PREDICTION OF LONG-MEMORY TIME SERIES: A TUTORIAL REVIEW *†

R. J. Bhansali
University of Liverpool, U.K,.

P. S. Kokoszka
Utah State University, U.S.A.

July 3, 2002

Abstract

Two different approaches, called *Type-I* and *Type-II*, to linear least-squares prediction of a long-memory time series are distinguished. In the former, no new theory is required and a long-memory time series is treated *on par* with a standard short-memory time series and its multistep predictions are obtained by using the existing modelling approaches to prediction of such time series. The latter, by contrast, seeks to model the long-memory stochastic characteristics of the observed time series by a fractional process such that its dth fractional difference, $0 < d < 0.5$, follows a standard short-memory process. The various approaches to constructing long-memory stochastic models are reviewed, and the associated question of parameter estimation for these models is discussed. Having fitted a long-memory stochastic model to a time series, linear multi-step forecasts of its future values are constructed from the model itself. The question of how to evaluate the multistep prediction constants is considered and three different methods proposed for doing so are outlined; it is further noted that, under appropriate regularity conditions, these methods apply also to the class of linear long memory processes with infinite variance. In addition, a brief review of the class of non-linear chaotic maps implying long-memory is given.

1 Introduction

In this section we introduce the framework for the subsequent sections dealing with prediction and extensions of linear long memory processes.

If a time series is understood to be a set of data recorded chronologically, then 'memory' is not a characteristic of such a record; memory instead should be recognised as a property associated with a stochastic model used for describing the generating structure of this record. The class of discrete-time weakly stationary stochastic processes provides a basic but rich family of models towards this end. Thus, suppose that an observed time series of length T, $\{X_t\}$ $(t = 1, ..., T)$, $T > 1$, is a part realization of a weakly stationary process, $\{x_t\}$ $(t = \ldots, -1, 0, 1, \ldots)$ with mean

* *AMS 2000 Subject classification:* 62M20, 62M10, 62M15
† *Keywords and phrases:* Long memory, Prediction.

$\mu_x = E(x_t)$, covariance function $R_x(u) = E[\{x_t - \mu_x\}\{x_{t+u} - \mu_x\}]$, $(t, u = \ldots, -1, 0, 1, \ldots)$, correlation function $r_x(u) = R_x(u)/R_x(0)$ and spectral density function

$$f_x(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} R_x(u) \exp(-iu\lambda). \tag{1.1}$$

As in Bhansali and Kokoszka (2002b), we say $\{x_t\}$ is a short-memory process if the following conditions hold:

a) The covariance function of $\{x_t\}$ is absolutely summable:

$$\sum_{j=-\infty}^{\infty} |R_x(u)| < \infty. \tag{1.2}$$

b) The spectral density function is non-vanishing :

$$f_x(\lambda) \neq 0, \text{ all } \lambda \in [-\pi, \pi]. \tag{1.3}$$

Suppose in addition that $\mu_x = 0$. Under these conditions, the spectral density function, $f_x(\lambda)$, is uniformly continuous and bounded: $0 < m < f_x(\lambda) < M < \infty$, all $\lambda \in [0, \pi]$ and $\{x_t\}$ admits an infinite moving average representation and an infinite autoregressive representation as follows:

$$x_t = \sum_{j=0}^{\infty} b(j)\epsilon_{t-j}, \quad b(0) = 1, \tag{1.4}$$

$$\sum_{j=0}^{\infty} a(j)x_{t-j} = \epsilon_t, \quad a(0) = 1, \tag{1.5}$$

where $\{\epsilon_t\}$ is a sequence of uncorrelated random variables, each with mean 0 and variance σ^2 and the coefficients, $b(j)$ and $a(j)$, are absolutely summable:

$$\sum_{j=0}^{\infty} |b(j)| < \infty, \tag{1.6}$$

$$\sum_{j=0}^{\infty} |a(j)| < \infty. \tag{1.7}$$

For a proof of these results, see Theorem 3.8.4 of Brillinger (1975), who also shows that if $f_x(\lambda)$ is treated as being known exactly, the coefficients, $b(j)$ and $a(j)$, may be determined by the Wiener-Hopf spectral factorization procedure, see Bhansali (1974).

A popular class of short-memory processes is defined by the standard Autoregressive Moving Average model of order (p, q), ARMA(p, q) model, for short:

$$\sum_{j=0}^{p} \phi(j)x_{t-j} = \sum_{j=0}^{q} \theta(j)\epsilon_{t-j}, \tag{1.8}$$

where $\phi(0) = \theta(0) = 1$, $\{\epsilon_t\}$ is a sequence of independent Normal random variables, each with mean 0 and variance σ^2, and the real-valued coefficients, $\phi(j)$ and $\theta(j)$, are such that if

$$\Phi(z) = \sum_{j=0}^{p} \phi(j) z^j, \quad \Theta(z) = \sum_{j=0}^{q} \theta(j) z^j, \tag{1.9}$$

repectively, denote their characteristic polynomials, then $\Phi(z) \neq 0$, $\Theta(z) \neq 0$, $|z| \leq 1$ and $\Phi(z)$ and $\Theta(z)$ do not have a common zero.

As is well-known, for this class of models, the spectral density function, $f_y(\lambda)$, of $\{y_t\}$ is specified to be a rational function:

$$f_y(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\Theta(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2}. \tag{1.10}$$

Also, the correlation function, $r_x(u)$, tends to zero, as $u \to \infty$, at an exponential rate, that is, there exist constants, B and ρ such that, as $u \to \infty$,

$$|r_x(u)| \leq B\rho^{|u|}, \quad 0 < \rho < 1. \tag{1.11}$$

However, the condition (1.11) is not necessary for a process to possess short-memory and in this sense the definition given above is more general and enables the inclusion of the autoregressive model fitting approach, see Berk (1974), Bhansali (1978), within the group of statistical procedures based on the class of short-memory processes.

For a long-memory process, by contrast, the condition (1.2) does not hold, that is, the covariance function, $R_x(u)$, is not summable:

$$\sum_{u=-\infty}^{\infty} R_x(u) = \infty, \tag{1.12}$$

and $f_x(\lambda)$ has a singularity at the origin, and, for a constant, d, such that $0 < d < 0.5$, we may write

$$f_x(\lambda) \sim \lambda^{-2d} \Delta(1/|\lambda|), \quad \text{as} \quad \lambda \to 0, \tag{1.13}$$

where $\Delta(\lambda)$ is a slowly-varying function at infinity.

Following Hosking (1981) and Granger and Joyeux (1980), and as in Bhansali and Kokoszka (2002b), we say that a discrete-time, weakly stationary process, $\{y_t\}$, has long-memory if the process, $\{x_t\}$, say, obtained by taking its dth fractional difference, $0 < d < 0.5$,

$$x_t = (1 - z)^d y_t, \tag{1.14}$$

where z denote the backshift operator, $z^j y_t = y_{t-j}$, has short-memory.

