PREDICTION OF LONG-MEMORY TIME SERIES: AN OVERVIEW *†

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

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

14 February, 2001

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 also discussed and a review of the various methods proposed for doing so is given. The predictive efficacy of the *Type-II* approaches relative to the *Type-I* approches is investigated by a simulation study in which the *Type-I* forecasts are obtained by autoregressive model fitting while the *Type-II* forecasts are constructed by fitting an $ARFIMA(1, d, 1)$ model to each generated time series, but the actual simulated processes follow a variety of $ARFIMA$ and $FEXP$ models and not necessarily an $ARFIMA(1, d, 1)$ model.

1 Introduction

The long-memory time series models have attracted much attention lately and there is now a growing realisation that time series possessing long-memory characteristics arise in subject areas as diverse as Economics, Geophysics and telecom traffic. In this paper, we consider the question of multistep prediction of long-memory time series and aim to provide a current state-of-the-art survey of the recent advances on this question.

In Section 2, the principal characteristics of long-memory time series models are contrasted with those of the more usual short-memory models and it is noted that since a long-memory

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

process admits an infinite autoregressive representation with square-summable coefficients, the classical Wiener-Kolmogorov prediction theory, see, e.g., Whittle (1963), may still be applied for its linear least-squares prediction.

A review of the Wiener-Kolmogorov prediction theory and the associated Wiener-Hopf spectral factorization procedure is given in Section 3.

In Section 4, a distinction between the *Type-I* and *Type-II* approaches for linear least-squares prediction of long memory time series is drawn. The former constructs the multistep forecasts of a long-memory time series without explicitly seeking to model its long-memory characteristics, that is, by approximating a long-memory time series by either a long autoregression or a standard ARMA model, and, in essence, by employing the well-established modelling strategies for predicting a short-memory time series. The latter, by contrast, explicitly seeks to model the long-memory stochastic characteristics by specifying that the dth difference, $0 < d < 0.5$, of an oberved time series follows a standard short-memory model.

Section 5 discusses the implementation of the *Type-I* approach with an observed long-memory time series while Section 6 gives a review of the different approaches, especially the parametric, non-parametric, semi-parametric and hierarchical model fitting approaches, to constructing suitable stochastic models for long-memory processes. In Section 7, the question of estimating the parameters of a long-memory model is discussed.

The various methods for computing the prediction constants when constructing the multistep forecasts of a long-memory time series by *Type-II* approaches are reviewed in Section 8, and in Section 9, which is the final section, the results of a simulation study designed to compare the *Type-I* and *Type-II* approaches to multistep prediction of long-memory time series are discuused.

2 Characteristic Features of Short- and Long-memory processes

The question of a formal definition of a long-memory process has received some attention in the literature; see Cox (1991), Heyde and Yang (1997), Hall (1997). Here, we take a more informal approach and distinguish in general terms between a short-memory and a long-memory process by comparing their characteristic features. The standard autoregressive-moving average and the related linear models for a discrete-time stationary process, $\{x_t\}$, say, provide examples of a short-memory process. Thus, let $\mu_x = E(x_t)$ and $R_x(u) = E[\{x_t - \mu_x\}\{x_{t+u} - \mu_x\}]$, $(t, u = \ldots, -1, 0, 1, \ldots)$ denote the mean and covariance function of $\{x_t\}$ and define its correlation function by $r_x(u) = R_x(u)/R_x(0)$ and the spectral density function by

$$f_x(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} R_x(u) \exp(-iu\lambda). \tag{2.1}$$

Then, a short-memory stationary process may be characterized by the following features:

$S1$). 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 $u \to \infty$,

$$r_x(u) \sim B\rho^{|u|}, \quad 0 < \rho < 1. \tag{2.2}$$

$S2$). The process, $\{x_t\}$, admits an infinite moving average representation and an infinite autoregressive representation as follows:

$$x_t = \sum_{j=0}^{\infty} b(j)\epsilon_{t-j}, \tag{2.3}$$

$$\sum_{j=0}^{\infty} a(j)x_{t-j} = \epsilon_t, \tag{2.4}$$

where $b(0) = 1$, $a(0) = 1$, and the coefficients, $b(j)$ and $a(j)$, also tend to zero at an exponential rate as j approaches ∞.

$S3$). The spectral density function, $f_x(\lambda)$, is continuous and bounded for all λ: $0 < m < f_x(\lambda) < M < \infty$, all $\lambda \in [0, \pi]$.

Note that $S1$) and $S2$) above imply that the covariance function, $R_x(u)$, and the coefficients, $b(j)$ and $a(j)$, occurring in (2.3) and (2.4) of a short-memory stationary process are absolutely summable:

$$\sum_{u=-\infty}^{\infty} |R_x(u)| < \infty, \tag{2.5}$$

$$\sum_{j=0}^{\infty} |b(j)| < \infty, \tag{2.6}$$

$$\sum_{j=0}^{\infty} |a(j)| < \infty. \tag{2.7}$$

By contrast, a long-memory stationary process is usually characterised by the following features, see Hosking (1981):

$L1$). The spectral density function, $f(\lambda)$, diverges to ∞ as $\lambda \to 0$, and, for a constant, d, such that $0 < d < 0.5$, we may write

$$f(\lambda) \sim \lambda^{-2d} L(1/|\lambda|), \quad \text{as} \quad \lambda \to 0, \tag{2.8}$$

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

$L2$). The covariance function, $R(u)$, is not summable, that is,

$$\sum_{u=-\infty}^{\infty} R_x(u) = \infty. \tag{2.9}$$

$L3$). The process, $\{x_t\}$, still admits an infinite moving average and an infinite autoregressive representation defined above in equations (2.3) and (2.4), but the coefficients $b(j)$ and $a(j)$ in these representations now decrease to zero at a geometric rate and satisfy:

$$\sum_{j=0}^{\infty} b(j)^2 < \infty, \tag{2.10}$$

$$\sum_{j=0}^{\infty} |a(j)| < \infty. \tag{2.11}$$

Now, since as $j \to \infty, b(j) \sim Bj^{d-1}$, the $b(j)$ are square-summable but not absolutely summable and thus (2.6) does not hold though since as $j \to \infty, a(j) \sim Aj^{-d-1}$, the $a(j)$ are absolutely summable and (2.7) still holds.

Henceforth, for brevity, we refer to a stationary process possessing the features $S1)$ - $S3)$ above more simply as a short-memory process and that possessing features $L1)$ - $L3)$ above as a long-memory process and we thus omit the qualifier 'stationary'. The constant, d, occurring in (2.8) is known as the memory-parameter of a long-memory process.

It should be noted that from the point of view of linear least-squares prediction of a long-memory process, since (2.10) and (2.11) hold, the classical Wiener-Kolmogorov prediction theory continues to apply; see, for example, Theorem 4.2 of Doob (1953), and, accordingly, we next give a brief outline of this theory.

3 The Wiener-Kolmogorov Prediction Theory

Let $\{x_t\}$ be a discrete-time (weakly) stationary process with mean 0 and such that it admits the one-sided infinite moving average and autoregressive representations defined by (2.3) and (2.4) in which $\{\epsilon_t\}$ is a sequence of uncorrelated random variables, each with mean 0 and variance σ^2 and the $b(j)$ and $a(j)$ are square-summable real coefficients, normalized to ensure that $b(0) = a(0) = 1$.

Suppose that for some integer, n, the infinite past, $\{x_{n+1-j},\ j \geq 1\}$, may be treated as known and it is desired to predict the future values, $\{x_{n+j},\ j \geq 1\}$, by linear least-squares methods. Thus, if, for some $h \geq 1$,

$$\hat{x}_n(h) = \sum_{j=0}^{\infty} \gamma_h(j) x_{n-j} \tag{3.1}$$

denotes the resulting predictor of x_{n+h}, the h-step prediction constants, $\gamma_h(j)$, are chosen to ensure that the h-step mean squared errror of prediction,

$$V(h) = E\left[\{\hat{x}_n(h) - x_{n+h}\}^2\right], \tag{3.2}$$

is as small as possible within the class of all linear predictors of x_{n+h}.

We consider the question of how to express the prediction constants, $\gamma_h(j)$, in terms of the coefficients, $a(j)$, of the autoregressive representation, (2.4), for $\{x_t\}$; in addition, we also discuss two related questions, namely the conditions under which the one-sided representations, (2.3), (2.4), exist, and how to find the $a(j)$ and $b(j)$ from a knowledge only of the spectral density function, $f_x(\lambda)$, of $\{x_t\}$.

