ESTIMATION OF LONG-MEMORY TIME SERIES MODELS: A SURVEY OF DIFFERENT LIKELIHOOD-BASED METHODS

Ngai Hang Chan and Wilfredo Palma

ABSTRACT

Since the seminal works by Granger and Joyeux (1980) and Hosking (1981), estimations of long-memory time series models have been receiving considerable attention and a number of parameter estimation procedures have been proposed. This paper gives an overview of this plethora of methodologies with special focus on likelihood-based techniques. Broadly speaking, likelihood-based techniques can be classified into the following categories: the exact maximum likelihood (ML) estimation (Sowell, 1992; Dahlhaus, 1989), ML estimates based on autoregressive approximations (Granger & Joyeux, 1980; Li & McLeod, 1986), Whittle estimates (Fox & Taqqu, 1986; Giraitis & Surgailis, 1990), Whittle estimates with autoregressive truncation (Beran, 1994a), approximate estimates based on the Durbin–Levinson algorithm (Haslett & Raftery, 1989), state-space-based maximum likelihood estimates for ARFIMA models (Chan & Palma, 1998), and estimation of stochastic volatility models (Ghysels, Harvey, & Renault, 1996; Breidt, Crato, & de Lima, 1998; Chan & Petris, 2000) among others. Given the diversified

Econometric Analysis of Financial and Economic Time Series/Part B
Advances in Econometrics, Volume 20, 89–121

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20023-3

applications of these techniques in different areas, this review aims at providing a succinct survey of these methodologies as well as an overview of important related problems such as the ML estimation with missing data (Palma & Chan, 1997), influence of subsets of observations on estimates and the estimation of seasonal long-memory models (Palma & Chan, 2005). Performances and asymptotic properties of these techniques are compared and examined. Inter-connections and finite sample performances among these procedures are studied. Finally, applications to financial time series of these methodologies are discussed.

1. INTRODUCTION

Long-range dependence has become a key aspect of time series modeling in a wide variety of disciplines including econometrics, hydrology and physics, among many others. Stationary long-memory processes are defined by autocorrelations decaying slowly to zero or spectral density displaying a pole at zero frequency. A well-known class of long-memory models is the autoregressive fractionally integrated moving average (ARFIMA) processes introduced by Granger and Joyeux (1980) and Hosking (1981). An ARFIMA process $\{y_t\}$ is defined by

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

where $\Phi(B) = 1 + \phi_1 B + \cdots + \phi_p B^p$ and $\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ are the autoregressive and moving average (ARMA) operators, respectively; $(1-B)^d$ is the fractional differencing operator defined by the binomial expansion $(1-B)^d = \sum_{j=0}^{\infty} \binom{d}{j} B^j$ and $\{\varepsilon_t\}$ a white noise sequence with variance σ_ε^2. Under the assumption that the roots of the polynomials $\Phi(B)$ and Θ(B) are outside the unit circle and $|d| < 1/2$, the ARFIMA(p,d,q) process is second-order stationary and invertible. The spectral density of this process is

$$f(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} |1 - e^{i\omega}|^{-2d} |1 - \Phi(e^{i\omega})|^{-2} |1 - \Theta(e^{i\omega})|^2$$

and its ACF may be written as

$$\gamma(k) = \int_{-\pi}^{\pi} f(\omega) e^{i\omega k} d\omega \tag{2}$$

Estimation of long-memory models has been considered by a large number of authors. Broadly speaking, most of the estimation methodologies

proposed in the literature can be classified into the time-domain and the spectral-domain procedures. In the first group, we have the exact maximum likelihood estimators (MLE) and the quasi maximum likelihood estimators (QMLE), see for example the works by Granger and Joyeux (1980), Sowell (1992) and Beran (1994a), among others. In the second group, we have for instance, the Whittle and the semi-parametric estimators (see Fox & Taqqu, 1986; Giraitis & Surgailis, 1990; Robinson, 1995) among others.

In this article, we attempt to give an overview of this plethora of estimation methodologies, examining their advantages and disadvantages, computational aspects such as their arithmetic complexity, finite sample behavior and asymptotic properties.

The remaining of this paper is structured as follows. Exact ML methods are reviewed in Section 2, including the Cholesky decomposition, the Levinson–Durbin algorithm and state-space methodologies. Section 3 discusses approximate ML methods based on truncations of the infinite autoregressive (AR) expansion of the long-memory process, including the Haslett and Raftery estimator and the Beran method. Truncations of the infinite moving average (MA) expansion of the process are discussed in Section 4 along with the corresponding Kalman filter recursions. The spectrum-based Whittle method and semi-parametric procedures are studied in Section 5. Extensions to the non-Gaussian case are also addressed in this section. The problem of parameter estimation of time series with missing values is addressed in Section 6 along with an analysis of the effects of data gaps on the estimates. Section 7 discusses methodologies for estimating time series displaying both persistence and cyclical behavior. Long-memory models for financial time series are addressed in Section 8, while final remarks are presented in Section 9.

2. EXACT MAXIMUM LIKELIHOOD METHOD

Under the assumption that the process $\{y_t\}$ is Gaussian and has zero mean, the log-likelihood function may be written as

$$\mathscr{L}(\boldsymbol{\theta}) = -\frac{1}{2}\log\ \det\ \Gamma_\theta - \frac{1}{2}\, Y'\Gamma_\theta^{-1}\, Y \tag{3}$$

where $Y = (y_1, \ldots, y_n)'$, $\Gamma_\theta = \mathrm{var}(Y)$ and θ is the model parameter vector. Hence, the ML estimate is given by $\hat{\theta} = \mathrm{argmax}_\theta \mathscr{L}(\boldsymbol{\theta})$.

Expression (3) involves the calculation of the determinant and the inverse of Γ_θ. As described next, the well-known Cholesky decomposition algorithm

can be used to carry out these calculations. Further details can be found in Sowell (1992). Here we give an overview of this and other related methodologies for computing (3) such as the Levinson–Durbin algorithm and the Kalman filter recursions.

2.1. Cholesky Decomposition

Since the variance–covariance matrix Γ_θ is positive definite, it can be decomposed as

$$\Gamma_\theta = L_1 L_1'$$

where L_1 is a lower triangular matrix. This Cholesky decomposition provides the determinant $\det \Gamma_\theta = (\det L_1)^2 = \prod_{i=1}^n l_{ii}^2$, where l_{ii} denotes the ith-diagonal element of L_1. Furthermore, the inverse of Γ_θ can be obtained as $\Gamma_\theta^{-1} = (L_1^{-1})' L_1^{-1}$, where the inverse of L_1 can be computed by means of a very simple procedure, see for example Press, Teukolsky, Vetterling, and Flannery (1992, 89ff).

Observe that while the inversion of a nonsingular square matrix $n \times n$ has arithmetic complexity of order $\mathcal{O}(n^3)$, the Cholesky algorithm is of order $\mathcal{O}(n^3/6)$, cf. Press et al. (1992, p. 34).

2.2. Levinson–Durbin Algorithm

Since for large sample sizes, the Cholesky algorithm could be inefficient, faster methods for calculating the log-likelihood function have been developed. These numerical procedures, designed to exploit the Toeplitz structure of the variance–covariance matrix of an second-order stationary process, are based on the seminal works by Levinson (1947) and Durbin (1960).

Let $\hat{y}_1 = 0$ and $\hat{y}_{t+1} = \phi_{t1} y_t + \cdots + \phi_{tt} y_1$ for $t = 1, \ldots, n-1$ be the one-step predictors of the process $\{y_t\}$ based on the finite past $\{y_1, \ldots y_{t-1}\}$, where the partial regression coefficients ϕ_{tj} are given by the recursive equations

$$\begin{aligned}
&\phi_{tt} = \left[\gamma(t) - \sum_{i=1}^{t-1} \phi_{t-1,i}\gamma(t-i)\right] / v_{t-1} \\
&\phi_{tj} = \phi_{t-1,j} - \phi_{tt}\phi_{t-1,t-j}, \qquad j = 1, \ldots, n-1 \\
&v_0 = \gamma(0) \\
&v_t = v_{t-1}[1 - \phi_{tt}^2], \qquad t = 1, \ldots, n-1
\end{aligned}$$

Now, if $e_t = y_t - \hat{y}_t$ is the prediction error and $\boldsymbol{e} = (e_1, \ldots, e_n)'$, then $\boldsymbol{e} = L_2 Y$, where L_2 is the lower triangular matrix

$$L_2 = \begin{pmatrix} 1 & & & & \\ -\phi_{11} & 1 & & & \\ -\phi_{22} & -\phi_{21} & 1 & & \\ -\phi_{33} & -\phi_{32} & -\phi_{31} & 1 & \\ \vdots & \vdots & & & \\ -\phi_{n-1,n-1} & -\phi_{n-1,n-2} & \cdots & -\phi_{n-1,1} & 1 \end{pmatrix}$$

Thus, Γ_θ may be decomposed as $\Gamma_\theta = L_2 D L_2'$, where $D = \text{diag}(v_0, \ldots, v_{n-1})$. Consequently, $\det \Gamma_\theta = \prod_{j=1}^{n} v_{j-1}$ and $Y'\Gamma_\theta^{-1} Y = \boldsymbol{e}' D^{-1} \boldsymbol{e}$. Hence, the log-likelihood function may be written as

$$\mathscr{L}(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{n} \log v_{t-1} - \frac{1}{2}\sum_{t=1}^{n} e_t^2 / v_{t-1}$$

The arithmetic complexity of this algorithm is $\mathcal{O}(2n^2)$ for a linear stationary process, see Ammar (1996). However, for some Markovian processes such as the family of ARMA models, the Levinson–Durbin algorithm can be implemented in only $\mathcal{O}(n)$ operations, see for example Section 5.3 of Brockwell and Davis (1991). Unfortunately, ARFIMA models are not Markovian and this reduction in operations count does not apply to them.

