CHAPTER 11

FRACTIONAL AUTOREGRESSIVE-MOVING AVERAGE MODELS

11.1 INTRODUCTION

As explained in detail in Chapter 10, the well known *Hurst Phenomenon* defined in Section 10.3.1 stimulated extensive research in the field of stochastic hydrology. One valuable by-product of this research was the development of *long memory models* (see Sections 2.5.3, 10.3.3 and 11.2.1 for a definition of long memory). In particular, the *fractional Gaussian noise (FGN)* model of Section 10.4 possesses long memory and was developed within stochastic hydrology as an attempt to explain the Hurst Phenomenon through the concept of *long term persistence.*

The FGN model is not the only kind of stochastic model having long memory. As a matter of fact, due to its rather inflexible design and the difficulties encountered when applying it to real data (see Section 10.4), researchers have studied a variety of long memory models. The objective of this chapter is to present the most flexible and useful class of long memory models that have currently been developed. More specifically, this family is called the *fractional autoregressive-moving average (FARMA) group of models* (Hosking, 1981; Granger and Joyeux, 1980) because it arises as a natural extension of the ARIMA(p,d,q) models of Chapter 4. By allowing the parameter d in an ARIMA(p,d,q) model to take on real values, the resulting FARMA model possesses long memory for d falling within the range $0 < d < 1/2$.

A sound explanation for the Hurst phenomenon is presented in Section 10.6. In particular, by properly fitting ARMA models to a variety of geophysical time series, it is shown using simulation that the *ARMA models statistically preserve the Hurst statistics* consisting of the *rescaled adjusted range (RAR)* and the *Hurst coefficient K*. Because FARMA models are simply extensions or generalizations of ARMA (Chapter 3) and ARIMA (Chapter 4) models, one could also consider FARMA models in statistical experiments similar to those given in Section 10.6. Nonetheless, from a physical viewpoint hydrologic phenomena such as annual riverflows do not possess long memory or persistence since current flows do not depend upon annual flows that took place hundreds or thousands of years ago. Hence, for these kinds of series, ARMA models can adequately explain the Hurst phenomenon. However, the reader should keep in mind that there may be series that have long term memory and for these data one can employ FARMA models.

In the next section, the FARMA model is defined and some of its main statistical properties are described. Within Section 11.3, it is explained how FARMA models can be fitted to time series by following the identification, estimation and diagnostic check stages of *model construction*. Although good model building tools are now available, further research is required for developing more comprehensive estimation procedures. Methods for *simulating* and *forecasting* with FARMA models are given in Section 11.4. Before the conclusions, FARMA models are fitted to hydrological time series to illustrate how they are applied in practice. Parts of the presentations provided in Sections 11.2 to 11.5 were originally given in a paper by Jimenez et al. (1990).

11.2 DEFINITIONS AND STATISTICAL PROPERTIES

11.2.1 Long Memory

Persistence or *long term memory* is the term used to describe a time series that has either an autocorrelation structure that decays to zero slowly with increasing lag or equivalently a spectral density that is highly concentrated at frequencies close to zero. This autocorrelation structure suggests that the present state of the process must be highly dependent on values of the time series lying far away in the past, and, hence, to model the process the whole past should be incorporated into the description of the process.

A variety of precise mathematical definitions for long memory are given by authors such as Eberlein and Taqqu (1986), Davison and Cox (1989) as well as other authors cited in this chapter and Chapter 10. A simple definition that captures the essence of persistence, is the one presented in Sections 2.5 and 10.3.3. More specifically, a time series process can be classified according to the behaviour of the memory of the process where memory is defined as

$$M = \sum_{k=-\infty}^{\infty} |\rho_k|, \qquad [11.2.1]$$

where ρ_k is the theoretical ACF at lag k for the process. A long term memory process is defined as a process with $M = \infty$, whereas a short term memory process has $M < \infty$. The M term is often used as a mixing coefficient for stationary time series (Brillinger, 1975) to indicate the rate at which the present values of the time series are independent of the far away past values. The asymptotic independence between values of the time series well spaced in time where the mixing rate is given by $M < \infty$, has been traditionally used by time series analysts to prove results relating to normality, asymptotic behaviour of a quantity like the sample ACF, parameter estimates obtained either by maximum likelihood estimation or by the method of moments, hypothesis testing, and Portmanteau tests. Hence, most of the findings usually used in time series analysis are not necessarily true for long term memory processes because these processes have an infinite memory.

Besides hydrology, meteorology and geophysics, the classification of time series according to short and long memory may be useful in other areas (Cox, 1984; Parzen, 1982) such as economics (Granger, 1980; Granger and Joyeux, 1980). This classification has been used even with other types of stochastic processes (Cox, 1984), although the memory has had other definitions. An alternative definition of long term memory, essentially equivalent to the definition given above, is to consider time series processes whose ACF decays as

$$\rho_k = O(k^{-a}), \qquad [11.2.2]$$

where a lies in the interval (0,1).

An advantage of the FARMA family of models, defined in the next subsection, is that it can describe both short and long term memory. Furthermore, it constitutes a direct generalization of the ARMA and ARIMA models of Chapters 3 and 4, respectively.

11.2.2 Definition of FARMA Models

As explained in Chapter 4, a device frequently used in time series modelling is differencing the series, if it is thought that its mean function is time dependent. A time dependent mean could produce sample autocorrelations that, rather than decaying to zero exponentially like the ACF's of ARMA models, decay to zero much more slowly. In fact, if the rate of decay of the ACF seems to depend linearly upon the lag the usual approach is to work with the first differences of the time series. For the type of processes studied in this chapter, the ACF decays to zero at a rate slower than exponential, but faster than linear. This suggests the use of a device similar to the usual differencing operator, to model time series having a slowly decaying ACF with long memory. In fact, FARMA models generalize in a natural form the concept of ARIMA time series models containing differencing operators.

The FARMA family of models is a generalization of the ARIMA models of Chapter 4 which in turn constitute an extension of the ARMA models of Chapter 3. To define FARMA models, the concept of differencing is generalized by means of the filter

$$\begin{aligned}\nabla^d(B) &= (1-B)^d = \sum_{j=0}^{\infty}\binom{d}{k}(-B)^k \\ &= 1 - dB - \frac{1}{2}d(1-d)B^2 - \frac{1}{6}d(1-d)(2-d)B^3 - \cdots \qquad [11.2.3]\end{aligned}$$

where B is the backward shift operator. For an ARIMA model, the values of d in the filter in [11.2.3] are restricted to be zero when the series being modelled is stationary and to be a positive integer when the series must be differenced to remove nonstationarity. When d can be *fractional,* and hence take on real values, the above filter becomes the one used with FARMA models. As is explained in Section 11.2.3 on the statistical properties of FARMA models, the value of d controls the memory of the process.

As originally suggested independently by Hosking (1981) and Granger and Joyeux (1980), a *FARMA(p,d,q) model* for modelling a series z_t is defined as

$$\phi(B)\nabla^d z_t = \theta(B)a_t \qquad [11.2.4]$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is the autoregressive (AR) operator of order p having the AR parameters $\phi_1, \phi_2, \ldots, \phi_p$; $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is the moving average (MA) operator of order q having the MA parameters $\theta_1, \theta_2, \ldots, \theta_q$; ∇^d is the fractional differencing operator defined in [11.2.3]; and a_t is a white noise process that is identically and independently distributed with a mean of zero and variance of σ_a^2 (i.e. IID$(0,\sigma_a^2)$). As is also the case for the standard ARMA model of Chapter 3, the operators $\phi(B)$ and $\theta(B)$ are assumed to have all roots lying outside the unit circle and to have no common roots. Finally, no mean level, μ, is written in [11.2.4] since $\nabla^d\mu = 0$ for positive d.

One can write the FARMA process in [11.2.4] as

$$\nabla^d z_t = \frac{\theta(B)}{\phi(B)}a_t \qquad [11.2.5]$$

One can interpret the short memory component of the FARMA process as being modelled by applying the usual ARMA filter given by $\theta(B)/\phi(B)$ to the a_t time series that is IID$(0,\sigma_a^2)$. The *fractional differencing filter* ∇^d handles the long memory part of the overall process.

For a nonseasonal FARMA model, the notation FARMA(p,d,q) is employed where p and q are the orders of the AR and MA operators, respectively, and d is a parameter in the filter in [11.2.3] and can take on real values. When d is a positive integer, the FARMA(p,d,q) model is equivalent to an ARIMA(p,d,q) model where the acronym ARIMA stands for autoregressive integrated moving average (see Chapter 4). If $d = 0$, the FARMA(p,d,q) model is identical to a short memory ARMA(p,q) model of Chapter 3. When $p = q = 0$, the FARMA(p,d,q) model reduces to

$$\nabla^d z_t = a_t \qquad [11.2.6]$$

which is called a *fractional differencing model*. The labels that can be used for the various types of FARMA, ARIMA and ARMA models are listed in Table 11.2.1. In this table, the FARMA(p,d,q) model is the most general and it contains all the other models as subsets.