Now, since $\{x_t\}$ admits the infinite moving average and autoregressive representations given by (1.4) and (1.5), with coefficients, $b(j)$, and $a(j)$, satisfying (1.6) and (1.7), respectively, the long-memory process, $\{y_t\}$, also admits one-sided infinite moving average and autoregressive representations as follows:

$$y_t = \sum_{j=0}^{\infty} \psi(j) \epsilon_{t-j}, \; \psi(0) = 1, \tag{1.15}$$

$$\sum_{j=0}^{\infty} \pi(j) y_{t-j} = \epsilon_t, \ \pi(0) = 1, \tag{1.16}$$

where $\{\epsilon_t\}$ is as in (1.4) and (1.5), and if

$$\Psi(z) = \sum_{j=0}^{\infty} \psi(j) z^j, \tag{1.17}$$

$$\Pi(z) = \sum_{j=0}^{\infty} \pi(j) z^j, \tag{1.18}$$

denote the generating functions of the coefficients, $\psi(j)$ and $\pi(j)$, respectively, $\Psi(z)$ nad $\Pi(z)$ are given by:

$$\Psi(z) = (1-z)^{-d} \sum_{j=0}^{\infty} b(j) z^j, \tag{1.19}$$

$$\Pi(z) = (1-z)^{d} \sum_{j=0}^{\infty} a(j) z^j. \tag{1.20}$$

Hence, it follows that the $\psi(j)$ and $\pi(j)$ decrease to zero at a polynomial (geometric) rate: $\psi(j) \sim Bj^{d-1}$, $\pi(j) \sim Aj^{-d-1}$, as $j \to \infty$, and the $\psi(j)$ are square-summable but not absolutely summable:

$$\sum_{j=0}^{\infty} \psi(j)^2 < \infty, \tag{1.21}$$

but the $\pi(j)$ are absolutely-summable:

$$\sum_{j=0}^{\infty} |\pi(j)| < \infty. \tag{1.22}$$

Moreover both the covariance function, $R_y(u)$, and the correlation function, $r_y(u)$ of $\{y_t\}$ also decrease to zero as $u \to \infty$ at a polynomial rate, $r_y(u) \sim Bu^{2d-1}$. Many authors, , see Brockwell and Davis (1991), define a long-memory process in terms of this last property.

An advantage of defining the class of long-memory processes by (1.14) is that the resulting class of processes may conveniently be related to the popular class of non-stationary integrated processes, see Box and Jenkins (1970), which is also defined as in (1.14) but with d allowed to take only integer values. On adopting the notation introduced by these authors, $\{y_t\}$ is said to follow an $ARFIMA(p,d,q)$ process, if it has long-memory but the fractionally-differenced process, $\{x_t\}$, in (1.14) follows an $ARMA(p,q)$ model defined by (1.8). For $d > 0.5$, however, the resulting process is non-stationary. If $-0.5 < d < 0$, then the process is stationary and it is often called an intermediate-memory or antipersistent process, see McLeod (1998).

In Section 4 we discuss the question of how to define a long-memory process for the class of infinite variance stable processes for which the correlation function does not exist, see also Heyde and Yang (1997); moreover, in Section 5, we discuss non-linear chaotic maps with long-memory, in the sense of having correlations which decrease to zero at a polynomial rate. First, however, we discuss the question of linear least-squares prediction for the class of (linear) long-memory processes introduced above.

2 The Type-I and Type-II approaches

It should be noted that from the point of view of linear least-squares prediction of a long-memory process, since (1.21) and (1.22) hold, the classical Wiener-Kolmogorov prediction theory continues to apply; see, for example, Theorem 4.2 of Doob (1953). On the other hand, however, the long-memory process, $\{y_t\}$, is related to the short-memory process, $\{x_t\}$ by (1.14). We may accordingly distinguish between a *Type-I* and a *Type-II* approach to the linear least-squares prediction of long memory time series. In the former, the multistep forecasts of a long-memory time series are constructed without explicitly seeking to model its long-memory characteristics, that is, by approximating a long-memory time series by a suitable short-memory model, and, in essence, by employing the well-established modelling strategies for predicting the latter. The *Type-II* approach by contrast, explicitly seeks to model the long-memory stochastic characteristics by specifying that the dth difference, $0 < d < 0.5$, of an observed time series follows a standard short-memory model.

Below, we briefly describe how these two approaches may be implemented; for further details see Bhansali and Kokoszka (2002b).

2.1 Implementation of The Type-I Approach

For implementing the *Type-I* approach with an observed time series, there is a choice of modelling strategies, as further described below:

1). *Parameteric ARMA models.* A standard autoregressive-moving average model of order (p, q), defined by equation (1.8), is fitted to the observed time series, with the values of p and q determined by a model selection criterion such as AIC or BIC. Crato and Ray (1996) have applied this approach for multistep prediction of long-memory time series. This approach can be computationally demanding, however, since the various ARMA models of increasing orders would need to be fitted by a non-linear maximum likelihood procedure; as discussed by Bhansali (1993) in his review of order selection procedures for linear time series models, the use of a three-stage Hannan-Rissanen procedure, see Hannan and Rissannen (1982), provides a way around this difficulty. A related but different possibility is investigated by Tiao and Tsay (1994), who use an $ARMA(1, 1)$ model for multistep prediction of a long-memory time series, but estimate its parameters by an adaptive procedure, that is, see Bhansali (2002), by a multistep method such that the estimates are constructed anew for each step of prediction by a least-squares procedure involving minimisation of the sum of squares of h-step prediction errors; the classical maximum likelihood approach to estimating the parameters of an $ARMA$ model is, in this terminology, referred to as the one-step method. Brodsky and Hurvich (1999) and Basak *et al.* (2001) are additional references on this approach; the former compares its effectiveness relative to the *Type-II* approach, while the latter examines the nature of approximation provided by an $ARMA(1, 1)$ model for a long-mmeory process.

2). *The autoregressive model fitting approach.* For short-memory time series, this approach has been used successfully for spectral estimation, linear prediction and related applications; see Parzen (1969), Berk (1974), Bhansali (1978), Shibata (1980), among others. The observed time series is assumed to be a realization of a weakly stationary process admitting a one-sided infinite autoregressive representation, (1.5), whose coefficients, $a(j)$, are absolutely summable and satisfy (1.7); a model of order k is fitted by least-squares or by solving the Yule-Walker equations and the value of k is determined by AIC, or a related criterion. Ray (1993) has investigated the use of the autoregressive model fitting approach for multistep prediction of a long-memory time series. In her implementation, a 'plug-in' method is used for generating the multistep forecasts and this involves repeatedly iterating the model fitted for one-step prediction and then replacing the unknown future values by their own forecasts. An alternative Direct method may also be considered according to which a new autoregressive model is selected and fitted separately for each step of prediction; Bhansali (1996a) has established an asymptotic optimality property for this method. It should be noted that this Direct method is related to yet slightly different from the adaptive method of Tiao and Tsay (1994) discussed above, in whose application a new model is not selected separately for each step of prediction, but rather the parameters of the model used for one-step prediction, namely an $ARMA(1,1)$ model, are re-estimated for each value of h; see Bhansali (1999) and Bhansali (2002) for a further discussion of the differences between these two methods and for an explanation of the motivation for considering the Direct method. Also, Bhansali and Kokoszka (2002b) compare the relative predictive performance of the *Type-I* plug-in and *Type-I* direct methods of fitting autoregressions on simulated long-memory time series.

