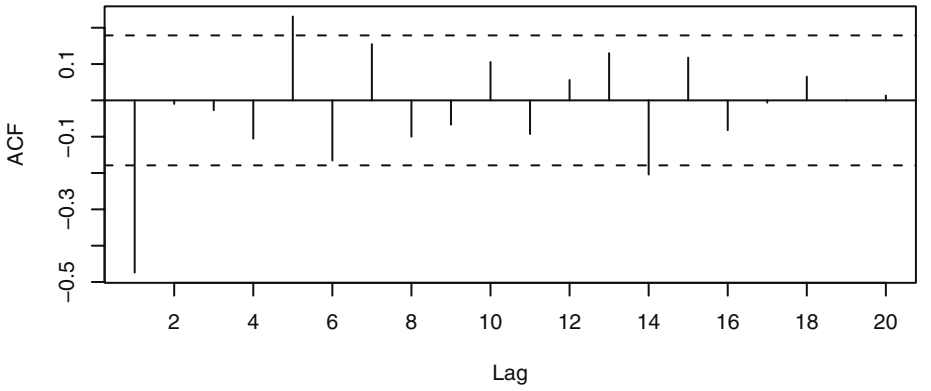
6.3 Specification of Some Simulated Time Series

To illustrate the theory of Sections 6.1 and 6.2, we shall consider the sample autocorrelation and sample partial correlation of some simulated time series.

Exhibit 6.5 displays a graph of the sample autocorrelation out to lag 20 for the simulated time series that we first saw in Exhibit 4.5 on page 61. This series, of length 120, was generated from an MA(1) model with $\theta = 0.9$. From Exhibit 4.1 on page 58, the theoretical autocorrelation at lag 1 is –0.4972. The estimated or sample value shown at lag 1 on the graph is –0.474. Using Exhibit 6.2 on page 112, the approximate standard error

of this estimate is $0.71/\sqrt{n} = 0.71/\sqrt{120} = 0.065$, so the estimate is well within two standard errors of the true value.

Exhibit 6.5 Sample Autocorrelation of an MA(1) Process with $\theta = 0.9$

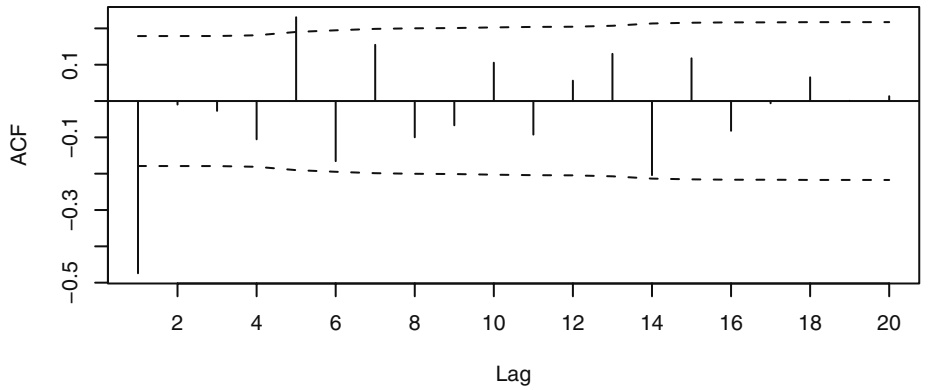



```
> data(ma1.1.s)
> win.graph(width=4.875,height=3,pointsize=8)
> acf(ma1.1.s,xaxp=c(0,20,10))
```

The dashed horizontal lines in Exhibit 6.5, plotted at $\pm 2/\sqrt{n} = \pm 0.1826$, are intended to give critical values for testing whether or not the autocorrelation coefficients are significantly different from zero. These limits are based on the approximate large sample standard error that applies to a white noise process, namely $1/\sqrt{n}$. Notice that the sample ACF values exceed these rough critical values at lags 1, 5, and 14. Of course, the true autocorrelations at lags 5 and 14 are both zero.

Exhibit 6.6 displays the same sample ACF but with critical bounds based on plus and minus two of the more complex standard errors implied by Equation (6.1.11) on page 112. In using Equation (6.1.11), we replace ρ's by r's, let q equal 1, 2, 3,… successively, and take the square root to obtain these standard errors.

Exhibit 6.6 Alternative Bounds for the Sample ACF for the MA(1) Process

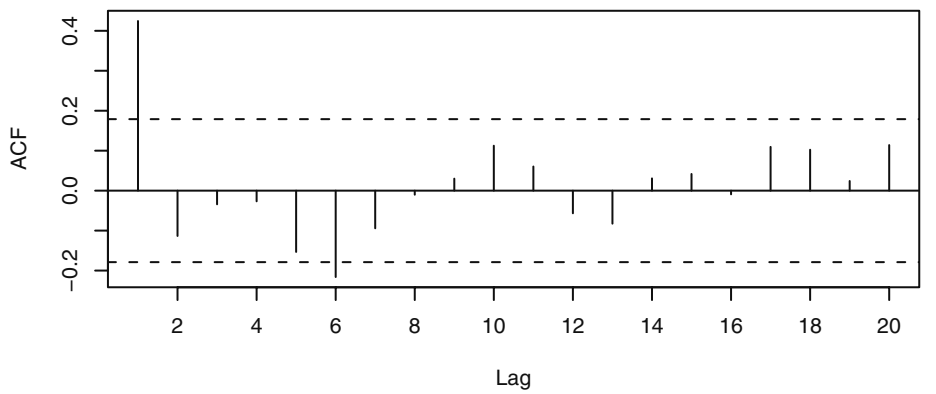



```
> acf(ma1.1.s,ci.type='ma',xaxp=c(0,20,10))
```

Now the sample ACF value at lag 14 is insignificant and the one at lag 5 is just barely significant. The lag 1 autocorrelation is still highly significant, and the information given in these two plots taken together leads us to consider an MA(1) model for this series. Remember that the model is tentative at this point and we would certainly want to consider other "nearby" alternative models when we carry out model diagnostics.

As a second example, Exhibit 6.7 shows the sample ACF for the series shown in Exhibit 4.2 on page 59, generated by an MA(1) model with $\theta = -0.9$. The critical values based on the very approximate standard errors point to an MA(1) model for this series also.

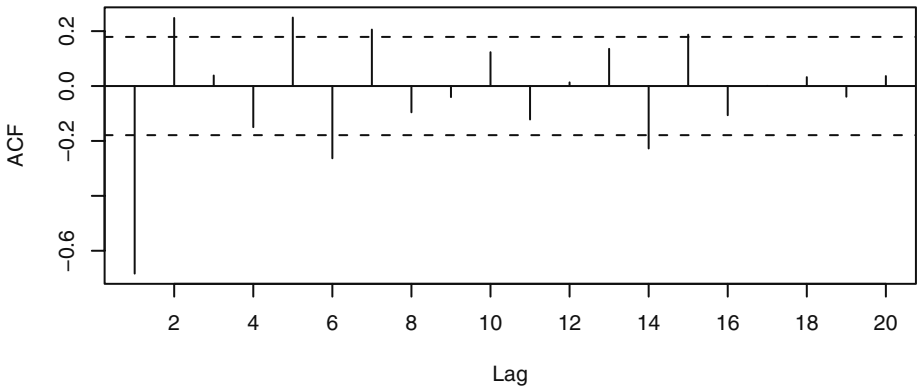
Exhibit 6.7 Sample Autocorrelation for an MA(1) Process with $\theta = -0.9$




```
> data(ma1.2.s); acf(ma1.2.s,xaxp=c(0,20,10))
```

For our third example, we use the data shown in Exhibit 4.8 on page 63, which were simulated from an MA(2) model with $\theta_1 = 1$ and $\theta_2 = -0.6$. The sample ACF displays significance at lags 1, 2, 5, 6, 7, and 14 when we use the simple standard error bounds.

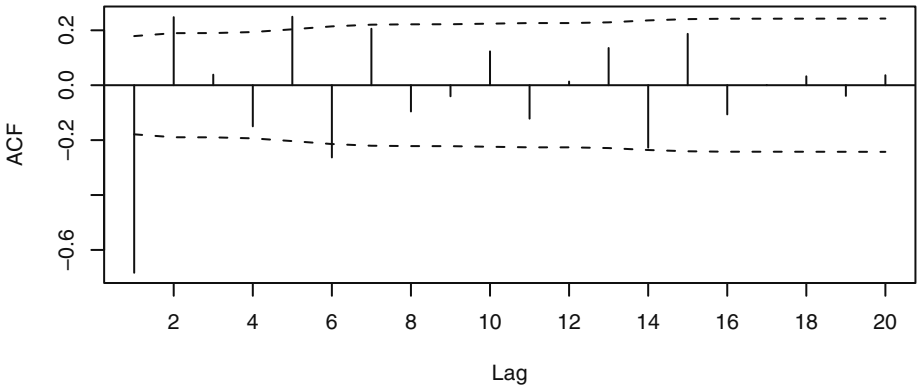
Exhibit 6.8 Sample ACF for an MA(2) Process with $\theta_1 = 1$ and $\theta_2 = -0.6$




```
> data(ma2.s); acf(ma2.s,xaxp=c(0,20,10))
```

Exhibit 6.9 displays the sample ACF with the more sophisticated standard error bounds. Now the lag 2 ACF is no longer significant, and it appears that an MA(1) may be applicable. We will have to wait until we get further along in the model-building process to see that the MA(2) model—the correct one—is the most appropriate model for these data.

