



ELSEVIER

International Journal of Forecasting 18 (2002) 167–179

international journal  
of forecasting

www.elsevier.com/locate/ijforecast

# Multistep forecasting of long memory series using fractional exponential models

Clifford M. Hurvich\*

*New York University, 100 Bleecker St., Apt. 2B, New York, NY 10012, USA*

---

## Abstract

We develop forecasting methodology for the fractional exponential (FEXP) model. First, we devise algorithms for fast exact computation of the coefficients in the infinite order autoregressive and moving average representations of a FEXP process. We also describe an algorithm to accurately approximate the autocovariances and to simulate realizations of the process. Next, we present a fast frequency-domain cross validation method for selecting the order of the model. This model selection method is designed to yield the model which provides the best multistep forecast for the given lead time, without assuming that the process actually obeys a FEXP model. Finally, we use the infinite order autoregressive coefficients of a fitted FEXP model to construct multistep forecasts of inflation in the United Kingdom. These forecasts are substantially different than those from a fitted ARFIMA model. © 2002 International Institute of Forecasters. Published by Elsevier Science B.V. All rights reserved.

*Keywords:* Fractional integration; Long-range dependence; Spectral factorization

---

## 1. Introduction

Multistep forecasting of long-memory series has received considerable attention in recent years. See, for example, Crato and Ray (1996), Ray (1993), Brodsky and Hurvich (1999), Tiao and Tsay (1994). Most of these works focus on the autoregressive fractionally integrated moving average (ARFIMA) model, which is a long-memory generalization of the well-known ARIMA models. The ARFIMA model, proposed independently by Hosking (1981) and

Granger and Joyeux (1980), is convenient for forecasting as the coefficients of the best linear predictor may be determined rather easily. In some cases, there is even an explicit formula for these coefficients. Nevertheless, estimation and model selection for ARFIMA models are somewhat problematic. Estimation is typically carried out by means of numerical optimization, which can only be reliably achieved when the dimension of the model is small. Further difficulties may arise from the need to constrain the autoregressive and moving average parameters for stationarity and invertibility. In addition, although there is some literature on the selection of ARFIMA models (e.g., Beran, Bhansali & Ocker, 1998), the focus in this literature is on

---

\*Tel.: +1-212-998-0449; fax: +1-212-995-4003.

*E-mail address:* churvich@stern.nyu.edu (C.M. Hurvich).

consistent selection of the true model order, assuming that the series actually obeys a low-dimensional ARFIMA model. No methodology, however, seems to have been proposed on selection of an ARFIMA model for the purpose of multistep prediction, in a setting where the ARFIMA models are viewed merely as approximations to the true process.

In this paper, we wish to develop forecasting methodology for a different class of long-memory models, known as the fractional exponential, or FEXP models. The FEXP model is a long-memory generalization of the exponential model originally proposed by Bloomfield (1973). Under the exponential model, the logarithm of the spectral density of the process is assumed to have a finite order Fourier series expansion. The parameters in the model are the coefficients in this expansion. The FEXP model has been discussed by Janacek (1982), Beran (1993, 1994) and Robinson (1994), among others.

Recently, there has been interest in using FEXP models for semiparametric estimation of the long-memory parameter  $d$ . Moulines and Soulier (1999) (see also Hurvich & Brodsky, 2001) determined the asymptotic properties of the FEXP estimator  $\hat{d}_{\text{FEXP}}$ , obtained by fitting the model to the log periodogram by least-squares, and allowing the dimension of the model to grow slowly with the sample size  $n$ . Assuming that the series is Gaussian but does not obey a finite-dimensional FEXP model, Moulines and Soulier (1999) showed that  $\hat{d}_{\text{FEXP}}$  is asymptotically normal, with a mean squared error that can be as small as  $O[(\log n)/n]$  if the true spectral density is sufficiently smooth. This compares favorably with the log-periodogram regression estimator  $\hat{d}_{\text{GPH}}$  of Geweke and Porter-Hudak (1983) (justified theoretically by Robinson, 1995; Hurvich, Deo & Brodsky, 1998), which has mean squared error no better than  $O(n^{-4/5})$  in general. The efficiency of both estimators can be improved if the periodogram

is pooled (averaged) before taking logs, as shown by Robinson (1995) for  $\hat{d}_{\text{GPH}}$  and by Moulines and Soulier (1999) for  $\hat{d}_{\text{FEXP}}$ .

The fitting of FEXP models by least-squares regression of the log periodogram provides us not only with an estimator of  $d$ , but also with an estimator of the spectral density at all frequencies. Moulines and Soulier (2000) showed if the dimension,  $p$ , of the FEXP model is chosen to minimize Mallows' (1973)  $C_L$  criterion (see Li, 1987), then the resulting spectral estimator is asymptotically optimal in terms of mean integrated squared error of the log spectrum, once again assuming that the true process is not a finite-order FEXP. In view of the results of Shibata (1980, 1981), it seems plausible that the  $C_L$  method provides an optimal selection of FEXP models for one-step ahead forecasting. The question then arises: How do we compute the forecasts, and how should we select a model if we are interested in multistep forecasting? A perhaps even more basic question is: How do we calculate the coefficients in the one-sided infinite order autoregressive (AR( $\infty$ )) and moving average (MA( $\infty$ )) representations of the process? We will provide answers to these questions here.

The outline of the paper is as follows. First, we define the exponential model, and give recursive formulas for its AR( $\infty$ ) and MA( $\infty$ ) coefficients, as well as a fast algorithm for approximating its autocovariances. Next, we define the FEXP model and use the results described above to give fast algorithms for exact calculation of its AR( $\infty$ ) and MA( $\infty$ ) coefficients, and approximation of its autocovariances. Next, we present a fast frequency-domain cross validation method for selecting the order  $p$  of the FEXP model. This model selection method is designed to yield the value of  $p$  which provides the best multistep forecast for the given lead time, without assuming that the process actually obeys a FEXP model. We view the suggestion of this method

as merely a first step and defer research into its theoretical and practical performance to a later paper. We do, however, explore the practical performance of multistep forecasts of inflation in the United Kingdom, based on FEXP and ARFIMA models, without model selection. The FEXP multistep forecasts are computed using our algorithm for the AR( $\infty$ ) coefficients of the fitted FEXP model, and these forecasts differ substantially from the ARFIMA ones.

## 2. The exponential model

Bloomfield's (1973) exponential model gives the spectral density of a process  $\{Z_t\}$  as

$$g(\lambda) = \exp \left[ \sum_{k=0}^p \theta_k \cos(k\lambda) \right],$$

$$\lambda \in [-\pi, \pi] \quad (1)$$

where the model order  $p$  is a positive integer, and  $\theta_0, \dots, \theta_p$  are real-valued parameters. The process is stationary and invertible for all parameter configurations. For forecasting purposes, it is essential to derive a one-sided moving average representation of the process, of form

$$Z_t = \sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j}, \quad (2)$$

where  $\beta_0 = 1$  and  $\{\varepsilon_t\}$  is a white noise process such that  $E[\varepsilon_t] = 0$  and  $\text{var}[\varepsilon_t] = \sigma^2$  for all  $t$ , and  $E[\varepsilon_t \varepsilon_s] = 0$  if  $s \neq t$ . For notational simplicity, we assume that the mean of the series is zero. In practice, we may add the sample mean to a forecast based on the de-meaned data.