3). *The spectral factorization procedure.* Bhansali (1974) has studied statistical properties of the Wiener-Hopf spectral factorization procedure when it is applied to a smooth periodogram estimator of the spectral density function of short-memory time series. Recently, Hidalgo and Yajima (1999) have considered its use for multistep prediction of long-memory time series.

2.2 Implementation of the Type-II Approach

As with the *Type-I* approach discussed above, there is a choice of different strategies concerning the nature of the stochastic model specified for $\{x_t\}$, the short-memory process defined by (1.14). Moreover, this choice is now wider because the use of the Exponential model, see Bloomfield (1973), and related models specified in terms of the logarithm of the spectral density function of $\{x_t\}$ may be considered, and, in addition, if the principal aim is to simply estimate the long-memory parameter, d, rather than to predict the future values of $\{y_t\}$, then a stochastic model for $\{x_t\}$ need not be specified. Below, we describe the various modelling approaches together with qualitative features of the various models suggested for a long- memory process; in Section 2.2.2, we discuss the associated question of how to estimate the parameters of the specified models from an observed long-memory time series.

2.2.1 Stochastic Models for Long-Memory Processes

1). *Parametric Models.* Here, the stochastic (covariance) structure of $\{x_t\}$ is fully specified, except for a finite number of parameters, but the values of these parameters are unknown and so is the actual number of parameters (the order of the model). Based on the success of the standard $ARMA(p,q)$ models for short-memory time series, the class of $ARFIMA(p,d,q)$

models is a popular choice of parametric models for long-memory processes. An alternative class of parametric models is obtained, see Beran (1993), by specifying that the spectral density function, $f_x(\lambda)$, of $\{x_t\}$ follows an exponential model of Bloomfield (1973), that is,

$$f_y(\lambda) = \frac{\sigma^2}{2\pi}|\Pi(\lambda)|^{-2}, \tag{2.1}$$

where

$$\Pi(\lambda) = \left|1 - e^{-i\lambda}\right|^{-2d} \left[\exp\{\sum_{j=1}^{n} c(j)\cos(j\lambda)\}\right], \tag{2.2}$$

and $n \geq 1$ is the order of the model. and, thus, in this model, the logarithm of the spectral density function of $\{x_t\}$ is postulated to have a finite Fourier series expansion. An alternative specification involves postulating that $\log f_x(\lambda)$ is a polynomial function, which is one of the $non - ARFIMA$ models simulated by Crato and Ray (1996); Beran (1993) also envisages the use of piecewise continuous functions in place of either Fourier cosine functions, or a polynomial function.

2). *Non-parametric Approach.* In this approach, only the behaviour of $f_y(\lambda)$ near the origin, as $\lambda \to 0$, is specified and little is explicitly assumed about the behaviour of $f_y(\lambda)$ away from the origin, that is, about the short-memory component of $\{y_t\}$.

Equation (1.13) gives the class of models being considered. A basic model of this class is obtained by specifying that $\Delta(1/|\lambda|)$ is a constant in a neighbourhood of the origin.

As discussed by Hall (1997), the specification (1.13) is equivalent to requiring that the covariance function, $R_y(s)$, of $\{y_t\}$ takes the following form:

$$R_y(s) \sim \Delta(s)s^{-(1-2d)}. \tag{2.3}$$

3). *Hierarchical Parametric Models.* According to this approach, which seeks to build on the success of the autoregressive model approach discussed in Section 2.1 for short-memory time series, $\{y_t\}$ is assumed to satisfy (1.14), but its dth fractional difference, $\{x_t\}$, is specified to follow a parameteric model with an infinite number of parameters. Hence, this approach is less general than the non-parametric approach but less restrictive than the parametric approach. Moreover, since this approach is based upon the existence of a one-sided infinite autoregressive and infinite moving average representations for $\{x_t\}$ discussed in Section 1, it is more plausible than a fully parametric approach, and at the same time, as discussed further in Section 3, it may still be applied for constructing linear least-squares predictors of the unknown future values of $\{y_t\}$. Two different families of hierarachical models may be considered:

a). Fractional autoregressive, FAR, models. In this approach, the short-memory process, $\{x_t\}$, of $\{y_t\}$, is assumed to admit representations (1.5) and (1.4), but with cefficients, $a(j)$ and $b(j)$ decreasing to zero at an exponential rate. Given a stretch of T observations from $\{y_t\}$, a fractional autoregressive model of order k, an $FAR(k, d)$ model, is fitted by a likelihood procedure discussed further in Section 2.2.2, where k is assumed to be a function of T and allowed to approach ∞ simultaneously but sufficiently slowly with it, see Bhansali and Kokoszka (2001).

b). Fractional exponential, FEXP, models. This approach is basically equivalent to the FAR approach described above in a), but the short-memory process, $\{x_t\}$, of $\{y_t\}$ is assumed to follow an infinite-order Exponential model of Bloomfield (1973). Thus, now the spectral density function, $f_y(\lambda)$, of $\{y_t\}$, is given by equation (2.1) but with $n = \infty$ and the coefficients $c(j)$ assumed to approach zero sufficiently fast, either at an exponential rate or to satisfy a slightly weaker summability condition ensuring that, for all $\lambda \in [0, \pi]$, $\log f_x(\lambda)$ admits bounded derivatives of order β, $\beta \geq 1$. There has been much interest recently in this family of models, see, Beran (1997), Moulines and Soulier (1999, 2000) Hurvich and Brodsky (2001), among others, and it is motivated by the intuition that, under (1.14), $\log f_y(\lambda)$ admits the decomposition (2.1), and hence the behaviour of $\{y_t\}$ could be conveniently characterized by specifying that $\log f_x(\lambda)$, which controls the short-memory component, admits an infinite Fourier expansion. For an observed long-memory time series, this approach is implemented in a similar fashion to that described above for the FAR approach; the Fourier expansion for $\log f_x(\lambda)$ is truncated at a finite value, k, say, and the parameters of the resulting $FEXP$ model of order k, whose spectral density function is now given by equation (2.1) with $n = k$, are estimated, though a regression procedure discussed further in Section 2.2.2 is used for parameter estimation.

2.2.2 Parameter Estimation for Long-Memory Models

Denote the observed time series by $y_1, \ldots, y_T$, $T > 1$, which we suppose is a part-realizaton of a long-memory process, $\{y_t\}$, satisfying (1.14). Below, the question of how to estimate the long-memory parameter, d, together with the parameters of a stochastic model specified for the short-memory process, $\{x_t\}$, is addressed. The discussion of various methods is however necessarily brief since several different surveys reviewing the current state-of-the-art on this question are already available; see, Bardet *et al.* (1999) and Bhansali and Kokoszka (2001), among others.

It is possible to group the underlying methods under two broad headings: A) Likelihood-based methods; B) Regression-based methods.

The periodogram function of the observed time series is given by

$$I_y(\lambda) = (2\pi T)^{-1} \left| \sum_{t=1}^{T} y_t \exp(-it\lambda) \right|^2, \tag{2.4}$$

and both the likelihood-based and regression-based methods adopt the periodogram as a basic statistic.