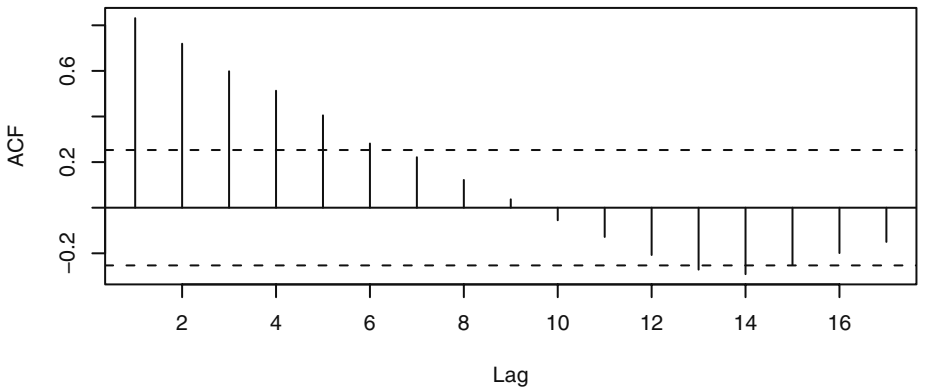
Exhibit 6.9 Alternative Bounds for the Sample ACF for the MA(2) Process




```
> acf(ma2.s,ci.type='ma',xaxp=c(0,20,10))
```

How do these techniques work for autoregressive models? Exhibit 6.10 gives the sample ACF for the simulated AR(1) process we saw in Exhibit 4.13 on page 68. The positive sample ACF values at lags 1, 2, and 3 reflect the strength of the lagged relationships that we saw earlier in Exhibits 4.14, 4.15, and 4.16. However, notice that the sample ACF decreases more linearly than exponentially as theory suggests. Also contrary to theory, the sample ACF goes negative at lag 10 and remains so for many lags.

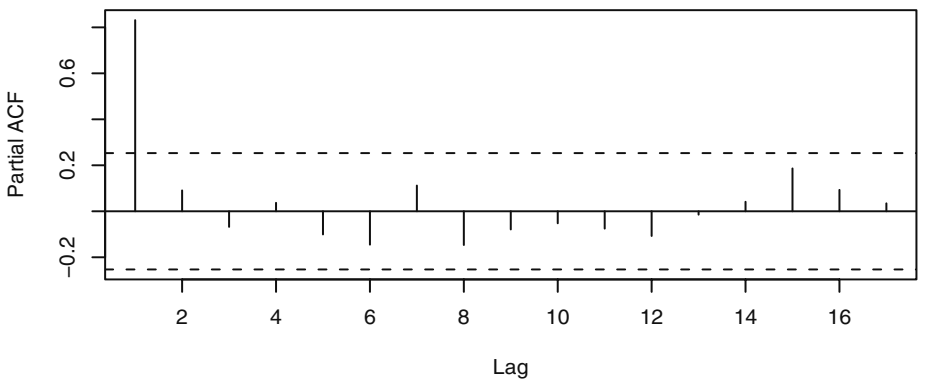
Exhibit 6.10 Sample ACF for an AR(1) Process with $\phi = 0.9$




```
> data(ar1.s); acf(ar1.s,xaxp=c(0,20,10))
```

The sample partial autocorrelation (PACF) shown in Exhibit 6.11, gives a much clearer picture about the nature of the generating model. Based on this graph, we would certainly entertain an AR(1) model for this time series.

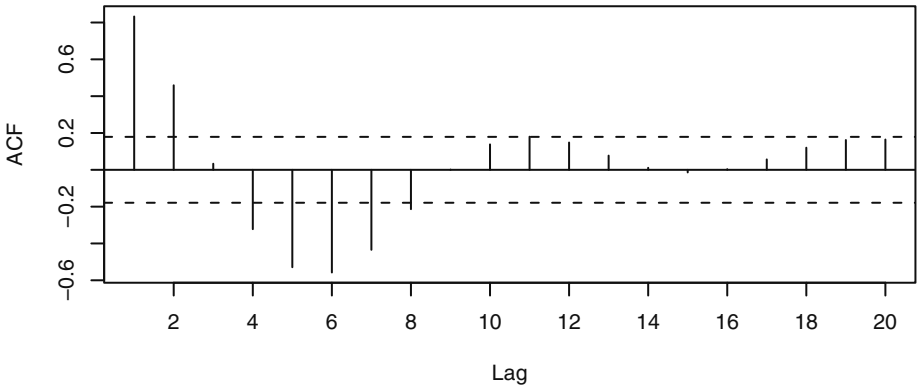
Exhibit 6.11 Sample Partial ACF for an AR(1) Process with $\phi = 0.9$



```
> pacf(ar1.s,xaxp=c(0,20,10))
```

Exhibit 6.12 displays the sample ACF for our AR(2) time series. The time series plot for this series was shown in Exhibit 4.19 on page 74. The sample ACF does look somewhat like the damped wave that Equation (4.3.17) on page 73, and Exhibit 4.18 suggest. However, the sample ACF does not damp down nearly as quickly as theory predicts.

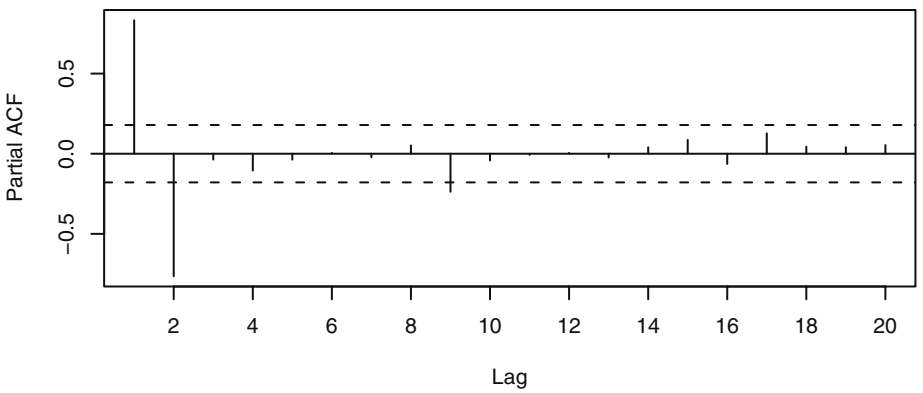
Exhibit 6.12 Sample ACF for an AR(2) Process with $\phi_1 = 1.5$ and $\phi_2 = -0.75$




```
> acf(ar2.s,xaxp=c(0,20,10))
```

The sample PACF in Exhibit 6.13 gives a strong indication that we should consider an AR(2) model for these data. The seemingly significant sample PACF at lag 9 would need to be investigated further during model diagnostics.

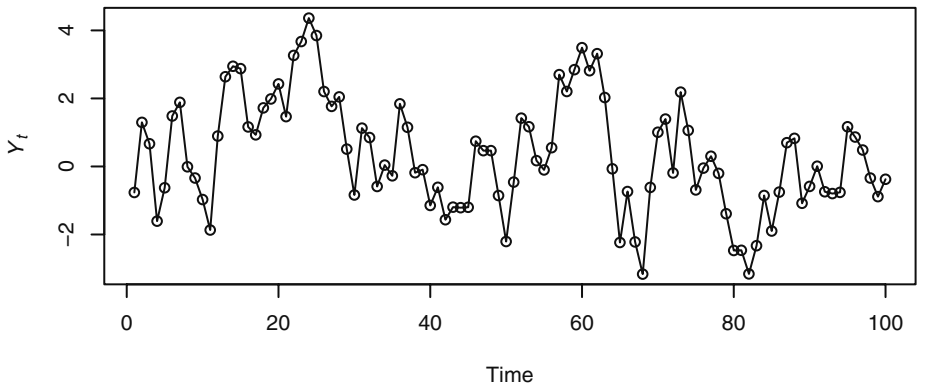
Exhibit 6.13 Sample PACF for an AR(2) Process with $\phi_1 = 1.5$ and $\phi_2 = -0.75$



```
> pacf(ar2.s,xaxp=c(0,20,10))
```

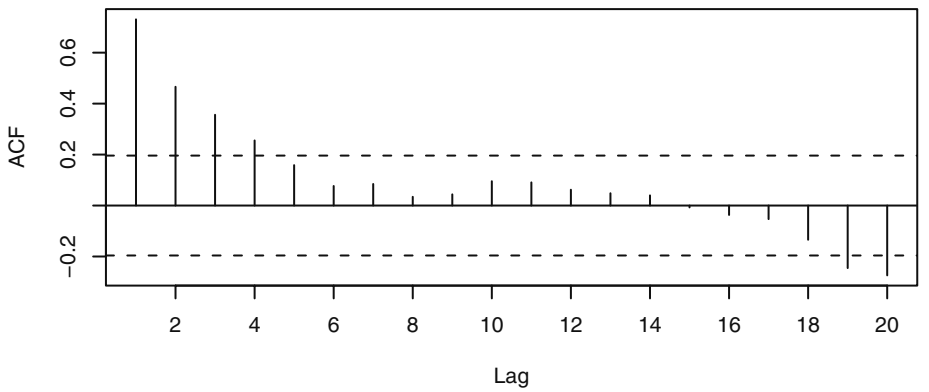
As a final example, we simulated 100 values of a mixed ARMA(1,1) model with $\phi = 0.6$ and $\theta = -0.3$. The time series plot is shown in Exhibit 6.14 and the sample ACF and PACFs are shown in Exhibit 6.15 and Exhibit 6.16, respectively. These seem to indicate that an AR(1) model should be specified.

Exhibit 6.14 Simulated ARMA(1,1) Series with $\phi = 0.6$ and $\theta = -0.3$.



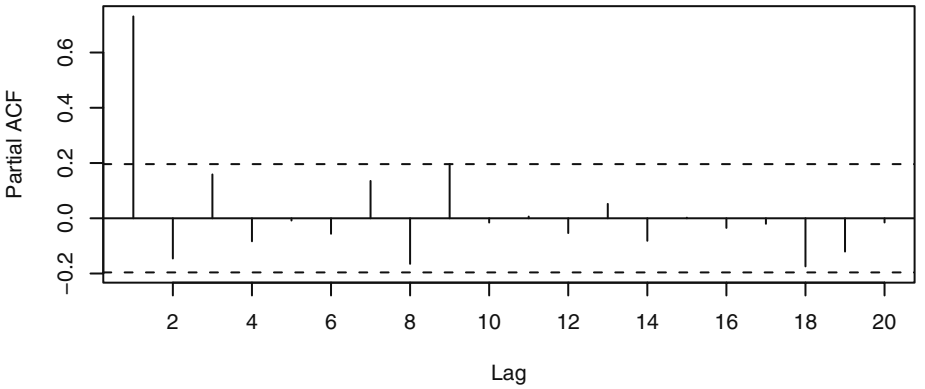

```
> data(arma11.s)
> plot(arma11.s, type='o',ylab=expression(Y[t]))
```

Exhibit 6.15 Sample ACF for Simulated ARMA(1,1) Series




```
> acf(arma11.s,xaxp=c(0,20,10))
```

Exhibit 6.16 Sample PACF for Simulated ARMA(1,1) Series




```
> pacf(arma11.s,xaxp=c(0,20,10))
```

However, the triangular region of zeros shown in the sample EACF in Exhibit 6.17 indicates quite clearly that a mixed model with $q = 1$ and with $p = 1$ or 2 would be more appropriate. We will illustrate further uses of the EACF when we specify some real series in Section 6.6.