As an example of how to write a specific FARMA(p,d,q) model, consider the case of a FARMA(0,0.3,1) for which $p = 0$, $q = 1$ and d has a real value of 0.3. From [11.2.4] this model is given as

$$(1 - B)^{0.3} z_t = (1 - \theta_1 B) a_t$$

Using [11.2.3], the fractional differencing operator is expanded as

$$(1 - B)^{0.3} = 1 - 0.3B - \frac{1}{2}(0.3)(1 - 0.3)B^2 - \frac{1}{6}0.3(1 - 0.3)(2 - 0.3)B^3 - \cdots$$

$$= 1 - 0.3B - 0.105B^2 - 0.060B^3 - \cdots$$

Substituting the expanded fractional differencing operator into the equation for the FARMA(0,0.3,1) model results in

$$(1 - 0.3B - 0.105B^2 - 0.059B^3 - \cdots) z_t = (1 - \theta_1 B) a_t$$

or

$$z_t - 0.3z_{t-1} - 0.105z_{t-2} - 0.059z_{t-3} - \cdots = (1 - \theta_1 B) a_t$$

or

$$z_t = 0.3z_{t-1} + 0.105z_{t-2} + 0.059z_{t-3} + \cdots + a_t - \theta_1 a_{t-1} \qquad [11.2.7]$$

From this equation, one can see that the weights for the z_t terms are decreasing as one goes further into the past.

For the theoretical definition of the FARMA(p,d,q) model in [11.2.4], the a_t series is assumed to be IID($0,\sigma_a^2$). In order to develop estimation and other model construction methods, usually the a_t's are assumed to be normally distributed. Recall that the assumption that the a_t's are NID($0,\sigma_a^2$) for application purposes is also invoked in Part III for the ARMA and ARIMA models of Chapters 3 and 4, respectively, as well as most other models presented in this book.

Table 11.2.1. Names of models.

Values of d	Values of p	Values of q	Equivalent Model Names	Chapters
Real value	p	q	FARMA(p,d,q)	11
Real value	0	0	FARMA(0,d,0), Fractional Differencing	11
Positive Integer	p	q	ARIMA(p,d,q), FARMA(p,d,q) for which d is a positive integer	4
Positive Integer	0	q	ARIMA(0,d,q), IMA(d,q), FARMA(0,d,q) for which d is a positive integer	4
0	p	q	ARMA(p,q), FARMA(p,0,q), ARIMA(p,0,q)	3
0	p	0	ARMA(p,0), AR(p), FARMA(p,0,0), ARIMA(p,0,0)	3
0	0	q	ARMA(0,q), MA(q), FARMA(0,0,q), ARIMA(0,0,q)	3
0	0	0	ARMA(0,0), FARMA(0,0,0), ARIMA(0,0,0), White Noise	3

When the residuals of a fitted FARMA model and, hence, the original series are not normally distributed, one approach to overcome this problem is to invoke the Box-Cox transformation defined in [3.4.30]. Subsequent to this, one can estimate the parameters of all the model parameters, including d, for the FARMA model fitted to the transformed series.

Three classes of seasonal models are given in Part VI of the book. The definition for nonseasonal FARMA models can be easily extended to create *long memory seasonal FARMA models* for each of the three kinds of seasonal models. To create a seasonal FARMA model, which is similar to the seasonal ARIMA model of Chapter 12, one simply has to incorporate a seasonal fractional differencing operator as well as seasonal AR and MA operators into the basic nonseasonal FARMA model in [11.2.4]. A deseasonalized FARMA model is formed by fitting a nonseasonal FARMA model to a series which has been first deseasonalized using an appropriate deseasonalization technique from Chapter 13. To obtain a periodic FARMA model that reflects the periodic ARMA model of Chapter 14, one simply defines a separate nonseasonal FARMA model for each season of the year. Future research could concentrate on developing comprehensive model building techniques, especially for the cases of seasonal FARMA and periodic FARMA models. Hui and Li (1988) have developed maximum likelihood estimators for use with periodic FARMA(0,d,0) (i.e. periodic fractional differencing) and periodic FARMA(p,d,0) models.

Another function for FARMA modelling is to allow the noise terms of transfer function-noise (Part VII), intervention (Part VIII), and multivariate ARMA (Part IX) models to follow a FARMA model. The definitions of these models are simple. However, the development of model construction techniques, especially efficient estimation methods, would be a formidable task. Hence, this should only be undertaken if practical applications using real world data indicate a need for these kinds of long memory models.

Keeping in mind that FARMA modelling can be expanded in many directions, the rest of this chapter is restricted to the case of nonseasonal FARMA models. In the next subsection, some theoretical properties of the FARMA(p,d,q) model in [11.2.4] are given.

11.2.3 Statistical Properties of FARMA Models

As explained by Hosking (1981), the FGN model of Section 10.4 is in fact a discrete-time analogue of continuous-time fractional noise. Another discrete time version of continuous-time fractional noise is the fractional differencing (i.e. FARMA(0,d,0)) model in [11.2.6]. An advantage of the fractional differencing model over FGN is that it can be expanded to become the comprehensive FARMA(p,d,q) model in [11.2.4], which in turn is a generalization of the ARIMA model.

The basic properties of FARMA processes are presented by Hosking (1981) and Granger and Joyeux (1980). As explained by Jimenez et al. (1990), they found among other things that:

(a) For the process to be stationary, $d < 0.5$ and all the roots of the characteristic equation $\phi(B) = 0$ must lie outside the unit circle.

(b) For the process to be invertible, $d > -0.5$ and all the roots of the characteristic equation $\theta(B) = 0$ must lie outside the unit circle.

(c) Because of (a) and (b), if $-\frac{1}{2} < d < \frac{1}{2}$, the FARMA($p$,$d$,$q$) process is both stationary and invertible.

(d) For $0 < d < \frac{1}{2}$, the process has long memory (see Section 11.2.1 for definitions of long memory).

(e) The ACF behaves as

$$\rho_k = 0(k^{-1+2d}). \qquad [11.2.8]$$

(f) The process is self-similar, which means that the stochastic properties of the process are invariant under changes of scale.

Several probabilists (Rosenblatt, 1961, 1979, 1981; Taqqu, 1975) have studied the behaviour of statistics derived from time series processes where the ACF behaves as in [11.2.8] and d is positive. They found that:

(g) The sample mean times $N^{1/2+d}$, where N is the number of observations, converges in law to a normal random variable.

(h) The sample autocovariances do not converge asymptotically to a normal random variable.

The result in (g) about the mean is of some interest to hydrologists because it has been found that processes thought to possess long term memory have a sample mean that seems to indicate slow changes in trend. This wandering of the sample mean can be explained in terms of (g) above. This shifting level process for modelling a changing mean is referred to in Section 10.3.3 and references are provided at the end of Chapter 10. Arguments to show that persistence in geophysical processes is due to a slowly changing trend cannot be based just on statistical behaviour but should use geophysical insight. Note also that the above results are not restricted to the FARMA process case but that they are valid for any time series whose ACF behaves as in [11.2.8].

An important, although seemingly trivial extension of the original definition of FARMA processes by Hosking (1981) and Granger and Joyeux (1980) is to relax the assumption that the mean of the time series is zero. The extension of the model given above to the case of a nonzero mean is straightforward. However, what is very important to note is that if a constant, in particular the mean, is passed through the filter ∇^d, the output, for the case of a positive d, is zero. Hence, the mean of the process does not have to appear in the equations that define the model. Nevertheless, it should be noted that the mean is a well defined quantity for this process when $d < 0.5$.

The aforementioned property is very important for determining the stochastic properties of the estimates for the parameters. This is because the sample mean can be used as an estimate for the mean of the time series and the slow rate of convergence of the sample mean as given in (g) above does not affect the asymptotic rate of convergence of the estimates for the other parameters to a Gaussian random variable, where this rate is the usual N^{-1}.

Another interesting feature is that the filter ∇^d can smooth some special trends as can be seen easily for the case $d = 1$ when the trend is a straight line. When $0 \le d < 0.5$ the filter ∇^d smooths slowly changing trends. Hence, even if the process mean is slowly changing, FARMA models could be used to model the time series in much the same way that ARIMA models are employed with a deterministic drift component.