The likelihood-based methods are used when either the parametric approach or the hierarchical model fitting approach involving the fitting of $FAR(k, d)$ models has been adopted. By contrast, the regression-based methods are used when either the non-parametric approach or the hierarchical model fitting approach involving the fitting of $FEXP$ models has been adopted. In addition, a semi-parametric approach is implemented by essentially adopting the non-parametric approach but using the Whittle likelihood for estimating the long-memory parameter, as further discussed below.

A convenient way of implementing the likelihood-based methods is to use the approximate Whittle likelihood function and to separately parametrize the variance, $\sigma^2 = \text{Var}(\epsilon_t)$, of the innovation process, $\{\epsilon_t\}$, see (1.15), from the remaining parameters of the model. Under this assumption, the Whittle likelihood function of the data is given by:

$$L(y|\tau) = \int_{-\pi}^{\pi} \frac{I_y(\lambda)}{g(\lambda,\tau)} d\lambda, \tag{2.5}$$

where

$$g(\lambda,\tau) = \left| \sum_{j=0}^{\infty} \psi(j) \exp(-ij\lambda) \right|^2, \tag{2.6}$$

and τ denotes a finite-dimensional vector of 'spectral parameters' such that the infinite sequence, $\{\psi(j)\}$ $(j = 1, 2, \ldots)$, of the coefficients in the moving average representation of $\{y_t\}$ may be expressed as a function of these parameters. If the order, (p, q), of the $ARFIMA(p, d, q)$ model is treated as known, the $g(\lambda, \tau)$ occurring in (2.5) is given by

$$g(\lambda,\tau) = \left|1 - e^{-i\lambda}\right|^{-2d} \frac{|\Theta(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2}, \tag{2.7}$$

where the parameter vector, τ, is given by $\tau = [d, \phi(1), ..., \phi(p), \theta(1), ..., \theta(q)]'$.

Similarly, for the hierarchical fitting of $FAR(k, d)$ models, $\tau = [d, a(1), ..., a(k)]'$, and

$$g(\lambda,\tau) = \left|1 - e^{-i\lambda}\right|^{-2d} |A_k(e^{-i\lambda})|^{-2}, \tag{2.8}$$

where

$$A_k(z) = 1 + a(1)z + ... + a(k)z^k. \tag{2.9}$$

Other approximations for the likelihood function of the data, including the use of the exact likelihood function, have also been suggested, and it is not necessary even to separately parametrize the innovation variance, σ^2; see Beran (1995), Dahlhaus (1989). However, we do not pursue these question here beyond noting that the software for implementing the likelihood-based methods is available in standard statistical packages; for example, the *Splus* function `arima.fracdiff` can be used.

Next, consider the regression-based procedures. For the non-parametric approach, $f_y(\lambda)$, is specified by (1.13), with $\Delta(\lambda) = C$, a constant for all λ, and since this specification implies that $\log f_y(\lambda)$ is a linear function of $\log \lambda$ in a neighbourhood of the origin, $\lambda = 0$, an estimate of d is obtained by regressing $\log I_y(\lambda)$ on $\log \lambda$, or λ, in a neighbourhood of $\lambda = 0$. A correction for the 'bias' of $\log I_y(\lambda)$ in estimating $\log f_y(\lambda)$, see Davis and Jones (1968), is however necessary and it is made. In addition, because the behaviour of the periodogram function for frequencies near 0 is non-standard for a long-memory time series, see Hurvich and Beltrao (1993), periodogram ordinates which are either too close to or too far away from the 0 frequency are usually omitted by choosing two trimming numbers which determine the band of frequencies over which the actual regression is performed. This procedure is often called the *GPH* method, since it was originally suggested by Geweke and Porter-Hudak (1983). A number of variations on this basic procedures have also been considered, especially the use of a 'smooth' periodogram in place of the 'raw' periodogram, and the inclusion of all frequencies near zero; we however do not discuss these variations here.

For the hierarchical model fitting approach involving the fitting of $FEXP$ models, an estimate of d is constructed by regressing $\log I_y(\lambda)$ on $\log|2\sin(\lambda/2)|$ and $\cos(j\lambda)$, $(j=0,1,...,k)$. Moreover, an equivalent estimate of d may also be obtained, see Bardet *et al.* (1999), from the estimated regression coefficient in a simple linear regression of $\log I_y(\lambda)$ on ξ, where ξ is the difference between $\log|2sin(\lambda/2)|$ and its orthogonal projection on the linear space spanned by $\cos(j\lambda)$, $(j=0,1,...,k)$.

As in the non-parametric approach, the specification (1.13) for $f_y(\lambda)$, with $\Delta(\lambda)=C$, a constant for all λ, is also adopted for the semi-parametric approach. However, see Künsch (1987), Robinson (1995b), estimates of d and C are constructed from the Whittle likelihood, which is now computed only over a neighbourhood of the 0 frequency.

For a discussion of the asymptotic statistical properties of the estimates provided by these four approaches, see Bhansali and Kokoszka (2001). Below, we indicate some of the shortcomings of each approach:

If the order, (p,q), of an $ARFIMA(p,d,q)$ model is treated as known, the maximum likelihood estimator, $\hat{\tau}$, of τ is known to be asymptotically Normal and Fisher efficient. In practice, however, the order (p,q) is invariably unknown, and there is much evidence to suggest that when the actual order is misspecified, the estimates could be badly biased, see Taqqu and Teverovsky (1998), among others. The question of model selection for the class of $ARFIMA(p,d,0)$ models has been considered by Beran *et al.* (1998). An additional conceptual difficulty is, see Hannan (1987), that an observed time series need not follow a 'true' $ARMA$ model.

A main difficulty with both non-parametric and semi-parametric approaches is that while each makes few prior assumptions concerning the stochastic structure of the observed time series, the resulting estimates have a slow rate of convergence, namely $T^{-\delta}$, where, no matter how smooth the short-memory spectral desity, $f_x(\lambda)$, is, $\delta<0.4$. Moreover, an important practical difficulty is how to choose in an optimal fashion the band of frequencies over which the log-periodogrm regression is performed.

The $FEXP$ method yields an estimator with a faster rate of convergence than the non-parametric or semi-parametric methods, but without requiring that the short-memory process follows a parametric model. On the other hand, however, in the situation where the parametric approach is applicable, this and the FAR approach are not as efficient and have a larger variance.

Bhansali and Kokoszka (2001) have studied the empirical behaviour of these approaches by a simulation study.

3 Multistep Prediction of the Long-Memory Time Series by the Type-II approaches

The parametric and the hierarchical model fitting approaches specify the behaviour of the spectral density function, $f_y(\lambda)$, of the process, for all $\lambda\in[0,\pi]$, and are therefore suitable for constructing the multistep forecasts of a long-memory process by linear least-squares methods.