Exhibit 6.17 Sample EACF for Simulated ARMA(1,1) Series

AR / MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	o	o	o	o	o	o	o	o	o	o
1	x	o	o	o	o	o	o	o	o	o	o	o	o	o
2	x	o	o	o	o	o	o	o	o	o	o	o	o	o
3	x	x	o	o	o	o	o	o	o	o	o	o	o	o
4	x	o	x	o	o	o	o	o	o	o	o	o	o	o
5	x	o	o	o	o	o	o	o	o	o	o	o	o	o
6	x	o	o	o	x	o	o	o	o	o	o	o	o	o
7	x	o	o	o	x	o	o	o	o	o	o	o	o	o

```
> eacf(arma11.s)
```

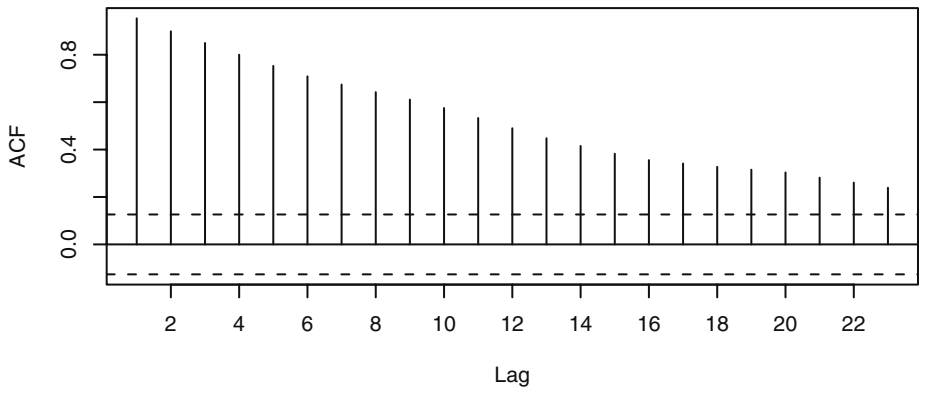
6.4 Nonstationarity

As indicated in Chapter 5, many series exhibit nonstationarity that can be explained by integrated ARMA models. The nonstationarity will frequently be apparent in the time series plot of the series. A review of Exhibits 5.1, 5.5, and 5.8 is recommended here.

The sample ACF computed for nonstationary series will also usually indicate the nonstationarity. The definition of the sample autocorrelation function implicitly *assumes* stationarity; for example, we use lagged products of deviations from the overall mean, and the denominator assumes a constant variance over time. Thus it is not at all clear what the sample ACF is estimating for a nonstationary process. Nevertheless, for nonstationary series, the sample ACF typically fails to die out rapidly as the lags increase. This is due to the tendency for nonstationary series to drift slowly, either up or down, with apparent "trends." The values of r_k need not be large even for low lags, but often they are.

Consider the oil price time series shown in Exhibit 5.1 on page 88. The sample ACF for the logarithms of these data is displayed in Exhibit 6.18. All values shown are "significantly far from zero," and the only pattern is perhaps a linear decrease with increasing lag. The sample PACF (not shown) is also indeterminate.

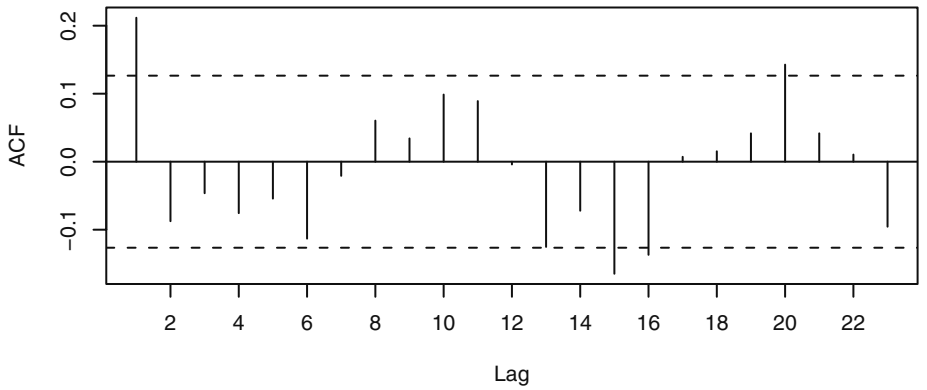
Exhibit 6.18 Sample ACF for the Oil Price Time Series




```
> data(oil.price)
> acf(as.vector(oil.price),xaxp=c(0,24,12))
```

The sample ACF computed on the first differences of the logs of the oil price series is shown in Exhibit 6.19. Now the pattern emerges much more clearly—after differencing, a moving average model of order 1 seems appropriate. The model for the original oil price series would then be a nonstationary IMA(1,1) model. (The "significant" ACF at lags 15, 16, and 20 are ignored for now.)

Exhibit 6.19 Sample ACF for the Difference of the Log Oil Price Series




```
> acf(diff(as.vector(log(oil.price))),xaxp=c(0,24,12))
```

If the first difference of a series and its sample ACF do not appear to support a stationary ARMA model, then we take another difference and again compute the sample ACF and PACF to look for characteristics of a stationary ARMA process. Usually one or at most two differences, perhaps combined with a logarithm or other transformation, will accomplish this reduction to stationarity. Additional properties of the sample ACF computed on nonstationary data are given in Wichern (1973), Roy (1977), and Hasza (1980). See also Box, Jenkins, and Reinsel (1994, p. 218).

Overdifferencing

From Exercise 2.6 on page 20, we know that the difference of any stationary time series is also stationary. However, overdifferencing introduces unnecessary correlations into a series and will complicate the modeling process.

For example, suppose our observed series, $\{Y_t\}$, is in fact a random walk so that one difference would lead to a very simple white noise model

$$\nabla Y_t = Y_t - Y_{t-1} = e_t$$

However, if we difference once more (that is, overdifference) we have

$$\nabla^2 Y_t = e_t - e_{t-1}$$

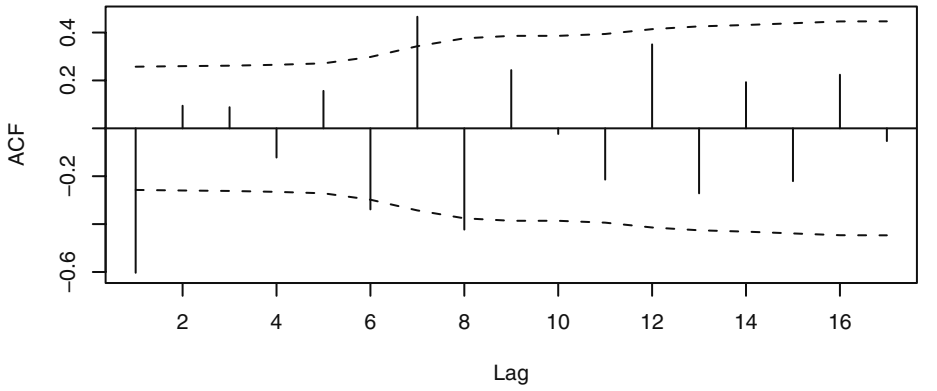
which is an MA(1) model but with $\theta = 1$. If we take two differences in this situation we unnecessarily have to estimate the unknown value of θ. Specifying an IMA(2,1) model would not be appropriate here. The random walk model, which can be thought of as IMA(1,1) with $\theta = 0$, is the correct model.[†] Overdifferencing also creates a noninvert-

[†] The random walk model can also be thought of as an ARI(1,1) with $\phi = 0$ or as a nonstationary AR(1) with $\phi = 1$.

ible model—see Section 4.5 on page 79.[†] Noninvertible models also create serious problems when we attempt to estimate their parameters—see Chapter 7.

To illustrate overdifferencing, consider the random walk shown in Exhibit 2.1 on page 14. Taking one difference should lead to white noise—a very simple model. If we mistakenly take two differences (that is, overdifference) and compute the sample ACF, we obtain the graph shown in Exhibit 6.20. Based on this plot, we would likely specify at least an IMA(2,1) model for the original series and then estimate the unnecessary MA parameter. We also have a significant sample ACF value at lag 7 to think about and deal with.

Exhibit 6.20 Sample ACF of Overdifferenced Random Walk

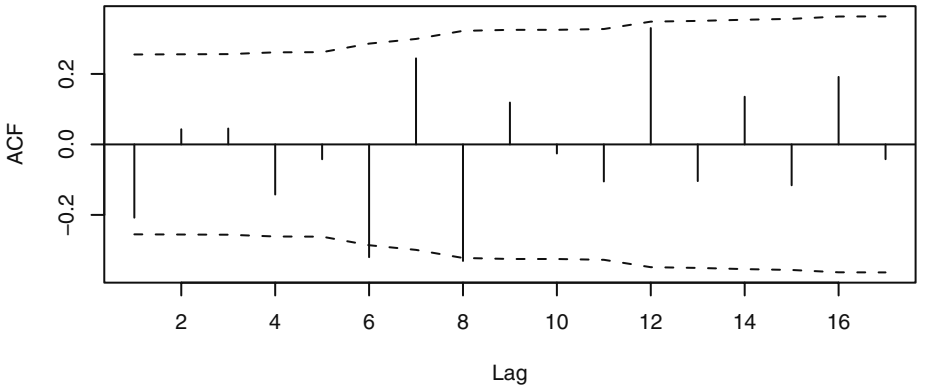



```
> data(rwalk)
> acf(diff(rwalk,difference=2),ci.type='ma', xaxp=c(0,18,9))
```

In contrast, Exhibit 6.21 displays the sample ACF of the *first* difference of the random walk series. Viewing this graph, we would likely want to consider the correct model—the first difference looks very much like white noise.