Another consequence of the fact that

$$\nabla^d(z_t - \mu) = \nabla^d z_t, \quad d > 0 \qquad [11.2.9]$$

where z_t is the value of the process at time t with a theoretical mean μ, is that the process behaviour is independent of the mean. In the stationary ARMA process, on the other hand, the local behaviour of the process does depend on the mean. This can be seen by considering the value of the process conditioned on the past as given by $E\{z_{t+1} | z_s, \ s \le t\}$. In the ARMA case, this quantity depends on $\mu = E\{z_t\}$ but in the FARMA case with $d > 0$, it does not. In the remaining parts of this section, unless stated to the contrary, the mean μ will be assumed equal to zero.

An important consequence of the slow rate of decay to zero of the ACF as given by [11.2.8] is that Bartlett's formula (Bartlett, 1946) for the variances and the covariances of the estimated autocovariance function (ACVF), $\{\hat{\gamma}_k\}$, has to be modified accordingly. In fact, the exact formula for the variance is given by

$$var(\hat{\gamma}_k) = N^{-1} \sum_{m=-(N-k)+k}^{(N-k-1)} \left\{1 - |m| + \frac{k}{N}\right\}\left\{\gamma_m^2 + \gamma_{m+k}\gamma_{m-k}\right\} \qquad [11.2.10]$$

Then, by [11.2.8]

$$var(\hat{\gamma}_k) = \begin{cases} 0(N^{-1}), & \text{if } d \le 0.25 \\ 0(N^{4d-2}), & \text{if } d > 0.25 \end{cases} \qquad [11.2.11]$$

Hence, if $d < 0.25$ then $var(\hat{\gamma}_k) = 0(N^{-1})$, which is the same order as in the case of a short memory process. However, if $0.25 < d < 0.5$ the order of $var(\hat{\gamma}_k)$ is larger than N^{-1}. In fact as d approaches 0.5 the variance approaches a quantity of order one. This implies that the stochastic variability of the estimated ACVF is higher for long term memory processes with $0.25 < d < 0.5$ than for short term memory processes. Moreover, the order of the variance depends on the unknown quantity d. Finally, similar results are valid for the covariances of the estimated ACF.

An interesting subset of the FARMA(p,d,q) family of processes in [11.2.4] is the FARMA(0,d,0) process in [11.2.6] which is referred to as the fractional differencing model. This model has been studied in some detail and expressions for the ACF, partial autocorrelations function (PACF), partial linear regression coefficients, and inverse autocorrelations are known (Hosking, 1981, 1984, 1985). One important fact about the stochastic behaviour of a FARMA(0,d,0) process is that all its autocorrelations are positive if d is positive, and they are negative otherwise. Also, all the partial autocorrelations of the FARMA(0,d,0) model have the same sign as the *persistence parameter* d, and their rate of decay to zero is of the same order as the inverse of the lag. Because of these limitations of the structure of the ACF, fractionally differenced noise is passed through an ARMA filter in order to obtain a richer autocorrelation structure within the framework of a FARMA(p,d,q) process.

As suggested by Jiminez et al. (1990), it is possible to generalize the filter $(1-B)$ in another form, which is closely related to the $(1-B)^d$ filter in [11.2.3]. In particular, this filter is defined by $(1+B)^d$. Note that the associated transfer function also has a root on the unit circle at $B=-1$. The coefficients of this filter are the same as those of the filter $(1-B)^d$ except for the sign and hence the process also has long term memory if $d > 0$, it is stationary if $d < 0.5$, and invertible if $d > -0.5$. However, the interesting fact is that although the absolute values of the autocorrelations are the same for both filters, the autocorrelations of the filter $(1+B)^d$ alternate in sign. More general autocorrelations structures could be obtained by generalizing the filters to accommodate complex roots on the unit circle. The class of processes studied in this chapter are particular cases of the more general processes that result by filtering white noise through the filters defined by $(1-\varepsilon B)^d$, where the parameter ε lies in the range $|\varepsilon| \le 1$. In this chapter it is assumed that $\varepsilon = 1$, or -1.

11.3 CONSTRUCTING FARMA MODELS

11.3.1 Overview

To fit a FARMA(p,d,q) model to a given time series, one can follow the usual identification, estimation and diagnostic check stages of model construction. Model building procedures are fairly well developed for the case of the fractional differencing (i.e. FARMA($0,d,0$) model in [11.2.6]). However, further research is required for obtaining a comprehensive set of tools for building the FARMA(p,d,q) models in [11.2.4]. Of particular importance is the need for good estimation techniques that are both computationally and statistically efficient, as well as capable of estimating the mean level along with the other FARMA model parameters. Unlike ARIMA(p,d,q) models where d is fixed at zero or some positive integer value prior to estimating the other model parameters for the differenced series, one must, of course, estimate d in the FARMA(p,d,q) model simultaneously with the other model parameters.

11.3.2 Identification

To identify a suitable ARMA model (Chapter 3) or ARIMA model (Chapter 4) to fit to a given time series, one can examine the characteristics of graphs of the sample ACF, PACF, IACF and IPACF (Chapter 5). By knowing the behaviour of the theoretical ACF, PACF, IACF and IPACF for ARMA or ARIMA models, one can determine from the sample plots which parameters to include in the model. If more than one model is fitted to the series, an automatic selection criterion such as the AIC (see Section 6.3) can be used to select the best one.

The sample ACF, PACF, IACF and IPACF can also be used to identify a FARMA(p,d,q) model for fitting to a series. If the series is stationary and the sample ACF dies off slowly, then d should be estimated to account for this long term persistence. Hosking (1981) gives formulae for the theoretical ACF, PACF and IACF for the case of the fractional differencing model in [11.2.6]. Further research is required to obtain formulae for the theoretical PACF, IACF and IPACF for FARMA(p,d,q) models. By comparing the behaviour of the sample graphs to the theoretical findings one can decide upon which parameters to include in the FARMA(p,d,q) model. Additional procedures for model identification are presented in Section 11.5 with the applications.

11.3.3 Estimation

This section follows the research findings of Jimenez et al. (1990). However, the reader may also wish to refer to the FARMA estimation procedures presented by Boes et al. (1989), and by Brockwell and Davis (1987, pp. 464-478), as well. As noted earlier in Section 11.2.3, the coefficients of the filter $(1-B)^d$ and $(1+B)^d$ only differ in sign. Because the estimation results of this section are valid for both filters, everything is described only for the filter $\nabla^d = (1-B)^d$.

There are several estimation procedures available in the literature. Frequency domain methods do not seem to be as efficient as estimators based on the time domain representation. Hence, only time domain methods are considered here.

Because of the slow rate of convergence of the sample mean to the true mean as can be seen in (g) in Section 11.2.3, it is of utmost importance to find a more efficient estimator of the mean. The most obvious candidate is the maximum likelihood estimate of the mean (McLeod and Hipel, 1978a), which is given by

$$\hat{\mu} = (\mathbf{z}^T\Sigma^{-1}\mathbf{1})(\mathbf{1}^T\Sigma^{-1}\mathbf{1}) \qquad [11.3.1]$$

where $z^T=(z_1, z_2, \ldots, z_N)$ is the $1\times N$ vector of observations, Σ is the autocorrelation matrix of the time series, and 1 represents a column vector of ones. However, it can be shown that the sample mean is efficient for the case when the persistence parameter d is nonnegative, and it is not efficient when the persistence parameter is negative. This agrees with the common knowledge that overdifferencing can lead to inefficient estimates. Although it is difficult to give a physical meaning to antipersistence, a negative value of d can be useful from a purely fitting point of view as it has been observed that sometimes FARMA models with negative d arise while fitting them to a time series, and, therefore, it is important in these cases to estimate the mean of the process using the maximum likelihood estimate as given by the above formula. The evaluation of the above formula can be performed efficiently using either Cholesky decomposition (Healy, 1968) of the inverse of Σ given by the partial linear regression coefficients, $\{\phi_{i,t}\}$ (which can be obtained easily by the Levison-Durbin algorithm (Durbin, 1960)), or by the Trench algorithm for the inverse of a Toeplitz matrix (Trench, 1964). For the particular case of fractionally differenced noise, Hosking (1981) gives a closed expression for the reflection coefficients or partial linear regression coefficients. Hence, in this case a closed expression for the maximum likelihood estimate of the mean is known. For the situation where $\varepsilon = -1$ in the filter, mentioned at the end of Section 11.2.3, this closed expression is still valid with appropriate sign changes. In terms of the partial linear regression coefficients, the following expression could be used to evaluate the maximum likelihood estimate $\hat{\mu}$

$$\hat{\mu} = N^{-1}\frac{\sum_{t=0}^{N-1}(z_t - \phi_{1,t}z_{t-1} - \phi_{2,t}z_{t-2} - \cdots - \phi_{t,t}z_o)}{\sum_{t=0}^{N-1}(1 - \phi_{1,t} - \phi_{2,t} - \cdots - \phi_{t,t})^2} \qquad [11.3.2]$$

In this section, it is assumed that the *persistence parameter d* is nonnegative, the sample mean is used as the estimate of the mean, and the sample mean has been subtracted from each observation.