The construction of the MA( $\infty$ ) representation (2) proceeds by spectral factorization techniques developed by Wiener (1949) and Kolmogorov (1941), and described for example, by Priestley (1981, pp. 730–735). Many of the details which we present here for the model (1) were given in Bloomfield's (1973) original paper, although we

will provide some additional algorithmic details. It follows from Kolmogorov's formula (see Brockwell & Davis, 1991, p. 191) that the innovation variance  $\sigma^2$  and  $\theta_0$  are related by  $\theta_0 = \log[\sigma^2/(2\pi)]$ , so we can express the spectral density as

$$g(\lambda) = \frac{\sigma^2}{2\pi} \exp \left[ \sum_{k=1}^p \theta_k \cos(k\lambda) \right].$$

Define the complex-valued function

$$\xi(z) = \exp \left( \frac{1}{2} \sum_{k=1}^p \theta_k z^k \right)$$

where  $z$  is a complex variable. Note that  $\xi(0) = 1$ . It can be shown that all roots of  $\xi(z)$  lie outside the unit circle. Denoting the complex conjugate of  $z$  by  $\bar{z}$ , we obtain

$$\overline{\xi(z)} = \exp \left( \frac{1}{2} \sum_{k=1}^p \theta_k \bar{z}^k \right),$$

so that

$$|\xi(z)|^2 = \exp \left[ \sum_{k=1}^p \theta_k \text{Re}(z^k) \right],$$

where  $\text{Re}(z^k)$  is the real part of  $z^k$ . Now, define

$$A(\lambda) = \xi(e^{-i\lambda}) = \exp \left( \frac{1}{2} \sum_{k=1}^p \theta_k e^{-i\lambda k} \right)$$

so that

$$\frac{\sigma^2}{2\pi} |A(\lambda)|^2 = \frac{\sigma^2}{2\pi} \exp \left[ \sum_{k=1}^p \theta_k \cos(k\lambda) \right] = g(\lambda).$$

The MA( $\infty$ ) coefficients  $\{\beta_j\}$  of the process are the Fourier coefficients of  $A(\lambda)$ , or equivalently, the coefficients in the Taylor expansion  $\xi(z) = \sum_{j=0}^{\infty} \beta_j z^j$ . Therefore,

$$\beta_j = \frac{\xi^{(j)}(0)}{j!}, \quad (3)$$

where  $\xi^{(j)}(0)$  denotes the  $j$ th derivative of  $\xi(z)$  evaluated at  $z = 0$ . We have  $\xi(z) = \exp[h(z)]$  where  $h(z) = 1/2 \sum_{k=1}^p \theta_k z^k$ . Since  $\xi'(z) =$

$h'(z)\xi(z)$ , it follows from the product rule for differentiation that for  $j > 0$ ,

$$\xi^{(j)}(0) = \sum_{k=1}^j \binom{j-1}{k-1} h^{(k)}(0)$$

$$\xi^{(j-k)}(0) = \sum_{k=1}^j \binom{j-1}{k-1} \frac{1}{2} k! \theta_k \xi^{(j-k)}(0),$$

where we set  $\theta_k = 0$  if  $k > p$ . We conclude from (3) that  $\beta_0 = \xi(0) = 1$ , and

$$\beta_j = \frac{1}{2j} \sum_{k=1}^j k \theta_k \beta_{j-k} \quad (j > 0). \tag{4}$$

Similarly, we can obtain the  $AR(\infty)$  representation

$$\sum_{j=0}^{\infty} \alpha_j Z_{t-j} = \varepsilon_t, \tag{5}$$

where  $\alpha_0 = 1$ . The  $AR(\infty)$  coefficients  $\alpha_j$  are the Fourier coefficients of  $1/A(\lambda)$ , or equivalently, the coefficients in the Taylor expansion  $\tilde{\xi}(z) = \sum_{j=0}^{\infty} \alpha^j z^j$  where

$$\tilde{\xi}(z) = 1/\xi(z) = \exp\left(-\frac{1}{2} \sum_{k=1}^p \theta_k z^k\right),$$

so that

$$\alpha_j = \frac{\tilde{\xi}^{(j)}(0)}{j!}. \tag{6}$$

A similar argument to that given above yields the conclusion that  $\alpha_0 = \tilde{\xi}(0) = 1/\xi(0) = 1$ , and

$$\alpha_j = -\frac{1}{2j} \sum_{k=1}^j k \theta_k \alpha_{j-k} \quad (j > 0). \tag{7}$$

It should be emphasized that if  $j > p$ , the upper limit of summation in Eqs. (4) and (7) is  $p$ .

Using the  $MA(\infty)$  representation (2), we conclude that  $\{Z_t\}$  has autocovariances  $\tilde{c}_r = E[Z_t Z_{t-r}]$  given by

$$\begin{aligned} \tilde{c}_r &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \beta_j \beta_k E[\varepsilon_{t-j} \varepsilon_{t-r-k}] \\ &= \sigma^2 \sum_{j=r}^{\infty} \beta_j \beta_{j-r} \quad (r \geq 0). \end{aligned} \tag{8}$$

Since it can be shown that  $\beta_j \rightarrow 0$  exponentially fast as  $j$  increases, we can obtain a very good approximation to  $\tilde{c}_r$  by using the truncated version  $\tilde{c}_{r,N} = \sigma^2 \sum_{j=r}^{N-1} \beta_j \beta_{j-r}$  by taking  $N$  sufficiently large. (In practice, it should suffice to check that  $\beta_j$  is very nearly zero for  $j$  near  $N$ .) We may compute the  $\tilde{c}_{r,N}$  efficiently using the fast Fourier transform (FFT). By Bloomfield (1976, p. 166) we have

$$\begin{aligned} \tilde{c}_{r,N} &= (\sigma^2/N') \sum_{t=0}^{N'-1} \left| \sum_{j=0}^{N'-1} \beta'_j \right. \\ &\quad \times \exp(-i2\pi jt/N') \left. \right|^2 \exp(i2\pi tr/N') \\ &\quad r = 0, \dots, N-1 \end{aligned} \tag{9}$$

where  $N' = 2N$ ,  $\beta'_j = \beta_j$  for  $0 \leq j \leq N-1$  and  $\beta'_j = 0$  for  $N \leq j \leq N'-1$ . Thus,  $\{\tilde{c}_{r,N}\}_{r=0}^{N-1}$  may be computed in  $O(N \log N)$  operations by two applications of the FFT.