2.3. Calculation of Autocovariances

A critical aspect in the implementation of the Cholesky and the Levinson–Durbin algorithms is the calculation of the ACF process. A closed form expression for the ACF of an ARFIMA model is given in Sowell (1992). Here, we briefly review this and other alternative methods of obtaining the ACF of a long-memory process.

Observe that the polynomial $\Phi(B)$ defined in (1) can be written as

$$\Phi(B) = \prod_{i=1}^{p} (1 - \rho_i B)$$

where $\{\rho_i\}$ are the roots of the polynomial $\Phi(z^{-1})$. Assuming that all these roots have multiplicity one, it can be deduced from (2) that

$$\gamma(k) = \sigma_\varepsilon^2 \sum_{i=-q}^{q} \sum_{j=1}^{p} \psi(i)\xi_j C(d, p + i - k,\ \rho_j)$$

with $\psi(i) = \sum_{j=\max(0,l)}^{\min(q,q-1)} \theta_j \theta_{j-l}$, $\xi_j = \left[\rho_j \prod_{i=1}^{p}(1-\rho_i\rho_j)\prod_{m\neq j}(\rho_j-\rho_m)\right]^{-1}$ and

$$c(d,h,\rho) = \frac{\Gamma(1-2d)\Gamma(d+h)}{\Gamma(1-d+h)\Gamma(1-d)\Gamma(d)} \times [\rho^{2p}F(d+h,\ 1,\ 1-d+h,\ \rho) + F(d-h,\ 1,\ 1-d-h,\ \rho) - 1]$$

where $\Gamma(\cdot)$ is the Gamma function and $F(a, b, c, x)$ is the Gaussian hypergeometric function, see (Gradshteyn & Ryzhik, 2000, Section 9.1).

An alternative method for calculating the ACF is the so-called splitting method, see Bertelli and Caporin (2002). This technique is based on the decomposition of the model into its ARMA and its fractional integrated (FI) parts. Let $\gamma_1(\cdot)$ be the ACF of the ARMA component and $\gamma_2(\cdot)$ be the ACF of the fractional noise. Then, the ACF of the corresponding ARFIMA process is given by the convolution of these two functions:

$$\gamma(k) = \sum_{h=-\infty}^{\infty} \gamma_1(h)\gamma_2(k-h)$$

If this infinite sum is truncated to m summands, we obtain the approximation

$$\gamma(k) \approx \sum_{h=-m}^{m} \gamma_1(h)\gamma_2(k-h)$$

Thus, the ACF can be efficiently calculated with a great level of precision, see for instance the numerical experiments reported by Bertelli and Caporin (2002) for further details.

2.4. Exact State-Space Method

Another method for computing exact ML estimates is provided by the state-space system theory. In this Section, we review the application of Kalman filter techniques to long-memory processes. Note that since these processes are not Markovian, all the state space representations are infinite dimensional as shown by Chan and Palma (1998). Despite this fact, the Kalman filter equations can be used to calculate the exact log-likelihood (3) in a finite number of steps.

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

$$y_t = \frac{\Theta(B)}{\Phi(B)}(1-B)^{-d}\varepsilon_t = \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j} \tag{4}$$

where φ_j are the coefficients of $\varphi(z) = \sum_{j=0}^{\infty} \varphi_j z^j = \Theta(z)\Phi(z)^{-1}(1-z)^{-d}$. An infinite-dimensional state-space representation may be constructed as follows. From Eq. (4), a state space system may be written as

$$X_{t+1} = FX_t + H\varepsilon_t \tag{5}$$

$$y_t = GX_t + \varepsilon_t \tag{6}$$

where

$$X_t = [y(t|t-1)\ y(t+1|t-1)\ y(t+2|t-1)\ \ldots\ ]' \tag{7}$$

$$y(t|j) = E[y_t|y_j, y_{j-1}, \ldots] \tag{8}$$

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \ddots & & \end{bmatrix}, \qquad H = [\varphi_1, \varphi_2, \ldots]', \text{ and } G = [1, 0, 0, \ldots] \tag{9}$$

The log-likelihood 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. The Kalman algorithm is as follows: 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

$$\hat{X}_1 = E[X_1] \tag{10}$$

$$\Omega_1 = E[X_1 X_1'] - E[\hat{X}_1 \hat{X}'_1] \tag{11}$$

$$\Delta_t = \omega_{11}^{(t)} + 1 \tag{12}$$

$$\omega_{ij}^{(t+1)} = \omega_{i+1,j+1}^{(t)} + \varphi_i \varphi_j - \frac{(\omega_{i+1,1}^{(t)} + \varphi_i)(\omega_{j+1,1}^{(t)} + \varphi_j)}{\omega_{11}^{(t)} + 1} \tag{13}$$

and

$$\hat{X}_{t+1} = (\hat{X}_1^{(t+1)}, \hat{X}_2^{(t+1)}, \ldots)' = (\hat{X}_i^{(t+1)})'_{i=1,2,\ldots} \tag{14}$$

where

$$\hat{X}_i^{(t+1)} = \hat{X}_{i+1}^{(t)} + \frac{(y_t - \hat{X}_1^{(t)})(\omega_{i+1,1}^{(t)} + \varphi_i)}{\omega_{11}^{(t)} + 1} \tag{15}$$

$$\hat{y}_{t+1} = G\hat{X}_{t+1} = \hat{X}_1^{(t+1)} \tag{16}$$

and the log-likelihood function is given by

$$\mathscr{L}(\boldsymbol{\theta}) = -\frac{1}{2}\left\{ n \log 2\pi + \sum_{t=1}^{n} \log \Delta_t + n \log \sigma_\varepsilon^2 + \frac{1}{\sigma_\varepsilon^2} \sum_{t=1}^{n} \frac{(y_t - \hat{y}_t)^2}{\Delta_t} \right\}$$

Although 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 due to Chan and Palma (1998).

Theorem 1. Let $\{y_1, \ldots, y_n\}$ be a finite sample of an ARFIMA(p,d,q) process. If Ω_1 is the variance of the initial state X_1 of the infinite-dimensional representation (5)–(9), then the computation of the exact likelihood function (3) depends only on the first n components of the Kalman Eqs. (10)–(16).

It is worth noting that as a consequence of Theorem 1, given a sample of n observations from an ARFIMA process, the calculation 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.

The arithmetic complexity of this algorithm is $\mathcal{O}(n^3)$. Therefore, it is comparable to the Cholesky decomposition but it is less efficient than the Levinson–Durbin procedure. The Kalman approach is advisable for moderate sample sizes or for handling time series with missing values. The state space framework provides a simple solution to this problem. Note that the Levinson–Durbin method is no longer appropriate when the series displays missing values since the variance–covariance matrix of the incomplete data does not have a Toeplitz structure.

2.5. Asymptotic Results for the Exact MLE

The consistency, asymptotic normality and efficiency of the MLE have been established by Yajima (1985) for the fractional noise process and by Dahlhaus (1989) for the general case including the ARFIMA model.

Let $\hat{\boldsymbol{\theta}}_n$ be the value that maximizes the exact log-likelihood where

$$\boldsymbol{\theta} = (\phi_1, \ldots, \phi_p, \phi_1, \ldots, \theta_q, d, \sigma_\varepsilon)'$$

is a $p + q + 2$ dimensional parameter vector and let $\boldsymbol{\theta}_0$ be the *true* parameter. Assume that the regularity conditions listed in Dahlhaus (1989) hold.

Theorem 2. (Consistency) $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ in probability as $n \to \infty$.
(Central Limit Theorem) $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to N(0, \Gamma^{-1}(\boldsymbol{\theta}_0))$, as $n \to \infty$, where $\Gamma(\boldsymbol{\theta}) = (\Gamma_{ij}(\boldsymbol{\theta}))$ with

$$\begin{aligned} &\Gamma_{ij}(\boldsymbol{\theta}) = \tfrac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial \log k(\omega,\boldsymbol{\theta})}{\partial\theta_i}\right\}\left\{\frac{\partial \log k(\omega,\boldsymbol{\theta})}{\partial\theta_j}\right\}dw \\ &\text{and } k(\omega,\boldsymbol{\theta}) = |\textstyle\sum_{j=0}^{\infty}\psi_j(\boldsymbol{\theta})e^{ij\omega}|^2 \end{aligned} \tag{17}$$

(Efficency) $\hat{\boldsymbol{\theta}}_n$ is an efficient estimator of $\boldsymbol{\theta}_0$.

The next two sections discuss ML methods based on AR and MA approximations.

3. AUTOREGRESSIVE APPROXIMATIONS

Since the computation of exact ML estimates is computationally highly demanding, several authors have considered the use of AR approximations to speed up the calculation of parameter estimates. In particular, this approach has been adopted by Granger and Joyeux (1980), Li and McLeod (1986), Haslett and Raftery (1989), Beran (1994b), Shumway and Stoffer (2000) and Bhansali and Kokoszka (2003), among others.

Most of these techniques are based on the following estimation strategy. Consider a long-memory process $\{y_t\}$ defined by the AR(∞) expansion

$$y_t - \pi_1(\boldsymbol{\theta})y_{t-1} - \pi_2(\boldsymbol{\theta})y_{t-2} - \cdots = \varepsilon_t$$

where $\pi_j(\boldsymbol{\theta})$ are the coefficients of $\Phi(B)\Theta^{-1}(B)(1-B)^d$. In practice, only a finite number of observations is available, $\{y_1, \ldots, y_n\}$, therefore the following truncated model is considered

$$y_t - \pi_1(\boldsymbol{\theta})y_{t-1} - \cdots - \pi_m(\boldsymbol{\theta})y_{y-m} = \tilde{\varepsilon}_t$$

for $m < t \le n$. The ML estimate $\hat{\boldsymbol{\theta}}_n$ is found by minimizing

$$\mathscr{L}_0(\boldsymbol{\theta}) = \sum_{t=m+1}^{n}[y_t - \pi_1(\boldsymbol{\theta})y_{t-1} - \cdots - \pi_m(\boldsymbol{\theta})y_{t-m}]^2$$

Upon this basic framework, many refinements can be made to improve the quality of these estimates. In what follows next, we describe some of these refinements. For simplicity, any estimator produced by the maximization of an approximation of the Gaussian likelihood function (3) will be called QMLE.

