

# Maximum likelihood estimation of stationary univariate fractionally integrated time series models\*

Fallaw Sowell

*Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA*

Received August 1988, final version received April 1991

To estimate the parameters of a stationary univariate fractionally integrated time series, the unconditional exact likelihood function is derived. This allows the simultaneous estimation of all the parameters of the model by exact maximum likelihood. Issues involved in obtaining maximum likelihood estimates are discussed. Particular attention is given to efficient procedures to evaluate the likelihood function, obtaining starting values, and the small sample properties of the estimators. Limitations of previous estimation procedures are also discussed.

## 1. Introduction

A time series implication of several economic and financial models is the absence of strong dependence between distant observations. Because nonzero values of the fractional differencing parameter implies this strong dependence, several researchers have tested these models by estimating fractional differencing parameters. The series that have been studied include: real gross national product [Diebold and Rudebusch (1989), Haubrich and Lo (1989)], interest rates [Shea (1990)], consumer and wholesale price indices [Geweke and Porter-Hudak (1982)], stock market returns [Aydogan and Booth (1988), Greene and Fielitz (1977)], stock market prices [Lo (1989)], futures prices for soybeans, soybean oil, and soybean meal [Helms et al. (1984)], and exchange rates [Cheung (1989), Booth et al. (1982)]. The wide use of the fractional

\* This is a generalization of Chapter 3 of the author's Ph.D. thesis, which was supported by the Alfred P. Sloan Foundation through a Doctoral Dissertation Fellowship. The additional work was supported by a Faculty Development Grant from Carnegie Mellon University. The author thanks John Geweke and two anonymous referees for helpful suggestions and Ken Murphy for computing assistance.

differencing model in empirical work highlights the importance of efficient estimation procedures for this model.

The importance of efficient estimation procedures has also been highlighted by recent developments in both econometric and economic theory. In econometric theory, both Gourieroux, Maurel, and Monfort (1987) and Sowell (1990a) have shown that the asymptotic distributions of regression estimators are very different for series that follow fractional differenced models. While in economic theory, Haubrach and Lo (1989), using a technique developed in Granger (1980), present a theoretical macroeconomic model which gives the implication that aggregate time series should be fractionally integrated. Both these lines of research highlight the need for efficient procedures to estimate fractionally integrated models.

The most efficient estimation procedure for fractionally differenced models is presumably maximum likelihood. Unfortunately, the perceived complexities associated with deriving the likelihood function and evaluating it repeatedly [see McLeod and Hipel (1978), Hosking (1984), and Brockwell and Davis (1987)] has meant that none of the empirical work cited above has used maximum likelihood. In this paper, this perception is shown to be unfounded. The unconditional exact likelihood function for a normally distributed stationary fractionally integrated time series is derived. Recursive procedures are noted which allow efficient evaluation of the likelihood function.

The next section is an introduction to fractionally integrated models. In section 3 previous estimation procedures are reviewed and their limitations are noted. The likelihood function is derived in section 4. The fifth section presents issues involved in obtaining maximum likelihood estimates. Particular attention is given to: efficient procedures to evaluate the likelihood function, obtaining starting values for the parameters, and the small sample properties of the estimators. The final section summarizes the paper and indicates directions for future research.

## 2. An introduction to fractional models

For  $d \in (-\frac{1}{2}, \frac{1}{2}]$ , the fractional difference operator  $(1-L)^d$  is defined by the Binomial Theorem as the series

$$(1-L)^d \equiv \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j L^j, \quad (1)$$

where the sequence of coefficients

$$\binom{d}{j} (-1)^j = \frac{\Gamma(d+1)(-1)^j}{\Gamma(d-j+1)\Gamma(j+1)} = \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)}$$

is square summable. For values of  $d > \frac{1}{2}$ ,  $(1-L)^d$  can be defined by combining integer differencing and eq. (1).

A time series,  $z_t$ , is said to follow a fractionally differenced model if, after applying the operator  $(1-L)^d$ , it follows an ARMA( $p, q$ ) model, where  $p$  and  $q$  are finite nonnegative integers. The time series  $z_t$  is said to be integrated of order  $d$ , which is denoted  $z_t \sim I(d)$ . Extending the standard notation  $z_t \sim I(d)$  can also be written  $z_t \sim \text{ARIMA}(p, d, q)$ . The general fractionally integrated time series model will be written as

$$\Phi(L)(1-L)^d z_t = \Theta(L)\varepsilon_t, \quad (2)$$

where

$$\Phi(L) = 1 + \phi_1 L + \phi_2 L^2 + \cdots + \phi_p L^p$$

and

$$\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$$

are polynomials in the lag operator  $L$ . Attention will be restricted to the class of models which satisfy the following assumptions:

*Assumption 1.*  $\Phi(L)$  is of order less than or equal to  $p$ ,  $\Theta(L)$  is of order less than or equal to  $q$ , the roots of  $\Theta(x)$  and  $\Phi(x)$  are outside the unit circle, and  $\varepsilon_t \sim \text{IIDN}(0, \sigma^2)$ .

*Assumption 2.*  $d < \frac{1}{2}$ .

*Assumption 3.* The roots of  $\Phi(x)$  are simple.

Assumption 1 is standard and requires no explanation. As noted below, Assumption 2 is required to restrict attention to stationary models. Assumption 3 is required for the procedure used to derive the likelihood function. The practical restrictions implied by Assumption 3 are investigated in section 5.4.

If Assumption 1 holds the distinguishing characteristics of the fractional  $z_t$  time series are:

- (1)  $z_t$  is stationary if  $d < \frac{1}{2}$ .
- (2)  $z_t$  possesses an invertible moving average representation if  $d > -\frac{1}{2}$ .

- (3) If the spectral density is denoted by  $f_z(\lambda)$ , then as  $\lambda \rightarrow 0$ ,  $f_z(\lambda) \sim \kappa^{-2d}$ , where  $\kappa$  is a constant and is independent of  $d$ .
- (4) If the autocovariances are denoted by  $\gamma(s) = E z_t z_{t-s}$ , then as  $s \rightarrow \infty$ ,  $\gamma(s) \sim k s^{2d-1}$ , where  $k$  is a function of  $d$ .

These results are proven in Granger and Joyeux (1980) or Hosking (1981) and the explicit forms of  $f_z(\lambda)$  and  $\gamma(s)$  are derived below.

The most useful feature of fractionally integrated time series are their long-range dependence. The dependence between observations produced by the ARMA structure of the model decays at a geometric rate, while the dependence produced by the fractional differencing parameter decays at the much slower hyperbolic rate. Hence the long-range dependence between observations is eventually determined only by the fractional differencing parameter. These characteristics can be seen in the shapes of the spectral density and the autocovariance function. If  $0 < d < \frac{1}{2}$ , the process eventually exhibits strong positive dependency between distant observations. This is noted in the frequency domain by the spectral density increasing to an infinite value as  $\lambda$  approaches zero. In the time domain, this long-range positive dependence is indicated by the fact that the autocovariances are eventual positive and decline slowly. If  $-\frac{1}{2} < d < 0$ , the process eventually exhibits negative dependency between distant observations. In the frequency domain, this is indicated by the decline of the spectral density to zero as  $\lambda$  approaches zero. In the time domain, this long-range negative dependence is indicated by the fact that the autocovariances are eventual negative and decline slowly.