[†] In backshift notation, if the correct model is $\phi(B)(1-B)Y_t = \theta(B)e_t$, overdifferencing leads to $\phi(B)(1-B)^2Y_t = \theta(B)(1-B)e_t = \theta'(B)e_t$, say, where $\theta'(B) = (1-B)\theta(B)$ and the "forbidden" root in $\theta'(B)$ at $B = 1$ is obvious.

Exhibit 6.21 Sample ACF of Correctly Differenced Random Walk




```
> acf(diff(rwalk),ci.type='ma',xaxp=c(0,18,9))
```

To avoid overdifferencing, we recommend looking carefully at each difference in succession and keeping the principle of parsimony always in mind—*models should be simple, but not too simple*.

The Dickey-Fuller Unit-Root Test

While the approximate linear decay of the sample ACF is often taken as a symptom that the underlying time series is nonstationary and requires differencing, it is also useful to quantify the evidence of nonstationarity in the data-generating mechanism. This can be done via hypothesis testing. Consider the model

$$Y_t = \alpha Y_{t-1} + X_t \text{ for } t = 1, 2, \ldots$$

where $\{X_t\}$ is a stationary process. The process $\{Y_t\}$ is nonstationary if the coefficient $\alpha = 1$, but it is stationary if $|\alpha| < 1$. Suppose that $\{X_t\}$ is an AR(k) process: $X_t = \phi_1 X_{t-1} + \cdots + \phi_k X_{t-k} + e_t$. Under the null hypothesis that $\alpha = 1$, $X_t = Y_t - Y_{t-1}$. Letting $a = \alpha - 1$, we have

$$\begin{aligned} Y_t - Y_{t-1} &= (\alpha - 1) Y_{t-1} + X_t \\ &= a Y_{t-1} + \phi_1 X_{t-1} + \cdots + \phi_k X_{t-k} + e_t \\ &= a Y_{t-1} + \phi_1 (Y_{t-1} - Y_{t-2}) + \cdots + \phi_k (Y_{t-k} - Y_{t-k-1}) + e_t \end{aligned} \tag{6.4.1}$$

where $a = 0$ under the hypothesis that Y_t is difference nonstationary. On the other hand, if $\{Y_t\}$ is stationary so that $-1 < \alpha < 1$, then it can be verified that Y_t still satisfies an equation similar to the equation above but with different coefficients; for example, $a = (1 - \phi_1 - \cdots - \phi_k)(1 - \alpha) < 0$. Indeed, $\{Y_t\}$ is then an AR($k + 1$) process whose AR characteristic equation is given by $\Phi(x)(1 - \alpha x) = 0$, where $\Phi(x) = 1 - \phi_1 x - \cdots - \phi_k x^k$. So, the null hypothesis corresponds to the case where the AR characteristic polynomial has a unit root and the alternative hypothesis states that it has no unit roots. *Consequently, the*

test for differencing amounts to testing for a unit root in the AR characteristic polynomial of $\{Y_t\}$.

By the analysis above, the null hypothesis that $\alpha = 1$ (equivalently $a = 0$) can be tested by regressing the first difference of the observed time series on lag 1 of the observed series and on the past k lags of the first difference of the observed series. We then test whether the coefficient $a = 0$—the null hypothesis being that the process is difference nonstationary. That is, the process is nonstationary but becomes stationary after first differencing. The alternative hypothesis is that $a < 0$ and hence $\{Y_t\}$ is stationary. The augmented Dickey-Fuller (ADF) test statistic is the t-statistic of the estimated coefficient of a from the method of least squares regression. However, the ADF test statistic is not approximately t-distributed under the null hypothesis; instead, it has a certain nonstandard large-sample distribution under the null hypothesis of a unit root. Fortunately, percentage points of this limit (null) distribution have been tabulated; see Fuller (1996).

In practice, even after first differencing, the process may not be a finite-order AR process, but it may be closely approximated by some AR process with the AR order increasing with the sample size. Said and Dickey (1984) (see also Chang and Park, 2002) showed that with the AR order increasing with the sample size, the ADF test has the same large-sample null distribution as the case where the first difference of the time series is a finite-order AR process. Often, the approximating AR order can be first estimated based on some information criteria (for example, AIC or BIC) before carrying out the ADF test. See Section 6.5 on page 130 for more information on the AIC and BIC criteria.

In some cases, the process may be trend nonstationary in the sense that it has a deterministic trend (for example, some linear trend) but otherwise is stationary. A unit-root test may be conducted with the aim of discerning difference stationarity from trend stationarity. This can be done by carrying out the ADF test with the detrended data. Equivalently, this can be implemented by regressing the first difference on the covariates defining the trend, the lag 1 of the original data, and the past lags of the first difference of the original data. The t-statistic based on the coefficient estimate of the lag 1 of the original data furnishes the ADF test statistic, which has another nonstandard large-sample null distribution. See Phillips and Xiao (1998) for a survey of unit root testing.

We now illustrate the ADF test with the simulated random walk shown in Exhibit 2.1 on page 14. First, we consider testing the null hypothesis of a unit root versus the alternative hypothesis that the time series is stationary with unknown mean. Hence, the regression defined by Equation (6.4.1) is augmented with an intercept to allow for the possibly nonzero mean under the alternative hypothesis. (For the alternative hypothesis that the process is a stationary process of zero mean, the ADF test statistic can be obtained by running the unaugmented regression defined by Equation (6.4.1).) To carry out the test, we must determine k.[†] Using the AIC with the first difference of the data, we find that $k = 8$, in which case the ADF test statistic becomes –0.601, with the p-value

[†] R code: `ar(diff(rwalk))`

being greater than 0.1.[†] On the other hand, setting $k = 0$ (the true order) leads to the ADF statistic −1.738, with p-value still greater than 0.1.[‡] Thus, there is strong evidence supporting the unit-root hypothesis. Second, recall that the simulated random walk appears to have a linear trend. Hence, linear trend plus stationary error forms another reasonable alternative to the null hypothesis of unit root (difference nonstationarity). For this test, we include both an intercept term and the covariate time in the regression defined by Equation (6.4.1). With $k = 8$, the ADF test statistic equals −2.289 with p-value greater than 0.1[††]; that is, we do not reject the null hypothesis of unit root. On the other hand, setting $k = 0$, the true order that is unknown in practice, the ADF test statistic becomes −3.49 with p-value equal to 0.0501.[‡‡] Hence, there is weak evidence that the process is linear-trend nonstationary; that is, the process equals linear time trend plus stationary error, contrary to the truth that the process is a random walk, being difference nonstationary! This example shows that with a small sample size, it may be hard to differentiate between trend nonstationarity and difference nonstationarity.

6.5 Other Specification Methods

A number of other approaches to model specification have been proposed since Box and Jenkins' seminal work. One of the most studied is **Akaike's** (1973) **Information Criterion** (**AIC**). This criterion says to select the model that minimizes

$$\text{AIC} = -2\log(\text{maximum likelihood}) + 2k \tag{6.5.1}$$

where $k = p + q + 1$ if the model contains an intercept or constant term and $k = p + q$ otherwise. Maximum likelihood estimation is discussed in Chapter 7. The addition of the term $2(p + q + 1)$ or $2(p + q)$ serves as a "penalty function" to help ensure selection of parsimonious models and to avoid choosing models with too many parameters.

The AIC is an estimator of the average Kullback-Leibler divergence of the estimated model from the true model. Let $p(y_1,y_2,\ldots,y_n)$ be the true pdf of $Y_1, Y_2, \ldots, Y_n$, and $q_\theta(y_1,y_2,\ldots,y_n)$ be the corresponding pdf under the model with parameter θ. The Kullback-Leibler divergence of q_θ from p is defined by the formula

$$D(p, q_\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(y_1, y_2, \ldots, y_n)\log\left[\frac{p(y_1, y_2, \ldots, y_n)}{q_\theta(y_1, y_2, \ldots, y_n)}\right]dy_1 dy_2 \ldots dy_n$$

The AIC estimates $E[D(p, q_{\hat{\theta}})]$, where $\hat{\theta}$ is the maximum likelihood estimator of the vector parameter θ. However, the AIC is a biased estimator, and the bias can be appreciable for large parameter per data ratios. Hurvich and Tsai (1989) showed that the bias can be approximately eliminated by adding another nonstochastic penalty term to the AIC, resulting in the corrected AIC, denoted by AIC_C and defined by the formula

---

[†] R code: `library(uroot); ADF.test(rwalk,selectlags=list(mode=c(1,2,3,4,5,6,7,8),Pmax=8),itsd=c(1,0,0))`

[‡] `ADF.test(rwalk,selectlags=list(Pmax=0),itsd=c(1,0,0))`

[††] `ADF.test(rwalk,selectlags=list(mode=c(1,2,3,4,5,6,7,8),Pmax=8),itsd=c(1,1,0))`

[‡‡] `ADF.test(rwalk,selectlags=list(Pmax=0),itsd=c(1,1,0))`

$$\mathrm{AIC_C} = \mathrm{AIC} + \frac{2(k+1)(k+2)}{n-k-2} \tag{6.5.2}$$

Here n is the (effective) sample size and again k is the total number of parameters as above excluding the noise variance. Simulation results by Hurvich and Tsai (1989) suggest that for cases with k/n greater than 10%, the $\mathrm{AIC_C}$ outperforms many other model selection criteria, including both the AIC and BIC.