Following Whittle (1963), and in view of the representation (2.3), we may rewrite (3.1) as

$$\hat{x}_n(h) = \sum_{j=0}^{\infty} \phi_h(j) \epsilon_{n-j}, \tag{3.3}$$

where

$$(3.4) \qquad \sum_{j=0}^{\infty} \phi_h(j) z^j = \left(\sum_{j=0}^{\infty} b(j) z^j \right) \left(\sum_{j=0}^{\infty} \gamma_h(j) z^j \right).$$

Hence, we may write (3.2) as

$$(3.5) \qquad V(h) = \sum_{j=0}^{h-1} b^2(j) + \sum_{j=0}^{\infty} \{b(j+h) - \phi_h(j)\}^2.$$

It is clear that as a function of the $\phi_h(j)$, the smallest value of $V(h)$ is obtained by setting $\phi_h(j) = b(j+h)$ $(j \geq 0)$, and then

$$(3.6) \qquad \hat{x}_n(h) = \sum_{j=0}^{\infty} b(j+h) \epsilon_{n-j},$$

$$(3.7) \qquad V(h) = \sum_{j=0}^{h-1} b^2(j).$$

Furthermore, if

$$(3.8) \qquad A(z) = \sum_{j=0}^{\infty} a(j) z^j,$$

denotes the characteristic polynomial of the $a(j)$ and

$$(3.9) \qquad B(z) = \sum_{j=0}^{\infty} b(j) z^j,$$

that of the $b(j)$, then, in view of the identity, $A(z) = [B(z)]^{-1}$, $|z| \leq 1$, we may write, for each $h \geq 1$,

$$(3.10) \qquad \hat{x}_n(h) = - \sum_{j=1}^{h-1} a(j) \hat{x}_n(h-j) - \sum_{j=h}^{\infty} a(j) x_{n+h-j}.$$

Note that

$$(3.11) \qquad \begin{aligned} x_{n+h} &= - \sum_{j=1}^{h-1} a(j) x_{n+h-j} - \sum_{j=h}^{\infty} a(j) x_{n+h-j} + \epsilon_{n+h}, \\ &= \sum_{j=1}^{h-1} b(j) \epsilon_{n+h-j} + \sum_{j=h}^{\infty} b(j) \epsilon_{n+h-j} + \epsilon_{n+h}. \end{aligned}$$

Thus, see Box and Jenkins (1970), for obtaining $\hat{x}_n(h)$, we replace the unknown future values x_{n+h-j} $(j = 1, \ldots, h-1)$ by their linear least-squares forecasts, replace the future 'innovations', ϵ_{n+h-j} $(j = 1, ..., h)$, by zero, and leave $\{x_{n+1-j}, j \geq 1\}$, and $\{\epsilon_{n-j}, j \geq 1\}$ unchanged since their values are treated as known.

Next, consider the question of conditions under which the representations (2.3) and (2.4) exist. As discussed by Whittle (1963), a necessary and sufficient condition under which the moving average representation, (2.3), with square-summable coefficients, $b(j)$, exists is that

$$(3.12) \qquad \int_{-\pi}^{\pi} \log f_x(\lambda) d\lambda > -\infty.$$

Moreover, by the Wiener-Levy theorem, Zygmund (1959), this moving average representation may be inverted to the autoregressive representation, (2.4), provided

$$B(z) \neq 0, \quad |z| \leq 1. \tag{3.13}$$

As in Whittle (1963), among others, under (3.12) and (3.13), the $a(j)$ and $b(j)$ may be determined by the Wiener-Hopf spectral factorization procedure, see also Bhansali (1974). An algorithmic statement of this procedure, without giving the underlying analytical justification of its validity, is given below:

Under (3.12), $\log f_x(\lambda)$ admits a Fourier expansion:

$$\log f_x(\lambda) = \sum_{v=-\infty}^{\infty} c(v) \exp(-iv\lambda), \tag{3.14}$$

where

$$c(v) = (2\pi)^{-1} \int_{-\pi}^{\pi} \log f_x(\lambda) \exp(iv\lambda) d\lambda. \tag{3.15}$$

Hence, we set

$$\sigma^2 = 2\pi \exp c(0) = 2\pi \exp \left\{ (2\pi)^{-1} \int_{-\pi}^{\pi} \log f_x(\lambda) d\lambda \right\}, \tag{3.16}$$

$$B\left(e^{-i\lambda}\right) = \exp \left\{ \sum_{v=1}^{\infty} c(v) \exp\left(-iv\lambda\right) \right\}, \tag{3.17}$$

$$A\left(e^{-i\lambda}\right) = \exp \left\{ -\sum_{v=1}^{\infty} c(v) \exp(-iv\lambda) \right\}. \tag{3.18}$$

Now, for all $j \geq 0$, the $b(j)$ and $a(j)$ are given by:

$$b(j) = (2\pi)^{-1} \int_{-\pi}^{\pi} B\left(e^{-i\lambda}\right) \exp(ij\lambda) d\lambda, \tag{3.19}$$

$$a(j) = (2\pi)^{-1} \int_{-\pi}^{\pi} A\left(e^{-i\lambda}\right) \exp(ij\lambda) d\lambda. \tag{3.20}$$

A useful set of sufficient conditions which ensure that (3.12) and (3.13) hold is the following:

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

$$\sum_{j=-\infty}^{\infty} |R_x(u)| < \infty. \tag{3.21}$$

b) The spectral density function is non-vanishing :

$$f_x(\lambda) \neq 0, \text{ all } \lambda \in [-\pi, \pi]. \tag{3.22}$$

As shown in Theorem 3.8.4 of Brillinger (1975), for a process satisfying (3.21) and (3.22), the representations (2.3) and (2.4) both exist and moreover if the coefficients, $b(j)$ and $a(j)$, are determined as in (3.19) and (3.20) by the spectral factorization procedure then they are absolutely summable and satisfy (2.6) and (2.7), respectively. However, the resulting $a(j)$ and $b(j)$ need not be exponentially decreasing. Baxter (1962) gives a necessary and sufficient condition on $f_x(\lambda)$ which ensures this, namely that the spectral density function, $f_x(\lambda)$, coincides almost everywhere in $-\pi \leq \lambda \leq \pi$ with a function which is analytic for real λ and has no real zeroes; see also (Bary (1964), p. 80). In the following lemma, a sufficient set of conditions for ensuring that the $a(j)$ and $b(j)$ decrease to zero exponentially fast is given directly in terms of the behaviour of the characteristic polynomial, $A(z)$, of the $a(j)$.

LEMMA 3.1 *Suppose for some $r > 1$, $A(z)$ is analytic and has no zeroes in $D_r := \{z : |z| < r\}$. Then there is a constant $0 < \rho < 1$ such that for all sufficiently large j, $|b(j)| \leq \rho^j$, $|a(j)| \leq \rho^j$.*

PROOF. For $z \in D_r$ define $B(z) = 1/A(z)$. By the assumptions $B(z)$ is also analytic in D_r and so has an expansion $B(z) = \sum_{j=0}^{\infty} b(j)z^j$. The radius of convergence of B, say r_B, satisfies $r_B \geq r$. Since, see e.g. Section 10.5 of Rudin (1987), $\limsup_{j\to\infty} |b(j)|^{1/j} = r_B^{-1}$, we conclude that $\limsup_{j\to\infty} |b(j)|^{1/j} \leq r^{-1}$. Fix $r^{-1} < \rho < 1$. Then for sufficiently large j, $|b(j)|^{1/j} \leq \rho$. For $z \in D_r$, $B(z)A(z) = 1$, so B does not have any zeroes in D_r either. Reversing the roles of A and B we conclude that $|a(j)|^{1/j} \leq \rho$ for sufficiently large j.

4 The Type-I and Type-II approaches

As discussed in Section 2, there are various approaches to defining a long-memory process, and especially to ensure that the definition encompasses not only the weakly stationary processes but also strictly stationary processes with infinite variance and more general classes of processes. However, from the point of view of linear least-squares prediction, we say that a discrete-time, weakly stationary process, $\{y_t\}$, possesses long-memory if the process, $\{x_t\}$, say, obtained by taking its *dth* fractional difference, $0 < d < 0.5$, is a short-memory process and such that its covariance and spectral density functions satisfy at least the conditions (3.21) and (3.22). The nature of conditions imposed on $\{x_t\}$ is made more precise in Section 6 when discussing the various long-memory stochastic models.