### 3. Previous estimation techniques

Several estimation procedures for fractionally integrated time series have been proposed. Unfortunately the AR and MA parameters are typically not accurately estimated. This problem is avoided in this paper by estimating all the parameters by maximum likelihood. Before deriving the likelihood function, the previous estimation techniques will be reviewed and their inability to estimate the general model is noted.

There are two types of estimation procedures: two-step procedures and one-step procedures. The earliest were the two-step procedures. In the first step, the fractional differencing parameter is estimated and in the second step the other parameters of the model were estimated. The two-step procedures only differ in their first step. In the second step, the observed series is transformed and the remaining parameters of the model were estimated by standard time series procedures applied to the transformed series.

There are four ways of performing the first step of estimating the fractional differencing parameter. The first is the estimated rescaled range exponent,  $H$  [see McLeod and Hipel (1978) for a survey of this statistic]. The differencing parameter can be identified as  $\hat{d} = \hat{H} - \frac{1}{2}$ . A generalization, which is designed to be robust to 'short-term dependence', is given in Lo (1989). Two alternative estimation procedures exploit the special shape of the spectral density of a fractionally integrated time series. In Janacek (1982) the differencing parameter estimate is obtained by numerical integration of the log periodogram. Geweke and Porter-Hudak (1983) exploits the behavior of the spectral density around zero. Using frequencies near zero, a univariate regression of the log periodogram on the log of the frequencies is performed. The slope estimate is an estimate of the differencing parameter. The fourth procedure presented in Shea (1990) is maximum likelihood estimation of the regression model used in Geweke and Porter-Hudak (1983).

These procedures only estimate the differencing parameter. In the second step, the estimated differencing parameter is used to transform the observed series into a series that presumably follows an ARMA( $p, q$ ) model. If the model of the observed series is  $\Phi(L)(1-L)^d z_t = \Theta(L)\varepsilon_t$ , the estimated differencing parameter  $\hat{d}$  is used to try and obtain the series  $u_t = (1-L)^{\hat{d}} z_t$ , which would follow the model  $\Phi(L)u_t = \Theta(L)\varepsilon_t$ .

A problem is that  $(1-L)^d$  is defined as an infinite lag polynomial, hence the series  $u_t = (1-L)^d z_t$  can only be obtained when there exists an infinite realization of  $z_t$ . Given a finite sample, two different procedures have been proposed to approximate  $u_t$ ; one in the time domain and the other in the frequency domain. In the time domain the Binomial Theorem representation of  $(1-L)^d$  is used to transform the observed series. If  $u_t = (1-L)^d z_t$ , then

$$u_t = \sum_{j=0}^{\infty} \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)} z_{t-j}. \quad (3)$$

This suggests approximating  $u_t$  by using  $\hat{d}$  in the truncated polynomial, i.e., use  $\hat{d}$  for  $d$  in the above expression and set  $z_{t-j} = 0$  for  $t-j$  outside of the sample.

The alternative transformation is the frequency domain approach presented in Geweke and Porter-Hudak (1983). The approach is to calculate the Fourier transform of the observed series, multiply by the Fourier transform of the fractional differencing operator based on  $\hat{d}$ , and then calculate the inverse Fourier transform. It was hoped that the effect of this transformation is a series that follows the ARMA( $p, q$ ) model of the  $u_t$  process. The actual effect is identical to the time domain truncated polynomial using  $\hat{d}$ .

The equivalence of the frequency transform and the truncated time transformation applies to general linear transformations. To see this consider an arbitrary linear operator  $\sum_{j=0}^{\infty} \delta_j L^j$ . The frequency domain transformed series (denoted  $\bar{u}_s$ ) is calculated by taking the inverse Fourier transform of the product of the Fourier transforms of a linear operator  $\sum_{j=0}^{\infty} \delta_j L^j$  and the sample  $z_1, z_2, \dots, z_T$ :

$$\begin{aligned} \bar{u}_s &= \frac{1}{2\pi} \int_0^{2\pi} e^{is\lambda} \sum_{j=0}^{\infty} \delta_j e^{-ij\lambda} \sum_{t=1}^T z_t e^{-it\lambda} d\lambda \\ &= \frac{1}{2\pi} \int_0^{2\pi} e^{is\lambda} \sum_{k=1}^{\infty} \sum_{t=1}^{\min\{T, k\}} \delta_{k-t} z_t e^{-ik\lambda} d\lambda \\ &= \begin{cases} 0 & \text{if } s \leq 0 \\ \sum_{t=1}^{\min\{T, s\}} \delta_{s-t} z_t & \text{if } s \geq 1. \end{cases} \end{aligned}$$

For  $s = 1, 2, \dots, T$ , this is the truncated polynomial:

$$\bar{u}_s = \sum_{t=1}^s \delta_{s-t} z_t = \sum_{j=0}^{s-1} \delta_j z_{s-j}.$$

This is a well-known result in Fourier analysis that has apparently been overlooked by researchers working with fractional ARIMA models.

Consider the truncated time domain implementation of the fraction differencing transformation. For any finite realization (ignoring any bias in  $\hat{d}$  which only further complicates the estimation of the ARMA parameters),

$$\begin{aligned} \bar{u}_t &= u_t - \sum_{j=t}^{\infty} \frac{\Gamma(-\hat{d}+j)}{\Gamma(-\hat{d})\Gamma(j+1)} z_{t-j} \\ &= u_t - \sum_{k=0}^{\infty} \frac{\Gamma(-\hat{d}+t+k)}{\Gamma(-\hat{d})\Gamma(t+k+1)} z_k. \end{aligned}$$

After applying either transformation, the resulting series is the sum of a series which follows an ARMA model and a linear combination of an infinite number of unobserved terms. The transformed series does not have the ARMA model of the  $u_t$  series, hence the ARMA parameters will not be correctly estimated in the second step. This problem can be avoided by estimating all the parameters of the model in one step.

Two one-step procedures have previously been proposed, one in Li and McLeod (1986) and the other in Fox and Taquq (1986). In Li and McLeod (1986) the procedure suggested is to truncate the infinite sum that defines  $(1-L)^d$  and use standard time series estimation procedures. Truncating the infinite series is equivalent to the two-step procedure which leads to misleading estimates. The procedure proposed in Fox and Taquq (1986) is an approximation to the maximization of the normal likelihood function. The estimates are consistent and asymptotically normal. However, as noted in Dahlhaus (1988), the small sample behavior of this type of estimator is poor if the spectrum of the series contains peaks near zero, which is known to exist for positive values of the fractional differencing parameter. Dahlhaus (1988) concludes that exact maximum likelihood is optimal.