Another approach to determining the ARMA orders is to select a model that minimizes the Schwarz **Bayesian Information Criterion** (**BIC**) defined as

$$\mathrm{BIC} = -2\log(\text{maximum likelihood}) + k\log(n) \tag{6.5.3}$$

If the true process follows an ARMA(p,q) model, then it is known that the orders specified by minimizing the BIC are consistent; that is, they approach the true orders as the sample size increases. However, if the true process is not a finite-order ARMA process, then minimizing AIC among an increasingly large class of ARMA models enjoys the appealing property that it will lead to an optimal ARMA model that is closest to the true process among the class of models under study.[†]

Regardless of whether we use the AIC or BIC, the methods require carrying out maximum likelihood estimation. However, maximum likelihood estimation for an ARMA model is prone to numerical problems due to multimodality of the likelihood function and the problem of overfitting when the AR and MA orders exceed the true orders. Hannan and Rissanen (1982) proposed an interesting and practical solution to this problem. Their procedure consists of first fitting a high-order AR process with the order determined by minimizing the AIC. The second step uses the residuals from the first step as proxies for the unobservable error terms. Thus, an ARMA(k,j) model can be approximately estimated by regressing the time series on its own lags 1 to k together with the lags 1 to j of the residuals from the high order autoregression; the BIC of this autoregressive model is an estimate of the BIC obtained with maximum likelihood estimation. Hannan and Rissanen (1982) demonstrated that minimizing the approximate BIC still leads to consistent estimation of the ARMA orders.

Order determination is related to the problem of finding the subset of nonzero coefficients of an ARMA model with sufficiently high ARMA orders. A subset ARMA(p,q) model is an ARMA(p,q) model with a subset of its coefficients known to be zero. For example, the model

$$Y_t = 0.8Y_{t-12} + e_t + 0.7e_{t-12} \tag{6.5.4}$$

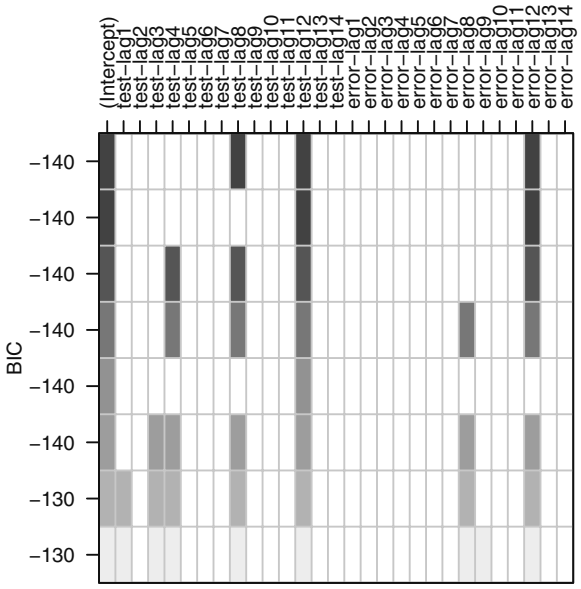
is a subset ARMA(12,12) model useful for modeling some monthly seasonal time series. For ARMA models of very high orders, such as the preceding ARMA(12,12) model, finding a subset ARMA model that adequately approximates the underlying process is more important from a practical standpoint than simply determining the ARMA orders. The method of Hannan and Rissanen (1982) for estimating the ARMA orders can be extended to solving the problem of finding an optimal subset ARMA model.

[†] Closeness is measured in terms of the Kullback-Leibler divergence—a measure of disparity between models. See Shibata (1976) and the discussion in Stenseth et al. (2004).

Indeed, several model selection criteria (including AIC and BIC) of the subset ARMA(p,q) models (2^{p+q} of them!) can be approximately, exhaustively, and quickly computed by the method of regression by leaps and bounds (Furnival and Wilson, 1974) applied to the subset regression of Y_t on its own lags and on lags of the residuals from a high-order autoregression of $\{Y_t\}$.

It is prudent to examine a few best subset ARMA models (in terms of, for example, BIC) in order to arrive at some helpful tentative models for further study. The pattern of which lags of the observed time series and which of the error process enter into the various best subset models can be summarized succinctly in a display like that shown in Exhibit 6.22. This table is based on a simulation of the ARMA(12,12) model shown in Equation (6.5.4). Each row in the exhibit corresponds to a subset ARMA model where the cells of the variables selected for the model are shaded. The models are sorted according to their BIC, with better models (lower BIC) placed in higher rows and with darker shades. The top row tells us that the subset ARMA(14,14) model with the smallest BIC contains only lags 8 and 12 of the observed time series and lag 12 of the error process. The next best model contains lag 12 of the time series and lag 8 of the errors, while the third best model contains lags 4, 8, and 12 of the time series and lag 12 of the errors. In our simulated time series, the second best model is the true subset model. However, the BIC values for these three models are all very similar, and all three (plus the fourth best model) are worthy of further study. However, lag 12 of the time series and that of the errors are the two variables most frequently found in the various subset models summarized in the exhibit, suggesting that perhaps they are the more important variables, as we know they are!

Exhibit 6.22 Best Subset ARMA Selection Based on BIC



```
> set.seed(92397)
> test=arima.sim(model=list(ar=c(rep(0,11),.8),
    ma=c(rep(0,11),0.7)),n=120)
> res=armasubsets(y=test,nar=14,nma=14,y.name='test',
    ar.method='ols')
> plot(res)
```

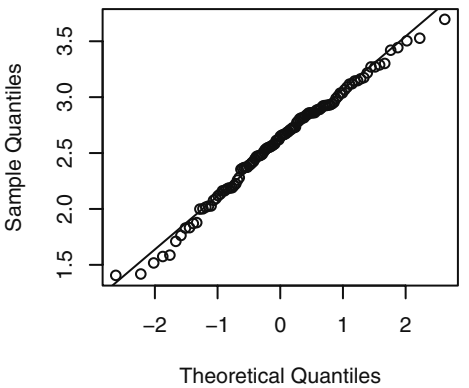
6.6 Specification of Some Actual Time Series

Consider now specification of models for some of the actual time series that we saw in earlier chapters.

The Los Angeles Annual Rainfall Series

Annual total rainfall amounts for Los Angeles were shown in Exhibit 1.1 on page 2. In Chapter 3, we noted in Exhibit 3.17 on page 50, that rainfall amounts were not normally distributed. As is shown in Exhibit 6.23, taking logarithms improves the normality dramatically.

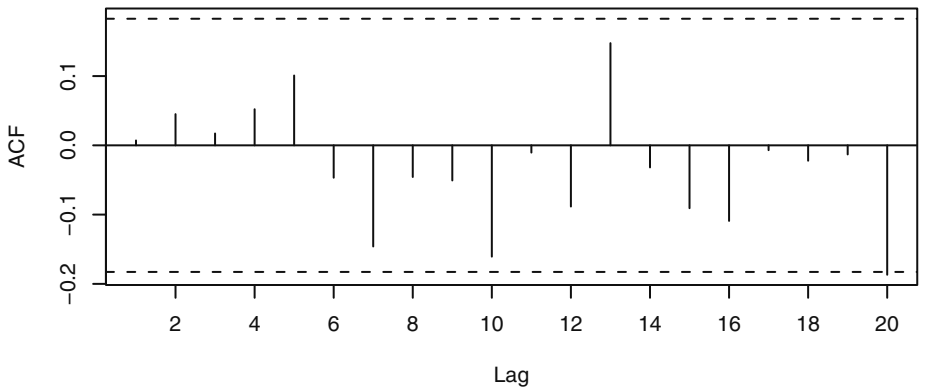
Exhibit 6.23 QQ Normal Plot of the Logarithms of LA Annual Rainfall




```
> data(larain); win.graph(width=2.5,height=2.5,pointsize=8)
> qqnorm(log(larain)); qqline(log(larain))
```

Exhibit 6.24 displays the sample autocorrelations for the logarithms of the annual rainfall series.

Exhibit 6.24 Sample ACF of the Logarithms of LA Annual Rainfall



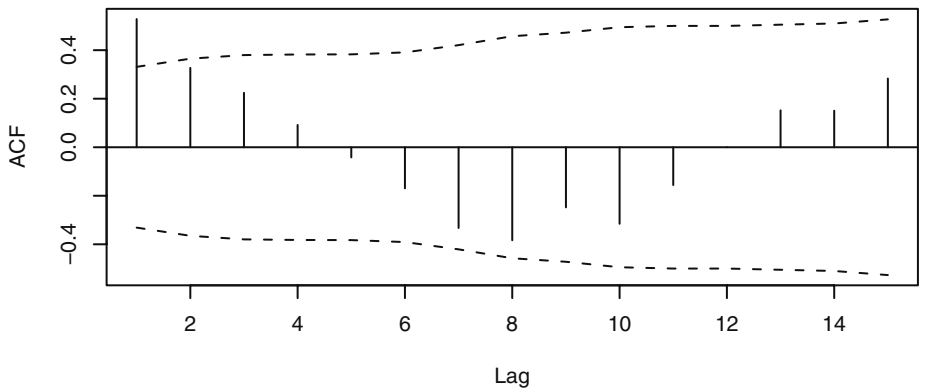

```
> win.graph(width=4.875,height=3,pointsize=8)
> acf(log(larain),xaxp=c(0,20,10))
```

The log transformation has improved the normality, but there is no discernable dependence in this time series. We could model the logarithm of annual rainfall amount as independent, normal random variables with mean 2.58 and standard deviation 0.478. Both these values are in units of log(inches).

The Chemical Process Color Property Series

The industrial chemical process color property displayed in Exhibit 1.3 on page 3, shows more promise of interesting time series modeling—especially in light of the dependence of successive batches shown in Exhibit 1.4 on page 4. The sample ACF plotted in Exhibit 6.25 might at first glance suggest an MA(1) model, as only the lag 1 autocorrelation is significantly different from zero.