It readily follows that the logarithm of the spectral density function of $\{y_t\}$, $f_y(\lambda)$, is given by

(4.1) $$\log f_y(\lambda) = -2d \log\{2|\sin(\lambda/2)|\} + \log f_x(\lambda).$$

Moreover, since (see, Tolstov (1976), p.92),

(4.2) $$\int_{-\pi}^{\pi} \log\{2|\sin(\lambda/2)|\}d\lambda = 0,$$

we have

(4.3) $$\int_{-\pi}^{\pi} \log f_y(\lambda)d\lambda = \int_{-\pi}^{\pi} \log f_x(\lambda)d\lambda > -\infty.$$

Indeed, on letting z denote the backshift operator, $z^j y_t = y_{t-j}$, we may write

$$y_t = (1-z)^{-d} x_t, \tag{4.4}$$

where $\{x_t\}$ admits the infinite moving average and autoregressive representations given by (2.3) and (2.4), with coefficients, $b(j)$, and $a(j)$, satisfying (2.6) and (2.7), respectively. Thus, the long-memory process, $\{y_t\}$, also admits one-sided infinite moving average and autoregressive representations as follow:

$$y_t = \sum_{j=0}^{\infty} \psi(j)\epsilon_{t-j}, \ \psi(0) = 1, \tag{4.5}$$

$$\sum_{j=0}^{\infty} \pi(j) y_{t-j} = \epsilon_t, \ \pi(0) = 1, \tag{4.6}$$

where $\{\epsilon_t\}$ is a sequence of uncorrelated random variables, each with mean 0 and variance σ^2, and the coefficients, $\psi(j)$ and $\pi(j)$ in the representations (4.5) and (4.6) are related to the coefficients $b(j)$ and $a(j)$ in the representations (2.3) and (2.4) by:

$$\sum_{j=0}^{\infty} \psi(j) z^j = (1-z)^{-d} \sum_{j=0}^{\infty} b(j) z^j, \tag{4.7}$$

$$\sum_{j=0}^{\infty} \pi(j) z^j = (1-z)^{d} \sum_{j=0}^{\infty} a(j) z^j. \tag{4.8}$$

Hence, it follows that the $\psi(j)$ are square-summable and satisfy:

$$\sum_{j=0}^{\infty} \psi(j)^2 < \infty, \tag{4.9}$$

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

$$\sum_{j=0}^{\infty} |\pi(j)| < \infty. \tag{4.10}$$

It is clear that in view of the representations, (4.5) and (4.6) and the relation (4.4) connecting $\{y_t\}$ with the short-memory process, $\{x_t\}$, we may distinguish between the following two approaches to multistep prediction of a long-memory process:

Type-I: Apply the Wiener-Kolmogorov prediction theory directly to the long- memory process, $\{y_t\}$, without appealing to (4.4), that is, without explicitly identifying the value of d and the structure of the short-memory process, $\{x_t\}$, as implied by the relations (4.7) and (4.8). In this approach, the linear least-squares predictor of the future values, $\{y_{n+h},\ h = 1, 2, \ldots\}$, based on the infinite past, $\{y_{n-j},\ j = 0, 1, \ldots\}$, of $\{y_t\}$ is still given by equations (3.6), but with the coefficients $b(j)$ replaced by the $\psi(j)$. Moreover, we may still equivalently write the predictor explicitly in the form (3.10), but now the $a(j)$ are replaced by the $\pi(j)$ and the $\{x_{n-j},\ j = 0, 1, \ldots\}$ replaced by the past $\{y_{n-j}, j = 0, 1, \ldots\}$ of y_t.

Type-II. Model $\{y_t\}$ explicitly in the form (4.4), with possibly a model specified also for $\{x_t\}$. Now the linear least-squares predictor of the unknown future values of $\{y_t\}$ is determined by first finding the value of d and the model postulated for $\{x_t\}$ and then using (4.7) and (4.8) for determining the values of the $\psi(j)$ and $\pi(j)$.

If the spectral density function, $f_y(\lambda)$, of $\{y_t\}$ is treated as known, these two approaches are equivalent and clearly give rise to the same linear least-squares predictor. In practice, however, when only a finite realization of $\{y_t\}$ has been observed, they give rise to two different approaches to estimating the linear least-squares predictor from the observed realization.

5 Implementation of The Type-I Approach

For implementing the *Type-I* approach with an observed time series, no new theory is required, and a long-memory time series is treated on par with a short-memory time series without recognising its special structure. There is nevertheless a choice of modelling strategies, as further described below:

1). *Parameteric ARMA models.* A standard autoregressive-moving average model of order (p, q), an $ARMA(p, q)$ model, defined later in Section 6 by equation (6.1), 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; see Bhansali (1993) for a review of order selection procedures for linear time series models. The use of a three-stage Hannan- Rissanen procedure, see Hannan and Rissannen (1982), may also be considered. Crato and Ray (1996) consider the use of this approach for multistep prediction of long-memory time series. A related but different possibility is investigated by Tiao and Tsay (1994), who point out that an $ARMA(1,1)$ model may be used for multistep prediction of a long-memory time series, provided the parameters of the fitted model are estimated by an adaptive procedure, that is, see Bhansali (2001), by a multistep method such that the estimates are constructed anew separately 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. For the situation in which $\{y_t\}$ is a fractionally-differenced white noise, that is, (4.4) holds, but $\{x_t\} = \{\epsilon_t\}$ is a purely random process, these authors compare the predictive performance of the adaptive multi-step procedure relative to the use of the actual long-meory model generating $\{y_t\}$, and report that the loss in predictive efficiency due to using the incorrect $ARMA(1,1)$ model in this situation is not serious. On the other hand, however, the loss in predictive efficiency due to using the classical one-step estimation method in this situation can be substantial, and thus the use of the adaptive multistep estimation procedure when an incorrect model is fitted to the observed time series is justified by these results. Nevertheless, the comparisons are based on an implausible assumption that the value of d, the long-memory parameter, generating $\{y_t\}$ is known, and thus ignore the effect of parameter estimation on the mean squared error of prediction of the two methods, see Brodsky and Hurvich (1999), who have repeated the above study without invoking the assumption of known parameters and report that, after taking account of the parameter estimation effects, the long-memory model can provide better multistep predictions than an incorrect $ARMA(1,1)$ model. The basic question considered in these studies, namely a comparison of the *Type-I* and *Type-II* approaches, is a pertinent one and we return to it in Section 9.

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, (2.4), whose coefficients, $a(j)$, are absolutely summable and satisfy (2.7); 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; thus the multistep forecasts are still given by (3.10), but with the $a(j)$ replaced by their estimates, $\hat{a}_k(j)$, and the second sum to the right of this equation truncated at k. 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. It should be noted that this direct approach 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 (2001) for a further discussion of the differences between these two methods and for an explanation of the motivation for considering the direct method. For short-memory time series, an asymptotic optimality property for the direct method has been established by Bhansali (1996). In Section 9, the predictive performance of the *Type-I* plug-in and *Type-I* direct methods of fitting autoregressions is compared with that of the *Type-II* approach on simulated long-memory time series.

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

6 Stochastic Models for Long-Memory Processes

As discussed in Section 4, the *Type-II* approach is implemented by explicitly recognising that the observed process, $\{y_t\}$, has long-memory and by postulating that (4.4) holds and that its dth fractional difference, $0 < d < 0.5$, $\{x_t\}$, follows a short-memory model. As with the *Type-I* approach discussed above, there is a choice of different strategies concerning the nature of the stochastic model specified for $\{x_t\}$. 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 7, we discuss the associated question of how to estimate the parameters of the specified models from an observed long-memory time series:

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. For this class, $\{y_t\}$ satisfies (4.4), but $\{x_t\}$ follows the standard $ARMA(p,q)$ model:

$$\sum_{j=0}^{p} \phi(j) x_{t-j} = \sum_{j=0}^{q} \theta(j) \epsilon_{t-j}, \tag{6.1}$$

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

$$\Phi(z) = \sum_{j=0}^{p} \phi(j) z^j, \Theta(z) = \sum_{j=0}^{q} \theta(j) z^j, \tag{6.2}$$

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:

$$f_y(\lambda) = \frac{\sigma^2}{2\pi} \left|1 - e^{-i\lambda}\right|^{-2d} \frac{|\Theta(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2}. \tag{6.3}$$

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) and thus

$$f_y(\lambda) = \frac{\sigma^2}{2\pi} \left|1 - e^{-i\lambda}\right|^{-2d} \left[\exp\{\sum_{j=1}^{n} c(j)\cos(j\lambda)\}\right], \tag{6.4}$$

where $n \geq 1$ is the order of the model. 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 (2.8) gives the class of models being considered. A basic model of this class is obtained by specifying that $L(1/|\lambda|)$ is a constant in a neighbourhood of the origin.