#### 4. The normal likelihood function

Consider a stationary normally distributed fractionally integrated time series  $z_t$  generated by the model given by the eq. (2) which satisfies Assumptions 1, 2, and 3. Let  $Z_T$  be a sample of  $T$  observations such that  $Z_T = [z_1 \ z_2 \ \dots \ z_T]'$  and  $Z_T \sim N_T(0, \Sigma)$ , with probability density function

$$f(Z_T, \Sigma) = (2\pi)^{-T/2} |\Sigma|^{-1/2} \exp\{-\frac{1}{2} Z_T' \Sigma^{-1} Z_T\}.$$

Stationarity implies that the covariance matrix is a Toeplitz form

$$\Sigma = [\gamma(i-j)] \quad \text{for } i, j = 1, 2, \dots, T.$$

To estimate the parameters of the model by maximum likelihood requires evaluating the likelihood function for a given set of parameter values. This requires writing the covariance matrix or, equivalently, the autocovariance function in terms of the parameters of the model. This parameterization of the autocovariance function is derived by writing the spectral density of  $z_t$  in term of the parameters of the model and then calculating the autocovariance function by

$$\gamma(s) = \frac{1}{2\pi} \int_0^{2\pi} f_z(\lambda) e^{i\lambda s} d\lambda. \quad (4)$$

The spectral density of  $z_t$  will be calculated in two steps, first the spectral density of  $u_t = (1-L)^d z_t$  is calculated and then the spectral density of  $z_t$ . Assumptions 1 and 2 imply  $u_t$  is generated by the ARMA( $p, q$ ) process

$\Phi(L)u_t = \Theta(L)\varepsilon_t$ . The Wold representation of  $u_t$  is

$$u_t = \frac{\Theta(L)}{\Phi(L)} \varepsilon_t.$$

The roots of  $\Phi(L)$  are assumed to be outside of the unit circle hence  $\Phi(x)$  can be written

$$\Phi(x) = \prod_{j=1}^p (1 - \rho_j x),$$

where  $|\rho_n| < 1$  for  $n = 1, 2, \dots, p$ . The spectral density of  $u_t$  is

$$f_u(\lambda) = \frac{|\Theta(\omega)|^2}{|\Phi(\omega)|^2} \sigma^2 = |\Theta(\omega)|^2 \sigma^2 \prod_{j=1}^p (1 - \rho_j \omega)^{-1} (1 - \rho_j \omega^{-1})^{-1},$$

where  $\omega = e^{i\lambda}$ . By Assumption 3 it is possible to use the partial fraction decomposition, to write

$$f_u(\lambda) = |\Theta(\omega)|^2 \sigma^2 \sum_{j=1}^p \omega^p \zeta_j \left[ \frac{\rho_j^{2p}}{(1 - \rho_j \omega)} - \frac{1}{(1 - \rho_j^{-1} \omega)} \right],$$

where

$$\zeta_j = \left[ \rho_j \prod_{i=1}^p (1 - \rho_i \rho_j) \prod_{m \neq j} (\rho_j - \rho_m) \right]^{-1}.$$

Because  $\Theta(\omega)$  is of finite order, each element of  $|\Theta(\omega)|^2$  can be written as a polynomial in  $\omega$  with both positive and negative values of the exponents and the spectral density of  $u_t$  can be written

$$f_u(\lambda) = \sigma^2 \sum_{l=-q}^q \psi(l) \omega^l \sum_{j=1}^p \omega^p \zeta_j \left[ \frac{\rho_j^{2p}}{(1 - \rho_j \omega)} - \frac{1}{(1 - \rho_j^{-1} \omega)} \right], \quad (5)$$

where

$$\psi(l) = \sum_{s=\max\{0, l\}}^{\min\{q, q-l\}} \theta_s \theta_{s-l}.$$



Eq. (5) is one representation of the spectral density of an ARMA( $p, q$ ) model where the AR process has simple roots.

Given the definition of  $u_t$ , the spectral density of  $z_t$  can be written

$$f_z(\lambda) = (1 - \omega)^{-d} (1 - \omega^{-1})^{-d} f_u(\lambda).$$

Upon substituting in for  $f_u(\lambda)$ ,

$$f_z(\lambda) = \sigma^2 \sum_{l=-q}^q \sum_{j=1}^p \psi(l) \zeta_j \left[ \frac{\rho_j^{2p}}{(1 - \rho_j \omega)} - \frac{1}{(1 - \rho_j^{-1} \omega)} \right] \\ \times (1 - \omega)^{-d} (1 - \omega^{-1})^{-d} \omega^{p+l}. \quad (6)$$

This is the general form of the spectral density for a time series which is generated by a stationary fractional ARMA( $p, d, q$ ) model where the roots of the AR polynomial are simple.

$\gamma(s)$  can be calculated by substituting eq. (6) into eq. (4). From the structure of the spectral density this reduces to the evaluation of the integral<sup>1</sup> (recall  $\omega = e^{-i\lambda}$ )

$$C(d, h, \rho) = \frac{1}{2\pi} \int_0^{2\pi} \left[ \frac{\rho^{2p}}{(1 - \rho e^{-i\lambda})} - \frac{1}{(1 - \rho^{-1} e^{-i\lambda})} \right] \\ \times (1 - e^{-i\lambda})^{-d} (1 - e^{i\lambda})^{-d} e^{-i\lambda h} d\lambda. \quad (7)$$

The autocovariance function can be written

$$\gamma(s) = \sigma^2 \sum_{l=-q}^q \sum_{j=1}^p \psi(l) \zeta_j C(d, p + l - s, \rho_j). \quad (8)$$

If  $p = 0$ , the partial fraction decomposition need not be considered and the autocovariances have the form

$$\gamma(s) = \sigma^2 \sum_{l=-q}^q \psi(l) \frac{\Gamma(1 - 2d) \Gamma(d + s - l)}{\Gamma(d) \Gamma(1 - d) \Gamma(1 - d - s + l)}.$$

<sup>1</sup>If  $p = 0$ , the term in brackets equals one; this reduces to the integral evaluated in appendix 2. Also, if  $p = 0$ , the  $j$  indexed sum and the terms  $\zeta_j$  will not in the expression of the spectral density.

