Quasi-maximum likelihood estimation for a class of continuous-time long-memory processes

HENGHSIU TSAI *

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan 115, R.O.C.

K. S. CHAN

Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242, U.S.A.

April 27, 2003

Abstract

Tsai and Chan (2003, an unpublished manuscript) has recently introduced the Continuous-time Auto-Regressive Fractionally Integrated Moving-Average (CARFIMA) Models useful for studying long-memory data. We consider the estimation of the CARFIMA Models with discrete-time data by maximizing the Whittle likelihood. We show that the quasi-maximum likelihood estimator is asymptotically normal and efficient. Finite-sample properties of the quasi-maximum likelihood estimator and those of the exact maximum likelihood estimator are compared by simulations. Simulations suggest that for finite samples, the quasi-maximum likelihood estimator of the Hurst parameter is less biased but more variable than the exact maximum likelihood estimator. We illustrate the method with a real application.

Keywords: asymptotic efficiency, asymptotic normality, CARFIMA models, fractional Brownian motion, Whittle likelihood.

**Corresponding Author*: Henghsiu Tsai, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan 115, R.O.C.; Phone: 886-2-2783-5611 ext. 215; Fax: 886-2-2783-1523; e-mail: htsai@stat.sinica.edu.tw

1 Introduction

It is well known that the long range dependence properties of time series data have found diverse applications in many fields including hydrology, economics and telecommunications; see Bloomfield (1992), Robinson (1993), Beran (1994), Baillie (1996) and Ray and Tsay (1997). Extensive work has been done for discrete time long memory processes in the literature; see Sowell (1992), Robinson (1995) and Chan and Palma (1998). A commonly used model for such processes is the autoregressive fractionally integrated moving-average (ARFIMA) model, see Granger and Joyeux (1980). For continuous-time long memory modelling, Comte and Renault (1996) developed a general class of linear continuous time processes that exhibit long range dependence. Comte (1996) studied the statistical inference of a first-order fractional stochastic differential equation (SDE).

Recently Tsai and Chan (2003, an unpublished manuscript) developed the continuous-time autoregressive fractionally integrated moving average (CARFIMA) models of general order that are based on the stochastic calculus of fractional Brownian motions developed by Duncan *et al.* (2000). Tsai and Chan (2003, an unpublished manuscript) studied the use of the innovations algorithm for maximum likelihood estimation of the CARFIMA models with discrete-time data. The innovations algorithm has the merit of being able to handle irregularly-spaced time series data. Moreover, the predictive residuals generated by the innovations algorithm can be used for model diagnostics. However, the innovations algorithm suffers from the drawback of being computer intensive so that its use becomes increasingly prohibitive with increasing large sample size. Here, we aim to study an alternative estimation method that is both fast and statistically efficient.

In the discrete-time framework, computations can be sped up via several possible approaches for approximating the Gaussian likelihood function; see Beran (1994). Comte (1996) studied two approximation methods for estimating a first-order continuous-time long-memory model with regularly-spaced data, namely the Whittle (1953) likelihood approach and a semi-parametric technique proposed by Geweke and Porter-Hudak (1983) that first estimates the long memory parameter by a log-periodogram regression and then estimates the other parameters by maximizing an associated approximating AR(1) process. While Comte (1996) mentioned that both methods can be generalized to fractional stochastic differential equations of higher order, this extension has not been explicitly reported in the literature. Although Comte (1996) mentioned that the later method is much faster than the former one, it seems that the Whittle likelihood method can be more readily generalized to the case of higher order CARFIMA models.

In this paper, we develop the Whittle likelihood approach to estimating a general-order stationary CARFIMA model (Tsai and Chan 2003, an unpublished manuscript) with regularly-spaced discrete time data. The paper is organized as follows. In section 2, we briefly review the CARFIMA model of Tsai and Chan (2003, an unpublished manuscript) and derive the spectral density function of a stationary CARFIMA model. Quasi-maximum likelihood estimation of a stationary CARFIMA model via maximizing the Whittle likelihood is discussed in section 3. Sections 4 and 5 consider respectively the large sample and finite sample properties of the quasi-maximum likelihood estimator. A real application is illustrated in section 6. We briefly conclude in section 7. All proofs are collected in an appendix.

2 Continuous-time fractionally integrated ARMA processes

Heuristically, a CARFIMA(p,H,q) process is the solution of a p-th order stochastic differential equation with suitable initial condition and driven by a standard fractional Brownian motion with Hurst parameter H and its derivatives up to and including order $0 \le q < p$. Specifically, for $t \ge 0$,

$$Y_t^{(p)} - \alpha_p Y_t^{(p-1)} - \cdots - \alpha_1 Y_t - \alpha_0 = \sigma\{B_{t,H}^{(1)} + \beta_1 B_{t,H}^{(2)} + \cdots + \beta_q B_{t,H}^{(q+1)}\}, \qquad (1)$$

where $0 \le q < p$ and $\{B_{t,H} = B_t^H, t \ge 0\}$ is the standard fractional Brownian motion with Hurst parameter $1/2 < H < 1$; the superscript $^{(j)}$ denotes j-fold differentiation with respect to t. We assume that $\sigma > 0$ and $\beta_q \neq 0$, $dY_t^{(j-1)} = Y_t^{(j)} dt, j = 1, ..., p-1$.

However, the fractional Brownian motion is nowhere differentiable (see Mandelbrot and Van Ness, 1968) so the stochastic equation (1) has to be appropriately interpreted as some integral equation as explained below. Analogous to the case of continuous-time ARMA processes (see, e.g., Brockwell, 1993), equation (1) can be equivalently cast in terms of the *observation* and *state* equations:

$$Y_t = \beta' X_t, \qquad t \geq 0, \tag{2}$$

$$dX_t = (AX_t + \alpha_0 \delta_p)dt + \sigma \delta_p dB_t^H, \tag{3}$$

where the superscript $'$ denotes taking transpose,

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_p \end{bmatrix}, \quad X_t = \begin{bmatrix} X_t^{(0)} \\ X_t^{(1)} \\ \vdots \\ X_t^{(p-2)} \\ X_t^{(p-1)} \end{bmatrix}, \quad \delta_p = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad \beta = \begin{bmatrix} 1 \\ \beta_1 \\ \vdots \\ \beta_{p-2} \\ \beta_{p-1} \end{bmatrix},$$

and $\beta_j = 0$ for $j > q$. Equation (3) for the state vector X_t is defined by the stochastic calculus developed by Duncan *et al.* (2000). In the case that $\beta_j = 0$, $j \geq 1$, the state vector X_t becomes the vector of derivatives of the continuous-time fractionally integrated AR(p) process $\{Y_t\}$.

The process $\{Y_t, t \geq 0\}$ is said to be a CARFIMA(p,H,q) process with parameter $(\theta, \sigma) = (\alpha_0, ..., \alpha_p, \beta_1, \cdots, \beta_q, H, \sigma)$ if $Y_t = \beta' X_t$, where X_t is the solution of (3) with the initial condition X_0 of finite variance. Tsai and Chan (2003, an unpublished manuscript) showed that the solution of (3) can be written as

$$X_t = e^{At} X_0 + \alpha_0 \int_0^t e^{A(t-u)} \delta_p du + \sigma \int_0^t e^{A(t-u)} \delta_p dB_u^H, \tag{4}$$

where $e^{At} = I + \sum_{n=1}^{\infty} \{(At)^n (n!)^{-1}\}$, and I is the identity matrix. Tsai and Chan (2003, an unpublished manuscript) derived an asymptotic stationarity condition for the CARFIMA model, namely all eigenvalues of A have negative real parts. Moreover, they showed that the stationary mean equals $\mu = -(\alpha_0/\alpha_1)\delta_1$, where $\delta_1 = [1, 0, \cdots, 0]'$ and that the autocovariance function of $\{Y_t, t \geq 0\}$ equals

$$\begin{aligned} \gamma_Y(h) &:= \operatorname{cov}(Y_{t+h}, Y_t) \\ &= C_H \beta' e^{Ah} \left(\int_0^h e^{-Au} u^{2H-2} du \right) V^* \beta + C_H \beta' e^{-Ah} \left(\int_h^\infty e^{Au} u^{2H-2} du \right) V^* \beta \\ &\quad + C_H \beta' e^{Ah} \left(\int_0^\infty e^{Au} u^{2H-2} du \right) V^* \beta. \end{aligned} \tag{5}$$

We now derive a closed-form expression for the spectral density function of $\{Y_t, t \geq 0\}$. First note that spectral density function is independent of the value of α_0. Therefore, without loss of generality, assume $\alpha_0 = 0$. Then equation (1) can be written as

$$\alpha(D)Y_t = \sigma\beta(D)DB_t^H, \tag{6}$$

where

$$\alpha(z) = z^p - \alpha_p z^{p-1} - \cdots - \alpha_1, \tag{7}$$

$$\beta(z) = 1 + \beta_1 z + \beta_2 z^2 + \cdots + \beta_q z^q. \tag{8}$$

By equations (4.12.45) and (4.12.47) of Priestley (1981), the (continuous time) spectral density of $\{Y_t, t \geq 0\}$ equals

$$h_Y(w) = \sigma^2 \frac{|\beta(iw)|^2}{|\alpha(iw)|^2} h_\epsilon(w), \tag{9}$$

where $\{h_\epsilon(w)\}$ is the spectral density function of $\{\epsilon_t^H, t \geq 0\}$ and $B_t^H = \int_0^t \epsilon_u^H du$. Note that

$$\text{cov}(B_t^H, B_s^H) = \text{cov}\left(\int_0^t \epsilon_u^H du, \int_0^s \epsilon_v^H dv\right) = \int_0^t \int_0^s \text{cov}(\epsilon_u^H, \epsilon_v^H) du dv. \tag{10}$$

But by Duncan (1999),

$$\text{cov}(B_t^H, B_s^H) = H(2H-1) \int_0^t \int_0^s |u - v|^{2H-2} du dv. \tag{11}$$

Comparing equations (11) and (10), we get $\text{cov}(\epsilon_u^H, \epsilon_v^H) = H(2H-1)|u-v|^{2H-2}$. Therefore, the covariance function for $\{\epsilon_u^H, u \geq 0\}$ equals $\gamma_\epsilon(\tau) = \text{cov}(\epsilon_{u+\tau}^H, \epsilon_u^H) = H(2H-1)|\tau|^{2H-2}$. Consequently, the spectral density function of $\{\epsilon_t^H, t \geq 0\}$ can be obtained as the Fourier transform of $\gamma_\epsilon(\tau)$, i.e.

$$\begin{aligned} h_\epsilon(w) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iw\tau} \gamma_\epsilon(\tau) d\tau \\ &= \frac{H(2H-1)}{2\pi} \int_{-\infty}^{\infty} e^{-iw\tau} |\tau|^{2H-2} d\tau \\ &= \frac{H(2H-1)}{2\pi} 2\Gamma(2H-1)|w|^{1-2H} \cos\left[\frac{\pi}{2}(2H-1)\right] \\ &= \frac{1}{2\pi}\Gamma(2H+1)|w|^{1-2H} \sin(\pi H), \end{aligned} \tag{12}$$

where $\Gamma()$ is the Gamma function. The spectral density function of $\{Y_t, t \geq 0\}$ derived heuristically above can be proved rigorously. We summarize the above result in the following theorem.

THEOREM 1 *The spectral density function of* $\{Y_t, t \geq 0\}$ *equals*

$$h_Y(w) \quad = \quad \frac{\sigma^2}{2\pi}\Gamma(2H+1)\sin(\pi H)|w|^{1-2H}\frac{|\beta(iw)|^2}{|\alpha(iw)|^2}. \tag{13}$$

Notice that for $H = 1/2$, $h_\epsilon(w) = 1/(2\pi)$, and the spectral density function given by equation (13) becomes that of the short-memory CARMA(p,q) process.

3 Quasi-maximum likelihood estimation

Let $\{Y_{ih}\}_{i=1,\dots,N}$ be the observations, where h being the step size. By the aliasing formula, the (discrete time) spectral density of Y equals

$$f_h(w,\theta,\sigma^2) = \frac{1}{h}\sum_{k\in Z} h_Y\left(\frac{w+2k\pi}{h}\right); \tag{14}$$

see Priestley (1981) and Comte (1996). Let r denote the largest integer smaller than or equal to $(N-1)/2$. The (negative) log-likelihood function of $\{Y_{ih}\}$ can be approximated by the (negative) Whittle log-likelihood function (see Beran, 1994 and Dahlhaus, 1989):

$$-\tilde{l}_w(\theta,\sigma^2) = \sum_{j=1}^{r}\left\{\log f_h(w_j,\theta,\sigma^2) + \frac{I_N(w_j)}{f_h(w_j,\theta,\sigma^2)}\right\}, \tag{15}$$

where $\theta := (\theta_1,\cdots,\theta_{p+q+1}) = (\alpha_1,\cdots,\alpha_p,\beta_1,\cdots,\beta_q,H)$, $w_j := 2\pi j/N \in (0,\pi)$ is the Fourier frequency, and

$$I_N(w) = \frac{1}{2\pi N}\left|\sum_{j=1}^{N}(Y_{jh}-\bar{Y})e^{ijw}\right|^2, \quad \bar{Y} = \frac{1}{N}\sum_{j=1}^{N}Y_{jh}.$$

Letting $g_h(w_j,\theta) = f_h(w_j,\theta)/\sigma^2$, the (negative) Whittle log-likelihood function (15) can be rewritten as

$$-\tilde{l}_w(\theta,\sigma^2) = \sum_{j=1}^{r}\left\{\log\sigma^2 + \log g_h(w_j,\theta) + \frac{I_N(w_j)}{\sigma^2 g_h(w_j,\theta)}\right\}. \tag{16}$$

Differentiating (16) with respect to σ^2 and equating to zero gives

$$\begin{aligned}\sigma^2 &= \frac{1}{r}\sum_{j=1}^{r}\frac{I_N(w_j)}{g_h(w_j,\theta)}\\ &= \frac{2\pi h}{r\Gamma(2H+1)\sin(\pi H)}\sum_{j=1}^{r}\frac{I_N(w_j)}{g_h^*(w_j,\theta)},\end{aligned} \tag{17}$$

where

$$g_h^*(w,\theta) = \sum_{k\in Z} g_Y^*\left(\frac{w+2k\pi}{h}\right) \approx \sum_{|k|\le M} g_Y^*\left(\frac{w+2k\pi}{h}\right) \tag{18}$$

for some large M, and

$$g_Y^*(w) \quad = \quad |w|^{1-2H}\frac{|\beta(iw)|^2}{|\alpha(iw)|^2}.$$

In practice, setting $M = 100$ yields sufficient accuracy; see Comte (1996) and Percival and Walden (2000, p. 280).

Substituting equation (17) into (16) yields the objective function

$$\begin{aligned} -\tilde{l}_w(\theta) &= \sum_{j=1}^{r}\log g_h(w_j,\theta) + r\log\left(\sum_{j=1}^{r}\frac{I_N(w_j)}{g_h(w_j,\theta)}\right) + \text{constant} \\ &= \sum_{j=1}^{r}\log g_h^*(w_j,\theta) + r\log\left(\sum_{j=1}^{r}\frac{I_N(w_j)}{g_h^*(w_j,\theta)}\right) + \text{constant}. \end{aligned} \tag{19}$$

The objective function is then minimized with respect to θ to get the quasi-maximum likelihood estimator (QMLE) $\hat{\theta}$. The estimator $\hat{\sigma}^2$ is then calculated by (17).

4 Large sample properties of the quasi-maximum likelihood estimator

For regularly-spaced time series data sampled from a stationary CARFIMA(p,H,q) process, the quasi-maximum likelihood estimator can be shown to be asymptotically normal and efficient by making use of some results in Dahlhaus (Theorems 2.1 and 4.1, 1989). For simplicity, in this section, we augment σ into $\theta = (\alpha_0, ..., \alpha_p, \beta_1, \cdots, \beta_q, H, \sigma)$. Let $\hat{\theta}$ be the quasi-maximum likelihood estimator that minimizes the negative Whittle log-likelihood function (15).

THEOREM 2 *Let $Y = \{Y_{t_i}\}_{i=1}^N$ be sampled from a stationary CARFIMA(p,H,q) process, where $t_i = ih$ with $h > 0$ being the step size. Let the quasi-maximum likelihood estimator $\hat{\theta} = \hat{\theta}_N \in \Theta$, a compact parameter space, and the true parameter θ_0 be in the interior of the parameter space. Then,*

$$\sqrt{N}(\hat{\theta} - \theta_0) \longrightarrow N(0, \Gamma_h^{-1}(\theta_0))$$

with $\Gamma_h(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}(\bigtriangledown \log f_h(x))(\bigtriangledown \log f_h(x))'dx,$

where $\bigtriangledown$ denotes taking the derivative w.r.t. θ. Moreover, $\hat{\theta}$ is asymptotically efficient.

We note that the compactness condition on the parameter space is taken from condition (A0) in Dahlhaus (1989) who pointed out that the quasi-maximum likelihood estimator may lie on the boundary of the compact parameter space.

5 Simulation studies

The purpose of this section is to evaluate the finite sample performance of the quasi-maximum likelihood estimator. We consider five CARFIMA(p,H,0) models for $p = 1$ and $p = 2$. For each model, we simulated regularly spaced time series data, $Y_i = X_i^{(0)}, i = 1, ..., N$, by the method of Davies and Harte's (1987). See also Comte (1996). The sample size is always $N = 512$. Each experiment is replicated 1000 times. The quasi-maximum likelihood estimates were obtained by constrained numerical optimization via the DBCONF subroutine of the IMSL package that approximates derivatives by a finite-difference scheme. The parameter H is constrained to be between 0.50 and 0.99, and the α's between -10^{30} and 0. For comparison purposes, we also compute the exact maximum likelihood estimators (exact MLEs) (see Tsai and Chan, 2003, an unpublished manuscript, for details). While theory predicts normality, the estimates of the α's appear to be non-normal and have a few outliers, perhaps because the sample size is not yet large enough. Consequently, we report in Table 1 robust summary statistics for the location and spread of the estimators. Specifically we use the median as a location measure and $s = \sqrt{\pi/2} \times$ the mean of the absolute deviations of the estimates from the median as a measure of the spread. We also report the theoretical asymptotic standard errors of the parameter estimators, which are computed from $\Gamma_h(\theta)$ defined in Theorem 2. Finally, we compute for each parameter of each model the empirical coverage rates of the 95% confidence intervals using the asymptotic standard errors; these confidence intervals take the form of estimate $\pm$ 1.96 $\times$ the asymptotic standard error.

The value M of equation (18) is set to be 100. We have tried different M in the program and the results are very robust to the choice of M. The initial values of the DBCONF are conveniently set to be the true values. We have tried different initial values in the program and the results are very robust to the choice of the initial values. From Table 1, it can be seen that the biases of the exact MLEs are generally larger than those of the Whittle MLEs, whereas the standard errors of the former are smaller than those of the latter. In terms of the empirical coverage rates, confidence intervals centered at the exact MLEs are closer to the nominal values for the parameters α's and σ but intervals centered at the Whittle MLEs are closer to the nominal values for the long memory parameter H.

6 Application

Example: We consider a series of annual tree ring measurements from New Mexico, USA, from 837 through 1989 AD, a total of 1,153 data. Each tree ring measurement represents the relative or normalized tree-ring width, in dimensionless units, which depicts the annual growth of a tree; the data are posted in the file NM560.DAT at the website `http://www-personal.buseco.monash.edu.au/~hyndman/TSDL`. Many tree ring data exhibit the long range dependence properties; see Baillie (1996). Figure 1a displays the time series plot of the tree ring measurements, whereas figure 1b exhibits the sample autocorrelation of the data. We have fitted $CARFIMA(p, H, 0)$ models to the tree ring data with the autoregressive order $0 \leq p \leq 4$. For $p \leq 2$, the initial value of H is set to be 0.75 and those of α's equal -1.0. For $p \geq 3$, the initial value of H is set to be the estimate of H from $p = 2$ and those of α's obtained by maximizing the Whittle likelihood with H fixed at the initial value. We report in Table 2 the quasi-maximum likelihood estimates of H for different p and the corresponding Akaike information criterion $AIC = -2(l_Y(\hat{\theta}) - r)$, where r is the number of parameters in the model, $-l_Y$ is replaced by $-\tilde{l}_w$ expressed in equation (19) without the constant term, and $\hat{\theta}$ is the quasi-maximum likelihood estimator of θ. Based on the AIC, the autoregressive order is selected as $p = 2$. The parameter estimates are $(\hat{H}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\sigma}) = (0.7858, -11.55, -3.427, 4.534)$. We have also checked the residuals computed by the innovations algorithm (Tsai and Chan, 2003, an unpublished manuscript) and they appear to be white, suggesting that this model provides

Table 1: Comparison between the quasi-maximum likelihood estimator and the exact maximum likelihood estimator. The table list for each estimation method the median, ($s = \sqrt{\pi/2} \times$ mean|estimates-median(estimates)|) and (c=empirical coverage rates of the 95% C.I. using the asymptotic std. error) of 1,000 simulated values. The sample size is 512.

p		true value	quasi-MLE median	(s)	(c)	exact MLE median	(s)	(c)	asymptotic std. error
1	H	0.6	0.5952	(0.0744)	(0.975)	0.5695	(0.0685)	(0.995)	0.0692
	α_1	-2	-1.9775	(1.0529)	(0.909)	-1.8231	(0.6282)	(0.955)	0.6048
	σ	2	1.9951	(0.6699)	(0.903)	1.8732	(0.3915)	(0.954)	0.3864
1	H	0.75	0.7475	(0.0781)	(0.910)	0.7182	(0.0785)	(0.871)	0.0640
	α_1	-2	-1.9832	(0.7423)	(0.891)	-1.8081	(0.5826)	(0.919)	0.5482
	σ	2	1.9995	(0.7532)	(0.899)	1.8257	(0.4696)	(0.961)	0.4618
1	H	0.9	0.9019	(0.0737)	(0.937)	0.8611	(0.0744)	(0.870)	0.0624
	α_1	-2	-2.0144	(0.6029)	(0.911)	-1.7596	(0.4899)	(0.937)	0.5247
	σ	2	2.0353	(1.6191)	(0.782)	1.6339	(0.6055)	(0.986)	0.8226
2	H	0.75	0.7518	(0.0510)	(0.923)	0.7364	(0.0490)	(0.918)	0.0475
	α_1	-2	-2.0096	(0.1559)	(0.936)	-1.9708	(0.1504)	(0.944)	0.1465
	α_2	-1	-0.9988	(0.1143)	(0.940)	-0.9890	(0.1122)	(0.945)	0.1083
	σ	2	2.0172	(0.1751)	(0.918)	1.9609	(0.1535)	(0.959)	0.1563
2	H	0.9	0.9072	(0.0630)	(0.949)	0.8701	(0.0580)	(0.903)	0.0584
	α_1	-3	-3.0205	(0.3758)	(0.915)	-2.8669	(0.3498)	(0.931)	0.3446
	α_2	-2	-1.9750	(0.3049)	(0.935)	-1.9893	(0.3107)	(0.928)	0.2905
	σ	2	2.0743	(1.0059)	(0.797)	1.7937	(0.3992)	(0.986)	0.5209

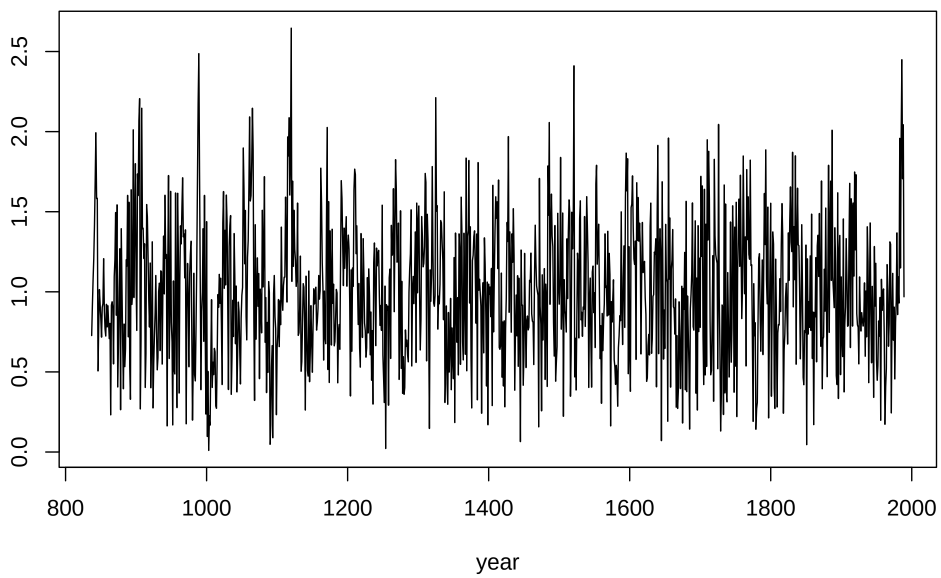
Table 2: AIC and Maximum likelihood estimates of H

	p	0	1	2	3	4
original data	AIC	-2178.74	-2178.93	-2179.88	-2178.69	-2176.04
(N=1153)	H	0.6452	0.7860	0.7858	0.6895	0.7737

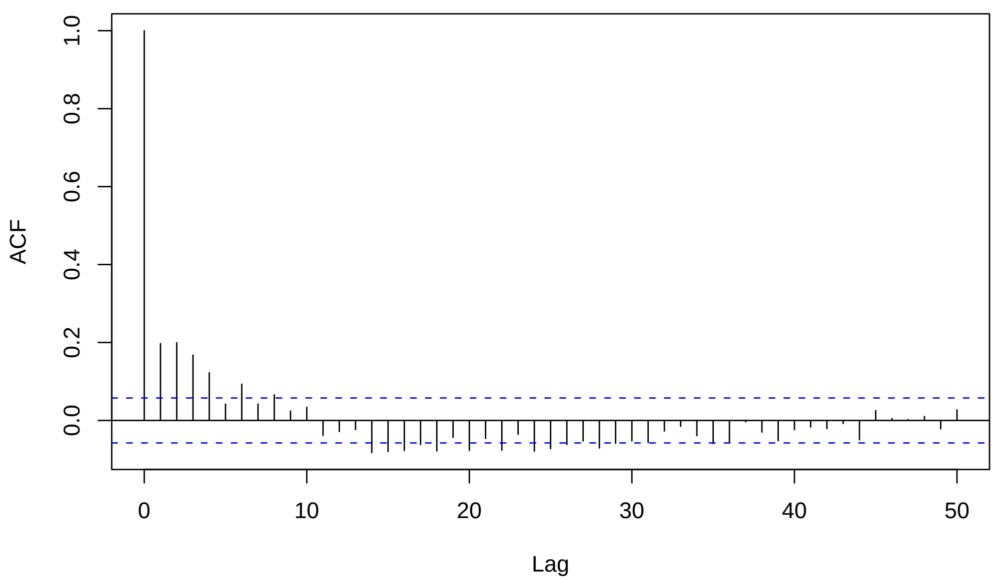
adequate fit to the data.

To assess the accuracy of the quasi-maximum likelihood estimator, we have carried out parametric bootstrap with bootstrap size equal to 1000. Except for H, The bootstrap estimates appear to have skewed distributions; consequently we report the 95% bootstrap confidence intervals for the estimates in Table 3; these confidence intervals are obtained by the percentile method (Efron and Tibshirani, 1993, Ch. 13) with the 2.5 and 97.5 percentiles of the bootstrap estimates as end points of the 95% confidence intervals. The 95% confidence interval of H extends from 0.7275 to 0.8617, suggesting that the data are indeed of long memory. We have also reported in Table 3 the asymptotic standard errors of the estimator which except for H, are substantially larger than their bootstrap counterparts. Therefore, we re-did the bootstrap study with larger sample sizes. Table 3 suggests that the relative differences between the bootstrap standard errors and their asymptotic counterparts become smaller with larger sample size, although the differences are still noticeable even with $N = 8192$. In particular, we recommend using parametric bootstrap to calibrate the uncertainty of the quasi-maximum likelihood estimator.

The population spectral density function can be estimated by (13) with the unknown parameters there replaced by the quasi-maximum likelihood estimates. While the estimated spectral density peaks only at 0 with a singularity there (Fig. 2, over the range $0 < w < 5$), the squared magnitude of the transfer function of the autoregressive filter, i.e. $1/|\alpha(iw)|^2$, peaks at $w = 2.38$, with the corresponding period equal to $2\pi/2.38 = 2.64$ years. Thus, besides the long-memory component, the tree ring dataset admits a short-memory component with a significant contribution from cycles of period about 2.64 years.



(a)



(b)

Figure 1: The annual tree ring measurements from New Mexico, USA; (a) time series plot and (b) sample autocorrelation function.

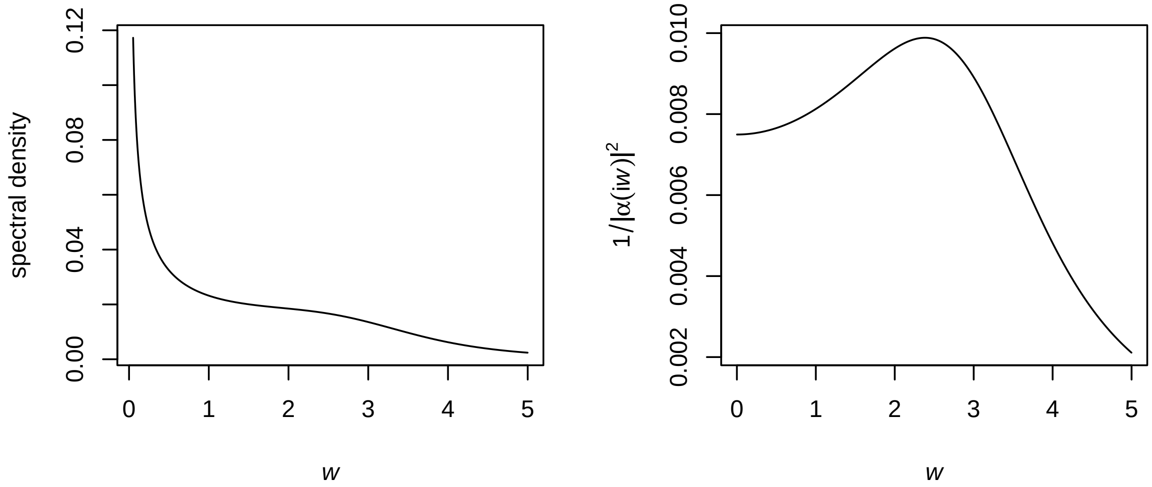


Figure 2: The spectral density of the fitted model and squared magnitude of the transfer function of the autoregressive part of the model.

Table 3: 95% parametric bootstrap confidence intervals of the fitted CARFIMA model for the tree ring data. We have also reported the median and a robust spread measure ($s = \sqrt{\pi/2} \times$ mean|estimates-median(estimates)|) of the bootstrap estimates as well as the asymptotic standard errors of the quasi-maximum likelihood estimators.

N		estimated value	median (s)	95% bootstrap confidence intervals	asymptotic std. error
1153	H	0.7858	0.7972 (0.0333)	(0.7275, 0.8617)	0.0334
	α_1	-11.55	-11.48 (4.843)	(-26.71, -7.248)	11.63
	α_2	-3.427	-3.228 (1.036)	(-5.443, -1.952)	2.917
	σ	4.534	4.475 (1.758)	(2.792, 9.731)	4.482
4096	H	0.7858	0.7902 (0.0178)	(0.7533, 0.8247)	0.0177
	α_1	-11.55	-11.37 (3.107)	(-20.93, -8.187)	6.172
	α_2	-3.427	-3.365 (0.7298)	(-4.913, -2.326)	1.547
	σ	4.534	4.473 (1.178)	(3.1702, 7.928)	2.378
8192	H	0.7858	0.7892 (0.0123)	(0.7651, 0.8142)	0.0125
	α_1	-11.55	-11.90 (2.874)	(-19.74, -8.676)	4.364
	α_2	-3.427	-3.516 (0.6526)	(-4.815, -2.559)	1.094
	σ	4.534	4.678 (1.088)	(3.394, 7.495)	1.682

7 Conclusion

We have shown that the quasi-maximum likelihood estimator provides an useful alternative to the exact maximum likelihood estimator of a stationary CARFIMA model. A limitation of the quasi-maximum likelihood approach is that the data must be sampled regularly. Extension of the method to irregularly spaced data will be of much interest especially if a speedy implementation is available.

APPENDIX

Proof of Theorem 1

The spectral density function of $\{Y_t, t \geq 0\}$ equals

$$\begin{aligned} h_Y(w) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iw\tau}\gamma_Y(\tau)d\tau \\ &= \frac{1}{\pi}\mathrm{Re}\left\{\int_0^{\infty} e^{-iw\tau}\gamma_Y(\tau)d\tau\right\}, \end{aligned}$$

where $\mathrm{Re}\{\cdot\}$ means the real part of the item in the parenthesis. Note that

$$\begin{aligned} & \int_0^{\infty} e^{-iw\tau}\gamma_Y(\tau)d\tau \\ = \ & C_H\int_0^{\infty} e^{-iw\tau}\beta' e^{A\tau}\left(\int_0^{\tau} e^{-Au}u^{2H-2}du\right)V^*\beta d\tau \\ & +C_H\int_0^{\infty} e^{-iw\tau}\beta' V^* e^{-A'\tau}\left(\int_{\tau}^{\infty} e^{A'u}u^{2H-2}du\right)\beta d\tau \\ & C_H\int_0^{\infty} e^{-iw\tau}\beta' e^{A\tau}\left(\int_0^{\infty} e^{Au}u^{2H-2}du\right)V^*\beta d\tau \\ = \ & -2C_H\beta'(A^2+w^2I)^{-1}AV^*\beta\left(\int_0^{\infty} e^{-iwu}u^{2H-2}du\right) \\ & -2iwC_H\beta'(A^2+w^2I)^{-1}\left(\int_0^{\infty} e^{Au}u^{2H-2}du\right)V^*\beta. \end{aligned}$$

Thus,

$$h_Y(w) = -\frac{2}{\pi}C_H\beta'(A^2+w^2I)^{-1}AV^*\beta R\left\{\int_0^{\infty} e^{-iwu}u^{2H-2}du\right\}. \tag{20}$$

By equation (6.20) of Karatzas and Shreve (1991), we have

$$AV^* + V^*A' = -\sigma^2\delta_p\delta_p', \tag{21}$$

which implies that

$$\begin{aligned} \beta'(A^2+w^2I)^{-1}AV^*\beta &= -\sigma^2\beta' A(A+iwI)^{-1}(A-iwI)^{-1}\delta_p\delta_p'(A'+iwI)^{-1}\beta \\ & \quad -\beta'(A+iwI)^{-1}AV^*(A'+iwI)^{-1}\beta, \end{aligned} \tag{22}$$

and

$$\beta'(A+iwI)^{-1}AV^*(A'+iwI)^{-1}\beta = -\frac{1}{2}\sigma^2\beta'(A+iwI)^{-1}\delta_p\delta_p'(A'+iwI)^{-1}\beta. \tag{23}$$

Substituting equation (23) into equation (22) we get

$$\begin{aligned} & \beta'(A^2+w^2I)^{-1}AV^*\beta \\ = \ & -\frac{\sigma^2}{2}\beta'(A-iwI)^{-1}\delta_p\delta_p'(A'+iwI)^{-1}\beta \\ = \ & -\frac{\sigma^2}{2}\frac{|\beta(iw)|^2}{|\alpha(iw)|^2}, \end{aligned} \tag{24}$$

the last equality following from the fact that $-\alpha(s)(A-sI)^{-1}\delta_p = [1, s, \cdots, s^{p-1}]'$. By equations (20) and (24), we have

$$
\begin{aligned}
h_Y(w) &= \frac{\sigma^2}{\pi} C_H \frac{|\beta(iw)|^2}{|\alpha(iw)|^2} R\left\{\int_0^\infty e^{-iwu} u^{2H-2} du\right\} \\
&= \frac{\sigma^2}{2\pi} C_H \frac{|\beta(iw)|^2}{|\alpha(iw)|^2} \int_{-\infty}^\infty e^{-iwu} |u|^{2H-2} du \\
&= \sigma^2 \frac{|\beta(iw)|^2}{|\alpha(iw)|^2} h_\epsilon(w).
\end{aligned}
$$

This completes the proof of Theorem 1.

Proof of Theorem 2

Theorem 2 follows from Theorems 2.1 and 4.1 of Dahlhaus (1989) if we can verify conditions (A0)-(A7) listed there. Incidentally, assumption (A8) there also holds as $\alpha(\theta)$ appearing in (A1-7) can be chosen to be $2H-1$ that is independent of θ. We now verify (A0-7).

(A0) The identifiability of the model with regularly spaced data can be checked as follows. Without loss of generality, let $h=1$. Denote the (discrete time) spectral density of $Y = \{Y_i, i \in Z\}$ by

$$
f(w) = f_c(w) + \sum_{k\neq 0} f_c(w + 2\pi k),
$$

where $f_c(w) = h_Y(w)$ is given by (13). Consider

$$
\begin{aligned}
\log f(w) &= \log f_c(w) + \log\left\{\frac{\sum_{k\neq 0} f_c(w+2\pi k)}{f_c(w)} + 1\right\} \\
&= (1-2H)\log|w| + \log\{g(H)\} + \log\{L(w)\} \\
&\quad + \log\left\{|w|^{2H-1} x(w) + 1\right\}, \qquad (25)
\end{aligned}
$$

where $g(H) = \sin(\pi H)\Gamma(2H+1)/(2\pi)$, $L(w) = \sigma^2|\beta(iw)|^2/|\alpha(iw)|^2$, and $x(w) = \sum_{k\neq 0} f_c(w+2\pi k)/\{g(H)L(w)\}$. It follows from $\lim_{w\longrightarrow 0} \log f(w)/\log|w| = 1-2H$ that the Hurst parameter H is identifiable. By equation (25), we have $\lim_{w\longrightarrow 0}\{\log f(w) - (1-2H)\log|w|\} - \log\{g(H)\} = \log\{L(0)\}$, which implies the identifiability of $L(0)$. Again from equation (25),

$$
\begin{aligned}
&\lim_{w\longrightarrow 0}\left[|w|^{2-2H}\frac{\partial}{\partial w}\{\log f(w) - (1-2H)\log|w|\}\right] \\
&= \lim_{w\longrightarrow 0}\left[|w|^{2-2H}\lim_{w\longrightarrow 0}\left\{\frac{1}{L(w)}\frac{\partial L(w)}{\partial w}\right\}\right] + \lim_{w\longrightarrow 0}\frac{|w|^{2-2H}S(w)}{R(w)} \\
&= (2H-1)x(0),
\end{aligned}
$$

where $R(w) = |w|^{2H-1}x(w) + 1$, $S(w) = (2H-1)|w|^{2H-2}x(w) + |w|^{2H-1}\dot{x}(w)$, $\dot{x}(w) = \partial x(w)/\partial w$ and $\dot{L}(0) = \partial L(w)/\partial w|_{w=0}$; hence $x(0)$ is identifiable. It can be checked that $x(w) - x(0) = wO(1)$ for w tending to 0. Now,

$$\begin{aligned} & \lim_{w \longrightarrow 0} \frac{\partial}{\partial w}\{\log f(w) - (1-2H)\log|w|\} \\ = & \lim_{w \longrightarrow 0}\left\{\frac{1}{L(w)}\frac{\partial L(w)}{\partial w}\right\} + \lim_{w \longrightarrow 0}\frac{S(w)}{R(w)} \\ = & \frac{\dot{L}(0)}{L(0)}, \end{aligned}$$

showing that $\dot{L}(0)$ is identifiable. Similarly, given the spectral density function of the discrete-time data, we can compute all the higher derivatives of x and L at 0. Because $L(w)$ is an analytic function for real w, it is uniquely determined by all of its derivatives at 0. By analytic continuation, $L(w)$ as a meromorphic function with complex w is determined for all w. The poles of L then determine uniquely the coefficients in $\alpha(z)$ and the zeroes of L then determine all the coefficients in $\beta(z)$. The value of L at a regular w, i.e. a non-zero or pole of L, then determines the instantaneous variance σ^2.

(A1) $u(\theta) = \int_{-\pi}^{\pi} \log f(x)dx$ can be differentiated twice under the integral.

Verification of (A1): We only give the proof that $u(\theta)$ is differentiable with respect to θ as the argument for the twice-differentiability of u is similar. Suppose differentiation and integration can be interchanged, so that

$$\begin{aligned} & \frac{\partial}{\partial\theta}\int_{-\pi}^{\pi}\log f(x)dx \\ = & \int_{-\pi}^{\pi}\frac{\partial}{\partial\theta}\log\left\{\sum_k f_c(x+2\pi k)\right\}dx \\ = & \int_{-\pi}^{\pi}\frac{1}{f(x)}\sum_k\left\{\frac{\partial}{\partial\theta}f_c(x+2\pi k)\right\}dx \\ = & \int_{-\pi}^{\pi}\frac{1}{f(x)}\sum_k\left\{f_c(x+2\pi k)\frac{\partial}{\partial\theta}\log f_c(x+2\pi k)\right\}dx. \end{aligned}$$

We can justify the interchange by showing that the integrand of the last integral is absolutely integrable as follows. Because $f(x)$ is bounded away from 0 on $[-\pi, \pi]$, there exists some $\delta > 0$ such that $f(x) > \delta > 0$ for all $x \in [-\pi, \pi]$. Therefore,

$$\begin{aligned} & \int_{-\pi}^{\pi}\left|\frac{1}{f(x)}\sum_k\left\{f_c(x+2\pi k)\frac{\partial}{\partial\theta}\log f_c(x+2\pi k)\right\}\right|dx \\ \leq & \frac{1}{\delta}\left[\int_{-\pi}^{\pi}\left|\sum_{k\neq 0}\left\{f_c(x+2\pi k)\frac{\partial}{\partial\theta}\log f_c(x+2\pi k)\right\}\right|dx + \int_{-\pi}^{\pi}\left|\frac{\partial}{\partial\theta}\log f_c(x)\right|dx\right]. \end{aligned}$$

Below $g_i(H)$, $1 \le i \le 5$, denote some generic function of H. Note that $\partial \log f_c(x+2\pi k)/\partial H = -2\log|x+2k\pi| + g_1(H)$, hence

$$\int_{-\pi}^{\pi} \left| \frac{\partial}{\partial H} \log f_c(x) \right| dx \le 2\int_{-\pi}^{\pi} \log|x| dx + g_2(H) < \infty,$$

and

$$\begin{aligned}
&\int_{-\pi}^{\pi} \left| \sum_{k\neq 0} \left\{ f_c(x+2\pi k) \frac{\partial}{\partial H} \log f_c(x+2\pi k) \right\} \right| dx \\
\le\ & 2\int_{-\pi}^{\pi} \left| \sum_{k\neq 0} f_c(x+2\pi k) \log|x+2k\pi| \right| dx + g_1(H) \int_{-\pi}^{\pi} \left| \sum_{k\neq 0} f_c(x+2\pi k) \right| dx \\
=\ & 2g_3(H) \int_{-\pi}^{\pi} \left| \sum_{k\neq 0} |x+2k\pi|^{1-2H} \log|x+2k\pi| \frac{|\beta(i(x+2k\pi))|^2}{|\alpha(i(x+2k\pi))|^2} \right| dx \\
& + g_4(H) \int_{-\pi}^{\pi} \left| \sum_{k\neq 0} |x+2k\pi|^{1-2H} \frac{|\beta(i(x+2k\pi))|^2}{|\alpha(i(x+2k\pi))|^2} \right| dx. \qquad (26)
\end{aligned}$$

Because $\lim_{k\longrightarrow\infty} |x+2k\pi|^{1-2H} \log|x+2k\pi| = 0$, and $\lim_{k\longrightarrow\infty} |x+2k\pi|^{1-2H} = 0$, there exists constants γ_1 and γ_2 such that for all $k \neq 0$ and $-\pi \le x \le \pi$, $|x+2k\pi|^{1-2H} \log|x+2k\pi| \le \gamma_1$, and $|x+2k\pi|^{1-2H} \le \gamma_2$. Therefore, from equation (26), we have

$$\int_{-\pi}^{\pi} \left| \sum_{k\neq 0} \left\{ f_c(x+2\pi k) \frac{\partial}{\partial \theta} \log f_c(x+2\pi k) \right\} \right| dx \ \le\ g_5(H) \int_{-\pi}^{\pi} \sum_{k\neq 0} \frac{|\beta(i(x+2k\pi))|^2}{|\alpha(i(x+2k\pi))|^2} dx < \infty.$$

This shows that $u(\theta)$ can be differentiated with respect to H under the integral. Note also that

$$\frac{\partial}{\partial \beta_i} \log f_c(x+2\pi k) = \frac{1}{|\beta(i(x+2k\pi))|^2} \frac{\partial |\beta(i(x+2k\pi))|^2}{\partial \beta_i},$$

which is a rational polynomial whose denominator is of higher or equal degree than the numerator, and $|\beta(i(x+2k\pi))|^2 > 0$, so $\partial \log f_c(x+2\pi k)/\partial \beta_i$ is bounded above. Therefore, it can be seen that $u(\theta)$ is differentiable with respect to the β_i's under the integral. Similarly, $u(\theta)$ is differentiable with respect to the α_i's.

There exists a function $d : \Theta \longrightarrow (0,1)$ such that for each $\delta > 0$:

(A2) $f(x)$ is continuous at all (x,θ), $x \neq 0$, $f(x)^{-1}$ is continuous at all (x,θ) and $f(x) = O(|x|^{-d(\theta)-\delta})$.

Verification of (A2): It is obvious that $f(x)$ is continuous at all (x,θ), $x \neq 0$ and $f(x)^{-1}$ is continuous at all (x,θ). Note that

$$f(x) \ = \ g(H)|x|^{1-2H} L(x) + \sum_{k\neq 0} f_c(x+2\pi k)$$

$$\begin{aligned} &\approx g(H)|x|^{1-2H}L(0), \text{ for } x \longrightarrow 0 \\ &= O(|x|^{1-2H-\delta}) \text{ for all } \delta > 0. \end{aligned}$$

Thus, we can choose $d(\theta) = 2H - 1$ here and also for the remaining conditions.

(A3)

$$\frac{\partial}{\partial\theta_j} f^{-1}(x), \frac{\partial^2}{\partial\theta_j\partial\theta_k} f^{-1}(x), \frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l} f^{-1}(x)$$

are continuous at all (x, θ) and

$$\frac{\partial}{\partial\theta_j} f^{-1}(x) = O(|x|^{d(\theta)-\delta}), 1 \le j \le p,$$
$$\frac{\partial^2}{\partial\theta_j\partial\theta_k} f^{-1}(x) = O(|x|^{d(\theta)-\delta}), 1 \le j, k \le p,$$
$$\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l} f^{-1}(x) = O(|x|^{d(\theta)-\delta}), 1 \le j, k, l \le p.$$

Verification of (A3): Note that

$$\frac{\partial}{\partial\theta_j} f^{-1}(x) = -f^{-2}(x)\frac{\partial f(x)}{\partial\theta_j},$$

and

$$\begin{aligned} \frac{\partial f(x)}{\partial H} &= g'(H)|x|^{1-2H}L(x) - 2g(H)|x|^{1-2H}L(x)\log|x| + \sum_{k\neq 0}\frac{\partial}{\partial H} f_c(x + 2k\pi) \\ &\approx -2g(H)|x|^{1-2H}L(0)\log|x|, \text{ for } x \longrightarrow 0. \end{aligned}$$

where $g'(H) = \partial g(H)/\partial H$. Thus,

$$\begin{aligned} \frac{\partial}{\partial H} f^{-1}(x) &\approx C\frac{|x|^{1-2H}\log|x|}{|x|^{2-4H}}, \text{ for } x \longrightarrow 0 \\ &= O(|x|^{d(\theta)-\delta}) \text{ for all } \delta > 0. \end{aligned}$$

Similarly,

$$\begin{aligned} \frac{\partial}{\partial\sigma^2} f^{-1}(x) &\approx \frac{g(H)|x|^{1-2H}}{\sigma^2|x|^{2-4H}}, \text{ for } x \longrightarrow 0 \\ &= O(|x|^{d(\theta)-\delta}) \text{ for all } \delta > 0, \end{aligned}$$

and for $\theta_j \in \{\alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q\}$,

$$\begin{aligned} \frac{\partial}{\partial\theta_j} f^{-1}(x) &\approx C\frac{|x|^{1-2H}}{|x|^{2-4H}}, \text{ for } x \longrightarrow 0 \\ &= O(|x|^{d(\theta)-\delta}) \text{ for all } \delta > 0. \end{aligned}$$

The higher derivatives can be treated in the same way with increasingly complicated expressions for the derivatives. Maximal order for second-order derivatives is $x^{2H-1}(\log x)^2$ and for third-order derivatives $x^{2H-1}(\log x)^3$ which are $O(x^{d(\theta)-\delta})$ for all $\delta > 0$. See also p. 35 of Comte(1996).

(A4) $(\partial/\partial x)f(x)$ is continuous at all (x,θ), $x \neq 0$, and

$$\frac{\partial}{\partial x}f(x) = O(|x|^{-d(\theta)-1-\delta}).$$

Verification of (A4):

$$\begin{aligned}
\frac{\partial}{\partial x}f(x) &= g(H)(1-2H)|x|^{-2H}L(x) + g(H)|x|^{1-2H}\dot{L}(x) \\
&\quad + \sum_{k\neq 0}\frac{\partial}{\partial x}f_c(x+2k\pi) \\
&\approx g(H)(1-2H)|x|^{-2H}L(0), \text{ for } x \longrightarrow 0 \\
&= O(|x|^{-d(\theta)-1-\delta}) \text{ for all } \delta > 0.
\end{aligned}$$

(A5) $(\partial^2/\partial x\partial\theta_j)f^{-1}(x)$ is continuous at all (x,θ), $x \neq 0$, and

$$\frac{\partial^2}{\partial x\partial\theta_j}f^{-1}(x) = O(|x|^{d(\theta)-1-\delta}), 1 \leq j \leq p.$$

Verification of (A5): Note that

$$\frac{\partial^2}{\partial x\partial\theta_j}f^{-1}(x) = 2f^{-3}(x)\frac{\partial f(x)}{\partial x}\frac{\partial f(x)}{\partial\theta_j} - f^{-2}(x)\frac{\partial^2}{\partial x\partial\theta_j}f(x),$$

and

$$\begin{aligned}
\frac{\partial^2}{\partial x\partial H}f(x) &= (1-2H)g'(H)|x|^{1-2H}L(x) - 2g(H)|x|^{-2H}L(x) \\
&\quad -2(1-2H)g(H)|x|^{-2H}L(x) + g'(H)|x|^{1-2H}L(x) \\
&\quad -2g(H)|x|^{-2H}\dot{L}(x)\log|x| + \sum_{k\neq 0}\frac{\partial^2}{\partial x\partial H}f_c(x+2k\pi).
\end{aligned}$$

(A6) $(\partial^3/\partial^2 x\partial\theta_j)f^{-1}(x)$ is continuous at all (x,θ), $x \neq 0$, and

$$\frac{\partial^3}{\partial^2 x\partial\theta_j}f^{-1}(x) = O(|x|^{d(\theta)-2-\delta}), 1 \leq j \leq p.$$

Verification of (A6): similar to that of (A5) and hence omitted.

(A7) $(\partial/\partial x)f^{-1}(x)$ and $(\partial^2/\partial x^2)f^{-1}(x)$ are continuous at all (x,θ), $x \neq 0$, and

$$\left(\frac{\partial}{\partial x}\right)^k f^{-1}(x) = O(|x|^{d(\theta)-k-\delta}), 0 \leq k \leq 2.$$

Verification of (A7): The results follow from the fact that

$$\begin{aligned}\frac{\partial}{\partial x} f^{-1}(x) &= -2 f^{-2}(x) \frac{\partial f(x)}{\partial x}, \\ \frac{\partial^2}{\partial x^2} f^{-1}(x) &= 2 f^{-3}(x) \left(\frac{\partial f(x)}{\partial x} \right)^2 - f^{-2}(x) \frac{\partial^2 f(x)}{\partial x^2},\end{aligned}$$

where

$$\begin{aligned}\frac{\partial}{\partial x} f(x) &= g(H)(1-2H)|x|^{-2H} L(x) + g(H)|x|^{1-2H} \dot{L}(x) + \sum_{k \neq 0} \frac{\partial}{\partial x} f_c(x + 2k\pi) \\ &\approx g(H)(1-2H) L(0)|x|^{-2H}, \text{ for } x \longrightarrow 0,\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial x^2} f(x) &= -2H(1-2H) g(H)|x|^{-2H-1} L(x) + (1-2H) g(H)|x|^{-2H} \dot{L}(x) \\ &\quad + (1-2H) g(H)|x|^{-2H} \dot{L}(x) + g(H)|x|^{1-2H} \ddot{L}(x) \\ &\quad + \sum_{k \neq 0} \frac{\partial^2}{\partial x^2} f_c(x + 2k\pi) \\ &\approx -2H(1-2H) g(H) L(0)|x|^{-2H-1}, \text{ for } x \longrightarrow 0.\end{aligned}$$

Acknowledgement

We thank Academia Sinica, the University of Iowa and the National Science Council (NSC 91-2118-M-001-011), R.O.C. for partial support.

References

[1] Baillie, R. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, **73**, 5-59.

[2] Beran, J. (1994). *Statistics for Long-Memory Processes.* New York: Chapman and Hall.

[3] Bloomfield, P. (1992). Trends in global temperature. *Climatic Change* **21**, 1-16.

[4] Brockwell, P. J. (1993). Threshold ARMA processes in continuous time, 170-90 in *Dimension Estimation and Models*, ed. by H. Tong, World Scientific Publishing, River Edge, NJ.

[5] Chan, N.H. and Palma. W. (1998). State-space modelling of long-memory processes. *Ann. Statist.*, **26**, 719-40.

[6] Comte, F. (1996). Simulation and estimation of long memory continuous time models. *Journal of Time Series Analysis*, **17**, 19-36.

[7] Comte, F. and Renault, E. (1996). Long memory continuous time models. *Journal of Econometrics*, **73**, no. 1, 101-149.

[8] Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Stat.* **17**, no. 4, 1749-66.

[9] Davies, R.B. and Harte, D.S. (1987). Tests for Hurst effect. *Biometrika*, **74**, no. 1, 95-101.

[10] Duncan, Tyrone E. (1999). Some applications of fractional Brownian motion to linear systems. In *System theory: modeling, analysis and control*, 97–105, Cambridge, MA.

[11] Duncan, Tyrone E., Hu, Yaozhong and Pasik-Duncan Bozenna (2000). Stochastic calculus for fractional Brownian motion. I. Theory. *SIAM J. Control Optim.* **38**, no. 2, 582–612.

[12] Efron, B. and Tibshirani R. J. (1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

[13] Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.*, **4**, no. 4, 221-38.

[14] Granger, C. W. J. and Joyeux, R. (1980). An introduction to long memory time series models. *J. Time Ser. Anal.*, **1**, no. 1, 15-29.

[15] Karatzas, I. and Shreve, S.E. (1991). *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York.

[16] Mandelbrot, B.B. and Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, **10**, no. 4, 422-37.

[17] Percival, Donald B. and Walden, Andrew T. (2000). *Wavelet Methods for Time Series Analysis*. New York: Cambridge University Press.

[18] Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press.

[19] Ray, B. K. and Tsay, R. S. (1997). Bandwidth selection for kernel regression with long-range dependent errors. *Biometrika*, **84**, 791-802.

[20] Robinson, P. M. (1993). Time series with strong dependency. In *Advances in Econometrics, 6th World Congress*, Ed. C A. Sims, 47-95. Cambridge: Cambridge University Press.

[21] Robinson, P.M. (1995). Log-periodogram regression of time series with long range dependence. *Ann. Stat.* **23**, 1048-72.

[22] Sowell, F. (1992). Maximum likelihood of stationary univariate fractionally integrated time series. *J. Econometrics*, **53**, 165-88.

[23] Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.*, **2**, 423-34.