As discussed by Hall (1997), the specification (2.8) is equivalent to requiring that the covariance function, $R_y(s)$, of $\{y_t\}$ takes the following form:

$$R(s) \sim L(s)s^{-(1-2d)}. \tag{6.5}$$

A principal difficulty with the non-parametric approach is that since in the specification (2.8), or (6.5), the behaviour of $f_y(\lambda)$ is not specified at all frquencies, the spectral factorization procedure discussed in Section 3 does not apply and hence in its existig form this approach is suitable only for the problem of estimating the long-memory parameter, d, and it is not suitable for multi-step prediction of $\{y_t\}$. Crato and Ray (1996) consider a modification of this approach in which the short-memory process obtained by filtering out the long-memory component follows a standard $ARMA$ model given by equation (6.1). Such a specification however corresponds to adopting a parametric approach.

3). *Hierarchical Parametric Models.* This approach builds on the success of the autoregressive model approach discussed in Section 5 for short-memory time series and seeks to generalise it to the class of long- memory processes. According to this approach, $\{y_t\}$ is assumed to satisfy (4.4), 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, this approach is based upon the existence of a one-sided infinite autoregressive and infinite moving average representations for $\{x_t\}$ discussed in Section 3 and it is thus more plausible than a fully parametric approach, and at the same time it may still be used 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 (2.4) and (2.3), but with cefficients, $a(j)$ and $b(j)$ decreasing to zero at an exponential rate. Given a stretch of N 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 7, where k is assumed to be a function of N and allowed to approach ∞ simultaneously but sufficiently slowly with it.

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 (6.4) but with $n = \infty$ and the coefficients $c(j)$ decreasing to 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 (4.4), $\log f_y(\lambda)$ admits the decomposition (4.1), 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 (6.4) with $n = k$, are estimated, though a regression procedure discussed further in Section 7 is used for parameter estimation.

7 Parameter Estimation for Long-Memory Models

Denote the observed time series by $y_1, \ldots, y_N$, $N > 1$, which we suppose is a part-realizaton a long-memory process, $\{y_t\}$, satisfying (4.4). 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 (2001a), among others.

It is possible to distinguish between *parametric*, *non-parametric*, *semi-parametric* and *hierarchical* model fitting approaches to the estimation of the parameters of a long-memory model. In addition, the underlying methods may be grouped further under two broad headings: A) Likelihood-based methods; B) Regression-based methods.

The periodogram function of the observed time series is given by

$$I_y(\lambda) = (2\pi N)^{-1} \left| \sum_{t=1}^{N} y_t \exp(-it\lambda) \right|^2, \tag{7.1}$$

and both the likelihood-based and regression-based methods adopt the periodogram as a basic statistic. A convenient way of implementing the likelihood-based methods is to use the approximate Whittle likelihood function and to separately parametrize the variance, $\sigma^2 = \mathrm{Var}(\epsilon_t)$, of the innovation process, $\{\epsilon_t\}$, see (4.5), from the remaining parameters of the model. Under this assumption, the Whittle likelihood function of the data is given by:

$$L(y|\tau) = \int_{-\pi}^{\pi} \frac{I_y(\lambda)}{g(\lambda, \tau)} d\lambda, \tag{7.2}$$

where

$$g(\lambda, \tau) = \left| \sum_{j=0}^{\infty} \psi(j) \exp(-ij\lambda) \right|^2, \tag{7.3}$$

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. 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 and, in the following exposition, only the use of the Whittle likelihood function is considered:

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.

Consider the parametric approach first. If the order, (p,q), of the $ARFIMA(p,d,q)$ model is treated as known, the $g(\lambda,\tau)$ occurring in (7.2) is given by

$$g(\lambda,\tau) = \left|1 - e^{-i\lambda}\right|^{-2d} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}, \tag{7.4}$$

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

$$g(\lambda,\tau) = \left|1 - e^{-i\lambda}\right|^{-2d} |A_k(e^{-i\lambda})|^{-2}, \tag{7.5}$$

where

$$A_k(z) = 1 + a(1)z + ... + a(k)z^k. \tag{7.6}$$

Next, consider the regression-based procedures. For the non-parametric approach, $f_y(\lambda)$, is specified by (2.8), with $L(\lambda) = C$, 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), who however adopted the specification (4.1) and suggested regressing $\log I_y(\lambda)$ on the first term to the right of this equation in a neighbourhood of the zero frequency for estimating d. The two procedures are nevertheless readily seen to be equivalent. Further 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.

As discussed earlier in Section 6, for the hierarchical model fitting approach involving the fitting of $FEXP$ models, the behaviour of $f_y(\lambda)$ for all $\lambda \in [0,\pi]$ is specified instead of only near the origin, which is the case with the non-parametric approach, and since N is finite, the infinite Fourier expansion, (3.14), for $\log f_x(\lambda)$ is truncated at k, a finite integer. An estimate of d is accordingly constructed, see Hurvich and Brodsky (2001), Moulines and Soulier (1999,2000) 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 (2.8) for $f_y(\lambda)$, with $L(\lambda) = C$, a constant for all λ, is also adopted for the semi-parametric approach. However, see Künsch (1987), Robinson (1995), estimates of d and C are constructed from the Whittle likelihood, which is now computed only over a neighbourhood of the 0 frequency.

Asymptotic statistical properties of the estimates provided by these four approaches have been extensively investigated and the following is a brief discussion of the results obtained and of some of the shortcomings of each approach:

If the order, (p, q), of an $ARFIMA(p, d, q)$ model may be treated as known, the maximum likelihood estimator, $\hat{\tau}$, of τ has been shown to be asymptotically Normal and Fisher efficient, in the sense that $\sqrt{N}(\hat{\tau} - \tau_0)$, where τ_0 denotes the true value of τ, is as $N \to \infty$, asymptotically Normally distributed with 0 mean and covariance matrix, W^{-1}, say, where W denotes the relevant Fisher information matrix applicable to this problem; see Fox and Taqqu (1986), Dahlhaus (1989), Giraitis and Surgailis (1990), among others. In practice, however, the order (p, q) is invariably unknown. There is much evidence to suggest that if the actual order is misspecified then 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) who show that whereas the BIC criterion provides a consistent order estimation, AIC does not do so. A related conceptual difficulty is, see Hannan (1987), that an observed time series need not follow a 'true' $ARMA$ model of a given, even though unknown, order.

The asymptotic distributions of both non-parametric and semi-parametric estimators, has been rigorously derived by Robinson(1995a, 1995b). It is clear from the results given there that while the non-parametric and semi-parametric approaches are quite general and make few prior assumptions concerning the stochastic structure of the observed time series, the estimates they give rise to have a slow rate of convergence, namely $N^{-\delta}$, where, no matter how smooth the short-memory spectral desity, $f_x(\lambda)$, is, $\delta < 0.4$; see Bardet *et al.* (1999) for a further discussion this property. 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.

Moulines and Soulier (1999), see also Hurvich and Brodsky (2001), have rigorously studied the asymptotic distribution of the $FEXP$ estimator, $\hat{d}_E$, say, and under regularity conditions including the assumption that the coefficients, $c(j)$, in the Fourier expansion (3.14) of $\log f_x(\lambda)$ decrese to zero at an exponential rate, and k, the truncation point used for constructing $\hat{d}_E$ is a function of N and approaches infinity simultaneously but sufficiently slowly with it, $\sqrt{N/\log N}\{\hat{d}_E - d^0\}$ is, as $N \to \infty$, asymptotically distributed as Normal with 0 mean and finite variance, whose value depends upon the degree of smoothing applied to the periodogram before performing the log-periodogram regression. The $FEXP$ method thus 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.

Bhansali and Kokoszka (2001a) have studied the empirical behaviour of the FAR method by a simulation study, and, also, established that, under certain regularity conditions, including the assumption that the order, k, of the autoregressive model is a function of N and converges to infinity simultaneously with it, this method provides a consistent estimator of d and, also, of the coefficients, $a(j)$, in the autoregressive representation for $\{x_t\}$.

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

If the parametric or the hierarchical model fitting approaches discussed in Sections 6 and 7 above are used for modelling the behaviour of an observed long-memory process, then having estimated the parameters of the postulated model, the corresponding estimate, $\hat{f}_y(\lambda)$, say, of the spectral density function, $f_y(\lambda)$, of the process, may be constructed from the fitted model by replacing the unknown model parameters by their estimates. Thus, if the parametric approach based on the fitting of an $ARFIMA(p, d, q)$ model is adopted, $f_y(\lambda)$, is given by equation (6.3) and its estimate may be constructed from this equation by replacing the $\theta(j), \phi(j), d$, and σ^2 on its right by their Whittle likelihood estimates. Also, for the $FEXP$ approach, noting, as discussed above, that d is estimated by truncating the infinite Fourier expansion of $\log f_x(\lambda)$ at a finite value, k, $f_y(\lambda)$, may be estimated from equation (6.4) by setting $n = k$ and on replacing the unknown d, σ^2 and the $c(j)$ by their estimates; while for the FAR approach, $f_y(\lambda)$, is effectively given equation (6.3), but with $q = 0, p = k$ and the polynomial $\Phi(e^{-i\lambda})$ replaced by the polynomial $A_k(e^{-i\lambda})$, where $A_k(z)$ is defined by equation (7.6).