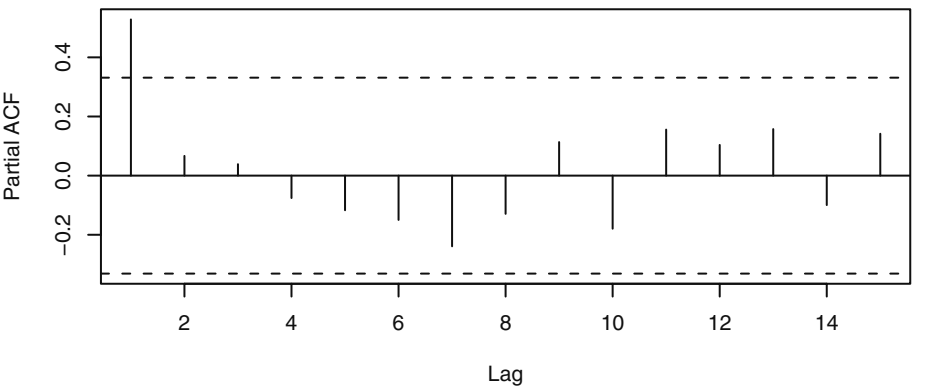
Exhibit 6.25 Sample ACF for the Color Property Series




```
> data(color); acf(color,ci.type='ma')
```

However, the damped sine wave appearance of the plot encourages us to look further at the sample partial autocorrelation. Exhibit 6.26 displays that plot, and now we see clearly that an AR(1) model is worthy of first consideration. As always, our specified models are tentative and subject to modification during the model diagnostics stage of model building.

Exhibit 6.26 Sample Partial ACF for the Color Property Series

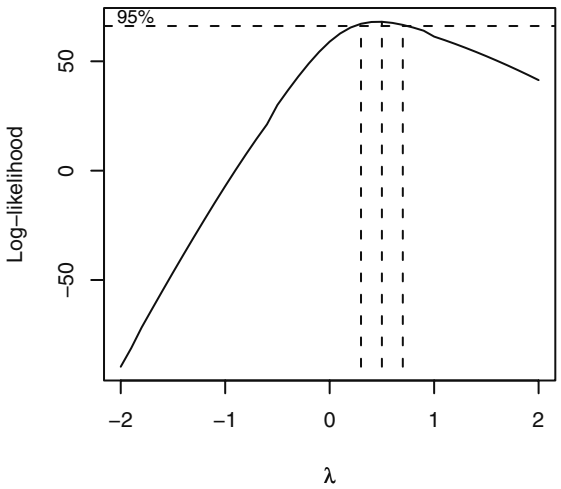



```
> pacf(color)
```

The Annual Abundance of Canadian Hare Series

The time series of annual abundance of hare of the Hudson Bay in Canada was displayed in Exhibit 1.5 on page 5, and the year-to-year dependence was demonstrated in Exhibit 1.6. It has been suggested in the literature that a transformation might be used to produce a good model for these data. Exhibit 6.27 displays the log-likelihood as a function of the power parameter, λ. The maximum occurs at $\lambda = 0.4$, but a square root transformation with $\lambda = 0.5$ is well within the confidence interval for λ. We will take the square root of the abundance values for all further analyses.

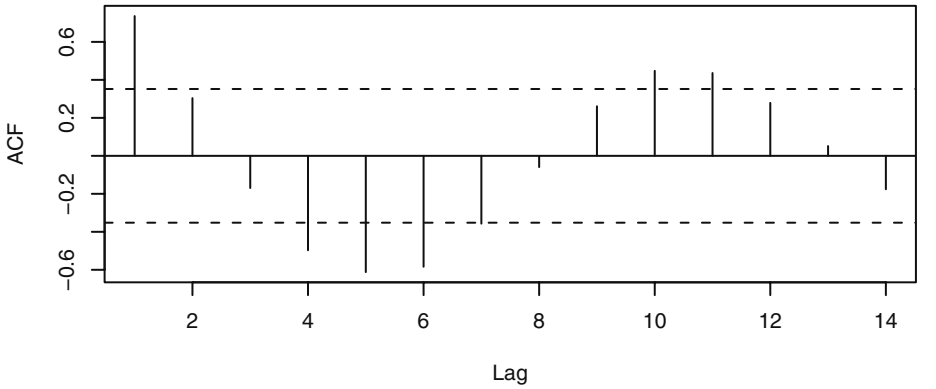
Exhibit 6.27 Box-Cox Power Transformation Results for Hare Abundance




```
> win.graph(width=3,height=3,pointsize=8)
> data(hare); BoxCox.ar(hare)
```

Exhibit 6.28 shows the sample ACF for this transformed series. The fairly strong lag 1 autocorrelation dominates but, again, there is a strong indication of damped oscillatory behavior.

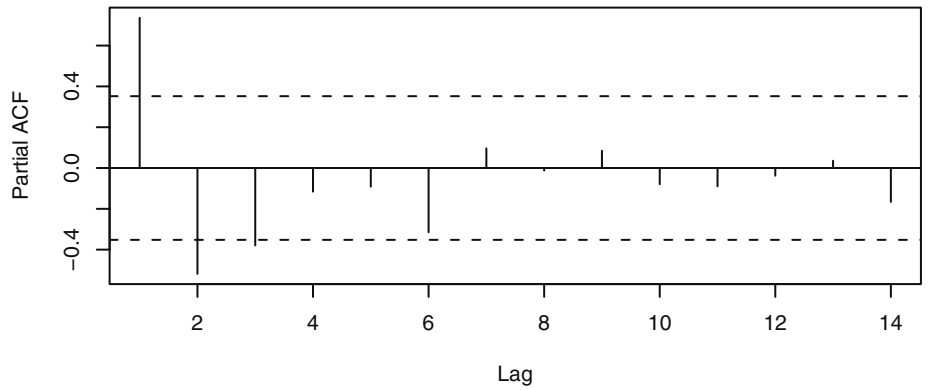
Exhibit 6.28 Sample ACF for Square Root of Hare Abundance




```
> acf(hare^.5)
```

The sample partial autocorrelation for the transformed series is shown in Exhibit 6.29. It gives strong evidence to support an AR(2) or possibly an AR(3) model for these data.

Exhibit 6.29 Sample Partial ACF for Square Root of Hare Abundance




```
> pacf(hare^.5)
```

The Oil Price Series

In Chapter 5, we began to look at the monthly oil price time series and argued graphically that the difference of the logarithms could be considered stationary—see Exhibit 5.1 on page 88. Software implementation of the Augmented Dickey-Fuller unit-root test applied to the logs of the original prices leads to a test statistic of −1.1119 and a p-value of 0.9189. With stationarity as the alternative hypothesis, this provides strong evidence of nonstationarity and the appropriateness of taking a difference of the logs. For this test, the software chose a value of $k = 6$ in Equation (6.4.1) on page 128 based on large-sample theory.

Exhibit 6.30 shows the summary EACF table for the differences of the logarithms of the oil price data. This table suggests an ARMA model with $p = 0$ and $q = 1$.

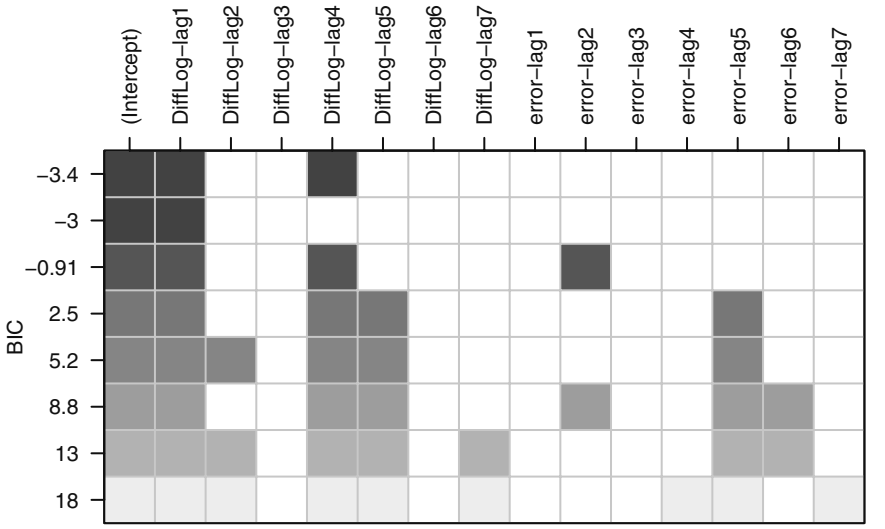
Exhibit 6.30 Extended ACF for Difference of Logarithms of Oil Price Series

AR / MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	o	o	o	o	o	o	o	o	o	o	o	o	o
1	x	x	o	o	o	o	o	o	o	o	x	o	o	o
2	o	x	o	o	o	o	o	o	o	o	o	o	o	o
3	o	x	o	o	o	o	o	o	o	o	o	o	o	o
4	o	x	x	o	o	o	o	o	o	o	o	o	o	o
5	o	x	o	x	o	o	o	o	o	o	o	o	o	o
6	o	x	o	x	o	o	o	o	o	o	o	o	o	o
7	x	x	o	x	o	o	o	o	o	o	o	o	o	o

```
> eacf(diff(log(oil.price)))
```

The results of the best subsets ARMA approach are displayed in Exhibit 6.31.

Exhibit 6.31 Best Subset ARMA Model for Difference of Log(Oil)

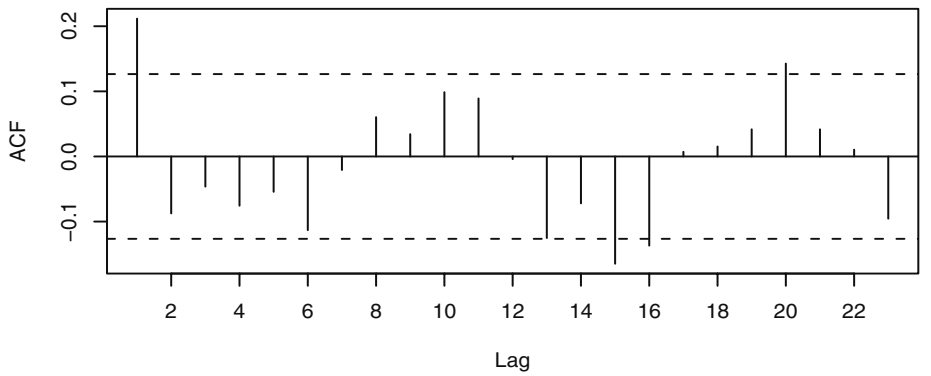



```
> res=armasubsets(y=diff(log(oil.price)),nar=7,nma=7,
    y.name='test', ar.method='ols')
> plot(res)
```

Here the suggestion is that $Y_t = \nabla\log(\text{Oil}_t)$ should be modeled in terms of Y_{t-1} and Y_{t-4} and that no lags are needed in the error terms. The second best model omits the lag 4 term so that an ARIMA(1,1,0) model on the logarithms should also be investigated further.

