11 Further Topics

In this chapter we touch on a variety of topics of special interest. In Section 11.1 we consider transfer function models, designed to exploit for predictive purposes the relationship between two time series when one acts as a leading indicator for the other. Section 11.2 deals with intervention analysis, which allows for possible changes in the mechanism generating a time series, causing it to have different properties over different time intervals. In Section 11.3 we introduce the very fast growing area of nonlinear time series analysis, and in Section 11.4 we discuss fractionally integrated ARMA processes, sometimes called "long-memory" processes on account of the slow rate of convergence of their autocorrelation functions to zero as the lag increases. In Section 11.5 we discuss continuous-time ARMA processes which, for continuously evolving processes, play a role analogous to that of ARMA processes in discrete time. Besides being of interest in their own right, they have proved a useful class of models in the representation of financial time series and in the modeling of irregularly spaced data.

11.1 Transfer Function Models

In this section we consider the problem of estimating the transfer function of a linear filter when the output includes added uncorrelated noise. Suppose that $\{X_{t1}\}$ and $\{X_{t2}\}$ are, respectively, the input and output of the transfer function model

$$X_{t2} = \sum_{j=0}^{\infty} \tau_j X_{t-j,\,1} + N_t, \tag{11.1.1}$$

P.J. Brockwell, R.A. Davis, *Introduction to Time Series and Forecasting*,
Springer Texts in Statistics, DOI 10.1007/978-3-319-29854-2_11

where $T = \{\tau_j, j = 0, 1, \ldots\}$ is a causal time-invariant linear filter and $\{N_t\}$ is a zero-mean stationary process, uncorrelated with the input process $\{X_{t1}\}$. We further assume that $\{X_{t1}\}$ is a zero-mean stationary time series. Then the bivariate process $\{(X_{t1}, X_{t2})'\}$ is also stationary. Multiplying each side of (11.1.1) by $X_{t-k,1}$ and then taking expectations gives the equation

$$\gamma_{21}(k) = \sum_{j=0}^{\infty} \tau_j \gamma_{11}(k-j). \tag{11.1.2}$$

Equation (11.1.2) simplifies a great deal if the input process happens to be white noise. For example, if $\{X_{t1}\} \sim \mathrm{WN}(0, \sigma_1^2)$, then we can immediately identify t_k from (11.1.2) as

$$\tau_k = \gamma_{21}(k)/\sigma_1^2. \tag{11.1.3}$$

This observation suggests that "prewhitening" of the input process might simplify the identification of an appropriate transfer function model and at the same time provide simple preliminary estimates of the coefficients t_k.

If $\{X_{t1}\}$ can be represented as an invertible ARMA(p, q) process

$$\phi(B)X_{t1} = \theta(B)Z_t, \qquad \{Z_t\} \sim \mathrm{WN}\left(0, \sigma_Z^2\right), \tag{11.1.4}$$

then application of the filter $\pi(B) = \phi(B)\theta^{-1}(B)$ to $\{X_{t1}\}$ will produce the whitened series $\{Z_t\}$. Now applying the operator $\pi(B)$ to each side of (11.1.1) and letting $Y_t = \pi(B)X_{t2}$, we obtain the relation

$$Y_t = \sum_{j=0}^{\infty} \tau_j Z_{t-j} + N_t',$$

where

$$N_t' = \pi(B)N_t,$$

and $\{N_t'\}$ is a zero-mean stationary process, uncorrelated with $\{Z_t\}$. The same arguments that led to (11.1.3) therefore yield the equation

$$\tau_j = \rho_{YZ}(j)\sigma_Y/\sigma_Z, \tag{11.1.5}$$

where ρ_{YZ} is the cross-correlation function of $\{Y_t\}$ and $\{Z_t\}$, σ_Z^2 =Var(Z_t), and σ_Y^2 =Var(Y_t).

Given the observations $\{(X_{t1}, X_{t2})', t = 1, \ldots, n\}$, the results of the previous paragraph suggest the following procedure for estimating $\{\tau_j\}$ and analyzing the noise $\{N_t\}$ in the model (11.1.1):

1. Fit an ARMA model to $\{X_{t1}\}$ and file the residuals $(\hat{Z}_1, \ldots, \hat{Z}_n)$ (using the `Export` button in ITSM to copy them to the clipboard and then pasting them into the first column of an Excel file). Let $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\theta}}$ denote the maximum likelihood estimates of the autoregressive and moving-average parameters and let $\hat{\sigma}_Z^2$ be the maximum likelihood estimate of the variance of $\{Z_t\}$.
2. Apply the operator $\hat{\pi}(B) = \hat{\phi}(B)\hat{\theta}^{-1}(B)$ to $\{X_{t2}\}$ to obtain the series $\left(\hat{Y}_1, \ldots, \hat{Y}_n\right)$. (After fitting the ARMA model as in Step 1 above, highlight the window containing the graph of $\{X_t\}$ and replace $\{X_t\}$ by $\{Y_t\}$ using the option `File>Import`. The residuals are then automatically replaced by the residuals of $\{Y_t\}$ under the model already fitted to $\{X_t\}$.) Export the new residuals to the clipboard, paste them into the second column of the Excel file created in Step 1, and save this as a text file, FNAME.TSM. The file FNAME.TSM then contains the bivariate series $\{(Z_t, Y_t)\}$. Let $\hat{\sigma}_Y^2$ denote the sample variance of $\hat{Y}_t$.

3. Compute the sample auto- and cross-correlation functions of $\{Z_t\}$ and $\{Y_t\}$ by opening the bivariate project FNAME.TSM in ITSM and clicking on the second yellow button at the top of the ITSM window. Comparison of $\hat{\rho}_{YZ}(h)$ with the bounds $\pm 1.96n^{-1/2}$ gives a preliminary indication of the lags h at which $\rho_{YZ}(h)$ is significantly different from zero. A more refined check can be carried out by using Bartlett's formula in Section 8.3.4 for the asymptotic variance of $\hat{\rho}_{YZ}(h)$. Under the assumptions that $\{\hat{Z}_t\} \sim \mathrm{WN}\left(0, \hat{\sigma}_Z^2\right)$ and $\left\{\left(\hat{Y}_t, \hat{Z}_t\right)'\right\}$ is a stationary Gaussian process,

$$n\mathrm{Var}(\hat{\rho}_{YZ}(h)) \sim 1 - \rho_{YZ}^2(h)\left[1.5 - \sum_{k=-\infty}^{\infty}(\rho_{YZ}^2(k) + \rho_{YY}^2(k)/2)\right]$$
$$+ \sum_{k=-\infty}^{\infty}\left[\rho_{YZ}(h+k)\rho_{YZ}(h-k) - 2\rho_{YZ}(h)\rho_{YZ}(k+h)\rho_{YY}^2(k)\right].$$

In order to check the hypothesis H_0 that $\rho_{YZ}(h) = 0, h \notin [a, b]$, where a and b are integers, we note from Corollary 8.3.1 that under H_0,

$$\mathrm{Var}\left(\hat{\rho}_{YZ}(h)\right) \sim n^{-1} \qquad \text{for } h \notin [a, b].$$

We can therefore check the hypothesis H_0 by comparing $\hat{\rho}_{YZ}, h \notin [a, b]$, with the bounds $\pm 1.96n^{-1/2}$. Observe that $\rho_{ZY}(h)$ should be zero for $h > 0$ if the model (11.1.1) is valid.

4. Preliminary estimates of τ_h for the lags h at which $\hat{\rho}_{YZ}(h)$ is significantly different from zero are

$$\hat{\tau}_h = \hat{\rho}_{YZ}(h)\hat{\sigma}_Y/\hat{\sigma}_Z.$$

For other values of h the preliminary estimates are $\hat{\tau}_h = 0$. The numerical values of the cross-correlations $\hat{\rho}_{YZ}(h)$ are found by right-clicking on the graphs of the sample correlations plotted in Step 3 and then on `Info`. The values of $\hat{\sigma}_Z$ and $\hat{\sigma}_Y$ are found by doing the same with the graphs of the series themselves. Let $m \geq 0$ be the largest value of j such that $\hat{\tau}_j$ is nonzero and let $b \geq 0$ be the smallest such value. Then b is known as the *delay parameter* of the filter $\{\hat{\tau}_j\}$. If m is very large and if the coefficients $\{\hat{\tau}_j\}$ are approximately related by difference equations of the form

$$\hat{\tau}_j - v_1\hat{\tau}_{j-1} - \cdots - v_p\hat{\tau}_{j-p} = 0, \qquad j \geq b + p,$$

then $\hat{T}(B) = \sum_{j=b}^{m} \hat{\tau}_j B^j$ can be represented approximately, using fewer parameters, as

$$\hat{T}(B) = w_0(1 - v_1 B - \cdots - v_p B_p)^{-1}B^b.$$

In particular, if $\hat{\tau}_j = 0,\ j < b$, and $\hat{\tau}_j = w_0 v_1^{j-b},\ j \geq b$, then

$$\hat{T}(B) = w_0(1 - v_1 B)^{-1}B^b. \tag{11.1.6}$$

Box and Jenkins (1976) recommend choosing $\hat{T}(B)$ to be a ratio of two polynomials. However, the degrees of the polynomials are often difficult to estimate from $\{\hat{\tau}_j\}$. The primary objective at this stage is to find a parametric function that provides an adequate approximation to $\hat{T}(B)$ without introducing too large a number of parameters. If $\hat{T}(B)$ is represented as $\hat{T}(B) = B^b w(B) v^{-1}(B) = B^b\left(w_0 + w_1 B + \cdots + w_q B^q\right)\left(1 - v_1 B - \cdots - v_p B^p\right)^{-1}$ with $v(z) \neq 0$ for $|z| \leq 1$, then we define $m = \max(q + b,\ p)$.

5. The noise sequence $\{N_t, t = m+1, \ldots, n\}$ is estimated as

$$\hat{N}_t = X_{t2} - \hat{T}(B)X_{t1}.$$

(We set $\hat{N}_t = 0, t \le m$, in order to compute $\hat{N}_t, t > m = \max(b+q, p)$). The calculations are done in ITSM by opening the bivariate file containing $\{(X_{t1}, X_{t2})\}$, selecting `Transfer>Specify Model`, and entering the preliminary model found in Step 4. Click on the fourth green button to see a graph of the residuals $\{N_t\}$. These should then be filed as, say, NOISE.TSM.

6. Preliminary identification of a suitable model for the noise sequence is carried out by fitting a causal invertible ARMA model

$$\phi^{(N)}(B)N_t = \theta^{(N)}(B)W_t, \qquad \{W_t\} \sim \text{WN}\left(0, \sigma_W^2\right), \tag{11.1.7}$$

to the estimated noise $\hat{N}_{m+1}, \ldots, \hat{N}_n$ filed as NOISE.TSM in Step 5.

7. At this stage we have the preliminary model

$$\phi^{(N)}(B)v(B)X_{t2} = B^b\phi^{(N)}(B)w(B)X_{t1} + \theta^{(N)}(B)v(B)W_t,$$

where $\hat{T}(B) = B^b w(B)v^{-1}(B)$ as in step (4). For this model we can compute $\hat{W}_t\left(\mathbf{w}, \mathbf{v}, \boldsymbol{\phi}^{(N)}, \boldsymbol{\theta}^{(N)}\right)$, $t > m^* = \max(p_2 + p, b + p_2 + q)$, by setting $\hat{W}_t = 0$ for $t \le m^*$. The parameters $\mathbf{w}$, $\mathbf{v}$, $\boldsymbol{\phi}^{(N)}$, and $\boldsymbol{\theta}^{(N)}$ can then be reestimated (more efficiently) by minimizing the sum of squares

$$\sum_{t=m^*+1}^{n} \hat{W}_t^2\left(\mathbf{w}, \mathbf{v}, \boldsymbol{\phi}^{(N)}, \boldsymbol{\theta}^{(N)}\right).$$

(The calculations are performed in ITSM by opening the bivariate project $\{(X_{t1}, X_{t2})\}$, selecting `Transfer>Specify model`, entering the preliminary model, and clicking `OK`. Then choose `Transfer>Estimation`, click `OK`, and the least squares estimates of the parameters will be computed. Pressing the fourth green button at the top of the screen will give a graph of the estimated residuals $\hat{W}_t$.)

8. To test for goodness of fit, the estimated residuals $\left\{\hat{W}_t, t > m^*\right\}$ and $\left\{\hat{Z}_t.t > m^*\right\}$ should be filed as a bivariate series and the auto- and cross correlations compared with the bounds $\pm 1.96/\sqrt{n}$ in order to check the hypothesis that the two series are uncorrelated white noise sequences. Alternative models can be compared using the AICC value that is printed with the estimated parameters in Step 7. It is computed from the exact Gaussian likelihood, which is computed using a state-space representation of the model, described in Brockwell and Davis (1991), Section 13.1.