The autocovariances for  $\rho \neq 0$  have been written in terms of  $C(d, h, \rho)$  and this function has been defined as an integral. To efficiently calculate the autocovariances, an alternative form will be derived. Note that the variable  $h$  can take on any integer value,  $\rho$  can be any complex number in the unit circle, and  $d$  is restricted to real values less than  $\frac{1}{2}$ . Using geometric series expansions,  $C(d, h, \rho)$  can be written

$$\begin{aligned} & \rho^{2p} \sum_{m=0}^{\infty} \rho^m \frac{1}{2\pi} \int_0^{2\pi} (1 - e^{-i\lambda})^{-d} (1 - e^{i\lambda})^{-d} e^{-i\lambda(h+m)} d\lambda \\ & + \sum_{n=1}^{\infty} \rho^n \frac{1}{2\pi} \int_0^{2\pi} (1 - e^{-i\lambda})^{-d} (1 - e^{i\lambda})^{-d} e^{-i\lambda(h-n)} d\lambda. \end{aligned}$$

The interchange of the sum and integral is justified in appendix 1. The problem is reduced to the following evaluation, which is derived in appendix 2,

$$\int_0^{2\pi} (1 - e^{-i\lambda})^{-d} (1 - e^{i\lambda})^{-d} e^{-i\lambda h} d\lambda = \frac{\Gamma(1-2d)(-1)^h}{\Gamma(1-d-h)\Gamma(1-d+h)}.$$

With this it is possible to write

$$\begin{aligned} C(d, h, \rho) &= \Gamma(1-2d) \left[ \rho^{2p} \sum_{m=0}^{\infty} \frac{\rho^m (-1)^{h+m}}{\Gamma(1-d+h+m)\Gamma(1-d-h-m)} \right. \\ & \quad \left. + \sum_{n=1}^{\infty} \frac{\rho^n (-1)^{h-n}}{\Gamma(1-d+h-n)\Gamma(1-d-h+n)} \right] \\ &= \frac{\Gamma(1-2d)\Gamma(d+h)}{\Gamma(1-d+h)\Gamma(1-d)\Gamma(d)} \\ & \quad \times [\rho^{2p} F(d+h, 1; 1-d+h; \rho) \\ & \quad + F(d-h, 1; 1-d-h; \rho) - 1], \end{aligned} \tag{9}$$

where  $F(a, b; c; x)$  is the hypergeometric function. The last equality is derived in appendix 3. The advantage of this final form is that the needed hypergeometric functions can be quickly and accurately approximated on a computer (see appendix 3).

## 5. Maximum likelihood estimates

### 5.1. Evaluating the log-likelihood function

Using the formulas presented above, the log-likelihood function is easily evaluated on a computer.<sup>2</sup> The main parts of the program are:

- (1) Factor the autoregressive polynomial.
- (2) Calculate the  $\zeta_j$ 's.
- (3) Calculate the different  $C(d, h, \rho)$  values.
- (4) Evaluate the covariance matrix.
- (5) Calculate the Cholesky decomposition and determinant of the inverse of the covariance matrix.
- (6) Calculate the log likelihood function value.

Important savings in time are made by separating steps 3 and 4. Because the same  $C(d, h, \rho)$  term appears in several autocovariances, all the  $C(d, h, \rho)$  terms are calculated, saved in memory, and then recalled in step 4. Step 3 is relatively quick because, as noted in appendix 3, the  $C(d, h, \rho)$  terms can be calculated by a recursive formula. In step 5, Levinson's algorithm [see Sowell (1989)] is used to recursively calculate the terms of the Cholesky decomposition and determinant of the inverse of the covariance matrix; hence as  $T$  grows the number of calculations grows at the rate  $T^2$  instead of  $T^3$ .

### 5.2. Starting values

When using a numerical optimization algorithm to maximize the log-likelihood function, concern must be given to the selection of starting values for the model's parameters.<sup>3</sup> The log-likelihood function is not globally concave; hence, the results of a search algorithm depend on the choice of starting values. No single procedure to obtain starting values will be optimal for all models; therefore different procedures are presented.

The procedures given below to obtain starting values for fractional ARIMA( $p, d, q$ ) models use a procedure to obtain starting values for a

<sup>2</sup>A FORTRAN program that calculates the maximum likelihood estimates is available from the author.

<sup>3</sup>These comments apply not only to the maximum likelihood procedure but also to the procedure presented in Fox and Taqqu (1986).

nonfractional ARMA( $p, q$ ) model. Several alternative procedures are available for the nonfractional ARMA( $p, q$ ) models [see ch. 6 of Hannan and Deistler (1988)] and any of these could be used. The specific procedure used below is described in Box and Jenkins (1976) and is widely accessible through the subroutine NSPE in the International Mathematical and Statistical Libraries (IMSL).

*Procedure 1.* First estimate  $d$  using the GPH procedure (or the R/S procedure, etc.). Start  $d$  at the estimated value and start the initial ARMA parameter and the innovation variance at the parameters estimated by NSPE for the series  $(1-L)^d z_t$ . Where the transformation  $(1-L)^d$  is applied by the truncated polynomial.

*Procedure 2.* Choose a grid of  $d$  values. For each value of  $d$  calculate the series  $(1-L)^d z_t$ , and estimate the ARMA parameters and the innovation variance by NSPE. Start the parameters at the values associated with the smallest estimate of the innovation variance.

For small samples Procedure 1 suffers from the fact that the individual estimates of  $d$  are extremely imprecise; while Procedure 2 suffers from the imprecise estimate of the innovation variance. Experience has shown that in samples sizes as large as 200, Procedure 1 and Procedure 2 (using NSPE) yield comparable final parameter estimates. In practice, it is important to choose several different starting values to be confident that the global maximum has been achieved.

### 5.3. Small sample properties

Sufficient conditions for the consistency and asymptotic normality of the exact maximum likelihood estimates are presented in Dahlhaus (1989).<sup>4</sup> However, because the fractional differencing parameter captures long-cycle characteristics of a series, asymptotic properties of maximum likelihood estimators may be of questionable use in small samples. To discover the small sample properties, Monte Carlo simulation was used to compare different estimation procedures. The parameter of simulated samples were estimated by the maximum likelihood procedure (MAX), the Fox and Taquq (1986) procedure (F&T), and the Geweke and Porter-Hudak (1983) procedure<sup>5</sup> (GPH). The GPH only estimates the  $d$  parameters, while both MAX and F&T estimate all the parameters of the model. The log-likelihood function and the negative of the objective function in Fox and Taquq were maximized

<sup>4</sup>All conditions are fulfilled for fractional Gaussian noise and fractional ARMA-processes [Dahlhaus (1989, p. 1751)].

<sup>5</sup>For the GPH procedure  $\alpha$  was set equal to  $\frac{1}{2}$ .

Table 1  
Estimated parameter bias and square root of the mean squared error for the ARIMA(0,  $d$ , 0) model ( $T = 100$ , 100 replications).

$d$	BIAS			$\sqrt{\text{MSE}}$		
	MAX	F&T	GPH	MAX	F&T	GPH
-0.4	-0.017	-0.043	0.032	0.101	0.104	0.301
-0.3	-0.013	-0.044	0.020	0.089	0.100	0.327
-0.2	-0.007	-0.040	0.013	0.077	0.088	0.311
-0.1	-0.009	-0.055	0.016	0.079	0.096	0.274
0.0	-0.009	-0.048	0.002	0.079	0.094	0.316
0.1	-0.017	-0.057	0.010	0.067	0.086	0.308
0.2	-0.012	-0.046	0.013	0.092	0.122	0.299
0.3	-0.015	-0.056	-0.032	0.077	0.097	0.326
0.4	-0.018	-0.034	0.085	0.060	0.097	0.254

using the Davidon-Fletcher-Powell algorithm [Powell (1971)] in the DFP subroutine of GQOPT.<sup>6,7</sup>

The first model considered is  $(1 - L)^d z_t = \varepsilon_t$ . For each  $d$  value samples with  $T = 100$  were generated. Table 1 presents the results for this model. All three procedures perform equally well in terms of bias. For MAX and F&T the level of the bias does not depend on the parameter value, while for GPH the bias appears to increase as the parameter value moves away from zero. For all the models considered the  $\sqrt{\text{MSE}}$  was smallest for the maximum likelihood estimators. A striking feature is that the  $\sqrt{\text{MSE}}$  of the GPH estimates are three to four times larger than for the maximum likelihood estimates.