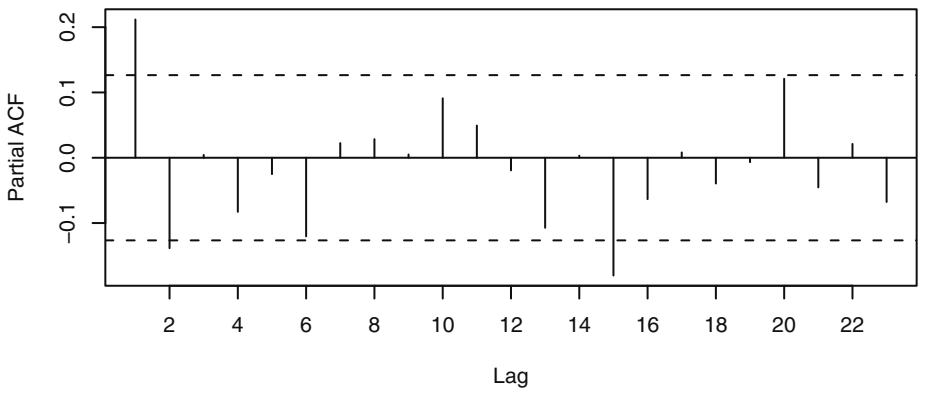
Exhibit 6.32 suggests that we specify an MA(1) model for the difference of the log oil prices, and Exhibit 6.33 says to consider an AR(2) model (ignoring some significant spikes at lags 15, 16, and 20). We will want to look at all of these models further when we estimate parameters and perform diagnostic tests in Chapters 7 and 8. (We will see later that to obtain a suitable model for the oil price series, the outliers in the series will need to be dealt with. (Can you spot the outliers in Exhibit 5.4 on page 91?)

Exhibit 6.32 Sample ACF of Difference of Logged Oil Prices




```
> acf(as.vector(diff(log(oil.price))),xaxp=c(0,22,11))
```

Exhibit 6.33 Sample PACF of Difference of Logged Oil Prices




```
> pacf(as.vector(diff(log(oil.price))),xaxp=c(0,22,11))
```

6.7 Summary

In this chapter, we considered the problem of specifying reasonable but simple models for observed times series. In particular, we investigated tools for choosing the orders (p, d, and q) for ARIMA(p,d,q) models. Three tools, the sample autocorrelation function, the sample partial autocorrelation function, and the sample extended autocorrelation function, were introduced and studied to help with this difficult task. The Dickey-Fuller unit-root test was also introduced to help distinguish between stationary and nonstationary series. These ideas were all illustrated with both simulated and actual time series.

EXERCISES

6.1 Verify Equation (6.1.3) on page 110 for the white noise process.
6.2 Verify Equation (6.1.4) on page 110 for the AR(1) process.
6.3 Verify the line in Exhibit 6.1 on page 111, for the values $\phi = \pm 0.9$.
6.4 Add new entries to Exhibit 6.1 on page 111, for the following values:
(a) $\phi = \pm 0.99$.
(b) $\phi = \pm 0.5$.
(c) $\phi = \pm 0.1$.
6.5 Verify Equation (6.1.9) on page 111 and Equation (6.1.10) for the MA(1) process.
6.6 Verify the line in Exhibit 6.2 on page 112, for the values $\theta = \pm 0.9$.
6.7 Add new entries to Exhibit 6.2 on page 112, for the following values:
(a) $\theta = \pm 0.99$.
(b) $\theta = \pm 0.8$.
(c) $\theta = \pm 0.2$.
6.8 Verify Equation (6.1.11) on page 112, for the general MA(q) process.
6.9 Use Equation (6.2.3) on page 113, to verify the value for the lag 2 partial autocorrelation function for the MA(1) process given in Equation (6.2.5) on page 114.
6.10 Show that the general expression for the partial autocorrelation function of an MA(1) process given in Equation (6.2.6) on page 114, satisfies the Yule-Walker recursion given in Equation (6.2.7).
6.11 Use Equation (6.2.8) on page 114, to find the (theoretical) partial autocorrelation function for an AR(2) model in terms of ϕ_1 and ϕ_2 and lag $k = 1, 2, 3, \ldots$.
6.12 From a time series of 100 observations, we calculate $r_1 = -0.49$, $r_2 = 0.31$, $r_3 = -0.21$, $r_4 = 0.11$, and $|r_k| < 0.09$ for $k > 4$. On this basis alone, what ARIMA model would we tentatively specify for the series?
6.13 A stationary time series of length 121 produced sample partial autocorrelation of $\hat{\phi}_{11} = 0.8$, $\hat{\phi}_{22} = -0.6$, $\hat{\phi}_{33} = 0.08$, and $\hat{\phi}_{44} = 0.00$. Based on this information alone, what model would we tentatively specify for the series?
6.14 For a series of length 169, we find that $r_1 = 0.41$, $r_2 = 0.32$, $r_3 = 0.26$, $r_4 = 0.21$, and $r_5 = 0.16$. What ARIMA model fits this pattern of autocorrelations?

6.15 The sample ACF for a series and its first difference are given in the following table. Here $n = 100$.

lag	1	2	3	4	5	6
ACF for Y_t	0.97	0.97	0.93	0.85	0.80	0.71
ACF for ∇Y_t	−0.42	0.18	−0.02	0.07	−0.10	−0.09

Based on this information alone, which ARIMA model(s) would we consider for the series?

6.16 For a series of length 64, the sample partial autocorrelations are given as:

Lag	1	2	3	4	5
PACF	0.47	−0.34	0.20	0.02	−0.06

Which models should we consider in this case?

6.17 Consider an AR(1) series of length 100 with $\phi = 0.7$.

(a) Would you be surprised if $r_1 = 0.6$?

(b) Would $r_{10} = -0.15$ be unusual?

6.18 Suppose the $\{X_t\}$ is a stationary AR(1) process with parameter ϕ but that we can only observe $Y_t = X_t + N_t$ where $\{N_t\}$ is the white noise measurement error independent of $\{X_t\}$.

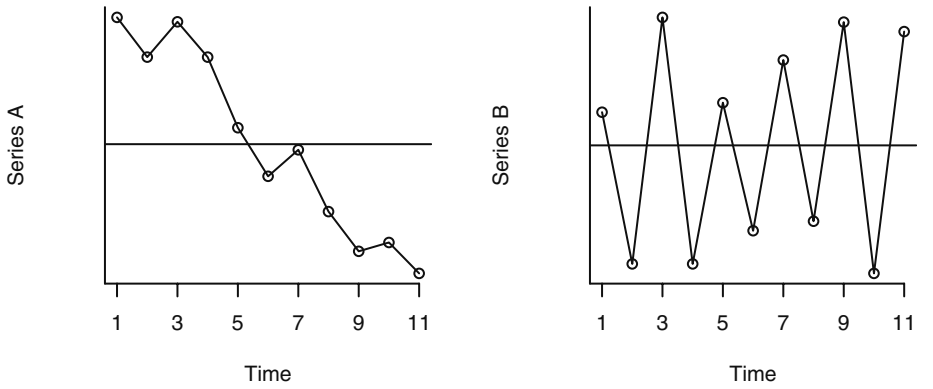
(a) Find the autocorrelation function for the observed process in terms of ϕ, σ_X^2, and σ_N^2.

(b) Which ARIMA model might we specify for $\{Y_t\}$?

6.19 The time plots of two series are shown below.

(a) For each of the series, describe r_1 using the terms strongly positive, moderately positive, near zero, moderately negative, or strongly negative. Do you need to know the scale of measurement for the series to answer this?

(b) Repeat part (a) for r_2.



6.20 Simulate an AR(1) time series with $n = 48$ and with $\phi = 0.7$.
(a) Calculate the theoretical autocorrelations at lag 1 and lag 5 for this model.
(b) Calculate the sample autocorrelations at lag 1 and lag 5 and compare the values with their theoretical values. Use Equations (6.1.5) and (6.1.6) page 111, to quantify the comparisons.
(c) Repeat part (b) with a new simulation. Describe how the precision of the estimate varies with different samples selected under identical conditions.
(d) If software permits, repeat the simulation of the series and calculation of r_1 and r_5 many times and form the sampling distributions of r_1 and r_5. Describe how the precision of the estimate varies with different samples selected under identical conditions. How well does the large-sample variance given in Equation (6.1.5) on page 111, approximate the variance in your sampling distribution?

6.21 Simulate an MA(1) time series with $n = 60$ and with $\theta = 0.5$.
(a) Calculate the theoretical autocorrelation at lag 1 for this model.
(b) Calculate the sample autocorrelation at lag 1, and compare the value with its theoretical value. Use Exhibit 6.2 on page 112, to quantify the comparisons.
(c) Repeat part (b) with a new simulation. Describe how the precision of the estimate varies with different samples selected under identical conditions.
(d) If software permits, repeat the simulation of the series and calculation of r_1 many times and form the sampling distribution of r_1. Describe how the precision of the estimate varies with different samples selected under identical conditions. How well does the large-sample variance given in Exhibit 6.2 on page 112, approximate the variance in your sampling distribution?