Example 11.1.1 Sales with a Leading Indicator

In this example we fit a transfer function model to the bivariate time series of Example 8.1.2. Let

$$X_{t1} = (1-B)Y_{t1} - 0.0228, \quad t = 2, \ldots, 150,$$

$$X_{t2} = (1-B)Y_{t2} - 0.420, \quad t = 2, \ldots, 150,$$

where $\{Y_{t1}\}$ and $\{Y_{t2}\}$, $t = 1, \ldots, 150$, are the leading indicator and sales data, respectively. It was found in Example 8.1.2 that $\{X_{t1}\}$ and $\{X_{t2}\}$ can be modeled as low-order zero-mean ARMA processes. In particular, we fitted the model

$$X_{t1} = (1 - 0.474B)Z_t, \qquad \{Z_t\} \sim \text{WN}(0, 0.0779),$$

to the series $\{X_{t1}\}$. We can therefore whiten the series by application of the filter $\hat{\pi}(B) = (1 - 0.474B)^{-1}$. Applying $\hat{\pi}(B)$ to both $\{X_{t1}\}$ and $\{X_{t2}\}$ we obtain

$$\hat{Z}_t = (1 - 0.474B)^{-1}X_{t1}, \quad \hat{\sigma}_Z^2 = 0.0779,$$

$$\hat{Y}_t = (1 - 0.474B)^{-1}X_{t2}, \quad \hat{\sigma}_Y^2 = 4.0217.$$

These calculations and the filing of the series $\{\hat{Z}_t\}$ and $\{\hat{Y}_t\}$ were carried out using ITSM as described in steps (1) and (2). Their sample auto- and cross-correlations, found as described in step (3), are shown in Figure 11-1. The cross-correlations $\hat{\rho}_{ZY}(h)$ (top right) and $\hat{\rho}_{YZ}(h)$ (bottom left), when compared with the upper and lower bounds $\pm 1.96(149)^{-1/2} = \pm 0.161$, strongly suggest a transfer function model for $\{X_{t2}\}$ in terms of $\{X_{t1}\}$ with delay parameter 3. Since $\hat{\tau}_j = \hat{\rho}_{YZ}(j)\hat{\sigma}_Y/\hat{\sigma}_Z$ is decreasing approximately geometrically for $j \geq 3$, we take $T(B)$ to have the form (11.1.6), i.e.,

$$T(B) = w_0(1 - v_1B)^{-1}B^3.$$

The preliminary estimates of w_0 and v_1 are $\hat{w}_0 = \hat{\tau}_3 = 4.86$ and $\hat{v}_1 = \hat{\tau}_4/\hat{\tau}_3 = 0.698$, the coefficients τ_j being estimated as described in step (4). The estimated noise sequence is determined and filed using ITSM as described in step (5). It satisfies the equations

$$\hat{N}_t = X_{t2} - 4.86B^3(1 - 0.698B)^{-1}X_{t1}, \qquad t = 5, 6, \ldots, 150.$$

Analysis of this univariate series with ITSM gives the MA(1) model

$$N_t = (1 - 0.364B)W_t, \qquad \{W_t\} \sim \text{WN}(0, 0.0590).$$

Substituting these preliminary noise and transfer function models into equation (11.1.1) then gives

$$X_{t2} = 4.86B^3(1-0.698B)^{-1}X_{t1} + (1-0.364B)W_t, \quad \{W_t\} \sim \text{WN}(0, 0.0590).$$

Now minimizing the sum of squares (11.1.7) with respect to the parameters $\left(w_0, v_1, \theta_1^{(N)}\right)$ as described in step (7), we obtain the least squares model

$$X_{t2} = 4.717B^3(1 - 0.724B)^{-1}X_{t1} + (1 - 0.582B)W_t, \tag{11.1.8}$$

where $\{W_t\} \sim \text{WN}(0, 0.0486)$ and

$$X_{t1} = (1 - 0.474B)Z_t, \quad \{Z_t\} \sim \text{WN}(0, 0.0779).$$

Notice the reduced white noise variance of $\{W_t\}$ in the least squares model as compared with the preliminary model.

The sample auto- and cross-correlation functions of the series $\hat{Z}_t$ and $\hat{W}_t$, $t = 5, \ldots, 150$, are shown in Figure 11-2. All of the correlations lie between the bounds $\pm 1.96/\sqrt{144}$, supporting the assumption underlying the fitted model that the residuals are uncorrelated white noise sequences.

□

11.1.1 Prediction Based on a Transfer Function Model

When predicting $X_{n+h,2}$ on the basis of the transfer function model defined by (11.1.1), (11.1.4), and (11.1.7), with observations of X_{t1} and X_{t2}, $t = 1, \ldots, n$, our aim is to find the linear combination of $1, X_{11}, \ldots, X_{n1}, X_{12}, \ldots, X_{n2}$ that predicts $X_{n+h,2}$ with minimum mean squared error. The exact solution of this problem can be found with the help of the Kalman recursions (see Brockwell and Davis (1991), Section 13.1 for details). The program ITSM uses these recursions to compute the predictors and their mean squared errors.

In order to provide a little more insight, we give here the predictors $\tilde{P}_nX_{n+h}$ and mean squared errors based on infinitely many past observations X_{t1} and X_{t2},

$-\infty < t \le n$. These predictors and their mean squared errors will be close to those based on X_{t1} and X_{t2}, $1 \le t \le n$, if n is sufficiently large.

The transfer function model defined by (11.1.1), (11.1.4), and (11.1.7) can be rewritten as

$$X_{t2} = T(B)X_{t1} + \beta(B)W_t, \tag{11.1.9}$$

$$X_{t1} = \theta(B)\phi^{-1}(B)Z_t, \tag{11.1.10}$$

where $\beta(B) = \theta^{(N)}(B)/\phi^{(N)}(B)$. Eliminating X_{t1} gives

$$X_{t2} = \sum_{j=0}^{\infty} \alpha_j Z_{t-j} + \sum_{j=0}^{\infty} \beta_j W_{t-j}, \tag{11.1.11}$$

where $\alpha(B) = T(B)\theta(B)/\phi(B)$.

Noting that each limit of linear combinations of $\{X_{t1}, X_{t2}, -\infty < t \le n\}$ is a limit of linear combinations of $\{Z_t, W_t, -\infty < t \le n\}$ and conversely and that $\{Z_t\}$ and $\{W_t\}$ are uncorrelated, we see at once from (11.1.11) that

$$\tilde{P}_n X_{n+h,2} = \sum_{j=h}^{\infty} \alpha_j Z_{n+h-j} + \sum_{j=h}^{\infty} \beta_j W_{n+h-j}. \tag{11.1.12}$$

Setting $t = n + h$ in (11.1.11) and subtracting (11.1.12) gives the mean squared error

$$E\left(X_{n+h,2} - \tilde{P}_n X_{n+h,2}\right)^2 = \sigma_Z^2 \sum_{j=0}^{h-1} \alpha_j^2 + \sigma_W^2 \sum_{j=0}^{h-1} \beta_j^2. \tag{11.1.13}$$

To compute the predictors $\tilde{P}_n X_{n+h,2}$ we proceed as follows. Rewrite (11.1.9) as

$$A(B)X_{t2} = B^b U(B)X_{t1} + V(B)W_t, \tag{11.1.14}$$

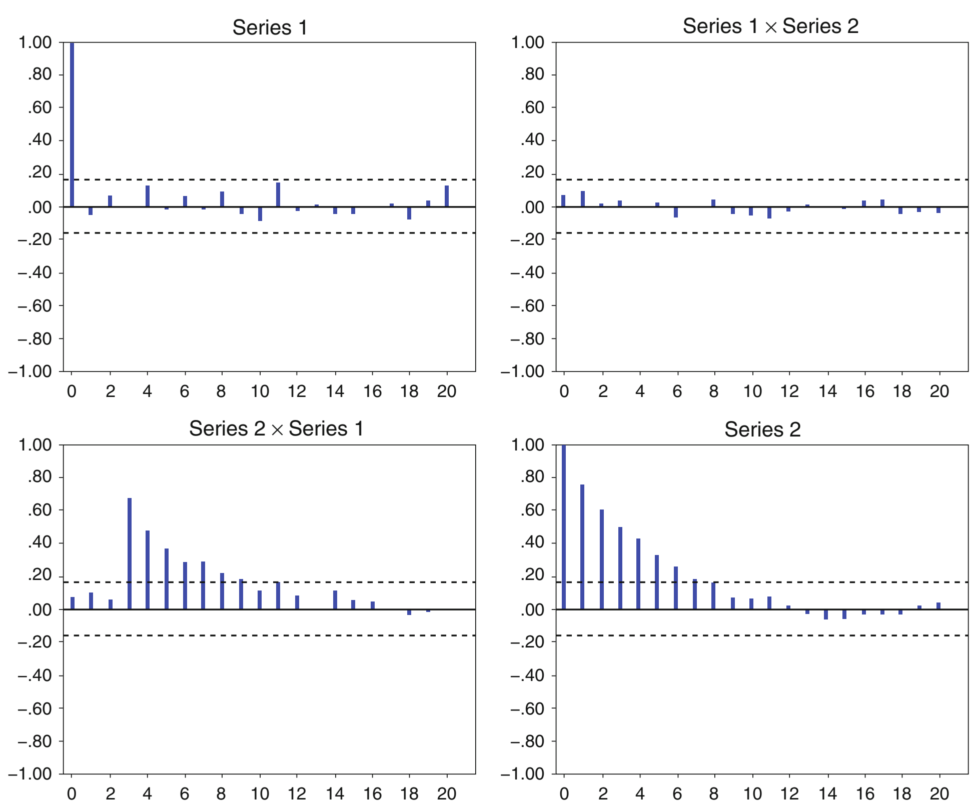

Figure 11-1 The sample correlation functions $\hat{\rho}_{ij}(h)$, of Example 11.1.1. Series 1 is $\{\hat{Z}_t\}$ and Series 2 is $\{\hat{Y}_t\}$

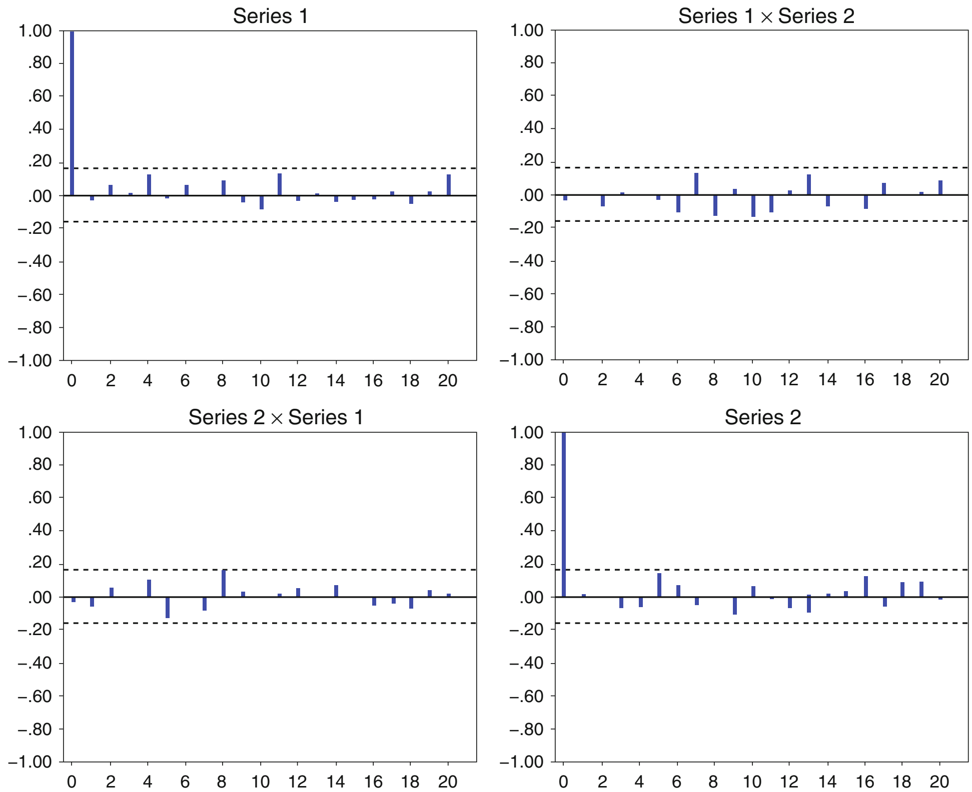

Figure 11-2 The sample correlation functions of the estimated residuals from the model fitted in Example 11.1.1. Series 1 is $\{\hat{Z}_t\}$ and Series 2 is $\{\hat{W}_t\}$

where A, U, and V are polynomials of the form

$$A(B) = 1 - A_1 B - \cdots - A_a B^a,$$

$$U(B) = U_0 + U_1 B + \cdots + U_u B^u,$$

$$V(B) = 1 + V_1 B + \cdots + V_v B^v.$$

Applying the operator $\tilde{P}_n$ to equation (11.1.14) with $t = n + h$, we obtain

$$\tilde{P}_n X_{n+h,2} = \sum_{j=1}^{a} A_j \tilde{P}_n X_{n+h-j,2} + \sum_{j=0}^{u} U_j \tilde{P}_n X_{n+h-b-j,1} + \sum_{j=h}^{v} V_j W_{n+h-j}, \qquad (11.1.15)$$

where the last sum is zero if $h > v$.