### 3. The FEXP model

We will say that the series  $\{X_t\}$  obeys a FEXP( $p$ ) model if the series has spectral density

$$f_p(\lambda) = \frac{\sigma^2}{2\pi} |1 - e^{-i\lambda}|^{-2d} \exp\left[\sum_{k=1}^p \theta_k \cos(k\lambda)\right],$$

where  $d \in (-0.5, 0.5)$ . Equivalently, we have  $X_t = (1 - B)^{-d} Z_t$  where  $\{Z_t\}$  is given by (2), and  $B$  is the backshift operator. Note that  $\{X_t\}$  has long memory if  $d \neq 0$ , while if  $d = 0$  then  $\{X_t\}$  has short memory and obeys an exponential model. Setting

$$\psi^*(B) = (1 - B)^{-d} = \sum_{j=0}^{\infty} \psi_j^* B^j$$

we obtain

$$X_t = \psi^*(B)Z_t = \sum_{j=0}^{\infty} \psi_j^* Z_{t-j}, \tag{10}$$

where  $\psi_j^*$  are the MA( $\infty$ ) coefficients of an ARFIMA(0,  $d$ , 0) model with unit innovation variance, given by Brockwell and Davis (1991, p. 522) as

$$\psi_j^* = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}.$$

The MA( $\infty$ ) representation for  $\{X_t\}$  is

$$X_t = \psi(B)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \tag{11}$$

where  $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ . Substituting (2) into (10) yields

$$X_t = \psi^*(B)\beta(B)\varepsilon_t, \tag{12}$$

where  $\beta(B) = \sum_{j=0}^{\infty} \beta_j B^j$ . Comparison of (11) and (12) yields

$$\psi(B) = \psi^*(B)\beta(B) = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \psi_k^* \beta_j B^{j+k}.$$

Thus, the MA( $\infty$ ) coefficients of  $\{X_t\}$  are

$$\psi_j = \sum_{k=0}^j \psi_k^* \beta_{j-k}. \tag{13}$$

We can use the FFT to efficiently compute  $\psi_0, \dots, \psi_{N-1}$  for any positive integer  $N$ . We start by evaluating  $\{\psi_j^*\}_{j=0}^{N-1}$  at a cost of  $O(N)$  operations, and evaluating  $\{\beta_j\}_{j=0}^{N-1}$  recursively from Eq. (4) at a cost of  $O(p^2 + pN)$  operations. Since the sequence  $\{\psi_j\}_{j=0}^{N-1}$  is the convolution of the sequences  $\{\psi_j^*\}_{j=0}^{N-1}$  and  $\{\beta_j\}_{j=0}^{N-1}$  it follows from Bloomfield (1976, p. 87) that we can now compute  $\psi_0, \dots, \psi_{N-1}$  in  $O(N \log N)$  operations using the FFT, as described in Appendix A.

Similarly, the AR( $\infty$ ) representation of  $\{X_t\}$  is  $\pi(B)X_t = \varepsilon_t$  where  $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$ . Setting  $\pi^*(B) = (1 - B)^d = \sum_{j=0}^{\infty} \pi_j^* B^j$ , we obtain

$$\pi^*(B)X_t = Z_t,$$

where  $\pi_j^*$  are the AR( $\infty$ ) coefficients of an ARFIMA(0,  $d$ , 0) model with unit innovation variance,

$$\pi_j^* = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}.$$

Writing (5) as  $\alpha(B)Z_t = \varepsilon_t$  where  $\alpha(B) = \sum_{j=0}^{\infty} \alpha_j B^j$  we conclude that  $\alpha(B)\pi^*(B)X_t = \pi(B)X_t$ , so that

$$\pi(B) = \alpha(B)\pi^*(B).$$

Thus, the AR( $\infty$ ) coefficients of  $\{X_t\}$  are

$$\pi_j = \sum_{k=0}^j \pi_k^* \alpha_{j-k}. \tag{14}$$

Again, the  $\pi_j$  can be efficiently calculated, as described in Appendix A.

Next, we present an algorithm to accurately approximate the autocovariances and to simulate realizations of  $\{X_t\}$ . The spectral density of  $\{X_t\}$  is  $f_p(\lambda) = |1 - e^{-i\lambda}|^{-2d} g(\lambda)$  where  $g(\lambda)$  is given by (1). We can express  $g(\lambda)$  in terms of the autocovariances of the exponential model, that is,

$$g(\lambda) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \tilde{c}_r \exp(ir\lambda),$$

where the  $\{\tilde{c}_r\}$  are given by (8) for  $r \geq 0$  and  $c_{-r} = c_r$ . Furthermore, we have

$$|1 - e^{-i\lambda}|^{-2d} = \sum_{s=-\infty}^{\infty} c_s^* \exp(is\lambda),$$

where the  $\{c_s^*\}$  are the autocovariances of an ARFIMA(0,  $d$ , 0) process with unit innovation variance, given (cf. Brockwell & Davis, 1991, p. 522) by

$$c_0^* = \frac{\Gamma(1 - 2d)}{\Gamma^2(1 - d)},$$

$$c_s^* = \frac{\Gamma(s + d)\Gamma(1 - 2d)}{\Gamma(s - d + 1)\Gamma(1 - d)\Gamma(d)}$$

$$(s > 0), \quad c_{-s}^* = c_s^*.$$

Thus,  $\{X_t\}$  has lag- $L$  autocovariance

$$\begin{aligned} c_L &= \int_{-\pi}^{\pi} f_p(\lambda) \exp(-iL\lambda) \, d\lambda \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \tilde{c}_r c_s^* \\ &\quad \times \exp\{i(r + s - L)\lambda\} \, d\lambda = \sum_{r=-\infty}^{\infty} \tilde{c}_r c_{L-r}^* \\ &= \tilde{c}_0 c_L^* + \sum_{r=1}^{\infty} \tilde{c}_r (c_{L-r}^* + c_{L+r}^*). \end{aligned} \tag{15}$$

The exponential model has short memory, and it can be shown that its autocovariances  $\{\tilde{c}_r\}$  decline to zero exponentially fast. Thus in practice it should be possible to determine a lag  $r_{\max}$  such that  $\sup_{r > r_{\max}} |\tilde{c}_r|$  is negligible for the purpose at hand. (We will not present a rigorous method for achieving this here, but in practice it should suffice to select  $r_{\max}$  such that  $|\tilde{c}_{r_{\max}}| < Tol$ , where  $Tol$  is a predetermined tolerance factor, e.g.,  $10^{-10}$ .) We can then approximate  $c_L$  by

$$c_{L, \text{Approx}} = \tilde{c}_0 c_L^* + \sum_{r=1}^{r_{\max}} \tilde{c}_r (c_{L-r}^* + c_{L+r}^*). \tag{16}$$

Since the choice of the truncation point  $r_{\max}$  in (16) is governed by the exponential decay of the  $\{\tilde{c}_r\}$ , we can obtain very accurate approximations to  $\{c_L\}_{L=0}^{n-1}$  using  $\{c_{L, \text{Approx}}\}_{L=0}^{n-1}$  for any integer  $n > 0$ . The idea described above was originally given by Bertelli and Caporin (in press) in the context of approximating the autocovariances of ARFIMA and other long-memory models. In this context, Bertelli and Caporin (in press) demonstrated that their algorithm is extremely accurate, giving results

which are correct to several decimal places. In practice, we would use  $\tilde{c}_{r,N}$  (9) in place of  $\tilde{c}_r$  in (16).