3.1. Haslett–Raftery Method

The following technique was proposed by Haslett and Raftery (1989). Consider the ARFIMA process (1). An approximate one-step predictor of y_t is given by

$$\hat{y}_t = \Phi(B)\Theta(B)^{-1} \sum_{j=1}^{t-1} \phi_{tj} y_{t-j} \tag{18}$$

with prediction error variance

$$v_t = \mathrm{var}(y_t - \hat{y}_t) = \sigma_y^2 \kappa \prod_{j=1}^{t-1} (1 - \phi_{jj}^2)$$

where $\sigma_y^2 = \mathrm{var}(y_t)$, κ is the ratio of the innovations variance to the variance of the ARMA(p,q) process as given by Eq. (3.4.4) of Box, Jenkins, and Reinsel (1994) and

$$\phi_{tj} = -\binom{t}{j} \frac{\Gamma(j-d)\Gamma(t-d-j+1)}{\Gamma(-d)\Gamma(t-d+1)}, \quad \text{for } j = 1, \ldots, t$$

To avoid the computation of a large number of coefficients ϕ_{tj}, the last term of the predictor (18) is approximated by

$$\sum_{j=1}^{t-1} \phi_{tj} y_{t-j} \approx \sum_{j=1}^{M} \phi_{tj} y_{t-j} - \sum_{j=M+1}^{t-1} \pi_j y_{t-j} \tag{19}$$

since $\phi_{tj} \sim -\pi_j$ for large j, cf. Hosking (1981), where for simplicity π_j denotes $\pi_j(\boldsymbol{\theta})$ and $a_j \sim b_j$ means that $a_j/b_j \to 1$, as $j \to \infty$.

An additional approximation is made to the second term on the right-hand-side of (19):

$$\sum_{j=M+1}^{t-1} \pi_j y_{t-j} \approx M \pi_M d^{-1} \left[1 - \left(\frac{M}{t} \right)^d \right] \bar{y}_{M+1,t-1-M}$$

where $\bar{y}_{M+1,t-1-M} = \frac{1}{t-1-2M} \sum_{j=M+1}^{t-1-M} y_j$. Hence, a QMLE $\hat{\boldsymbol{\theta}}_n$ is obtained by maximizing

$$\mathscr{L}_1(\boldsymbol{\theta}) = \text{constant} \; - \frac{1}{2} n \; \log[\hat{\sigma}_\varepsilon^2(\boldsymbol{\theta})]$$

with

$$\hat{\sigma}_\varepsilon^2(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\frac{(y_t - \hat{y}_t)^2}{v_t}$$

The arithmetic complexity of the Haslett and Raftery method is of order $\mathcal{O}(nM)$. For a fixed M, the algorithm is of order $\mathcal{O}(n)$, which is much faster compared to the Levinson–Durbin method. Haslett and Raftery (1989) suggest $M = 100$. Note that, when $M = n$, the exact ML estimated is obtained. However, the arithmetic complexity in that case becomes $\mathcal{O}(n^2)$ and no gain is obtained as compared to the Levinson–Durbin approach.

3.2. Beran Method

Beran (1994a) proposed the following version of the AR approximation approach. Assume that the following Gaussian innovation sequence

$$\varepsilon_t = y_t - \sum_{j=1}^{\infty}\pi_j(\boldsymbol{\theta})y_{t-j}$$

Since the values $\{y_t,\ t \le 0\}$ are not observed, an approximate innovation sequence $\{u_t\}$ can be obtained by assuming that $y_t = 0$ for $t \le 0$,

$$u_t(\boldsymbol{\theta}) = y_t - \sum_{j=1}^{t-1}\pi_j(\boldsymbol{\theta})y_{t-j}$$

for $t = 2, \ldots, n$. Let $r_t(\boldsymbol{\theta}) = u_t(\boldsymbol{\theta})/\theta_1$ and $\boldsymbol{\theta} = (\sigma_\varepsilon,\ \phi_1, \ldots, \phi_p,\ \theta_1, \ldots, \theta_q,\ d)$. Then, a QMLE for $\boldsymbol{\theta}$ is provided by the minimization of

$$\mathcal{L}_2(\boldsymbol{\theta}) = 2n\ \log(\theta_1) + \sum_{t=2}^{n} r_t^2(\boldsymbol{\theta})$$

Now, by taking partial derivatives with respect to $\boldsymbol{\theta}$, the minimization problem is equivalent to solving the non-linear equations

$$\sum_{t=2}^{n}\{r_t(\boldsymbol{\theta})\dot{r}_t(\boldsymbol{\theta}) - E[r_t(\boldsymbol{\theta})\dot{r}_t(\boldsymbol{\theta})]\} = 0 \tag{20}$$

where $\dot{r}_t(\boldsymbol{\theta}) = \left(\frac{\partial r_t(\boldsymbol{\theta})}{\partial\theta_1}, \ldots, \frac{\partial r_t(\boldsymbol{\theta})}{\partial\theta_r}\right)'$.

The arithmetic complexity of this method is $\mathcal{O}(n^2)$, that is, comparable to the Levinson–Durbin algorithm. Unlike the Haslett and Raftery method, the Beran approach uses the same variance for all the errors u_t. Hence, its performance may be poor for short time series.

3.2.1. Asymptotic Behavior

The QMLE based on the AR approximations share the same asymptotic properties with the exact MLE. The following results are due to Beran (1994a). Let $\tilde{\boldsymbol{\theta}}_n$ be the value that solves (20). Then,

Theorem 3. (Consistency) $\tilde{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$, in probability, as $n \to \infty$.
(Central Limit Theorem) $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to N(0, \Gamma^{-1}(\boldsymbol{\theta}_0))$, as $n \to \infty$, with $\Gamma(\boldsymbol{\theta}_0)$ is given in (17).
(Efficiency) $\tilde{\boldsymbol{\theta}}_n$ is an efficient estimator of $\boldsymbol{\theta}_0$.

4. MOVING AVERAGE APPROXIMATIONS

A natural alternative to AR approximations is the truncation of the infinite MA expansion of a long-memory process. The main advantage of this approach is the easy implementation of the Kalman filter recursions and the simplicity of the analysis of the theoretical properties of the ML estimates. Furthermore, if differencing is applied to the series, then the resulting truncation has less error variance than the AR approximation.

A causal representation of an ARFIMA(p, d, q) process $\{y_t\}$ is given by

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \tag{21}$$

On the other hand, we may consider an approximate model for (21) given by

$$y_t = \sum_{j=0}^{m} \psi_j \varepsilon_{t-j} \tag{22}$$

which corresponds to a MA(m) process in contrast to the MA(∞) process (21). A canonical state space representation of the MA(m) model (22) is given by

$$X_{t+1} = FX_t + H\varepsilon_t$$
$$y_t = GX_t + \varepsilon_t$$

with

$$F = \begin{bmatrix} 0 & I_{m-1} \\ 0 \ldots & 0 \end{bmatrix}, \qquad G = [1\ 0\ 0\ \ldots\ 0], \qquad H = [\psi_1\ \ldots\ \psi_m]'$$

$$X_t = [y(t|t-1),\ y(t+1|t-1),\ \ldots,\ y(t+m-1|t-1)]'$$
$$y(t+j|t-1) = E[y_{t+j}|y_{t-1}, y_{t-2}, \ldots]$$

The approximate representation of a causal ARFIMA(p, d, q) has computational advantages over the exact one. In particular, the order of the MLE algorithm is reduced from $\mathcal{O}(n^3)$ to $\mathcal{O}(n)$. A brief discussion about the Kalman filter implementation of this state space system follows.

4.1. Kalman Recursions

Let the initial conditions be $\hat{X}_1 = E[X_1]$ and $\Omega_1 = E[X_1 X_1'] - E[\hat{X}_1 \hat{X}'_1]$. The recursive Kalman equations may be written as follows (cf. Chan, 2002, Section 11.3):

$$\Delta_t = G\Omega_t G' + \sigma_\varepsilon^2 \tag{23}$$

$$\Theta_t = F\Omega_t G' + S \tag{24}$$

$$\Omega_{t+1} = F\Omega_t F' + Q - \Theta_t \Delta_t^{-1} \Theta_t' \tag{25}$$

$$\begin{aligned} \hat{X}_{t+1} &= F\hat{X}_t + \Theta_t \Delta_t^{-1}(y_t - G\hat{X}_t) \\ \hat{y}_t &= G\hat{X}_t \end{aligned} \tag{26}$$

for $t = 1, 2, \ldots, n$, where $Q = \mathrm{var}(H\varepsilon_t)$, $\sigma_\varepsilon = \mathrm{var}(\varepsilon_t)$ and $S = \mathrm{cov}(H\varepsilon_t, \varepsilon_t)$. The log-likelihood function, excepting a constant, is given by

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2}\left\{ \sum_{t=1}^{n} \log\, \Delta_t(\theta) + \sum_{t=1}^{n} \frac{(y_t - \hat{y}_t(\theta))^2}{\Delta_t(\theta)} \right\}$$

where $\boldsymbol{\theta} = (\phi_1, \ldots \phi_p, \theta_1, \ldots, \theta_q, d, \sigma^2)$ is the parameter vector associated with the ARFIMA representation (1). In order to evaluate the log-likelihood function $\mathcal{L}(\boldsymbol{\theta})$, we may choose the initial conditions $\hat{X}_1 = E[X_1] = 0$ and $\Omega_1 = E[X_1\, X_1'] = [\omega(i, j)]_{i,j=1,2,\ldots,}$ where $\omega(i, j) = \sum_{k=0}^{\infty} \psi_{i+k}\psi_{j+k}$.