Since $\{X_{t1}\}$ is uncorrelated with $\{W_t\}$, the predictors appearing in the second sum in (11.1.15) are therefore obtained by predicting the univariate series $\{X_{t1}\}$ as described in Section 3.3 using the model (11.1.10). In keeping with our assumption that n is large, we can replace $\tilde{P}_n X_{j1}$ for each j by the finite-past predictor obtained from the program ITSM. The values $W_j, j \le n$, are replaced by their estimated values $\hat{W}_j$ from the least squares estimation in step (7) of the modeling procedure.

Equation (11.1.15) can now be solved recursively for the predictors $\tilde{P}_n X_{n+1,2}$, $\tilde{P}_n X_{n+2,2}$, $\tilde{P}_n X_{n+3,2}, \ldots$.

Example 11.1.2 Sales with a Leading Indicator

Applying the preceding results to the series $\{X_{t1}, X_{t2}, 2 \le t \le 150\}$ of Example 11.1.1, and using the values $X_{148,1} = -0.093$, $X_{150,2} = 0.08$, $\hat{W}_{150} = -0.0706$, $\hat{W}_{149} = 0.1449$, we find from (11.1.8) and (11.1.15) that

$$\tilde{P}_{150} X_{151,2} = 0.724 X_{150,2} + 4.717 X_{148,1} - 1.306 W_{150} + 0.421 W_{149} = -0.228$$

and, using the value $X_{149,1} = 0.237$, that

$$\tilde{P}_{150}X_{152,2} = 0.724\tilde{P}_{150}X_{151,2} + 4.717X_{149,1} + 0.421W_{150} = 0.923.$$

In terms of the original sales data $\{Y_{t2}\}$ we have $Y_{149,2} = 262.7$ and

$$Y_{t2} = Y_{t-1,2} + X_{t2} + 0.420.$$

Hence the predictors of actual sales are

$$P^*_{150}Y_{151,\,2} = 262.70 - 0.228 + 0.420 = 262.89,$$
$$P^*_{150}Y_{152,\,2} = 262.89 + 0.923 + 0.420 = 264.23,$$

where P^*_{149} is based on $\{1, Y_{11}, Y_{12}, X_{s1}, X_{s2}, -\infty < s \le 150\}$, and it is assumed that Y_{11} and Y_{12} are uncorrelated with $\{X_{s1}\}$ and with $\{X_{s2}\}$. The predicted values are in close agreement with those based on the finite number of available observations that are computed by ITSM. Since our model for the sales data is

$$(1-B)Y_{t2} = 0.420+4.717B^3(1-0.474B)(1-0.724B)^{-1}Z_t+(1-0.582B)W_t,$$

it can be shown, using an argument analogous to that which gave (11.1.13), that the mean squared errors are given by

$$E(Y_{150+h,2} - P_{150}Y_{150+h,2})^2 = \sigma_Z^2 \sum_{j=0}^{h-1} \alpha_j^{*2} + \sigma_W^2 \sum_{j=0}^{h-1} \beta_j^{*2},$$

where

$$\sum_{j=0}^{\infty} \alpha_j^* z^j = 4.717z^3(1 - 0.474z)(1 - 0.724z)^{-1}(1 - z)^{-1}$$

and

$$\sum_{j=0}^{\infty} \beta_j^* z^j = (1 - 0.582z)(1 - z)^{-1}.$$

For $h = 1$ and 2 we obtain

$$E(Y_{151,\,2} - P^*_{150}Y_{151,\,2})^2 = 0.0486,$$
$$E(Y_{152,\,2} - P^*_{150}Y_{152,\,2})^2 = 0.0570,$$

in close agreement with the finite-past mean squared errors obtained by ITSM.

It is interesting to examine the improvement obtained by using the transfer function model rather than fitting a univariate model to the sales data alone. If we adopt the latter course, we obtain the model

$$X_{t2} - 0.249X_{t-1,\,2} - 0.199X_{t-2,2} = U_t,$$

where $\{U_t\} \sim \text{WN}(0, 1.794)$ and $X_{t2} = Y_{t2} - Y_{t-1,2} - 0.420$. The corresponding predictors of $Y_{151,\,2}$ and $Y_{152,\,2}$ are easily found from the program ITSM to be 263.14 and 263.58 with mean squared errors 1.794 and 4.593, respectively. These mean squared errors are much worse than those obtained using the transfer function model. □

11.2 Intervention Analysis

During the period for which a time series is observed, it is sometimes the case that a change occurs that affects the level of the series. A change in the tax laws may, for example, have a continuing effect on the daily closing prices of shares on the stock market. In the same way construction of a dam on a river may have a dramatic effect on the time series of streamflows below the dam. In the following we shall assume that the time T at which the change (or "intervention") occurs is known.

To account for such changes, Box and Tiao (1975) introduced a model for intervention analysis that has the same form as the transfer function model

$$Y_t = \sum_{j=0}^{\infty} \tau_j X_{t-j} + N_t, \tag{11.2.1}$$

except that the input series $\{X_t\}$ is not a random series but a deterministic function of t. It is clear from (11.2.1) that $\sum_{j=0}^{\infty} \tau_j X_{t-j}$ is then the mean of Y_t. The function $\{X_t\}$ and the coefficients $\{\tau_j\}$ are therefore chosen in such a way that the changing level of the observations of $\{Y_t\}$ is well represented by the sequence $\sum_{j=0}^{\infty} \tau_j X_{t-j}$. For a series $\{Y_t\}$ with $EY_t = 0$ for $t \le T$ and $EY_t \to 0$ as $t \to \infty$, a suitable input series is

$$X_t = I_t(T) = \begin{cases} 1 & \text{if } t = T, \\ 0 & \text{if } t \ne T. \end{cases} \tag{11.2.2}$$

For a series $\{Y_t\}$ with $EY_t = 0$ for $t \le T$ and $EY_t \to a \ne 0$ as $t \to \infty$, a suitable input series is

$$X_t = H_t(T) = \sum_{k=T}^{\infty} I_t(k) = \begin{cases} 1 & \text{if } t \ge T, \\ 0 & \text{if } t < T. \end{cases} \tag{11.2.3}$$

(Other deterministic input functions $\{X_t\}$ can also be used, for example when interventions occur at more than one time.) The function $\{X_t\}$ having been selected by inspection of the data, the determination of the coefficients $\{\tau_j\}$ in (11.2.1) then reduces to a regression problem in which the errors $\{N_t\}$ constitute an ARMA process. This problem can be solved using the program ITSM as described below.

The goal of intervention analysis is to estimate the effect of the intervention as indicated by the term $\sum_{j=0}^{\infty} \tau_j X_{t-j}$ and to use the resulting model (11.2.1) for forecasting. For example, Wichern and Jones (1978) used intervention analysis to investigate the effect of the American Dental Association's endorsement of Crest toothpaste on Crest's market share. Other applications of intervention analysis can be found in Box and Tiao (1975), Atkins (1979), and Bhattacharyya and Layton (1979). A more general approach can also be found in West and Harrison (1989), Harvey (1990), and Pole et al. (1994).

As in the case of transfer function modeling, once $\{X_t\}$ has been chosen (usually as either (11.2.2) or (11.2.3)), estimation of the linear filter $\{\tau_j\}$ in (11.2.1) is simplified by approximating the operator $T(B) = \sum_{j=0}^{\infty} \tau_j B^j$ with a rational operator of the form

$$T(B) = \frac{B^b W(B)}{V(B)}, \tag{11.2.4}$$

where b is the delay parameter and $W(B)$ and $V(B)$ are polynomials of the form

$$W(B) = w_0 + w_1 B + \cdots + w_q B^q$$

and

$$V(B) = 1 - v_1 B - \cdots - v_p B^p.$$

By suitable choice of the parameters b, q, p and the coefficients w_i and v_j, the intervention term $T(B)X_t$ can made to take a great variety of functional forms.

For example, if $T(B) = wB^2/(1 - vB)$ and $X_t = I_t(T)$ as in (11.2.2), the resulting intervention term is

$$\frac{wB^2}{(1 - vB)} I_t(T) = \sum_{j=0}^{\infty} v^j w I_{t-j-2}(T) = \sum_{j=0}^{\infty} v^j w I_t(T + 2 + j),$$

a series of pulses of sizes $v^j w$ at times $T + 2 + j$, $j = 0, 1, 2, \ldots$. If $|v| < 1$, the effect of the intervention is to add a series of pulses with size w at time $T + 2$, decreasing to zero at a geometric rate depending on v as $t \to \infty$. Similarly, with $X_t = H_t(T)$ as in (11.2.3),

$$\frac{wB^2}{(1 - vB)} H_t(T) = \sum_{j=0}^{\infty} v^j w H_{t-j-2}(T) = \sum_{j=0}^{\infty} (1 + v + \cdots + v^j) w I_t(T + 2 + j),$$

a series of pulses of sizes $(1 + v + \cdots + v^j)w$ at times $T + 2 + j$, $j = 0, 1, 2, \ldots$. If $|v| < 1$, the effect of the intervention is to bring about a shift in level of the series X_t, the size of the shift converging to $w/(1 - v)$ as $t \to \infty$.

An appropriate form for X_t and possible values of b, q, and p having been chosen by inspection of the data, the estimation of the parameters in (11.2.4) and the fitting of the model for $\{N_t\}$ can be carried out using steps (6)–(8) of the transfer function modeling procedure described in Section 11.1. Start with step (7) and assume that $\{N_t\}$ is white noise to get preliminary estimates of the coefficients w_i and v_j by least squares. The residuals are filed and used as estimates of $\{N_t\}$. Then go to step (6) and continue exactly as for transfer function modeling with input series $\{X_t\}$ and output series $\{Y_t\}$.

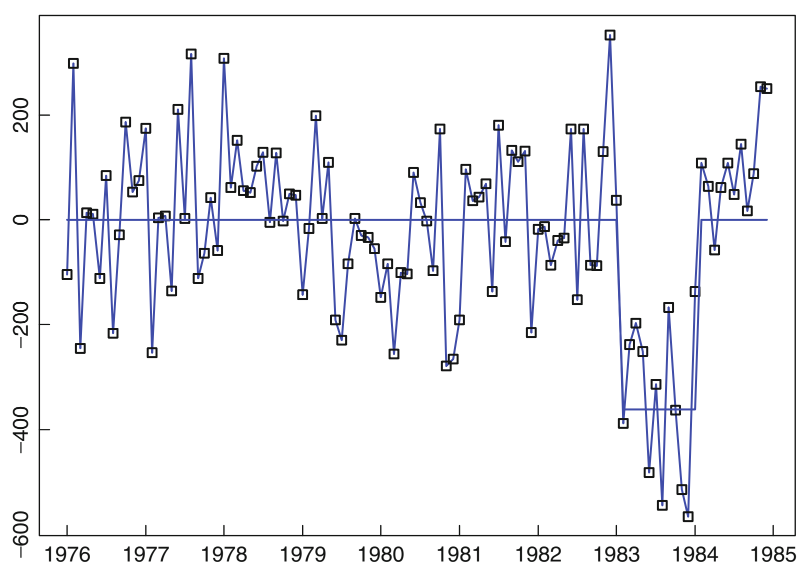

Figure 11-3 The differenced series of Example 11.2.1 (showing also the fitted intervention term accounting for the seat-belt legislation of 1983)

Example 11.2.1 Seat-Belt Legislation

In this example we reanalyze the seat-belt legislation data, SBL.TSM of Example 6.6.3 from the point of view of intervention analysis. For this purpose the bivariate series $\{(f_t, Y_t)\}$ consisting of the series filed as SBLIN.TSM and SBL.TSM respectively has been saved in the file SBL2.TSM. The *input* series $\{f_t\}$ is the deterministic step-function defined in Example 6.6.3 and Y_t is the number of deaths and serious injuries on UK roads in month t, $t = 1, \ldots, 120$, corresponding to the 10 years beginning with January 1975.

To account for the seat-belt legislation, we use the same model (6.6.15) as in Example 6.6.3 and, because of the apparent non-stationarity of the residuals, we again difference both $\{f_t\}$ and $\{Y_t\}$ at lag 12 to obtain the model (6.6.15), i.e.,

$$X_t = bg_t + N_t, \tag{11.2.4}$$

where $X_t = \nabla_{12} Y_t$, $g_t = \nabla_{12} f_t$, and $\{N_t\}$ is a zero-mean stationary time series. This is a particularly simple example of the general intervention model (11.2.1) for the series $\{X_t\}$ with intervention $\{bg_t\}$. Our aim is to find a suitable model for $\{N_t\}$ and at the same time to estimate b, taking into account the autocorrelation function of the model for $\{N_t\}$. To apply intervention analysis to this problem using ITSM, we proceed as follows:

(1) Open the bivariate project SBL2.TSM and difference the series at lag 12.

(2) Select `Transfer>Specify model` and you will see that the default input and noise are white noise, while the default transfer model relating the input g_t to the output X_t is $X_t = bg_t$ with $b = 1$. Click `OK`, leaving these settings as they are. The input model is irrelevant for intervention analysis and estimation of the transfer function with the default noise model will give us the ordinary least squares estimate of b in the model (10.2.4), with the residuals providing estimates of N_t. Now select `Transfer>Estimation` and click `OK`. You will then see the estimated value -346.9 for b. Finally, press the red Export button (top right in the ITSM window) to export the residuals (estimated values of N_t) to a file and call it, say, NOISE.TSM.