By their very nature, however, the non-parametric and semi-parametric approaches do not seek to model $f_y(\lambda)$ for all $\lambda\in[0,\pi]$ and hence in their original form they are not suitable for this purpose. Crato and Ray (1996) use the non-parametric approach for multistep prediction of a long-memory process by adopting a two-stage approach, in which at the first stage the long-memory component is estimated by a non-parametric method and then 'filtered' out and subsequently at the second stage a parametric $ARMA$ model is fitted to the resulting output series. This two-stage procedure is still parametric, however, and, even though its optimality is unclear, in the discussion below, it may be used in the same way as the parametric approach.

Bhansali and Kokoszka (2002a) consider the question of how to compute the prediction constants for multistep prediction of $\{y_t\}$ by linear least-squares methods. Three different methods are identified, namely the *Truncation Method*, *Type-II Plug-In Method* and *Type-II Direct Method*. Moreover, as the focus now is on the computation of the prediction constants from the estimated parameters, in the following exposition, we do not explicitly distinguish between the parametric or the hierarchical model fitting approaches.

In all three methods of computing the forecast coefficients, an estimate, $\hat{\Pi}(\lambda)$, say, of the transfer function, $\Pi(e^{-i\lambda})$, of the coefficients, $\pi(j)$, in the infinite autoregressive representation, (1.16), for $\{y_t\}$ is obtained first. Thus, for the likelihood-based estimation methods involving the fitting of either an $ARFIMA(p,d,q)$ model or an $FAR(k,d)$ model,

$$\hat{\Pi}(\lambda) = \hat{g}(\lambda,\tau)^{-1}, \tag{3.1}$$

where $\hat{g}(\lambda,\tau)$ is obtained from either (2.7) or (2.8) by replacing the unknown parameter vector, τ, by its estimate, $\hat{\tau}$, constructed by either of these two methods. Similarly, for the $FEXP$ method, $\hat{\Pi}(\lambda)$ is obtained from (2.2) by replacing the unknown parameters by their estimates constructed as described in Section 2.2.2.

For the Truncation method, which is similar to that suggested by Box and Jenkins (1970), an estimate, $\hat{\pi}(j)$, say, of $\pi(j)$ is constructed by a discrete Fourier transform of the sequence $\hat{\Pi}(2\pi t Q^{-1}),\ t = 0,1,\ldots,Q-1$, as follows:

$$\hat{\pi}_y(j) = (2\pi Q)^{-1} \sum_{t=0}^{Q-1} \hat{\Pi}(2\pi t Q^{-1}) \exp(it2\pi t Q^{-1}) \quad (j = 0,1,\ldots Q-1), \tag{3.2}$$

where Q is a large power of 2, e.g. $Q = 1024$. The Splus function `fft` can be used for this purpose. However, since only a finite past of $\{y_t\}$ has been observed, the rather long sequence of $\hat{\pi}(j)$ is truncated at a suitable finite value, K, say.

Next, consider the *Type-II Plug-In* method. This method is based on the innovations algorithm, see Brockwell and Davis (1991) and Crato and Ray (1996), and it is motivated by the observation that for each finite truncation point, K, the truncation method is not optimal in a maximum entropy sense, see Priestley (1981), since it does not use all the information in the first K correlations of the fitted model about the unknown future values, and this method may be improved upon by using the appropriate finite order linear least-squares forecast coefficients based on the correlation function of the fitted model. For applying this method, and the *Type-II Direct* method, the estimated covariance and correlation functions, $\hat{R}_y(u)$, and $\hat{r}_y(u)$, say, implied by the fitted long-memory model is required. Let $\hat{f}_y(\lambda)$ denote the estimated spectral density corresponding to the fitted long-memory model, which may be computed from (2.1), but by replacing $\Pi(\lambda)$ by its estimate constructed as described above. Then, as in Bhansali and Kokoszka (2002a), the $\hat{R}_y(u)$ may be computed by a discrete Fourier transform of the sequence $\hat{f}(2\pi t Q^{-1})$, $t = 0, 1, \ldots, Q-1$, that is, by using the formula (3.2), but with $\hat{\Pi}(2\pi t Q^{-1})$ replaced by $\hat{f}(2\pi t Q^{-1})$. Also, $\hat{r}_y(u)$, may be computed as $\hat{r}_y(u) = \hat{R}_y(u)/\hat{R}_y(0)$.

Now let $S \geq 1$, denote a large integer, which is large enough for approximating the fitted model by an approprite finite order autoregression using the AIC criterion given below. As is well-known, the coefficients, $\hat{\pi}_{Ps}(j)$ $(j = 1, \ldots, s;\ s = 1, \ldots, S)$ and the residual error variance, $\hat{\sigma}^2(s)$, of an sth order autoregressive approximation to the fitted model are the solutions of the following equations, with $\hat{\pi}_{Ps}(0) = 1$:

$$\sum_{j=0}^{s} \hat{\pi}_{Ps}(j)\hat{r}_y(u-j) = 0, \quad (u = 1, \ldots, s), \tag{3.3}$$

$$\sum_{j=0}^{s} \hat{\pi}_{Ps}(j)\hat{R}_y(j) = \hat{\sigma}^2(s). \tag{3.4}$$

The value, k, say, of the order of an appropriate finite order autoregressive approximation to the fitted model is chosen by minimising the Akaike Information Criterion:

$$AIC(s) = T \ln \hat{\sigma}^2(s) + 2s \quad (s = 0, 1, \ldots, S), \tag{3.5}$$

and thus k is such that $AIC(s)$ takes its smallest value for $s = k$. For the chosen value of k, the *Type-II Plug-In* forecasts of the unknown future values are constructed by repeatedly iterating the fitted autoregressive approximation and replacing the unknown future values by their own forecasts; the details are omitted in the interest of space.

By contrast, the *Type-II Direct* forecasts are constructed, as discussed in Section 1, by developing a new autoregressive approximation, that is, the h-step prediction constants and the order of the model to be fitted, separately for each prediction lead time, h. A motivation for considering this approach in the present context is that if $h > 1$ and s is finite, the plug-in prediction constants obtained by solving equations (3.3) do not correspond to the sth order linear least-squares coefficients based on the $\hat{r}_y(u)$, and hence the corresponding h-step forecast may be improved upon by using the latter set of coefficients. As is readily verified, for an $h \geq 1$ and each finite s, the relevant h -step linear least-squares coefficients, $\hat{\pi}_{Dhs}(j)$ $(j = 1, \ldots, s;\ s = 1, \ldots, S)$ are the solutions of the following equations:

$$\sum_{j=1}^{s} \hat{\pi}_{Dhs}(j)\hat{r}_y(u-j) = -\hat{r}_y(h+u-1), \quad (u = 1, \ldots, s). \tag{3.6}$$

Also, the corresponding estimate of the h-step prediction error variance, $\hat{V}_{Dh}(s)$, is given by:

$$\sum_{j=1}^{s} \hat{\pi}_{Dhs}(j)\hat{R}_y(h+j-1) + \hat{R}_y(0) = \hat{V}_{Dh}(s). \tag{3.7}$$

For $h = 1$, equations (3.6) and (3.7) clearly reduce to equations (3.3) and (3.4); for $h > 1$, however, the two sets of prediction constants differ from each other.

