The Annals of Statistics
1998, Vol. 26, No. 2, 719–740

STATE SPACE MODELING OF LONG-MEMORY PROCESSES[1]

BY NGAI HANG CHAN AND WILFREDO PALMA

Carnegie Mellon University and
Pontificia Universidad Católica de Chile

This paper develops a state space modeling for long-range dependent data. Although a long-range dependent process has an infinite-dimensional state space representation, it is shown that by using the Kalman filter, the exact likelihood function can be computed recursively in a finite number of steps. Furthermore, an approximation to the likelihood function based on the truncated state space equation is considered. Asymptotic properties of these approximate maximum likelihood estimates are established for a class of long-range dependent models, namely, the fractional autoregressive moving average models. Simulation studies show rapid converging properties of the approximate maximum likelihood approach.

1. Introduction. Long-range dependent or long-memory processes have been receiving considerable attention among researchers from various disciplines, ranging from econometrics [Cheung and Diebold (1994)] to meteorology [Bloomfield (1992)]. The monograph by Beran (1994) and the reviewing article by Robinson (1994) provide two updated surveys of recent developments of long-memory processes in statistics and economics, respectively.

One of the most commonly used long-memory models is the fractionally integrated autoregressive moving average process (ARFIMA). An ARFIMA(p, d, q) process $\{y_t\}$ is defined by

$$\Phi(B)(1-B)^d y_t = \Theta(B)\varepsilon_t, \tag{1.1}$$

where $|d| < 1/2$, $\{\varepsilon_t\}$ is a Gaussian white noise sequence (independent, normally distributed with zero mean and constant variance, σ_ε^2); B is the backshift operator $By_t = y_{t-1}$; $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is the AR operator; $\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ is the MA operator such that $\Phi(B)$ and $\Theta(B)$ have all their roots outside the unit circle with no common factors and $(1-B)^d = \sum_{k=0}^{\infty} \pi_k (-B)^k$ is the fractional difference operator, with $\pi_k = \Gamma(k-d)/\Gamma(k+1)\Gamma(-d)$ and Γ denoting the Gamma function. If $d = 0$, the usual ARMA model is obtained.

A distinctive feature of a long-memory process is that its spectral density $f(\omega)$ is unbounded at the frequency zero. The spectral density function of an

---

Received February 1997; revised October 1997.

[1]Supported in part by Office of Naval Research Grant N00014-89-J-1851, an Earmarked Research Grant HKUST 603/95P from the RGC of Hong Kong and Project 97/07E from the DIPUC of Chile.

AMS 1991 *subject classifications*. Primary 62M10, 62E20; secondary 60F17.

Key words and phrases. ARFIMA, asymptotic normality, consistency, efficiency, long-memory, MLE, truncated state space.

ARFIMA model is given by

$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{|\Theta(e^{i\omega})|^2}{|\Phi(e^{i\omega})|^2} |1 - e^{i\omega}|^{-2d}, \qquad \omega \in [-\pi, \pi]. \tag{1.2}$$

Thus, for $0 < d < 1/2$, $f(\omega)$ has a singularity at $\omega = 0$. A particular case of an ARFIMA model is the *fractional noise* process described by $(1 - B)^d y_t = \varepsilon_t$ or, equivalently,

$$y_t = (1 - B)^{-d} \varepsilon_t = \sum_{j=0}^{\infty} \eta_j \varepsilon_{t-j}, \tag{1.3}$$

where $\eta_j = \Gamma(j + d)/\Gamma(j + 1)\Gamma(d)$.

For an ARFIMA model, several methods for estimating the fractional difference parameter d are available in the literature. The so-called R/S algorithm, introduced by Hurst (1951) in hydrology, is studied by Mandelbrot and Wallis (1969) and Mandelbrot and Taqqu (1979). A regression method based on the spectral density of the fractional ARIMA processes is developed by Geweke and Porter-Hudak (1983). It is known that this method results in an asymptotically biased estimate of d which is less efficient than the estimates based on maximum likelihood procedures; see, for example, Hurvich and Beltrao (1993). Semiparametric modifications to this method can be found in Robinson (1995). For Gaussian maximum likelihood estimates (MLE), several asymptotic results are available. In particular, asymptotic distributional properties such as consistency, asymptotic normality and asymptotic efficiencies of the Gaussian parameter estimates of the MLE of an ARFIMA model are established by Fox and Taqqu (1986), Dahlhaus (1989) and Giraitis and Surgailis (1990).

Despite all these desirable properties, implementing exact maximum likelihood procedures presents serious computational problems in practice. There are two kinds of problems. First, since the correlations of a long-memory model decay slowly, all autocorrelations including those with large lags have to be included in the computation. This requires extensive memory capacity on the computer. Second, contrary to the standard ARIMA model, in the ARFIMA case, no recursive schemes have been constructed to facilitate the computing of the log-likelihood function. Discussions on some of these computational problems can be found in Li and McLeod (1986) and Sowell (1992).

In this paper, a state space approach is proposed to compute the exact and approximate ML estimates for an ARFIMA model. By means of the Kalman filter, it is shown that the exact MLE can be evaluated in a finite number of steps. By truncating the dimension of the state in a state space representation of an ARFIMA model, an approximate ML procedure is proposed. It is shown that parameter estimates obtained from this approach are consistent, asymptotically normal and asymptotically efficient.

One of the main objectives in this paper is, therefore, to provide a computationally efficient approach to calculate the MLE for an ARFIMA model. The proposed approach not only possesses some of the desirable asymptotic properties of the exact MLE, but is also easily computable. Simulation studies show that the approximate MLE approach gives satisfactory results for a variety of situations. Furthermore, a state space approach provides an appropriate framework to study missing values problems, Kalman smoothing and predictions and recursive evaluations of the likelihood functions for a long-memory model.

This paper is organized as follows. Section 2 introduces the state space approach and its application to an ARFIMA model. The Kalman filter theory and calculations of exact MLE are also discussed. In particular, it is shown in this section that although the state space representation of an ARFIMA model is infinite-dimensional, the exact likelihood function can be computed in a finite number of steps. Section 3 proposes an approximate maximum likelihood estimate based on a truncated state space representation. It is shown that this approximate MLE is consistent, asymptotically normal, and efficient. In Section 4, a simulation study is conducted on the approximate MLE procedure which demonstrates its rapid convergence properties. Concluding remarks are given in Section 5.

2. State space models. This section introduces the state space system, the Kalman filter and the exact likelihood function for an ARFIMA model. It is shown in Corollary 2.1 that there exists only an infinite-dimensional representation for a long-memory process. Despite this fact, the exact likelihood functions can still be evaluated in a finite number of steps, as stated in Theorem 2.2. In one sense, this result is not surprising since the likelihood function for a short-memory stationary series can be evaluated by means of the modified Cholesky factorization of the covariance matrix in a finite number of steps; see, for example, Newton (1988). Earlier discussions on the Kalman filter formulation of the likelihood function in the time series context can be found in Jones (1980).

2.1. *State space representations of ARFIMA models.* For a standard ARIMA model, it is always possible to find a finite-dimensional state space representation. For an ARFIMA process, it is proved in Theorem 2.1 and Corollary 2.1 that there are no finite-dimensional state space representations.

Consider a finite-dimensional state space system

$$X_{t+1} = FX_t + \varepsilon_t, \tag{2.1}$$

$$y_t = GX_t + \eta_t, \tag{2.2}$$

where $y_t \in \mathbb{R}$ and $X_t \in \mathbb{R}^m$, and m is the dimension of the state space representation. For any given causal time series $\{y_t\}$, the following results characterize the dimension of its state space representation.

LEMMA 2.1. *Consider the finite-dimensional state space system given by* (2.1) *and* (2.2). *If this system is observable and* $\{y_t\}$ *is stationary*, *then the sequence of states*, $\{X_t\}$, *is also stationary*.

PROOF. The state X_t satisfies the equation

$$O_m X_t = \tilde{y}_t - \tilde{\eta}_t - M_m \tilde{\varepsilon}_t,$$

where $\tilde{y}_t = (y_t, y_{t+1}, \dots, y_{t+m-1})'$, $\tilde{\eta}_t = (\eta_t, \dots, \eta_{t+m-1})'$, $\tilde{\varepsilon}_t = (\varepsilon_t', \dots, \varepsilon_{t+m-1}')'$, $O_m = (G', (GF)', \dots, (GF^{m-1})')'$ is the observability matrix, and

$$M_m = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ G & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \\ GF^{m-2} & \cdots & G & 0 \end{bmatrix}.$$

Since the system is *observable*, the matrix O_m is of full rank; see page 44 of Hannan and Deistler (1988). Hence, the matrix $O_m' O_m$ is also of full rank and the state X_t can be written as $X_t = (O_m' O_m)^{-1} O_m (\tilde{y}_t - \tilde{\eta}_t - M_m \tilde{\varepsilon}_t)$. As the process $\tilde{y}_t - \tilde{\eta}_t - M_m \tilde{\varepsilon}_t$ is stationary for a fixed m, the sequence of states X_t is also stationary. □

THEOREM 2.1. *Any causal time series with a finite-dimensional state space representation is a short memory process, that is,* $\gamma_k \sim a^k$, $|a| < 1$, *where* γ_k *is its autocovariance function.*

PROOF. Let $\{y_t\}$ and $\{X_t\}$ be the time series and the states satisfying the state space equations (2.1) and (2.2), respectively. Since it is always possible to find an equivalent *observable* representation for any given system [see page 45 of Hannan and Deistler (1988)], one may assume this system to be observable. By Lemma 2.1, as the process $\{y_t\}$ is causal, the state process $\{X_t\}$ is also causal, which in turn implies $|\lambda_m(F)| < 1$, where $\lambda_m(F)$ is the eigenvalue of F with maximal modulus [see Brockwell and Davis (1991), pages 417–418].

Consider the autocovariance function of $\{y_t\}$,

$$\gamma_y(k) = \sum_{j=1}^{\infty} GF^{j-1} \Sigma_\varepsilon (F')^{j-1} (F')^k G' + F^k S(\eta, \varepsilon),$$

where $\Sigma_\varepsilon = \mathrm{Var}[\varepsilon_t]$ and $S(\eta, \varepsilon) = \mathrm{Cov}[\eta_t, \varepsilon_t]$. Thus,

$$|\gamma_y(k)| \le \| F \|^k \left[\sum_{j=1}^{\infty} \| G \|^2 \| F \|^{2(j-1)} \| \Sigma_\varepsilon \| + \| S(\eta, \varepsilon) \| \right],$$

where $\|A\| = |\lambda_m(A)|$ denotes the spectral norm of the matrix A. Since $\|F\| < 1$, $|\gamma_y(k)| \to 0$, at exponential rate, as $k \to \infty$. □

COROLLARY 2.1. *There is no finite-dimensional state space representation of an ARFIMA process with long-memory parameter* $d \neq 0$.

The proof follows directly from the preceding theorem.

Recall that a causal ARFIMA(p, d, q) process $\{y_t\}$ has a linear process representation given by

$$y_t = \frac{\Theta(B)}{\Phi(B)}(1-B)^{-d}\varepsilon_t = \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j}, \tag{2.3}$$

where φ_j are the coefficients of $\varphi(z) = \sum_{j=0}^{\infty} \varphi_j z^j = (\Theta(z)/\Phi(z))(1-z)^{-d}$. An infinite-dimensional state space representation may be constructed as follows. From equation (2.3), a state space system may be written [cf. page 22 of Hannan and Deistler (1988)] as

$$X_{t+1} = FX_t + H\varepsilon_t, \tag{2.4}$$

$$y_t = GX_t + \varepsilon_t, \tag{2.5}$$

where

$$X_t = \begin{bmatrix} y(t|t-1) \\ y(t+1|t-1) \\ y(t+2|t-1) \\ \vdots \end{bmatrix}, \qquad y(t|j) = E[y_t | y_j, y_{j-1}, \dots], \tag{2.6}$$

$$F = \begin{bmatrix} 0 & 1 & 0 & & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & & \ddots & \end{bmatrix}, \qquad H = [\varphi_1, \varphi_2, \dots]' \quad \text{and} \quad G = [1, 0, 0, \dots]. \tag{2.7}$$

2.2. *Exact likelihood function.* Recall that if the noise is Gaussian, the exact log-likelihood function is given (omitting a constant) by

$$l(\boldsymbol{\theta}) = -\tfrac{1}{2}\log\det T_n(\boldsymbol{\theta}) - \tfrac{1}{2} Y_n' T_n^{-1}(\boldsymbol{\theta}) Y_n, \tag{2.8}$$

where $[T_n(\boldsymbol{\theta})]_{r,s=1,\dots,n} = \int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\omega) e^{i\omega(r-s)}\, d\omega$ is the covariance matrix of $Y_n = (y_1, \dots, y_n)'$, $\boldsymbol{\theta} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, d, \sigma_\varepsilon)'$ and $f_{\boldsymbol{\theta}}(\omega)$ is the spectral density of the process given in (1.2). This function can be evaluated by directly applying the Kalman recursive equations in Proposition 12.2.2 of Brockwell and Davis (1991) to the infinite-dimensional system. Although Corollary 2.1 states that the state space representation of an ARFIMA model is infinite-dimensional, the exact likelihood function can be evaluated in a finite number of steps. Specifically, we have the following theorem.

THEOREM 2.2. *Let $\{y_1, \dots, y_n\}$ be a finite sample of an ARFIMA(p, d, q) process. If Ω_1 is the variance of the initial state X_1 for the infinite-dimensional representation, then the computation of the exact likelihood function* (2.8) *depends only on the first n components of the Kalman equations.*

PROOF. Let $\Omega_t = (\omega_{ij}^{(t)})$ be the state estimation error covariance matrix at time t. The Kalman equations for the infinite-dimensional system are given by

$$\hat{X}_1 = E[X_1], \tag{2.9}$$

$$\Omega_1 = E[X_1 X_1'] - E[\hat{X}_1 \hat{X}_1'], \tag{2.10}$$

$$\Delta_t = \omega_{11}^{(t)} + 1, \tag{2.11}$$

$$\omega_{ij}^{(t+1)} = \omega_{i+1,\, j+1}^{(t)} + \varphi_i \varphi_j - \frac{(\omega_{i+1,\,1}^{(t)} + \varphi_i)(\omega_{j+1,\,1}^{(t)} + \varphi_j)}{\omega_{11}^{(t)} + 1} \tag{2.12}$$

and

$$\hat{X}_{t+1} = \big(\hat{X}_1^{(t+1)}, \hat{X}_2^{(t+1)}, \dots\big)' = \big(\hat{X}_i^{(t+1)}\big)_{i=1,\,2,\dots}', \tag{2.13}$$

where

$$\hat{X}_i^{(t+1)} = \hat{X}_{i+1}^{(t)} + \frac{(y_t - \hat{X}_1^{(t)})(\omega_{i+1,\,1}^{(t)} + \varphi_i)}{\omega_{11}^{(t)} + 1}, \tag{2.14}$$

$$\hat{y}_{t+1} = G\hat{X}_{t+1} = \hat{X}_1^{(t+1)}, \tag{2.15}$$

and the log-likelihood function is given by

$$l_n = -\frac{1}{2}\left\{ n \log 2\pi + \sum_{t=1}^{n} \log \Delta_t + n \log \sigma_\varepsilon^2 + \frac{1}{\sigma_\varepsilon^2} \sum_{t=1}^{n} \frac{(y_t - \hat{y}_t)^2}{\Delta_t} \right\}. \tag{2.16}$$

In order to evaluate this likelihood, the initial state is estimated by $\hat{X}_1 = 0$ and the initial prediction error covariance matrix is estimated by $\Omega_1 = E[X_1 X_1'] = (\omega_{ij}^1(\boldsymbol{\theta}))_{i,\, j=1,2,\dots}$, with $\omega_{ij}^1(\boldsymbol{\theta}) = \sum_{k=0}^{\infty} \varphi_{i+k}(\boldsymbol{\theta}) \varphi_{j+k}(\boldsymbol{\theta})$. Expressions for the coefficients $\varphi_i(\boldsymbol{\theta})$ can be found on page 92 of Brockwell and Davis (1991). The optimization process starts with an initial value $\boldsymbol{\theta}_0$ and proceeds to maximize the likelihood function.

Let $\tilde{X}_1$ be the vector consisting of the first n components of $\hat{X}_1$, that is,

$$\tilde{X}_1 = E\big(X_1^{(1)}, \dots, X_n^{(1)}\big)' = \big(E[X_i^{(1)}]\big)_{i=1,\dots,n}', \tag{2.17}$$

where

$$X_1 = \big(X_1^{(1)}, X_2^{(1)}, \dots\big)' = \big(X_i^{(1)}\big)_{i=1,\,2,\dots}'; \tag{2.18}$$

$$\tilde{\Omega}_1 = \big(E[X_i^{(1)} X_j^{(1)}] - E[\tilde{X}_i^{(1)} \tilde{X}_j^{(1)}]\big)_{i,\, j=1,\dots,n}; \tag{2.19}$$

$$\tilde{\Delta}_t = \tilde{\omega}_{11}^{(t)} + 1; \tag{2.20}$$

$$\tilde{\omega}_{ij}^{(t+1)} = \begin{cases} \tilde{\omega}_{i+1,\,j+1}^{(t)} + \varphi_i\varphi_j - \dfrac{(\tilde{\omega}_{i+1,\,1}^{(t)} + \varphi_i)(\tilde{\omega}_{j+1,\,1}^{(t)} + \varphi_j)}{\tilde{\omega}_{11}^{(t)} + 1}, & i,\, j \le n-1, \\ \dfrac{\varphi_i\varphi_j\tilde{\omega}_{11}^{(t)}}{\tilde{\omega}_{11}^{(t)} + 1}, \qquad i = n \text{ or } j = n; \end{cases} \tag{2.21}$$

$$\tilde{X}_{t+1} = \big(\tilde{X}_i^{(t+1)}\big)_{i=1,\,2,\ldots,n}, \tag{2.22}$$

with

$$\tilde{X}_i^{(t+1)} = \begin{cases} \tilde{X}_{i+1}^{(t)} + \dfrac{(y_t - \tilde{X}_1^{(t)})(\tilde{\omega}_{i+1,\,1}^{(t)} + \varphi_i)}{\tilde{\omega}_{11}^{(t)} + 1}, & i \le n-1, \\ \dfrac{(y_t - \tilde{X}_1^{(t)})\varphi_n}{\tilde{\omega}_{11}^{(t)} + 1}, & i = n; \end{cases} \tag{2.23}$$

$$\tilde{y}_{t+1} = G\tilde{X}_{t+1} = \tilde{X}_1^{(t+1)}; \tag{2.24}$$

and the log-likelihood function is given by

$$\tilde{l}_n = -\frac{1}{2}\left\{ n\log 2\pi + \sum_{t=1}^{n} \log\tilde{\Delta}_t + n\log\sigma_\varepsilon^2 + \frac{1}{\sigma_\varepsilon^2}\sum_{t=1}^{n}\frac{(y_t - \tilde{y}_t)^2}{\tilde{\Delta}_t} \right\}. \tag{2.25}$$

From these equations, observe that since the first n rows (columns) of Ω_1 and $\tilde{\Omega}_1$ are equal, then $\Delta_t = \tilde{\Delta}_t$ for $t = 1, \ldots n$. Thus, the matrices Ω_t and $\tilde{\Omega}_t$ are the same except for the last $n - t + 1$ rows (columns). Similarly, the states $\hat{X}_t$ and $\tilde{X}_t$ are the same except for the last $n - t + 1$ components, and $\hat{y}_t = \tilde{y}_t$, $t = 1, \ldots, n$. Since the log-likelihood functions l_n and $\tilde{l}_n$ depend only on $\{\Delta_t\}$, $\{\hat{y}_t\}$ and $\{\tilde{\Delta}_t\}$, $\{\tilde{y}_t\}$ which are identical, $l_n = \tilde{l}_n$. □

It is worth noting that as a consequence of Theorem 2.2, given a sample of n observations from an ARFIMA process, the evaluation of the exact likelihood function is based only on the first n components of the state vector. Therefore, the remaining infinitely many components of the state vector can be omitted from the computations.

3. Approximate maximum likelihood estimation. Although Theorem 2.2 guarantees that the exact likelihood function can be computed in a finite number of steps, such a computation may be cumbersome. According to the Kalman filter equations, the evaluation of the likelihood function consists of n iterations (sample size) and each iteration consists of a number of matrix evaluations. For the exact method, these matrices are of dimensions $n \times n$, so there are n^2 evaluations. The resulting algorithm is then of order n^3.

Alternatively, consider the moving average expansion of the differenced process $z_t = (1 - B)y_t$:

$$z_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \tag{3.1}$$

where $\psi_j = \varphi_j - \varphi_{j-1}$. If we truncate this expansion after m components, an approximate model can be written as

$$z_t = \sum_{j=0}^{m} \psi_j \varepsilon_{t-j}. \tag{3.2}$$

A main advantage of this approach is that the coefficients ψ_j converge to zero faster than the coefficients φ_j so that a smaller truncation parameter m results with the differenced data. On the other hand, the differenced approach suffers from the drawback that it would reduce each stretch of contiguous data by one observation. This would pose a problem for time series with missing values. In such cases, one may have to resort to using the *exact* Kalman filter to the nondifferenced data. Alternatively, the *truncated* state space can also be applied directly to the nondifferenced data.

A state space representation of this truncated model is given by [see, for example, page 470 of Brockwell and Davis (1991)],

$$X_{t+1} = \begin{bmatrix} 0 & & I_{m-1} \\ 0 & \cdots & 0 \end{bmatrix} X_t + \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_m \end{bmatrix} \varepsilon_t, \tag{3.3}$$

$$z_t = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix} X_t + \varepsilon_t. \tag{3.4}$$

The Gaussian log-likelihood of the truncated model (3.2) may be written (omitting a constant) as

$$\tilde{l}_n(\boldsymbol{\theta}) = -\tfrac{1}{2} \log \det \tilde{T}_{n,m}(\boldsymbol{\theta}) - \tfrac{1}{2} Z_n' \tilde{T}_{n,m}(\boldsymbol{\theta})^{-1} Z_n, \tag{3.5}$$

where $[\tilde{T}_{n,m}(\boldsymbol{\theta})]_{r,s=1,\dots,n} = \int_{-\pi}^{\pi} \tilde{f}_{m,\boldsymbol{\theta}}(\omega) e^{i\omega(r-s)} d\omega$, is the covariance matrix of $Z_n = (z_1, \dots, z_n)'$ with $\tilde{f}_{m,\boldsymbol{\theta}}(\omega) = \sigma^2 |\Psi_m(e^{i\omega})|^2$, and $\Psi_m(e^{i\omega}) = 1 + \psi_1 e^{i\omega} + \cdots + \psi_m e^{mi\omega}$.

In this case, the matrices involved in the truncated Kalman equations are of order $m \times m$. Therefore, only m^2 evaluations are required for each iteration and the algorithm has an order $n \times m^2$. For a fixed truncation parameter m, the calculation of the likelihood function is only of order n for the approximate ML method. Thus, for very large samples, it may be desirable to consider truncating the Kalman recursive equations after m components. With this truncation, the number of operations required for a single evaluation of the log-likelihood function is reduced to an order of n. Asymptotic behaviors of these approximate state space estimators are established in this section. The consistency of the approximate MLE is addressed in Theorem 3.1. Asymptotic normality is given in Theorem 3.2 and its efficiency is discussed in Theorem 3.3.

3.1. *Asymptotic properties of the approximate MLE.* Before stating the main theorems, we first introduce some definitions, regularity conditions, and notation. Let $\tilde{\boldsymbol{\theta}}_{n,m}$ be the value that maximizes the log-likelihood of the truncated model where $\boldsymbol{\theta} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, d, \sigma_\varepsilon)'$ is an $r = p + q + 2$-dimensional parameter vector and let $\boldsymbol{\theta}_0$ be the *true* parameter. Assume that the regularity conditions listed in Dahlhaus (1989) hold.

Define the partial derivatives

$$\nabla f(\boldsymbol{\theta}) = \left(\frac{\partial}{\partial \theta_j} f(\boldsymbol{\theta})\right)_{j=1,\dots,r}$$

and

$$\nabla^2 f(\boldsymbol{\theta}) = \left(\frac{\partial^2}{\partial \theta_j \partial \theta_k} f(\boldsymbol{\theta})\right)_{j,k=1,\dots,r},$$

and the matrices for $i, j = 1, \dots, r$:

$$T_{\partial i} = T\left(\frac{\partial}{\partial \theta_i} f_{\boldsymbol{\theta}}\right), \qquad T_{\partial i \partial j} = T\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} f_{\boldsymbol{\theta}}\right),$$

$$A_i^{(1)} = T^{-1} T_{\partial i} T^{-1}, \qquad A^{(1)} = T^{-1} T_{\nabla} T^{-1},$$

$$\tilde{A}^{(1)} = \tilde{T}^{-1} \tilde{T}_{\nabla} \tilde{T}^{-1}, \qquad \tilde{A}_i^{(1)} = \tilde{T}^{-1} \tilde{T}_{\partial i} \tilde{T}^{-1},$$

$$A_{ij}^{(2)} = T^{-1} T_{\partial i} T^{-1} T_{\partial j} T^{-1}, \qquad A^{(2)} = T^{-1} T_{\nabla} T^{-1} T_{\nabla} T^{-1},$$

$$\tilde{A}^{(2)} = \tilde{T}^{-1} \tilde{T}_{\nabla} \tilde{T}^{-1} \tilde{T}_{\nabla} \tilde{T}^{-1}, \qquad \tilde{A}_{ij}^{(2)} = \tilde{T}^{-1} \tilde{T}_{\partial i} \tilde{T}^{-1} \tilde{T}_{\partial j} \tilde{T}^{-1},$$

$$A_{ij}^{(3)} = T^{-1} T_{\partial i \partial j} T^{-1}, \qquad A^{(3)} = T^{-1} T_{\nabla^2} T^{-1},$$

$$\tilde{A}^{(3)} = \tilde{T}^{-1} \tilde{T}_{\nabla^2} \tilde{T}^{-1}, \qquad \tilde{A}_{ij}^{(3)} = \tilde{T}^{-1} \tilde{T}_{\partial i \partial j} \tilde{T}^{-1},$$

where $T = T_n(f_{\boldsymbol{\theta}})$, $T_\nabla = T_n(\nabla f_{\boldsymbol{\theta}})$, $T_{\nabla^2} = T_n(\nabla^2 f_{\boldsymbol{\theta}})$, $\tilde{T} = T_n(\tilde{f}_{\boldsymbol{\theta},m})$, $\tilde{T}_\nabla = T_n(\nabla \tilde{f}_{\boldsymbol{\theta},m})$, $\tilde{T}_{\nabla^2} = T_n(\nabla^2 \tilde{f}_{\boldsymbol{\theta},m})$.

The next three results show that the approximate maximum likelihood estimates obtained by the truncation approach are consistent, asymptotically normal, and efficient. Their proofs are given in the next subsection.

THEOREM 3.1 (Consistency). *Assume that $m = n^\beta$ with $\beta > 0$, then as $n \to \infty$,*

$$\tilde{\boldsymbol{\theta}}_{n,m} \to \boldsymbol{\theta}_0 \quad \textit{in probability.}$$

THEOREM 3.2 (Central limit theorem). *Suppose that $m = n^\beta$ with $\beta \geq 1/2$, then as $n \to \infty$,*

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_{n,m} - \boldsymbol{\theta}_0) \to_{\mathscr{L}} N(0, \Gamma^{-1}(\boldsymbol{\theta}_0)),$$

where "$\to_{\mathscr{L}}$" denotes convergence in distribution and $\Gamma(\boldsymbol{\theta}) = (\Gamma_{ij}(\boldsymbol{\theta}))$ *with*

$$\Gamma_{ij}(\boldsymbol{\theta}) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial \log k(\omega,\boldsymbol{\theta})}{\partial\theta_i}\right\}\left\{\frac{\partial \log k(\omega,\boldsymbol{\theta})}{\partial\theta_j}\right\}d\omega,$$

and

$$k(\omega,\boldsymbol{\theta}) = \left|\sum_{j=0}^{\infty}\psi_j(\boldsymbol{\theta})e^{ij\omega}\right|^2.$$

THEOREM 3.3 (Efficiency). *Assume that* $m = n^{\beta}$ *with* $\beta \geq 1/2$, *then* $\tilde{\boldsymbol{\theta}}_{n,m}$ *is an efficient estimator of* $\boldsymbol{\theta}_0$.

Before proving these theorems, we first establish some auxiliary lemmas. In what follows, the symbol K denotes a generic constant which may vary from line to line.

LEMMA 3.1. *Let* ψ_j *be the coefficients given in* (3.1), *then for* j *large and* $\varepsilon > 0$,

(i) $$|\psi_j(\boldsymbol{\theta})| \leq K j^{d-2},$$

(ii) $$\left|\frac{\partial}{\partial\theta_i}\psi_j(\boldsymbol{\theta})\right| \leq K j^{d-2+\varepsilon}, \qquad i = 1,\dots,r,$$

(iii) $$\left|\frac{\partial^2}{\partial\theta_i\partial\theta_l}\psi_j(\boldsymbol{\theta})\right| \leq K j^{d-2+2\varepsilon}, \qquad i,l = 1,\dots,r.$$

PROOF. (i) $\psi(z) = (\Theta(z)/\Phi(z))(1-z)^{-d+1} = \sum_{j=0}^{\infty}\psi_j z^j$. Define $\varphi(z) = (\Theta(z)/\Phi(z)) = \sum_{k=0}^{\infty}\varphi_k z^k$ and $\eta(z) = (1-z)^{-d+1} = \sum_{k=0}^{\infty}\eta_k z^k$. The coefficients φ_k can be written [cf. Brockwell and Davis (1991), page 92] as

$$\varphi_k = \sum_{i=1}^{m}\sum_{l=0}^{r_i-1}\alpha_{il}k^l\xi_i^k, \qquad k \geq \max(p, q+1) - p,$$

where m is the number of distinct roots of $\Phi(z)$, $|\xi_i| < 1$ is the inverse of the ith root of $\Phi(z)$ with multiplicity r_i and α_{il} are constants. Let $L \geq \max(p, q+1) - p$, then

$$\begin{aligned}|\psi_j| &= \left|\sum_{k=0}^{L}\varphi_k\eta_{j-k} + \sum_{k=L+1}^{\infty}\varphi_k\eta_{j-k}\right|\\ &= \left|\sum_{k=0}^{L}\varphi_k\eta_{j-k} + \sum_{k=L+1}^{\infty}\left[\sum_{i=1}^{m}\sum_{l=0}^{r_i-1}\alpha_{il}k^l\xi_i^k\right]\eta_{j-k}\right|\\ &\leq \sum_{k=0}^{L}|\varphi_k||\eta_{j-k}| + \sum_{i=1}^{m}\sum_{l=0}^{r_i-1}|\alpha_{il}|\left[\sum_{k=L+1}^{\infty}k^l|\xi_i|^k|\eta_{j-k}|\right].\end{aligned}$$

Observe that for $k \geq L$, there exist constants $c_l \geq 0$ and $0 < a_i < 1$ such that $k^l|\xi_i|^k \leq c_l a_i^k$. This can be seen as follows. Let $a_i = |\xi_i| + \varepsilon_i$ where $\varepsilon_i > 0$. Then $o < a_i < 1$ (since $|\xi_i| < 1$, it is always possible to find such an ε_i). For $l \geq 0$, $k^l(|\xi_i|/a_i)^k \to 0$ as $k \to \infty$. Thus, for $k \geq L$, there is a constant c_l such that $k^l(|\xi_i|/a_i)^k \leq c_l$, which implies $k^l|\xi_i|^k \leq c_l a_i^k$. Therefore,

$$\begin{aligned}|\psi_j| &\leq \sum_{k=0}^{L} |\varphi_k|\,|\eta_{j-k}| + \sum_{i=1}^{m}\sum_{l=0}^{r_i-1} |\alpha_{il}|\left[\sum_{k=L+1}^{\infty} c_l a_i^k |\eta_{j-k}|\right] \\ &\leq \sum_{k=0}^{L} |\varphi_k|\,|\eta_{j-k}| + \sum_{i=1}^{m}\sum_{l=0}^{r_i-1} |\alpha_{il}| c_l\left[\sum_{k=0}^{\infty} a_i^k |\eta_{j-k}|\right] \\ &= \sum_{k=0}^{L} |\varphi_k|\,|\eta_{j-k}| + |\eta_j|\sum_{i=1}^{m}\sum_{l=0}^{r_i-1} |\alpha_{il}| c_l\left[\sum_{k=0}^{\infty} a_i^k P(k, j, d)\right],\end{aligned}$$

where

$$P(k, j, d) = \frac{\eta_{j-k}}{\eta_j} = \frac{j!(j-k+d-2)!}{(j-k)!(j+d-2)!},$$

since

$$\eta_{j-k} = \frac{(j-k+d-2)!}{(j-k)!(d-2)!} \quad \text{and} \quad \eta_j = \frac{(j+d-2)!}{j!(d-2)!}.$$

Thus,

$$|\psi_j| \leq \sum_{k=0}^{L} |\varphi_k|\,|\eta_{j-k}| + |\eta_j|\sum_{i=1}^{m}\sum_{l=0}^{r_i-1} |\alpha_{il}| c_l F(1, -j, 2-d-j, a_i),$$

where F is the hypergeometric function [cf. Hosking (1981)]. Moreover, $|\eta_{j-k}| \leq |\eta_{j-L}|$, for $0 \leq k \leq L$. This implies

$$|\psi_j| \leq |\eta_{j-L}|\sum_{k=0}^{L} |\varphi_k| + |\eta_j|\sum_{i=1}^{m}\sum_{l=0}^{r_i-1} |\alpha_{il}| c_l F(1, -j, 2-d-j, a_i).$$

As $j \to \infty$, $F(1, -j, 2-d-j, a_i) \to (1-a_i)^{-1}$ [cf. Hosking (1981)], and for j large, $|\eta_j| \sim K j^{d-2}$ and $|\eta_{j-L}| \sim K j^{d-2}$ (for fixed L). Therefore, for large j,

$$|\psi_j| \leq K j^{d-2}\left[\sum_{k=0}^{L} |\varphi_k| + \sum_{i=1}^{m}\sum_{l=0}^{r_i-1} \frac{|\alpha_{il}|}{1-a_i}\right].$$

Since $[\sum_{k=0}^{L} |\varphi_k| + \sum_{i=1}^{m}\sum_{l=0}^{r_i-1}(|\alpha_{il}|/1-a_i)]$ is a rational function of the parameter $\boldsymbol{\theta} \in \Theta$, it is continuous. Given that the parameter space Θ is compact, there is a constant K such that

$$\sum_{k=0}^{L} |\varphi_k| + \sum_{i=1}^{m}\sum_{l=0}^{r_i-1} \frac{|\alpha_{il}|}{1-a_i} \leq K.$$

Thus,

$$|\psi_j| \le K j^{d-2}.$$

Parts (ii) and (iii) are proved analogously to part (i). □

LEMMA 3.2. (i) *Let* $\{a_k\}$ *be a sequence of numbers such that*

$$\lim_{m\to\infty} a_m m^{1-\beta} = |\beta| c,$$

then

$$\lim_{m\to\infty} \frac{1}{m^\beta} \sum_{k=0}^{m} a_k = c.$$

(ii) *If* γ_k *is the autocovariance function of a differenced ARFIMA process, then*

$$\lim_{m\to\infty} \frac{\sum_{k=0}^{m} \gamma_k(d)}{m^{2d-2}} = K.$$

(iii) *If* $\eta(k)$ *are the coefficients of the MA*(∞) *expansion of the differenced fractional noise,* $\eta(z) = \sum_{k=1}^{\infty} \eta_k z^k = (1-z)^{-d+1}$, *then*

$$\lim_{m\to\infty} \frac{\sum_{k=m}^{\infty} \eta_k^2}{m^{2d-3}} = K.$$

(iv) *For large* n,

$$\frac{\sum_{k=n}^{\infty} k^{2d-4}}{n^{2d-3}} \le K.$$

PROOF. (i) For simplicity, assume $\beta > 0$; the proof for $\beta < 0$ is analogous. By the assumption, for all $\varepsilon > 0$, there exists a k_0 such that for all $k > k_0$, $|a_k k^{1-\beta} - \beta c| < \varepsilon$. Thus,

$$\left|\frac{\sum_{k=0}^{m} a_k}{m^\beta} - c\right| \le \frac{\sum_{k=0}^{k_0-1} |a_k|}{m^\beta} + \frac{\sum_{k=k_0}^{m} |a_k - c\beta k^{\beta-1}|}{m^\beta} + \left|\frac{c\beta \sum_{k=k_0}^{m} k^{\beta-1}}{m^\beta} - c\right|.$$

Taking limits in m on both sides, with fixed k_0,

$$\lim_{m\to\infty}\left|\frac{\sum_{k=0}^{m} a_k}{m^\beta} - c\right| \le \varepsilon \lim_{m\to\infty} \frac{\sum_{k=k_0}^{m} k^{\beta-1}}{m^\beta} + \lim_{m\to\infty} c\beta \left|\frac{\sum_{k=k_0}^{m} k^{\beta-1}}{m^\beta} - \frac{1}{\beta}\right|.$$

Since

$$\lim_{m\to\infty} \frac{\sum_{k=k_0}^{m} k^{\beta-1}}{m^\beta} = \lim_{m\to\infty} \frac{1}{m} \sum_{k=k_0}^{m} (k/m)^{\beta-1} \le \int_0^1 x^{\beta-1}\, dx = \frac{1}{\beta},$$

for all $\varepsilon > 0$,

$$\lim_{m\to\infty}\left|\frac{\sum_{k=0}^{m} a_k}{m^\beta} - c\right| \le \frac{\varepsilon}{\beta}.$$

Since $\gamma_k \sim k^{2d-3}$, statement (ii) follows from (i). Parts (iii) and (iv) are proved as follows. Let $\beta = 2d-3 < 0$. Then

$$\lim_{n\to\infty} \frac{\sum_{k=n}^{\infty} k^{\beta-1}}{n^{\beta}} = \lim_{n\to\infty} \sum_{k=n}^{\infty} (k/n)^{\beta-1} \frac{1}{n}$$
$$= \lim_{n\to\infty} \sum_{j=0}^{\infty} (j/n+1)^{\beta-1} \frac{1}{n}.$$

By Pólya and Szegö [(1992), page 53], the last sum equals $\int_0^{\infty} (x+1)^{\beta-1}\,dx = 1/|\beta|$. □

LEMMA 3.3. *Let $m = n^{\beta}$ with $\beta > 0$. Then, for $\varepsilon > 0$ and for large n, the following inequalities hold*:

(i) $$\|T - \tilde{T}\| \le K n^{\beta(2d-2)},$$

(ii) $$\|T_{\nabla} - \tilde{T}_{\nabla}\| \le K n^{\beta(2d-2+\varepsilon)},$$

(iii) $$\|T_{\nabla^2} - \tilde{T}_{\nabla^2}\| \le K n^{\beta(2d-2+2\varepsilon)},$$

where K is a constant which is independent of $\boldsymbol{\theta}$ and n, and $\|A\| = \sup_x((x'A'Ax)/x'x)$ is the spectral norm of A.

PROOF. (i) The covariance matrices T and $\tilde{T}$ are of dimensions $n \times n$ satisfying

$$[T - \tilde{T}]_{l,j} = \int_{-\pi}^{\pi} [f_{\boldsymbol{\theta}}(\omega) - \tilde{f}_{m,\boldsymbol{\theta}}(\omega)] e^{i\omega(l-j)}\,d\omega = T(h),$$

with $h(\omega) = f_{\boldsymbol{\theta}}(\omega) - \tilde{f}_{m,\boldsymbol{\theta}}(\omega) = \sum_{j=0}^{\infty} (\gamma_j - \tilde{\gamma}_j) e^{i\omega j}$, where γ_j and $\tilde{\gamma}_j$ are the autocovariance functions of models (3.1) and (3.2), respectively. Therefore, $|h(\omega)| \le \sum_{j=0}^{m} |\gamma_j - \tilde{\gamma}_j| + \sum_{j=m+1}^{\infty} |\gamma_j| \le K m^{2d-2} = K n^{\beta(2d-2)}$, by Lemma 3.2 (ii). Furthermore, $T(h) = T(h^+ - h^-) = T(h^+) - T(h^-)$ and then $\|T(h)\| \le \|T(h^+)\| + \|T(h^-)\|$. Since $T(h^+) \ge 0$ and $T(h^-) \ge 0$, $\|T(h)\| \le \|T(h^+)^{1/2}\|^2 + \|T(h^-)^{1/2}\|^2$. On the other hand, $\|T(h^+)^{1/2}\|^2 = \sup_x((x'T(h^+)x)/x'x)$ and

$$x'T(h^+)x = \int_{-\pi}^{\pi} h^+(\omega) \left| \sum_{j=1}^{n} x_j e^{i\omega j} \right|^2 d\omega$$
$$\le \int_{-\pi}^{\pi} |h(\omega)| \left| \sum_{j=1}^{n} x_j e^{i\omega j} \right|^2 d\omega$$
$$\le K m^{2d-2} \int_{-\pi}^{\pi} \left| \sum_{j=1}^{n} x_j e^{i\omega j} \right|^2 d\omega$$
$$\le K m^{2d-2} x'x.$$

Hence, $\|T(h^+)^{1/2}\|^2 \le K m^{2d-2}$. Analogously, $\|T(h^-)^{1/2}\|^2 \le K m^{2d-2}$. Thus, $\|T - \tilde{T}\| \le K m^{2d-2} = K n^{\beta(2d-2)}$. Parts (ii) and (iii) are proved analogously. □

LEMMA 3.4. *Let $m = n^{\beta}$ with $\beta > 0$. Then as $n \to \infty$, uniformly in $\boldsymbol{\theta}$, we have the following*:

(i) *For* $\beta > 0$, $\|T - \tilde{T}\| \to 0$,
(ii) *for* $\beta > 0$, $\|T^{-1} - \tilde{T}^{-1}\| \to 0$,
(iii) *for* $\beta > 0$, $\|T_{\partial i} - \tilde{T}_{\partial i}\| \to 0$, *for* $i = 1, \dots, r$,
(iv) *for* $\beta \geq 1/2$, $\sqrt{n}\|A_i^{(1)} - \tilde{A}_i^{(1)}\| \to 0$, *for* $i = 1, \dots, r$,
(v) *for* $\beta \geq 1/2$, $\|A_{ij}^{(2)} - \tilde{A}_{ij}^{(2)}\| \to 0$, *for* $i, j = 1, \dots, r$,
(vi) *for* $\beta \geq 1/2$, $\|A_{ij}^{(3)} - \tilde{A}_{ij}^{(3)}\| \to 0$, *for* $i, j = 1, \dots, r$.

PROOF. (i) From Lemma 3.3(i), $\|T - \tilde{T}\| \leq Kn^{\beta(2d-2)}$. Thus, for $d < 1/2$, the result holds.

(ii) The matrix T satisfies $x'Tx \geq Kx'x$ uniformly in $\boldsymbol{\theta}$, where K is a constant. By part (i), for $\varepsilon > 0$, there exists n_0, independent of $\boldsymbol{\theta}$, such that for $n \geq n_0$, $\|T - \tilde{T}\| \leq \varepsilon$, where $\|A\|^2 = \sup_{x \in R^n}((x'AA'x)/x'x)$ is the square of the spectral norm of the matrix A. Thus, $|x'Tx - x'\tilde{T}x| \leq \varepsilon x'x$, and

$$x'\tilde{T}x \geq x'Tx - \varepsilon x'x \geq (k - \varepsilon)x'x.$$

It follows that $\|\tilde{T}^{-1}\| \leq K$ uniformly in $\boldsymbol{\theta}$. Using the property $\|AB\| \leq \|A\|\|B\|$ [see Dahlhaus (1989)], $\|T^{-1} - \tilde{T}^{-1}\| \leq \|T^{-1}\|\|\tilde{T}^{-1}\|\|T - \tilde{T}\| \leq K\|T - \tilde{T}\|$. By part (i), the last term tends to zero uniformly in $\boldsymbol{\theta}$ and hence (ii).

(iii) The proof of (iii) is analogous to (i) by means of the property $\|T_{\partial i} - \tilde{T}_{\partial i}\| \leq Kn^{\beta(2d-2+\varepsilon)}$ from Lemma 3.3(ii).

(iv) Observe that

$$\begin{aligned}
\sqrt{n}\|A_i^{(1)} - \tilde{A}_i^{(1)}\| &= \sqrt{n}\|T^{-1}T_{\partial i}T^{-1} - \tilde{T}^{-1}\tilde{T}_{\partial i}\tilde{T}^{-1}\| \\
&\leq \sqrt{n}\|T^{-1}T_{\partial i}T^{-1} - T^{-1}T_{\partial i}\tilde{T}^{-1}\| \\
&\quad + \sqrt{n}\|T^{-1}T_{\partial i}\tilde{T}^{-1} - \tilde{T}^{-1}T_{\partial i}\tilde{T}^{-1}\| \\
&\quad + \sqrt{n}\|\tilde{T}^{-1}(T_{\partial i} - \tilde{T}_{\partial i})\tilde{T}^{-1}\| \\
&\leq \sqrt{n}\|T^{-1}T_{\partial i}\|\|T^{-1} - \tilde{T}^{-1}\| \\
&\quad + \sqrt{n}\|\tilde{T}^{-1}T_{\partial i}\|\|T^{-1} - \tilde{T}^{-1}\| + \sqrt{n}\|\tilde{T}^{-1}\|^2\|T_{\partial i} - \tilde{T}_{\partial i}\| \\
&\leq K\sqrt{n}\|T^{-1} - \tilde{T}^{-1}\| + K\sqrt{n}\|T_{\partial i} - \tilde{T}_{\partial i}\| \\
&\leq Kn^{(2d-2)\beta+1/2}.
\end{aligned}$$

Thus, for $d < 1/2$ and $\beta \geq 1/2$, uniform convergence is obtained and hence (iv). Parts (v) and (vi) are proved analogously to part (iv). □

LEMMA 3.5. *For $\beta \geq 1/2$, as $n \to \infty$, uniformly in $\boldsymbol{\theta}$*,

$$\text{(i)} \qquad \frac{1}{\sqrt{n}}|\operatorname{tr}[T^{-1}T_{\partial i} - \tilde{T}^{-1}\tilde{T}_{\partial i}]| \to 0 \quad \textit{for } i = 1, \dots, r,$$

(ii) $$\frac{1}{n}|\operatorname{tr}[T^{-1}T_{\partial i}T^{-1}T_{\partial j} - \tilde{T}^{-1}\tilde{T}_{\partial i}\tilde{T}^{-1}\tilde{T}_{\partial j}]| \to 0 \quad \textit{for } i,\, j = 1, \dots, r,$$

(iii) $$\frac{1}{\sqrt{n}}|\operatorname{tr}[T^{-1}T_{\partial i\partial j} - \tilde{T}^{-1}\tilde{T}_{\partial i\partial j}]| \to 0 \quad \textit{for } i,\, j = 1, \dots, r.$$

PROOF. (i) Observe that

$$\begin{aligned}
\frac{1}{\sqrt{n}}|\operatorname{tr}[T^{-1}T_{\partial i} - \tilde{T}^{-1}\tilde{T}_{\partial i}]| &\le \frac{1}{\sqrt{n}}|\operatorname{tr}[T^{-1}T_{\partial i} - T^{-1}\tilde{T}_{\partial i}]| \\
&\quad + \frac{1}{\sqrt{n}}|\operatorname{tr}[(T^{-1} - \tilde{T}^{-1})\tilde{T}_{\partial i}]| \\
&\le \frac{1}{\sqrt{n}}\|T^{-1}\|\|T_{\partial i} - \tilde{T}_{\partial i}\| + \frac{1}{\sqrt{n}}\|T^{-1} - \tilde{T}^{-1}\|\|\tilde{T}_{\partial i}\| \\
&\le K\sqrt{n}\|T_{\partial i} - \tilde{T}_{\partial i}\| + K\sqrt{n}\|T^{-1} - \tilde{T}^{-1}\| \\
&\le Kn^{(2d-2)\beta+1/2}.
\end{aligned}$$

Thus, by parts (ii) and (iii) of Lemma 3.4, for $d < 1/2$ and $\beta \ge 1/2$, (i) holds. Parts (ii) and (iii) are proved analogously to part (i). □

LEMMA 3.6. *Let* $m = n^\beta$ *with* $\beta > 0$. *Then uniformly in* $\boldsymbol{\theta}$,

$$\lim_{n\to\infty} \frac{1}{n}\log\det\{\tilde{T}T^{-1}\} = 0.$$

PROOF.

$$\frac{1}{n}\log\det\{\tilde{T}T^{-1}\} \le \log\left\{\frac{1}{n}\operatorname{tr}\{\tilde{T}T^{-1}\}\right\} = \log\left\{\frac{1}{n}\operatorname{tr}\{T^{-1}[\tilde{T} - T]\} + 1\right\}.$$

Since $T^{-1} \le KI_n$ uniformly in $\boldsymbol{\theta}$,

$$\left|\left\{\frac{1}{n}\operatorname{tr}\{T^{-1}[\tilde{T} - T]\}\right\}\right| \le K\left|\left\{\frac{1}{n}\operatorname{tr}\{\tilde{T} - T\}\right\}\right| = K\sum_{k=m+1}^{\infty}\psi_k^2(\boldsymbol{\theta}).$$

By Lemma 3.1(i), for large k, $\psi_k(\boldsymbol{\theta})^2 \le Kk^{2d-4}$. Thus, for large m,

$$\sum_{k=m+1}^{\infty}\psi_k^2(\boldsymbol{\theta}) \le K\sum_{k=m+1}^{\infty}k^{2d-4}.$$

By Lemma 3.2(iv), $\sum_{k=m+1}^{\infty}k^{2d-4} \le Km^{2d-3}$. Thus, $\sum_{k=m+1}^{\infty}\psi_k^2(\boldsymbol{\theta}) \le Km^{2d-3} = Kn^{\beta(2d-3)}$. Since $2d-3 < 0$ and $\beta > 0$, $n^{\beta(2d-3)} \to 0$ as $n \to \infty$. This completes the proof. □

The ML estimates of the differenced fractional ARIMA model are obtained by maximizing

$$L_n(\boldsymbol{\theta}) = -\frac{1}{n}\log\det\tilde{T}_n(\boldsymbol{\theta}) - \frac{1}{n}Z_n'\tilde{T}_n^{-1}(\boldsymbol{\theta})Z_n,$$

and the *truncated* ML estimates of the differenced fractional ARIMA model are given by the maximization of

$$\tilde{L}_n(\boldsymbol{\theta}) = -\frac{1}{n}\log\det\tilde{T}_{n,m}(\boldsymbol{\theta}) - \frac{1}{n}Z_n'\tilde{T}_{n,m}^{-1}(\boldsymbol{\theta})Z_n.$$

LEMMA 3.7. *Let $m = n^{\beta}$, with $\beta > 0$. Then as $n \to \infty$,*

$$\sup_{\boldsymbol{\theta}}|\tilde{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})| \to 0 \quad a.s.$$

PROOF. Observe that

$$|\tilde{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})| \le \frac{1}{2n}\log\det\{\tilde{T}T^{-1}\} + \frac{1}{2n}Z_n'Z_n\|\tilde{T}^{-1} - T^{-1}\|.$$

By Lemma 3.6, $(1/2n)\log\det\{\tilde{T}T^{-1}\} \to 0$ and by page 133 of Hannan (1973), $(1/2n)Z_n'Z_n \to \gamma_0/2$ a.s. as $n \to \infty$ where γ_0 is the variance of the process $\{Z_t\}$. From Lemma 3.4(ii), $\|\tilde{T}^{-1} - T^{-1}\| \to 0$ uniformly in $\boldsymbol{\theta}$ as $n \to \infty$. This completes the proof of our result. □

3.2. *Proofs of Theorems.*

PROOF OF THEOREM 3.1 (Consistency). By Lemma 3.7, $\sup_{\boldsymbol{\theta}}|\tilde{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})| \to 0$ a.s., as n tends to infinity. Clearly,

$$-\tilde{L}_n(\boldsymbol{\theta}) \le |\tilde{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})| - L_n(\boldsymbol{\theta}).$$

Then,

$$\sup_{\boldsymbol{\theta}}\{-\tilde{L}_n(\boldsymbol{\theta})\} \le \sup_{\boldsymbol{\theta}}|\tilde{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}}\{-L_n(\boldsymbol{\theta})\}.$$

Equivalently,

$$-\inf_{\boldsymbol{\theta}}\tilde{L}_n(\boldsymbol{\theta}) \le \sup_{\boldsymbol{\theta}}|\tilde{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})| - \inf_{\boldsymbol{\theta}}L_n(\boldsymbol{\theta}),$$

or

$$\inf_{\boldsymbol{\theta}}L_n(\boldsymbol{\theta}) - \inf_{\boldsymbol{\theta}}\tilde{L}_n(\boldsymbol{\theta}) \le \sup_{\boldsymbol{\theta}}|\tilde{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})|.$$

Similarly,

$$\inf_{\boldsymbol{\theta}}\tilde{L}_n(\boldsymbol{\theta}) - \inf_{\boldsymbol{\theta}}L_n(\boldsymbol{\theta}) \le \sup_{\boldsymbol{\theta}}|\tilde{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})|.$$

Thus,

$$\left|\inf_{\boldsymbol{\theta}}\tilde{L}_n(\boldsymbol{\theta}) - \inf_{\boldsymbol{\theta}}L_n(\boldsymbol{\theta})\right| \le \sup_{\boldsymbol{\theta}}|\tilde{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})|.$$

Therefore,

$$\left|\inf_{\boldsymbol{\theta}}\tilde{L}_n(\boldsymbol{\theta}) - \inf_{\boldsymbol{\theta}}L_n(\boldsymbol{\theta})\right| \to 0 \quad \text{a.s.,} \tag{3.6}$$

as n goes to infinity. Let $U(\boldsymbol{\theta}_0)$ be a neighborhood of $\boldsymbol{\theta}_0$ with radius δ_0. By virtue of the triangle inequality,

$$\left|\inf_{\boldsymbol{\theta}} \tilde{L}_n(\boldsymbol{\theta}) - \inf_{U(\boldsymbol{\theta}_0)} L_n(\boldsymbol{\theta})\right| \le \left|\inf_{\boldsymbol{\theta}} \tilde{L}_n(\boldsymbol{\theta}) - \inf_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta})\right| + \left|\inf_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}) - \inf_{U(\boldsymbol{\theta}_0)} L_n(\boldsymbol{\theta})\right|.$$

By (3.6), the first term on the right-hand side converges to zero a.s., and hence in probability. On the other hand, $|\inf_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}) - \inf_{U(\boldsymbol{\theta}_0)} L_n(\boldsymbol{\theta})| \to 0$, in probability, since the MLE, $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ in probability, by Theorem 3.1 of Dahlhaus (1989). Thus,

$$\left|\inf_{\boldsymbol{\theta}} \tilde{L}_n(\boldsymbol{\theta}) - \inf_{U(\boldsymbol{\theta}_0)} L_n(\boldsymbol{\theta})\right| \to 0 \quad \text{in probability,}$$

which implies $\tilde{\boldsymbol{\theta}}_{n,m} \to \boldsymbol{\theta}_0$ in probability, as required. □

Proof of Theorem 3.2 (Central limit theorem). It suffices to prove the following results:

(i) $$\sup_{\boldsymbol{\theta}\in\Theta} \sqrt{n}|\nabla L_n(\boldsymbol{\theta}) - \nabla \tilde{L}_n(\boldsymbol{\theta})| \to 0 \quad \text{a.s.,}$$

(ii) $$\sup_{\boldsymbol{\theta}\in\Theta} |\nabla^2 L_n(\boldsymbol{\theta}) - \nabla^2 \tilde{L}_n(\boldsymbol{\theta})| \to 0 \quad \text{a.s.}$$

(i) The gradient of $L_n(\boldsymbol{\theta})$ can be written as

$$\nabla L_n(\boldsymbol{\theta}) = \frac{1}{2n}\operatorname{tr}[T^{-1}T_\nabla] - \frac{1}{2n} Z_n' A^{(1)} Z_n,$$

and

$$\nabla \tilde{L}_n(\boldsymbol{\theta}) = \frac{1}{2n}\operatorname{tr}[\tilde{T}^{-1}\tilde{T}_\nabla] - \frac{1}{2n} Z_n' \tilde{A}^{(1)} Z_n.$$

Hence,

$$\sqrt{n}|\nabla L_n(\boldsymbol{\theta}) - \nabla \tilde{L}_n(\boldsymbol{\theta})| \le \frac{1}{2\sqrt{n}}|\operatorname{tr}[T^{-1}T_\nabla - \tilde{T}^{-1}\tilde{T}_\nabla]| + \frac{Z_n' Z_n}{2n}\sqrt{n}\|A^{(1)} - \tilde{A}^{(1)}\|.$$

Observe that

$$\frac{1}{2\sqrt{n}}|\operatorname{tr}[T^{-1}T_\nabla - \tilde{T}^{-1}\tilde{T}_\nabla]| = \frac{1}{2\sqrt{n}}\left(\sum_{i=1}^{r} |\operatorname{tr}[T^{-1}T_{\partial i} - \tilde{T}^{-1}\tilde{T}_{\partial i}]|^2\right)^{1/2}.$$

By Lemma 3.5(i), $(1/2\sqrt{n})|\operatorname{tr}[T^{-1}T_{\partial i}] - \operatorname{tr}[\tilde{T}^{-1}\tilde{T}_{\partial i}]|$ goes to zero uniformly in $\boldsymbol{\theta}$, as n goes to infinity. Hence, the same result holds for $(1/2\sqrt{n})|\operatorname{tr}[T^{-1}T_\nabla - \operatorname{tr}\tilde{T}^{-1}\tilde{T}_\nabla]|$. On the other hand, $Z_n' Z_n/2n$ is asymptotically equal to $\gamma_0/2$ and

$$\sqrt{n}\|A^{(1)} - \tilde{A}^{(1)}\| = \sqrt{n}\left(\sum_{i=1}^{r} \|A_i^{(1)} - \tilde{A}_i^{(1)}\|^2\right)^{1/2}.$$

By Lemma 3.4(iv), for $i = 1, \dots, r$, $\sqrt{n}\|A_i^{(1)} - \tilde{A}_i^{(1)}\| \to 0$ as $n \to \infty$ uniformly in $\boldsymbol{\theta}$. Thus (i) is established.

(ii) The second derivatives of $L_n(\boldsymbol{\theta})$ can be written as

$$\nabla^2 L_n(\boldsymbol{\theta}) = -\frac{1}{2n}\operatorname{tr}[T^{-1}T_\nabla T^{-1}T_\nabla] + \frac{1}{2n}\operatorname{tr}[T^{-1}T_{\nabla^2}] + \frac{1}{2n}Z_n' A^{(2)} Z_n - \frac{1}{2n}Z_n' A^{(3)} Z_n,$$

and

$$\nabla^2 \tilde{L}_n(\boldsymbol{\theta}) = -\frac{1}{2n}\operatorname{tr}[\tilde{T}^{-1}\tilde{T}_\nabla \tilde{T}^{-1}\tilde{T}_\nabla] + \frac{1}{2n}\operatorname{tr}[\tilde{T}^{-1}\tilde{T}_{\nabla^2}] + \frac{1}{2n}Z_n' \tilde{A}^{(2)} Z_n - \frac{1}{2n}Z_n' \tilde{A}^{(3)} Z_n.$$

Hence,

$$\begin{aligned}|\nabla^2 L(\boldsymbol{\theta}) - \nabla^2 \tilde{L}(\boldsymbol{\theta})| &\le \frac{1}{2n}|\operatorname{tr}\{T^{-1}T_\nabla T^{-1}T_\nabla - \tilde{T}^{-1}\tilde{T}_\nabla \tilde{T}^{-1}\tilde{T}_\nabla\}| \\ &\quad + \frac{1}{2n}|\operatorname{tr}\{T^{-1}T_{\nabla^2} - \tilde{T}^{-1}\tilde{T}_{\nabla^2}\}| \\ &\quad + \frac{1}{2n}Z_n' Z_n \|A^{(2)} - \tilde{A}^{(2)}\| + \frac{1}{2n}Z_n' Z_n \|A^{(3)} - \tilde{A}^{(3)}\| \\ &= I + II + III, \quad \text{say.}\end{aligned}$$

Now

$$I = \frac{1}{2}\left(\sum_{i=1}^{r}\left[\frac{1}{n}|\operatorname{tr}\{T^{-1}T_{\partial i}T^{-1}T_{\partial i} - \tilde{T}^{-1}\tilde{T}_{\partial i}\tilde{T}^{-1}\tilde{T}_{\partial i}\}|\right]^2\right)^{1/2}.$$

By Lemma 3.5(ii), this term converges to zero for $\beta \ge 1/2$ uniformly in $\boldsymbol{\theta}$. Also,

$$\begin{aligned}II &= \frac{1}{2n}\left|\sum_{i=1}^{r}\operatorname{tr}\{T^{-1}T_{\partial i\partial i} - \tilde{T}^{-1}\tilde{T}_{\partial i\partial i}\}\right| \\ &\le \frac{1}{2}\sum_{i=1}^{r}\frac{1}{n}|\operatorname{tr}\{T^{-1}T_{\partial i\partial i} - \tilde{T}^{-1}\tilde{T}_{\partial i\partial i}\}|,\end{aligned}$$

and by virtue of Lemma 3.5(iii), the right-hand side converges to zero for $\beta \ge 1/2$ uniformly in $\boldsymbol{\theta}$. Finally, since ${Z_n}'Z_n/2n \sim \gamma_0/2$, $\|A^{(2)} - \tilde{A}^{(2)}\| = (\sum_{i,\,j=1}^{r}\|A_{ij}^{(2)} - \tilde{A}_{ij}^{(2)}\|^2)^{1/2} \to 0$ as $n \to \infty$ from Lemma 3.4 (v), and $\|A^{(3)} - \tilde{A}^{(3)}\| = (\sum_{i,\,j=1}^{r}\|A_{ij}^{(3)} - \tilde{A}_{ij}^{(3)}\|^2)^{1/2} \to 0$ according to Lemma 3.4(vi), $III \to 0$ and hence (ii) is established. □

PROOF OF THEOREM 3.3 (Efficiency). Let $I_n(\boldsymbol{\theta}_0)$ and $\tilde{I}_n(\boldsymbol{\theta}_0)$ be the Fisher information matrices of the exact and truncated model evaluated at $\boldsymbol{\theta}_0$, respectively. Then,

$$\frac{1}{n}I_n(\boldsymbol{\theta}_0) - \frac{1}{n}\tilde{I}_n(\boldsymbol{\theta}_0) = n\operatorname{tr}\{T^{-1}T_\nabla T^{-1}T_\nabla - \tilde{T}^{-1}\tilde{T}_\nabla \tilde{T}^{-1}\tilde{T}_\nabla\}.$$

An application of Lemma 3.5(ii) yields

$$\lim_{n\to\infty}\left(\frac{1}{n}I_n(\boldsymbol{\theta}_0)-\frac{1}{n}\tilde{I}_n(\boldsymbol{\theta}_0)\right)=0.$$

Therefore, by Theorem 5.1 of Dahlhaus (1989),

$$\lim_{n\to\infty}\frac{1}{n}\tilde{I}_n(\boldsymbol{\theta}_0)=\Gamma(\boldsymbol{\theta}_0). \qquad \square$$

4. Monte Carlo experiments. In order to assess the performance of the approximate state space methodology, a number of Monte Carlo simulations are carried out. The data are simulated from a fractional noise process

$$y_t=(1-B)^{-d}\varepsilon_t, \tag{4.1}$$

and from an ARFIMA$(1,d,1)$ process

$$(1-\phi B)y_t=(1-\theta B)(1-B)^{-d}\varepsilon_t. \tag{4.2}$$

In both cases, the white noise sequences $\{\varepsilon_t\}$ are independent standard normal random variables.

The values of the long-memory parameter d vary from 0.05 to 0.40, and the sample sizes vary from 100 to 10,000 observations. The simulation results are reported in Tables 1 to 5. In Tables 1 to 3, the first column represents the *true* value of the long-memory parameter d. The second column corresponds to the average of the estimated parameter $\hat{d}$, whereas the third column is the estimated standard deviation of $\hat{d}$. The fourth column shows the average of the MLE $\hat{\sigma}$ of the standard deviation of the white noise σ, and the fifth column gives the estimated standard deviation of $\hat{\sigma}$. Similarly, Tables 4 and 5 give the results for an ARFIMA$(1,d,1)$ model. All averages and standard deviations are based on 100 repetitions.

Table 1 displays the results for the exact ML estimation with a sample size of 100 observations. As shown in the second column, the means of the estimated parameters $\hat{d}$ are close to their expected value, d. The sample standard deviations (see third column) are also close to their expected value of 0.077. Similarly, the sample means and standard deviations for $\hat{\sigma}$ are as expected, 1 and 0.071, respectively.

TABLE 1

Exact maximum likelihood estimation of d and σ, $n=100$

d	$\hat{d}$	S.D. $(\hat{d})$	$\hat{\sigma}$	S.D. $(\hat{\sigma})$
0.10	0.098	0.071	0.986	0.067
0.20	0.199	0.072	1.01	0.066
0.30	0.289	0.067	0.984	0.071
0.40	0.391	0.072	0.996	0.069

TABLE 2
Truncated maximum likelihood estimation of d and σ, $m = 6$, and $n = 100$

d	$\hat{d}$	S.D. ($\hat{d}$)	$\hat{\sigma}$	S.D. ($\hat{\sigma}$)
0.10	0.104	0.075	1.01	0.070
0.20	0.203	0.078	0.995	0.069
0.30	0.297	0.069	0.997	0.065
0.40	0.410	0.074	1.02	0.072

Simulation results for the approximate MLE are given in Table 2. The sample size is 100 and the truncation parameter is $m = 6$. The observed means for $\hat{d}$ are close to their expected values, and the standard deviations are, as expected, close to 0.077. Analogously, the sample mean and standard deviations of $\hat{\sigma}$ are also in close agreements with their expected values, 1 and 0.071 respectively. The estimated values in Table 2 compare favorably with the results reported in Li and McLeod (1986) and Sowell (1992) where the exact MLE approach is considered. The computational time, however, is much faster for the truncated MLE approach used in Table 2.

Table 3 consists of simulations of 10, 000 observations from the same class of models. As indicated clearly in this case, the truncated MLE works extremely well with the truncation parameter m as small as 14.

Tables 4 and 5 display the results for the ARFIMA$(1, d, 1)$ experiments with a sample size of 1,000 observations. Simulation results for the exact and approximate MLE are shown in Table 4 and Table 5, respectively. For the approximate MLE, the truncation parameter is $m = 30$. Again, with such a small m relative to a sample of size 1,000, the values reported in Tables 4 and 5 are in close agreement, demonstrating the efficient performance of the truncated MLE.

These Monte Carlo studies suggest that the exact and the approximate ML state space methodology developed in this paper performs well even for samples with only 100 observations. In the fractional noise case, as the sample size increases from 100 to 10,000, the truncation parameter only has to be increased from 6 to 14. This indicates that when applying the approximate

TABLE 3
Truncated maximum likelihood estimation of d and σ, $m = 14$, and $n = 10,000$

d	$\hat{d}$	S.D. ($\hat{d}$)	$\hat{\sigma}$	S.D. ($\hat{\sigma}$)
0.05	0.048	0.009	1.00	0.007
0.10	0.100	0.008	1.00	0.005
0.20	0.202	0.009	1.00	0.005
0.30	0.306	0.009	1.00	0.006
0.40	0.398	0.009	1.00	0.006

TABLE 4
Exact maximum likelihood estimation of ARFIMA$(1, d, 1)$ *model with* $\phi = -0.50$, $\theta = -0.30$ *and* $n = 1,000$

d	$\hat{d}$	$\hat{\phi}$	$\hat{\theta}$	$\hat{\sigma}$	S.D. ($\hat{d}$)	S.D. ($\hat{\phi}$)	S.D. ($\hat{\theta}$)	S.D. ($\hat{\sigma}$)
0.10	0.103	−0.501	−0.294	0.999	0.034	0.112	0.140	0.023
0.20	0.196	−0.497	−0.299	1.01	0.038	0.135	0.163	0.021
0.30	0.299	−0.508	−0.306	1.00	0.037	0.130	0.139	0.022
0.40	0.401	−0.505	−0.307	1.01	0.036	0.128	0.155	0.023

TABLE 5
Truncated maximum likelihood estimation of ARFIMA$(1, d, 1)$ *model with* $\phi = -0.50$, $\theta = -0.30$, $m = 30$ *and* $n = 1,000$

d	$\hat{d}$	$\hat{\phi}$	$\hat{\theta}$	$\hat{\sigma}$	S.D. ($\hat{d}$)	S.D. ($\hat{\phi}$)	S.D. ($\hat{\theta}$)	S.D. ($\hat{\sigma}$)
0.10	0.105	−0.507	−0.309	0.998	0.035	0.115	0.138	0.022
0.20	0.209	−0.511	−0.292	1.02	0.037	0.138	0.144	0.023
0.30	0.301	−0.491	−0.289	1.00	0.037	0.129	0.151	0.024
0.40	0.397	−0.508	−0.305	1.01	0.040	0.131	0.162	0.022

ML technique, a small truncation parameter m relative to the sample size n still provides very satisfactory results.

5. Conclusions. State space systems are applied to the analysis of long-memory processes. Unlike the standard ARMA case, a state space representation for an ARFIMA model has to be infinite-dimensional. Despite this fact, computation of the likelihood function can be carried out in a finite number of steps. An approximation to the likelihood functions via truncation provides an efficient means to calculate the MLE. This procedure possesses similar desirable asymptotic properties as the exact MLE. Furthermore, simulation studies suggest that the proposed approach can be extremely efficient. This new approach offers a promising alternative to estimating the parameters of a long-memory process.

Acknowledgments. We thank the Editors, two anonymous referees and Dr. G. Petris for helpful comments.

REFERENCES

BERAN, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, New York.

BLOOMFIELD, P. (1992). Trends in global temperature. *Climatic Change* **21** 1–16.

BROCKWELL P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.

CHEUNG, Y. W. and DIEBOLD, F. (1994). On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. *J. Econometrics* **62** 317–350.

DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749–1766.

FOX, R. and TAQQU, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532.

GEWEKE, J. and PORTER-HUDAK, S. (1983). The estimation and application of long-memory time series models. *J. Time Ser. Anal.* **4** 221–237.

GIRAITIS, L. and SURGAILIS, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. *Probab. Theory Related Fields* **86** 87–104.

HANNAN, E. J. (1973). The asymptotic theory of linear time-series models. *J. Appl. Probab.* **10** 130–145.

HANNAN, E. J. and DEISTLER, M. (1988). *Statistical Theory of Linear Systems*. Wiley, New York.

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

HURST, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* **116** 770–779.

HURVICH, C. M. and BELTRAO, K. I. (1993). Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series. *J. Time Ser. Anal.* **14** 455–472.

JONES, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **22** 389–395.

LI, W. K. and MCLEOD, A. I. (1986). Fractional time series modelling. *Biometrika* **73** 217–221.

MANDELBROT, B. B. and TAQQU, M. S. (1979). Robust R/S analysis of long run serial correlation. In *Proceedings of the 42nd Session of ISI, Manila* **2** 69–100.

MANDELBROT, B. B. and WALLIS, J. R. (1969). Computer experiments with fractional Gaussian noises. *Water Resources Research* **6** 225–267.

NEWTON, H. J. (1988). *TIMESLAB: A Time Series Analysis Laboratory*. Wadsworth, Belmont, CA.

PÓLYA, G. and SZEGÖ, G. (1992). *Problems and Theorems in Analysis I*. Springer, New York.

ROBINSON, P. M. (1994). Time series with strong dependence. In *Advances in Econometrics, Sixth World Congress* (C. A. Sims, ed.) **1** 47–95. Cambridge Univ. Press.

ROBINSON, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661.

SOWELL, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *J. Econometrics* **53** 165–188.

DEPARTMENT OF STATISTICS
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213-3890
E-MAIL: chan@stat.cmu.edu

DEPARTAMENTO DE PROBABILIDAD Y ESTADÍSTICA
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
CASILLA 306, SANTIAGO 22
CHILE
E-MAIL: wilfredo@mat.puc.cl