(3) Without closing the bivariate project, open the univariate project NOISE.TSM. The sample ACF and PACF of the series suggests either an MA(13) or AR(13) model. Fitting AR and MA models of order up to 13 (with no mean-correction) using the option `Model>Estimation>Autofit` gives an MA(12) model as the minimum AICC fit.

(4) Return to the bivariate project by highlighting the window labeled SBL2.TSM and select `Transfer>Specify model`. The transfer model will now show the estimated value -346.9 for b. Click on the `Residual Model` tab, enter 12 for the MA order and click `OK`. Select `Transfer>Estimation` and again click `OK`. The parameters in both the noise and transfer models will then be estimated and printed on the screen. Repeating the minimization with decreasing step-sizes, 0.1, 0.01 and then 0.001, gives the model,

$$X_t = -362.5g_t + N_t,$$

where $N_t = W_t + 0.207W_{t-1} + 0.311W_{t-2} + 0.105W_{t-3} + 0.040W_{t-4} + 0.194W_{t-5} + 0.100W_{t-6} + 0.299W_{t-7} + 0.080W_{t-8} + 0.125W_{t-9} + 0.210W_{t-10} + 0.109W_{t-11} + 0.501W_{t-12}$, and $\{W_t\} \sim \text{WN}(0,17289)$. File the residuals (which are now estimates of $\{W_t\}$) as RES.TSM. The differenced series $\{X_t\}$ and the fitted intervention term, $-362.5g_t$, are shown in Figure 11-3.

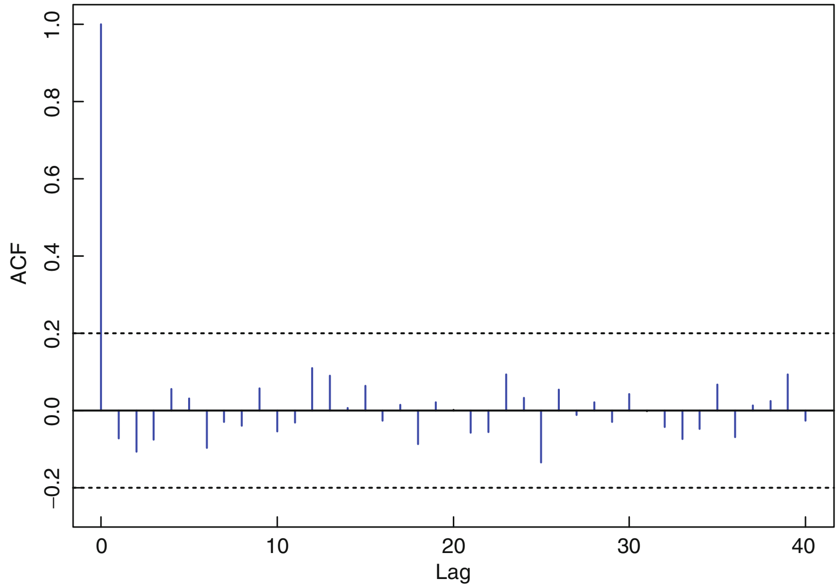

Figure 11-4
The sample ACF of the residuals from the model in Example 11.2.1

(5) Open the univariate project RES.TSM and apply the usual tests for randomness by selecting `Statistics>Residual Analysis`. The tests are all passed at level 0.05, leading us to conclude that the model found in step (4) is satisfactory. The sample ACF of the residuals is shown in Figure 11-4.

□

11.3 Nonlinear Models

A time series of the form

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim \text{IID}\left(0, \sigma^2\right), \tag{11.3.1}$$

where Z_t is expressible as a mean square limit of linear combinations of $\{X_s, \infty < s \leq t\}$, has the property that the best mean square predictor $E(X_{t+h}|X_s, -\infty < s \leq t)$ and the best linear predictor $\tilde{P}_t X_{t+h}$ in terms of $\{X_s, -\infty < s \leq t\}$ are identical. It can be shown that if iid is replaced by WN in (11.3.1), then the two predictors are identical if and only if $\{Z_t\}$ is a **martingale difference sequence** relative to $\{X_t\}$, i.e., if and only if $E(Z_t|X_s, -\infty < s < t) = 0$ for all t.

The Wold decomposition (Section 2.6) ensures that every purely nondeterministic stationary process can be expressed in the form (11.3.1) with $\{Z_t\} \sim \text{WN}\left(0, \sigma^2\right)$. The process $\{Z_t\}$ in the Wold decomposition, however, is generally not an iid sequence, and the best mean square predictor of X_{t+h} may be quite different from the best linear predictor.

In the case where $\{X_t\}$ is a purely nondeterministic Gaussian stationary process, the sequence $\{Z_t\}$ in the Wold decomposition is Gaussian and therefore iid. Every stationary purely nondeterministic Gaussian process can therefore be generated by applying a causal linear filter to an iid Gaussian sequence. We shall therefore refer to such a process as a **Gaussian linear process**.

In this section we shall use the term linear process to mean a process $\{X_t\}$ of the form (11.3.1). This is a more restrictive use of the term than in Definition 2.2.1.

11.3.1 Deviations from Linearity

Many of the time series encountered in practice exhibit characteristics not shown by linear processes, and so to obtain good models and predictors it is necessary to look to models more general than those satisfying (11.3.1) with iid noise. As indicated above, this will mean that the minimum mean squared error predictors are not, in general, linear functions of the past observations.

Gaussian linear processes have a number of properties that are often found to be violated by observed time series. The former are reversible in the sense that $\left(X_{t_1}, \cdots, X_{t_n}\right)'$ has the same distribution as $\left(X_{t_n}, \cdots, X_{t_1}\right)'$. (Except in a few special cases, ARMA processes are reversible if and only if they are Gaussian (Breidt and Davis 1992).) Deviations from this property by observed time series are suggested by sample paths that rise to their maxima and fall away at different rates (see, for example, the sunspot numbers filed as SUNSPOTS.TSM). Bursts of outlying values are frequently observed in practical time series and are seen also in the sample paths of nonlinear (and infinite-variance) models. They are rarely seen, however, in the sample paths of Gaussian linear processes. Other characteristics suggesting deviation from a Gaussian linear model are discussed by Tong (1990).

Many observed time series, particularly financial time series, exhibit periods during which they are "less predictable" (or "more volatile"), depending on the past history of the series. This dependence of the predictability (i.e., the size of the prediction mean squared error) on the past of the series cannot be modeled with a linear time series, since for a linear process the minimum h-step mean squared error is independent of the past history. Linear models thus fail to take account of the possibility that certain past histories may permit more accurate forecasting than others, and cannot identify the circumstances under which more accurate forecasts can be expected. Nonlinear models, on the other hand, do allow for this. The ARCH and GARCH models considered in Section 7.2 are in fact constructed around the dependence of the conditional variance of the process on its past history.

11.3.2 Chaotic Deterministic Sequences

To distinguish between linear and nonlinear processes, we need to be able to decide in particular when a white noise sequence is also iid. Sequences generated by nonlinear deterministic difference equations can exhibit sample correlation functions that are very close to those of samples from a white noise sequence. However, the deterministic nature of the recursions implies the strongest possible dependence between successive observations. For example, the celebrated logistic equation (see May, 1976, and Tong 1990) defines a sequence $\{x_n\}$, for any given x_0, via the equations

$$x_n = 4x_{n-1}(1 - x_{n-1}), \quad 0 < x_0 < 1.$$

The values of x_n are, for even moderately large values of n, extremely sensitive to small changes in x_0. This is clear from the fact that the sequence can be expressed explicitly as

$$x_n = \sin^2\left(2^n \arcsin\left(\sqrt{x_0}\right)\right), \quad n = 0, 1, 2, \ldots.$$

A very small change δ in $\arcsin\left(\sqrt{x_0}\right)$ leads to a change $2^n\delta$ in the argument of the sine function defining x_n. If we generate a sequence numerically, the generated sequence will, for most values of x_0 in the interval (0,1), be random in appearance, with a

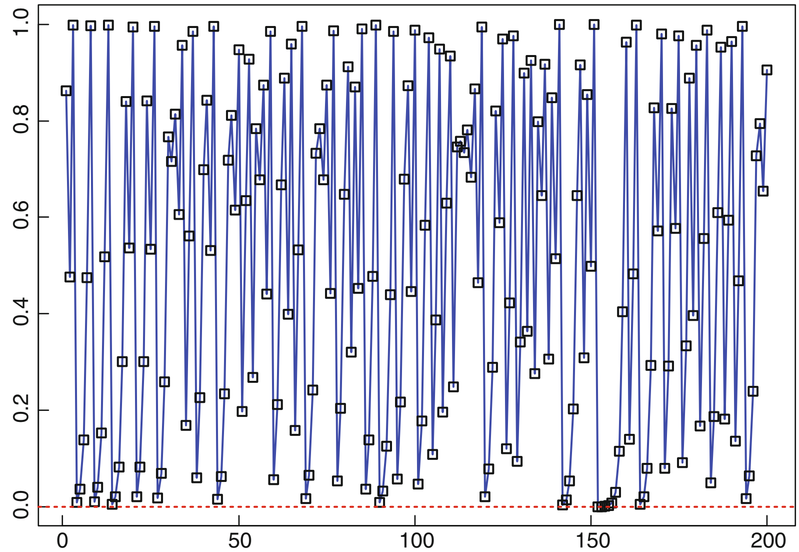

Figure 11-5
A sequence generated by the recursions $x_n = 4x_{n-1}(1 - x_{n-1})$

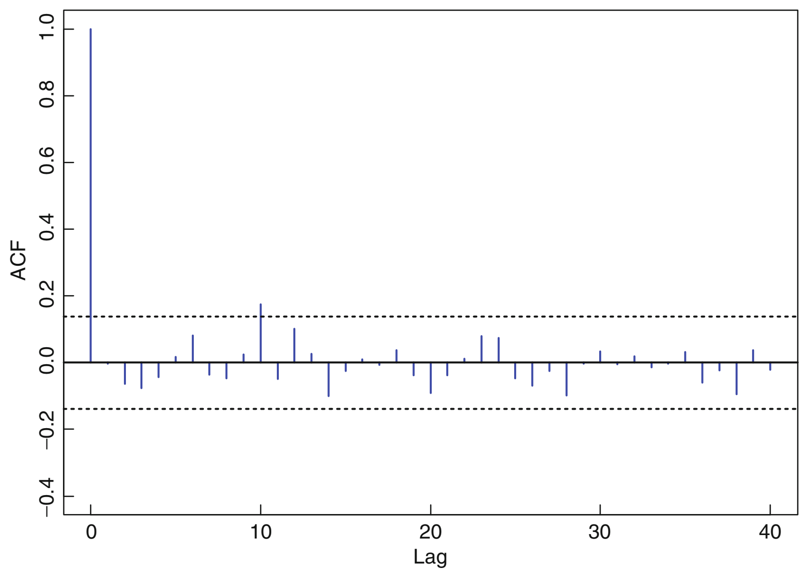

Figure 11-6
The sample autocorrelation function of the sequence in Figure 11-5

sample autocorrelation function similar to that of a sample from white noise. The data file CHAOS.TSM contains the sequence $x_1, \ldots, x_{200}$ (correct to nine decimal places) generated by the logistic equation with $x_0 = \pi/10$. The calculation requires specification of x_0 to at least 70 decimal places and the use of correspondingly high precision arithmetic. The series and its sample autocorrelation function are shown in Figures 11-5 and 11-6. The sample ACF and the AICC criterion both suggest white noise with mean 0.4954 as a model for the series. Under this model the best linear predictor of X_{201} would be 0.4954. However, the best predictor of X_{201} to nine decimal places is, in fact, $4x_{200}(1 - x_{200}) = 0.016286669$, with zero mean squared error.

Distinguishing between iid and non-iid white noise is clearly not possible on the basis of second-order properties. For insight into the dependence structure we can examine sample moments of order higher than two. For example, the dependence in the data in CHAOS.TSM is reflected by a significantly nonzero sample autocorrelation at lag 1 of the *squared* data. In the following paragraphs we consider several approaches to this problem.

11.3.3 Distinguishing Between White Noise and iid Sequences

If $\{X_t\} \sim \text{WN}\left(0, \sigma^2\right)$ and $E|X_t|^4 < \infty$, a useful tool for deciding whether or not $\{X_t\}$ is iid is the ACF $\rho_{X^2}(h)$ of the process $\left\{X_t^2\right\}$. If $\{X_t\}$ is iid, then $\rho_{X^2}(h) = 0$ for all $h \neq 0$, whereas this is not necessarily the case otherwise. This is the basis for the test of McLeod and Li described in Section 1.6.