By contrast, however, by their very nature, the non-parametric and semi-parametric approaches do not seek to yield an estimate of $f_y(\lambda)$, for all $\lambda \in [0, \pi]$ and hence in their original form they are not suitable for the problem of constructing multistep predictors of the future values of $\{y_t\}$. Although Crato and Ray (1996) have used this approach for multistep prediction by 'filtering' out the long-memory component and then fitting a parametric $ARMA$ model to the resulting output series, this two-stage approach is still parametric and, even though its optimality is unclear, in the discussion below, it may be used in the same way as the parametric approach for multistep prediction of a long-memory process.

Bhansali and Kokoszka (2001b) consider the question of how to compute the prediction constants for multistep prediction of $\{y_t\}$ by applying the Wiener-Kolmogorov prediction theory, discussed earlier in Sections 3 and 4, to this problem. Three different methods for implementing the *Type-II* approach 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 spectral density function, $\hat{f}_y(\lambda)$, of the model specified for $\{y_t\}$, for brevity, in the following exposition, we do not explicitly distinguish between the parametric or the hierarchical model fitting approaches to estimating its parameters and thus do not emphasise the actual method used for constructing $\hat{f}_y(\lambda)$.

Consider the Truncation method first. This method is similar to that suggested by Box and Jenkins (1970) and it is based on the infinite autoregressive representation, (4.6), for $\{y_t\}$. Estimates, $\hat{\pi}(j)$, say, of the $\pi(j)$ may be constructed in a straightforward fashion, see Bhansali and

Kokoszka (2001b) for details, from the estimated parameters of the specified model. However, since only a finite past of $\{y_t\}$ has been observed, the infinite sequence of $\hat{\pi}(j)$ will be truncated at a suitable finite value, K, say. For each lead time, h, the multistep prediction may then be constructed as in (3.10), but with the $a(j)$ replaced by the $\hat{\pi}(j)$, the past values of $\{x_t\}$ replaced by those of $\{y_t\}$ and the second sum to the right of this equation truncated at K. Bhansali and Kokoszka (2001b) suggest a computationally- effective procedure for computing the $\hat{\pi}(j)$ by a fast Fourier transform algorithm.

The Type-II Plug-In method is based on the innovations algorithm, see Brockwell and Davis (1991) and Crato and Ray (1996), who use this algorithm in their simulation study for constructing multistep predictions of long-memory time series. A motivation for adopting the innovations algorithm is 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:

Thus, let $L \geq 1$, denote a large integer. The question of how to choose the value of L is further discussed below and it is associated with the use of the AIC criterion for approximating the fitted model by an approprite finite order autoregression. For $j = 0, 1, \dots, L$, let

$$\hat{R}_y(j) = \int_{-\pi}^{\pi} \hat{f}_y(\lambda) \exp(ij\lambda) d\lambda, \tag{8.1}$$

and

$$\hat{r}(j) = \frac{\hat{R}_y(j)}{\hat{R}_y(0)}, \tag{8.2}$$

denote the covariance and correlation functions of the fitted model, respectively. As is well-known, the coefficients, $\hat{\pi}_{Ps}(j)$ $(j = 1, \dots, s;\ s = 1, \dots, L)$ 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 $\hat{\pi}_{Ps}(0) = 1$:

$$\sum_{j=0}^{s} \hat{\pi}_{Ps}(j) \hat{r}_y(u-j) = 0, \quad (u = 1, \dots, s), \tag{8.3}$$

$$\sum_{j=0}^{s} \hat{\pi}_{Ps}(j) \hat{R}_y(j) = \hat{\sigma}^2(s). \tag{8.4}$$

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:

$$AIC(s) = T \ln \hat{\sigma}^2(s) + 2s \quad (s = 0, 1, \dots, L), \tag{8.5}$$

and thus k is such that $AIC(s)$ takes its smallest value for $s = k$. Having chosen the value of k, the *Type-II Plug-In* forecasts of the unknown future values are constructed as in (3.10) but with the $a(j)$, replaced by the $\hat{\pi}_{Ps}(j)$, the second sum to the right of this equation truncated at k and the past values, $\{x_{n+1-j},\ j=1,2,\ldots\}$, of $\{x_t\}$ replaced by the relevant past values of $\{y_t\}$; the details are omitted in the interest of space.

In the *Type-II Direct Method*, as discussed in Section 1, a new autoregressive approximation, that is, the h-step prediction constants and the order of the model to be fitted, is determined 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 (8.3) 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, $\hat{\pi}_{Dhs}(j)$ $(j=1,\ldots,s;\ s=1,\ldots,L)$ are the solutions of the following equations:

$$\sum_{j=1}^{s} \hat{\pi}_{Dhs}(j)\hat{r}_y(u-j) = -\hat{r}_y(h+u-1), \quad (u=1,\ldots,s). \tag{8.6}$$

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

$$\sum_{j=1}^{s} \hat{\pi}_{Dhs}(j)\hat{R}_y(h+j-1) + \hat{R}_y(0) = \hat{V}_{Dh}(s). \tag{8.7}$$

For $h=1$, equations (8.6) and (8.7) clearly reduce to equations (8.3) and (8.4); however, for $h > 1$ the corresponding h-step direct prediction constants obtained as a solution of (8.5) would differ from the corresponding h-step plug-in prediction constants.

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 (1996):

$$AIC(h,s) = T\ln\hat{V}_{Dh}(s) + 2s \quad (s=0,1,\ldots,L). \tag{8.8}$$

Thus, an autoregression of order $k=\tilde{k}_h$ will be fitted, if the criterion (8.8) attains its smallest value for $s=k$. The corresponding h-step predictor is then computed separately from (3.11), but with the $a(j)$, replaced by the $\hat{\pi}_{Ps}(j)$, the second sum to the right of this equation truncated at k, the past values, $\{x_{n+1-j},\ j=1,\ldots\}$, of $\{x_t\}$ replaced by the relevant past values of $\{y_t\}$ and ϵ_{n+h} set identically equal to zero; the details are omitted in the interest of space.

Bhansali and Kokoszka (2001b) have investigated the numerical accuracy of an approximation for $\hat{R}_y(j)$ obtained by replacing the integral occurring on the right of (8.1) by the trapezoidal sum and evaluated using a fast Fourier transform algorithm. Bhansali and Kokoszka (2001b) also compare the relative predictive efficacy of the Truncation, $Type-IIPlug-In$ and $Type-IIDirect$ 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.

9 Simulation Results

The relative predictive performance of the *Type-I* and *Type-II* approaches was compared on simulated time series. The following four $ARFIMA(p,d,q)$ models were simulated, with the values of p, q and the type of $ARMA$ model specified for the short-memory process $x_t = (1-B)^d y_t$ described below:

Model 1. $ARFIMA(0, 0.3, 0)$, $x_t = \epsilon_t$.

Model 2. $ARFIMA(1, 0.3, 0)$, $x_t = 0.5x_{t-1} + \epsilon_t$.

Model 3. $ARFIMA(0, 0.3, 1)$, $x_t = \epsilon_t - 0.5\epsilon_{t-1}$.

Model 4. $ARFIMA(1, 0.3, 1)$, $x_t + 0.5x_{t-1} = 0.5\epsilon_{t-1} + \epsilon_t$.

We also simulated $FEXP$ models with spectral density function given by (6.4) and the following parameters:

Model 5. $\sigma_2 = 1,\ d = 0.4,\ n = 1,\ c(1) = 1$.

Model 6. $\sigma_2 = 1,\ d = 0.4,\ n = 1,\ c(1) = 0.5$.

Time series of length $N = 1000$ from each of these six models were generated, with the innovations process, ϵ_t, simulated as a sequence of standard Normal deviates. The number of time series, NRP, say, generated from each model was 10,000.

For each simulated long-memory time series, $y_1, \ldots, y_N$, say, the multistep forecasts up to $H = 15$ steps given by the *Type-I* and *Type-II* approaches were evaluated. The *Type-I* approach was implemented by fitting autoregressive models as discussed in Section 5. The multistep forecasts according to the $Type-IPlug-In$ method and the $Type-IDirect$ methods were then computed. The specific algorithms used for applying these two methods are described below for convenience and clarity:

Type-I Plug-In Method.