There are two methods available to estimate the remaining parameters in the time domain: exact maximum likelihood estimation or an approximation of the filter ∇^d. Most of the maximum likelihood estimation algorithms depend on computing the one step ahead prediction errors, a_t, which can be computed in terms of the partial linear regression coefficients. These coefficients can be computed efficiently by the Durbin-Levinson algorithm. Finally, with these values of e_t the estimates of the parameters are obtained by minimizing the modified sum of squares function given by:

$$\ln l = \sum_{t=1}^{N}(N - t + 1)\ln(1 - \phi_{t,t}^2) + \sum_{t=1}^{N}a_t^2/\sigma_t^2. \qquad [11.3.3]$$

Although the computation of estimates by maximum likelihood is statistically attractive, the amount of computations involved in the above scheme makes algorithms having fewer numbers of computations competitive alternatives.

The algorithm proposed by Li and McLeod (1986) is computationally economical and is presented in Appendix A11.1. The algorithm consists of approximating the filter ∇^d by the filter ∇_M^d where ∇_M^d is defined as the filter resulting by taking the first M terms of the filter ∇^d, i.e. by approximating the process by a "long" autoregression. Then the algorithm minimizes the sum of the squared residuals, where the residuals are obtained as the output of the filters ∇_M^d and the ARMA filter. To compute the residuals, an algorithm such as the one given by McLeod and Sales (1983) could be used. Also, as recommended by Box and Jenkins (1976) the sum of squared residuals could be extended back in time by backforecasting. Note that the approximation of ∇^d by ∇_M^d is not the optimal approximation in a least squares sense. However, since the order to M is comparable with N, it has to be very close to the optimal approximation. The order of approximation necessary to obtain consistence estimates has been found to be of the order of $N^{1/2}$ and an ad hoc rule is to fit time series with at least 50 observations. The order of truncation M is chosen as a number of between $N/4$ and $N^{1/2}$, by trying to balance the degree of approximation of the filter ∇_M^d to the filter ∇^d and the amount of computations involved. Nonetheless, for N close to 50, M is taken as half the number of observations. The amount of computations using this algorithm is much smaller than that for the maximum likelihood approach. Moreover, estimates obtained in this form are asymptotically equivalent to the maximum likelihood estimates and it seems that the finite sample estimates are generally close enough to the maximum likelihood estimates. Li and McLeod (1986) studied the asymptotical distributions of the estimates when the mean of the time series is known. They derived closed form expressions for the variances of the asymptotically normal distributions of the estimates. It can be demonstrated that the estimation of the mean by the sample mean does not affect the above asymptotic results. However, these results are not likely to hold for a finite sample size because of the long term persistence and the parameter d is constrained to lie in the open interval (-0.5, 0.5). In practice, the interval is closed and it can be observed using simulation that if the persistence parameter is close to 0.5, there is a high probability for the estimate of d to be equal to 0.5. A similar phenomenon was observed for the ARMA(0,1) model by Cryer and Ledolter (1981). Hence, the rate of convergence of the estimates depends on the parameters even for relatively large sample sizes of more than 200. Additionally, it should be noted that the above method is very similar to fitting an autoregressive process of order one if d is not close to 0.5, say less than 0.3.

Bootstrapping a Time Series Model

Because the FARMA model is an infinite autoregression and, moreover, is nonstationary when $d \geq 0.5$, it is expected that finite sample properties of the estimates are different than the large sample approximations. Consequently, it is interesting to obtain further information about these finite sample distributions. One interesting possibility to increase one's knowledge of the finite sample distribution of the estimates is by using the bootstrapping technique proposed by Cover and Unny (1986).

Since Efron (1979) proposed the bootstrap, there have been several proposals to extend the original technique to time series analysis. However, most of them have used a straightforward generalization of the original bootstrap with the consequence that what they did was to use distorted models. The idea of Cover and Unny (1986) is to inject randomness into the loss function by resampling the positions of the residuals and not the observations themselves (i.e. the time lags are resampled with replacement and with the same probability). This resampling of the time lags is interesting because of the nature of data that depends strongly on the time coordinates. Note also that unlike other resampling plans, the assumption that the fitted model is the true model is not crucial. Also, it can be applied to any time series model and not just to a FARMA model. Moreover, the idea is valid for other stochastic processes.

The technique can then be described as follows:

(a) Draw a random sample of size N with replacements from the integers between 1 and N;

(b) Obtain estimates of the parameters by minimizing the sum of the squared residuals a_t^2 with weights equal to the number of times that the number t appeared in the random sample in (a);

(c) Repeat (a) and (b) a large enough number of times to obtain reliable estimates of the distribution characteristics of the estimated parameters.

This technique can greatly increase one's information about the parameter estimates as can be seen in the applications. However, further theoretical results are required to confirm theoretically the finite sample validity of the bootstrap approach.

11.3.4 Diagnostic Checks

To ascertain if a calibrated FARMA(p,d,q) model adequately fits a given series, one can employ diagnostic checks similar to those given in Chapter 7 for ARMA and ARIMA models. The innovations of the FARMA model in [11.2.4] are assumed to be Gaussian, homoscedastic (i.e. have constant variance) and white. When, in practice, the residuals of the fitted model are not always normal/homoscedastic, this can often be overcome by transforming the data using the Box-Cox transformation of [3.4.30]. The parameters of the FARMA(p,d,q) model can then be estimated for the transformed series and the residuals once again subjected to diagnostic checks.

The most important innovation assumption is independence. If the residuals of the fitted model are correlated and not white, then a different FARMA model or, perhaps, some other type of model, should be fitted to the series. The best check for whiteness is to examine the residual autocorrelation function (RACF) for the calibrated model, as is also the case for ARMA and ARIMA models (see Chapter 7). The large-sample distribution of the RACF for a FARMA model is given by Li and McLeod (1986) who also present a modified Portmanteau test statistic to check for whiteness.

11.4 SIMULATION AND FORECASTING

11.4.1 Introduction

After a FARMA(p,d,q) model has been fitted to a given series, the calibrated model can be employed for applications such as simulation and forecasting. The purpose of this section is to present simulation and forecasting procedures for use with FARMA models. Techniques for simulating and forecasting with ARMA and ARIMA models are presented in Part IV of the book. Finally, forecasting experiments in which fractional differencing models are used, in addition to other kinds of models, are given in Section 8.3.

11.4.2 Simulating with FARMA Models

Based upon a knowledge of closed expressions for the partial linear regression coefficients, $\phi_{k,j}$, fast algorithms for generating synthetic sequences from FARMA models can be given. Partial linear regression coefficients are defined as the values of α_k that minimize

$$E\{z_t - \alpha_1 z_{t-1} - \cdots - \alpha_t z_0\}^2 \qquad [11.4.1]$$

where E is the expectation operator. Thus, they are the values that minimize the one step ahead forecast errors. As is well known, the time series process can be written in terms of the innovations as:

$$z_t = a_t + \phi_{1,t} z_{t-1} + \cdots + \phi_{t,t} z_0, \qquad [11.4.2]$$

where the innovations $\{a_t\}$ are a sequence of independent Gaussian random variables with mean 0 and variance $\sigma_t^2 = \prod_{j=1}^{t}(1 - \phi_{j,j}^2)$. First consider the case of simulating fractionally differenced noise. Expressions for $\phi_{k,t}$ are presented by Hosking (1981), and recursive expressions are given by:

$$\phi_{t,t} = d/(t-d)$$

$$\phi_{j,t} = \phi_{j+1,t}(j+1)(t-j-d)/((j-1-d)(t-j)), \qquad [11.4.3]$$

Consequently, to simulate a FARMA(0,d,0) noise model it is only necessary to compute recursively $\phi_{k,t}$, generate a normal random variable and then use [11.4.2].

To simulate using a FARMA(0,d,q) model, the fractionally differenced noise is generated and then passed through the moving average filter. When generating synthetic data using a FARMA(p,d,0) model, one possible approach is to simulate the FARMA(0,d,0) model using above algorithm and after choosing p initial values, which can be done using the method in McLeod and Hipel (1978b), as explained below, generate recursively the other simulated values. Finally, the general FARMA(p,d,q) case can be obtained by a combination of the above methods.

Another possible method to generate synthetic sequences (McLeod and Hipel, 1978a,b) is to obtain the Cholesky decomposition of the matrix of the theoretical autocorrelations, Σ, and to multiply this decomposition matrix by a vector of independent Gaussian variables with mean zero and desired variance (see Section 9.4 for the case of ARMA models). Finally, a mean correction is added to the series. However, although this method is attractive for other models, it may be less desirable than the method described above because it involves the computation of the matrix of autocorrelations and the theoretical autocorrelations are given in terms of hypergeometric functions (Hosking, 1981). Thus, the computation task time necessary to compute the autocorrelations is much bigger than the computation time necessary to pass FARMA(0,d,0) noise through the different filters. However, once the ACF has been calculated and the required Cholesky decomposition obtained, this method is useful if many independent realizations of the process are to be simulated. Finally, both methods are equivalent in the case of the FARMA(0,d,0) model.

11.4.3 Forecasting with FARMA Models

Forecasting by using ARMA models is generally most useful when the forecaster is just interested in one step ahead or two steps ahead forecasts. This is because the forecast functions produced by ARMA models converge exponentially fast to the mean of the time series. Hence, in ARMA models long term forecasts are given by the mean μ or some estimate of it. This is not the case when the process has a long term memory, as is clear from the definition of persistence. For the case of a long memory process, the forecasting functions still converge to the mean μ;

however, the rate of convergence is not exponential but slower. For persistent time series, the rate of decay of the forecast function depends on the degree of persistence that the process possesses.

Another consequence of persistence in forecasting is that the variance of the forecast function of a persistent process decays to the variance of the process, σ_z^2 at a rate that could be substantially slower than exponential, depending on the degree of persistence. Therefore, confidence bounds for the l-step ahead forecasts of persistent processes are smaller than those of short term memory processes, if l is bigger than two or three. This can be seen if the time series model is written as a linear process (Box and Jenkins, 1976)

$$z_t = \sum_{k=0}^{\infty} \alpha_k a_{t-k}. \qquad [11.4.4]$$

Then, the l-step ahead forecast, $\hat{z}_t[l]$, is given by

$$\hat{z}_t[l] = \sum_{k=0}^{\infty} \alpha_{k+l} a_{t-k} \qquad [11.4.5]$$

but, for a FARMA model, $\alpha_k \approx k^{-1-d}$. Therefore,

$$var\{\hat{z}_t[l]\} = \sigma_z^2 \approx 0(l^{-1+2d}) \qquad [11.4.6]$$

Equation [11.4.5] is most helpful for forecasting if estimates of a_t are available and if the coefficients α_k decay to zero fast enough so that the necessary truncation involved in the computation of $\hat{z}_t[l]$ as given in [11.4.5] produces a negligible error. However, for FARMA models these coefficients do not decay fast enough, and, hence, expressions for the forecast function $\hat{z}_t[l]$ that do not involve approximations could be useful. The method proposed is based on the AR form of the time series as given by [11.4.2]. The forecast function is given by

$$\hat{z}_t[l] = \phi_{1,t}(l)z_t + \phi_{2,t}(l)z_{t-1} + \cdots + \phi_{t,t}(l)z_0 \qquad [11.4.7]$$

where

$$\phi_{i,t}(l) = \phi_{l,t+l} + \sum_{j=1}^{l-1} \phi_{j,t+l-1}\phi_{i,t}(j). \qquad [11.4.8]$$

This expression has advantages over the formula given in [11.4.5] because it does not involve approximation either by truncation of an infinite series or in the computation of the residuals. Moreover, by using [11.4.7] it is possible to show that

$$\hat{z}_t[l] \approx \phi_{l,t+l}z_t + \cdots + \phi_{t+l,t+l}z_0. \qquad [11.4.9]$$

Hence, as discussed above the forecast function decays to the theoretical mean μ at a rate slower than exponential. For example, for the FARMA(0,d,0) model

$$\hat{z}_t(l) \approx \frac{l^{-d-1}z_t + \cdots + (t+l)^{-d-1}z_0}{(-d-1)!}. \qquad [11.4.10]$$

11.5 FITTING FARMA MODELS TO ANNUAL HYDROLOGICAL TIME SERIES

To demonstrate how FARMA models are applied in practice, FARMA models are fitted to the fourteen hydrological time series listed in Table 11.5.1. The data consists of eleven annual river flows in m^3 /s from different parts of the world, two records of average annual rainfall in mm, and an annual temperature series in degrees Celsius. Because efficient estimation procedures are available for use with FARMA(0,d,0) (i.e. fractional differencing) and FARMA(p,0,q) (i.e. ARMA(p,q)) models, these are the models which are considered for fitting to the series. Estimation methods for use with FARM(0,d,0) and ARMA(p,q) models are presented in Sections 11.3.3 and Chapter 5, respectively.

Table 11.5.1. Annual time series used in the applications of FARMA models.

	Descriptions	Geographical Locations	Time Spans	Lengths
(1)	Saugeen River	Walkerton, Ontario, Canada	1915-1976	62
(2)	Dal River	near Norslund, Sweden	1852-1922	70
(3)	Danube River	Orshava, Romania	1837-1957	120
(4)	French Broad River	Asheville, N. Carolina	1880-1900	70
(5)	Gota River	near Sjotop-Vannersburg, Sweden	1807-1957	150
(6)	McKenzie River	McKenzie Bridge, Oregon	1900-1956	56
(7)	Mississippi River	St. Louis, Missouri	1861-1957	96
(8)	Neumunas River	Smalininkai, Lithuania	1811-1943	132
(9)	Rhine River	Basle, Switzerland	1807-1957	150
(10)	St. Lawrence River	Ogdensburg, New York	1800-1930	131
(11)	Thames River	Teddington, England	1883-1954	71
(12)	Rainfall	Fortaleza, Brasil	1849-1979	131
(13)	Rainfall	Philadelphia	1800-1898	99
(14)	Average temperature	Central England	1723-1970	248

In practice, the definition of long term memory in terms of $M = \infty$ in [11.2.1] is difficult to check and instead the persistence criterion given in [11.2.8] is used. Hence, a sample ACF that decays slowly to zero could indicate that the time series has long term memory. For those time series whose sample ACF decays to zero at a hyperbolic rate, the possibility of modelling them by FARMA models is considered. Within the fourteen data sets, the St. Lawrence Riverflows and the Philadelphia Rainfall series show an estimated ACF that seems to decay to zero hyperbolically. Therefore, these two data sets present evidence that suggests the use of FARMA models to fit them. The graph of the St. Lawrence Riverflow series against time and its sample ACF are shown in Figures II.1 and 3.2.1, respectively. For other records such as the Saugeen Riverflows and Rainfall at Fortaleza, the evidence, as given by the estimated ACF's, in favour of a persistence parameter is not so strong but it is a possibility. However, it should be remarked that if the persistence parameter d is close to zero and, hence, between 0 and 0.2, detection of long term memory by visual inspection of the autocorrelations can be difficult. Moreover, because Bartlett's formula needs to be multiplied by a factor of order N^{-1+4d} ($d \geq 0.25$), for the case of a FARMA(p,d,q) process, visual inspection of the sample ACF should be used with care when it is suspected that the process under analysis could have long term memory.

If the process belongs to the FARMA family of models, the PACF should decay to zero at a hyperbolic rate. This rate is independent of the degree of persistence. However, for the case of a FARMA(0,d,0) process, long term memory implies that all the values of the PACF should be positive. This behaviour of the PACF for the FARMA(0,d,0) process suggests that to detect persistence, not only a hyperbolic decay of the PACF is of interest, but also the behaviour of the signs of the PACF. This suggests the use of a nonparametric sign test to test the signs of the estimated PACF. None of the estimated PACF's of the fourteen data sets show strong evidence of a hyperbolic rate of decay to zero. However, some of them like those for the St. Lawrence Riverflows and Philadelphia Rainfall series show PACF structures that are generally positive. A sign test of these PACF's gives further support for the conjecture that these time series demonstrate signs of persistence.

Another characteristic of a time series that could indicate the presence of persistence is the behaviour of the partial sample means, $\bar{z}_k$, of the process that are defined as

$$\bar{z}_k = \frac{(z_{k-1} + z_{k-2} + \cdots + z_o)}{k}. \qquad [11.5.1]$$

For a short memory time series, a plot of $\bar{z}_k$ against k should show great stochastic variability for the first values of k, but after k reaches a moderate value the graph should decay to an almost constant value and should show small stochastic variability. However, for the case of a long memory time series, the plot of $\bar{z}_k$ against k should display great stochastic variability for the first few values of k. For moderate values of k the graph should show a gentle trend that should oscillate around a constant value as k increases, and after k reaches a very large value, which depends on the degree of persistence, $\bar{z}_k$ should reach a constant value. Furthermore, because the present values of the time series are correlated with the past, the current values of $\bar{z}_k$ are highly correlated with the past and, therefore, a plot of $\bar{z}_k$ against k could show local trends. To detect persistence, the rate of decay towards a constant value of the local trends is of interest, as is also the presence of an overall gentle trend. However, the presence of local trends in the plot of $\bar{z}_k$ against k by itself does not indicate the presence of persistence. Within the fourteen data sets, the St. Lawrence Riverflows have an overall decreasing trend. This trend is gentle enough to assume that it could be due to the presence of persistence in the time series and not due to non-stationarity. The graph of $\bar{z}_k$ against k of the St. Lawrence Riverflows is displayed in Figure 11.5.1. For some of the other data sets, local trends in $\bar{z}_k$ seemed to be present even at the end of the series. Finally, for most of the data sets the behaviour of the partial means is consistent with what could be expected in time series having a short term memory, consisting of a rapid decay of the graph to a constant value.

All of the FARMA models considered for fitting to the series in Table 11.5.1 are subsets of the FARMA(2,d,1) model given by

$$(1 - \phi_1 B - \phi_2 B^2)\nabla^d(B)(z_t - \mu) = (1 - \theta_1 B)a_t, \qquad [11.5.2]$$

where ϕ_i is the ith AR parameter, θ_1 is the first MA parameter, and μ is not present in [11.5.2] for positive d. For the St. Lawrence Riverflows the additional constrained AR(3) model given by

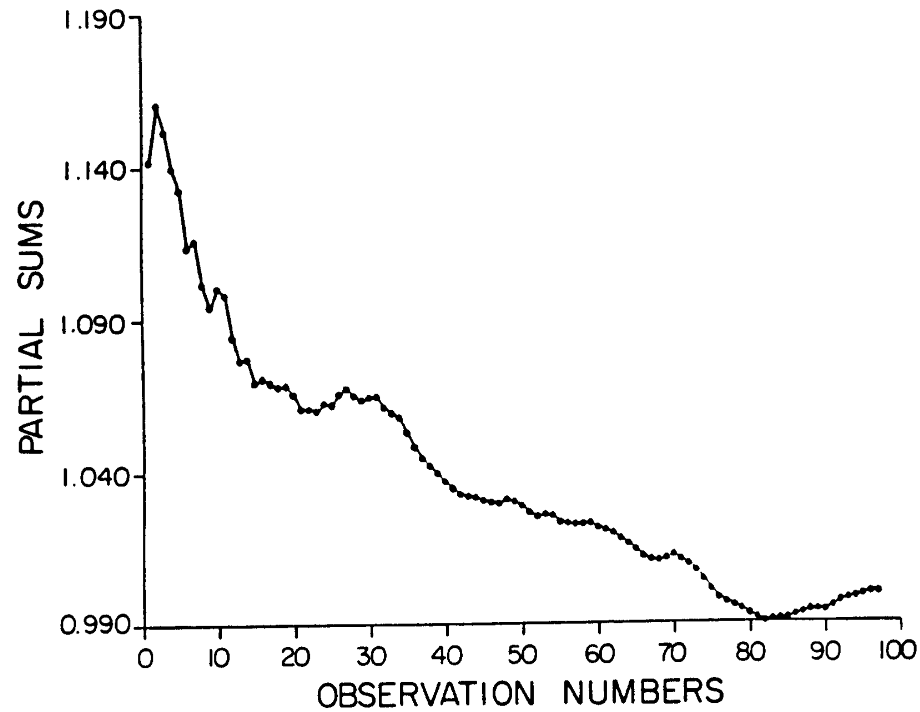

Figure 11.5.1. Partial sums of the St. Lawrence at Ogdensburg, New York from 1860-1957.

$$(1 - \phi_1 B - \phi_3 B^3)(z_t - \mu) = a_t, \qquad [11.5.3]$$

was considered, because this is the model used in Chapter 3 and Part III, within the class of ARMA models. The most appropriate FARMA model from [11.2.4] to fit each series was selected according to the minimum AIC (see Section 6.3), considering only those models that passed tests for whiteness of the fitted model residuals. The maximum likelihood estimates (MLE's) of the model parameters for each series along with the standard errors (SE's) given in brackets are displayed in Table 11.5.2. Those time series for which the estimates of d given in Table 11.5.2 are positive, portray persistent behaviour. Also, because the degree of persistence depends on the magnitude of d, those series having higher values of d possess greater degrees of persistence. For example, the model for the St. Lawrence River was estimated as a FARMA(0,d,0) with $d = 0.4999$. This indicates that the flows of the St. Lawrence are highly persistent, and, hence the far away past strongly influences the present. A consequence of this influence is the slow rate of convergence of the sample mean to the true value. For the case of the St. Lawrence River this rate of decay is of order $O(N^{-0.0001})$, where N is the number of observations. This order of convergence is also true for the forecasting function and the estimated ACF of the St. Lawrence Riverflows. An interesting feature of the St. Lawrence River is that it is associated with great masses of water which perhaps suggests a model having a reservoir term whose time step is larger than the time step used to measure the series. All the models that exhibit persistence in Table 11.5.2 are FARMA(0,d,0). The data sets for which it was appropriate to fit a FARMA(0,d,0) model are the Mckenzie, St. Lawrence and Thames annual riverflows plus the Philadelphia rainfall series. There were other data sets for which the AIC selected ARMA models but the differences between minimum AIC's for the ARMA models and the AIC's for FARMA(0,d,0) models were very small. Finally, note that some rivers do not show any sign of second order correlation structure, as the optimal model according to the AIC was simply the mean. These data sets are the Dal, Danube and Rhine Rivers. Keep in mind that most of these findings are consistent with the models suggested by the sample ACF, sample PACF and behaviour of the partial means.

Table 11.5.2. Parameter estimates and standard errors in brackets for FARMA models fitted to the hydrological time series.

Series	Parameter Estimates and SE's			
	ϕ_1	ϕ_2	d	θ_1
Saugeen	-	-	-	-
Dal	-	-	-	-
Danube	-	-	-	-
French	-0.234 (0.12)	-	-	-
Gota	0.59 (0.08)	-0.27 (0.08)	-	-
Mckenzie	-	-	0.27 (0.10)	-
Mississippi	0.29 (0.10)	-	-	-
Neumunas	-	-	-	-0.19 (0.08)
Rhine	-	-	-	-
St. Lawrence	-	-	0.499 (0.08)	-
Thames	-	-	0.12 (0.10)	-
Fortaleza	0.24 (0.08)	-	-	-
Philadelphia	-	-	0.23 (0.08)	-
Temperature	0.12 (0.06)	0.2 (0.06)	-	-

The bootstrapping technique of Cover and Unny (1986) was used to increase the finite sample information about the estimates of the persistence parameter d. For some data sets, the AIC does not provide a clear cut separation between models with and without the persistence parameter d. Also, most of the time the best FARMA model with a persistent parameter was the FARMA(0,d,0) model. Because of these two remarks, it is interesting to obtain information on how reliable are the estimates of d and the estimates of its SE. For these reasons, the bootstrap technique proposed by Cover and Unny was used with the FARMA(0,d,0) model for all the data sets. Although this model is not appropriate for some data sets, the information about the behaviour of the estimates of d is valuable. The information given by the bootstrap was summarized by two methods. First, the sample mean and standard deviation of the estimates of d using the bootstrap technique for each data set were computed. Second, the distribution of the estimates of d for each data set was estimated using a nonparametric kernel estimate (Fryer, 1977). Using these two pieces of information, it is possible to decide if the data set exhibits any evidence of persistence. The means and standard deviations are given in Table 11.5.3, together with the estimates of d obtained using the approximate maximum likelihood method. Plots of the density of the estimates of d for the St. Lawrence riverflows and for rainfall at Philadelphia are given in Figures 11.5.2 and 11.5.3, respectively. From these tables and graphs, one can draw a number of conclusions:

(a) The large sample approximations are not necessarily valid for finite sample sizes. For example, the distribution of the estimate of the parameter d for the St. Lawrence seems not to have tails, and the density seems to be concentrated on the interval 0.4 to 0.54. Moreover, the density seems somewhat skewed.

(b) It appears that the asymptotic standard deviations are smaller than the bootstrap estimates for values of d not very close to 0.5, and this behaviour is reversed for values of d close to 0.5.

(c) Note also that the means of the estimates of the parameter d obtained by resampling and the standard deviations of these estimates do not necessarily represent the data.

If the threshold a is assumed known, the estimate of a by least squares can be easily found, and it can be shown using standard techniques that it has an asymptotic normal distribution with the inverse of the information matrix as the asymptotic variance. This information matrix is given by the expected value of the truncated variable $z_t^2\{|z_t| > a\}$. If the threshold a is unknown, the usual techniques cannot be used because of the nondifferentiability of the sum of squares function with respect to a.

11.6 CONCLUSIONS

As a direct result of research on long memory modelling motivated by the controversy surrounding the Hurst phenomenon defined in Section 10.3.1, Hosking (1981) originally proposed the generalization of ARIMA models so that long term persistence could be effectively modelled. In particular, Hosking (1981) and independently, Granger and Joyeux (1980), suggested the FARMA(p,d,q) model in [11.2.4] as a flexible approach for describing persistence. The FARMA model is especially appealing to researchers and practitioners in hydrology, economics and elsewhere, because it can model both long and short term behaviour within the confines of a single model. Bloomfield (1992), for example, employs FARMA models for investigating trends in annual global temperature data. The fractional differencing filter in [11.2.3] can account for long term behaviour or persistence while the ARMA component of the overall FARMA model in [11.2.4] takes care of the short memory aspects of the series being modelled. Because of these and other reasons, the FARMA(p,d,q) model of this chapter constitutes a more flexible approach to modelling persistence than the FGN model of Section 10.4.

Model construction techniques are available for fitting FARMA(p,d,q) models to data sets. However, as noted in Section 11.3 improved estimation techniques should be developed and further contributions to model identification could be made. An approximate maximum likelihood estimation algorithm is presented in Appendix A11.1. The applications of Section 11.5 demonstrate how FARMA(0,d,0) models can be fitted in practice to yearly time series. After obtaining MLE's for the model parameters and subjecting the fitted model to diagnostic checks

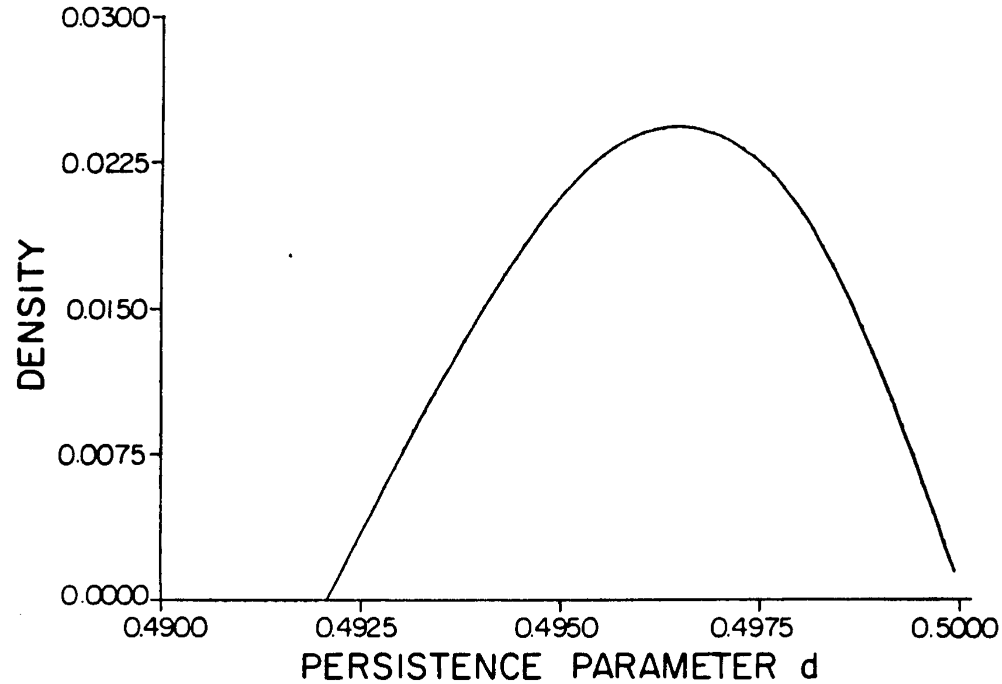

Figure 11.5.2. Probability density of the persistence parameter d obtained by bootstrapping for the St. Lawrence Riverflows.

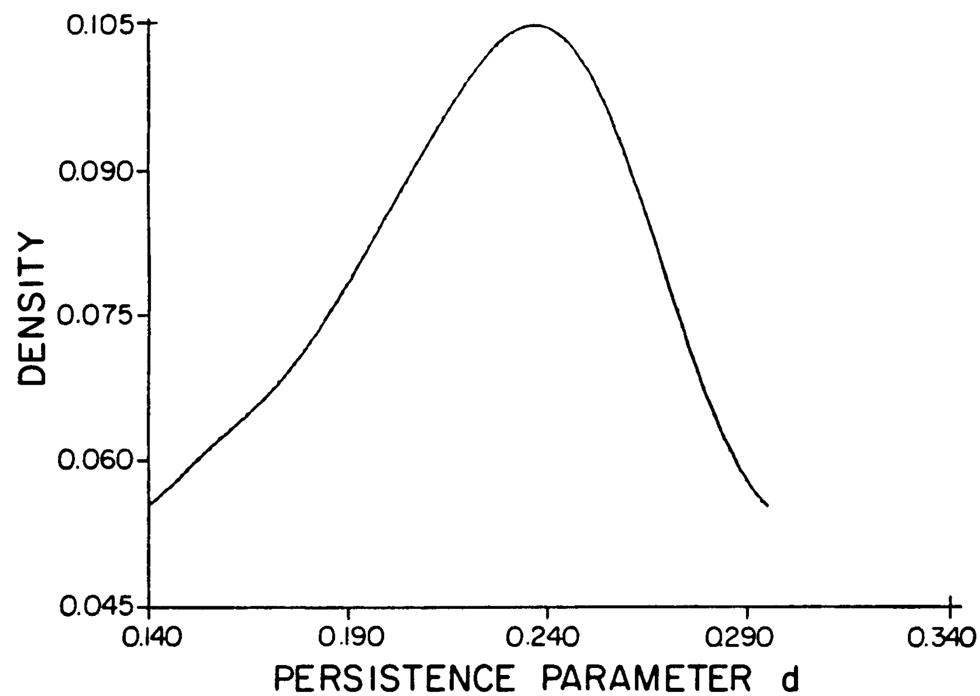

Figure 11.5.3. Probability density of the persistence parameter d obtained by bootstrapping for the Philadelphia rainfall.

Table 11.5.3. Estimation of the parameter d using bootstrapping.

Data Set Identification	Means of $\tilde{d}$	St. Deviations	$\hat{d}$	St. Deviations
(1) Saugeen	0.110	0.212	0.108	0.100
(2) Dal	0.028	0.177	0.024	0.093
(3) Danube	0.069	0.157	0.059	0.072
(4) French	0.148	0.168	0.134	0.093
(5) Gota	0.365	0.245	0.388	0.634
(6) Mckenzie	0.234	0.142	0.274	0.105
(9) Neumunas	0.105	0.137	0.103	0.068
(10) Rain Phil.	0.210	0.110	0.229	0.078
(12) St. Lawrence	0.475	0.055	0.499	0.079
(13) Thames	0.139	0.149	0.120	0.093
(14) Temperature	0.153	0.079	0.151	0.050

Using the bootstrapping technique described in Section 11.3.3, the value of the persistence parameter d in the model $\nabla^d(B)z_t=a_t$ was estimated by the mean value of the estimates obtained using the bootstrap, and is given in the second column. The third column gives the standard deviation of the estimate obtained by bootstrapping. The fourth column lists $\hat{d}$ which is the estimate of d obtained by the appropriate maximum likelihood method described in the Section 11.3.3 and the last column gives the asymptotic standard deviation of $\hat{d}$.

(Section 11.3.4), the calibrated model can be used for simulation and forecasting. Techniques for simulating and forecasting with a FARMA(p,d,q) model are presented in Section 11.4.

APPENDIX A11.1

ESTIMATION ALGORITHM FOR FARMA MODELS

This appendix presents an algorithm for obtaining approximate MLE's for the parameters of a FARMA(p,d,q) model. This estimation algorithm was originally presented by Jimenez et al. (1990) and constitutes an extension of the estimation algorithm of Li and McLeod (1986). In the algorithm, it is assumed that the estimated mean of the series has been subtracted from each observation in order to produce a series having a zero mean. The mean can be estimated using [11.3.2] or some other appropriate technique.