Now suppose that $\{X_t\}$ is a strictly stationary time series such that $E|X_t|^k \le K < \infty$ for some integer $k \ge 3$. The kth-order cumulant $C_k(r_1, \ldots, r_{k-1})$ of $\{X_t\}$ is then defined as the joint cumulant of the random variables, $X_t, X_{t+r_1}, \ldots, X_{t+r_{k-1}}$, i.e., as the coefficient of $i^k z_1 z_2 \cdots z_k$ in the Taylor expansion about $(0, \ldots, 0)$ of

$$\chi(z_1, \ldots, z_k) := \ln E[\exp(iz_1 X_t + iz_2 X_{t+r_1} + \cdots + iz_k X_{t+r_{k-1}})]. \qquad (11.3.2)$$

(Since $\{X_t\}$ is strictly stationary, this quantity does not depend on t.) In particular, the third-order cumulant function C_3 of $\{X_t\}$ coincides with the third-order central moment function, i.e.,

$$C_3(r, s) = E[(X_t - \mu)(X_{t+r} - \mu)(X_{t+s} - \mu)], \qquad r, s \in \{0, \pm 1, \ldots\},$$

where $\mu = EX_t$. If $\sum_r \sum_s |C_3(r, s)| < \infty$, we define the third-order polyspectral density (or bispectral density) of $\{X_t\}$ to be the Fourier transform

$$f_3(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} C_3(r, s) e^{-ir\omega_1 - is\omega_2}, \qquad -\pi \le \omega_1, \omega_2 \le \pi,$$

in which case

$$C_3(r, s) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{ir\omega_1 + is\omega_2} f_3(\omega_1, \omega_2) d\omega_1\, d\omega_2.$$

[More generally, if the kth order cumulants $C_k(r_1, \cdots, r_{k-1})$, of $\{X_t\}$ are absolutely summable, we define the kth order polyspectral density as the Fourier transform of C_k. For details see Rosenblatt (1985) and Priestley (1988).]

If $\{X_t\}$ is a Gaussian linear process, it follows from Problem 10.3 that the cumulant function C_3 of $\{X_t\}$ is identically zero. (The same is also true of all the cumulant functions C_k with $k > 3$.) Consequently, $f_3(\omega_1, \omega_2) = 0$ for all $\omega_1, \omega_2 \in [-\pi, \pi]$. Appropriateness of a Gaussian linear model for a given data set can therefore be checked by using the data to test the null hypothesis $f_3 = 0$. For details of such a test, see Subba-Rao and Gabr (1984).

If $\{X_t\}$ is a linear process of the form (11.3.1) with $E|Z_t|^3 < \infty$, $EZ_t^3 = \eta$, and $\sum_{j=0}^{\infty} |\psi_j| < \infty$, it can be shown from (11.3.2) (see Problem 11.3) that the third-order cumulant function of $\{X_t\}$ is given by

$$C_3(r, s) = \eta \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+r} \psi_{i+s} \qquad (11.3.3)$$

(with $\psi_j = 0$ for $j < 0$), and hence that $\{X_t\}$ has bispectral density

$$f_3(\omega_1, \omega_2) = \frac{\eta}{4\pi^2} \psi\left(e^{i(\omega_1 + \omega_2)}\right) \psi\left(e^{-i\omega_1}\right) \psi\left(e^{-i\omega_2}\right), \qquad (11.3.4)$$

where $\psi(z) := \sum_{j=0}^{\infty} \psi_j z^j$. By Proposition 4.3.1, the spectral density of $\{X_t\}$ is

$$f(\omega) = \frac{\sigma^2}{2\pi} \left|\psi\left(e^{-i\omega}\right)\right|^2.$$

Hence,

$$\phi(\omega_1, \omega_2) := \frac{|f_3(\omega_1,\ \omega_2)|^2}{f(\omega_1)f(\omega_2)f(\omega_1+\omega_2)} = \frac{\eta^2}{2\pi\sigma^6}.$$

Appropriateness of the linear process (11.3.1) for modeling a given data set can therefore be checked by using the data to test for constancy of $\phi(\omega_1, \omega_2)$ (Subba-Rao and Gabr 1984).

11.3.4 Three Useful Classes of Nonlinear Models

If it is decided that a linear Gaussian model is not appropriate, there is a choice of several families of nonlinear processes that have been found useful for modeling purposes. These include bilinear models, autoregressive models with random coefficients, and threshold models. Excellent accounts of these are available in Subba-Rao and Gabr (1984), Nicholls and Quinn (1982), and Tong (1990), respectively.

The bilinear model of order (p, q, r, s) is defined by the equations

$$X_t = Z_t + \sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=1}^{q} b_j Z_{t-j} + \sum_{i=1}^{r}\sum_{j=1}^{s} c_{ij} X_{t-i} Z_{t-j}, \tag{11.3.5}$$

where $\{Z_t\} \sim \text{iid}\left(0, \sigma^2\right)$. A sufficient condition for the existence of a strictly stationary solution of these equations is given by Liu and Brockwell (1988).

A random coefficient autoregressive process $\{X_t\}$ of order p satisfies an equation of the form

$$X_t = \sum_{i=1}^{p} \left(\phi_i + U_t^{(i)}\right) X_{t-i} + Z_t,$$

where $\{Z_t\} \sim \text{IID}\left(0, \sigma^2\right)$, $\left\{U_t^{(i)}\right\} \sim \text{IID}\left(0, \nu^2\right)$, $\{Z_t\}$ is independent of $\left\{U_t\right\}$, and $\phi_1, \ldots, \phi_p \in \mathbb{R}$.

Threshold models can be regarded as piecewise linear models in which the linear relationship varies with the values of the process. For example, if $R^{(i)}$, $i = 1, \ldots, k$, is a partition of $\mathbb{R}^p$, and $\{Z_t\} \sim \text{IID}(0, 1)$, then the k difference equations

$$X_t = \sigma^{(i)} Z_t + \sum_{j=1}^{p} \phi_j^{(i)} X_{t-j}, \quad (X_{t-1}, \cdots, X_{t-p}) \in R^{(i)}, \quad i = 1, \cdots, k, \tag{11.3.6}$$

define a threshold AR(p) model. Model identification and parameter estimation for threshold models can be carried out in a manner similar to that for linear models using maximum likelihood and the AIC criterion.

11.4 Long-Memory Models

The autocorrelation function $\rho(\cdot)$ of an ARMA process at lag h converges rapidly to zero as $h \to \infty$ in the sense that there exists $r > 1$ such that

$$r^h \rho(h) \to 0 \qquad \text{as} \quad h \to \infty. \tag{11.4.1}$$

Stationary processes with much more slowly decreasing autocorrelation function, known as **fractionally integrated ARMA processes**, or more precisely as ARIMA (p, d, q) processes with $0 < |d| < 0.5$, satisfy difference equations of the form

$$(1 - B)^d \phi(B) X_t = \theta(B) Z_t, \tag{11.4.2}$$

where $\phi(z)$ and $\theta(z)$ are polynomials of degrees p and q, respectively, satisfying

$$\phi(z) \neq 0 \quad \text{and} \quad \theta(z) \neq 0 \quad \text{for all } z \text{ such that } |z| \leq 1,$$

B is the backward shift operator, and $\{Z_t\}$ is a white noise sequence with mean 0 and variance σ^2. The operator $(1 - B)^d$ is defined by the binomial expansion

$$(1 - B)^d = \sum_{j=0}^{\infty} \pi_j B^j,$$

where $n_0 = 1$ and

$$\pi_j = \prod_{0<k\leq j} \frac{k - 1 - d}{k}, \qquad j = 1, 2, \ldots.$$

The autocorrelation $\rho(h)$ at lag h of an ARIMA(p, d, q) process with $0 < |d| < 0.5$ has the property

$$\rho(h) h^{1-2d} \to c \neq 0 \qquad \text{as} \quad h \to \infty. \tag{11.4.3}$$

This implies (see (11.4.1)) that $\rho(h)$ converges to zero as $h \to \infty$ at a much slower rate than $\rho(h)$ for an ARMA process. Consequently, fractionally integrated ARMA processes are said to have "long memory." In contrast, stationary processes whose ACF converges to 0 rapidly, such as ARMA processes, are said to have "short memory."

A fractionally integrated ARIMA(p, d, q) process can be regarded as an ARMA (p, q) process driven by fractionally integrated noise; i.e., we can replace equation (11.4.2) by the two equations

$$\phi(B) X_t = \theta(B) W_t \tag{11.4.4}$$

and

$$(1 - B)^d W_t = Z_t. \tag{11.4.5}$$

The process $\{W_t\}$ is called **fractionally integrated white noise** and can be shown (see, e.g., Brockwell and Davis (1991), Section 13.2) to have variance and autocorrelations given by

$$\gamma_W(0) = \sigma^2 \frac{\Gamma(1 - 2d)}{\Gamma^2(1 - d)} \tag{11.4.6}$$

and

$$\rho_W(h) = \frac{\Gamma(h + d)\Gamma(1 - d)}{\Gamma(h - d + 1)\Gamma(d)} = \prod_{0<k\leq h} \frac{k - 1 + d}{k - d}, \quad h = 1, 2, \ldots, \tag{11.4.7}$$

where $\Gamma(\cdot)$ is the gamma function (see Example (*d*) of Section A.1). The exact autocovariance function of the ARIMA(p, d, q) process $\{X_t\}$ defined by (11.4.2) can therefore be expressed, by Proposition 2.2.1, as

$$\gamma_X(h) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \psi_k \gamma_W(h + j - k), \tag{11.4.8}$$

where $\sum_{i=0}^{\infty} \psi_i z^i = \theta(z)/\phi(z)$, $|z| \le 1$, and $\gamma_W(\cdot)$ is the autocovariance function of fractionally integrated white noise with parameters d and σ^2, i.e.,

$$\gamma_W(h) = \gamma_W(0)\rho_W(h),$$

with $\gamma_W(0)$ and $\rho_W(h)$ as in (11.4.6) and (11.4.7). The series (11.4.8) converges rapidly as long as $\phi(z)$ does not have zeros with absolute value close to 1.

The spectral density of $\{X_t\}$ is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{\left|\theta(e^{-i\lambda})\right|^2}{\left|\phi(e^{-i\lambda})\right|^2} \left|1 - e^{-i\lambda}\right|^{-2d}. \quad (11.4.9)$$

Calculation of the exact Gaussian likelihood of observations $\{x_1, \ldots, x_n\}$ of a fractionally integrated ARMA process is very slow and demanding in terms of computer memory. Instead of estimating the parameters d, $\phi_1, \ldots, \phi_p$, $\theta_1, \ldots, \theta_q$, and σ^2 by maximizing the exact Gaussian likelihood, it is much simpler to maximize the Whittle approximation L_W, defined by

$$-2\ln(L_W) = n\ln(2\pi) + 2n\ln\sigma + \sigma^{-2}\sum_j \frac{I_n(\omega_j)}{g(\omega_j)} + \sum_j \ln g(\omega_j), \quad (11.4.10)$$

where I_n is the periodogram, $\sigma^2 g/(2\pi)(= f)$ is the model spectral density, and $\sum_j$ denotes the sum over all nonzero Fourier frequencies $\omega_j = 2\pi j/n \in (-\pi, \pi]$. The program ITSM estimates parameters for ARIMA(p, d, q) models in this way. It can also be used to predict and simulate fractionally integrated ARMA series and to compute the autocovariance function of any specified fractionally integrated ARMA model.

Example 11.4.1 Annual Minimum Water Levels; NILE.TSM

The data file NILE.TSM consists of the annual minimum water levels of the Nile river as measured at the Roda gauge near Cairo for the years 622–871. These values are plotted in Figure 11-7 with the corresponding sample autocorrelations shown in Figure 11-8. The rather slow decay of the sample autocorrelation function suggests the possibility of a fractionally intergrated model for the mean-corrected series $Y_t = X_t - 1119$.