The evolution of the state estimation and its variance, Ω_t, is given by the following recursive equations. Let $\delta_i = 1$ if $i \in \{0, 1, \ldots, m-1\}$ and $\delta_i = 0$ otherwise. Furthermore, let $\delta_{ij} = \delta_i \delta_j$. Then, the elements of Ω_{t+1} and $\hat{X}_{t+1}$ in (25) and (26) are as follows:

$$\begin{aligned} \omega_{t+1}(i, j) = {} & \omega_t(i+1, j+1)\delta_{ij} + \psi_i \psi_j \\ & - \frac{[\omega_t(i+1, 1)\delta_i + \psi_i][\omega_t(j+1, 1)\delta_j + \psi_j]}{\omega_t(1, 1) + 1} \end{aligned}$$

and

$$\hat{X}_{t+1}(i) = \hat{X}_t(i+1)\delta_i + \frac{(\omega_t(i+1,1)\delta_i + \psi_i)(y_t - \hat{X}_t(1))}{\omega_t(1,1)+1}$$

In addition,

$$\hat{y}_t = G\hat{X}_t = \hat{X}_t(1)$$

In order to speed up the algorithm, we can difference the series $\{y_t\}$ so that the MA expansion of the differenced series converges faster. To this end, consider the MA expansion of the differenced process

$$z_t = (1-B)y_t = \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j} \tag{27}$$

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

$$z_t = \sum_{j=0}^{m} \varphi_j \varepsilon_{t-j} \tag{28}$$

The main advantage of this approach is that the coefficients φ_j converge to zero faster than the coefficients ψ_j do. Thus, a smaller truncation parameter is necessary to achieve a good level of approximation.

A state space representation of this truncated model is given by, see for example Section 12.1 of Brockwell and Davis (1991),

$$X_{t+1} = \begin{bmatrix} 0 & & I_{m-1} \\ 0 \cdots & & 0 \end{bmatrix} X_t + \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_m \end{bmatrix} \varepsilon_t$$

and

$$z_t = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \end{bmatrix} X_t + \varepsilon_t$$

The Gaussian log-likelihood of the truncated model (28) may be written as

$$\mathscr{L}_n(\boldsymbol{\theta}) = \frac{1}{2n} \log \det T_{n,m}(\boldsymbol{\theta}) - \frac{1}{2n} Z_n' T_{n,m}(\boldsymbol{\theta})^{-1} Z_n$$

where $[T_{n,m}(\boldsymbol{\theta})]_{r,s=1,\dots,n} = \int_{-\pi}^{\pi} \tilde{f}_{m,\boldsymbol{\theta}}(\omega) e^{i\omega(r-s)} d\omega$ is the covariance matrix of $Z_n = (z_1 \dots z_n)'$ with $\tilde{f}_{m,\boldsymbol{\theta}}(\omega) = \sigma_\varepsilon^2 |\varphi_m(e^{i\omega})|^2$ and $\varphi_m(e^{i\omega}) = 1 + \varphi_1 e^{i\omega} + \cdots + \varphi_{1_m} e^{mi\omega}$.

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 $\mathcal{O}(n)$ 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 $\mathcal{O}(n)$.

This approach is discussed in Chan and Palma (1998), where they show the following result.

Theorem 4. (Consistency) Assume that $m = n^\beta$ with $\beta > 0$, then as $n \to \infty$, $\tilde{\boldsymbol{\theta}}_{n,m} \to \boldsymbol{\theta}_0$, in probability.
(Central Limit Theorem) Suppose that $m = n^\beta$ with $\beta \geq 1/2$, then as $n \to \infty$, $\sqrt{n}(\tilde{\boldsymbol{\theta}}_{n,m} - \boldsymbol{\theta}_0) \to N(0, \Gamma^{-1}(\boldsymbol{\theta}_0))$,
where $\Gamma(\boldsymbol{\theta})$ is given in (17).
(Efficiency) Assume that $m = n^\beta$ with $\beta \geq 1/2$, then $\tilde{\boldsymbol{\theta}}_{n,m}$, is an efficient estimator of θ_0.

Observe that both AR(m) and MA(m) approximations produce algorithms with arithmetic complexity of order $\mathcal{O}(n)$. However, the quality of the approximation is governed by the truncation parameter m. Bondon and Palma (2005) prove that the variance of the truncation error for an AR(m) approximation is of order $\mathcal{O}(1/n)$. On the other hand, it can be easily shown that for the MA(m) case, this quantity is of order $\mathcal{O}(n^{2d-1})$. Furthermore, when the differenced approach is used, the truncation error variance is of order $\mathcal{O}(n^{2d-3})$.

5. WHITTLE APPROXIMATIONS

Another approach to obtain approximate ML estimates is based on the calculation of the periodogram by means of the Fast Fourier Transform (FFT) and the use of the Whittle approximation of the Gaussian log-likelihood function. This approach produces fast numerical algorithms for computing parameter estimates, since the calculation of the FFT has an arithmetic complexity $\mathcal{O}(n \log_2(n))$ (cf. Press et al., 1992, p. 498).

5.1. Whittle Approximation of the Gaussian Likelihood Function

Consider the Gaussian process $\boldsymbol{Y} = (y_1, \ldots, y_n)'$ with zero mean and variance $\Gamma_{\boldsymbol{\theta}}$. Then the log-likelihood function divided by the sample size is given by

$$\mathscr{L}(\boldsymbol{\theta}) = -\frac{1}{2n}\log\ \det\ \Gamma_\theta - \frac{1}{2n}\, Y'\Gamma_\theta^{-1}\, Y \tag{29}$$

Observe that the variance–covariance matrix $\Gamma_{\boldsymbol{\theta}}$ can be expressed in terms of the spectral density of the process $f_{\boldsymbol{\theta}}(\cdot)$ as follows:

$$(\Gamma_{\boldsymbol{\theta}})_{ij} = \gamma_{\boldsymbol{\theta}}(i-j)$$

where

$$\gamma_{\boldsymbol{\theta}}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\omega)\ \exp\,(i\omega k)\, d\omega$$

In order to obtain the method proposed by Whittle (1951), two approximations are made. Following the result by Grenander and Szegö (1958) that

$$\frac{1}{n}\log\ \det\ \Gamma_{\boldsymbol{\theta}} \to \frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_{\boldsymbol{\theta}}(\omega)\, d\omega$$

as $n \to \infty$, the first term in (29) is approximated by

$$\frac{1}{2n}\log\ \det\ \Gamma_\theta \approx \frac{1}{4\pi}\int_{-\pi}^{\pi} \log f_{\boldsymbol{\theta}}(\omega)\, d\omega$$

On the other hand, the second term in (29) is approximated by

$$\begin{aligned}
\frac{1}{2\pi}\, Y'\Gamma_\theta^{-1}\, Y &\approx \sum_{l=1}^{n}\sum_{j=1}^{n} y_l \left\{\frac{1}{4\pi n}\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}^{-1}(\omega)\ \exp\,[i\omega(l-j)]d\omega\right\} y_j \\
&= \frac{1}{4\pi n}\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}^{-1}(\omega) \sum_{l=1}^{n}\sum_{j=1}^{n} y_l y_j\ \exp\,[i\omega(l-j)]d\omega \\
&= \frac{1}{4\pi n}\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}^{-1}(\omega) \Big|\sum_{j=1}^{n} y_j\ \exp\,(i\omega j)\Big|^2 d\omega \\
&= \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{I(\omega)}{f_{\boldsymbol{\theta}}(\omega)} d\omega
\end{aligned}$$

where $I(\omega) = \frac{1}{n}\left|\sum_{j=1}^{n}(y_j - \bar{y})\ \exp\,(i\omega j)\right|^2$ is the periodogram of the series $\{y_t\}$. Thus, the log-likelihood function is approximated by

$$\mathscr{L}_3(\boldsymbol{\theta}) = -\frac{1}{4\pi}\left\{\int_{-\pi}^{\pi} \log f_{\boldsymbol{\theta}}(\omega) d\omega + \int_{-\pi}^{\pi} \frac{I(\omega)}{f_{\boldsymbol{\theta}}(\omega)} d\omega\right\} \tag{30}$$

5.2. Discrete Version

The evaluation of the log-likelihood function (30) requires the calculation of integrals. To simplify this computation, the integrals can be substituted by Riemann sums as follows:

$$\int_{-\pi}^{\pi} \log f_{\boldsymbol{\theta}}(\omega) d\omega \approx \frac{2\pi}{n} \sum_{j=1}^{n} \log f_{\boldsymbol{\theta}}(\omega_j)$$

and

$$\int_{-\pi}^{\pi} \frac{I(\omega)}{f_{\boldsymbol{\theta}}(\omega)} d\omega\} \approx \frac{2\pi}{n} \sum_{j=1}^{n} \frac{I(\omega_j)}{f_{\boldsymbol{\theta}}(\omega_j)}$$

where $\omega_j = \frac{2\pi j}{n}$ are the Fourier frequencies. Thus, a discrete version of the log-likelihood function (30) is

$$\mathscr{L}_4(\boldsymbol{\theta}) = -\frac{1}{2n} \left\{ \sum_{j=1}^{n} \log f_{\boldsymbol{\theta}}(\omega_j) + \sum_{j=1}^{n} \frac{I(\omega_j)}{f_{\boldsymbol{\theta}}(\omega_j)} \right\}$$

5.3. Alternative Versions

Further simplifications of the Whittle log-likelihood function can be made. For example, by assuming that the spectral density is normalized as

$$\int_{-\pi}^{\pi} \log f_{\boldsymbol{\theta}}(\omega) d\omega = 0 \tag{31}$$

then the Whittle log-likelihood function is reduced to

$$\mathscr{L}_5(\boldsymbol{\theta}) = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{I(\omega)}{f_{\boldsymbol{\theta}}(\omega)} d\omega$$

with the discrete version

$$\mathscr{L}_6(\boldsymbol{\theta}) = -\frac{1}{2n} \sum_{j=1}^{n} \frac{I(\omega_j)}{f_{\boldsymbol{\theta}}(\omega_j)}$$

Observe that by virtue of the well-known Szegö–Kolmogorov formula

$$\sigma_\varepsilon^2 = 2\pi \exp\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_{\boldsymbol{\theta}}(\omega) d\omega \right]$$

the normalization (31) is equivalent to setting $\sigma_\varepsilon^2 = 2\pi$.