The order of the autoregression to be fitted for each h may be determined by minimising the following h-step extension of the Akaike information criterion, see Bhansali (1996a):

$$AIC(h,s) = T \ln \hat{V}_{Dh}(s) + 2s \quad (s = 0, 1, \ldots, S). \tag{3.8}$$

Thus, an autoregression of order $k = \tilde{k}_h$ will be fitted, if the criterion (3.8) attains its smallest value for $s = k$.

The corresponding h-step predictor is then computed as follows:

$$\hat{y}_n(h) = -\sum_{j=1}^{k} \hat{\pi}_{Dhs}(j) y_{n+1-j}, \tag{3.9}$$

where $k = \tilde{k}_h$ is the autoregressive order selected by minimising the criterion (3.8).

Bhansali and Kokoszka (2002a) compare the relative predictive efficacy of the Truncation, *Type-II Plug-In* and *Type-II Direct* methods of evaluating the prediction constants for multistep prediction of several actual long-memory models and report that no one method necessrily outperforms the other two for all models and at all prediction lead times. The *Type II Direct* method may nevertheless be recommended because it frequently outperforms the other two methods and, even in the situations where it does not do so, the resulting loss in predictive efficiency is negligible.

4 Linear long memory processes with infinite variance

The observed long-memory process, $\{y_t\}$, has so far been assumed to possess finite variance. However, many observed time series, notably those occurring in finance and hydrology, display both persistence and heavy tails, see Mittnik and Rachev (2000), Samorodnitsky and Taqqu (1994) and Adler *et al.* (1998) for a detailed exposition of various empirical properties of such time series and a discussion of the many important theoretical challanges they pose.

A main problem is that since the covariances do not exist for such time series, the class of weakly stationary processes does not provide a suitable theoretical model and the class of strictly stationary processes should be considered instead. A difficulty now is how to define 'long memory' for the latter class of processes, since the definitions discussed in Section 1 do not apply. However, currently, there is no such universally accepted definition. One approach has been proposed by Heyde and Yang (1997), according to whose definition, a strictly stationary process, $\{y_t\}$, is a long memory process if its normalized sums, $(\sum_{j=1}^{m} y_j^2)^{-1}(\sum_{j=1}^{m} y_j)^2$, diverge in probability to infinity as $m \to \infty$. A different approach has been proposed by Rachev and Samorodnitsky (2001) and Mansfield *et al.* (2001) whose definition recognises the fact that there is a tendency for the realizations of long memory time series to stay above or below the mean for extended periods of time and hence focuses on "long strange segments". While definitions of this type are useful and allow broad classes of processes to be considered, it is not clear how they could be exploited for constructing mutltistep forecasts. The notion of long strange segments might be relevant for this purpose, but no theoretical results are currently availble in this direction.

A more fruitful approach, which does enable the development of effective prediction methods, is to focus instead on a subclass, $M_{d,\delta}$, say, of stictly stationary stable processes, $\{y_t\}$, satisfying (1.14) and with $\{x_t\}$ admitting a representation of the form (1.4), but in which $\{\epsilon_t\}$ is a sequence of independent identically distributed random variables with the common distribution belonging to the domain of attraction of a δ-stable law, $1 < \delta < 2$, and the $b(j)$ are absolutely summable coefficients satisfying (1.6). As discussed by Bhansali (1996b), this subclass of processes does admit parameteric models like $ARFIMA, FAR$ and $FEXP$, discussed earlier in Sections 1 and 2. A limitation of this approach, however, is that, see Chapters 6 and 7 of Samorodnitsky and Taqqu (1994) within the class of all stictly stationary stable processes, the sub-class $M_{d,\delta}$ of processes forms only a subset which is disjoint from the sub-class of harmonizable processes, and moreover there are processes which do not belong to either of these two subclasses. Hence, it is not clear whether the $M_{d,\delta}$ class of processes is large enough to be suitable for all observed long-memory time series with infinite variance. This limitation notwithstanding, which does not apply to the class of Gaussian stationary processes because such processes admit both moving average and harmonizable representations, below we restrict attention to the $M_{d,\delta}$ class of processes. Kokoszka and Taqqu (1995) show that a sufficient condition for a member of the $M_{d,\delta}$ class of processes to be strictly stationary is that $0 < d < (1 - \delta^{-1})$. The use of $Type - I$ approach for this class may be justified by the results of Cline and Brockwell (1985) who show that if instead of the criterion of minimum mean squared error of prediction, the optimality of the linear predictor is judged by the criterion of minimum dispersion, then the classical Wiener-Kolmogorov prediction theory still continues to apply. As regards the $Type - II$ approach, there has been much development on the question of parameter estimation and prediction for the family of $ARFIMA(p, d, q)$ models belonging to the $M_{d,\delta}$ class. Thus, Kokoszka and Taqqu (1995) investigate certain stochastic properties of this family of models, while, Kokoszka and Taqqu (1996) consider the approximate maximum likelihood estimation of the parameter of this family of models, but treat the order (p, q) as known. In addition, Kokoszka (1996) has generalised the earlier results of Cline and Brockwell (1985) to this family of models and shown that the optimal linear predictor based on the infinite past coincides with the corresponding predictor in the Gaussian case. This result, and especially his Theorem 3.2 , provides a justification for using the *Truncation method* in the infinite variance case. An experimental justification for using the *Type II Direct and Plug-In* methods is provided by Bhansali and Kokoszka (2002a), who investigate the behaviour of these two methods and the truncation method by a simulation study. Also, Bhansali and Kokoszka (2001) establish a consistency property for the parameters estimated by the FAR approach when $\{y_t\}$ is a member of the $M_{d,\delta}$ class of processes.

5 Long memory chaotic time series

By a chaotic time series we mean a deterministic sequence, $\{w_t\}$ $(t = 0, 1, \ldots)$, generated by iteratively applying a one-dimensional map of the following form:

(5.1)
$$w_{t+1} = \zeta(w_t) \ (t = 0, 1, \ldots),$$

where w_0 is real-valued and specifies the *initial condition* for the iterative scheme specified by (5.1), $\zeta : J \to J$ is a non-linear map (function) and J denotes a closed interval of the real line, $(-\infty, \infty)$. Such maps have attracted much attention in recent years and it is recognised that they can generate complex mathematical behaviour such as *strange attractors*, *period doubling*, *limit points* and *chaos*, and can be used to describe phenomena which appear random. We however only emphasise statistical time series aspects of such maps. Thus, we only consider measure-preserving and ergodic maps admitting a time-invariant density, $\chi(w)$, over the interval J. A main result in Ergodic theory states that if $\eta : J \to J$ is an integrable function, then, see Chatterjee and Yilmaz (1992), as $N \to \infty$,

$$\frac{1}{N}\sum_{t=0}^{N-1} \eta(\zeta^t(w)) \to \int_J \eta(w)\chi(w)dw, \tag{5.2}$$

where the convergence is almost surely with respect to the density $\chi(w)$. Thus, Ergodic theory enables a study of the long-term behaviour of a chaotic map. To relate the map (5.1) to a weakly stationary process, suppose that the initial condition, w_0, is generated from a distribution with probability density function, $\chi(w)$, the invariant density, and assume that the first two moments of this distribution exist. Then, $\{w_t, t = 0, 1, \ldots\}$, defines a weakly stationary process, with mean

$$\mu_w = \int_J w\chi(w)dw, \tag{5.3}$$

and covariance function

$$R_w(u) = \int_J \zeta^{u+2}(w)\chi(w)dw - (\mu_w)^2. \tag{5.4}$$

Closed-form analytic expressions for the invariant density and correlation function are now known for a range of different maps; see Lawrance (2001) for an expository discussion. Three well-known examples being the following:

1). *The Tent map.* Here $J = [0, 1]$ and

$$\zeta(w) = 1 - |2w - 1|. \tag{5.5}$$

The invariant density for this map coincides with that for the $Uniform[0,1]$distribution:

$$\chi(w) = 1, \quad \text{all } w \in J, \tag{5.6}$$

and its correlation function is given by:

$$r_w(u) = 0, u = 1, 2, \ldots. \tag{5.7}$$

2). *The Logistic map.* Here $J = [0, 1]$ and

$$\zeta(w) = 1 - |2w - 1|^2 = 4w(1 - w). \tag{5.8}$$

The invariant density for this map coincides with that for a $Beta(\frac{1}{2}, \frac{1}{2})$ distribution:

$$\chi(w) = \frac{1}{\sqrt{\pi^2 w(1 - w)}}, \quad \text{all } w \in J, \tag{5.9}$$

and its correlation function is given by (5.7).

3). *The Bernoulli map.* Here $J = [0, 1]$ and

$$\zeta(w) = 2w \bmod(1) \tag{5.10}$$

The invariant density for this map coincides with that for the *Uniform[0, 1]* distribution given by (5.6) and its correlation function is given by:

$$r_w(u) = 2^{-u}, u = 1, 2, \ldots. \tag{5.11}$$

Additional references include Lawrance and Spencer (1998), Lawrance and Balakrishna (2001), Sakai and Tokumuru (1980).

These results point to the limitations of using linear $ARMA$ models discussed in Section 1 for non-Gaussian and non-linear time series described by these maps and, for the Bernoulli map, for example, these methods lead to identifying an $AR(1)$ model but the mechanism generating the process is non-linear, with a similar remark applying to both the Tent and the Logistic map for which the traditional methods would lead to a purely random process as a stochastic model.

The chaotic maps mentioned above however still imply that the generating sequence has a short memory and has a correlation function, $r(u)$, which decays exponentially to 0, as $u \to \infty$. There has however lately been much development on devising chaotic *intermittency* maps exhibiting long or intermediate memory, that is, maps exhibiting correlation functions which decay at a polynomial rate, $r_y(u) \sim Bu^{-\alpha}$, $0 < \alpha < 2$, or even more slowly. A detailed expository discussion of such maps is given by Bhansali *et al.* (2002). Suffice to say here that early work on such maps is due to Manneville and Pomeau (1980) whose motivation was to provide a dynamical model implying intermittency, that is, a phenomenon displaying long periods of laminar behaviour together with short bursts of erratic behaviour - see Ott (1993) for a further discussion of the latter concept. Following is a slightly generalised version of the *intermittency* map studied by these authors:

$$\zeta_\beta(w) = \begin{cases} w(1 + 2^\beta w^\beta) & (0 \leq w \leq \frac{1}{2}, \\ 2w - 1 & (\frac{1}{2} < w \leq 1). \end{cases} \tag{5.12}$$

where $\beta \in (0, 1)$ is a free parameter of the map.

The invariant density for this map is of the following general form, with $J = [0, 1]$:

$$\chi(w) = \frac{H(w)}{w^\beta}, \quad \text{all } w \in J, \tag{5.13}$$

where $H(w)$ is a uniformly bounded but unknown function of w. Moreover while a precise closed-form expression for the correlation function, $r_w(u)$, of $\{w_t\}$ is as yet unknown, it is known that if $\kappa_t = \Omega(w_t)$, where Ω denotes a Hölder continuous function, then there exists at least one such function, Ω, for which $r_\kappa(u) = O(u^{-\alpha}), \alpha = \frac{1}{\beta} - 1$; hence, $r_\kappa(u) = O(u^-a), a \in (0, \infty)$. It should be noted that for $\beta = 0$, the intermittency map, (5.12), reduces to the Bernoulli map, (5.10); for $\beta = 1$, however, the invariant density does not exist.

It is also possible to devise intermittency maps which imply a logarithmic rate of decay of correlations, see Bhansali *et al.* (2002) for details. For an application of intermittency maps for modelling telecom network traffic, see Mandragon *et al.* (2001).

6 Acknowledgement

Thanks are due to Mark Holland for helpful discussions relating to Section 5 and especially for introducing the authors to the intermittency maps discussed in that section.

References

Adler, R., Feldman, R. and Taqqu, M. S. (1998). *A practical guide to heavy tails: Statistical techniques for analyzing heavy tailed distributions.* Birkhauser, Boston.

Bardet, J-M., Moulines, E. and Soulier, P. (1999). Recent advances on the semi–parametric estimation of the long–range dependence coefficient. In *ESAIM Proceedings*, pp. 23–43. Société de Mathématiques Appliquées et Industrielles.

Basak, G. K., Chan, N. H. and Palma, W. (2001). The approximation of long-memory processes by an ARMA model. *Journal of Forecasting*, **20,** 367–390.

Beran, J. (1993). Fitting long–memory models by generalized linear regression. *Biometrica*, **80,** 817–822.

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. Royal Statist. Soc.* B, **57,** 659–673.

Beran, J. (1997). Discussion of 'Heavy tail modelling and teletraffic data'. *The Annals of Statistics*, **25,** 1852–1856.

Beran, J., Bhansali, R. J. and Ocker, D. (1998). On unified model selection for stationary and nonstationary short– and long–memory autoregressive processes. *Biometrika*, **85,** 921–934.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics*, **2,** 489–502.

Bhansali, R. J. (1974). Asymptotic properties of the Wiener–Kolmogorov predictor I. *Journal of the Royal Statistical Society,* B, **36,** 61–73.

Bhansali, R. J. (1978). Linear prediction by autoregressive model fitting in the time domain. *The Annals of Statistics*, **6,** 224–231.

Bhansali, R. J. (1993). Order selection for linear time series models: a review. In *Developments in Time Series Analysis*, *London* (ed. T. Subba Rao), pp. 50–6. Chapman and Hall.

Bhansali, R. J. (1996a). Asymptotically efficient autoregrassive model selection for multistep prediction. *Ann. Inst. Statist. Math.*, **48,** 577–602.

Bhansali, R. J. (1996b). Recent developments in analysis of time series with infinite variance: a review. In *Athens Conference on Applied Probability and Time Series.* (eds P. M. Robinson and M. Rosenblatt), pp. 42–54. Springer. Volume 2.

Bhansali, R. J. (1999). Parameter estimation and model selection for multistep prediction of a time series: a review. In *Asymptotics, Nonparametrics and Time Series*, *New York* (ed. S. Ghosh), pp. 201–225. Marcel Dekker.