$PI.1$). For $u = 0, \pm 1, \ldots, \pm N-1$, compute the standard *positive-definite* estimators of the autocovariance and autocorrelation functions of $\{y_t\}$ by

$$R^{(N)}(u) = \frac{1}{N}\sum_{t=1}^{N-|u|} y_t y_{t+|u|}, \tag{9.1}$$

$$r^{(N)}(u) = \frac{R^{(N)}(u)}{R^{(N)}(0)}. \tag{9.2}$$

$PI.2$). The sth order *Type-I Plug-In* estimates, $\pi^{(NPs)}(j),\ j = 1, ..., s$ and $\sigma^{(NP)}(s)^2$ of the prediction constants and residual error variance are then solutions of equations (8.3) and (8.4), but with the $\hat{r}_y(u)$ and $\hat{R}_y(0)$ replaced by $r^{(N)}(u)$ and $R^{(N)}(0)$, respectively.

$PI.3$). Repeat the calculations described at *Step PI.2)* above for $s = 0, 1, \ldots, L$, where we set $L = 20$, and choose the order, k, of the autoregression to fit by minimising the AIC criterion defined by equation (8.5), but with $\sigma^{(NP)}(s)^2$ replacing $\hat{\sigma}^2(s)$.

$PI.4$). For the value of k selected at *Step PI.3)* above, compute the h-step *Type-I Plug-In* multistep forecasts by equation (3.10), but with the $\pi^{(NPk)}(j)$ replacing the $a(j)$, the second sum to its right truncated at k and the past values $\{x_{n+1-j},\ j \geq 1\}$ replaced by the corresponding past values of the simulated long-memory series, $\{y_t\}$.

Type-I Direct Method.

$DI.1$). Compute $R^{(N)}(u)$ and $r^{(N)}(u), u = 0, \pm 1, \ldots, \pm N-1$, as in *Step PI.1)* of the *Type-I Plug-In method* above by equations (9.1) and (9.2), respectively.

$DI.2$). For a fixed $h > 1$, compute the corresponding h-step *Type-I Direct* estimates, $\pi^{(NDhs)}(j), j = 1, .., s$ and $V^{(NDh)}(s)$, of order s by (8.6) and (8.7), but with the $r^{(N)}(u)$ and $R^{(N)}(0)$ replacing the $\hat{r}_y(u)$ and $\hat{R}_y(0)$, respectively.

$DI.3$). Repeat the calculations described at *Step DI.2* above for $s = 0, 1, \ldots, L$, with $L = 20$, and choose the order, $k = k_h$, say, of the autoregression to fit for $h-step$ prediction by minimising the $AIC(h, s)$ criterion defined by equation (8.8), but with $V^{(NDh)}(s)$ replacing $\hat{V}_{Dh}(s)$.

$DI.4$). For the value of k selected at *Step DI.3* above and if $h \leq k$, compute the h-step *Type-I Direct* forecast by equation (3.11), but with $\epsilon_{n+h} = 0$, the $\pi^{(NDhk)}(j)$ replacing the $a(j)$, the second sum to its right truncated at k and the past values $\{x_{n+1-j},\ j \geq 1\}$ replaced by the corresponding past values of the simulated long-memory series, $\{y_t\}$. If $h > k$, then the first sum to the right of (3.11) is truncated at k and the second sum is set identically equal to zero, but the basic procedure for computing the *Type-I Direct h-step* forecast from equation (3.11) described above does not change.

$DI.5$). Increase the value of h by setting $h = h + 1$ and repeat the procedure described in *Steps DI.2) - DI.4* above until $h = H$, when the procedure ends.

Type-II Plug-In Method. The main difference now as compared with the *Type-I Plug-In* method described above is that the 'parametric' estimates, $\hat{R}_y(j)$ and $\hat{r}_y(j)$, defined by (8.1) and (8.2), of the autocovariance and autocorrelation functions of $\{y_t\}$ are used in place of the positive-definite estimators, (9.1) and (9.2). Hence, the step $PI.1$ of the *Type-I Plug-In Method* described above is replaced by the step $PII.1$ described below while the steps $PI.2) - PI.4)$ are modified accordingly:

$PII.1$). a) For the simulated $ARFIMA(p, d, q)$, or $FEXP$, model defined above, estimate the parameters of an $ARFIMA(1, d, 1)$ model using the approximate maximum likelihood estimators provided by the *S-Plus* function `arima.fracdiff`.

b) Obtain the corresponding estimate, $\hat{f}_y(\lambda)$, of the spectral density function, $f_y(\lambda)$, of the generated model from (6.3), with $p = q = 1$, by replacing the unknown model parameters by their estimates constructed as in a) above.

c) Evaluate $\hat{R}_{yQ}(j)$ as in (8.1), but by replacing the integral occurring on its right by a trapezoidal sum, that is, by forming a discrete Fourier transform of $\hat{f}_y(\lambda)$ as follows:

$$\hat{R}_{yQ}(u) = \frac{2\pi}{Q} \sum_{j=0}^{Q-1} \hat{f}_y(\lambda_j) \exp(iu\lambda_j), \quad u = 0, 1, \ldots, Q-1. \tag{9.3}$$

The corresponding estimate of $r_y(u)$ is then given by

$$\hat{r}_{yQ}(u) = \frac{\hat{R}_{yQ}(u)}{\hat{R}_{yQ}(0)}. \tag{9.4}$$

As discussed by Bhansali and Kokoszka (2001b), a main difficulty in using the approximation (9.3) is that $f(\lambda)$ diverges to infinity as $\lambda \to 0$, and hence there is a certain amount of arbitrariness in assigning a value for $f(0)$. We set $f(0) = 0$, and, in effect, form only the discrete Fourier transform of the sequence $\{f(\lambda_1), \ldots, f(\lambda_{Q-1})\}$. Moreover, we put $Q = 2048$.

$PII.2)$. The sth order *Type-I Plug-In* estimates, $\hat{\pi}_Q(j),\ j = 1, ..., s$ and $\hat{\sigma}_Q(s)^2$ of the prediction constants and residual error variance are then solutions of equations (8.3) and (8.4), but with the $\hat{r}_y(u)$ and $\hat{R}_y(0)$ replaced by $\hat{r}_{yQ}(u)$ and $\hat{R}_{yQ}(0)$, respectively.

$PII.3)$. Repeat the calculations described at *Step PII.2)* above for $s = 0, 1, \ldots, L$, with $L = 20$, and choose the order, k, of the autoregression to fit by minimising the AIC criterion defined by equation (8.5), but with $\hat{\sigma}_Q(s)^2$ replacing $\hat{\sigma}^2(s)$.

$PII.4)$. For the value of k selected at *Step PII.3)* above, compute the h-step Type-II Plug-In multistep forecasts by equation (3.10), but with the $\hat{\pi}_Q(j)$ replacing the $a(j)$, the second sum to its right truncated at k and the past values $\{x_{n+1-j},\ j \geq 1\}$ replaced by the corresponding past values of the simulated long-memory series, $\{y_t\}$.

Type-II Direct Method.

$DII.1)$. Compute $\hat{R}_{yQ}(u)$ and $\hat{r}_{yQ}(u)(u)$, as described in *Step PII.1)* above by equations (9.3) and (9.4), respectively, with $Q = 2048$ and $f(0) = 0$.

$DII.2)$. For a fixed $h > 1$, compute the corresponding h-step sth order *Type-II Direct* estimates, $\hat{\pi}_{DhQs}(j), j = 1, .., s$ and $\hat{V}_{DhQ}(s)$ of order s by (8.6) and (8.7), but with the $\hat{r}_{yQ}(u)$ and $\hat{R}_{yQ}(0)$ replacing the $\hat{r}_y(u)$ and $\hat{R}_y(0)$, respectively.

$DII.3)$. Repeat the calculations described at *Step DII.2)* above for $s = 0, 1, \ldots, L$, with $L = 20$, and choose the order, $k = k_h$, say, of the autoregression to fit for $h - step$ prediction by minimising the $AIC(h, s)$ criterion defined by equation (8.8), but with $\hat{V}_{DhQ}(s)$ replacing $\hat{V}_{Dh}(s)$.

$DII.4)$. For the value of k selected at *Step DII.3* above and if $h \leq k$, compute the *Type-II Direct h-step* forecast by equation (3.11), but with $\epsilon_{n+h} = 0$, the $\hat{\pi}_{DhQk}(j)$ replacing the $a(j)$, the second sum to its right truncated at k and the past values $\{x_{n+1-j},\ j \geq 1\}$ replaced by the corresponding past values of the simulated long-memory series, $\{y_t\}$. If $h > k$, then the first sum to the right of (3.11) is truncated at k and the second sum is set identically equal to zero, but the basic procedure for computing the *Type-II Direct h-step* forecast from equation (3.11) described above does not change.