5.4. Asymptotic Results

The asymptotic behavior of the Whittle estimator is similar to that of the exact MLE. The following theorem combines results by Fox and Taqqu (1986) and Dahlhaus (1989).

Theorem 5. Let $\hat{\boldsymbol{\theta}}_n^{(i)}$ be the value that maximizes the log-likelihood function $\mathscr{L}_i(\boldsymbol{\theta})$ for $i = 3, \ldots, 6$ for a Gaussian process $\{y_t\}$. Then, under some regularity conditions, $\hat{\boldsymbol{\theta}}_n^{(i)}$ is consistent and $\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{(i)} - \boldsymbol{\theta}_0) \to N(0, \Gamma_{\boldsymbol{\theta}_0}^{-1})$, as n $\to \infty$.

5.5. Non-Gaussian Processes

All of the above methods apply to Gaussian processes. When this assumption in dropped, it is still possible to find well-behaved Whittle estimates. In particular, Giraitis and Surgailis (1990) have studied the estimates based on the maximization of the log-likelihood function $\mathscr{L}_5(\boldsymbol{\theta})$ for a general class of linear processes with independent innovations.

Consider the process $\{y_t\}$ generated by the Wold decomposition

$$y_t = \sum_{j=0}^{\infty} \psi_j(\boldsymbol{\theta}) \varepsilon_{t-j}$$

where ε_t is an i.i.d. sequence with finite four cumulant and $\sum_{j=0}^{\infty} \psi_j^2(\boldsymbol{\theta}) < \infty$. The following result, due to Giraitis and Surgailis (1990), establishes the consistency and the asymptotic normality of the Whittle estimate under these circumstances.

Theorem 6. Let $\hat{\boldsymbol{\theta}}_n$ be the value that maximizes the log-likelihood function $\mathscr{L}_5(\boldsymbol{\theta})$. Then, under some regularity conditions, $\hat{\boldsymbol{\theta}}_n$ is consistent and $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \to N(0, \Gamma_{\boldsymbol{\theta}_0}^{-1})$, as $n \to \infty$.
Note that this theorem does not assume the normality of the process.

5.6. Semi-Parametric Methods

Another generalization of the Whittle estimator is the Gaussian semi-parametric estimation method proposed by Robinson (1995). This estimation approach does not require the specification of a parametric model for the data. It only relies on the specification of the shape of the spectral density of the time series.

Consider the stationary process $\{y_t\}$ with spectral density satisfying

$$f(\omega) \sim G\omega^{1-2H}$$

as $\omega \to 0+$, with $G \in (0, \infty)$ and $H \in (0, 1)$. Note that for an ARFIMA model, the terms G and H correspond to $\sigma^2/2\pi[\theta(1)/\phi(1)]^2$ and $1/2+\mathrm{d}$, respectively. H is usually called the *Hurst parameter.*

Let the objective function $Q(G, H)$ to be minimized be given by

$$Q(G, H) = \frac{1}{m}\sum_{j=1}^{m}\left\{\log\ G\omega_j^{1-2H} + \frac{\omega_j^{2H-1}}{G} I(\omega_j)\right\}$$

where m is an integer satisfying $m < n/2$. Let $(\hat{G}, \hat{H})$ be the value that minimizes $Q(G, H)$. Then, under some regularity conditions of the spectral density and

$$\frac{1}{m} + \frac{m}{n} \to 0$$

as $n \to \infty$, the following result due to Robinson (1995) holds:

Theorem 7. Let H_0 be the true value of the Hurst parameter. The estimator $\hat{H}$ is consistent and $\sqrt{m}(\hat{H} - \hat{H}_0) \to N\left(0, \frac{1}{4}\right)$, as $n \to \infty$.

5.7. Numerical experiments

Table 1 displays the results from several simulations comparing five ML estimation methods for Gaussian processes: Exact MLE, Haslett and Raftery's approach, AR(40) approximation, MA(40) approximation and the Whittle method. The process considered is a fractional noise ARFIMA(0, d, 0) with three values of the long-memory parameter:

Table 1. Finite Sample Behavior of ML Estimates of ARFIMA(0, d, 0) Models. Sample Size $n = 250$ and Truncation $m = 40$ for AR(m) and MA(m) Approximations.

d		Exact	HR	AR	MA	Whittle
0.40	mean	0.371210	0.372320	0.376730	0.371310	0.391760
	stdev	0.047959	0.048421	0.057392	0.050048	0.057801
0.25	mean	0.229700	0.230400	0.229540	0.229810	0.221760
	stdev	0.051899	0.051959	0.056487	0.051475	0.060388
0.10	mean	0.082900	0.083240	0.083910	0.084410	0.065234
	stdev	0.049260	0.049440	0.052010	0.049020	0.051646

$d = 0.1, 0.25, 0.4$, Gaussian innovations with mean zero and variance 1 and sample size $n = 250$. The mean and standard deviations of the estimates are based on 1,000 repetitions. All the simulations were carried out by means of Splus programs, available upon request.

From this table, it seems that all estimates are somewhat downward biased for the three values of d considered. All the estimators, excepting the Whittle, seem to behave similarly in terms of both bias and standard error. On the other hand, the Whittle method has less bias for $d = 0.4$, but greater bias for $d = 0.1$. Besides, this procedure seems to have greater standard error than the other estimators, for the three values of d. The empirical parameter estimate standard deviations of all the method considered are close to its theoretical value 0.04931.

In the next section we discuss the application of the ML estimation methodology to time series with missing values.

6. ESTIMATION OF INCOMPLETE SERIES

The Kalman filter recursive Eqs. (23)–(26) can be modified to calculate the log-likelihood function for incomplete series, as described in Palma and Chan (1997). In this case, we have

$$\Delta_t = G\Omega_t G' + \sigma_w^2$$
$$\Theta_t = F\Omega_t G' + S$$

$$\Omega_{t+1} = \begin{cases} F\Omega_t F' + Q - \Theta_t \Delta_t^{-1} \Theta_t' & y_t \text{ known} \\ F\Omega_t F' + Q & y_t \text{ missing} \end{cases}$$

$$\hat{X}_{t+1} = \begin{cases} F\hat{X}_t + \Theta_t \Delta_t^{-1}(y_t - G\hat{X}_t) & y_t \text{ known} \\ F\hat{X}_t & y_t \text{ missing} \end{cases}$$

Let K_n be the set indexing the observed values of the process $\{y_t\}$. The Kalman recursive log-likelihood function is given by

$$\mathscr{L}(\boldsymbol{\theta}) = -\frac{1}{2}\left\{ r \log 2\pi + \sum_{t \in K_n} \log \Delta_t + r \log \sigma_\varepsilon^2 \frac{1}{\sigma_\varepsilon^2} \sum_{t \in K_n} \frac{(y_t - \hat{y}_t)^2}{\Delta_t} \right\}$$

where r is the number of observed values and $\Delta_t = \omega_{11}^{(t)} + 1$ is the variance of the best predictor $\hat{y}_t$. This form of the likelihood function may be used to efficiently calculate Gaussian ML estimates. One-step predictions and their corresponding standard deviations are obtained directly from the recursive

Kalman filter equations without further computation, see Palma and Chan (1997) for details.

6.1. Effect of Data Irregularities and Missing Values on ML Estimates

To illustrate the potential dramatic effects of data irregularities such as repeated or missing values on the parameter estimation of a long-memory process, consider the well-known Nile river data, cf. Beran (1994b), shown in Fig. 1. Panel (a) displays the original data. From this plot, it seems that in several periods, the data were repeated year after year in order to complete the series. Panel (b) shows the same series, but without those repeated values (filtered series). This procedure has been discussed in Palma and Del Pino (1999).

Table 2 shows the fitted parameters of an ARFIMA(0, d, 0) using an AR(40) approximation along the Kalman filter, for both the original and the data without repetitions.