6.22 Simulate an AR(1) time series with $n = 48$, with
(a) $\phi = 0.9$, and calculate the theoretical autocorrelations at lag 1 and lag 5;
(b) $\phi = 0.6$, and calculate the theoretical autocorrelations at lag 1 and lag 5;
(c) $\phi = 0.3$, and calculate the theoretical autocorrelations at lag 1 and lag 5.
(d) For each of the series in parts (a), (b), and (c), calculate the sample autocorrelations at lag 1 and lag 5 and compare the values with their theoretical values. Use Equations (6.1.5) and 6.1.6, page 111, to quantify the comparisons. In general, describe how the precision of the estimate varies with the value of ϕ.

6.23 Simulate an AR(1) time series with $\phi = 0.6$, with
(a) $n = 24$, and estimate $\rho_1 = \phi = 0.6$ with r_1;
(b) $n = 60$, and estimate $\rho_1 = \phi = 0.6$ with r_1;
(c) $n = 120$, and estimate $\rho_1 = \phi = 0.6$ with r_1.
(d) For each of the series in parts (a), (b), and (c), compare the estimated values with the theoretical value. Use Equation (6.1.5) on page 111, to quantify the comparisons. In general, describe how the precision of the estimate varies with the sample size.

6.24 Simulate an MA(1) time series with $\theta = 0.7$, with
(a) $n = 24$, and estimate ρ_1 with r_1;
(b) $n = 60$, and estimate ρ_1 with r_1;
(c) $n = 120$, and estimate ρ_1 with r_1.
(d) For each of the series in parts (a), (b), and (c), compare the estimated values of ρ_1 with the theoretical value. Use Exhibit 6.2 on page 112, to quantify the comparisons. In general, describe how the precision of the estimate varies with the sample size.

6.25 Simulate an AR(1) time series of length $n = 36$ with $\phi = 0.7$.
(a) Calculate and plot the theoretical autocorrelation function for this model. Plot sufficient lags until the correlations are negligible.
(b) Calculate and plot the sample ACF for your simulated series. How well do the values and patterns match the theoretical ACF from part (a)?
(c) What are the theoretical partial autocorrelations for this model?
(d) Calculate and plot the sample ACF for your simulated series. How well do the values and patterns match the theoretical ACF from part (a)? Use the large-sample standard errors reported in Exhibit 6.1 on page 111, to quantify your answer.
(e) Calculate and plot the sample PACF for your simulated series. How well do the values and patterns match the theoretical PACF from part (c)? Use the large-sample standard errors reported on page 115 to quantify your answer.

6.26 Simulate an MA(1) time series of length $n = 48$ with $\theta = 0.5$.
(a) What are the theoretical autocorrelations for this model?
(b) Calculate and plot the sample ACF for your simulated series. How well do the values and patterns match the theoretical ACF from part (a)?
(c) Calculate and plot the theoretical partial autocorrelation function for this model. Plot sufficient lags until the correlations are negligible. (Hint: See Equation (6.2.6) on page 114.)
(d) Calculate and plot the sample PACF for your simulated series. How well do the values and patterns match the theoretical PACF from part (c)?

6.27 Simulate an AR(2) time series of length $n = 72$ with $\phi_1 = 0.7$ and $\phi_2 = -0.4$.
(a) Calculate and plot the theoretical autocorrelation function for this model. Plot sufficient lags until the correlations are negligible.
(b) Calculate and plot the sample ACF for your simulated series. How well do the values and patterns match the theoretical ACF from part (a)?
(c) What are the theoretical partial autocorrelations for this model?
(d) Calculate and plot the sample ACF for your simulated series. How well do the values and patterns match the theoretical ACF from part (a)?
(e) Calculate and plot the sample PACF for your simulated series. How well do the values and patterns match the theoretical PACF from part (c)?

6.28 Simulate an MA(2) time series of length $n = 36$ with $\theta_1 = 0.7$ and $\theta_2 = -0.4$.
- **(a)** What are the theoretical autocorrelations for this model?
- **(b)** Calculate and plot the sample ACF for your simulated series. How well do the values and patterns match the theoretical ACF from part (a)?
- **(c)** Plot the theoretical partial autocorrelation function for this model. Plot sufficient lags until the correlations are negligible. (We do not have a formula for this PACF. Instead, perform a very large sample simulation, say $n = 1000$, for this model and calculate and plot the sample PACF for this simulation.)
- **(d)** Calculate and plot the sample PACF for your simulated series of part (a). How well do the values and patterns match the "theoretical" PACF from part (c)?

6.29 Simulate a mixed ARMA(1,1) model of length $n = 60$ with $\phi = 0.4$ and $\theta = 0.6$.
- **(a)** Calculate and plot the theoretical autocorrelation function for this model. Plot sufficient lags until the correlations are negligible.
- **(b)** Calculate and plot the sample ACF for your simulated series. How well do the values and patterns match the theoretical ACF from part (a)?
- **(c)** Calculate and interpret the sample EACF for this series. Does the EACF help you specify the correct orders for the model?
- **(d)** Repeat parts (b) and (c) with a new simulation using the same parameter values and sample size.
- **(e)** Repeat parts (b) and (c) with a new simulation using the same parameter values but sample size $n = 36$.
- **(f)** Repeat parts (b) and (c) with a new simulation using the same parameter values but sample size $n = 120$.

6.30 Simulate a mixed ARMA(1,1) model of length $n = 100$ with $\phi = 0.8$ and $\theta = 0.4$.
- **(a)** Calculate and plot the theoretical autocorrelation function for this model. Plot sufficient lags until the correlations are negligible.
- **(b)** Calculate and plot the sample ACF for your simulated series. How well do the values and patterns match the theoretical ACF from part (a)?
- **(c)** Calculate and interpret the sample EACF for this series. Does the EACF help you specify the correct orders for the model?
- **(d)** Repeat parts (b) and (c) with a new simulation using the same parameter values and sample size.
- **(e)** Repeat parts (b) and (c) with a new simulation using the same parameter values but sample size $n = 48$.
- **(f)** Repeat parts (b) and (c) with a new simulation using the same parameter values but sample size $n = 200$.

6.31 Simulate a nonstationary time series with $n = 60$ according to the model ARIMA(0,1,1) with $\theta = 0.8$.

(a) Perform the (augmented) Dickey-Fuller test on the series with $k = 0$ in Equation (6.4.1) on page 128. (With $k = 0$, this is the Dickey-Fuller test and is not augmented.) Comment on the results.

(b) Perform the augmented Dickey-Fuller test on the series with k chosen by the software—that is, the "best" value for k. Comment on the results.

(c) Repeat parts (a) and (b) but use the differences of the simulated series. Comment on the results. (Here, of course, you should reject the unit root hypothesis.)

6.32 Simulate a stationary time series of length $n = 36$ according to an AR(1) model with $\phi = 0.95$. This model is stationary, but just barely so. With such a series and a short history, it will be difficult if not impossible to distinguish between stationary and nonstationary with a unit root.

(a) Plot the series and calculate the sample ACF and PACF and describe what you see.

(b) Perform the (augmented) Dickey-Fuller test on the series with $k = 0$ in Equation (6.4.1) on page 128. (With $k = 0$ this is the Dickey-Fuller test and is not augmented.) Comment on the results.

(c) Perform the augmented Dickey-Fuller test on the series with k chosen by the software—that is, the "best" value for k. Comment on the results.

(d) Repeat parts (a), (b), and (c) but with a new simulation with $n = 100$.

6.33 The data file named deere1 contains 82 consecutive values for the amount of deviation (in 0.000025 inch units) from a specified target value that an industrial machining process at Deere & Co. produced under certain specified operating conditions.

(a) Display the time series plot of this series and comment on any unusual points.

(b) Calculate the sample ACF for this series and comment on the results.

(c) Now replace the unusual value by a much more typical value and recalculate the sample ACF. Comment on the change from what you saw in part (b).

(d) Calculate the sample PACF based on the revised series that you used in part (c). What model would you specify for the revised series? (Later we will investigate other ways to handle outliers in time series modeling.)

6.34 The data file named deere2 contains 102 consecutive values for the amount of deviation (in 0.0000025 inch units) from a specified target value that another industrial machining process produced at Deere & Co.

(a) Display the time series plot of this series and comment on its appearance. Would a stationary model seem to be appropriate?

(b) Display the sample ACF and PACF for this series and select tentative orders for an ARMA model for the series.

6.35 The data file named deere3 contains 57 consecutive measurements recorded from a complex machine tool at Deere & Co. The values given are deviations from a target value in units of ten millionths of an inch. The process employs a control mechanism that resets some of the parameters of the machine tool depending on the magnitude of deviation from target of the last item produced.
(a) Display the time series plot of this series and comment on its appearance. Would a stationary model be appropriate here?
(b) Display the sample ACF and PACF for this series and select tentative orders for an ARMA model for the series.

6.36 The data file named robot contains a time series obtained from an industrial robot. The robot was put through a sequence of maneuvers, and the distance from a desired ending point was recorded in inches. This was repeated 324 times to form the time series.
(a) Display the time series plot of the data. Based on this information, do these data appear to come from a stationary or nonstationary process?
(b) Calculate and plot the sample ACF and PACF for these data. Based on this additional information, do these data appear to come from a stationary or nonstationary process?
(c) Calculate and interpret the sample EACF.
(d) Use the best subsets ARMA approach to specify a model for these data. Compare these results with what you discovered in parts (a), (b), and (c).

6.37 Calculate and interpret the sample EACF for the logarithms of the Los Angeles rainfall series. The data are in the file named larain. Do the results confirm that the logs are white noise?

6.38 Calculate and interpret the sample EACF for the color property time series. The data are in the color file. Does the sample EACF suggest the same model that was specified by looking at the sample PACF?

6.39 The data file named days contains accounting data from the Winegard Co. of Burlington, Iowa. The data are the number of days until Winegard receives payment for 130 consecutive orders from a particular distributor of Winegard products. (The name of the distributor must remain anonymous for confidentiality reasons.)
(a) Plot the time series, and comment on the display. Are there any unusual values?
(b) Calculate the sample ACF and PACF for this series.
(c) Now replace each of the unusual values with a value of 35 days—much more typical values—and repeat the calculation of the sample ACF and PACF. What ARMA model would you specify for this series after removing the outliers? (Later we will investigate other ways to handle outliers in time series modeling.)