$DII.5)$. Increase the value of h by setting $h = h + 1$ and repeat *Steps DII.2) - DII.4)* above until $h = H$, when the procedure ends.

We compare the efficacy of the following methods for multistep prediction of long-memory time series:

1) $Type - IPlug - In$ method v/s $Type - IIPlug - In$ method;

2) $Type - IDirect$ method v/s $Type - IIDirect$ method;

3) $Type - IPlug - In$ method v/s $Type - IDirect$ method.

The multistep predictions by the $Type{-}II$ method are constructed from an $ARFIMA(1, d, 1)$ model fitted to the simulated time series. For Model 4 this corresponds to assuming that the stochastic model generating the simulated long-memory time series is known, and hence, as discussed in Section 7, the maximum likelihood estimates of the parameters of the fitted model are, asymptotically Normally distributed and Fisher efficient. Moreover, for Models 2 and 3 this assumption still corresponds to fitting the simulated model but with just one additional redundant parameter, namely setting $(p = 1, q = 1)$, for both models, instead of the actual $(p = 1, q = 0)$ for Model 2 and $(p = 0, q = 1)$ for Model 3, and, for these two models, also the maximum likelihood estimators have the same optimality property as described above when the correct model is fitted. Thus, for these three models, apart from the possible effects of approximating the integral occurring in (8.1) by a finite sum, the *Type-II* methods may be expected to outperform the *Type-I* methods, especially because the latter do not utilise the information about the stochastic model generating the simulated time series. In the light of this in-built advantage of the *Type-II* methods for Models 2, 3, and 4, an objective of the simulation study was to examine the extent of potential disadvantage, if any, the *Type-I* methods may suffer relative to the *Type-II* methods in this situation. Also, for Model 1, an $ARFIMA(0, d, 0)$ process is simulated, but an $ARFIMA(1, d, 1)$ model is fitted. Thus, a second objective of the simulation study was to obtain an indication of how the *Type-II* methods perform when the fitted values of both p and q are greater than their actual values. Models 5 and 6 both follow an $FEXP$ model and there is no true value of the $ARMA$ order for these two models; furthermore, the simulated value of the long-memory $d = 0.4$ is much greater for these two models than that for Models 1 - 4. Therefore, a third objective of the simulation study was to compare the relative behaviour of the methods when an $ARFIMA$ model is fitted as a rational transfer function approximation to an observed long-memory time series and whose generating process does not actually follow the fitted $ARFIMA$ model. A final objective of the simulation study was to compare the relative behaviour of the *Type-I Plug-In* and *Type-I Direct* methods of fitting autoregressions for multistep prediction of long- memory time series.

We use the criterion of the proportionate change in the cumulative simulated mean squared errors of prediction up to H steps ahead, $H = 1, 2, \ldots, 15$, for making the comparisons discussed above. Accordingly, a definition of how this measure was computed is given in generic terms for two different methods of multistep prediction, called Methods A and B respectively:

For a given simulated model and each $h = 1, \ldots, H$, let $\hat{y}_{jNA}(h)$ denote the h-step forecast provided by a Method A for the jth simulated time series from that model, $j = 1, \ldots, NRP$, and let $\hat{y}_{jNB}(h)$ denote the corresponding h-step forecast according to a Method B. The simulated h-step mean squared error of prediction, $SMSE_A(h)$, for the Method A is given by

$$(9.5) \qquad SMSE_A(h) = (NRP)^{-1} \sum_{j=1}^{NRP} \{\hat{y}_{jNA}(h) - y_{j,N+h}\}^2,$$

where $y_{j,N+h}$ denotes the actual value of y_{N+h} at the jth simulation of the relevant model, and the corresponding cumulative simulated mean squared error of H-step prediction for Method A is given by:

$$(9.6) \qquad CSMSE_A(H) = \sum_{h=1}^{H} SMSE_A(h) \qquad (H = 1, 2, \ldots, 15).$$

The cumulative simulated mean squared errors of prediction, $CSMSE_B(H)$ for Method B is defined analogously to (9.6) but with $\hat{y}_{jNA}(h)$ in (9.5) replaced by the corresponding forecast, $\hat{y}_{jNB}(h)$ given by the Method B.

The proportionate change in the cumulative simulated mean squared error of prediction for Method A relative to Method B is then given by

$$(9.7) \qquad PCSMSE(H) = \frac{\{CSMSE_A(H) - CSMSE_B(H)\}}{CSMSE_B(H)},$$

and it measures the change in the cumulative simulated mean squared error of prediction when using Method A in preference to the Method B, as a proportion of the cumulative simulated mean squared error of prediction of the Method B; for convenience, the dependence of the left hand side of (9.7) on the Methods A and B is suppressed. If, for a given value of H, $PCSMSE(H)$ is positive then this signals that the Method B has an advantage over Method A in that the latter has a larger cumulative simulated mean squared error of prediction up to H steps ahead than the former method, and, conversely, a negative value of $PCSMSE(H)$ indicates that the Method A has an advantage over Method B in terms of their cumulative simulated mean squared errors of prediction up to H steps ahead.

Consider first the *Type-I* and *Type-II* methods. For *Models 1 - 6*, Figure 9.1 shows the graphs of $PCSMSE(H)$ of the *Type-II Plug-In* method relative to the *Type-I Plug-In* method, and, also, of the *Type-II Direct* method relative to the *Type-I Direct* Method. These figures enable the comparisons 1) and 2) listed above to be made. It may be gleaned that for *Models 1 and 3*, the *Type-I Plug-In* method has an advantage over the *Type-II Plug-In* method in that the graph of $PCSMSE(H)$ is positive at all lead times for these two models. Moreover, a somewhat similar finding holds also for *Model 4*, where, except for $H=3$ and 5, $PCSMSE(H)$ is positive at all other lead times. Thus for these models, it is not clear that the estimation of d and the relevant short-memory parameters is necessarily beneficial in improving the multistep forecasts. However, for the other three models, and especially for *Models 5 and 6*, the two *non-ARFIMA* models, the corresponding values of $PCSMSE(H)$ remain negative for all large values of H, and for these models the *Type-II Plug-In* method is seen to have an advantage over the *Type-I Plug-In* method. The simulation results, thus, do not unequivocally point to the superiority of one method over the other. This finding is rather surprising especially because, as discussed above, for *Model 4*, the latter method has an in-built advantage over the former method for multistep prediction. A possible explanation may be that the simulated value of the long-memory parameter for this model, namely $d = 0.3$, is on the low side and it is possible to approximate its correlation structure by an appropriate short-memory process.

Next, consider the comparison 2), namely that between the *Type-I Direct* and *Type-II Direct* methods. The simulation results here broadly appear to favour the *Type-II Direct* method, since for all models, except *Model 4*, the graph of $PCSMSE(H)$ takes negative values at higher lead times. This finding is not surprising and supports the use of the *Direct* method for constructing multistep forecasts discussed earlier.

Finally consider the comparison between the *Type-I Plug-In* and *Type-I Direct* methods. Figure 9.2 shows the values of $PCSMSE(H)$ for making this comparison. The simulation results by and large support the use of *Type-I Direct method*: for all 6 models, the graph of $PCSMSE(H)$ nearly always takes a positive value. On the other hand, however, the numerical values of the difference is not large and thus the loss in predictive efficiency due to using the *Type-I Plug-In* method in preference to the *Type-I Direct* method may not necessarily be great in this situation; a similar finding has earlier been also reported by Bhansali (1996).

References

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.

Bary, N. K. (1964). *A Treatise on Trigonometric Series.* Pergamon Press, Oxford.

Baxter, G. (1962). An asymptotic result for the finite predictor. *Math. Scand.*, **10,** 137–144.

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. (1996). Asymptotically efficient autoregressive model selection for multistep prediction. *Ann. Inst. Statist. Math.*, **48,** 577–602.

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. (2001). Multi-step forecasting. In *Companion to Economic Forecasting* (eds M. P. Clements and D. Hendry). Blackwell. forthcoming.

Bhansali, R. J. and Kokoszka, P.S. (2001a). Estimation of the long-memory parameter: a review and an extension. submitted.

Bhansali, R. J. and Kokoszka, P.S. (2001b). Computation of the forecast coefficients for multistep prediction of long-range dependent time series. submitted.

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.

Cox, D. R. (1991). Long-range dependence, non-linearity and time irreversibility. *J. Time Series Analysis*, **12,** 329–335.

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.

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, **14,** 517–532.

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.

Giraitis, L. and Surgailis, D. (1990). CLT for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle's estimate. *Prob. Th. Rel. Fields*, **86**.

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. forthcoming in *Journal of Time Series Analysis.*

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

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.