The results for the ARIMA(1,  $d$ , 0) model  $(1 + \phi L)(1 - L)^d x_t = \varepsilon_t$  are presented in table 2 and the results for the ARIMA(0,  $d$ , 1) model  $(1 - L)^d x_t = (1 + \theta L)\varepsilon_t$  are presented in table 3. On average all three procedures appear to have bias of comparable orders of magnitude. For all three procedures the bias in  $d$  was greater for negative values of  $\phi$  and  $\theta$  than for positive values. This occurred for all the values of  $d$ . The source of the problem is that as  $\phi$  or  $\theta$  approach negative one, the operators  $(1 + \phi L)$  and  $(1 + \theta L)$  tend to either cancel out or reinforce the  $(1 - L)^d$  operator.

For the ARIMA(1,  $d$ , 0) model the smallest bias was achieved by GPH for seven of the models, by MAX for five of the models, and by F&T for one

<sup>6</sup>GQOPT is a FORTRAN library of numerical optimization algorithms distributed by Richard Quandt at the Department of Economics, Princeton University.

<sup>7</sup>The initial parameter values were chosen by fixing  $d = 0$  and starting the ARMA parameters and the innovation variance selected by NSPE. The final results reported are not sensitive to the starting values. These starting values were chosen to save time for the thousands of different starting values required for the Monte Carlo. In practice researchers should use either Procedure 1 or Procedure 2.

Table 2  
Estimated parameter bias and square root of the mean squared error for the ARIMA(1, d, 0) model (T = 100, 100 replications).

d	φ	BIAS						√MSE					
		MAX		F&T		GPH		MAX		F&T		GPH	
		$\hat{d}$	$\hat{\phi}$	$\hat{d}$	$\hat{\phi}$	$\hat{d}$	$\hat{\phi}$	$\hat{d}$	$\hat{\phi}$	$\hat{d}$	$\hat{\phi}$	$\hat{d}$	$\hat{\phi}$
-0.3	-0.7	0.102	0.138	-0.035	0.042	0.311	0.233	0.225	0.233	0.258	0.232	0.345	
-0.3	-0.2	-0.118	-0.094	-0.408	-0.337	0.046	0.264	0.242	0.264	0.352	0.342	0.296	
-0.3	0.3	-0.037	-0.047	-0.130	-0.128	0.040	0.206	0.185	0.206	0.296	0.297	0.334	
-0.3	0.8	-0.021	-0.017	-0.047	-0.034	0.013	0.062	0.095	0.062	0.114	0.060	0.318	
0.0	-0.7	0.002	0.022	-0.108	-0.050	0.283	0.157	0.172	0.157	0.142	0.093	0.308	
0.0	-0.2	-0.100	-0.064	-0.325	-0.253	0.055	0.248	0.235	0.248	0.352	0.325	0.292	
0.0	0.3	-0.019	-0.006	-0.121	-0.088	-0.017	0.114	0.102	0.114	0.229	0.235	0.301	
0.0	0.8	0.000	0.005	-0.046	0.024	-0.009	0.064	0.088	0.064	0.102	0.070	0.314	
0.3	-0.7	-0.137	-0.085	-0.145	-0.059	0.283	0.124	0.155	0.124	0.147	0.103	0.276	
0.3	-0.2	-0.096	-0.081	-0.521	-0.452	0.069	0.232	0.197	0.232	0.303	0.289	0.275	
0.3	0.3	-0.026	-0.017	-0.109	-0.088	-0.010	0.125	0.096	0.125	0.240	0.248	0.274	
0.3	0.8	-0.033	-0.018	-0.073	-0.035	-0.020	0.067	0.089	0.067	0.103	0.072	0.305	

Table 3  
Estimated parameter bias and square root of the mean squared error for the ARIMA( $\theta, d, 1$ ) model ( $T = 100, 100$  replications).

$d$	$\theta$	MAX		BIAS		F&T		GPH		MAX		$\sqrt{\text{MSE}}$		F&T		GPH	
		$\hat{d}$	$\hat{\theta}$	$\hat{d}$	$\hat{\theta}$	$\hat{d}$	$\hat{\theta}$	$\hat{d}$	$\hat{\theta}$	$\hat{d}$	$\hat{\theta}$	$\hat{d}$	$\hat{\theta}$	$\hat{d}$	$\hat{\theta}$	$\hat{d}$	$\hat{\theta}$
-0.3	-0.8	-0.029	0.019	-0.094	0.142	-0.242	0.193	0.215	0.193	0.260	0.262	0.311	0.260	0.262	0.311	0.260	0.262
-0.3	-0.3	0.016	-0.019	-0.135	0.125	-0.004	0.248	0.228	0.248	0.210	0.212	0.300	0.210	0.212	0.300	0.210	0.212
-0.3	0.5	-0.016	0.012	-0.070	0.032	0.044	0.111	0.096	0.111	0.105	0.109	0.278	0.105	0.109	0.278	0.105	0.109
-0.3	0.9	-0.010	0.001	-0.048	-0.030	-0.013	0.048	0.097	0.048	0.103	0.068	0.259	0.103	0.068	0.259	0.103	0.068
0.0	-0.8	-0.077	0.067	-0.328	0.295	-0.461	0.206	0.223	0.206	0.353	0.352	0.282	0.353	0.352	0.282	0.353	0.352
0.0	-0.3	-0.019	-0.004	-0.190	0.156	-0.086	0.237	0.198	0.237	0.192	0.182	0.307	0.192	0.182	0.307	0.192	0.182
0.0	0.5	-0.013	0.012	-0.064	0.027	-0.015	0.118	0.100	0.118	0.119	0.117	0.275	0.119	0.117	0.275	0.119	0.117
0.0	0.9	0.012	-0.011	-0.037	-0.052	0.017	0.080	0.080	0.017	0.080	0.049	0.295	0.080	0.079	0.295	0.080	0.079
0.3	-0.8	-0.180	0.173	-0.515	0.481	-0.486	0.259	0.245	0.259	0.277	0.305	0.281	0.277	0.305	0.281	0.277	0.305
0.3	-0.3	-0.064	0.059	-0.178	0.166	-0.091	0.142	0.142	0.091	0.185	0.185	0.294	0.185	0.197	0.294	0.185	0.197
0.3	0.5	-0.012	0.023	-0.050	0.030	0.044	0.088	0.088	0.044	0.117	0.118	0.320	0.117	0.118	0.320	0.117	0.118
0.3	0.9	-0.007	-0.005	-0.043	0.064	0.039	0.065	0.065	0.039	0.096	0.075	0.303	0.096	0.075	0.303	0.096	0.075

model. For the ARIMA(0,  $d$ , 1) model the smallest bias was achieved by MAX for eleven of the models and by GPH for the other model. The  $\sqrt{\text{MSE}}$  of the  $d$  parameter tended to be smallest for the maximum likelihood estimator. The  $\sqrt{\text{MSE}}$  for the maximum likelihood estimates was always smaller than for the GPH estimates, and for 21 of the 24 models considered the  $\sqrt{\text{MSE}}$  for MAX was smaller than for F&T. Generally, the maximum likelihood estimator was superior to the procedures presented in Fox and Taquq (1986) and in Geweke and Porter-Hudak (1982).