Bhansali, R. J. (2002). Multi-step forecasting. In *Companion to Economic Forecasting*, *New York* (eds M. P. Clements and D. Hendry), pp. 206–221. Blackwell.

Bhansali, R. J., Holland, M. and Kokoszka, P. S. (2002). Statistical properties of chaotic intermittency maps. Technical Report. University of Liverpool. In preparation.

Bhansali, R. J. and Kokoszka, P. S. (2001). Estimation of the long-memory parameter: a review and an extension. In *Proceedings of the Symposium on Inference for Stochastic Processes* (eds I.V. Basawa, C.C.Heyde and R.L. Taylor), IMS Lecture Notes, pp. 125–150. IMS.

Bhansali, R. J. and Kokoszka, P. S. (2002a). Computation of the forecast coefficients for multistep prediction of long-range dependent time series. *International Journal of Forecasting*, **18**, 181–206.

Bhansali, R. J. and Kokoszka, P. S. (2002b). Prediction of long-memory time series: an overview. *Estadistica*, **00**, 000–000; forthcoming. Preprint available at `http://www.liv.ac.uk/maths/SOR/HOME/RJBhansali.html`.

Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika*, **60**, 217–226.

Box, G. E. P. and Jenkins, G. M. (1970). *Time series analysis; forecasting and control.* Holden Day, New York.

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

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer-Verlag, New York.

Brodsky, J. and Hurvich, C. M. (1999). Multi-step forecasting for long-memory processes. *J. Forecasting*, **18**, 59–75.

Chatterjee, S. and Yilmaz, M. R. (1992). Chaos, fractals and statistics. *Statistical Science*, **7**, 49–121.

Cline, D. B. H. and Brockwell, P. J. (1985). Linear prediction of ARMA processes with infinite variance. *Stochastic Processes and their Applications*, **19**, 281–296.

Crato, N. and Ray, B. K. (1996). Model selection and forecasting of long–range dependent processes: results of a simulation study. *Journal of Forecasting*, **15**, 107–125.

Dahlhaus, R. (1989). Efficient parameter estimation for self similar processes. *Ann. Stat.*, **17**, number 4, 1749–1766.

Davis, H. T. and Jones, R. H. (1968). Estimation of the innovation variance of a stationary time series. *J. Amer. Statist. Assoc.*, **63**, 141–149.

Doob, J. L. (1953). *Stochastic Processes.* Wiley, New York.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221–238.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series and fractional differencing. *J. Time Series Anal.*, **1**, 15–30.

Hall, P. (1997). Defining and measuring long-range dependence. *Fields Institute Communications*, **11**, 153–160.

Hannan, E. J. (1987). Rational transfer function approximation. *Statistical Science*, **2**, 135–161.

Hannan, E. J. and Rissannen, J (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, **69**, 81–94; Correction (1983) **70**, 303.

Heyde, C. C. and Yang, Y. (1997). On defining long-range dependence. *J. Appl. Probab.*, **34**, 939–944.

Hidalgo, J. and Yajima, Y. (1999). Prediction of strong dependent processes in the frequency domain with application to signal processing. preprint.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, **68,** number 1, 165–176.

Hurvich, C. M. and Beltrao, K. I. (1993). Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series. *Journal of Time Series Analysis*, **14,** 455–472.

Hurvich, C. M. and Brodsky, J. (2001). Broadband semiparametric estimation of the memory parameter of a long–memory time series using fractional exponential models. *Journal of Time Series Analysis*, **22,** 221–249.

Kokoszka, P. S. (1996). Prediction of infinite variance fractional ARIMA. *Probability and Mathematical Statistics*, **16/1,** 65–83.

Kokoszka, P. S. and Taqqu, M. S. (1995). Fractional ARIMA with stable innovations. *Stochastic Processes and their Applications*, **60,** 19–47.

Kokoszka, P. S. and Taqqu, M. S. (1996). Parameter estimation for infinite variance fractional ARIMA. *The Annals of Statistics*, **24,** 1880–1913.

Künsch, H. (1987). Statistical aspects of self–similar processes. *Bernoulli*, **1,** 67–74.

Lawrance, A. J. (2001). Statistical aspects of chaos. a series of 7 lectures for MSM4S2. Technical Report. School of Mathematics and Statistics, University of Birmingham.

Lawrance, A. J. and Balakrishna, N. (2001). Statistical aspects of chaotic maps with negative dependence in a communications setting. *J. Royal Statist. Soc.,* B, **63,** 843–853.

Lawrance, A. J. and Spencer, N. M. (1998). Curved chaotic map time series models and their stochastic reversal. *Scandinavian Journal of Statistics*, **25,** 371–382.

Mandragon, R. J., Arrowsmith, D. K. and Pitts, J. M. (2001). Chaotic maps for traffic modelling and queueing performance analysis. *Performance Evaluation*, **43,** 223–240.

Manneville, P. and Pomeau, M. (1980). Intermittent transition to turbulence in dissipative dynamical systems. *Comm. Math. Phys.*, **74,** 189–197.

Mansfield, P., Rachev, S. T. and Samorodnitsky, G. (2001). Long strange segments of a stochastic process and long range dependence. *Annals of Applied Probability*, **11,** 878–921.

McLeod, A. I. (1998). Hyperbolic decay of time series. *J. Time Series Analysis*, **19,** 473–484.

Mittnik, S. and Rachev, S. (2000). *Stable Paretian Models in Finance.* Wiley, New York.

Moulines, E. and Soulier, P. (1999). Broad band log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, **27,** 1415–1439.

Moulines, E. and Soulier, P. (2000). Data–driven order selection for projection estimator of the spectral density of time series with long range dependence. *Journal of Time Series Analysis*, **21,** 193–218.

Ott, E. (1993). *Chaos in Dynamical Systems.* Cambridge University Press.

Parzen, E. (1969). Multiple time series modelling. In *Multivariate Analysis II, New York* (ed. P. R. Krishnaiah). Academic Press.

Priestley, M. B. (1981). *Spectral Analysis and Time Series: Volume 1.* Academic Press.

Rachev, S. T. and Samorodnitsky, G. (2001). Long strange segments in a long range-dependent moving average. *Stochastic Processes and their Applications*, **93,** 119–148.

Ray, B. K. (1993). Modeling long-memory processes for optimal long-range prediction. *Journal of Time Series Analysis*, **14,** 511–525.

Robinson, P. M. (1995b). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, **23,** 1630–1661.

Sakai, H. and Tokumuru, H. (1980). Autocorrelations of a certain chaos. *IEEE Trans. Acoustics, Speech and Signal Processing* A, **SSP-28,** 588–590.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* Chapman & Hall.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *The Annals of Statistics*, **8,** 147–164.

Taqqu, M. S. and Teverovsky, V. (1998). On estimating the intensity of long-range dependence in finite and infinite variance series. In *A practical guide to heavy tails: Statistical techniques for analyzing heavy tailed distributions*, *Boston* (eds R. Adler, R. Feldman and M. S. Taqqu), pp. 177–217. Birkhauser.

Tiao, G. C. and Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time-series analysis. *J. Forecasting*, **13,** 109–131.