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.

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

Robinson, P. M. (1995). Log–periodogram regression of time series with long range dependence. *The Annals of Statistics*, **23,** 1048–1072.

Robinson, P. M. (1995a). Log–periodogram regression of time series with long range dependence. *The Annals of Statistics*, **23,** 1048–1072.

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

Rudin, W. (1987). *Real and Complex Analysis*, International edn. McGraw-Hill, Singapore.

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. Feldman R. Adler and M. S. Taqqu), pp. 000–000. 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.

Tolstov, G. P. (1976). *Fourier Series.* Dover, New York.

Whittle, P. (1963). *Prediction and Regulation by Linear Least–Squares Methods.* English University Press, London.

Zygmund, A. (1959). *Trigonometric Series*, volume I, II. Cambridge University Press.

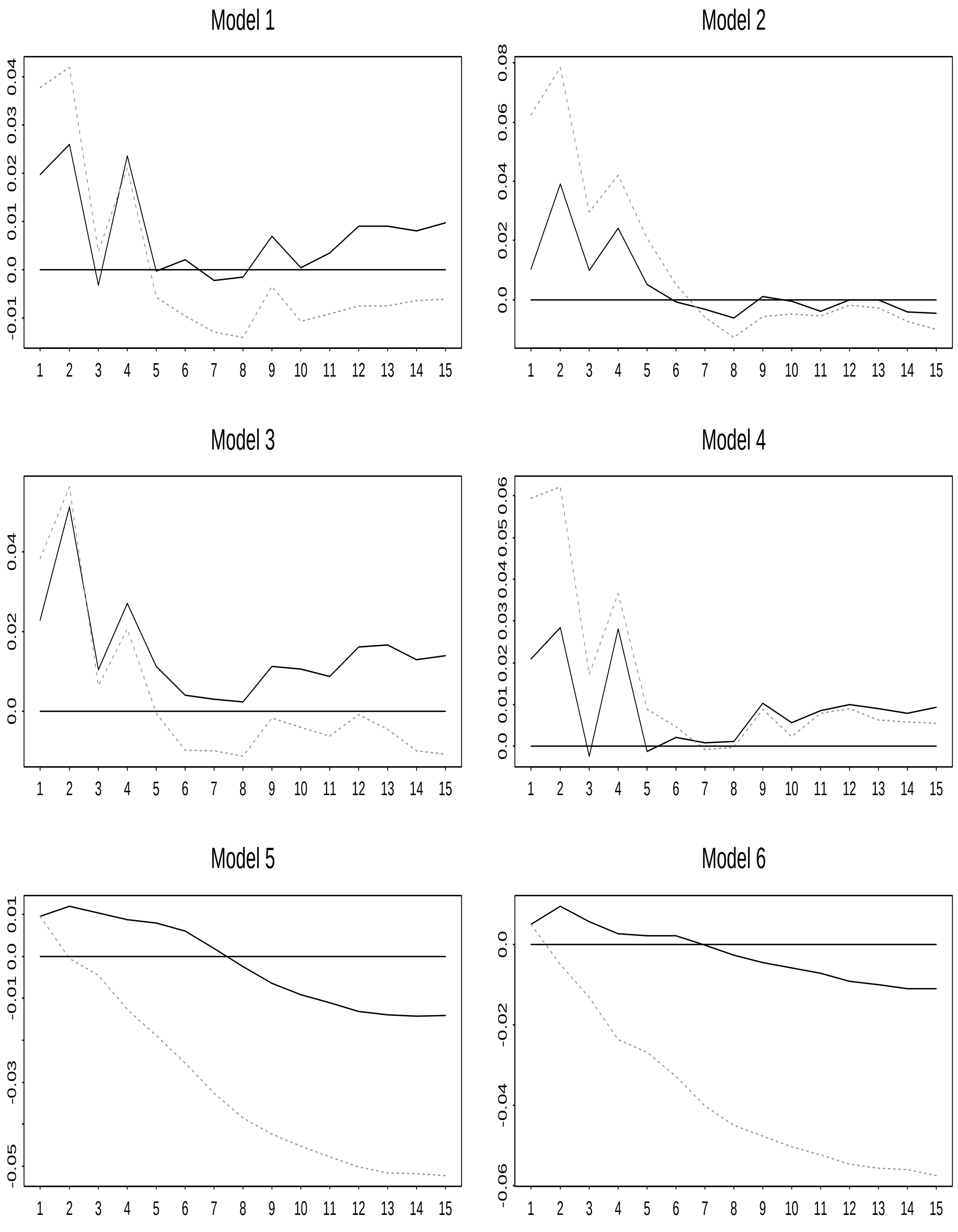

FIGURE 9.1. PCSMSE(h) for the Type II method relative to Type I method. The continuous line corresponds to Plug-In methods and the dotted tine to direct methods.

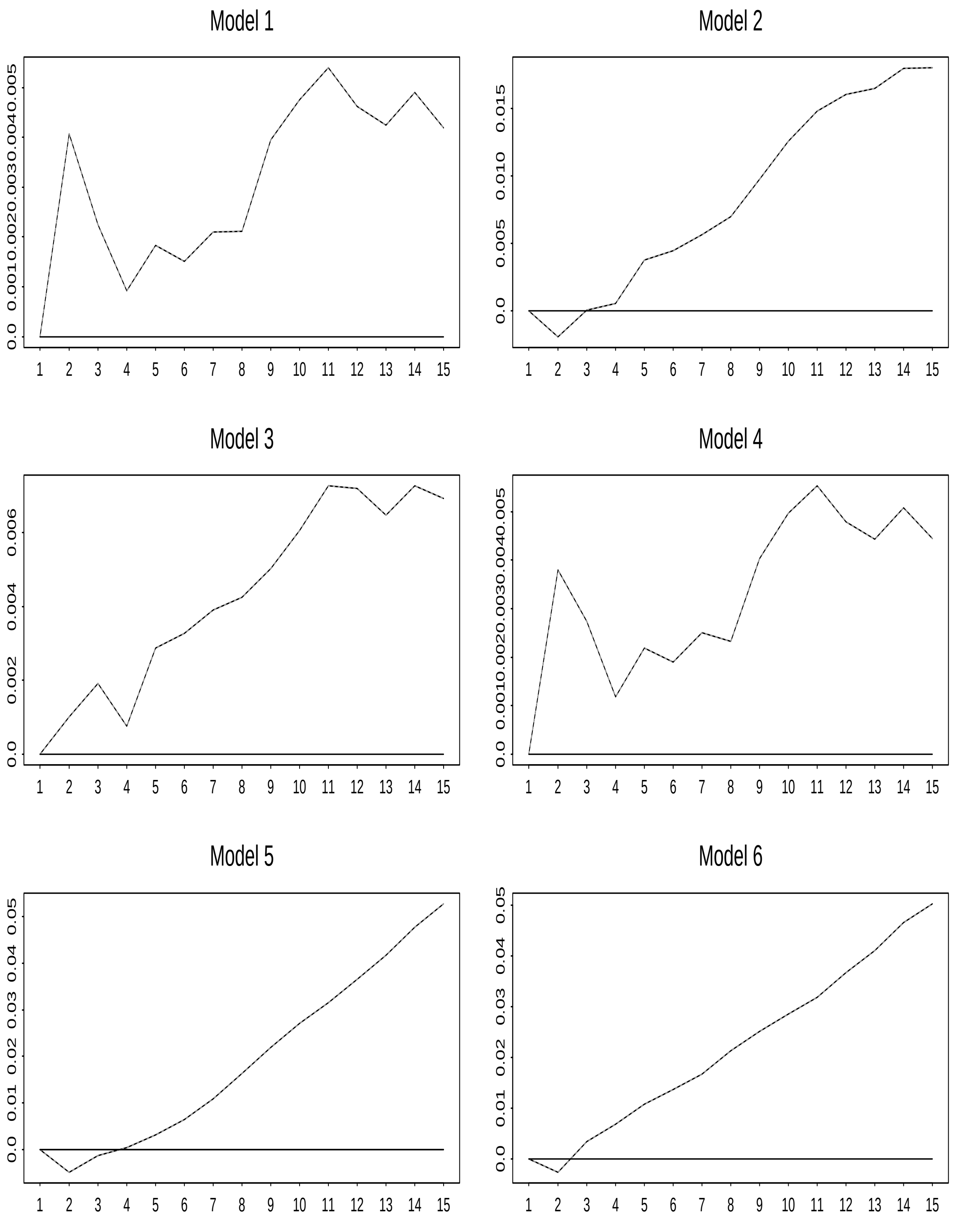

FIGURE 9.2. PCSMSE(h) for the Type I Plug-In relative to type I direct method.