#### 5.4. Roots of the autoregressive polynomial

One possible limitation of the maximum likelihood estimation procedure is that at a *theoretical level* it is limited to models which do not have repeated roots in the autoregressive polynomial, see Assumption 3. This is because the partial fraction decomposition used to derive the spectral density of the  $u_t$  process is restricted to simple roots in the AR polynomial. If the population parameters of the model actually contains repeated roots, there is the possibility that the procedure used to evaluate the likelihood function may break down. This problem is unique to the procedure outlined in this paper and is not a problem for some alternative estimation procedures such as F&T or GPH.

This problem was investigated by simulation. Series were generated from models which contained repeated roots in the AR polynomial and their parameters estimated with the maximum likelihood procedure. The results of the simulation are reported in table 4. The MAX procedure did not break down for any of simulated samples. This was not a result of imprecise estimation; of twelve parameters estimated, the smallest bias was achieved by MAX for eleven of the parameters and the smallest  $\sqrt{\text{MSE}}$  was achieved by MAX for nine of the parameters. As with the ARIMA(1,  $d$ , 0) model the bias and the  $\sqrt{\text{MSE}}$  increased as the roots of the AR polynomial approached negative one.

The restriction that the roots of the autoregressive polynomial be simple does not appear to be a binding restriction at an *empirical level*. This is because in the space of polynomials of a given order, the subset which has repeated roots is a set with zero Lebesgue measure. Hence, numerical search routines choose AR parameter values with repeated roots with probability zero. The maximum likelihood procedure is well defined over a large enough subset of the parameter space to accurately estimate models with repeated roots.

#### 5.5. Computation time

The addition of the fractional differencing parameter to the ARMA model significantly alters the covariance matrix so that efficient procedures to



Table 4  
 Estimated parameter bias and square root of the mean squared error for the ARIMA(2, d, 0) model with repeated roots in the autoregressive polynomial ( $T = 200$ , 100 replications).

$d$	BIAS															
	MAX			F&T			GPH			MAX			$\sqrt{\text{MSE}}$			
	$\phi_1$	$\phi_2$	$\hat{d}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{d}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{d}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{d}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{d}$	
-0.2	-1.2	0.36	-0.099	-0.076	0.075	-0.280	-0.217	0.172	0.189	0.225	0.219	0.136	0.234	0.205	0.131	0.253
0.0	0.6	0.09	-0.014	-0.007	0.001	-0.077	-0.063	-0.039	-0.017	0.089	0.116	0.106	0.135	0.152	0.128	0.189
0.0	-1.0	0.25	-0.159	-0.148	0.108	-0.348	-0.303	0.208	0.103	0.256	0.241	0.133	0.297	0.271	0.160	0.223
0.3	0.8	0.16	-0.024	-0.030	-0.014	-0.077	-0.079	-0.050	-0.036	0.085	0.102	0.095	0.165	0.185	0.148	0.253

Table 5  
Seconds of CPU time required to evaluate the likelihood function for the fractional ARIMA( $p, d, q$ ) model and the nonfractional ARIMA( $p, 0, q$ ).

Numr of AR parameters ( $p$ )	Numr of MA parameters ( $q$ )											
	Fractional ARIMA( $p, d, q$ )					ARIMA( $p, 0, q$ )						
	0	1	2	3	4	5	0	1	2	3	4	5
$T = 100$												
0	0.08	0.08	0.10	0.23	0.24	0.35	0.01	0.03	0.07	0.13	0.22	0.34
1	0.11	0.11	0.13	0.25	0.27	0.38	0.01	0.03	0.07	0.13	0.22	0.34
2	0.14	0.15	0.18	0.30	0.33	0.47	0.03	0.03	0.07	0.12	0.21	0.34
3	0.29	0.30	0.33	0.47	0.50	0.63	0.06	0.06	0.06	0.12	0.22	0.34
4	0.32	0.34	0.37	0.51	0.55	0.69	0.12	0.12	0.12	0.12	0.21	0.34
5	0.49	0.50	0.54	0.70	0.74	0.89	0.21	0.22	0.21	0.21	0.22	0.34
$T = 150$												
0	0.17	0.17	0.20	0.32	0.33	0.46	0.01	0.04	0.10	0.19	0.33	0.52
1	0.20	0.21	0.23	0.36	0.40	0.49	0.01	0.04	0.10	0.19	0.33	0.52
2	0.25	0.26	0.28	0.42	0.45	0.59	0.04	0.04	0.10	0.19	0.33	0.52
3	0.40	0.42	0.44	0.59	0.64	0.78	0.09	0.09	0.10	0.19	0.32	0.51
4	0.45	0.47	0.50	0.66	0.72	0.87	0.18	0.18	0.19	0.19	0.32	0.51
5	0.63	0.66	0.70	0.87	0.94	1.10	0.32	0.32	0.32	0.32	0.32	0.52
$T = 200$												
0	0.31	0.31	0.33	0.46	0.47	0.59	0.01	0.05	0.14	0.25	0.43	0.68
1	0.34	0.34	0.37	0.50	0.53	0.65	0.01	0.05	0.14	0.25	0.43	0.68
2	0.40	0.41	0.43	0.59	0.64	0.77	0.05	0.05	0.13	0.25	0.43	0.68
3	0.57	0.58	0.61	0.78	0.83	0.99	0.12	0.12	0.13	0.25	0.43	0.68
4	0.62	0.64	0.69	0.86	0.93	1.11	0.24	0.24	0.25	0.25	0.43	0.68
5	0.81	0.84	0.92	1.08	1.16	1.37	0.42	0.42	0.42	0.43	0.43	0.68

evaluate the autocovariance function of ARMA models [e.g., Harvey and Phillips (1979)] cannot be used. The loss of these procedures raises the question of the practicality of repeated evaluation of the fractional likelihood function as is required for numerical optimization algorithms. The concern is the computation time required for one evaluation of the fractional ARIMA( $p, d, q$ ) likelihood function.

Table 5 presents the seconds of CPU time<sup>8</sup> required for one evaluation of the fractional ARIMA likelihood function for samples of size  $T = 100$ ,  $T = 150$ , and  $T = 200$ . Also presented is the time required to evaluate the nonfractional ARMA models using the Kalman filter based procedure presented in Harvey and Phillips (1979). Except for the simplest models, the time required to evaluate the fractional likelihood function is of the same order of magnitude as for the nonfractional models. The times required for the slowest fractional models are only between two and three times as long as for the comparable nonfractional ARMA model.