The sequence  $\{c_{L, \text{Approx}}\}_{L=0}^{n-1}$  can then be used in conjunction with the fast algorithm of Davies and Harte (1987) (see also Beran, 1994, pp. 216–217) to simulate a zero-mean Gaussian realization  $X_1, \dots, X_n$  from the FEXP process  $\{X_t\}$ . The algorithm is exact, in the sense that the autocovariances of the simulated realizations exactly match those given as input to the algorithm.

#### 4. Multistep forecasting using FEXP models with fixed coefficients

It is shown in Koopmans (1974, p. 229–231) that if a weakly stationary zero-mean series  $\{X_t\}$  has moving average coefficients  $\psi_j$  for  $j \geq 0$ , then the minimum mean squared error  $L$  step ahead linear forecast may be expressed as  $\hat{X}_L(t) = [\psi_L(B)/\psi(B)]X_t$ , where  $\psi_L(B) = \sum_{j=0}^{\infty} \psi_{j+L} B^j$ . This result, which may be proved using the spectral representation of  $\{X_t\}$ , gives the forecast as a linear combination of present and past values of the series, where the coefficient of  $X_{t-j}$  is equal to the coefficient of  $B^j$  in a Taylor series (about zero) for the function  $\psi_L(B)/\psi(B)$ . It follows that the mean squared error of the forecast  $\hat{X}_L(t)$  may be expressed as

$$\begin{aligned} E[X_{t+L} - \hat{X}_L(t)]^2 &= \int_{-\pi}^{\pi} \left| \sum_{j=0}^{L-1} \psi_j e^{-i\lambda(j-L)} / \psi(e^{-i\lambda}) \right|^2 f(\lambda) \, d\lambda, \end{aligned} \tag{17}$$

where  $f(\lambda)$  is the spectral density of  $\{X_t\}$ . It is noteworthy that (17) remains valid even if  $\psi(B)$  is not the  $MA(\infty)$  transfer function of  $\{X_t\}$ . As long as the function  $\psi(B)$  has a Taylor series expansion with coefficients  $\psi_j$ , and the function  $\psi_L(e^{-i\lambda})/\psi(e^{-i\lambda})$  has a one-sided Fourier series expansion whose partial sums converge in the space  $L^2(f)$  to  $\psi_L(e^{-i\lambda})/\psi(e^{-i\lambda})$ , then the pre-

dicator  $\hat{X}_L(t) = [\psi_L(B)/\psi(B)]X_t$  for any zero-mean weakly stationary series  $\{X_t\}$  with spectral density  $f(\lambda)$  is well-defined (but not necessarily optimal), and yields mean squared error given by (17). (See Theorem 4.10.1 of Brockwell & Davis, 1991, p. 154).

In particular, if we take  $\psi(B)$  to be the  $MA(\infty)$  transfer function of a FEXP( $p$ ) model with  $\{\psi_j\}$  given by (13), then  $|\psi(e^{-i\lambda})|^2 = (2\pi/\sigma^2)f_p(\lambda)$ , and we obtain

$$E[X_{t+L} - \hat{X}_L(t)]^2 = \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{j=0}^{L-1} \psi_j e^{-i\lambda(j-L)} \right|^2 f(\lambda) / f_p(\lambda) d\lambda, \quad (18)$$

as long as  $\{X_t\}$  is weakly stationary with mean zero and spectral density  $f(\lambda)$ , and the conditions on the  $\{\psi_j\}$  mentioned above are satisfied. It should be emphasized that (18) remains valid even if  $\{X_t\}$  is not FEXP( $p$ ).

### 5. Multistep forecasting of fitted FEXP models

Suppose we have data  $X_1, \dots, X_n$  from a weakly stationary process  $\{X_t\}$  with spectral density

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} f^*(\lambda) = |2 \sin(\lambda/2)|^{-2d} f^*(\lambda)$$

where  $f^*(\lambda)$  is a positive and continuous function on  $[-\pi, \pi]$ . The process  $\{X_t\}$  does not necessarily follow a FEXP model. Correspondingly,  $f^*(\lambda)$  is not necessarily the spectral density of an exponential model. Nevertheless, we are going to fit FEXP models to the data. Define the periodogram by

$$I_j = I(\omega_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^n X_t \exp(-i\omega_j t) \right|^2,$$

$$j = 1, \dots, \tilde{n},$$

where  $\omega_j = 2\pi j/n$  and  $\tilde{n} = \lfloor (n-1)/2 \rfloor$ . Let  $y =$

$\log I(\underline{\omega}) + \gamma$ , where  $\underline{\omega} = (\omega_1, \dots, \omega_{\tilde{n}})'$  and  $\gamma = 0.577216\dots$  is Euler's constant. Then we can write  $y = \mu + e$ , where  $\mu = \log f(\underline{\omega}) = -2d \log|2 \sin(\underline{\omega}/2)| + \log f^*(\underline{\omega})$  and  $e = \log[I(\underline{\omega}) / f(\underline{\omega})] + \gamma$ . In the case where  $\{X_t\}$  is Gaussian white noise, the entries  $e_j$  ( $j = 1, \dots, \tilde{n}$ ) of  $e$  are iid with mean zero and variance  $\pi^2/6$ .

We assume that

$$\log f^*(\lambda) = \sum_{k=0}^{\infty} \gamma_k \cos(k\lambda), \quad \lambda \in [-\pi, \pi],$$

where the  $\{\gamma_k\}$  are unknown constants. Thus,  $\log f^*(\underline{\omega}) = \sum_{k=0}^{\infty} \gamma_k V_k$ , where  $V_k = \cos(k\underline{\omega})$ . In the special case where  $\{X_t\}$  is FEXP( $p$ ), then  $\gamma_k = \theta_k$  for  $k = 0, \dots, p$  where  $\theta_k$  are the parameters in the exponential model (1) and  $\gamma_k = 0$  for  $k > p$ , so that  $\log f^*(\underline{\omega}) = \sum_{k=0}^p \theta_k V_k$ . The above discussion serves to motivate a regression estimator, constructed as follows.

For any given positive integer  $p$ , define  $X = [V_0, V_1, \dots, V_p, \tilde{V}]$ , where  $\tilde{V} = \log|2 \sin(\underline{\omega}/2)|$ . Define

$$\hat{\beta}_p^* = (X'X)^{-1}X'y = (\hat{\theta}_0, \dots, \hat{\theta}_p, -2\hat{d})', \quad (19)$$

which may be viewed as the ordinary least squares estimator of the parameter  $\beta_p^* = (\gamma_0, \gamma_1, \dots, \gamma_p, -2d)'$ . We will sometimes refer to  $\hat{\theta}_0$  as  $\hat{\sigma}^2/(2\pi)$ . A FEXP( $p$ ) forecast of  $X_{n+L}$  is now given by

$$\hat{X}_L(n, \hat{\beta}_p^*) = [\hat{\psi}_L(B)/\hat{\psi}(B)]X_n, \quad (20)$$

where  $\hat{\psi}_L(B)$  and  $\hat{\psi}(B)$  are versions of  $\psi_L(B)$  and  $\psi(B)$  as defined earlier, with the FEXP parameters  $(\theta_0, \dots, \theta_p, -2d)'$  replaced by  $\hat{\beta}_p^*$ .

We use the form (20) for the forecast in deriving our model selection criterion, but an explicit formula for the predictor coefficients in (20) would be difficult to obtain. Therefore, to actually calculate the forecast we suggest to use the estimated  $AR(\infty)$  coefficients  $\hat{\pi}_j = \sum_{k=0}^j \hat{\pi}_k^* \hat{\alpha}_{j-k}$  where  $\hat{\pi}_k^* = [\Gamma(k - \hat{d}) / \Gamma(k + 1)\Gamma(-\hat{d})]$ ,  $\hat{\alpha}_0 = 1$  and  $\hat{\alpha}_j = -1/(2j) \sum_{k=1}^j k \hat{\theta}_k \hat{\alpha}_{j-k}$  ( $j > 0$ ). (The  $\hat{\pi}_j$  may be efficient-

ly calculated as described in Appendix A). The forecast takes the form

$$\begin{aligned} \tilde{X}_1(n, \hat{\beta}_p^*) &= -\sum_{j=1}^n \hat{\pi}_j X_{n+1-j}, \\ \tilde{X}_L(n, \hat{\beta}_p^*) &= -\sum_{j=1}^{L-1} \hat{\pi}_j \tilde{X}_{L-j}(n, \hat{\beta}_p^*) \\ &\quad - \sum_{j=L}^{n+L-1} \hat{\pi}_j X_{n+L-j} \quad (L > 1) \end{aligned}$$

The two predictors  $\hat{X}_L$  and  $\tilde{X}_L$  may differ somewhat, due to truncation effects. We will ignore any differences here, however.

A practical question which arises at this point is: How should we select  $p$ ? Here, we describe a model selection method based on cross-validation in the frequency domain, originally proposed by Hurvich (1987) in the context of Multistep forecasting. A useful feature of this method is that the selected value of  $p$  is allowed to depend on the lead time,  $L$ .

We use the definition of the (final) prediction error as given by Akaike (1969). That is, the parameters of the model are first estimated on the series  $\{X_t\}$ . Then these estimated parameters are used to construct forecasts of an independent series  $\{Y_t\}$  having the same stochastic structure as  $\{X_t\}$ . The mean squared error for predicting  $Y_{n+L}$  on the basis of  $Y_n, Y_{n-1}, \dots$  is then computed, where the expectation is taken with respect to both  $\{X_t\}$  and  $\{Y_t\}$ . This results in a mean squared error of

$$\text{MSE}_L(p) = E E_Y [Y_{n+L} - \{\hat{\psi}_L(B)/\hat{\psi}(B)\} Y_n]^2,$$

where  $E$  and  $E_Y$  are the expectations with respect to  $\{X_t\}$  and  $\{Y_t\}$ . We conclude from Eq. (18) that

$$\begin{aligned} \text{MSE}_L(p) &= E \frac{\hat{\sigma}^2}{2\pi} \int_{-\pi}^{\pi} \\ &\times \left| \sum_{j=0}^{L-1} \hat{\psi}_j e^{-i\lambda(j-L)} \right|^2 f(\lambda)/\hat{f}_p(\lambda) d\lambda, \end{aligned} \quad (21)$$

where  $\hat{f}_p(\lambda) = \hat{\sigma}^2/2\pi |2 \sin(\lambda/2)|^{-2d} \exp[\sum_{k=1}^p \hat{\theta}_k \cos(k\lambda)]$ ,  $\hat{\psi}_j = \sum_{k=0}^j \hat{\psi}_k^* \hat{\beta}_{j-k}$ ,  $\hat{\psi}_k^* = [\Gamma(k + \hat{d})/\Gamma(k + 1)\Gamma(\hat{d})]$ ,  $\hat{\beta}_0 = 1$  and  $\hat{\beta}_j = 1/(2j) \sum_{k=1}^j k \hat{\theta}_k \hat{\beta}_{j-k}$  ( $j > 0$ ).

We now construct an approximately unbiased estimator of  $\text{MSE}_L(p)$  in (21) using frequency-domain cross validation. We start by defining  $\hat{\beta}_{p(j)}^*$ , obtained by omitting the  $j$ th observation  $y_j$  as well as the  $j$ th row of the  $X$  matrix from the regression in (19), that is,

$$\begin{aligned} \hat{\beta}_{p(j)}^* &= (X'_{(j)} X_{(j)})^{-1} X'_{(j)} y_{(j)} \\ &= (\hat{\theta}_{0(j)}, \dots, \hat{\theta}_{p(j)}, -2\hat{d}_{(j)})', \end{aligned}$$

where  $X_{(j)}$  is obtained from  $X$  by omitting the  $j$ th row, denoted by  $v_j'$ , and  $y_{(j)} = (y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_n)'$ . It follows from Chatterjee and Hadi (1988, p. 115) that

$$\hat{\beta}_{p(j)}^* = \hat{\beta}_p^* - (X'X)^{-1} v_j r_j / (1 - h_{jj}), \quad (22)$$

where  $r_j$  is the  $j$ th entry of the residual vector  $r = y - Hy$ ,  $H = X(X'X)^{-1}X'$  is the hat matrix, and  $h_{jj}$  is the  $j$ th diagonal element of  $H$ . If the series were Gaussian white noise, then  $\hat{\beta}_{p(j)}^*$  would be independent of  $I_j$ , and  $I_j$  would be exactly unbiased for  $f_j = f(\omega_j)$ . Thus, it may be hoped that an approximately unbiased estimator of  $\text{MSE}_L(p)$  is given by

$$\begin{aligned} \widehat{\text{MSE}}_L(p) &= \frac{4\pi}{n} \sum_{j=1}^{\tilde{n}} \frac{\hat{\sigma}_{(j)}^2}{2\pi} \\ &\times \left| \sum_{k=0}^{L-1} \hat{\psi}_{k(j)} e^{-i\omega_j(k-L)} \right|^2 I_j / \hat{f}_{p(j)}, \end{aligned} \quad (23)$$

where  $\hat{f}_{p(j)} = \hat{\sigma}_{(j)}^2/(2\pi) |2 \sin(\omega_j/2)|^{-2d_{(j)}} \exp[\sum_{k=1}^p \hat{\theta}_{k(j)} \cos(k\omega_j)]$ ,  $\hat{\sigma}_{(j)}^2/(2\pi) = \hat{\theta}_{0(j)}$ , and  $\{\hat{\psi}_{k(j)}\}_{k=0}^{L-1}$  are calculated as the MA coefficients of a FEXP model with FEXP parameters  $\hat{\beta}_{p(j)}^*$ . We may therefore select  $p$  to minimize  $\widehat{\text{MSE}}_L(\cdot)$ . For the sake of brevity, we will postpone the study of the empirical and theoretical properties of this model selection method to a later paper.



We conclude by discussing the efficient computation of (23). First, the matrix  $X$  can be computed in  $O(np)$  operations, and using the FFT,  $y$  may be computed in  $O(n \log n)$  operations. Next,  $\hat{\beta}_p^*$  (as well as  $\{h_{jj}\}_{j=1}^{\tilde{n}}$  and  $r$ ) may be computed in  $O(np^2 + p^3)$  operations. For each  $j$  in  $1, \dots, \tilde{n}$ , we may next compute  $\hat{\beta}_{p(j)}^*$  from (22) in  $O(p^2)$  operations, and then compute  $\{\hat{\psi}_{k(j)}\}_{k=0}^{L-1}$  in  $O(p^2 + Lp + L \log L)$  operations, and finally compute  $\hat{\sigma}_{(j)}^2 / (2\pi)$  and  $\hat{f}_{p(j)}$  in  $O(p)$  operations. Thus, the cost of computing (23) is  $O(n \log n + np^2 + p^3 + nLp + nL \log L)$  operations.

## 6. Example: UK inflation

In this section, we present an analysis of the Consumer Price Index (CPI)-based inflation rate for the United Kingdom, a monthly series

covering the dates from February 1957 to December 1997 (a total of 491 observations). The data were taken from the International Monetary Fund's International Financial Statistics CD-ROM. The inflation rate is defined as first difference of the logarithm of the CPI.

Our goal in this section is simply to exhibit an application of some of the methodology of the paper. We will not, however, attempt to carry out any model selection or diagnostics.

Model parameters were estimated from a subseries extending from February 1957 to April 1990 (the 399th observation of the full series). Fig. 1 plots the full series, with a vertical line at  $t = 399$  to mark the end of the subseries.

Prior to estimating model parameters, we seasonally adjusted the inflation data by removing a month effect. This was done by taking out the (subseries) average of all observations for

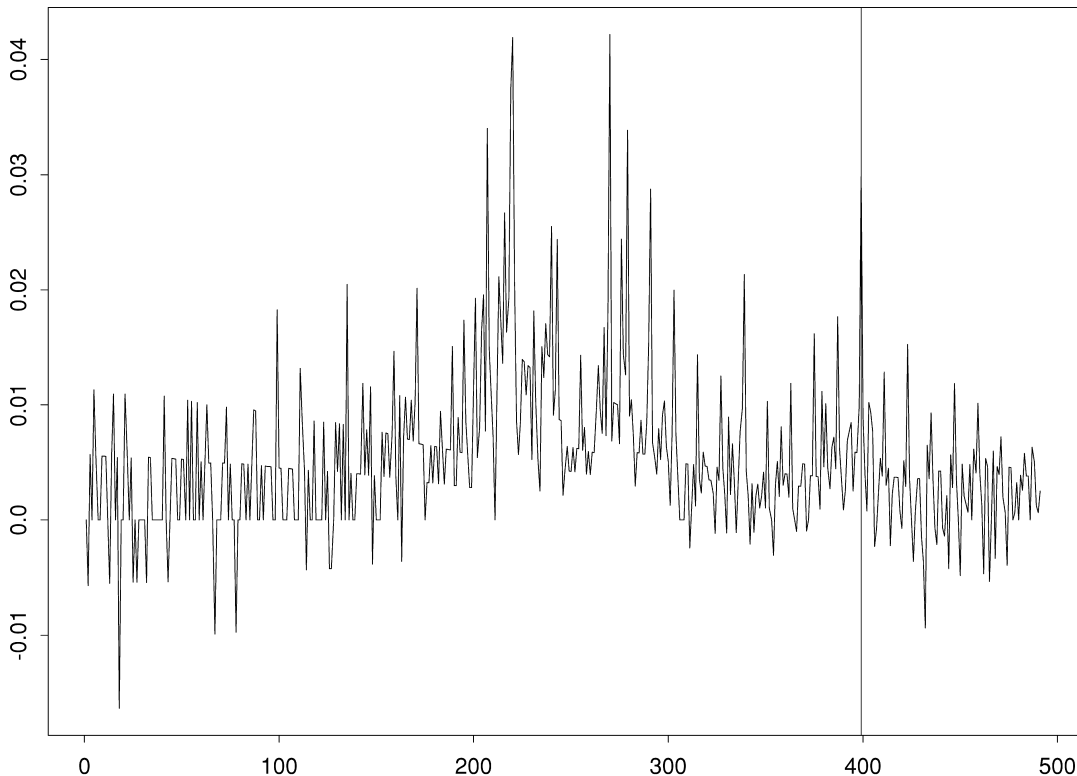


Fig. 1. Consumer Price Index (CPI)-based inflation rate for UK Monthly, February 1957 – December 1997 ( $n = 491$ ).

the given month. An ARFIMA(1,  $d$ , 0) model was fitted to the seasonally adjusted subseries, using the S-Plus function, `arma.fracdiff`. The estimated value of  $d$  was 0.175 and the estimated AR(1) parameter was 0.338. Next, a FEXP(1) model was fitted to the seasonally adjusted subseries, using log-periodogram regression on all available Fourier frequencies except the 133rd, which was omitted. This frequency corresponds to a period of 3 months, that is, one quarter year. The frequency was omitted on account of robustness considerations, as the log periodogram of the seasonally adjusted subseries was extremely negative at that ordinate. The resulting FEXP estimator of  $d$  was

0.496, which is much larger than the ARFIMA(1,  $d$ , 0) estimate of  $d$ . The estimated EXP(1) parameter in the exponential model was  $-0.292$ .

Multistep forecasts were constructed from a forecast horizon of April 1990, extending all the way to the end of the full series, December 1997. First, the seasonally adjusted series was forecasted, and then the (subseries) averages were added back to yield forecasts of inflation. For both models, the forecasts of the seasonally adjusted series were constructed from the  $AR(\infty)$  weights corresponding to the fitted model. Fig. 2 plots the actual (out of sample) inflation values, together with the FEXP and

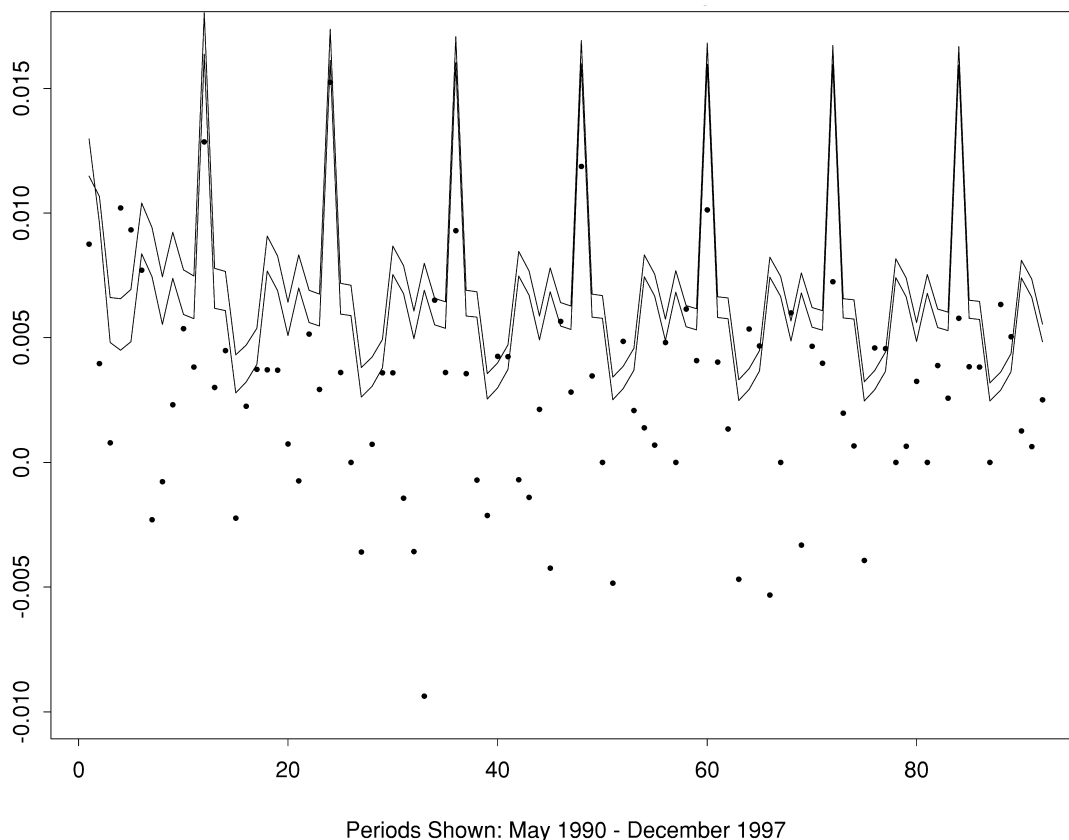


Fig. 2. Inflation, and Multistep FEXP and ARFIMA Forecasts from Forecast Horizon of April 1990.

ARFIMA forecasts. There are noticeable differences between the two forecasts. Except for the very first lead (May 1990), the FEXP forecast always lies above the ARFIMA forecast. This is presumably due to differences in the estimated memory parameters for the two models. The estimated seasonal effect has a noticeable peak in the month of April, and this effect continued for 4 to 5 years into the post-subseries observations.

Fig. 3 plots the seasonally adjusted (out of sample) inflation series, as well as the multistep FEXP and ARFIMA forecasts of this adjusted series. Here, it can be seen clearly that the FEXP forecasts decay much more slowly than

the ARFIMA forecasts as the lead time increases. Again, this occurs because the estimated value of  $d$  is much larger for the FEXP model than for the ARFIMA model.

It is not possible to judge which model provides better forecasts on the basis of one example. For the given data set and forecast horizon, in terms of average squared forecast error (ASE), the FEXP forecast was better than the ARFIMA model for the first five leads ( $ASE=2.11 \times 10^{-5}$  for FEXP,  $ASE=2.37 \times 10^{-5}$  for ARFIMA), while for the entire set of 92 leads, the ARFIMA model was much better ( $ASE=3.59 \times 10^{-5}$  for FEXP,  $ASE=2.75 \times 10^{-5}$  for ARFIMA). The overall superiority of

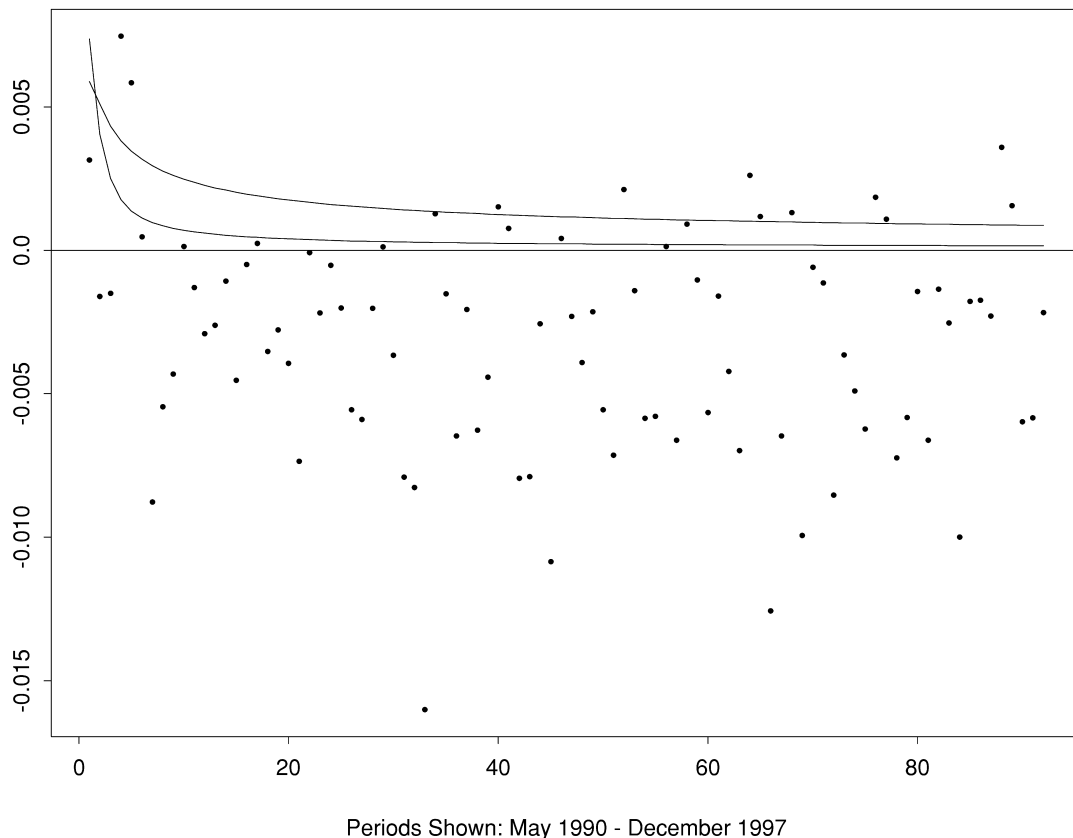


Fig. 3. Seasonally Adjusted Inflation, and Multistep FEXP and ARFIMA Forecasts from Forecast Horizon of April 1990.

the ARFIMA forecasts here is due to the fact that the seasonally adjusted inflation values were persistently negative for most of the out-of-subsample period. Since the inflation value at the time of the forecast horizon was quite high, the slower decay of the FEXP forecast was a liability, although it would have been an advantage if a later forecast horizon with a negative (adjusted) inflation value had been chosen.

## 7. Postscript

After the original version of this paper was submitted, we learned of the paper of Pourahmadi (1983). That paper contains recursive formulas which are generalizations of our formulas (4) and (7). We thank Professor M. Pourahmadi for pointing this out to us.

## Appendix A. Computing convolutions using the fast Fourier transform

Here, we show how to efficiently compute the convolutions in Eqs. (13) and (14), or the versions of these which use estimated parameters. For definiteness, we consider here computation of  $\psi_0, \dots, \psi_{N-1}$  given by (13). Note that for maximum computational efficiency,  $N$  should be chosen to be the smallest power of 2 that exceeds the number of coefficients actually desired. The algorithm proceeds as follows.

1. Append  $N$  zeros at the end of each of the sequences  $\{\psi_j^*\}_{j=0}^{N-1}$  and  $\{\beta_j\}_{j=0}^{N-1}$ .
2. Using an FFT program, compute the discrete Fourier transform of each of the resulting sequences. Here, the discrete Fourier transform of a sequence  $\{\delta_j\}_{j=0}^{N'-1}$  ( $N' = 2N$ ) is defined by  $\{J_{\delta,t}\}_{t=0}^{N'-1}$  where  $J_{\delta,t} = \sum_{j=0}^{N'-1} \delta_j \exp(-i2\pi jt/N')$ .
3. Compute the componentwise product of the two discrete Fourier transform sequences from step (2).

4. Take the inverse Fourier transform of the sequence in Step (3), where the inverse Fourier transform of a sequence  $\{J_t\}_{t=0}^{N'-1}$  is defined as the sequence  $\{W_j\}_{j=0}^{N'-1}$ , with  $W_j = (1/N') \sum_{t=0}^{N'-1} J_t \exp(i2\pi tj/N')$ .
5. The first  $N$  entries of the resulting sequence are  $\psi_0, \dots, \psi_{N-1}$ .

## References

- Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* 21, 243–247.
- Beran, J. (1993). Fitting long-memory models by generalized linear regression. *Biometrika* 80, 817–822.
- Beran, J. (1994). *Statistics for long-memory processes*, Chapman and Hall, New York.
- Beran, J., Bhansali, R. J., & Ocker, D. (1998). On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika* 85, 921–934.
- Bertelli, S., & Caporin, M. (in press). A note on calculating autocovariances of long-memory processes. *Journal of Time Series Analysis*.
- Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* 60, 217–226.
- Bloomfield, P. (1976). *Fourier analysis of time series: an introduction*, Wiley, New York.
- Brockwell, P. J., & Davis, R. A. (1991). *Time series: theory and methods*, 2nd ed., Springer-Verlag, New York.
- Brodsky, J., & Hurvich, C. M. (1999). Multi-step forecasting for long-memory processes. *Journal of Forecasting* 18, 59–75.
- Chatterjee, S., & Hadi, A. (1988). *Sensitivity analysis in linear regression*, Wiley, New York.
- Crato, N., & Ray, B. (1996). Model selection and forecasting for long-range dependent processes. *Journal of Forecasting* 15, 107–125.
- Davies, R. B., & Harte, D. S. (1987). Tests for Hurst effect. *Biometrika* 74, 95–101.
- Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221–237.
- Granger, C. W. J., & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–29.

- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* 68, 165–176.
- Hurvich, C. M. (1987). Automatic selection of a linear predictor through frequency domain cross-validation. *Communications in Statistical Theory and Methods* 16, 3199–3234.
- Hurvich, C. M., & Brodsky, J. (2001). Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential models. *Journal of Time Series Analysis* 22, 221–249.
- Hurvich, C. M., Deo, R. S., & Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. *Journal of Time Series Analysis* 19, 19–46.
- Janacek, G. J. (1982). Determining the degree of differencing for time series via the log spectrum. *Journal of Time Series Analysis* 3, 177–183.
- Kolmogorov, A. N. (1941). Interpolation und extrapolation von stationären zufälligen folgen (Russian, German summary). *Bulletin of the Academy of Science (Nauk), U.S.S.R., Mathematics Series* 5, 3–14.
- Koopmans, L. H. (1974). *The spectral analysis of time series*, Academic Press, New York.
- Li, K. -C. (1987). Asymptotic optimality for  $C_p$ ,  $C_L$ , cross-validation and generalized cross-validation. *The Annals of Statistics* 15, 958–975.
- Mallows, C. L. (1973). Some comments on  $C_p$ . *Technometrics* 15, 661–675.
- Moulines, E., & Soulier, P. (1999). Broad band log-periodogram regression of time series with long range dependence. *The Annals of Statistics* 27, 1415–1439.
- Moulines, E., & 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.
- Pourahmadi, M. (1983). Exact factorization of the spectral density and its application to forecasting and time series analysis. *Communications on Statistical Theory Methods* 12, 2085–2094.
- Priestley, M. B. (1981). *Spectral analysis and time series*, Academic Press, New York.
- Ray, B. K. (1993). Modeling long-memory processes for optimal long-range prediction. *Journal of Time Series Analysis* 14, 511–525.
- Robinson, P. M. (1994). Time series with strong dependence. In: *Advances in econometrics, sixth world congress, vol. I*, Cambridge University Press, pp. 47–95.
- Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics* 23, 1048–1072.
- 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.
- Shibata, R. (1981). An optimal autoregressive spectral estimate. *The Annals of Statistics* 9, 300–306.
- Tiao, G. C., & Tsay, R. S. (1994). Some advances in non-linear and adaptive modeling in time-series. *Journal of Forecasting* 13, 109–131.
- Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series*, Wiley, New York.

**Biography:** Clifford M. HURVICH is a professor of Statistics at New York University. He obtained a Ph.D. in Statistics from Princeton University in 1985 under John W. Tukey on the topic of time series. His research interests include forecasting, model selection and long-memory time series analysis.