Observe that for the original data, the estimate of the long-memory parameter is 0.5, indicating that the model has reached the non-stationary boundary. On the other hand, for the filtered data, the estimate of d belongs to the stationary region. Thus, in this particular case, the presence of data

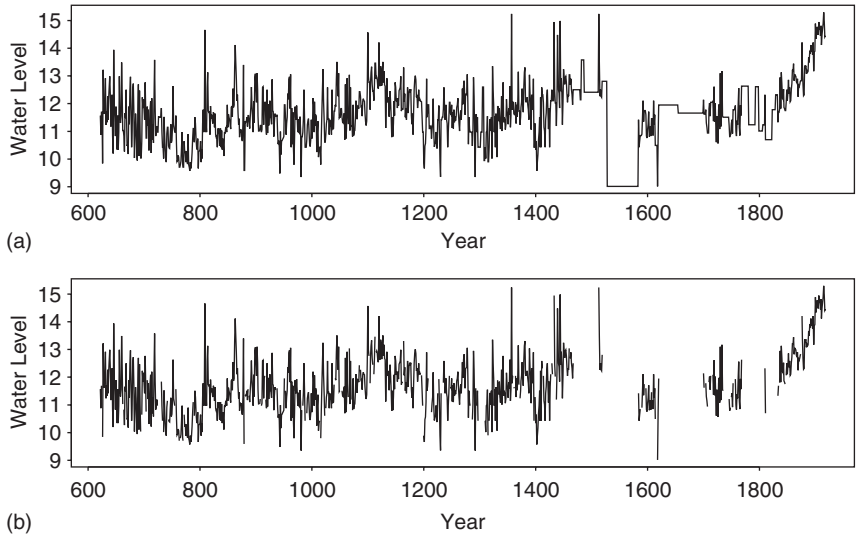

Fig. 1. Nile River Data (622 A.D. – 1921 A.D.). Panel (a) Original Data and panel (b) Filtered Data.

Table 2. Nile River Data: ML Estimates for the Fitted ARFIMA(0, *d,* 0) Model.

Series	$\hat{d}$	$t_{\hat{d}}$	$\hat{\sigma}_\varepsilon$
Panel (a)	0.5000	25.7518	0.6506
Panel (b)	0.4337	19.4150	0.7179

Table 3. Finite Sample Behavior of ML Estimates of ARFIMA(0, *d,* 0) Processes with Missing Values. Sample Size $n = 250$ and Truncation $m = 40$ for AR(m) and MA(m) Approximations.

d	NA's		AR	MA	Expected stdv
	0%	mean	0.380470	0.374390	
		stdev	0.056912	0.048912	0.049312
0.40	15%	mean	0.376730	0.369120	
		stdev	0.060268	0.055014	0.053550
	30%	mean	0.366650	0.363860	
		stdev	0.065166	0.059674	0.058935
	0%	mean	0.225600	0.226900	
		stdev	0.056810	0.052110	0.049312
0.25	15%	mean	0.223500	0.224800	
		stdev	0.065440	0.060040	0.053550
	30%	mean	0.219800	0.220800	
		stdev	0.072920	0.062510	0.058935
	0%	mean	0.081145	0.081806	
		stdev	0.054175	0.051013	0.049312
0.10	15%	mean	0.086788	0.081951	
		stdev	0.058106	0.052862	0.053550
	30%	mean	0.080430	0.081156	
		stdev	0.067788	0.061118	0.058935

irregularities such as the replacement of missing data with repeated values induces non-stationarity. On the other hand, when the missing value is appropriately taken care of, the resulting model is stationary (cf. Palma & Del Pino, 1999).

Table 3 displays the results from Monte Carlo simulations of approximate ML estimates for fractional noise ARFIMA(0, *d,* 0) with missing values at

random. The sample size chosen was $n = 250$ and the AR and MA truncations are $m = 40$ for both cases. The long-memory parameters are $d = 0.1, 0.25, 0.40$ and $\sigma_\varepsilon^2 = 1$. The number of missing values are 38 (15% of the sample) and 75 (30% of the sample) and were selected randomly for each sample.

Note that the bias and the standard deviation of the estimates seem to increase as the number of missing values increases. On the other hand, the sample standard deviation of the estimates seems to be close to the expected values, for the MA approximation. For the AR approximation, these values are greater than expected. The expected standard deviation used here is $\sqrt{6/\pi^2 n^*}$, where n^* is the number of observed values.

7. ESTIMATION OF SEASONAL LONG-MEMORY MODELS

In practical applications, many researchers have found time series exhibiting both long-range dependence and cyclical behavior. For instance, this phenomenon occurs for the inflation rates studied by Hassler and Wolters (1995), revenue series analyzed by Ray (1993), monetary aggregates considered by Porter-Hudak (1990), quarterly gross national product and shipping data discussed by Ooms (1995) and monthly flows of the Nile River studied by Montanari, Rosso, and Taqqu (2000).

Several statistical methodologies have been proposed to model this type of data. For instance, Gray, Zhang, and Woodward (1989) propose the generalized fractional or Gegenbauer (GARMA) processes, Porter-Hudak (1990) discusses seasonal fractionally integrated autoregressive moving average (SARFIMA) models, Hassler (1994) introduces the flexible seasonal fractionally integrated processes (flexible ARFISMA) and Woodward, Cheng, and Gray (1998) introduce the k-GARMA processes. Furthermore, the statistical properties of these models have been investigated by Giraitis and Leipus (1995), Chung (1996), Arteche and Robinson (2000), Velasco and Robinson (2000) and Giraitis, Hidalgo, and Robinson (2001), among others.

A rather general class of Gaussian seasonal long-memory processes is specified by the spectral density

$$f(\omega) = g(\omega)|\omega|^{-\alpha} \prod_{i=1}^{r} \prod_{j=1}^{m_i} |\omega - \omega_{ij}|^{-\alpha_i} \tag{32}$$

where $\omega \in (-\pi, \pi], 0 \leq \alpha, \alpha_i < 1, i = 1, \ldots, r$, $g(\omega)$ is a symmetric, strictly positive, continuous and bounded function and $\omega_{ij} \neq 0$ are known poles for $j = 1, \ldots, m_i$, $i = 1, \ldots, r$. To ensure the symmetry of f, it is assumed that for any $i = 1, \ldots, r$, $j = 1, \ldots m_i$, there is one and only one $1 \leq j' \leq m_i$ such that $\omega_{ij} = \omega_{ij'}$. The spectral density of many widely used models such as SARFIMA and k-factor GARMA satisfy specification (32).

The exact ML estimation of processes satisfying (32) has been recently studied by Palma and Chan (2005) who have established the following result:

Theorem 8. Let $\hat{\boldsymbol{\theta}}_n$ be the exact MLE for a process satisfying (32) and $\boldsymbol{\theta}_0$ the true parameter. Then, under some regularity conditions we have (Consistency) $\hat{\boldsymbol{\theta}}_{n \to p} \boldsymbol{\theta}_0$ *as* $n \to \infty$. (Central limit theorem) The ML estimate, $\hat{\boldsymbol{\theta}}_n$, satisfies the following limiting distribution as $n \to \infty$: $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to N(0, \Gamma(\boldsymbol{\theta}_0)^{-1})$, where $\Gamma(\boldsymbol{\theta}_0)$ is given by (17).
(Efficiency) The ML estimate, $\hat{\boldsymbol{\theta}}_n$, is asymptotically an efficient estimate of $\boldsymbol{\theta}_0$.

7.1. Monte Carlo Studies

In order to assess the finite sample performance of the ML estimates in the context of long-memory seasonal series, a number of Monte Carlo simulations were conducted for the class of SARFIMA$(p, d, q) \times (P, d_s, Q)_s$ models described by the following difference equation (cf. Porter-Hudak, 1990):

$$\phi(B)\Phi(B^s)(1 - B)^d y_t = \theta(B)\Theta(B^s)\varepsilon_t$$

where $\{\varepsilon_t\}$ are standard normal random variables, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, $\Phi(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{sP}$, $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$, $\Theta(B^s) = 1 + \Theta_1 B^s + \cdots + \Theta_Q B^{sQ}$, polynomials $\phi(B)$ and $\theta(B)$ have no common zeros, $\Phi(B^s)$ and $\Theta(B^s)$ have no common zeros, and the roots of these polynomials are outside the unit circle.

Table 4 shows the results from simulations for SARFIMA$(0, d, 0) \times (0, d_s, 0)_s$ models. The estimates of $\hat{d}$ and $\hat{d}_s$ reported in columns seven and eight of Table 4 and the estimated standard deviations displayed in the last two columns of the table are based on 1,000 repetitions. The MLE are computed by means of an extension to SARFIMA models of the state space representations of long-memory processes, see Chan and Palma (1998) for details. The theoretical values of the standard deviations of the estimated parameters are based on the formula (17). In general, analytic expressions for the

Table 4. Finite Sample Performance of ML Estimates of SARFIMA(0, d, 0) $\times$ (0, d_s, 0)$_s$ Models for Several Values of s, n, d and d_s.

Period	n	d	d_s	$\sigma(d)$	$\sigma(d_s)$	$\hat{d}$	$\hat{d}_s$	$\hat{\sigma}(\hat{d})$	$\hat{\sigma}(\hat{d}_s)$
4	1000	0.100	0.300	0.025	0.025	0.090	0.299	0.027	0.028
6	500	—	0.300	—	0.035	—	0.286	—	0.036
6	1000	0.150	0.200	0.025	0.025	0.140	0.198	0.026	0.026
6	2000	0.200	0.200	0.018	0.018	0.195	0.196	0.018	0.018
12	3000	—	0.250	—	0.014	—	0.252	—	0.016
12	1000	0.200	0.100	0.025	0.025	0.194	0.094	0.025	0.027

integral in (17) are difficult to obtain for an arbitrary period s. For an SARFIMA(0, d, 0) $\times$ (0, d_s, 0)$_s$ model, the matrix $\Gamma(\boldsymbol{\theta})$ can be written as

$$\Gamma(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\pi^2}{6} & c \\ c & \frac{\pi^2}{6} \end{pmatrix}$$

with $c = \frac{1}{\pi}\int_{-\pi}^{\pi}\{\log|2\sin(\frac{\omega}{2})|\}\{\log|2\sin(s\frac{\omega}{2})|\}d\omega$. An interesting feature of the asymptotic variance of the parameters is that for an SARFIMA(0, d, 0) $\times$ (0, d_s, 0)$_s$ process, the variance of $\hat{d}$ is the same as the variance of $\hat{d}_s$.