The time required to obtain a single set of maximum likelihood parameter estimates using a numerical optimizing algorithm is a function of more than simply the time required for one evaluation of the likelihood function. It also depends on the computer used, the number of parameters being estimated, the particular numerical optimization algorithm used, the starting values for the algorithm, and the population parameter for the model. In extreme situations (e.g., large number of parameters, large number of observations, roots near the unit circle) the time can be quite lengthy. However, for most models currently encountered in practice, the time required to obtain the exact maximum likelihood estimated for a fractional model is clearly of the same order of magnitude as the time required to obtain the exact maximum likelihood estimated for a nonfractional model. The general superiority of exact maximum likelihood estimates for time series models has been noted in Hillmer and Tiao (1979), Ansley and Newbold (1980), and Dahlhaus (1988).

## 6. Summary and extensions

This paper derived the unconditional normal likelihood function for a stationary fractionally integrated time series model. Recursive procedures, which allow quick evaluation of the likelihood function, were presented. These models make it possible to obtain maximum likelihood estimates. The small sample properties of the maximum likelihood estimates were compared to previous estimation techniques. The maximum likelihood estimates generally had smaller bias and MSE.

<sup>8</sup>The computations reported in this paper were performed in double precision VAX Fortran on a VAX workstation 3200.

The work in this paper can be extended in several directions. Currently, the approach is being generalized to the multivariate model. Also, the Monte Carlo work reported in this paper is only suggestive. More extensive simulations need to be performed to better understand the small sample properties of the maximum likelihood estimator. For empirical work, much of the early research can now be extended by using the more efficient estimation procedure and estimating all the parameters of the model.

One additional application of the maximum likelihood procedure is as a unit root test. The null hypothesis of a unit root in the AR process of a variable can be tested by first differencing the series and estimating the best fractional ARIMA( $p, d, q$ ) model. The null hypothesis can be rejected if  $\hat{d}$  is significantly different from zero. This approach to unit root testing has an advantage over previous tests because it explicitly accounts for nuisance parameters. The problems caused by nuisance parameters in unit root tests has been noted in Schwert (1987a, b). For an application of this approach to real United States GNP see Sowell (1990b).

#### Appendix I

##### *Justification of interchange of integrals*

This is a demonstration of the integrability of

$$\int_0^{2\pi} \sum_{n=0}^{\infty} \rho^n (1 - e^{-i\lambda})^{-d} (1 - e^{i\lambda})^{-d} e^{i\lambda h} d\lambda$$

to allow interchanging the summation and the integral. The parameters satisfy the restrictions:  $d$  is in the open interval  $(-\frac{1}{2}, \frac{1}{2})$ ,  $h$  is an integer, and  $|\rho| < 1$ ,

$$\begin{aligned} & \int_0^{2\pi} \sum_{n=0}^{\infty} |\rho|^n (1 - e^{-i\lambda})^{-d} (1 - e^{i\lambda})^{-d} e^{-i\lambda h} d\lambda \\ & \leq \int_0^{2\pi} \sum_{n=0}^{\infty} |\rho|^n |(1 - e^{-i\lambda})^{-d} (1 - e^{i\lambda})^{-d}| d\lambda \\ & = \frac{1}{1 - |\rho|} \int_0^{2\pi} |(1 - e^{-i\lambda})^{-d} (1 - e^{i\lambda})^{-d}| d\lambda. \end{aligned}$$

The integrand is either bounded or has poles at 0 and  $2\pi$ . At the poles the integrand behaves like  $\lambda^{-2d}$  near zero, which is integrable for  $d < \frac{1}{2}$ .

## Appendix 2

### Evaluation of the integral

This is the evaluation of the integral

$$\begin{aligned} & \frac{1}{2\pi} \int_0^{2\pi} (1 - e^{-i\lambda})^{-d} (1 - e^{i\lambda})^{-d} e^{i\lambda h} d\lambda \\ &= \frac{2^{-1-2d}}{\pi} \int_0^{2\pi} [\sin(\lambda/2)]^{-2d} e^{-i\lambda h} d\lambda, \end{aligned}$$

because

$$\begin{aligned} (1 - e^{i\xi}) &= 2 \sin(\xi/2) e^{i(\xi+3\pi)/2} = 2 \sin(\xi/2) e^{i(\xi-\pi)/2} \\ &= \frac{2^{-2d}}{\pi} \int_0^\pi [\sin(z)]^{-2d} e^{-i2hz} dz \end{aligned}$$

[see Erdelyi et al. (1953, p. 12, (1.5.29))]

$$= \frac{\Gamma(1-2d)(-1)^h}{\Gamma(1-d-h)\Gamma(1-d+h)}$$

## Appendix 3

### An alternative form of $C(d, h, \rho)$

Some of the autocovariances are written in terms of the function  $C(d, h, \rho)$ . The variable  $h$  can take on any integer value,  $\rho$  can any complex numr in the unit circle, while  $d$  is restricted to real values on the open interval  $(-\frac{1}{2}, \frac{1}{2})$ . An alternative form of the function, which is easier to evaluate, will derived using the following definition and equation. The hypergeometric function is defined by

$$F(a, b; c; x) \equiv \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c+n)\Gamma(n+1)} x^n.$$

For all real values of  $y$  and integer values of  $n$ , as long as the arguments of

the gamma function are never nonpositive integers, then

$$\frac{\Gamma(y+n)}{\Gamma(y)} = \frac{\Gamma(1-y)(-1)^n}{\Gamma(1-y-n)}. \quad (10)$$

This follows from eq. (8.334.3) of Gradshteyn and Ryzhik (1980) and cause  $\sin(\pi(n+y)) = (-1)^n \sin(\pi y)$  for real  $y$  and integer  $n$ .

The terms in the first sum of eq. (9) can be rewritten using eq. (10),

$$\begin{aligned} & \frac{(-1)^{h+m}}{\Gamma(1-d+h+m)\Gamma(1-d-h-m)} \\ &= \frac{(-1)^h}{\Gamma(1-d-h)\Gamma(1-d-h)} \left[ \frac{\Gamma(d+h+m)\Gamma(1-d+h)}{\Gamma(d+h)\Gamma(1-d+h+m)} \right]. \end{aligned}$$

Similarly, the terms in the second sum can be rewritten

$$\begin{aligned} & \frac{(-1)^{h-n}}{\Gamma(1-d+h-n)\Gamma(1-d-h+n)} \\ &= \frac{(-1)^h}{\Gamma(1-d-h)\Gamma(1-d-h)} \left[ \frac{\Gamma(d-h+n)\Gamma(1-d-h)}{\Gamma(d-h)\Gamma(1-d-h+n)} \right]. \end{aligned}$$

