Journal of Forecasting
J. Forecast. **20**, 367–389 (2001)
DOI: 10.1002/for.799

The Approximation of Long-memory Processes by an ARMA Model

GOPAL K. BASAK[1], NGAI HANG CHAN[2] AND WILFREDO PALMA[3]*
[1] *University of Bristol, UK*
[2] *Chinese University of Hong Kong, Hong Kong*
[3] *P. Universidad Católica de Chile, Chile*

ABSTRACT
A mean square error criterion is proposed in this paper to provide a systematic approach to approximate a long-memory time series by a short-memory ARMA(1, 1) process. Analytic expressions are derived to assess the effect of such an approximation. These results are established not only for the pure fractional noise case, but also for a general autoregressive fractional moving average long-memory time series. Performances of the ARMA(1,1) approximation as compared to using an ARFIMA model are illustrated by both computations and an application to the Nile river series. Results derived in this paper shed light on the forecasting issue of a long-memory process. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS ARMA(1,1); forecast error; long memory

INTRODUCTION

In a seminal paper, Hosking (1984) observes that a long-memory process can be approximated by an ARMA(1,1) process reasonably well when the approximating ARMA process has both roots close to the unit circle. Although no rigorous justification of this assertion is given in his paper, simulation studies conducted in Hosking (1984) indicate the validity of this assertion. Since then, the idea of approximating a long-memory process by a short-memory time series has been receiving considerable attention in the literature.

One of the main reasons for the continued interest in this problem is its practical implications. Although long-memory processes are widely applicable in econometrics and other fields (see, for example, Baillie, 1996), actual implementation of these models often requires intricate approximated likelihood procedures (see, for example, Sowell, 1992; Beran, 1994; Chan and Palma, 1998). When one wants to use a long-memory model to forecast future values, many of the traditional time series

* Correspondence to: W. Palma, Department of Statistics, P. Universidad Católica de Chile, Casilla 306, Santiago 22, Chile.
Contract/grant sponsor: National Security Agency; Contract/grant number: 006-124.
Contract/grant sponsor: Hong Kong Research Grants Council; Contract/grant number: CUHK 6082/98P.
Contract/grant sponsor: Fondecyt; Contract/grant number: 8000004/7980021.

Received June 1999
Accepted September 2000

techniques are no longer easily applicable. Therefore, if one can identify situations where long-memory processes can be reasonably approximated by certain short-memory ARMA models, then one can use these approximating models to perform forecasting and the quality of these forecasts can be assessed by usual ARMA techniques. One important advantage of ARMA models is the existence of efficient algorithms for calculating estimates and one-step predictions. For example, maximum likelihood calculations for ARMA processes are $O(n)$ whereas for ARFIMA processes, they are $O(n^2)$ (see, for example, Chan and Palma, 1998).

Motivated by a forecasting consideration given in Tiao and Xu (1993), Tiao and Tsay (1994) proposed an adaptive scheme to approximate certain long-memory processes by an ARMA(1,1) time series. By minimizing the l-step-ahead forecast error variance, they propose a procedure which estimates the parameters of the approximating ARMA(1,1) model adaptively. Their simulations show that the variance of this l-step-ahead forecast lies within a 5% margin with the optimal forecast based on an actual long-memory model. Further discussions on adaptive procedures are given in Tong (1997).

There are two objectives in this paper. The first studies the question of when an ARMA(1,1) model can be used to approximate a long-memory ARFIMA model adequately. This objective is addressed by comparing the best ARMA(1,1) forecast with the best ARFIMA(p,d,q) forecast using the mean square criterion. The second objective of this paper is to study when the gain from using an adaptive forecast scheme rather than a non-adaptive scheme is large. This goal is achieved by characterizing the relationships between the parameters ϕ and θ of the ARMA(1,1) model and the long-memory parameter d of the underlying long-memory process. This systematic characterization not only provides an explanation of why the adaptive scheme of Tiao and Tsay (1994) works, but also sheds light on the forecasting properties of a long-memory model based on an approximated short-memory model as studied in Brodsky and Hurvich (1999).

This paper is organized as follows. In the next section the approximation of a simple long-memory model, i.e. a fractional noise ARFIMA(0, d, 0) model, by an ARMA(1,1) process is first studied. Characterizations of the relationships between the ARMA(1,1) parameters and the long-memory parameter are given. The case of approximating a general ARFIMA(p, d, q) model is studied in the third section. Applications of these methodologies to the Nile river data are presented in the fourth section while conclusions are given in the final section.

APPROXIMATION OF FRACTIONAL NOISE

In this section we study in detail the mean square error of the l-step-ahead forecast and use it to characterize the relationship between the ARMA(1,1) parameters and the long-memory parameter d for a fractional noise model. Specifically, let $\{X_t\}$ be a fractional noise process satisfying

$$X_t = (1-B)^{-d}\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \tag{1}$$

where B is the backshift operator $BX_t = X_{t-1}$, $d \in (-0.5, 0.5)$, $\{\varepsilon_t\}$ is a sequence of independent standard normal random variables, and the coefficients $\{\psi_j\}$ are given by

$$\psi_j = \begin{cases} d(d+1)\cdots(d+j-1)/j!, & j \geq 1 \\ 1 & j = 0 \end{cases}$$

Some authors distinguish the behaviour of $\{X_t\}$ for different values of $d \in (-0.5, 0.5)$. It is well known (see, for example, Hosking, 1981; Samorodnitsky and Taqqu, 1994; or Brockwell and Davis, 1991) that for $d \in (-0.5, 0)$, the process is negative dependent and it is often referred to as an intermediate memory process since its autocorrelation function $\rho(k)$ is always negative, of order k^{2d-1} as $k \to \infty$, and $\sum_k |\rho(k)| < \infty$. On the other hand, for $d \in (0, 0.5)$, the process is long-memory since $\rho(k) = O(k^{2d-1})$ as $k \to \infty$ and $\sum_k \rho(k) = \infty$. For $d = 0$, the process $\{X_t\}$ is equal to a white-noise process. In this paper, $\{X_t\}$ will simply be referred to as a long-memory process as long as $d \in (-0.5, 0.5)$.

Let $\{Y_t\}$ be an ARMA(1,1) process satisfying

$$Y_t = \frac{1-\theta B}{1-\phi B}\varepsilon_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \tag{2}$$

where ϕ and θ are parameters lying in $(-1, 1)$, and $a_j = (\phi - \theta)\phi^{j-1}$, for $j \geq 1$ with $a_0 = 1$. Let $\widetilde{Y}_t(l)$ be the l-step-ahead prediction of Y_t based on the history of the process $\{Y_i : i \leq t\}$ given by equation (2). Then $\widetilde{Y}_t(l) = \sum_{j=l}^{\infty} a_j \varepsilon_{t+l-j}$. Now consider the problem of forecasting a future value of the long-memory process $\{X_t\}$ in (1) by the ARMA(1,1) process $\{Y_t\}$ defined in (2). A useful way to forecast $\{X_t\}$ in this context is the adaptive scheme given in Tiao and Tsay (1994) which selects an ARMA(1,1) model $\{Y_t\}$ that minimizes the l-step-ahead forecast error adaptively.

In order to gain a better understanding of the forecasting issue and the associated forecasting error, we first define the following notions. Let $G_l(\phi, \theta) = E(X_{t+l} - \widetilde{Y}_t(l))^2$ be defined as the mean square error of the l-step-ahead prediction based on a selected ARMA(1,1) model $\{Y_t\}$. Then, for $l \geq 1$

$$\begin{aligned} G_l(\phi, \theta) &= E(X_{t+l})^2 + E(\widetilde{Y}_t(l))^2 - 2E(X_t\widetilde{Y}_t(l)) \\ &= \sum_{j=0}^{\infty} \psi_j^2 + \sum_{j=l}^{\infty} a_j^2 - 2\sum_{j=l}^{\infty} \psi_j a_j \\ &= \sum_{j=0}^{\infty} \psi_j^2 + (\phi-\theta)^2 \sum_{j=l}^{\infty} \phi^{2(j-1)} - 2(\phi-\theta)\sum_{j=l}^{\infty} \psi_j \phi^{j-1} \\ &= \frac{\Gamma(1-2d)}{\Gamma(1-d)^2} + \frac{(\phi-\theta)^2\phi^{2(l-1)}}{1-\phi^2} - 2(\phi-\theta)p_l(\phi) \end{aligned} \tag{3}$$

where $p_l(\phi) = \sum_{j=l}^{\infty} \psi_j \phi^{j-l}$ for all $\phi \in (-1, 1)$. The quantity $G_l(\phi, \theta)$ defined in equation (3) plays an important role in understanding the relationships between the two processes $\{X_t\}$ and $\{Y_t\}$. Note that when (ϕ, θ) is chosen to minimize $G_l(\phi, \theta)$ as a function of l, it corresponds to the adaptive scheme proposed in Tiao and Tsay (1994). In order to see how large is the gain by using an adaptive scheme, we first characterize the relationship between the ARMA parameters (ϕ, θ) and the long-memory parameter d for different forecasting horizon l when an adaptive scheme is used. This is achieved by studying the underlying behaviour of $G_l(\phi, \theta)$. To this end, we first establish the following lemma.

Lemma 1 For $|\phi| < 1$ and $l \geq 1$,

$$(p_l(\phi))^2 \leq (p_{l-1}(\phi))^2 \text{for } l \geq 2 \tag{4}$$

Proof. Note that $p_{l-1}(\phi) = \psi_{l-1} + \phi p_l(\phi)$ and $\psi_j = d(d+1)\cdots(d+j-1)/j!$ The proof of this lemma can be divided into four cases according to the values of ϕ and d.