From Table 4, note that the estimates and their standard deviations are close to the theoretical values, for all the sample sizes and combinations of parameters investigated.

8. HETEROSKEDASTIC TIME SERIES

Evidence of long-memory behavior in returns and/or empirical volatilities has been observed by several authors, see for example Robinson (1991) and references therein. Accordingly, several models have been proposed in the econometric literature to explain the combined presence of long-range dependence and conditional heteroskedasticity. In particular, a class of models that has received considerable attention is the ARFIMA–GARCH (generalized autoregressive conditional heteroskedastic) process, see for example Ling and Li (1997). In this model, the returns have long memory and the noise has a conditional heteroskedasticity structure. A related class of interesting models is the extension of the ARCH(p) processes first introduced by Engle (1982) to the ARCH(∞) models to encompass the longer dependence observed in many squared financial series. On the other hand, extensions of the stochastic volatility processes to the long-memory case

have produced the so-called long-memory stochastic volatility models (LMSV). In this section, we discuss briefly some of these well-known econometric models.

8.1. ARFIMA–GARCH Model

An ARFIMA(p,d,q)–GARCH(r,s) process is defined by the discrete-time equation

$$\begin{aligned}\Phi(B)(1-B)^d(y_t-\mu) &= \Theta(B)\varepsilon_t \\ \varepsilon|\mathscr{F}_{t-1} &\sim N(0,\ h_t) \\ h_t &= \alpha_0 + \sum_{i=1}^{r}\alpha_i\varepsilon_t^2 + \sum_{j=1}^{s}\beta_j h_{t-j}\end{aligned}$$

where $\mathscr{F}_{t-1}$ is the σ-algebra generated by the past observations $y_{t-1}, y_{t-2}, \ldots$.

Most econometric models dealing with long memory and heteroskedastic behaviors are non-linear, in the sense that the noise sequence is not necessarily independent. An approximate MLE $\hat{\boldsymbol{\theta}}$ is obtained by maximizing the conditional log-likelihood

$$\mathscr{L}(\boldsymbol{\theta}) = -\frac{1}{2n}\sum_{t=1}^{n}\left\{\log\ h_t + \frac{\varepsilon_t^2}{h_t}\right\} \tag{33}$$

The asymptotic behavior of this estimate was formally established by Ling and Li (1997). Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\ \boldsymbol{\theta}_2)'$, where $\boldsymbol{\theta}_1 = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, d)'$ is the parameter vector involving the ARFIMA components and $\boldsymbol{\theta}_2 = (\alpha_0, \ldots, \alpha_r, \beta_1, \ldots, \beta_s)'$ is the parameter vector containing the GARCH component. The following result correspond to Theorem 3.2 of Ling and Li (1997).

Theorem 9. Let $\hat{\boldsymbol{\theta}}_n$ be the value that maximizes the conditional log-likelihood function (33). Then, under some regularity conditions, $\hat{\boldsymbol{\theta}}_n$ is a consistent estimate and $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to N(0,\ \Omega^{-1})$, as $n \to \infty$, where $\Omega = \mathrm{diag}(\Omega_1, \Omega_2)$ with

$$\Omega_1 = E\left[\frac{1}{h_t}\frac{\partial\varepsilon_t}{\partial\boldsymbol{\theta}_1}\frac{\partial\varepsilon_t}{\partial\boldsymbol{\theta}_1'} + \frac{1}{2h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}_1}\frac{\partial h_t}{\partial\boldsymbol{\theta}_1'}\right]$$

and

$$\Omega_2 = E\left[\frac{1}{2h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\theta}_2}\frac{\partial h_t}{\partial \boldsymbol{\theta}_2'}\right]$$

8.2. Arch-Type Models

The ARFIMA–GARCH process described in the previous subsection is adequate for modeling long-range dependence in returns of financial time series. However, as described by Rosenblatt (1961) and Palma and Zevallos (2004), the squares of an ARFIMA–GARCH process have only *intermediate memory* for $d \in (0, 1/4)$. In fact, for any $d \in (0, 1/2)$, the ACF of the squared series behaves like $k^{2\tilde{d}-1}$, where k denotes the k-*th* lag and $\tilde{d} = 2d - 1/2$. Consequently, the long-memory parameter of the squared series $\tilde{d}$ is always smaller than the long-memory parameter of the original series d, i.e. $\tilde{d} < d$ for $d < 1/2$.

Since in many financial applications the squared returns have the same or greater level of autocorrelation, the theoretical reduction in the memory that affects the squares of an ARFIMA–GARCH process may not be appropriate in practice. This situation leads one to consider other classes of processes to model the dependence of the squared returns directly. For instance, Robinson (1991) proposed the following extension of the ARCH(p) introduced by Engle (1982),

$$y_t = \sigma_t \xi_t$$
$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j y_{t-j}^2$$

which can be formally written as

$$y_t^2 = \alpha_0 + v_t + \sum_{j=1}^{\infty} \alpha_j y_{t-j}^2 \tag{34}$$

where $\sigma_t^2 = E[y_t^2 | y_{t-1}, y_{t-2}, \ldots]$, $\quad v_t = y_t^2 - \sigma_t^2$ is a martingale difference sequence, $\{\xi_t\}$ a sequence of independent and identically distributed random variables and α_0 a positive constant, cf. Eqs. (1.31), (1.33) and (1.35) of Robinson (2003), respectively.

When the coefficients $\{\alpha_j\}$ of (34) are specified by an ARFIMA(p,d,q) model, the resulting process corresponds to the (fractionally integrated

GARCH) FIGARCH(p,d,q) model which is defined by

$$\Phi(B)(1-B)^d y_t^2 = \omega + \Theta(B)v_t$$

where ω is a positive constant, cf. Baillie, Bollerslev, & Mikkelsen, (1996). As noted by Karanasos, Psaradakis, & Sola (2004), this process is strictly stationary and ergodic but not square integrable.

8.2.1. Estimation

Consider the quasi log-likelihood function

$$\mathscr{L}(\boldsymbol{\theta}) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left\{\log\ \sigma_t^2 + \frac{\varepsilon_t^2}{\sigma_t^2}\right\} \tag{35}$$

where $\boldsymbol{\theta} = (\omega, d, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_1)$. A QMLE $\hat{\boldsymbol{\theta}}_n$ can be obtained by maximizing (35). But, even though this estimation approach has been widely used in many practical applications, to the best of our knowledge, asymptotic results for these estimators remain an open issue. For a recent study about this problem, see for example Caporin (2002).

8.3. Stochastic Volatility

Stochastic volatility models have been addressed by Harvey, Ruiz, and Shephard (1994), Ghysels et al. (1996) and Breidt et al. (1998), among others. These processes are defined by

$$r_t = \sigma_t \xi_t$$

and

$$\sigma_{\mathrm{t}} = \sigma\ \exp(v_t/2) \tag{36}$$

where $\{\xi_t\}$ is a independent, identically distributed sequence with mean zero and variance one and $\{v_t\}$ is a stationary process independent of $\{\xi_t\}$. In particular, $\{v_t\}$ can be specified as a long-memory ARFIMA(p,d,q) process. The resulting process is called LMSV model.

From (36), we can write

$$\log(r_t^2) = \log(\sigma_t^2) + \log(\xi_t^2)$$
$$\log(\sigma_t^2) = \log(\sigma^2) + v_t$$

Let $y_t = \log(r_t^2)$, $\mu = \log(\sigma^2) + E[\log(\xi_t^2)]$ and $\varepsilon_t = \log(\xi_t^2) - E[\log(\xi_t^2)]$. Then

$$y_t = \mu + v_t + \varepsilon_t \tag{37}$$

Consequently, the transformed process $\{y_t\}$ corresponds to a stationary long-memory process plus an independent noise.

The ACF of (37) is given by

$$\gamma_y(\kappa) = \gamma_v(\kappa) + \sigma_\varepsilon^2 \delta_0(\kappa)$$

where $\delta_0(k) = 1$ for $k = 0$ and $\delta_0(k) = 0$ otherwise. Furthermore, the spectral density of $\{y_t\}$, f_y, is given by

$$f_y(\omega) = f_v(\omega) + \frac{\sigma_\varepsilon^2}{2\pi}$$

where f_v is the spectral density of the long-memory process $\{v_t\}$.

In particular, if the process $\{v_t\}$ is an ARFIMA(p,d,q) model

$$\Phi(B)(1 - B)^d v_t = \Theta(B)\eta_t \tag{38}$$

and $\boldsymbol{\theta} = (d, \sigma_\eta^2, \sigma_\varepsilon^2, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)'$ is the parameter vector that specifies model (38), then the spectral density is given by

$$f_{\boldsymbol{\theta}}(\omega) = \frac{\sigma_\eta^2}{2\pi} \frac{\left|\Theta(\exp(i\omega)\right|^2}{\left|1 - \exp(i\omega)\right|^{2d} \left|\Phi(\exp(i\omega)\right|^2} + \frac{\sigma_\varepsilon^2}{2\pi}$$

Breidt et al. (1998) consider the estimation of the parameter $\boldsymbol{\theta}$ by means of the spectral-likelihood estimator obtained by minimizing

$$\mathscr{L}_7(\boldsymbol{\theta}) = \frac{2\pi}{n} \sum_{j=1}^{n/2} \left\{ \log f_{\boldsymbol{\theta}}(\omega_j) + \frac{I(\omega_j)}{f_{\boldsymbol{\theta}}(\omega_j)} \right\}$$

where $f_{\boldsymbol{\theta}}(\omega)$ is given by (39).

Let $\hat{\boldsymbol{\theta}}$ be the value that minimizes $\mathscr{L}_7(\boldsymbol{\theta})$ over the parameter space Θ. Breidt et al. (1998) prove the following result.