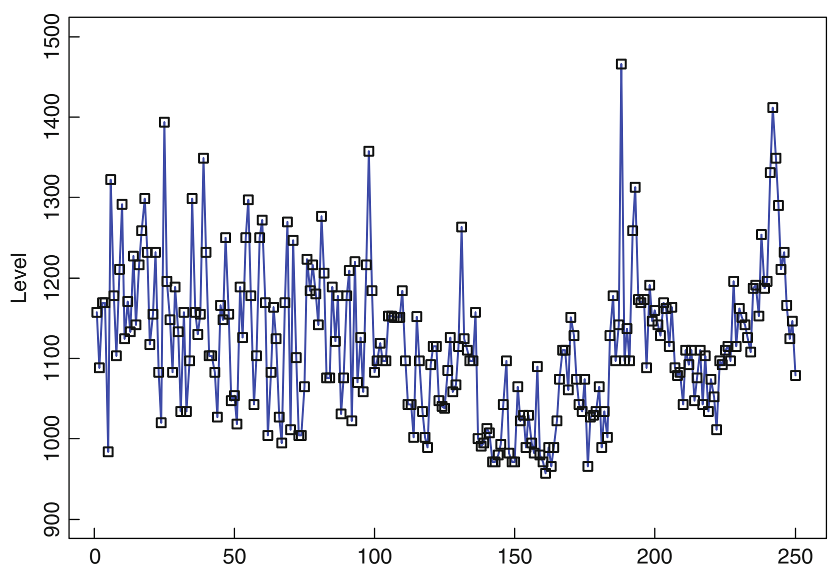

Figure 11-7 Annual minimum water levels of the Nile river for the years 622–871

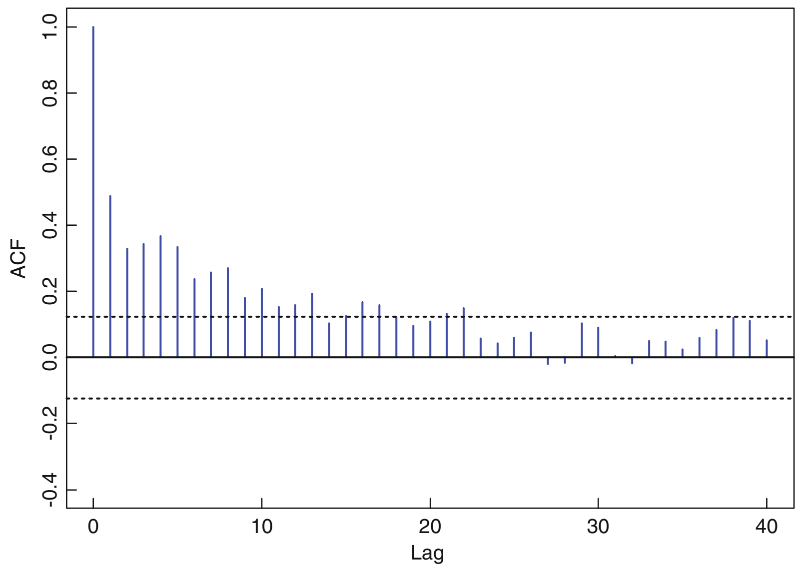

Figure 11-8
The sample correlation function of the data in Figure 11-7

The ARMA model with minimum (exact) AICC value for the mean-corrected series $\{Y_t\}$ is found, using `Model>Estimation>Autofit`, to be

$$Y_t = -0.323Y_{t-1} - 0.060Y_{t-2} + 0.633Y_{t-3} + 0.069Y_{t-4} + 0.248Y_{t-5} + Z_t + 0.702Z_{t-1} + 0.350Z_{t-2} - 0.419Z_{t-3}, \quad (11.4.11)$$

with $\{Z_t\} \sim \text{WN}(0, 5663.6)$ and AICC= 2889.9.

To fit a fractionally integrated ARMA model to this series, select the option `Model>Specify`, check the box marked `Fractionally integrated model`, and click on `OK`. Then select `Model>Estimation>Autofit`, and click on `Start`. This estimation procedure is relatively slow so the specified ranges for p and q should be small (the default is from 0 to 2). When models have been fitted for each value of (p, q), the fractionally integrated model with the smallest modified AIC value is found to be

$$(1 - B)^{0.3830}(1 - 0.1694B + 0.9704B^2)Y_t = (1 - 0.1800B + 0.9278B^2)Z_t, \quad (11.4.12)$$

with $\{Z_t\} \sim \text{WN}(0, 5827.4)$ and modified AIC= 2884.94. (The modified AIC statistic for estimating the parameters of a fractionally integrated ARMA(p, q) process is defined in terms of the Whittle likelihood L_W as $-2\ln L_W + 2(p + q + 2)$ if d is estimated, and $-2\ln L_W + 2(p+q+1)$ otherwise. The Whittle likelihood was defined in (11.4.10).)

In order to compare the models (11.4.11) and (11.4.12), the modified AIC value for (11.4.11) is found as follows. After fitting the model as described above, select `Model>Specify`, check the box marked `Fractionally integrated model`, set $d = 0$ and click on `OK`. Next choose `Model>Estimation>Max likelihood`, check `No optimization` and click on `OK`. You will then see the modified AIC value, 2884.58, displayed in the `ML estimates` window together with the value 2866.58 of $-2\ln L_W$.

The ARMA(5,3) model is slightly better in terms of modified AIC than the fractionally integrated model and its ACF is closer to the sample ACF of the data than is the ACF of the fractionally integrated model. (The sample and model autocorrelation functions can be compared by clicking on the third yellow button at the top of the ITSM window.) The residuals from both models pass all of the ITSM tests for randomness.

Figure 11-9 shows the graph of $\{x_{200}, \ldots, x_{250}\}$ with predictors of the next 20 values obtained from the model (11.4.12) for the mean-corrected series.

□

11.5 Continuous-Time ARMA Processes

Time series frequently consist of observations of a continuous-time process $\{Y(t), t \geq 0\}$ or $\{Y(t), t \in \mathbb{R}\}$ at a discrete sequence of observation times. It is then natural, even though the observations are made at discrete times, to model the data by fitting the underlying continuous-time process.

Even if there is no underlying continuous-time process, it may still be advantageous to model the data as observations of a continuous-time process sampled at discrete times. For example, the analysis of time series data observed at irregularly spaced times can be handled very conveniently by regarding the data as sampled values of a continuous-time process (see Jones 1980 and equation (11.5.6) below).

Continuous-time models also provide a unifying framework for data collected when a time series is observed at different frequencies, i.e., with different spacings between the observation times. Instead of requiring different discrete-time models to represent observations collected at different frequencies, continuous-time modelling provides a single model which can be sampled at any frequency whatsoever.

When very high-frequency observations are available (as in many financial and turbulence studies), the relation between the high-frequency sequence and the underlying continuous-time process is also of interest since the high-frequency observations provide a natural source of information regarding the continuous-time process of which the discrete observations are a sample.

Stationarity of a continuous-time process $\{Y(t)\}$ (cf. Definition 1.4.2) means that $EY(t)$ and $\text{Cov}(Y(t+h), Y(t))$ are defined and independent of t for all $h \geq 0$. Strict stationarity means that $(Y(t_1), \ldots, Y(t_n))$ and $(Y(t_1+h), \ldots, Y(t_n+h))$ have the same joint distributions for all $t_1, \ldots, t_n$, all $h \geq 0$ and all positive integers n.

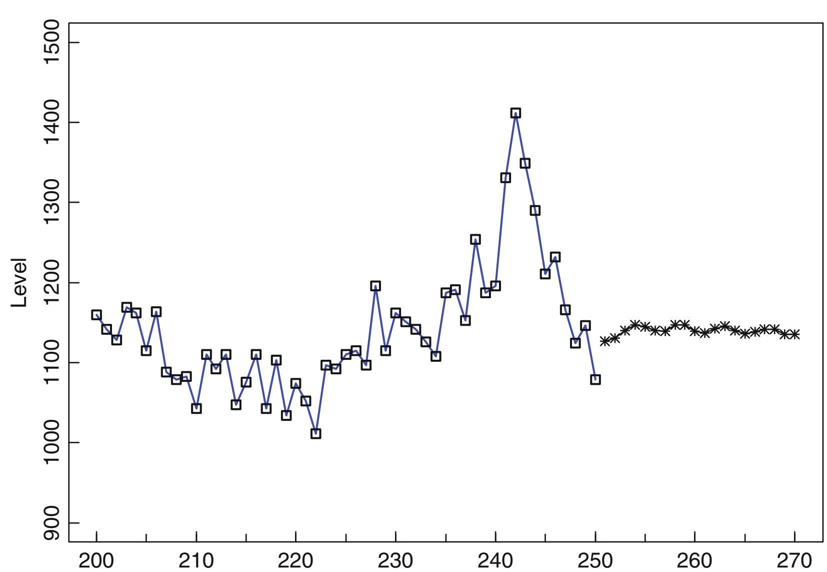

Figure 11-9
The minimum annual Nile river levels for the years 821–871, with 20 forecasts based on the model (11.4.12)

Continuous-time ARMA (or CARMA) processes are defined as stationary solutions of stochastic *differential* equations analogous to the difference equations that are used to define discrete-time ARMA processes. They play a role in continuous-time modelling analogous to that of ARMA processes in discrete time.

We shall begin with the Gaussian continuous-time AR(1) process, also known as the stationary Gaussian Ornstein-Uhlenbeck process.

11.5.1 The Gaussian CAR(1) Process, $\{Y(t), t \geq 0\}$

The Gaussian CAR(1) process $\{Y(t), t \geq 0\}$ is defined as a strictly stationary solution of the first-order stochastic differential equation,

$$DY(t) + aY(t) = \sigma DB(t) + c, \ t > 0, \tag{11.5.1}$$

where the operator D denotes differentiation with respect to t, $\{B(t), \ t \in \mathbb{R}\}$ is standard Brownian motion (see Example 7.5.1), a, c, and σ are parameters and $Y(0)$ is a normally distributed random variable independent of $\{B(t) - B(s), 0 \leq s \leq t < \infty\}$. The derivative $DB(t)$ does not exist in the usual sense, so equation (11.5.1) is interpreted as the Itô differential equation (see Appendix D.4),

$$dY(t) + aY(t)dt = \sigma dB(t) + c\,dt, \quad t > 0, \tag{11.5.2}$$

with $dY(t)$ and $dB(t)$ denoting the increments of Y and B in the time interval $(t, t+dt)$.

Standard theory of deterministic linear differential equations suggests multiplying this equation by e^{at} in which case the left-hand side would become $d(e^{at}Y(t))$. We therefore apply Itô's formula (Appendix, equation (D.3.7)) to $d(e^{at}Y(t))$ with $g(t, x) := e^{at}x$ and we obtain exactly the same result since the second partial derivative g_{xx} is zero. Hence we can rewrite (11.5.2) as

$$d(e^{at}Y(t)) = \sigma e^{at}dB(t) + ce^{at}dt,$$

or equivalently,

$$e^{at}Y(t) - Y(0) = \sigma \int_0^t e^{au}dB(u) + c\int_0^t e^{au}du.$$

Thus

$$Y(t) = e^{-at}Y(0) + \sigma \int_0^t e^{-a(t-u)}dB(u) + c\int_0^t e^{-a(t-u)}\,du. \tag{11.5.3}$$

Remark 1. The Itô integral $\int_0^t e^{-a(t-u)}dB(u)$ in (11.5.3) is of a special type in which the integrand is *deterministic*. This permits the application of integration by parts to obtain a *pathwise* representation of Y as

$$Y(t) = e^{-at}Y(0) + \sigma B(t) - \sigma \int_0^t ae^{-a(t-u)}B(u)du + c\int_0^t e^{-a(t-u)}du.$$ □

If $a > 0$ and $Y(0)$ has mean c/a and variance $\sigma^2/(2a)$, it is easy to check, using the properties of $I_{0,t}(f) = \int_0^t f(u)dB(u)$ in Remark 3 of Appendix D.3 and the independence of $Y(0)$ and $\{B(t) - B(s), 0 \leq s \leq t < \infty\}$ (Problem 11.4), that $\{Y(t)\}$ as defined by (11.5.3) is stationary with

$$E(Y(t)) = \frac{c}{a} \quad \text{and} \quad \mathrm{Cov}(Y(t+h), Y(t)) = \frac{\sigma^2}{2a}e^{-ah}, \quad t, h \geq 0. \tag{11.5.4}$$

Since $\{Y(t)\}$ is Gaussian it is also strictly stationary. Conversely, if $\{Y(t)\}$ is strictly stationary, then by equating the variances of both sides of (11.5.3), we find that $\left(1 - e^{-2at}\right) \mathrm{Var}(Y(0)) = \sigma^2 \int_0^t e^{-2au}\,du$ for all $t \geq 0$, and hence that $a > 0$ and $\mathrm{Var}(Y(0)) = \sigma^2/(2a)$. Equating the means of both sides of (11.5.3) then gives $E(Y(0)) = c/a$. Necessary and sufficient conditions for $\{Y(t)\}$ to be strictly stationary are therefore $a > 0$, $E(Y(0)) = c/a$, and $\mathrm{Var}(Y(0)) = \sigma^2/(2a)$.

If $a > 0$ and $0 \leq s \leq t$, it follows from (11.5.3) that $Y(t)$ satisfies the relation

$$Y(t) = e^{-a(t-s)}Y(s) + \frac{c}{a}\left(1 - e^{-a(t-s)}\right) + \sigma \int_s^t e^{-a(t-u)}dB(u), \quad t \geq s \geq 0. \tag{11.5.5}$$

This shows that the process is Markovian, i.e., that the distribution of $Y(t)$ given $Y(u), u \leq s$, is the same as the distribution of $Y(t)$ given $Y(s)$. It also shows that the conditional mean and variance of $Y(t)$ given $Y(s)$ are

$$E(Y(t)|Y(s)) = e^{-a(t-s)}Y(s) + c/a\left(1 - e^{-a(t-s)}\right)$$

and

$$\mathrm{Var}(Y(t)|Y(s)) = \frac{\sigma^2}{2a}\left[1 - e^{-2a(t-s)}\right].$$

We can now use the Markov property and the moments of the stationary distribution to write down the likelihood of observations $y(t_1), \ldots, y(t_n)$ at times $t_1, \ldots, t_n$ of the Gaussian CAR(1) process. This is just the joint density of $(Y(t_1), \ldots, Y(t_n))'$ at $(y(t_1), \ldots, y(t_n))'$, which can be expressed as the product of the stationary density at $y(t_1)$ and the transition densities of $Y(t_i)$ given $Y(t_{i-1}) = y(t_{i-1})$, $i = 2, \ldots, n$. The joint density g is therefore given by

$$g\left(y(t_1), \ldots, y(t_n); a, c, \sigma^2\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{v_i}} f\left(\frac{y(t_i) - m_i}{\sqrt{v_i}}\right), \tag{11.5.6}$$

where $f(y) = n(y; 0, 1)$ is the standard normal density, $m_1 = c/a$, $v_1 = \sigma^2/(2a)$, and for $i > 1$,

$$m_i = e^{-a(t_i - t_{i-1})}y(t_{i-1}) + \frac{c}{a}\left(1 - e^{-a(t_i - t_{i-1})}\right)$$

and

$$v_i = \frac{\sigma^2}{2a}\left[1 - e^{-2a(t_i - t_{i-1})}\right].$$

The maximum likelihood estimators of a, c, and σ^2 are the values that maximize $g\left(y(t_1), \ldots, y(t_n); a, c, \sigma^2\right)$. These can be found with the aid of a nonlinear maximization algorithm. Notice that the times t_i appearing in (11.5.6) are quite arbitrarily spaced. It is this feature that makes the CAR(1) process so useful for modeling irregularly spaced data.

If the observations are regularly spaced, say $t_i = i$, $i = 1, \ldots, n$, then the joint density g is exactly the same as the joint density of observations of the discrete-time Gaussian AR(1) process

$$Y_n - \frac{c}{a} = e^{-a}\left(Y_{n-1} - \frac{c}{a}\right) + Z_n, \quad \{Z_t\} \sim \mathrm{IIDN}\left(0, \frac{\sigma^2\left(1 - e^{-2a}\right)}{2a}\right).$$

This shows that the "embedded" (or sampled) discrete-time process $\{Y(i), i = 1, 2, \ldots\}$ of the CAR(1) process is a discrete-time AR(1) process with coefficient e^{-a}.

This coefficient is clearly positive, immediately raising the question of whether there is a continuous-time ARMA process for which the embedded process is a discrete-time AR(1) process with negative coefficient. It can be shown (Chan and Tong 1987) that the answer is yes and that given a discrete-time AR(1) process with negative coefficient, it can always be embedded in a suitably chosen continuous-time ARMA(2,1) process.

11.5.2 The Gaussian CARMA(p, q) Process, $\{Y(t), t \in \mathbb{R}\}$

We define a zero-mean Gaussian CARMA(p, q) process $\{Y(t), t \in \mathbb{R}\}$ (with $0 \le q < p$) to be a strictly stationary Gaussian process satisfying the pth-order linear differential equation,

$$D^pY(t) + a_1D^{p-1}Y(t) + \cdots + a_pY(t)$$
$$= b_0DB(t) + b_1D^2B(t) + \cdots + b_qD^{q+1}B(t), \tag{11.5.7}$$

where D^j denotes j-fold differentiation with respect to t, $\{B(t),\ t \in \mathbb{R}\}$ is standard Brownian motion, and $a_1, \ldots, a_p$ and $b_0, \ldots, b_q$ are constants. We assume that $b_q \neq 0$ and define $b_j := 0$ for $j > q$. We shall also assume that the polynomials, $a(z) := z^p + a_1z^{p-1} + \cdots + a_p$ and $b(z) := b_0 + b_1z + \cdots + b_qz^q$, have no common zeroes. Since the derivatives $D^jB(t), j > 0$, do not exist in the usual sense, we interpret (11.5.7) as being equivalent (see Remark 2 below) to the *observation* and *state* equations

$$Y(t) = \mathbf{b}'\mathbf{X}(t), \tag{11.5.8}$$

and

$$d\mathbf{X}(t) = A\mathbf{X}(t)\,dt + \mathbf{e}\,dB(t), \tag{11.5.9}$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \cdots & -a_1, \end{bmatrix},$$

$\mathbf{e} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix}'$, $\mathbf{b} = \begin{bmatrix} b_0 & b_1 & \cdots & b_{p-2} & b_{p-1} \end{bmatrix}'$, and (11.5.9) is an Itô differential equation for the state vector $\mathbf{X}(t)$ (see Appendix D.4).

Remark 2. Denoting the components of $\mathbf{X}(t)$ by $X_j(t), j = 0, \ldots, p-1$, the first $p-1$ component equations of (11.5.9) are

$$dX_j(t) = X_{j+1}(t)dt,\ j = 0, \ldots, p-2,$$

showing that $X_j(t)$ is just the j^{th} derivative of $X_0(t)$, $j = 1, \ldots, p-1$. The last component equation of (11.5.9) is

$$dX_{p-1}(t) = -(a_1X_{p-1} + a_2X_{p-2} + \cdots + a_pX_0(t))dt + dB(t),$$

which is the Itô form of the stochastic differential equation,

$$D^pX_0(t) + a_1D^{p-1}X_0(t) + \cdots + a_pX_0(t) = DB(t). \tag{11.5.10}$$

The state equation (11.5.9) is thus the Itô equation for the vector whose first component $X_0(t)$ satisfies the CARMA(p, 0) (or CAR(p)) equation (11.5.10) and whose remaining components are successively higher derivatives, up to order $p-1$, of $X_0(t)$. It is clear from the linearity of equation (11.5.7) that if $\mathbf{X}(t)$ satisfies (11.5.9) then $X_0(t)$ satisfies (11.5.10) and the linear combination,

$$Y(t) = b_0X_0(t) + b_1X_1(t) + \cdots + b_{p-1}X_{p-1}(t) = \mathbf{b}'\mathbf{X}(t),$$

of $X_0(t)$ and its derivatives satisfies (11.5.7). This explains the replacement of (11.5.7) by the observation and state equations (11.5.8) and (11.5.9). □

If $\mathbf{X}(0)$ is a normally distributed random vector independent of $\{B(t) - B(s), 0 \le s \le t < \infty\}$, then equation (11.5.9) is simply a vector form of equation (11.5.2) and its unique solution for $t \ge 0$, as specified in Appendix D.4, Theorem D.4.1, satisfies

$$\mathbf{X}(t) = e^{At}\mathbf{X}(0) + \int_0^t e^{A(t-u)}\mathbf{e}\, dB(u), \quad 0 \le t < \infty,$$

where the matrix e^{At} is defined in the usual way as $e^{At} := \sum_{j=0}^{\infty} \frac{A^j}{j!} t^j$.

More generally (see Appendix D, equation (D.4.6)), if for each $S \in \mathbb{R}$, $\mathbf{X}(S)$ has finite second moments and is independent of $\{B(t) - B(s),\ S \le s \le t < \infty\}$, then the unique solution of (11.5.9) specified by Theorem D.4.1 satisfies

$$\mathbf{X}(t) = e^{A(t-S)}\mathbf{X}(S) + \int_S^t e^{A(t-u)}\mathbf{e}\, dB(u), \quad t \ge S, \text{ for all } S \in \mathbb{R}. \tag{11.5.11}$$

If the real parts of the eigenvalues $\lambda_1, \ldots, \lambda_p$ of A (which are also the zeroes of the autoregressive polynomial $a(z)$) satisfy

$$\mathscr{R}e(\lambda_r) < 0, \quad r = 1, \ldots, p, \tag{11.5.12}$$

and if $\{\mathbf{X}(t)\}$ is a *stationary* solution of (11.5.11) then, taking mean square limits as $S \to -\infty$ in (11.5.11), we see that it must satisfy

$$\mathbf{X}(t) = \int_{-\infty}^t e^{A(t-u)}\mathbf{e}\, dB(u), \quad t \in \mathbb{R}. \tag{11.5.13}$$

Conversely, if $\{\mathbf{X}(t)\}$ is given by (11.5.13) then it is a stationary solution of (11.5.11) and for each $S \in \mathbb{R}$, $\mathbf{X}(S)$ has finite second moments and is independent of $\{B(t) - B(s),\ S \le s \le t < \infty\}$. Hence it is the unique solution of the type specified in Theorem D.4.1 with these properties. The property that, for each S, $\mathbf{X}(S)$ is independent of $\{B(t) - B(s),\ S \le s \le t < \infty\}$, corresponds to the discrete-time concept of causality introduced in Section 3.1.

Assuming that condition (11.5.12) is satisfied, we define the zero-mean causal Gaussian CARMA(p, q) process $\{Y(t),\ t \in \mathbb{R}\}$, with parameters $(a_1, \ldots, a_p, b_0, \ldots, b_q)$, by

$$Y(t) = \mathbf{b}'\mathbf{X}(t), \tag{11.5.14}$$

where $\{\mathbf{X}(t)\}$ is given by (11.5.13). A Gaussian CARMA process with mean m is obtained by simply adding the constant value m to Y.

Remark 3. For the zero-mean causal Gaussian CAR(1) process defined by (11.5.1), with $c = 0$ and with index set $\mathbb{R}$ instead of $[0, \infty)$ as in Section 11.5.1, we have $\mathbf{b} = \sigma$ and $A = -a$, so that

$$Y(t) = \sigma \int_{-\infty}^{t} e^{-a(t-u)} dB(u), \quad t \in \mathbb{R}. \qquad \square$$

The autocovariance function of the process $\mathbf{X}(t)$ at lag h is easily found from (11.5.13) to be

$$\text{Cov}(\mathbf{X}(t+h), \mathbf{X}(t)) = e^{A|h|}\Sigma,$$

where

$$\Sigma := \int_0^\infty e^{Ay}\mathbf{e}\,\mathbf{e}' e^{A'y} dy.$$

The mean and autocovariance function of the CARMA(p, q) process $\{Y(t)\}$ are therefore given by

$$EY(t) = 0$$

and

$$\text{Cov}(Y(t+h), Y(t)) = \mathbf{b}' e^{A|h|}\Sigma\mathbf{b}.$$

Inference for a CARMA(p, q) process with $p > 1$ is more complicated than for a CAR(1) process because the former is not Markovian, so the simple argument that led to (11.5.6) no longer holds. However, the Gaussian likelihood of observations at times $t_1, \ldots, t_n$ can still easily be computed using the discrete-time Kalman recursions as pointed out by Jones (1980).

Simulation and estimation, not only for Gaussian, but also for Lévy-driven CARMA processes (as introduced in the following subsection) can be carried out using the Yuima package, a package for use in the R environment, which can be downloaded from https://cran.r-project.org/web/packages/yuima. A detailed account of its application to CARMA processes is contained in the paper of Iacus and Mercuri (2015). A simulated Gaussian (3,2) process and the components of its state-vector, generated in R by the Yuima package, is shown in Figure 11-10.

Rather than examining Gaussian CARMA processes in more detail, we next introduce the more general class of Lévy-driven CARMA(p, q) processes, whose marginal distributions can be both heavy-tailed and asymmetric and whose sample-paths are continuous if $q < p-1$ and have the same jumps as the driving Lévy process if $q = p - 1$.

11.5.3 Lévy-driven CARMA Processes, $\{Y(t), t \in \mathbb{R}\}$

In Section 11.5.2, under the assumption (11.5.12), we defined the zero-mean causal Gaussian CARMA process $\{Y(t), t \in \mathbb{R}\}$ as the strictly stationary linear combination (11.5.14) of components of the state-vector $\mathbf{X}(t)$ given by (11.5.13). In this section we wish to extend the definition by replacing the driving process B by a Lévy process L in order to allow a much broader class of possible marginal distributions for $Y(t)$. As in Section 11.5.2 we shall assume that the polynomials $a(z)$ and $b(z)$ in the defining stochastic differential equation,

$$a(D)Y(t) = b(D)DL(t),$$

have no common zeroes. We use the same state-space representation of the process as in Section 11.5.2 to obtain a rigorous interpretation of this equation.

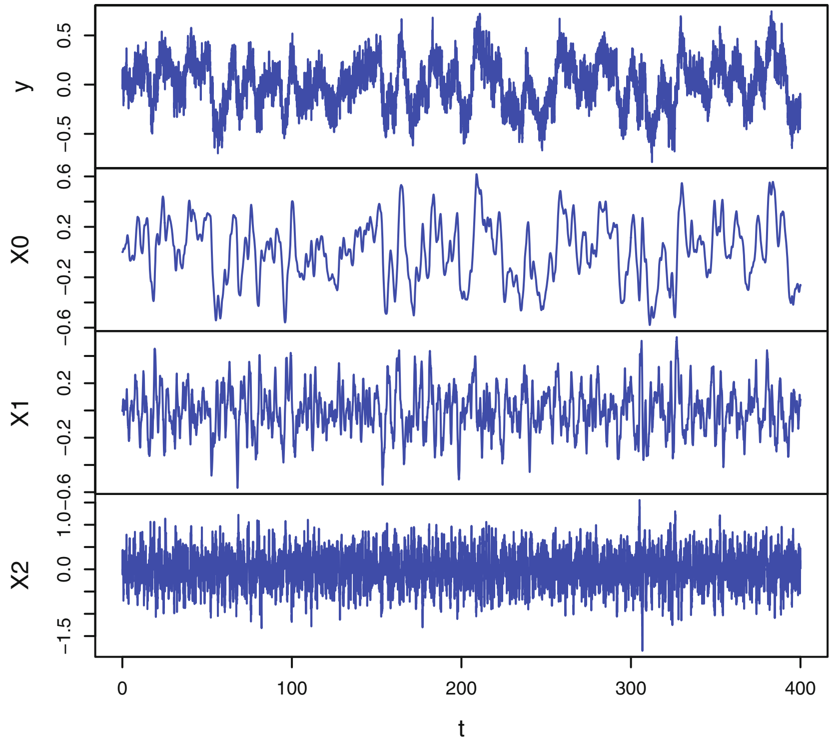

Figure 11-10 Simulated CARMA(3,2) process y and state-vector X driven by standard Brownian motion with $a_1 = 4,\ a_2 = 4.5,\ a_3 = 1.5,\ b_0 = 1,\ b_1 = .23,\ b_2 = .35$

Replacing the Brownian motion $\{B(t)\}$ in (11.5.11) by the Lévy process $\{L(t)\}$ gives

$$\mathbf{X}(t) = e^{A(t-S)}\mathbf{X}(S) + \int_{(S,t]} e^{A(t-S)}dL(u),\ t \geq S,\ \text{for all } S \in \mathbb{R}. \tag{11.5.15}$$

where the integral is now interpreted in the sense of Protter (2010). We then define the **CARMA**(p, q) **process driven by** L**, with coefficients** $(a_1, \ldots, a_p, b_1, \ldots, b_q)$, to be a strictly stationary solution $\{Y(t)\}$ of the equation (11.5.15) and

$$Y(t) = \mathbf{b}'\mathbf{X}(t). \tag{11.5.16}$$

The matrix A and the vector $\mathbf{b}$ are defined as in Section 11.5.2, except that we now define $b_q := 1$ since there is no variance constraint on $L(1)$ as there was on $B(1)$ in the definition of the Gaussian special cases in Sections 11.5.1 and 11.5.2. In the case when L is a subordinator (i.e., a Lévy process with non-decreasing sample paths), the integral in (11.5.15) can also be interpreted as a pathwise Stieltjes integral with respect to L.

Brockwell and Lindner (2009) have established necessary and sufficient conditions for the existence of a strictly stationary solution $\{Y(t),\ t \in \mathbb{R}\}$ of (11.5.15) and (11.5.16). If we assume that $a(z)$ and $b(z)$ have no common zeroes and L is not deterministic, then the necessary and sufficient conditions are

$$E \max(0. \log |L(1)|) < \infty \tag{11.5.17}$$

and

$$\mathscr{R}e(\lambda_r) \neq 0,\quad r = 1, \ldots, p, \tag{11.5.18}$$

where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of A (which are also the zeroes of the autoregressive polynomial $a(z)$).

The strictly stationary solution is unique, and under the stronger conditions,

$$\mathscr{R}e(\lambda_r) < 0,\quad r = 1, \ldots, p, \tag{11.5.19}$$

it is causal, i.e., for every s, t and u such that $s \le t \le u$, $Y(s)$ is independent of $L(u) - L(t)$ and can be expressed as

$$Y(t) = \int_{(-\infty,t]} \mathbf{b}' e^{A(t-u)} \mathbf{e} dL(u). \tag{11.5.20}$$

(The representation (11.5.20) is easily obtained formally by letting $S \to -\infty$ in (11.5.15) and substituting the resulting expression for $\mathbf{X}(t)$ in (11.5.16).) Thus, under the causality condition (11.5.19), equations (11.5.15) and (11.5.16) have the unique strictly stationary solution (11.5.20). This solution is the causal CARMA(p, q) process with parameters $(a_1, \ldots, a_p, b_1, \ldots, b_q := 1)$ driven by the Lévy process L.

From equation (11.5.20) we find that, if $E(L(1)^2) < \infty$, $EY(t) = \mu b_0/a_p$, where $\mu = EL(1)$, and

$$\gamma_Y(h) := \text{Cov}[Y(t+h), Y(t)] = \sigma^2 \mathbf{b}' e^{A|h|} \Sigma \mathbf{b}, \tag{11.5.21}$$

where $\sigma^2 = \text{Var}(L(1))$ and $\Sigma = \int_0^\infty e^{Ay} \mathbf{e}\mathbf{e}' e^{A'y} dy$.

Remark 4. A result of Eller (1987) was used by Brockwell and Lindner (2009) to rewrite (11.5.20) as

$$Y(t) = \int_{(-\infty,t]} \sum_{\lambda} \sum_{k=0}^{m(\lambda)-1} c_{\lambda k}(t-u)^k e^{\lambda(t-u)} dL(u), \tag{11.5.22}$$

where Σ_λ denotes summation over the zeroes λ of $a(z)$, $m(\lambda)$ is the multiplicity of the zero λ and $\sum_{k=0}^{m(\lambda)-1} c_{\lambda k} t^k e^{\lambda t}$ is the residue at λ of the mapping $z \mapsto e^{zt} b(z)/a(z)$. If the zeroes, $\lambda_1, \ldots, \lambda_p$, of $a(z)$ each have multiplicity one, then the expression (11.5.22) simplifies to

$$Y(t) = \int_{(-\infty,t]} \sum_{r=1}^{p} \alpha_r e^{\lambda_r(t-u)} dL(u), \tag{11.5.23}$$

where $\alpha_r = b(\lambda_r)/a'(\lambda_r)$. Hence $\{Y(t)\}$ has a corresponding *canonical representation* as a linear combination of (possibly complex-valued) CAR(1) processes,

$$Y(t) = \sum_{r=1}^{p} \alpha_r Y_r(t), \tag{11.5.24}$$

where $Y_r(t) = \int_{(-\infty,t]} e^{\lambda_r(t-u)} dL(u)$. Notice that the driving process L is the same for each of the component processes $\{Y_r(t)\}$, so they are not independent. Corresponding to the canonical decomposition (11.5.24), if $E(L(1)^2) < \infty$, there is an analogous representation of the autocovariance function when $E(L(1)^2) < \infty$, namely

$$\gamma(h) = \sum_{r=1}^{p} \beta_r e^{\lambda_r |h|}, \tag{11.5.25}$$

where $\beta_r = \sigma^2 b(\lambda_r) b(-\lambda_r)/[a(-\lambda_r) a'(\lambda_r)]$. □

Example 11.5.1. Stochastic Volatility

The stochastic volatility process, h, appearing in Example 7.5.4 was defined as

$$h(t) = \int_{(-\infty,t]} e^{\lambda(t-u)}\, dL(u), \quad \text{where } \lambda < 0, \tag{11.5.26}$$

i.e., as a Lévy-driven CARMA(1, 0) process with $a(z) = z - \lambda$ and $b(z) = 1$. For non-negativity of h, the Lévy process L is required to be a subordinator, for example, a gamma process with characteristic function $Ee^{i\theta L(t)} = (1 - i\theta/\beta)^{-\alpha t}$, $EL(t) = \alpha t/\beta$ and $\text{Var}(L(1)) = \alpha t/\beta^2$. Then $Eh(t) = \alpha t/(\beta|\lambda|)$ and the autocovariance function of h is, from (11.5.27), $\gamma_h(s) = e^{\lambda|s|}\alpha/(2\beta^2|\lambda|)$. For any finite-variance Lévy-driven CARMA(1,0) model for stochastic volatility, the autocorrelation function is necessarily of the form $\rho(s) = e^{\lambda|s|}$ for some negative λ. In order to relax this constraint a non-negative CARMA(p, q) model for h can be employed (see e.g., Brockwell and Lindner 2012).

□

Remark 5. Marginal distributions. The condition (11.5.17) clearly does not require $L(1)$ (and consequently $Y(t)$) to have finite variance. In fact the condition $E|L(1)|^r < \infty$ for *some* $r > 0$ is sufficient to ensure that (11.5.17) holds. Given a CARMA(p, q) process Y driven by a Lévy process L with characteristic function $E(e^{i\theta L(t)}) = \exp(t\xi(\theta)), \theta \in \mathbb{R}$, (see Appendix D.1), the joint characteristic function of $Y(t_1), \ldots, Y(t_n)$ can be expressed in terms of the coefficients of the polynomials a and b and the characteristic exponent $\xi(\cdot)$ of L (see Brockwell (2014)). In particular the logarithm of the marginal characteristic function of $Y(t)$ is

$$1\text{n}Ee^{i\theta Y(t)} = \int_0^\infty \xi(\theta \mathbf{b}' e^{Au}\mathbf{e})du, \theta \in \mathbb{R}. \tag{11.5.27}$$

For the CAR(1) process h defined by (11.5.26) this simplifies to

$$1\text{n}Ee^{i\theta h(t)} = \int_0^\infty \xi(\theta e^{\lambda u})du = |\lambda|^{-1}\int_0^\theta y^{-1}\xi(y)dy, \tag{11.5.28}$$

(where $\int_0^\theta := -\int_\theta^0$ if $\theta < 0$). If, for example, $L(1)$ has a symmetric stable distribution with $1\text{n}Ee^{i\theta L(1)} = -c|\theta|^\alpha, c > 0, 0 < \alpha \le 2$, then $E|L(1)|^r < \infty$ for all $r \in (0, \alpha)$ and from (11.5.28) we find at once (Problem 11.8) that,

$$1\text{n}Ee^{i\theta h(t)} = -\frac{c}{\alpha|\lambda|}|\theta|^\alpha, \tag{11.5.29}$$

in other words $h(t)$ has a symmetric stable distribution with the same exponent α as $L(1)$ but with the parameter c replaced by $c/(\alpha|\lambda|)$.

Problems

11.1 Find a transfer function model relating the input and output series X_{t1} and X_{t2}, $t = 1, \ldots, 200$, contained in the ITSM data files APPJ.TSM and APPK.TSM, respectively. Use the fitted model to predict $X_{201,2}$, $X_{202,2}$, and $X_{203,2}$. Compare the predictors and their mean squared errors with the corresponding predictors and mean squared errors obtained by modeling $\{X_{t2}\}$ as a univariate ARMA process and with the results of Problem 8.7.

11.2 Verify the calculations of Example 11.2.1 to fit an intervention model to the series SB.TSM.

11.3 If $\{X_t\}$ is the linear process (11.3.1) with $\{Z_t\} \sim \text{IID}\left(0, \sigma^2\right)$ and $\eta = EZ_t^3$, show that the third-order cumulant function of $\{X_t\}$ is given by

$$C_3(r, s) = \eta \sum_{i=-\infty}^{\infty} \psi_i\psi_{i+r}\psi_{i+s}.$$

Use this result to establish equation (11.3.4). Conclude that if $\{X_t\}$ is a Gaussian linear process, then $C_3(r,s) \equiv 0$ and $f_3(\omega_1, \omega_2) \equiv 0$.

11.4 If $a > 0$ and $Y(0)$ has mean b/a and variance $\sigma^2/(2a)$, show that the process defined by (11.5.3) is stationary and evaluate its mean and autocovariance function.

11.5 The file TRINGS.TSM contains normalized tree-ring widths of a Colorado pine for the years 525–774 (Donald Graybill 1984) from the file CO522.DAT at http://www-personal.buseco.monash.edu.au/hyndman/TSDL/.

a. Use exact maximum likelihood estimation to fit a fractionally integrated ARMA model to the first 230 tree-ring widths and use the model to generate forecasts and 95% prediction bounds for the last 20 observations (corresponding to $t = 231, \ldots, 250$). Plot the entire data set with the forecasts and prediction bounds superposed on the graph of the data.
b. Repeat part (a), but this time fitting an appropriate ARMA model. Compare the performance of the two sets of predictors.

11.6 The tent map with parameter $s \in (1, \infty)$ is the function

$$g(x) = sxI_{[0,1/s)}(x) + \frac{s}{s-1}(1-x)I_{[1/s,1]}(x), \quad x \in [0,1],$$

where I_A denotes the indicator function of the set A. If X_0 has the uniform distribution on $[0, 1]$ (written more concisely as $X_0 \sim U$) and if $\{X_n\}$ is the sequence defined by $X_n = g(X_{n-1})$, $n = 1, 2, \ldots$, then $\{X_n\}$ is a Markov chain and $X_n \sim U$ for all $n \in \{0, 1, 2, \ldots\}$, so that $\{X_n\}$ is strictly (and weakly) stationary.

a. Show that in the symmetric case ($s = 2$), $\{X_n\} \sim \text{WN}(0, 1/12)$.
b. In the general case, $X_n - 0.5 = \phi(X_{n-1} - 0.5) + Z_n$, $n = 1, 2, \ldots$, where $\phi = (2/s) - 1$ and $\{Z_n\}$ is an uncorrelated (but strongly dependent) sequence of random variables with mean zero and variance $(1-\phi^2)/12$. (See Sakai and Tokumaru 1980.)

11.7 A Lévy-driven CARMA(2,1) process is defined by the stochastic differential equation,

$$(D^2 + 1.5D + .5)Y(t) = (D + .2)DL(t), t \in \mathbb{R},$$

where L is a Poisson process with jump-rate ρ.

a. Calculate $EY(t)$.
b. Use (11.5.23) to determine the canonical decomposition of $\{Y(t)\}$.
c. Use (11.5.25) to determine the autocovariance function of $\{Y(t)\}$.

11.8 Use (11.5.27) to verify (11.5.28) and (11.5.29). If $L(1) \sim \text{N}(0, \sigma^2)$, use (11.5.29) to determine the distribution of $h(t)$, as defined by (11.5.26).