To compute the unconditional sum of squares of the residuals obtained assuming that the model can be represented by a long autoregressive approximation of

$$\phi(B)\nabla_M^d(B)z_t = \theta(B)a_t, \qquad [A11.1.1]$$

a backforecasting algorithm similar to the one used by McLeod and Sales (1983) to compute the unconditional residual sum of squares for seasonal ARMA models, can be used. The unconditional sum of squares of the residuals is given by

$$S = \sum_{t=-\infty}^{N} [a_t]^2, \qquad \text{[A11.1.2]}$$

where $[\cdot]$ denotes expectation with respect to the observations is approximated by

$$S = \sum_{t=1-Q}^{N} [a_t]^2, \qquad \text{[A11.1.3]}$$

where Q is a fairly large truncation point. The conditional form of [A11.1.1] is given by

$$\phi(B)\nabla_M^d(B)[z_t] = \theta(B)[a_t], \qquad \text{[A11.1.4]}$$

where $[a_t] = 0, t > N$. This can be expressed by a two stage model

$$\nabla_M^d(B)[z_t] = [c_t], \qquad \text{[A11.1.5]}$$

and

$$\phi(B)[c_t] = \theta(B)[a_t] \qquad \text{[A11.1.6]}$$

The Box-Jenkins backforecasting approach needs also the forward form of [A11.1.1] such that

$$\phi(F)\nabla_M^d(F)z_t = \theta(F)e_t, \qquad \text{[A11.1.7]}$$

where F is the forward time shift operator such that $Fz_t = z_{t+1}$ and e_t is a sequence of normal independent random variables with mean 0. The method uses the conditional form of [A11.1.7] given by

$$\nabla_M^d(F)[z_t] = [b_t], \qquad \text{[A11.1.8]}$$

and

$$\phi(F)[b_t] = \theta(F)[e_t] \qquad \text{[A11.1.9]}$$

where $[e_t] = 0, t < 1$.

In summary, the unconditional sum of squares can be obtained iteratively through the following steps.

Step 0. Select Q and M.

Step 1. Compute the autoregressive coefficients of ∇_M^d.

Step 2. Compute $[b_t]$, using [A11.1.9] for $t = N + Q, \ldots, 1$. Initially set $[b_t] = 0$.

Step 3. Backforecast the $[b_t]$ series using [A11.1.9]. This can be accomplished using the SARMAS algorithm of McLeod and Sales (1983).

Step 4. Backforecast the $[z_t]$ series using [A11.1.8].

Step 5. Compute the $[c_t]$ for $t = 1 - Q, \ldots, N$ series using [A11.1.5].

Step 6. Compute the $[a_t]$ for $t = 1 - Q, \ldots, N$ series using [A11.1.6].

Step 7. Compute S using [A11.1.2].

Steps 1 to 7 can be repeated until a previously specified tolerance limit is achieved. The parameters are obtained by minimizing S as given in [A11.1.2]. The minimization algorithm given by Powell (1964) can be employed to minimize S.

PROBLEMS

11.1 By referring to appropriate literature cited in Chapters 10 and 11, make a list of the range of related definitions for long memory, or persistence. Compare the similarities and differences among these definitions. Which definition is most clear to you?

11.2 The definition for a FARMA(p,d,q) model is presented in [11.2.4]. Employ [11.2.3] to write the expanded forms of the following FARMA models:

a) FARMA(1,0.4,1),

b) FARMA(0,-0.3,2),

c) FARMA(1,0.8,1).

11.3 Long memory models have been applied to time series in a variety of different fields. Find three different applications of long memory models by referring to publications in fields of your choice. For each application, write down the complete reference and provide a brief summary. Do not use applications from references given in Chapters 10 and 11.

11.4 By referring to the paper by Hosking (1981), write down the formula for the theoretical ACF of a fractional differencing model and comment upon the general properties of the ACF.

11.5 Outline the purposes of bootstrapping and how it is implemented in practice. Describe in some detail how the bootstrapping technique of Cover and Unny (1986) can be employed when estimating parameters in ARMA, ARIMA and FARMA models.

11.6 Select an annual time series which you think may possess long term memory. Explain reasons for suspecting persistence based upon your physical understanding of the problem. Following the approaches suggested in Sections 11.3 and 11.5, use statistical identification methods to justify your suspicions. Fit a fractional differencing or FARMA(0,d,0) as well as the most appropriate ARMA model to the series. Comment upon your findings.

REFERENCES

BOOTSTRAPPING

Cover, K. A. and Unny, T. E. (1986). Application of computer intensive statistics to parameter uncertainty in streamflow synthesis. *Water Resources Bulletin*, 22(3):495-507.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7:1-26.

Fryer, M. J. (1977). A review of some nonparametric methods of density estimation. *Journal of the Institute of Mathematics and Applications*, 20:335-354.

FARMA MODELLING

Bloomfield, P. (1992). Trends in global temperature. *Climatic Change* 21:1-16.

Boes, D. C., Davis, R. A. and Gupta, S. N. (1989). Parameter estimation in low order fractionally differenced ARMA processes. *Stochastic Hydrology and Hydraulics*, 3:97-110.

Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods*. Springer-Verlag, New York.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68:165-176.

Hosking, J. R. M. (1984). Modeling persistence of hydrological time series using fractional differencing. *Water Resources Research*, 20(12):1898-1908.

Hosking, J. R. M. (1985). Fractional differencing modelling in hydrology. *Water Resources Bulletin*, 21(4):677-682.

Hui, Y. V. and Li, W. K. (1988). On fractional differenced periodic processes. Technical report, Chinese University of Hong Kong.

Jimenez, C., Hipel, K. W. and McLeod, A. I. (1990). D´evelopements recents dans la modelisations de la persistance à long terme. *Revue des Sciences de l'Eau*, 3(1):55-81.

Li, W. K. and McLeod, A. I. (1986). Fractional time series modelling. *Biometrika*, 73(1):217-221.

LONG MEMORY

Brillinger, D. R. (1975). *Time Series Data Analysis and Theory*. Holt, Rinehart and Winston, New York.

Cox, D. R. (1984). Long range dependence: A review. In *Statistics: An Appraisal*, David, H. A. and David, H. T., editors, Proceedings of the 50th Anniversary, Iowa State Statistical Laboratory, Iowa State University Press, Ames, Iowa, pp. 55-74.

Davison, A. C. and Cox, D. R. (1989). Some simple properties of sums of random variables having long-range dependence. *Proceedings of the Royal Society of London*, A424:255-262.

Eberlein, E. and Taqqu, M. S., editors (1986). *Dependence in Probability and Statistics: A Survey of Recent Results*. Birkhauser, Boston.

Granger, C. W. J. (1980). Long-memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14:227-238.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1(1):15-29.

Rosenblatt, M. (1961). Independence and dependence. In *Proceedings of the 4th Berkley Symposium on Mathematical Statistics and Probability*. 431-443.

Rosenblatt, M. (1979). Some limit theorems for partial sums of quadratic forms in stationary Gaussian variables. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 49:125-132.

Rosenblatt, M. (1981). Limit theorems for Fourier transforms of functional Gaussian sequences. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 55:123-132.

Taqqu, M. (1975). Weak convergence to fractional Brownian motion and the Rosenblatt process. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 31:287-302.

TIME SERIES ANALYSIS

Bartlett, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society*, Series B, 8:27-41.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, Oakland, California, revised edition.

Cryer, J. D. and Ledolter, J. (1981). Small-sample properties of the maximum likelihood estimator in the first-order moving average model. *Biometrika*, 68:691-694.

Durbin, J. (1960). The fitting of time series models. *Revue de L'Institut International de Statistique*, 28(3):233-244.

Healy, M. J. R. (1968). Algorithm AS6, triangular decomposition of a symmetric matrix. *Journal of the Royal Statistical Society*, Series C (Applied Statistics), 17:195-197.

McLeod, A. I. and Hipel, K. W. (1978a). Comment on modelling monthly hydrologic persistence by G. K. Young and R. U. Jettmar. *Water Resources Research*, 14(4):699-702.

McLeod, A. I. and Hipel, K. W. (1978b). Simulation procedures for Box-Jenkins models. *Water Resources Research*, 14(5):969-975.

McLeod, A. I. and Sales, P. R. H. (1983). An algorithm for approximate likelihood calculation of ARMA and seasonal ARMA models. *Journal of the Royal Statistical Society*, Series C (Applied Statistics), 32:211-223.

Parzen, E. (1982). ARARMA models for time series analysis and forecasting. *Journal of Forecasting*, 1:67-82.

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal*, 7:155-162.

Trench, W. F. (1964). An algorithm for the inversion of finite Toeplitz matrices. *SIAM Journal of Applied Mathematics*, 12:515-521.