Theorem 10. Assume that the parameter vector $\boldsymbol{\theta}$ is an element of the compact parameter space Θ and assume that $f_{\boldsymbol{\theta}_1} = f_{\boldsymbol{\theta}_2}$ implies that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. Let $\boldsymbol{\theta}_0$ be the true parameter value. Then $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ in probability as $n \to \infty$.

Other estimation procedures for LMSV using state space systems can be found in Chan and Petris (2000) and Section 11 of Chan (2002).

Table 5. Quasi ML Estimation of Long-Memory Stochastic Volatility Models with an ARFIMA(0, d, 0) Specification for v_t, for Different Values of d.

d	$\hat{d}$	$\hat{\sigma}_\eta$	S.D.($\hat{d}$)	S.D.($\hat{\sigma}_\eta$)
0.10	0.0868	9.9344	0.0405	0.4021
0.25	0.2539	10.0593	0.0400	0.4199
0.40	0.4139	10.1198	0.0415	0.3773

8.4. Numerical Experiments

The finite sample performance of the spectral-likelihood estimator is analyzed here by means of Monte Carlo simulations. The model investigated is the LMSV with an ARFIMA(0,d,0) structure, $\sigma_\varepsilon = \pi/\sqrt{2}, \xi_t$ follows a standard normal distribution, $\sigma_\eta = 10$ and the sample size is $n = 400$. The results displayed in Table 5 are based on 1,000 replications.

From Table 5, observe that estimates of both the long-memory parameter d and the scale parameter σ_η are close to their true values. On the other hand, the standard deviations of $\hat{d}$ and $\hat{\sigma}_\eta$ seem to be similar for all the values of d simulated. However, to the best of our knowledge there are no formally established results for the asymptotic distribution of these QMLE yet.

9. SUMMARY

In this article, a number of estimation techniques for long-memory time series have been reviewed together with their corresponding asymptotic results. Finite sample behaviors of these techniques were studied through Monte Carlo simulations. It is found that they are relatively comparable in terms of finite sample performance. However, in situations like missing data or long-memory seasonal time series, some approaches such as the MLE or truncated MLE seems to be more efficient than their spectral domain counterparts such as the Whittle approach.

Clearly, long-memory time series is an exciting and important topic in econometrics as well as many other disciplines. This article does not attempt to cover all of the important aspects of this exciting field. Interested readers may find many actively pursued topics in this area in the recent monograph of Robinson (2003). It is hoped that this article offers a focused and practical introduction to the estimation of long-memory time series.

ACKNOWLEDGMENT

We would like to thank an anonymous referee for helpful comments and Mr. Ricardo Olea for carrying out some of the simulations reported in this paper. This research is suppored in part by HKSAR-RGC Grants CUHK 4043/02P, 400305 and 2060268 and Fondecyt Grant 1040934. Part of this research was conducted when the second author was visiting the Institute of Mathematical Sciences at the Chinese University of Hong Kong. Support from the IMS is gratefully acknowledged.

REFERENCES

Ammar, G. S. (1996). Classical foundations of algorithms for solving positive definite Toeplitz equations. *Calcolo*, *33*, 99–113.

Arteche, J., & Robinson, P. M. (2000). Semiparametric inference in seasonal and cyclical long memory processes. *Journal of Time Series Analysis*, *21*, 1–25.

Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*, 3–30.

Beran, J. (1994a). On a class of *M*-estimators for Gaussian long-memory models. *Biometrika*, *81*, 755–766.

Beran, J. (1994b). *Statistics for long-memory processes*. New York: Chapman and Hall.

Bertelli, S., & Caporin, M. (2002). A note on calculating autocovariances of long-memory processes. *Journal of Time Series Analysis*, *23*, 503–508.

Bhansali, R. J., & Kokoszka, P. S. (2003). Prediction of long-memory time series. In: P. Doukhan, G. Oppenheim & M. Taqqu (Eds), *Theory and applications of long-range dependence* (pp. 355–367). Boston: Birkhäuser.

Bondon, P., & Palma, W. (2005). Prediction of strongly dependent time series. Preprint, Pontificia Universidad Catolica de Chile, Santiago, Chile.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis, forecasting and control*. Englewood Cliffs: Prentice-Hall.

Breidt, F. J., Crato, N., & de Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, *83*, 325–348.

Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods*. New York: Springer-Verlag.

Caporin, M. (2002). Long memory conditional heteroskedasticity and second order causality. Ph.D. Dissertation, Universita de Venezia, Venezia.

Chan, N. H. (2002). *Time series, applications to finance*. Wiley Series in Probability and Statistics. New York: Wiley

Chan, N. H., & Palma, W. (1998). State space modeling of long-memory processes. *The Annals of Statistics*, *26*, 719–740.

Chan, N. H. & Petris, G. (2000). Recent developments in heteroskedastic financial series. In: W. S. Chan, W. K. Li & H. Tong (Eds), *Statistics and finance* (pp. 169–184). (Hong Kong, 1999). London: Imperial College Press.

Chung, C. F. (1996). A generalized fractionally integrated autoregressive moving-average process. *Journal of Time Series Analysis*, *17*, 111–140.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, *17*, 1749–1766.

Durbin, J. (1960). The fitting of time series models. *Review of The International Statistical Institute*, *28*, 233–244.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

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

Ghysels, E., Harvey, A. C., & Renault, E. (1996). Stochastic volatility. In: G. S. Maddala & C. R. Rao (Eds), *Statistical methods in finance* (Vol. 14 of Handbook of Statistics. pp. 119–191). Amsterdam: North-Holland.

Giraitis, L., Hidalgo, J., & Robinson, P. M. (2001). Gaussian estimation of parametric spectral density with unknown pole. *The Annals of Statistics*, *29*, 987–1023.

Giraitis, L., & Leipus, R. (1995). A generalized fractionally differencing approach in long-memory modeling. *Lietuvos Matematikos Rinkinys*, *35*, 65–81.

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

Gradshteyn, I. S., & Ryzhik, I. M. (2000). *Table of integrals, series, and products*. San Diego, CA: Academic Press Inc.

Granger, C. W. J., & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, *1*, 15–29.

Gray, H. L., Zhang, N. F., & Woodward, W. A. (1989). On generalized fractional processes. *Journal of Time Series Analysis*, *10*, 233–257.

Grenander, U., & Szegö, G. (1958). Toeplitz forms and their applications. California monographs in mathematical sciences. Berkeley: University of California Press.

Harvey, A. C., Ruiz, E., & Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, *61*, 247–265.

Hassler, U. (1994). (Mis)specification of long memory in seasonal time series. *Journal of Time Series Analysis*, *15*, 19–30.

Hassler, U., & Wolters, J. (1995). Long memory in inflation rates: International evidence. *Journal of Business and Economic Statistics*, *13*, 37–45.

Haslett, J., & Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Journal of Applied Statistics*, *38*, 1–50.

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

Karanasos, M., Psaradakis, Z., & Sola, M. (2004). On the autocorrelation properties of long-memory GARCH processes. *Journal of Time Series Analysis*, *25*, 265–281.

Levinson, N. (1947). The Wiener RMS (root mean square) error criterion in filter design and prediction. *Journal of Mathematical Physics of the Massachusetts Institute of Technology*, *25*, 261–278.

Li, W. K., & McLeod, A. I. (1986). Fractional time series modelling. *Biometrika*, *73*, 217–221.

Ling, S., & Li, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association*, *92*, 1184–1194.

Montanari, A., Rosso, R., & Taqqu, M. S. (2000). A seasonal fractional ARIMA model applied to Nile River monthly flows at Aswan. *Water Resources Research*, *36*, 1249–1259.

Ooms, M. (1995). *Flexible seasonal long memory and economic time series*. Rotterdam. Technical report Econometric Institute, Erasmus University.

Palma, W., & Chan, N. H. (1997). Estimation and forecasting of long-memory processes with missing values. *Journal of Forecasting*, *16*, 395–410.

Palma, W., & Chan, N. H. (2005). Efficient estimations of seasonal long range dependent processes. *Journal of Time Series Analysis*, *26*, 863–892.

Palma, W., & Del Pino, G. (1999). Statistical analysis of incomplete long-range dependent data. *Biometrika*, *86*, 965–972.

Palma, W., & Zevallos, M. (2004). Analysis of the correlation structure of square time series. *Journal of Time Series Analysis*, *25*, 529–550.

Porter-Hudak, S. (1990). An application of the seasonal fractionally differenced model to the monetary aggregates. *Journal of the American Statistical Association, Applications and Case Studies*, *85*, 338–344.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in FORTRAN*. Cambridge: Cambridge University Press.

Ray, B. K. (1993). Long-range forecasting of IBM product revenues using a seasonal fractionally differenced ARMA model. *International Journal of Forecasting*, *9*, 255–269.

Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, *47*, 67–84.

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

Robinson, P. M. (2003). *Time series with long memory*. Oxford: Oxford University Press.

Rosenblatt, M. (1961). Independence and dependence. In: J. Neyman (Ed.), *Proceeding of the 4th Berkeley Symposium On Mathematics Statistics and Probability* (Vol. II, pp. 431–443). Berkeley, CA: University of California Press.

Shumway, R. H., & Stoffer, D. S. (2000). *Time series analysis and its applications*. New York: Springer-Verlag.

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

Velasco, C., & Robinson, P. M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association*, *95*, 1229–1243.

Whittle, P. (1951). *Hypothesis testing in time series analysis*. New York: Hafner.

Woodward, W. A., Cheng, Q. C., & Gray, H. L. (1998). A *k*-factor GARMA long-memory model. *Journal of Time Series Analysis*, *19*, 485–504.

Yajima, Y. (1985). On estimation of long-memory time series models. *The Australian Journal of Statistics*, *27*, 303–320.