Case 1. $0 \le \phi < 1$ and $0 < d < \frac{1}{2}$. In this case, $\psi_j > 0$, $p_l(\phi) > 0$, and $p_{l-1}(\phi) > 0$. Proving equation (4) is equivalent to proving $p_l(\phi) \le p_{l-1}(\phi)$, i.e. $\sum_{j=l}^{\infty} \psi_j \phi^{j-l} \le \sum_{j=l-1}^{\infty} \psi_j \phi^{j-l+1}$, which in turn is equivalent to showing $\psi_j \le \psi_{j-1}$ for $j \ge 2$. Observe that

$$\psi_j = \frac{j-1+d}{j}\psi_{j-1}$$

Therefore,

$$\psi_j - \psi_{j-1} = \frac{d-1}{j}\psi_{j-1} \tag{5}$$

As the last quantity is negative for $0 < d < 1/2$, Case 1 is established.

Case 2. $0 \le \phi < 1$ and $-\frac{1}{2} < d < 0$. In this case, $\psi_j < 0$, $p_l(\phi) < 0$, and $p_{l-1}(\phi) < 0$. Therefore, in order to prove equation (4), it suffices to prove $p_l(\phi) \ge p_{l-1}(\phi)$, i.e, $\psi_j \ge \psi_{j-1}$. From equation (5),

$$\psi_j - \psi_{j-1} = [d(d+1)\cdots(d+j-2)/j!][d-1] > 0$$

as both $d < 0$ and $(d-1) < 0$, proving equation (4).

Case 3. $-1 < \phi < 0$ and $0 < d < \frac{1}{2}$. Although $0 < \psi_j < \psi_{j-1}$ in this case, determining the sign of $p_l(\phi)$ is more tricky. To this end, consider

$$\begin{aligned} p_l(\phi) &= \sum_{k=0}^{\infty} \psi_{2k+l}\phi^{2k} + \sum_{k=0}^{\infty} \psi_{2k+l+1}\phi^{2k+1} \\ &= \sum_{k=0}^{\infty} (\psi_{2k+l} + \phi\psi_{2k+l+1})\phi^{2k} \end{aligned} \tag{6}$$

Since $-1 < \phi < 0$ and $\psi_j > 0$ in this case,

$$\psi_{2k+l} + \phi\psi_{2k+l+1} > \psi_{2k+l} - \psi_{2k+l+1} > 0$$

Consequently, $p_l(\phi) > 0$. Similarly, the same argument shows that

$$p_{l-1}(\phi) = \sum_{k=0}^{\infty} \psi_{2k+l-1}\phi^{2k} + \sum_{k=0}^{\infty} \psi_{2k+l}\phi^{2k+1} > 0 \tag{7}$$

Therefore, to prove equation (4), it suffices to show $p_{l-1}(\phi) \ge p_l(\phi)$. Taking the difference of equations (6) and (7),

$$\begin{aligned} p_{l-1}(\phi) - p_l(\phi) &= \sum_{k=0}^{\infty} (\psi_{2k+l-1} - \psi_{2k+l})\phi^{2k} + \sum_{k=0}^{\infty} (\psi_{2k+l} - \psi_{2k+l+1})\phi^{2k+1} \\ &> \sum_{k=0}^{\infty} [(\psi_{2k+l-1} - \psi_{2k+l}) - (\psi_{2k+l} - \psi_{2k+l+1})]\phi^{2k} \end{aligned} \tag{8}$$

Since $-1 < \phi$ in this case, the last inequality of equation (8) follows from the two facts that $\psi_j > 0$ and $\psi_j \downarrow j$ for $d > 0$. Accordingly, $p_{l-1}(\phi) \geq p_l(\phi)$ if the summands in equation (8) are non-negative. From equation (5),

$$\psi_{2k+l-1} - \psi_{2k+l} = d(d+1)\cdots(d+2k+l-2)(1-d)/(2k+l)! \tag{9}$$

Replacing l by $l+1$ in this equation,

$$\psi_{2k+l} - \psi_{2k+l+1} = d(d+1)\cdots(d+2k_l-1)(1-d)/(2k+l+1)! \tag{10}$$

Therefore, taking the difference of equations (9) and (10),

$$\begin{aligned}(\psi_{2k+l-1} - \psi_{2k+l}) - (\psi_{2k+l} - \psi_{2k+l+1}) &= d(d+1)\cdots \\ &\times (d+2k+l-2)(1-d)(2-d)/(2k+l+1)! > 0\end{aligned} \tag{11}$$

as $0 < d < 1/2$. Hence, $p_{l-1}(\phi) \geq p_l(\phi)$, establishing equation (4).

Case 4. $-1 < \phi < 0$ and $-1/2 < d < 0$. In this case, $\psi_j < 0$ and $\psi_j > \psi_{j-1}$. By (6),

$$\begin{aligned}p_l(\phi) &= \sum_{k=0}^{\infty}(\psi_{2k+l} + \phi\psi_{2k+l+1})\phi^{2k} \\ &< \sum_{k=0}^{\infty}(\psi_{2k+l} - \psi_{2k+l+1})\phi^{2k} \\ &< 0\end{aligned}$$

as $\psi_{2k+l} < \psi_{2k+l+1}$ and $-1 < \phi < 0$. Similarly, $p_{l-1}(\phi) < 0$. Therefore, proving equation (4) is equivalent to proving $p_{l-1}(\phi) < p_l(\phi)$. From equations (8) and (11), it follows that

$$\begin{aligned}p_{l-1}(\phi) - p_l(\phi) &< \sum_{k=0}^{\infty}[d(d+1)\cdots(d+2k+l+2)(1-d)(2-d)/(2k+l+1)!]\phi^{2k} \\ &< 0,\end{aligned}$$

as $d < 0$. Hence, $p_{l-1}(\phi) < p_l(\phi)$ proving equation (4). Combining all these four cases, the proof of Lemma 1 is completed. □

With the aid of this lemma, we are now ready to prove the main result of this section. The next theorem states that as far as the mean square error criterion $G_l(\phi, \theta)$ is concerned, fixing an ARMA(1, 1) model $\{Y_t\}$ *a priori* and using it to predict X_{t+l} is always inferior to choosing a model which minimizes the l-step-ahead prediction error directly. Although intuitive, this result quantifies the fact that when a forecasted value $\widetilde{Y}_t(l)$ is used, we pay a price in the MSE. In addition, this result demonstrates the monotonicity of $G_l(\phi_l, \theta_l)$ which will be used to characterize the relationship between the parameters.

Theorem 1 Let $\{X_t\}$ be the long-memory process that follows equation (1) and $\{Y_t\}$ be an ARMA(1,1) process that follows equation (2). Then

$$\min_{\phi,\theta} E(X_t - Y_t)^2 = \min_{\phi,\theta} E(X_{t+l} - Y_{t+l})^2 \leq \min_{\phi,\theta} E(X_{t+l} - \widetilde{Y}_t(l))^2 \quad \text{for all } l \geq 1 \tag{12}$$

Proof. Observe that by definition, $E(X_{t+l} - Y_{t+l})^2 = E(X_t - Y_t)^2$ for any $l \geq 1$. Thus we need to show $\min_{\phi,\theta} E(X_t - Y_t)^2 \leq \min_{\phi,\theta} E(X_{t+l} - \widetilde{Y}_t(l))^2$ for all $l \geq 1$. Since, $E(X_t - Y_t)^2 = \sum_{j=0}^{\infty} \psi_j^2 + \sum_{j=0}^{\infty} a_j^2 - 2\sum_{j=0}^{\infty} \psi_j a_j = E(X_{t+1} - \widetilde{Y}_t(l))^2 - 1$, it is sufficient to show that $\min_{\phi,\theta} G_1(\phi, \theta) \leq \min_{\phi,\theta} G_l(\phi, \theta)$ for $l \geq 1$. In order to show the preceding inequality, we minimize $G_l(\phi, \theta) = K(\psi) - 2(\phi - \theta)\sum_{j=l}^{\infty} \psi_j \phi^{j-1} + (\phi - \theta)^2 \phi^{2(l-1)}/(1 - \phi^2)$, first with respect to θ and then with respect to ϕ. Here $K(\psi) = \sum_{j=0}^{\infty} \psi_j^2$. Differentiating with respect to θ, we have

$$\frac{\partial G_l(\phi, \theta)}{\partial \theta} = 2\sum_{j=l}^{\infty} \psi_j \phi^{j-1} - 2(\phi - \theta)\sum_{j=l}^{\infty} \phi^{2(j-1)}$$

Setting this equal to zero,

$$(\phi - \theta_l) = \frac{\sum_{j=l}^{\infty} \psi_j \phi^{j-1}}{\sum_{j=l}^{\infty} \phi^{2(j-1)}} = (1 - \phi^2)\frac{\sum_{j=l}^{\infty} \psi_j \phi^{j-1}}{\phi^{2(l-1)}}$$

where θ_l is the point of minimum for the $G_l(\phi, \theta)$ for every ϕ since $\partial^2 G_l(\phi, \theta)/\partial\theta^2 = 2\sum_{j=l}^{\infty} \phi^{2(j-1)} > 0$ for $\phi \neq 0$. Define,

$$f_l(\phi) = G_l(\phi, \theta_l) = K(\psi) - (1 - \phi^2)(p_l(\phi))^2$$

where $p_l(\phi)$ is defined as in Lemma 1. For $|\phi| < 1$ and $l \geq 2$, it follows from Lemma 1 that $p_{l-1}(\phi)^2 \geq p_l(\phi)^2$ and hence

$$f_l(\phi) - f_{l-1}(\phi) = (1 - \phi^2)(p_{l-1}(\phi)^2 - p_l(\phi)^2) \geq 0$$

Therefore, $f_l(\phi_l) \geq f_{l-1}(\phi_l) \geq f_{l-1}(\phi_{l-1})$, where ϕ_l is the value at which $f_l(\phi)$ attains its minimum. Repeating this inequality for each l, we conclude that $f_l(\phi_l) \geq f_1(\phi_1)$ which implies $G_l(\phi_l, \theta_l) \geq G_1(\phi_1, \theta_1)$. Therefore,

$$\min_{\phi,\theta} E(X_{t+l} - Y_{t+l})^2 \leq \min_{\phi,\theta} E(X_{t+1} - \widetilde{Y}_t(l))^2 \qquad \text{for all } l \geq 1.$$

This completes the proof of Theorem 1. □

Although Theorem 1 provides a way to forecast X_{t+l} by Y_{t+l}, it is not directly applicable in practice because we do not observe a future value Y_{t+l}. We have to estimate Y_{t+l} by a forecasted value $\widetilde{Y}_t(l)$. Observe that according to Theorem 1, $G_l(\phi_l, \theta_l)$ is an increasing function of l. In particular, for $l \geq 1$,

$$G_1(\phi_1, \theta_1) \leq G_l(\phi_l, \theta_l) \tag{13}$$

By definition, since (ϕ_l, θ_l) minimizes the function G_l, it is clear that

$$G_l(\phi_l, \theta_l) \leq G_l(\phi, \theta) \qquad \text{for all } \phi \text{ and } \theta$$

In particular, when we substitute the non-adaptive value (ϕ_1, θ_1) into the right-hand side, this last inequality says that when $l \geq 1$ is fixed, it is always better to use an adaptive procedure than a non-adaptive one. However, as suggested in Theorem 1, any approximation of Y_{t+l} would automatically incur in estimation error. In order to assess the effect of the approximation, it would be important to find the relationship between the long-memory parameter d and the optimal value of (ϕ_l, θ_l) which minimizes the MSE $G_l(\phi, \theta)$. Using the monotonicity of G_l given in equation (13), the next theorem characterizes this relationship.

Theorem 2 With the same notation as in Theorem 1, the value ϕ_l that minimizes $G_l(\phi, \theta_l)$ satisfies $\phi_l = (-p_l(\phi_l) \pm [(p_l(\phi_l))^2 + 4(p'_l(\phi_l))^2]^{\frac{1}{2}})/(2p'_l(\phi_l))$ where $p'_l(\phi) = \mathrm{d}p_l(\phi)/\mathrm{d}\phi$ and $\phi_l > 0$. In addition, the (ϕ_l, θ_l) that minimizes $G_l(\phi, \theta)$ satisfies the relationship $\theta_l < \phi_l$ for $d > 0$ and $\theta_l > \phi_l$ for $d < 0$.

Proof. To minimize $f_l(\phi) = G_l(\phi, \theta_l)$ with respect to ϕ, observe that

$$
\begin{aligned}
0 = \frac{\mathrm{d}f_l(\phi)}{\mathrm{d}\phi} &= -2(1-\phi^2)p_l(\phi)p'_l(\phi) + 2\phi(p_l(\phi))^2 \\
&= 2p_l(\phi)(\phi p_l(\phi) - (1-\phi^2)p'_l(\phi))
\end{aligned}
\tag{14}
$$

From the proof of Lemma 1, recall that $p_l(\phi) > 0$ for $d > 0$ and $p_l(\phi) < 0$ for $d < 0$. Therefore, the solution to equation (14) is attained at ϕ_l which solves the equation

$$
\frac{\phi}{1-\phi^2} = \frac{p'_l(\phi)}{p_l(\phi)}, \tag{15}
$$

i.e. ϕ_l satisfies

$$
\phi_l = (-p_l(\phi_l) \pm [(p_l(\phi_l))^2 + 4(p'_l(\phi_l))^2]^{\frac{1}{2}})/(2p'_l(\phi_l)) \tag{16}
$$

To prove that $\phi_l > 0$, it suffices to show that $p'_l(\phi)$ and $p_l(\phi)$ have the same sign for all $|\phi| < 1$. Since it is clear that for $0 \leq \phi < 1$, $p'_l(\phi) > 0$ when $0 < d < 1/2$ and $p'_l(\phi) < 0$ when $-1/2 < d < 0$, it remains to prove that $p'_l(\phi)$ and $p_l(\phi)$ have the same sign for $-1 < \phi < 0$. To this end, we show that $p_l(\phi)$ is an increasing function of ϕ for $0 < d < 1/2$, and $p_l(\phi)$ is a decreasing function of ϕ for $-1/2 < d < 0$.

Let $-1 < \phi_1 < \phi_2 < 0$ be given. Observe that

$$
\begin{aligned}
p_l(\phi_2) - p_l(\phi_1) &= \sum_{j=0}^{\infty} \psi_{j+l}(\phi_2^j - \phi_1^j) \\
&= \sum_{j=1}^{\infty} \psi_{j+l}(\phi_2 - \phi_1)\left(\sum_{k=0}^{j-1} \phi_2^{j-1-k}\phi_1^k\right) \\
&= (\phi_2 - \phi_1)\sum_{j=0}^{\infty} \psi_{j+l+1}\left(\sum_{k=0}^{j} \phi_2^{j-k}\phi_1^k\right)
\end{aligned}
$$

$$= (\phi_2 - \phi_1)\sum_{k=0}^{\infty}\sum_{j=k}^{\infty}\psi_{j+l+1}\phi_2^{j-k}\phi_1^k$$

$$= (\phi_2 - \phi_1)\sum_{k=0}^{\infty}\phi_1^k\sum_{j=0}^{\infty}\psi_{j+l+k+1}\phi_2^j$$

$$= (\phi_2 - \phi_1)\sum_{k=0}^{\infty}\phi_1^k p_{l+k+1}(\phi_2) \tag{17}$$

First, for $0 < d < 1/2$, we only have to establish $\sum_{k=0}^{\infty}\phi_1^k p_{l+k+1}(\phi_2) > 0$. To see this, observe that

$$\sum_{k=0}^{\infty}\phi_1^k p_{l+k+1}(\phi_2) = \sum_{k=0}^{\infty}(\phi_1^{2k} p_{l+2k+1}(\phi_2) + \phi_1^{2k+1} p_{l+2k+2}(\phi_2))$$

$$= \sum_{k=0}^{\infty}\phi_1^{2k}(p_{l+2k+1}(\phi_2) + \phi_1 p_{l+2k+2}(\phi_2))$$

$$\geq \sum_{k=0}^{\infty}\phi_1^{2k}(p_{l+2k+1}(\phi_2) - p_{l+2k+2}(\phi_2)) \geq 0$$

The last inequality follows from cases 1 and 3 of Lemma 1 that for $p_{m-1}(\phi) \geq p_m(\phi)$ for $|\phi| < 1$.

Second, for $-1/2 < d < 0$, from equation (17), it remains to show that $\sum_{k=0}^{\infty}\phi_1^k p_{l+k+1}(\phi_2) < 0$. By the same token, using cases 2 and 4 of Lemma 1, we have

$$\sum_{k=0}^{\infty}\phi_1^k p_{l+k+1}(\phi_2) = \sum_{k=0}^{\infty}\phi_1^{2k}(p_{l+2k+1}(\phi_2) + \phi_1 p_{l+2k+2}(\phi_2))$$

$$\leq \sum_{k=0}^{\infty}\phi_1^{2k}(p_{l+2k+1}(\phi_2) - p_{l+2k+2}(\phi_2)) \leq 0$$

since $p_{m-1}(\phi) \leq p_m(\phi) < 0$. This completes the proof of the fact that $\phi_l > 0$.

Finally, notice that (ϕ_l, θ_l) which minimizes $G_l(\phi, \theta)$ satisfies

$$\theta_l = \phi_l - (1 - \phi_l^2)\frac{p_l(\phi_l)}{(\phi_l)^{l-1}} \tag{18}$$

Therefore, $\theta_l < \phi_l$ for $d > 0$ as $p_l(\phi) > 0$ and $\theta_l > \phi_l$ for $d < 0$ as $p_l(\phi) < 0$. □

As an application, we now demonstrate how a characterization between d and (ϕ, θ) can be established. Specifically, consider $\phi = (-p_l(\phi) \pm [(p_l(\phi))^2 + 4(p_l'(\phi))^2]^{\frac{1}{2}})/(2p_l'(\phi))$. This value can be solved through iterative procedures such as the Newton–Raphson with a given starting value. Recall from the definition

$$p_{l-1}(\phi) = \sum_{j=l-1}^{\infty}\psi_j\phi^{j-(l-1)} = \psi_{l-1} + \phi p_l(\phi)$$

therefore,

$$p'_{l-1}(\phi) = p_l(\phi) + \phi p'_l(\phi)$$

It follows from this equality and equation (15) that

$$\begin{aligned}\frac{p'_l(\phi)}{p_l(\phi)} &= \left(\frac{p'_{l-1}(\phi)}{p_l(\phi)} - 1\right)\phi^{-1} \\ &= \frac{p'_{l-1}(\phi)}{p_{l-1}(\phi)}\left(\frac{p_{l-1}(\phi)}{p_{l-1}(\phi) - \psi_{l-1}}\right) - \frac{1}{\phi}\end{aligned}$$

As a result, we can evaluate the solution $p'_l(\phi)/p_l(\phi)$ from this recursion iteratively. For example, by putting $l = 0$ in equation (15), we have

$$\frac{\phi}{1-\phi^2} = \frac{p'_0(\phi)}{p_0(\phi)} = \frac{d(1-\phi)^{-d-1}}{(1-\phi)^{-d}} = \frac{d}{1-\phi}$$

which leads to the solution $\phi = d/1 - d$. It is immediate from this expression that as d tends to 1/2, ϕ tends to 1. Similarly, by putting $l = 1$ in equation (15), we obtain

$$\frac{\phi}{1-\phi^2} = \frac{p'_1(\phi)}{p_1(\phi)} = \frac{d}{1-\phi}\left(\frac{(1-\phi)^{-d}}{(1-\phi)^{-d} - 1}\right) - \frac{1}{\phi} \tag{19}$$

For a given d, the solution to the above equation can be studied graphically by plotting the difference of the two sides of equation (19) and locating its intersection with the axis $\phi = 0$. These functions are plotted in Figure 1. Explicit solutions of equations (18) for θ and (19) for ϕ are displayed in Figure 2 for $l = 1$. For $d = 0.25$, we can obtain the solutions for the values of ϕ and θ from Figure 2. In other words, as far as the one-step-ahead forecast is concerned, for $l = 1$ and $d = 0.25$, we observe from Figure 2 that the best ARMA(1,1) model is given by $\phi = 0.884$ and $\theta = 0.722$. Similarly, we observe from Figure 2 that as the long-memory parameter d tends to 1/2, the autoregressive parameter ϕ tends to 1 and hence the moving average parameter θ tends to 1 from equation (18).

In summary, Theorem 2 provides a numerical algorithm to calculate the relationship between d and (ϕ, θ) for a given forecast horizon l. By solving equations similar to (18) and (19) numerically, we can obtain the optimal values of ϕ and θ in terms of d. This characterization has two important consequences. First, it provides an analytic explanation of the well-known phenomenon reported in Hosking (1984) that as d tends to 1/2, the approximating ARMA(1,1) process would have both parameters close to the unit circle. Second, Theorem 2 provides a means to obtain the best approximating ARMA(1,1) model for different values of d.

APPROXIMATION OF ARFIMA PROCESSES

In this section, instead of a fractional noise model, we extend our study to the case where $\{X_t\}$ is assumed to be a general ARFIMA process. Specifically, let $\{X_t\}$ be a ARFIMA(p, d, q) process given by $\phi(B)X_t = (1-B)^{-d}\theta(B)\varepsilon_t$, where $\phi(B)$ and $\theta(B)$ are polynomials in B of pth and qth

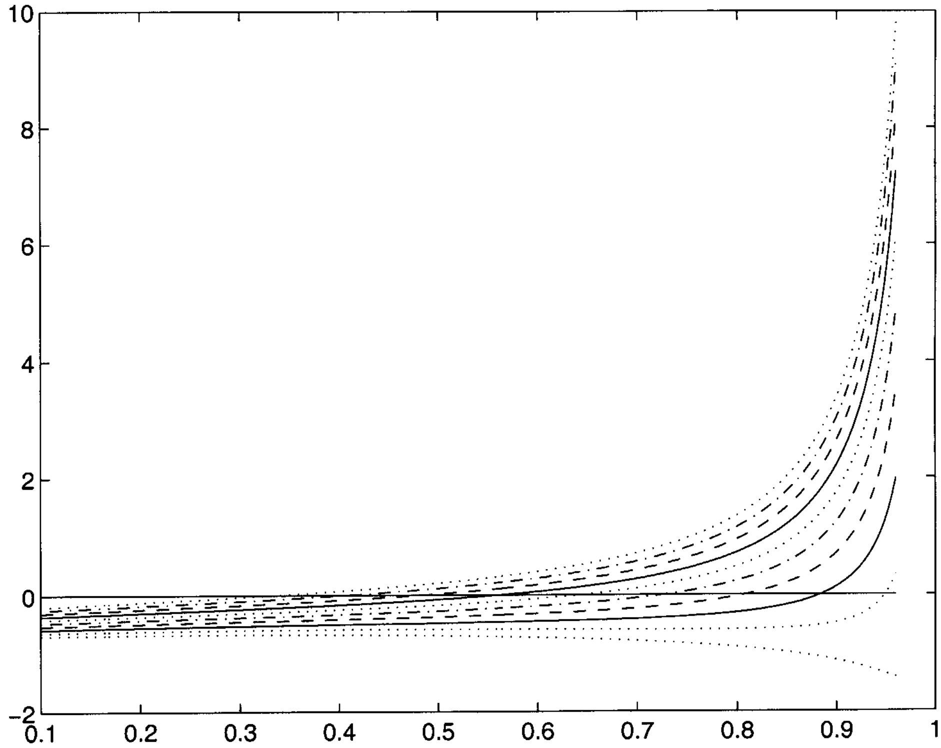

Figure 1. Plots of solutions of ϕ to equation (19). These solutions are obtained by the intersections of the plotted curves with the zero horizontal axis. The top curve starts at $d = -0.45$ with an increment of 0.1 moving downward, resulting in a total of 10 curves ranging from $d = -0.45, -0.35, \ldots, 0.35$, and 0.45 with the bottom curve being at $d = 0.45$

degrees respectively. We further assume that all roots of $\phi(z) = 0, \theta(z) = 0$ lie outside the unit disk and $\phi(z)$ and $\theta(z)$ have no common zeros. Under these assumptions, $\{X_t\}$ can be expressed as

$$X_t = \sum_{j=0}^{\infty} \pi_j \varepsilon_{t-j} \tag{20}$$

where the coefficients $\{\pi_j\}$ are functions of the autoregressive parameters, the moving average parameters, and the long-memory parameter d. The long-memory parameter d takes values in $(-0.5, 0.5)$ and ε_t is a sequence of independent standard normal random variables. Recall

$$\begin{aligned} G_l(\phi, \theta) &= E(X_{t+l} - \widetilde{Y}_t(l))^2 \\ &= \sum_{j=0}^{\infty} \pi_j^2 + \sum_{j=l}^{\infty} a_j^2 - 2\sum_{j=l}^{\infty} \pi_j a_j \qquad \text{for } l \geq 1 \end{aligned} \tag{21}$$

Analogous to Theorem 1, there exists an ARMA(1,1) process $\{Y_t\}$ which minimizes the mean square error when it is used to approximate an ARFIMA(*p, d, q*) process $\{X_t\}$. Before stating the main result, we need to establish two lemmas.

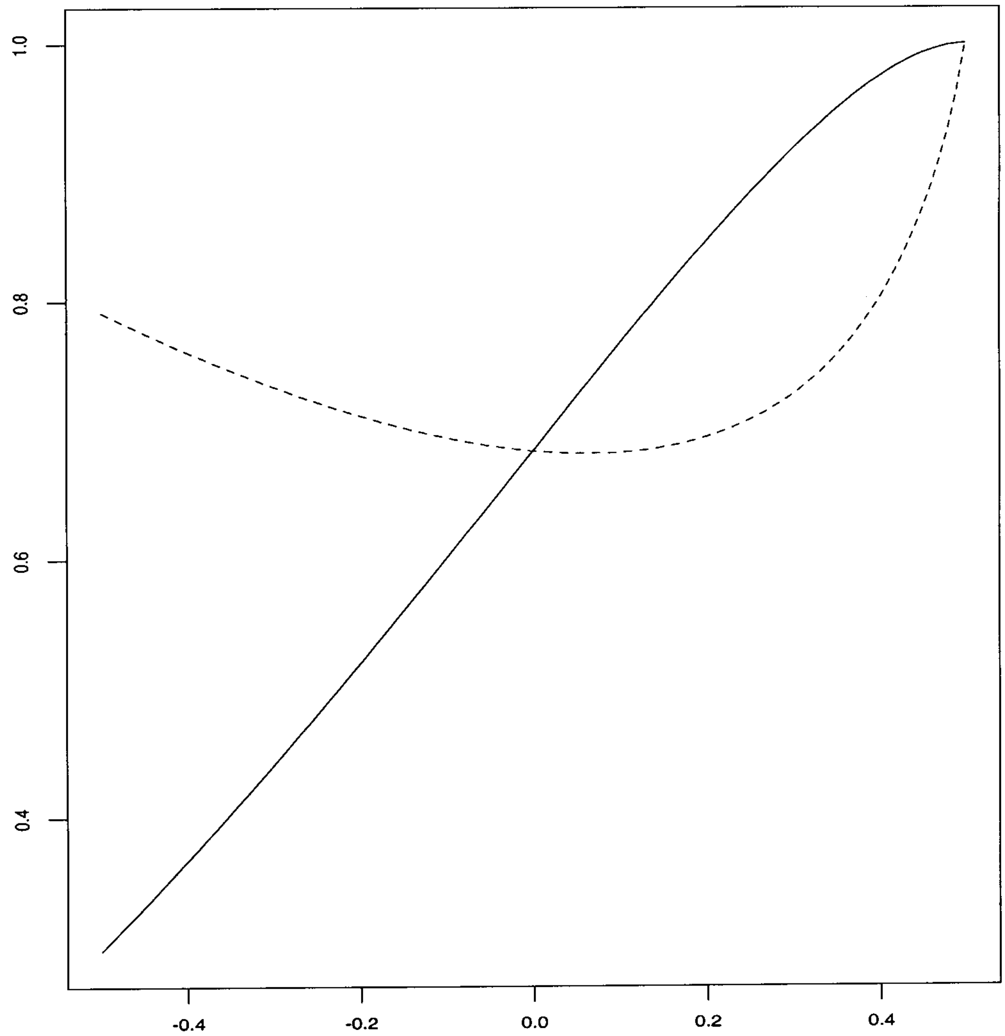

Figure 2. Plots of solutions of ϕ to equation (19) (heavy line) and θ to (18) (broken line) for different values of d (horizontal axis) for $l = 1$

Lemma 2 For $l > 1$ and $|\phi| < 1$, define $q_l(\phi) = \sum_{j=l}^{\infty} \pi_j \phi^{j-l}$. Assume that

(i) $|\pi_j| \downarrow j$
(ii) Every π_j has the same sign for $j \geq 1$
(iii) $|\pi_j - \pi_{j+1}| \downarrow j$ and
(iv) $\sum_{j=0}^{\infty} \pi_j^2 < \infty$;

then $(q_l(\phi))^2 \leq (q_{l-1}(\phi))^2$, for all $l \geq 2$ and $\phi \in (-1, 1)$.

Proof. By (iv), $q_l(\phi)$ is an absolutely summable power series (in ϕ) with radius of convergence equalling 1. So we consider two cases, $\phi \in [0, 1)$ and $\phi \in (-1, 0)$. Let $0 \leq \phi < 1$. By (ii), each π_j

has the same sign for $j \geq 1$. Under (i), if $0 \leq \pi_j \leq \pi_{j-1}$, $j \geq 2$, then $0 \leq q_l(\phi) \leq q_{l-1}(\phi)$, $l \geq 2$. Therefore,

$$(q_l(\phi))^2 \leq (q_{l-1}(\phi))^2 \qquad l \geq 2$$

On the other hand, if $\pi_{j-1} \leq \pi_j \leq 0$, $j \geq 2$, then $q_{l-1}(\phi) \leq q_l(\phi) \leq 0$, $l \geq 2$. Therefore,

$$(q_l(\phi))^2 \leq (q_{l-1}(\phi))^2 \qquad l \geq 2$$

For $-1 < \phi < 0$, let $\delta = -\phi$. First assume each π_j is positive, then by (i) and (iii),

$$\begin{aligned} 0 &\leq \sum_{k=0}^{\infty}(\pi_{2k+l} - \pi_{2k+l+1})\delta^{2k+1} \\ &\leq \sum_{k=0}^{\infty}(\pi_{2k+l-1} - \pi_{2k+l})\delta^{2k+1} \\ &\leq \sum_{k=0}^{\infty}(\pi_{2k+l-1} - \pi_{2k+l})\delta^{2k}. \end{aligned}$$

This implies, for $l \geq 2$,

$$\begin{aligned} 0 &\leq q_l(\phi) \\ &= \sum_{k=0}^{\infty}\pi_{2k+l}\delta^{2k} - \sum_{k=0}^{\infty}\pi_{2k+l+1}\delta^{2k+1} \\ &\leq \sum_{k=0}^{\infty}\pi_{2k+l-1}\delta^{2k} - \sum_{k=0}^{\infty}\pi_{2k+l}\delta^{2k+1} \\ &= q_{l-1}(\phi) \end{aligned}$$

Thus, $(q_l(\phi))^2 \leq (q_{l-1}(\phi))^2$.

Similarly, if we assume the π_j's to be negative, then by (i) and (iii), for $l \geq 2$,

$$\begin{aligned} \sum_{k=0}^{\infty}(\pi_{2k+l} - \pi_{2k+l-1})\delta^{2k} &\geq \sum_{k=0}^{\infty}(\pi_{2k+l+1} - \pi_{2k+l})\delta^{2k} \\ &\geq \sum_{k=0}^{\infty}(\pi_{2k+l+1} - \pi_{2k+l})\delta^{2k+1} \\ &\geq 0 \end{aligned}$$

This implies, for $l \geq 2$,

$$\begin{aligned} q_{l-1}(\phi) &= \sum_{k=0}^{\infty}\pi_{2k+l-1}\delta^{2k} - \sum_{k=0}^{\infty}\pi_{2k+l}\delta^{2k+1} \\ &\leq \sum_{k=0}^{\infty}\pi_{2k+l}\delta^{2k} - \sum_{k=0}^{\infty}\pi_{2k+l+1}\delta^{2k+1} \\ &= q_l(\phi) \leq 0 \end{aligned}$$

Hence, $(q_l(\phi))^2 \leq (q_{l-1}(\phi))^2$.

Under assumptions (i) and (iii) and using the arguments in Lemma 1, $(q_l(\phi))^2 \le (q_{l-1}(\phi))^2$, for all $\phi \in (-1, 1)$ and for all $l \ge 2$. □

Lemma 3 Let $q_l(\phi) = \sum_{j=l}^{\infty} \pi_j \phi^{j-l}$, for $l \ge 1$ and $\phi \in (-1, 1)$. Assume that

(i) $|\pi_j| \downarrow j$
(ii)′ Consecutive π_js have opposite signs for all $j \ge 1$,
(iii)′ $(|\pi_j| - |\pi_{j+1}|) \downarrow j$, and
(iv) $\sum_{j=0}^{\infty} \pi_j^2 < \infty$;

then $(q_l(\phi))^2 \le (q_{l-1}(\phi))^2$, for all $l \ge 2$ and $\phi \in (-1, 1)$.

Proof. As in Lemma 2, notice that by (iv), $q_l(\phi)$ is an absolutely summable power series (in ϕ) with radius of convergence equalling 1. Let $0 \le \phi < 1$. By (ii)′, consecutive π_js have opposite signs for $j \ge 1$. So, under assumptions (ii)′ and (iii)′,

$$\begin{aligned} 0 &\le \sum_{k=0}^{\infty} (|\pi_{2k+l}| - |\pi_{2k+l+1}|)\phi^{2k+1} \\ &\le \sum_{k=0}^{\infty} (|\pi_{2k+l-1}| - |\pi_{2k+l}|)\phi^{2k+1} \\ &\le \sum_{k=0}^{\infty} (|\pi_{2k+l-1}| - |\pi_{2k+l}|)\phi^{2k} \end{aligned}$$

This implies that for $l \ge 2$ with π_l positive, each π_{2k+l} is positive and

$$\begin{aligned} 0 &\le q_l(\phi) \\ &= \sum_{k=0}^{\infty} \pi_{2k+l}\phi^{2k} + \sum_{k=0}^{\infty} \pi_{2k+l+1}\phi^{2k+1} \\ &\le -\left(\sum_{k=0}^{\infty} \pi_{2k+l-1}\phi^{2k} + \sum_{k=0}^{\infty} \pi_{2k+l}\phi^{2k+1} \right) \\ &= -q_{l-1}(\phi) \end{aligned}$$

Therefore, we obtain $(q_l(\phi))^2 \le (q_{l-1}(\phi))^2$. Also, when π_l is negative, each π_{2k+l} is negative and

$$\begin{aligned} 0 &\le -q_l(\phi) \\ &= -\left(\sum_{k=0}^{\infty} \pi_{2k+l}\phi^{2k} + \sum_{k=0}^{\infty} \pi_{2k+l+1}\phi^{2k+1} \right) \\ &\le \sum_{k=0}^{\infty} \pi_{2k+l-1}\phi^{2k} + \sum_{k=0}^{\infty} \pi_{2k+l}\phi^{2k+1} \\ &= q_{l-1}(\phi) \end{aligned}$$

we have $(q_l(\phi))^2 \leq (q_{l-1}(\phi))^2$. For $-1 < \phi < 0$, let $\delta = -\phi$. By assumptions (i) and (ii)$'$ and for $l \geq 2$, observe

$$\begin{aligned}\sum_{k=0}^{\infty} |\pi_{2k+l-1}|\delta^{2k} &\geq \sum_{k=0}^{\infty} |\pi_{2k+l}|\delta^{2k} \\ &\geq \sum_{k=0}^{\infty} |\pi_{2k+l}|\delta^{2k+1} \\ &\geq \sum_{k=0}^{\infty} |\pi_{2k+l+1}|\delta^{2k+1} \\ &\geq 0\end{aligned}$$

This implies that for $l \geq 2$ and π_l positive,

$$\begin{aligned}q_{l-1}(\phi) &= \sum_{k=0}^{\infty} \pi_{2k+l-1}\delta^{2k} - \sum_{k=0}^{\infty} \pi_{2k+l}\delta^{2k+1} \\ &\leq -\sum_{k=0}^{\infty} \pi_{2k+l}\delta^{2k} + \sum_{k=0}^{\infty} \pi_{2k+l+1}\delta^{2k+1} \\ &= -q_l(\phi) \leq 0\end{aligned}$$

Therefore, $(q_l(\phi))^2 \leq (q_{l-1}(\phi))^2$. Similarly, when π_l is negative, by (i) and (ii)$'$ we have

$$\begin{aligned}q_{l-1}(\phi) &= \sum_{k=0}^{\infty} \pi_{2k+l-1}\delta^{2k} - \sum_{k=0}^{\infty} \pi_{2k+l}\delta^{2k+1} \\ &\geq -\sum_{k=0}^{\infty} \pi_{2k+l}\delta^{2k} + \sum_{k=0}^{\infty} \pi_{2k+l+1}\delta^{2k+1} \\ &= -q_l(\phi) \geq 0\end{aligned}$$

This implies $(q_l(\phi))^2 \leq (q_{l-1}(\phi))^2$. Thus, by (i), (ii)$'$, and (iii)$'$, $(q_l(\phi))^2 \leq (q_{l-1}(\phi))^2$ for all $\phi \in (-1, 1)$ and for all $l \geq 2$. □

With these lemmas, we now state the main theorem.

Theorem 3 Let $\{X_t\}$ be the ARFIMA(p, d, q) process following equation (20). Let $\{Y_t\}$ be an ARMA(1, 1) process defined in equation (22). Let $\tilde{Y}_t(l)$ be the l-step predictor based on Y_t. Assume that π_j satisfies either:

(i) $|\pi_j| \downarrow j$
(ii) For each fixed parameter value, each π_j has the same sign for all $j \geq 1$
(iii) $|\pi_j - \pi_{j+1}| \downarrow j$
(iv) $\sum_{j=0}^{\infty} \pi_j^2 < \infty$

or

(i) and (iv) as above,
(ii)′ for each fixed parameter value, consecutive π_j has opposite sign for all $j \geq 1$,
(iii)′ $(|\pi_j| - |\pi_{j+1}|) \downarrow j$; then

$$\min_{\phi,\theta} E(X_t - Y_t)^2 = \min_{\phi,\theta} E(X_{t+l} - Y_{t+l})^2 \leq \min_{\phi,\theta} E(X_{t+l} - \widetilde{Y}_t(l))^2 \qquad \text{for } l \geq 1.$$

Furthermore, the MSE is monotonic in l, i.e.

$$G_1(\phi_1, \theta_1) \leq G_l(\phi_l, \theta_l) \leq G_{l+1}(\phi_{l+1}, \theta_{l+1}) \text{ for } l \geq 1$$

Proof. Let $q_l(\phi) = \sum_{j=l}^{\infty} \pi_j \phi^{j-l}$, for $l \geq 1$. Observe that $E(X_t - Y_t)^2 = \sum_{j=0}^{\infty} \pi_j^2 + \sum_{j=0}^{\infty} a_j^2 - 2\sum_{j=0}^{\infty} \pi_j a_j = E(X_{t+l} - \widetilde{Y}_t(l))^2 - 1$. So it suffices to prove $\min_{\phi,\theta} G_1(\phi, \theta) \leq \min_{\phi,\theta} G_l(\phi, \theta)$, for $l \geq 1$, where $G_l(\phi, \theta)$ is defined in equation (21).

In order to establish this inequality, we minimize $G_l(\phi, \theta) = K(\pi) - 2(\phi - \theta)\sum_{j=l}^{\infty} \pi_j \phi^{j-1} + (\phi - \theta)^2 \phi^{2(l-1)}/(1 - \phi^2)$ with respect to θ and ϕ respectively, where $K(\pi) = \sum_{j=0}^{\infty} \pi_j^2$. Differentiating $G_l(\phi, \theta)$ with respect to θ and equating it to zero,

$$(\phi - \theta_l) = \frac{\sum_{j=l}^{\infty} \pi_j \phi^{j-1}}{\sum_{j=l}^{\infty} \phi^{2(j-1)}} = (1 - \phi^2)\frac{\sum_{j=l}^{\infty} \pi_j \phi^{j-1}}{\phi^{2(l-1)}} \tag{22}$$

where θ_l is a minimum point of $G_l(\phi, \theta)$ for every ϕ as $\partial^2 G_l(\phi, \theta)/\partial\theta^2 = 2\sum_{j=l}^{\infty} \phi^{2(j-1)} > 0$ for $\phi \neq 0$. Define

$$h_l(\phi) = G_l(\phi, \theta_l) = K(\pi) - (1 - \phi^2)(q_l(\phi))^2, \tag{23}$$

where $q_l(\phi)$ is defined in Lemma 2.

Using the preceding argument, we conclude that for $l \geq 2, h_l(\phi) \geq h_{l-1}(\phi)$ since $|\phi| < 1$, and from Lemmas 2 and 3 $(q_l(\phi))^2 \leq (q_{l-1}(\phi))^2$, i.e. $(1 - \phi^2)(q_l(\phi))^2 \leq (1 - \phi^2)(q_{l-1}(\phi))^2$. Thus, $h_l(\phi) \geq h_{l-1}(\phi)$, for $l \geq 2$ which is equivalent to $(1 - \phi^2)(q_l(\phi))^2 \leq (1 - \phi^2)(q_{l-1}(\phi))^2$, for $l \geq 2$, i.e.

$$(q_l(\phi))^2 \leq (q_{l-1}(\phi))^2 \qquad \text{for } l \geq 2 \qquad \text{as } |\phi| < 1 \tag{24}$$

Therefore, for each $l \geq 2$, if at $\phi = \phi_l$ the function $h_l(\phi)$ attains its minimum, it is clear that $h_l(\phi_l) \geq h_{l-1}(\phi_l) \geq h_{l-1}(\phi_{l-1})$. Repeating the inequality for each l until $l = 2$, we have $h_l(\phi_l) \geq h_1(\phi_1)$. This implies $G_l(\phi_l, \theta_l) \geq G_1(\phi_1, \theta_1)$ for $l \geq 1$.

By definition, $E(X_{t+l} - Y_{t+l})^2 = E(X_t - Y_t)^2$, for any $l \geq 1$. Hence, $\min_{\phi,\theta} E(X_{t+l} - Y_{t+l})^2 = \min_{\phi,\theta} E(X_t - Y_t)^2 \leq \min_{\phi,\theta} E(X_{t+l} - \widetilde{Y}_t(l))^2$, for all $l \geq 1$. This completes the proof of Theorem 3. □

Like Theorem 1, Theorem 3 states the intuitive fact that as far as mean square error is concerned, forecasting X_{t+l} by Y_{t+l} directly is always better than forecasting X_{t+l} by a forecasted value $\widetilde{Y}_t(l)$ based on a fixed ARMA(1,1) model $\{Y_t\}$ a *priori*. However, contrary to the pure fractional noise case studied in Theorem 1, when the underlying process $\{X_t\}$ is an ARFIMA(p, d, q), certain conditions on the weights π_j in equation (30) need to be satisfied in order for this theorem to hold. This result also suffers from the same drawback as Theorem 1 since the future value Y_{t+l} is usually not available at time t. Instead, we end up forecasting Y_{t+l} from an ARMA(1,1) model either adaptively or non-adaptively and using this forecasted value of Y_{t+l} to approximate X_{t+l}. As shown in equation (13), although Theorem 3 indicates that a adaptive forecast is always better than a non-adaptive forecast, it will be useful to find which ARMA(1,1) model provides a better approximating model for the adaptive procedure. Theorems 1, 2 and 3 together provide a theoretical guideline to look for the best approximating ARMA(1,1) model and quantify the relationship between d and ϕ and θ.

Tables I–IV illustrate the differences between the adaptive and non-adaptive forecast under four different ARFIMA(p, d, q) models for $d = 0.25$ and $d = 0.45$. The orders p, q and the parameters ϕ, θ of these models are given in these tables which are organized as follows. The first column consists of the length of future horizons l, the second column lists the actual forecast error variance, denoted by σ_l^2, when X_{t+l} is predicted from the generating fractional model X_t. This column represents the smallest forecast error variance in the best scenario. The third column contains the values of the AR parameter ϕ_l when using an ARMA(1,1) model to predict X_{t+l}, the fourth column is the ratio of the adaptive forecast error variance with respect to the actual forecast error variance, while the fifth column is the ratio of the variances of the nonadaptive forecast error based on a fixed ARMA(1,1) with respect to the actual forecast error variance. Note that these tables are computed by minimizing the respective quantities in Theorem 3. Similar to solving for ϕ in Theorem 2, we use the Newton–Raphson algorithm to solve for the ϕ_l which minimizes equation (23). After solving for ϕ_l, the value of θ_l is obtained from equation (22).

As can be seen, the largest gains are attained when d is close to 0.5. For a moderate value of d, 0.25 say, the gain of using an adaptive forecast is marginal with a horizon as large as $l = 20$. Depending on the goal, it may be prudent to use an adaptive scheme when the underlying model

Table I. $X_t = \dfrac{(1-B)^{-d}(1-0.725B)}{1-0.275B}\varepsilon_t \qquad d = 0.45$

l	σ_l^2	ϕ_l	$h_l(\phi_l)/\sigma_l^2$	$h_l(\phi_1)/\sigma_l^2$
1	1.000000	0.999230	1.116697	1.116697
2	1.000000	0.999204	1.116520	1.116523
3	1.000894	0.999176	1.115338	1.115350
4	1.002907	0.999210	1.113724	1.113726
5	1.005432	0.999273	1.111922	1.111931
6	1.008066	0.999338	1.110104	1.110162
7	1.010626	0.999398	1.108352	1.108504
8	1.013049	0.999450	1.106697	1.106982
9	1.015318	0.999495	1.105147	1.105594
10	1.017436	0.999534	1.103699	1.104332
11	1.019415	0.999568	1.102346	1.103181
12	1.021266	0.999597	1.101080	1.102131
13	1.023002	0.999623	1.099893	1.101168
14	1.024634	0.999646	1.098777	1.100283
15	1.026172	0.999666	1.097725	1.099468
16	1.027626	0.999684	1.096732	1.098714
17	1.029003	0.999700	1.095792	1.098015
18	1.030311	0.999715	1.094900	1.097365
19	1.031555	0.999728	1.094052	1.096759
20	1.032741	0.999740	1.093244	1.096193

Table II. $X_t = \dfrac{(1-B)^{-d}(1-0.625B)}{1-0.375B}\varepsilon_t \qquad d = 0.25$

l	σ_l^2	ϕ_l	$h_l(\phi_l)/\sigma_l^2$	$h_l(\phi_1)/\sigma_l^2$
1	1.000000	0.989311	1.005378	1.005378
2	1.000000	0.988054	1.004985	1.005006
3	1.000381	0.986340	1.004147	1.004243
4	1.001240	0.986082	1.003660	1.003763
5	1.002290	0.986846	1.003406	1.003465
6	1.003339	0.987982	1.003258	1.003276
7	1.004307	0.989133	1.003153	1.003153
8	1.005174	0.990169	1.003064	1.003072
9	1.005945	0.991065	1.002982	1.003018
10	1.006630	0.991830	1.002906	1.002982
11	1.007241	0.992483	1.002833	1.002959
12	1.007789	0.993045	1.002763	1.002945
13	1.008284	0.993532	1.002698	1.002938
14	1.008734	0.993957	1.002635	1.002935
15	1.009144	0.994330	1.002576	1.002936
16	1.009520	0.994661	1.002521	1.002939
17	1.009866	0.994956	1.002468	1.002944
18	1.010187	0.995221	1.002418	1.002950
19	1.010484	0.995459	1.002371	1.002957
20	1.010761	0.995675	1.002326	1.002964

Table III. $X_t = \dfrac{(1-B)^{-d}}{1+0.45B}\varepsilon_t \qquad d = 0.45$

l	σ_l^2	ϕ_l	$h_l(\phi_l)/\sigma_l^2$	$h_l(\phi_1)/\sigma_l^2$
1	1.000000	0.997013	1.520834	1.520834
2	1.106439	0.996556	1.372042	1.372210
3	1.120749	0.998271	1.378889	1.381494
4	1.151715	0.998540	1.348118	1.352403
5	1.167424	0.998878	1.338657	1.346684
6	1.184162	0.999045	1.325026	1.335755
7	1.196922	0.999187	1.315915	1.329766
8	1.208712	0.999285	1.307060	1.323638
9	1.218874	0.999365	1.299718	1.319003
10	1.228104	0.999428	1.293027	1.314823
11	1.236409	0.999480	1.287102	1.311304
12	1.244013	0.999523	1.281710	1.308183
13	1.250992	0.999559	1.276808	1.305447
14	1.257451	0.999591	1.272306	1.303003
15	1.263452	0.999618	1.268155	1.300815
16	1.269057	0.999642	1.264305	1.298839
17	1.274312	0.999663	1.260720	1.297046
18	1.279256	0.999681	1.257368	1.295410
19	1.283922	0.999698	1.254223	1.293912
20	1.288340	0.999713	1.251264	1.292532

Table IV. $X_t = \frac{(1-B)^{-d}}{1+0.25B}\varepsilon_t \qquad d = 0.25$

l	σ_l^2	ϕ_l	$h_l(\phi_l)/\sigma_l^2$	$h_l(\phi_1)/\sigma_l^2$
1	1.000000	0.964062	1.025366	1.025366
2	1.024414	0.939702	0.995702	0.996855
3	1.030518	0.970999	1.005703	1.005873
4	1.036246	0.976657	1.005084	1.005675
5	1.040091	0.981681	1.005591	1.006880
6	1.043151	0.984701	1.005540	1.007420
7	1.045591	0.986921	1.005474	1.007897
8	1.047612	0.988569	1.005354	1.008243
9	1.049316	0.989852	1.005223	1.008517
10	1.050780	0.990876	1.005088	1.008730
11	1.052055	0.991712	1.004955	1.008900
12	1.053179	0.992408	1.004827	1.009033
13	1.054178	0.992997	1.004706	1.009139
14	1.055075	0.993501	1.004590	1.009222
15	1.055886	0.993937	1.004481	1.009287
16	1.056624	0.994319	1.004379	1.009336
17	1.057298	0.994655	1.004282	1.009372
18	1.057918	0.994954	1.004190	1.009398
19	1.058491	0.995221	1.004104	1.009415
20	1.059022	0.995461	1.004022	1.009424

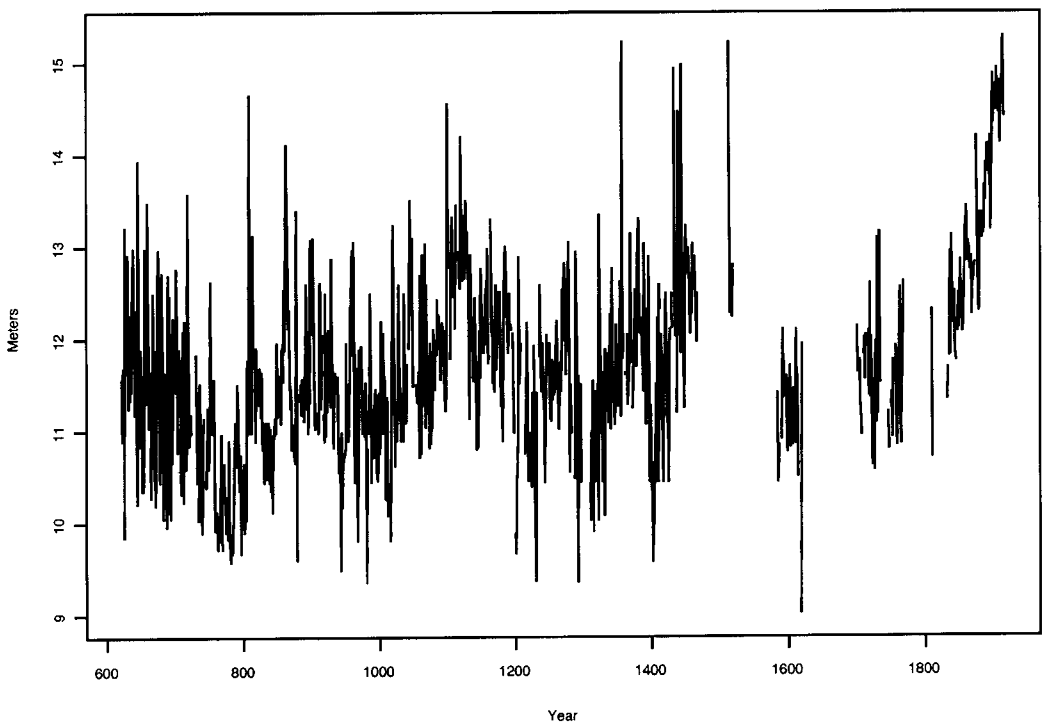

Figure 3. Nile river data

has a substantial long-memory parameter. On the other hand, if the long-memory effect is small, the difference of using an adaptive and a non-adaptive scheme may only be marginal when we do a medium-term forecast.

AN EXAMPLE

As an illustration, we apply the results from the preceding sections to the well-studied Nile river data set. Figure 3 displays the annual minimum level of the Nile river measured at the Roda gauge from AD 622 to AD 1921. This series has been studied by a number of people detecting long-memory behavior (see, for example, Beran, 1994). In order to compare the performance of the forecasting techniques discussed above we fit a non-adaptive ARMA(1,1) and a ARFIMA(0, d, 0) to the data. Following Beran (1994), the fitted fractional model has $d = 0.41$ and $\sigma_\varepsilon^2 = 0.54$. Similary, the fitted non-adaptive ARMA(1,1) model has $\phi = 0.96, \theta = 0.68$ and $\sigma_\varepsilon^2 = 0.54$.

From Figure 3, it is seen that the Nile river data exhibit long-range dependency as indicated in Beran (1994). One-step-ahead predictions using a non-adaptive ARMA(1,1) model are shown in Figure 4. Their standard deviations (see Figure 5) show a clear increase in values during periods of missing observations. A similar behaviour is observed in the prediction standard deviations of the ARFIMA model (see Figure 6).

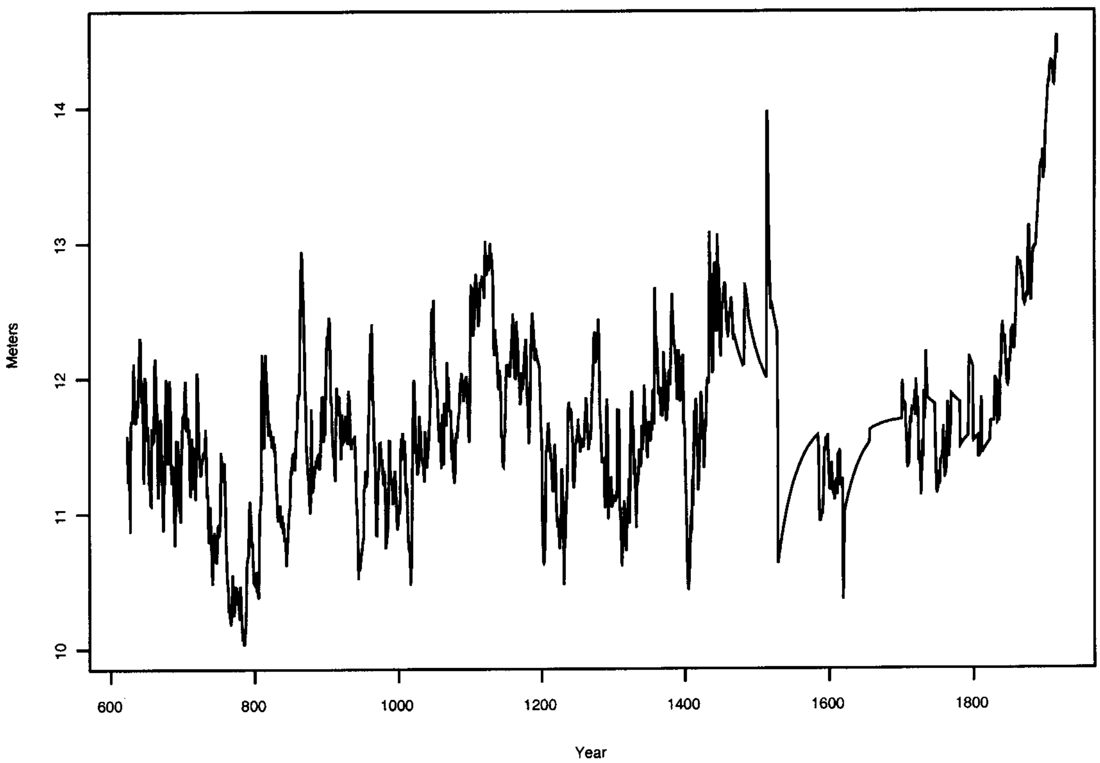

Figure 4. Nile river data: predictions using a non-adaptive ARMA(1,1)

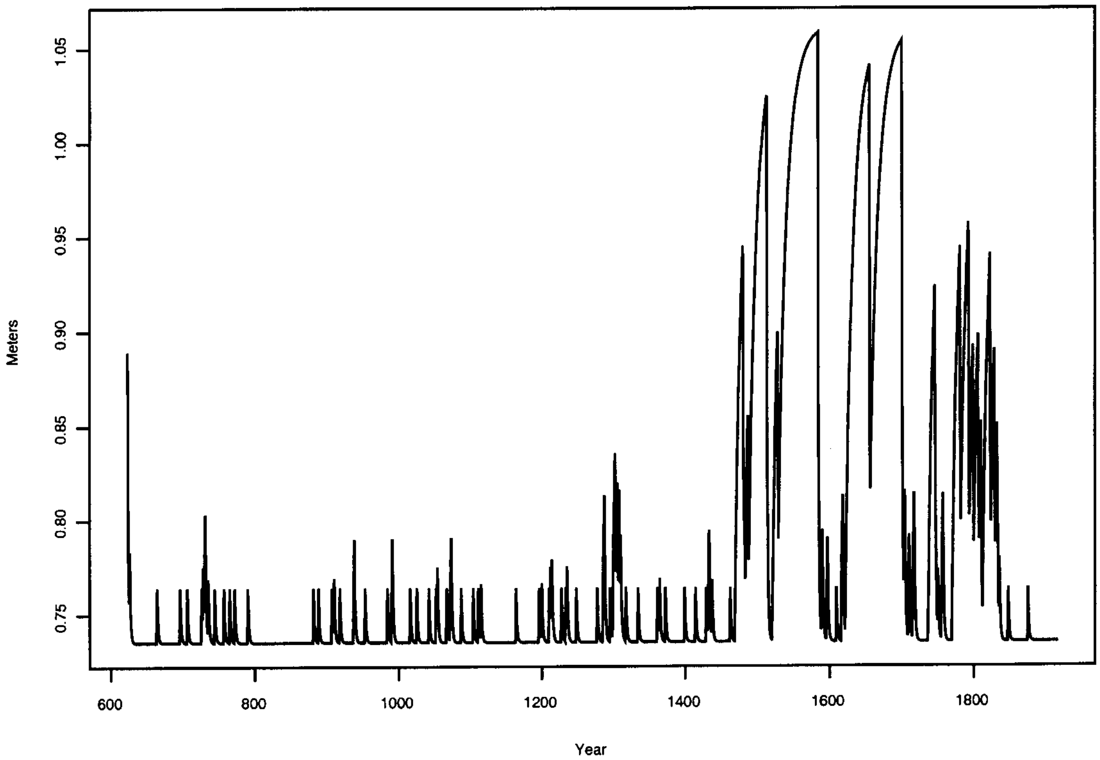

Figure 5. Nile river data: prediction of standard deviations based on the ARMA(1,1) model used in Figure 4

In order to assess the performance of the adaptive ARMA approach for l-step forecasts, we conduct a study similar to that of Brodsky and Hurvich (1999) for the Nile river data. Given the large number of missing values in the later period of the series, we only consider the first period, from AD 622 to AD 1281, which is the same time span used by Beran (1994). The estimated parameters are $d = 0.39$ and $\sigma^2 = 0.49$ and the results of the corresponding forecasts errors are given in Table V. In the table the first column lists the number of steps l being forecasted, the second column lists the prediction error variance from the fitted ARFIMA model, the third column displays the estimated parameter ϕ for the adaptive ARMA(1,1) model based on the Brodsky and Hurvich (1999) approach, the fourth column presents the ratio between the prediction error variance of the adaptive ARMA and the ARFIMA model, while the fifth column displays the same ratio between the non-adaptive ARMA(1,1) and the ARFIMA process. Observe from Table V that the ARFIMA and the adaptive ARMA models work well for predicting the Nile river data at different time horizons. On the other hand, the performance of the non-adaptive ARMA model is deficient, as compared to the other two approaches. Given that an estimated value of d is close to 0.4 for this series, the underperformance of the non-adaptive scheme is in accord with the computational studies reported earlier.

For certain data that exhibit strong long-memory behaviour such as the Nile river data, it seems prudent to model the data by means of an ARFIMA model or an adaptive ARMA(1,1) model as suggested by Hosking (1984) and Tiao and Tsay (1996) respectively. When the underlying

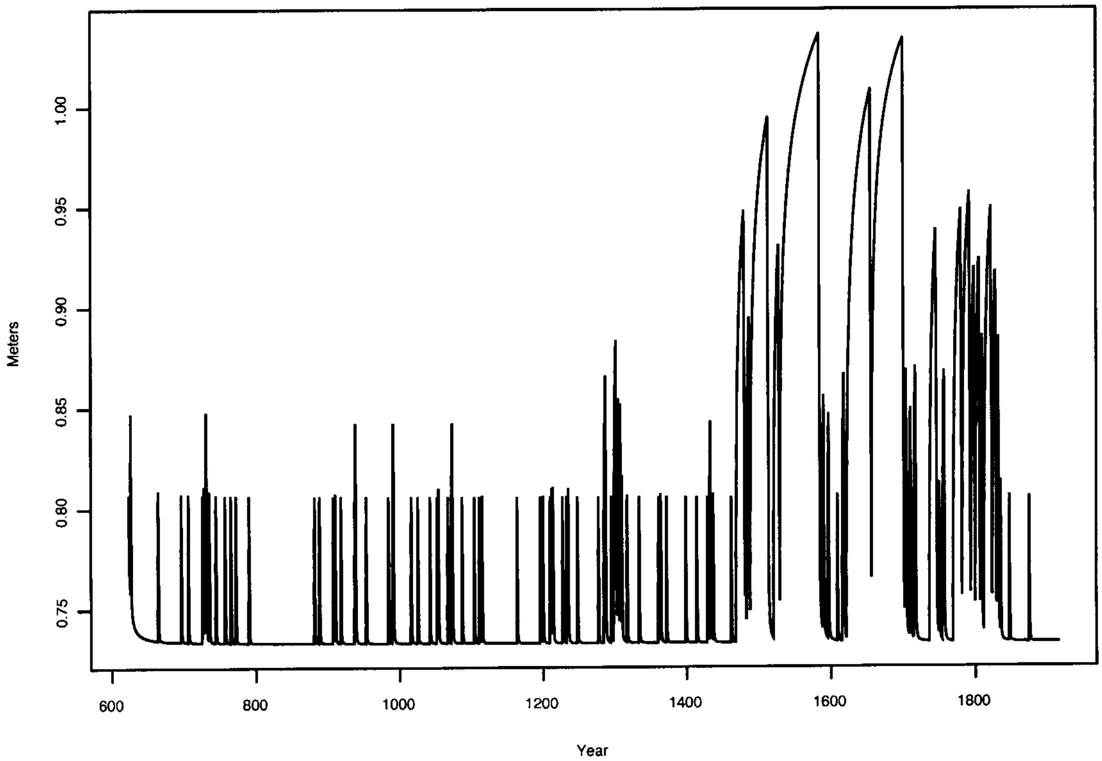

Figure 6. Nile river data: prediction of standard deviations based on ARFIMA

series is long, it may be desirable to fit an adaptive ARMA(1,1) model since it may often be time consuming to fit an ARFIMA model at the initial stage. Once we have some ideas about the approximating ARMA(1,1), Theorem 2 can be used to entertain a more focused ARFIMA model.

CONCLUSIONS

It has been folklore among time series analysts that one can use an ARMA(1,1) to approximate a long-memory model. The question is when and how? To answer this question, one needs a good theoretical understanding on what is to be approximated and what are the properties of the MSE incurred. This paper provides an answer to this type of questions and demonstrates, through Theorems 2 and 3, a numerical algorithm to characterize the relationship between d and ϕ and θ. This relationship is used to compute the MSE of different forecasting horizons for various ARFIMA models. From Tables I–IV, we assess under what circumstances an adaptive forecasting scheme gains most comparing it with a non-adaptive scheme. In addition, the Nile river data are analysed in Table V which demonstrates the usefulness of the adaptive scheme in a real application. The results of this paper should be of great interest to time series analysts who need to find ways to forecast a long-memory model.

Table V. Prediction square errors for a fitted ARFIMA(0,0.39,0) model and an adaptive ARMA(1,1) model for the Nile river data

l	σ_l^2	ϕ_l	$h_l(\phi_l)/\sigma_l^2$	$h_l(\phi_1)/\sigma_l^2$
1	0.489247	0.871365	1.026570	1.026570
2	0.563661	0.944291	1.031044	1.015476
3	0.599606	0.955950	1.015253	1.043132
4	0.622418	0.963265	1.009925	1.069446
5	0.638804	0.969950	1.004298	1.089612
6	0.651435	0.972120	0.994832	1.103810
7	0.661629	0.971804	0.990743	1.113127
8	0.670123	0.974780	0.994277	1.118688
9	0.677371	0.977053	0.997197	1.121444
10	0.683670	0.978081	0.994325	1.122155
11	0.689225	0.979064	0.996824	1.121404
12	0.694180	0.978608	0.994412	1.119629
13	0.698644	0.978211	0.993319	1.117160
14	0.702699	0.979063	0.995893	1.114238
15	0.706408	0.979254	0.997443	1.111041
16	0.709822	0.979233	0.997333	1.107697
17	0.712980	0.979206	0.995803	1.104296
18	0.715916	0.978881	0.991033	1.100903
19	0.718657	0.978382	0.989674	1.097561
20	0.721224	0.977256	0.988964	1.094300

ACKNOWLEDGEMENTS

Research was supported in part by the National Security Agency Grant No. 006-124, Hong Kong Research Grants Council Grant CUHK6082/98P and Fondecyt Grant 8000004/7980021. We would like to thank the referee and the departmental editor, Professor M. Ogaki, for insightful comments which led to an improved version of this paper.

REFERENCES

Bailllie RT. 1996. Long-memory processes and fractional integration in econometrics. *J. Econometrics* **73**: 5–59.

Beran J. 1994. *Statistics for Long-Memory Processes*. Chapman and Hall: New York.

Brockwell PJ, Davis RA. 1991. *Time Series: Theory and Methods*, 2nd edn. Springer-Verlag: New York.

Brodsky J, Hurvich CM. 1999. Multi-step forecasting for long-memory processes. *J. Forecasting* **18**: 59–75.

Chan NH, Palma W. 1998. State space modeling of long-memory processes. *Ann. Statist.* **26**: 719–740.

Hosking JRM. 1981. Fractional differencing. *Biometrika* **68**: 165–176.

Hosking JRM. 1984. Modelling persistence in hydrological time series using fractional differencing. *Water Resources Research* **20**: 1898–1908.

Samorodnitsky G, Taqqu MS. 1994. *Stable Non-Gaussian Random Processes*. Chapman and Hall: New York.

Sowell F. 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *J. Econometrics* **53**: 165–188.

Tiao GC, Tsay RS. 1994. Some advances in non-linear and adaptive modeling in time series. *J. Forecasting* **13**: 109–131.

Tiao GC, Xu D. 1993. Robustness of maximum likelihood estimates for multi-step predictions: The exponential smoothing case. *Biometrika* **80**: 623–641.

Tong H. 1997. Some comments on nonlinear time series analysis. In *Nonlinear Dynamics and Time Series*, Cutler CD, Kaplan DT (eds). Fields Institute Communications **11**: 17–27. American Mathematical Society: Rhode Island.

Authors' biographies:

Gopal K. Basak is a Lecturer in the Department of Mathematics, University of Bristol. His research interests include asymptotics of Markov processes, stochastic modelling and control, inference for stochastic processes and time series.

Ngai Hang Chan is Professor of Statistics at Carnegie Mellon University and Professor of Statistics and Director of the Risk Management Science Program at the Chinese University of Hong Kong. His research interests include finance and econometrics, risk management and time series modelling of long-range dependent data.

Wilfredo Palma is an Assistant Professor in the Department of Statistics, Pontificia Universidad Católica de Chile. His research interests include time series analysis and forecasting methodologies for long-range dependent data.

Author's addresses:
Gopal K. Basak, Department of Mathematics, University of Bristol, Bristol BS8 ITW, UK.

Ngai Hang Chan, Department of Statistics, Chinese University of Hong Kong, Shatin, NT, Hong Kong.

Wilfredo Palma, Department of Statistics, P. Universidad Católica de Chile, Casilla 306, Santiago 22, Chile.