Using these relations, eq. (9) can be written

$$\begin{aligned} C(d, h, \rho) &= \frac{\Gamma(1-2d)(-1)^h}{\Gamma(1-d-h)\Gamma(1-d+h)} \\ & \times \left[ \rho^{2\rho} \sum_{m=0}^{\infty} \frac{\rho^m \Gamma(d+h+m)\Gamma(1-d+h)}{\Gamma(1-d+h+m)\Gamma(d+h)} \right. \\ & \left. + \sum_{n=1}^{\infty} \frac{\rho^n \Gamma(d-h+n)\Gamma(1-d-h)}{\Gamma(d-h)\Gamma(1-d-h+n)} \right] \\ &= \frac{\Gamma(1-2d)\Gamma(d+h)}{\Gamma(1-d+h)\Gamma(1-d)\Gamma(d)} \\ & \times \left[ \rho^{2\rho} F(d+h, 1; 1-d+h; \rho) \right. \\ & \left. + F(d-h, 1; 1-d-h; \rho) - 1 \right]. \end{aligned}$$

When calculating the likelihood function for a fractional model many different values of  $C(d, h, \rho)$  must be evaluated. Fortunately, this does not imply multiple evaluation of the hypergeometric function. The  $d$  and  $\rho$  parameters remain fixed for all the autocovariances. Only  $h$  varies and it only varies by integer values. Using the recursive relationship,  $y\Gamma(y) = \Gamma(y+1)$  and the definition of the hypergeometric function, it is straightforward to show

$$F(a, 1; c; \rho) = \frac{c-1}{\rho(a-1)} [F(a-1, 1; c-1; \rho) - 1].$$

For each evaluation of the likelihood function only  $p$  different hypergeometric functions need to be evaluated. This recursive calculation of the hypergeometric functions allows quick evaluations of the  $C(d, h, \rho)$  functions.

### References

- Ansley, Craig F. and Paul Newbold, 1980, Finite sample properties of estimators for autoregressive moving average models, *Journal of Econometrics* 13, 159-183.
- Aydogan, Kursat and G. Geoffrey Booth, 1988, Are there long cycles in common stock returns?, *Southern Economic Journal*, 141-149.
- Booth, G. Geoffrey, Fred R. Kaen, and Peter E. Koveos, 1982, R/S analysis of foreign exchange rates under two international monetary regimes, *Journal of Monetary Economics* 10, 407-415.
- Box, G.E.P. and G.M. Jenkins, 1976, *Time series analysis forecasting and control* (Holden-Day, San Francisco, CA).
- Brockwell, Peter J. and Richard A. Davis, 1987, *Time series: Theory and methods* (Springer-Verlag, New York, NY).
- Cheung, Yin-Wong, 1989, Long-term memory in foreign exchange rates, Unpublished manuscript (Department of Economics, University of Pennsylvania, Philadelphia, PA).
- Dahlhaus, Rainer, 1988, Small sample effects in time series analysis: A new asymptotic theory and a new estimate, *Annals of Statistics* 16, 808-841.
- Dahlhaus, Rainer, 1989, Efficient parameter estimation for self-similar processes, *Annals of Statistics* 17, 1749-1766.
- Diebold, Francis X. and Glenn D. Rudebusch, 1989, Long memory and persistence in aggregate output, *Journal of Monetary Economics* 24, 189-209.
- Erdelyi, Arthur et al., 1953, *Higher transcendental functions*, Vol. I (McGraw-Hill, New York, NY).
- Fox, Rort and Murad S. Taqqu, 1986, Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Annals of Statistics* 14, 517-532.
- Geweke, John F. and Susan Porter-Hudak, 1982, The estimation and application of long memory time series models, *Journal of Time Series Analysis* 4, 221-238.
- Gourieroux, Christian, Francoise Maurel, and Alain Monfort, 1987, Regression and non stationarity, Working paper no. 8708 (Institut National de la Statistique et des Etudes Economiques, France).
- Gradshteyn, I.S. and I.M. Ryzhik, 1980, *Tables of integrals, series and products* (Academic Press, New York, NY).
- Granger, C.W.J., 1980, Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics* 14, 227-238.
- Granger, C.W.J. and Roselyne Joyeux, 1980, An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis* 1, 15-29.
- Greene, Myron T. and Bruce D. Fielitz, 1977, Long-term dependence in common stock returns, *Journal of Financial Economics* 4, 339-349.

- Hannan, E.J. and Manfred Deistler, 1988, *The statistical theory of linear systems* (Wiley, New York, NY).
- Harvey, A.C. and G.D.A. Phillips, 1979, Maximum likelihood estimation of regression models with autoregressive-moving average disturbances, *Biometrika* 66, 49-58.
- Haubrich, Joseph G. and Andrew W. Lo, 1989, The sources and nature of long-term memory in the business cycle, Working paper (Department of Finance, Wharton School, University of Pennsylvania, Philadelphia, PA).
- Helms, Billy P., Fred R. Kaen, and Rort E. Rosenman, 1984, Memory in commodity futures contracts, *Journal of Futures Markets* 4, 559-567.
- Hillmer, Steven C. and George C. Tiao, 1979, Likelihood function of stationary multiple autoregressive moving average models, *Journal of the American Statistical Association* 74, 652-660.
- Hosking, J.R.M., 1981, Fractional differencing, *Biometrika* 68, 165-176.
- Hosking, J.R.M., 1984, Modeling persistence in hydrological time series using fractional differencing, *Water Resources Research* 20, 1898-1908.
- Janacek, G.J., 1982, Determining the degree of differencing for time series via the long spectrum, *Journal of Time Series Analysis* 3, 177-183.
- Lo, Andrew, 1989, Long-term memory in stock market prices, Working paper 3014-89-EFA (Alfred P. Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA).
- McLeod, Ian, 1975, Derivation of the theoretical autocovariance function of autoregressive-moving average time series, *Applied Statistics* 24, 255-256 (Correction 26, 194).
- McLeod, Angus Ian and Keith William Hipel, 1978, Preservation of the rescaled adjusted range I: A reassessment of the Hurst phenomenon, *Water Resources Research* 14, 491-508.
- Powell, M.J.D., 1971, Recent advances in constrained optimization, *Mathematical Programming* 1, 26-57.
- Schwert, G. William, 1987a, Tests for unit roots: A Monte Carlo investigation, Working paper GPB 87-01 (Center for Research in Government Policy and Business, University of Rochester, Rochester, NY).
- Schwert, G. William, 1987b, Effects of model specification on tests for unit roots in macroeconomic data, *Journal of Monetary Economics* 20, 73-103.
- Shea, Gary S., 1990, Uncertainty and implied variance bounds in long-memory model of the interest rate term structure, *Empirical Economics*, forthcoming.
- Sowell, Fallaw, 1989, A decomposition of block Toeplitz matrices with applications to vector time series, Mimeo. (Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA).
- Sowell, Fallaw, 1990a, The fractional unit root distribution, *Econometrica* 58, 495-505.
- Sowell, Fallaw, 1990b, Modeling long run havior with the fractional ARIMA model, Working paper (Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA).