Maximum likelihood estimation of linear continuous-time long-memory processes with discrete-time data

Henghsiu Tsai†

Academia Sinica, Taipei, Republic of China

K. S. Chan

University of Iowa, Iowa City, U.S.A.

Revised, May, 2005

Summary. We develop a new class of Continuous-time Auto-Regressive Fractionally Integrated Moving-Average (CARFIMA) models which are useful for modelling regularly-spaced and irregularly-spaced discrete-time long-memory data. We derive the autocovariance function of a stationary CARFIMA model, and study maximum likelihood estimation of a regression model with CARFIMA errors, based on discrete-time data and via the innovations algorithm. It is shown that the maximum likelihood estimator is asymptotically normal, and its finite-sample properties are studied through simulation. The efficacy of the proposed approach is demonstrated with a dataset from an environmental study.

Keywords: CARFIMA models; fractional Brownian motion; innovations algorithm; irregularly-spaced data; polynomial trend.

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), Sowell (1992), Robinson (1993), Beran (1994), Baillie (1996) and Ray and Tsay (1997). A well-known class of discrete-time long-memory processes are the

†*Address for correspondence*: Henghsiu Tsai, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan 115, R.O.C.
E-mail: htsai@stat.sinica.edu.tw

autoregressive fractionally integrated moving average (ARFIMA) models; see Granger and Joyeux (1980) and Hosking (1981). Maximum likelihood estimation and forecasting of the ARFIMA models with missing values have been considered by Palma and Chan (1997) and Palma and Del Pino (1999).

For irregularly-spaced time series data, it is often more convenient to analyze the data by assuming that they are sampled from an underlying continuous-time process. The underlying continuous-time process may be modeled as driven by some stochastic differential equations, e.g. the linear continuous-time autoregressive moving average (ARMA) models. Continuous-time long memory modelling have been studied by a number of authors. Comte and Renault (1996) developed a general class of linear continuous time processes that exhibit long range dependence. Viano *et al.* (1994) studied the probabilistic properties of a class of continuous-time fractional ARMA processes, but the estimation problem is not considered. Chambers (1996) considered the estimation of the long memory parameter of a continuous-time fractional ARMA process with discrete-time data, under the assumption of model identifiability. Comte (1996) studied two estimation methods for a first-order fractional stochastic differential equation (SDE), namely the Whittle likelihood method and the semi-parametric approach, with regularly spaced data. These methods, however, can not be directly applied to irregularly-spaced data. While Comte (1996) mentioned that extension to general SDEs of higher order is straightforward, this extension, however, has not been explicitly reported in the literature. Recently, Brockwell and Marquardt (2005) introduced the fractionally integrated Lévy-driven continuous-time ARMA (CARMA) process, which is obtained by fractional integration of the kernel of the CARMA process. Their model is useful for modelling time series which exhibit both heavy-tailed and long-memory behaviour. However, statistical inference of such models has not been considered by Brockwell and Marquardt (2005).

In this paper, we develop the continuous-time autoregressive fractionally integrated moving average (CARFIMA) models of general order. Our model is based on the stochastic calculus for fractional Brownian motions developed by Duncan *et al.* (2000). Duncan *et al.* (2000) defined the Ito integral for a family of integrands so that the integral has zero mean and admits an explicit expression for the second moment. Their approach makes use of the Wick product and a derivative in the path sense; however, for deterministic integrands, the integral can be defined in terms of the limit of Riemann sums. Duncan *et al.* (2000)

derived an Ito formula, useful for solving SDEs. More importantly, the fact that their stochastic calculus framework admits an Ito formula and an analogue of the Girsanov theorem (see e.g. Valkeila, 1999, Elliott and van der Hoek, 2003, and Hu and Øksendal, 2003) holds the potential for developing nonlinear fractionally integrated processes, e.g. extending the continuous-time threshold ARMA processes (Stramer *et al.*, 1996) to continuous-time threshold fractionally integrated ARMA processes. See Øksendal (2003) for a survey of other variants of stochastic calculus with a fractional Brownian integrator. If the driving Lévy process is the standard Brownian motion, Brockwell and Marquardt's model is essentially same as ours. Thus, their model can be considered as an extension of ours. However, unlike the stochastic calculus framework fundamental to our model specification, Brockwell and Marquardt's approach is restricted to linear processes and cannot be easily extended to nonlinear processes.

Our new class of models is useful for modelling regularly or irregularly spaced discrete-time long-memory data. We study the use of the innovations algorithm for maximum likelihood estimation of the continuous-time long-memory models with (possibly) irregularly-spaced discrete-time data. The rest of the paper is organized as follows. In Section 2, we introduce the CARFIMA models. Implementation and large sample properties of the maximum likelihood estimator are discussed in Section 3. In Section 4, we study the finite sample performance of the maximum likelihood estimator. The new approach is illustrated by a real application in Section 5. Finally, conclusions are presented in Section 6.

2. Continuous-time fractionally integrated ARMA processes

Heuristically, a CARFIMA(p,H,q) process $\{Y_t\}$ is defined as 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 (if they existed) up to and including order $0 \leq q < p$. Specifically, for $t \geq 0$,

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

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

We now recall the definition of fractional Brownian motions. Let $0 < H < 1$ be a fixed number. It is well known (see, e.g., Duncan *et al.*, 2000) that there exists a Gaussian stochastic process $\{B_t^H, t \geq 0\}$ satisfying the following three properties, namely (i) with the initial condition $B_0^H = 0$, (ii) of zero mean, i.e. $E(B_t^H) = 0$ for all $t \geq 0$, and (iii) with the covariance kernel defined as

$$E(B_t^H B_s^H) = \frac{1}{2}\{|t|^{2H} + |s|^{2H} - |t-s|^{2H}\}, \tag{2}$$

for all $s, t \geq 0$. The Gaussian process $\{B_t^H\}$ is called the (standard) fractional Brownian motion with Hurst parameter H. The standard fractional Brownian motion with $H = 1/2$ equals the standard Brownian motion. On the other hand, for $H > 1/2$, increments of the standard fractional Brownian motion, $\{B_t^H - B_{t-1}^H, t = 1, 2, \cdots\}$, exhibits long-range dependence in the sense that its spectral density function is asymptotically linear on the log-log scale (equivalently, the autocovariances are not summable); see (8) below. Henceforth, the Hurst parameter H is fixed and $1/2 < H < 1$.

The fractional Brownian motion is nowhere differentiable (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:

$$Y_t = \beta^{'} X_t, \qquad t \geq 0, \tag{3}$$

$$dX_t = (AX_t + \alpha_0 \delta_p)dt + \sigma \delta_p dB_t^H, \tag{4}$$

where superscript $'$ denotes 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 $j > q$. The model can be made more useful by an extension that incorporates covariates; see Section 3. The stochastic differential equation defined by (4) is interpreted via the framework developed by Duncan *et al.* (2000) who used Wick products to define stochastic integrals with a fractional Brownian integrator, in the Ito sense.

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 (4) with the initial condition X_0. The solution of (4) can be shown to be

$$X_t = e^{At} X_0 + \alpha_0 \int_0^t e^{A(t-u)} \delta_p du + \sigma \int_0^t e^{A(t-u)} \delta_p dB_u^H, \tag{5}$$

where $e^{At} = I_p + \sum_{n=1}^{\infty} \{(At)^n (n!)^{-1}\}$, and I_p is the identity matrix. For a random initial X_0, the mean vector of $\{X_t\}$, denoted by $\mu_{X,t}$, satisfies the equation:

$$\mu_{X,t} = e^{At} \mu_{X,0} + \frac{\alpha_0}{\alpha_1}(e^{At} - I)\delta_1, \tag{6}$$

where $\delta_1 = [1, 0, \cdots, 0]'$. If $\mu_{X,0}$ is chosen to be $-(\alpha_0/\alpha_1)\delta_1$, then $\mu_{X,t}$ becomes $-(\alpha_0/\alpha_1)\delta_1$, which is independent of t. If all the eigenvalues of A have negative real parts, it can be readily checked that (5) implies that as $t \to \infty$, X_t converges in distribution to $N(-(\alpha_0/\alpha_1)\delta_1, V_X)$ where $V_X = H(2H-1)\sigma^2 \int_0^\infty \int_0^\infty e^{Au} \delta_p \delta_p' e^{A' v} |u-v|^{2H-2} du dv$ so that the stationary solution, if it exists, must be Gaussian. The stationary CARFIMA process defined over non-negative t can be extended so that it is a stationary process over all real t, see Appendix 1. The stationarity and the autocovariance function of the stationary CARFIMA process are summarized in the following theorem. (Below, $C_H = H(2H-1)$.)

THEOREM 1. *Let* $1/2 < H < 1$.

(a) Equation (1) with a deterministic initial condition admits an asymptotically stationary solution if and only if all the eigenvalues of A have negative real parts. Moreover, under the preceding eigenvalue condition on A and assuming the solution is stationary, Y_0 and $\{B_t^H, t \geq 0\}$ are jointly Gaussian with the covariances given by

$$cov(Y_0, B_t^H) = H\sigma\beta' \int_0^\infty e^{Au} \delta_p \{(u+t)^{2H-1} - u^{2H-1}\} du, \tag{7}$$

and the mean function of the stationary process equals $\mu_Y = -\alpha_0/\alpha_1$.

(b) Under the stationarity condition, the autocovariance function of $\{Y_t\}$ equals

$$\gamma_Y(h) := cov(Y_{t+h}, Y_t) = C_H \sigma^2 \beta' e^{Ah} \int_0^\infty \int_{-h}^\infty e^{Au} \delta_p \delta_p' e^{A'v} \beta |u-v|^{2H-2} du dv.$$

(c) If $h > 0$, the autocovariance matrix $\gamma_Y(h)$ in (b) can be written as

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

where $V^* = \sigma^2 \int_0^\infty e^{Au}\delta_p\delta_p' e^{A' u} du.$

It can be verified that for a random initial condition Y_0 that may be correlated with the fractional Brownian innovation process, the sufficiency part of the theorem stated in (a) continues to hold if Y_0 has finite variance. Furthermore, part (a) of the preceding theorem implies that under stationarity, Y_0 and the fractional Brownian innovations $B_t^H, t \geq 0$, are generally correlated, in contrast to the case that when $H = 1/2$, the stationary distribution of Y_0 is independent of the standard Brownian motion. The spectral density function of $\{Y_t, t \geq 0\}$ can be shown to be

$$h_Y(\omega) \quad = \quad \frac{\sigma^2}{2\pi}\Gamma(2H+1)\sin(\pi H)|\omega|^{1-2H}\frac{|\beta(i\omega)|^2}{|\alpha(i\omega)|^2}, \qquad -\infty < \omega < \infty, \tag{8}$$

where $\Gamma()$ is the Gamma function, $\alpha(z) = z^p - \alpha_p z^{p-1} - \cdots - \alpha_1$ and $\beta(z) = 1 + \beta_1 z + \beta_2 z^2 + \cdots + \beta_q z^q$; see Tsai and Chan (2005) for a proof.

3. Maximum likelihood estimator and its large sample properties

Let $Y = \{Y_{t_i}\}_{i=1}^N$ be a time series sampled with possibly unequal time intervals. For simplicity, we assume $\alpha_0 = \mu_Y = 0$. This assumption is justified if the data are pre-processed with mean deletion before carrying out the maximum likelihood estimation. For example, the stationary mean can be estimated by the sample mean. Alternatively, under the stationarity assumption, we can incorporate μ_Y as a parameter in the likelihood calculation obtained with Y_{t_i} below replaced by $Y_{t_i} - \mu_Y$. The log likelihood function of Y equals

$$l_Y(\theta, \sigma^2) = -\frac{1}{2}\sum_{j=1}^N \frac{(Y_{t_j} - \hat{Y}_{t_j})^2}{v_{j-1}} - \frac{1}{2}\sum_{j=1}^N \log v_{j-1} - \frac{N}{2}\log(2\pi), \tag{9}$$

where $\hat{Y}_{t_j} = E(Y_{t_j}|Y_{t_i}, i \leq j-1)$, $\tilde{Y}_j = Y_{t_j} - \hat{Y}_{t_j}$ and $v_{j-1} = \mathrm{var}(\tilde{Y}_j|Y_{t_i}, i \leq j-1)$. The predictive means and variances can be computed by the innovations algorithm (see Proposition 5.2.2 of Brockwell and Davis, 1991), namely

$$\hat{Y}_{t_{n+1}} = \begin{cases} 0 & if\ n = 0, \\ \sum_{j=1}^n \phi_{n,j}(Y_{t_{n+1-j}} - \hat{Y}_{t_{n+1-j}}) & if\ n \geq 1, \end{cases} \tag{10}$$

where $\phi_{n,j}$ are computed recursively by the following formulas:

$$\begin{aligned} \nu_0 &= \gamma_Y(0), \\ \phi_{n,n-k} &= v_k^{-1}\left(\gamma_Y(t_{n+1} - t_{k+1}) - \textstyle\sum_{j=0}^{k-1}\phi_{k,k-j}\phi_{n,n-j}v_j\right), \quad k = 0, \cdots, n-1, \\ v_n &= \gamma_Y(0) - \textstyle\sum_{k=0}^{n-1}\phi_{n,n-k}^2 v_k. \end{aligned} \tag{11}$$

Note that $\{\gamma_Y(h)\}$ is the autocovariance function of $\{Y_t, t \geq 0\}$ which can be computed by a numerically stable recursive procedure detailed in Appendix 3. With $v_j^* = v_j/\sigma^2$, the log likelihood function (9) can be rewritten as

$$l_Y(\theta, \sigma^2) = -\frac{1}{2}\sum_{j=1}^{N} \frac{(Y_{t_j} - \hat{Y}_{t_j})^2}{\sigma^2 v_{j-1}^*} - \frac{1}{2}\sum_{j=1}^{N} \log(\sigma^2 v_{j-1}^*) - \frac{N}{2}\log(2\pi). \tag{12}$$

Differentiating (12) with respect to σ^2 and equating to zero gives

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{j=1}^{N} \frac{(Y_{t_j} - \hat{Y}_{t_j})^2}{v_{j-1}^*}. \tag{13}$$

Upon substituting this equation into (12), minus twice the objective function becomes

$$-2l_Y(\theta) = \sum_{j=1}^{N} \log v_{j-1}^* + N\log\left(\sum_{j=1}^{N} \frac{(Y_{t_j} - \hat{Y}_{t_j})^2}{v_{j-1}^*}\right) + N\{1 - \log N + \log(2\pi)\}, \tag{14}$$

which is then minimized with respect to θ to get the maximum likelihood estimate $\hat{\theta}$. The parameter estimate $\hat{\sigma}^2$ is then calculated by (13).

We now extend the above procedure to the case when the model includes d-dimensional covariates $\{W_{t_i}\}$ as follows.

$$Y_{t_i} = \gamma' W_{t_i} + X_{t_i}^{(0)} \quad (i = 1, ..., N), \tag{15}$$

$$X_t^{(p)} - \alpha_p X_t^{(p-1)} - \cdots - \alpha_1 X_t^{(0)} = \sigma\{B_{t,H}^{(1)} + \beta_1 B_{t,H}^{(2)} + \cdots + \beta_q B_{t,H}^{(q+1)}\}, \tag{16}$$

where γ is the vector of regression coefficients. Maximum likelihood estimation can be done by adapting the above procedure with equation (10) replaced by

$$\hat{Y}_{t_{n+1}} = \begin{cases} \gamma' W_{t_{n+1}} & if\ n = 0, \\ \gamma' W_{t_{n+1}} + \sum_{j=1}^{n} \phi_{n,j}(Y_{t_{n+1-j}} - \hat{Y}_{t_{n+1-j}}) & if\ n \geq 1; \end{cases} \tag{17}$$

c.f. So (1999).

For large sample properties of the estimator, we first consider the simpler case of regularly spaced time series data sampled from a stationary CARFIMA(p,H,q) process with a polynomial trend of known degree, i.e., W_t in (15) equals $(1, t, \cdots, t^{d-1})'$ with a fixed d. Dahlhaus (1989) gave a set of regularity conditions under which the maximum likelihood estimator of a discrete-time stationary long memory model is $\sqrt{N}$-consistent, asymptotically normal and asymptotically efficient. Dahlhaus (2004), furthermore, relaxes condition

(A9) in Dahlhaus(1989) to (A9'): α is assumed to be continuous. (The notation α is defined in Dahlhaus(1989, 2004). Here, (A9') amounts to assuming that H is a continuous function of the parameter, which holds trivially.) For simplicity, in this section, let $\theta = (\alpha_1, ..., \alpha_p, \beta_1, \cdots, \beta_q, H, \sigma)$ and denote the regression coefficient by γ.

THEOREM 2. *Let $Y = \{Y_{t_i}\}_{i=1}^N$ be sampled from a model consisting of a polynomial trend of degree $d-1$ with stationary CARFIMA(p,H,q) noise process, where $t_i = ih$ with $h > 0$ being the step size, and $1/2 < H < 1$. Let the maximum likelihood estimator $\hat{\theta} = \hat{\theta}_N \in \Theta$, a compact parameter space, and the true parameter θ^* be in the interior of the parameter space; similarly, denote $\hat{\gamma} = \hat{\gamma}_N$ as the maximum likelihood estimator of γ with the true value being $\gamma^* \in R^d$. Assume that conditions (A0)-(A8) of Dahlhaus (1989) and (A9') are valid. Then, $\hat{\gamma}$ is asymptotically independent of $\hat{\theta}$. Moreover,*

$$N^{1-H} P_N(\hat{\gamma} - \gamma^*) \to N(0, c^2\Lambda^{-1}),$$

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

$$\text{where } P_N = \mathrm{diag}(N^0, N^1, \cdots, N^{d-1}),$$

$$c^2 = \sigma^2\Gamma(2H+1)\sin(\pi H)/(2\pi\alpha_1^2),$$

$$\Lambda = (\Lambda_{ij}), \Lambda_{ij} = \frac{\Gamma(0.5 - H + i)\Gamma(0.5 - H + j)}{\Gamma(1 - 2H + i)\Gamma(1 - 2H + j)(i + j - 2H)}, \text{ and}$$

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

where $\bigtriangledown$ denotes taking the derivative w.r.t. θ, and $f_h()$ is the spectral density of the equally spaced time series $\{Y_{ih}, i \in Z\}$ that is given by

$$f_h(\omega) = \frac{1}{h}\sum_{k \in Z} h_Y\left(\frac{\omega + 2k\pi}{h}\right), \qquad -\pi \le \omega \le \pi. \tag{18}$$

Remarks

1. Conditions (A0)-(A7) of Dahlhaus can be shown to be valid for a stationary CARFIMA model; see Tsai and Chan (2005).

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

3. See Priestley (1981) for a derivation of (18).

4. Often, h can be taken as 1. Note that using different h for the same data set will

change the short-memory parameter estimates, but the estimate of the Hurst parameter H is invariant, owing to self-similarity.

The proof of the theorem is deferred to the appendix. The preceding large-sample properties of the maximum likelihood estimator may be extended to irregularly-spaced data, under suitable regularity conditions. For example, if the sampling intervals are independent and identically distributed positive random variables, then the time series can be made regularly spaced by relabeling Y_{t_i} as Y_i. The corresponding spectral density function can be derived from that of the underlying continuous-time process, see Shapiro and Silverman (1960, equation (19)). As this extension requires non-trivial theoretical and data analysis, we shall not pursue this point further.

4. Simulation studies

We now report some simulation results about the finite sample performance of the maximum likelihood estimator. We have experimented with both regularly and irregularly spaced data from CARFIMA(p,H,0) models with $p = 0, 1$, and 2, and a CARFIMA(2,H,1) model. For regularly spaced data, $Y_{t_i} = X_i^{(0)}, i = 1, 2, ..., N$, were simulated by the method of Davies and Harte's (1987). Irregularly spaced data were sampled as follows. First, we simulated s_i, $i = 1, ..., N$, independently from the exponential distribution with mean equal to 0.5, and recursively set the sampling epochs with $t_0 = 0$, and $t_i = t_{i-1} + s_i + 0.5$, $i = 1, ..., N$; hence the sampling intervals are independent, identically distributed and of unit mean. Second, we simulated the irregularly spaced time series data, $Y_{t_i} = X_{t_i}^{(0)}, i = 1, ..., N$, from the joint stationary multivariate normal distribution induced by the CARFIMA model under study.

For the case of $p = 0$, the data are modeled as fractional Gaussian noises (FGN) defined by $Y_{t_i} = \sigma(B_{t_i}^H - B_{t_{i-1}}^H)$, $i = 1, ..., N$. See Section 5 for more details of the fractional Gaussian noise. All optimizations were done numerically by a constrained optimization procedure, the DBCONF subroutine of the IMSL package which uses a quasi-Newton method with the derivatives approximated by a finite-difference scheme. The H parameter was constrained to be between 0.5 and 1.0, and the α's being negative.

We have tried three sample sizes, $N = 100, 200$ and 400. For each model, the averages and the standard deviations of 1,000 replicates of the estimators are summarized in Tables 1 and 2. In general, the bias of the maximum likelihood estimator of H appears to vanish

Table 1. Averages (standard errors) {empirical coverage rates of the 95% C.I.'s using the asymptotic std. errors} [asymptotic standard errors] of 1,000 simulations of the ML estimators of the parameters

model		p	q		true value	N=100	N=200	N=400
1	regularly	0	0	H	0.75	0.7145 (0.0713)	0.7305 (0.0510)	0.7405 (0.0351)
	spaced					{0.893} [0.0660]	{0.909} [0.0467]	{0.927} [0.0330]
				σ	2	1.9457 (0.1997)	1.9686 (0.1449)	1.9838 (0.1042)
						{0.957} [0.2062]	{0.943} [0.1458]	{0.945} [0.1031]
2	regularly	1	0	H	0.6	0.6162 (0.0885)	0.6078 (0.0737)	0.6048 (0.0560)
	spaced					{0.973} [0.1074]	{0.954} [0.0759]	{0.951} [0.0537]
				α_1	-0.1	-0.1607 (0.0964)	-0.1311 (0.0699)	-0.1170 (0.0513)
						{0.858} [0.0787]	{0.895} [0.0557]	{0.910} [0.0394]
				σ	2	2.0842 (0.2735)	2.0508 (0.2281)	2.0308 (0.2231)
						{0.890} [0.1954]	{0.912} [0.1382]	{0.925} [0.0977]
3	regularly	1	0	H	0.75	0.7455 (0.0835)	0.7468 (0.0709)	0.7503 (0.0541)
	spaced					{0.995} [0.1119]	{0.967} [0.0791]	{0.952} [0.0559]
				α_1	-0.1	-0.1426 (0.0757)	-0.1219 (0.0549)	-0.1134 (0.0437)
						{0.912} [0.0797]	{0.927} [0.0564]	{0.926} [0.0398]
				σ	2	2.0879 (0.3549)	2.0641 (0.3099)	2.0464 (0.2689)
						{0.928} [0.3503]	{0.926} [0.2477]	{0.926} [0.1751]
4	regularly	1	0	H	0.9	0.8631 (0.0628)	0.8749 (0.0560)	0.8871 (0.0427)
	spaced					{0.995} [0.1150]	{0.982} [0.0813]	{0.986} [0.0575]
				α_1	-0.1	-0.1202 (0.0566)	-0.1070 (0.0411)	-0.1043 (0.0337)
						{0.977} [0.0805]	{0.983} [0.0569]	{0.969} [0.0402]
				σ	2	1.9000 (0.4548)	1.9708 (0.4891)	2.0277 (0.4717)
						{0.999} [1.0460]	{0.988} [0.7396]	{0.953} [0.5230]

Table 2. Averages (standard errors) {empirical coverage rates of the 95% C.I.'s using the asymptotic std. errors} [asymptotic standard errors] of 1,000 simulations of the ML estimators of the parameters

model		p	q		true value	N=100	N=200	N=400
5	regularly	2	0	H	0.75	0.6957 (0.0921)	0.7184 (0.0670)	0.7342 (0.0450)
	spaced					{0.883} [0.0875]	{0.899} [0.0619]	{0.932} [0.0437]
				α_1	-0.3	-0.2961 (0.0429)	-0.2970 (0.0299)	-0.2980 (0.0216)
						{0.935} [0.0418]	{0.938} [0.0295]	{0.944} [0.0209]
				α_2	-0.2	-0.1953 (0.0727)	-0.1988 (0.0507)	-0.1999 (0.0366)
						{0.959} [0.0748]	{0.958} [0.0529]	{0.958} [0.0374]
				σ	2	1.9376 (0.2553)	1.9582 (0.1844)	1.9780 (0.1232)
						{0.968} [0.2628]	{0.958} [0.1858]	{0.960} [0.1314]
6	regularly	2	1	H	0.75	0.6359 (0.1166)	0.6679 (0.0999)	0.7033 (0.0850)
	spaced					{1.000} [0.1730]	{0.939} [0.1224]	{0.924} [0.0865]
				α_1	-0.1	-0.0988 (0.0263)	-0.0964 (0.0182)	-0.0973 (0.0142)
						{0.968} [0.0294]	{0.977} [0.0208]	{0.955} [0.0147]
				α_2	-0.2	-0.1845 (0.0824)	-0.1855 (0.0555)	-0.1908 (0.0409)
						{0.965} [0.0852]	{0.977} [0.0603]	{0.969} [0.0426]
				β_1	0.3	0.1622 (0.1825)	0.1907 (0.1721)	0.2370 (0.1406)
						{0.994} [0.2625]	{0.992} [0.1856]	{0.845} [0.1312]
				σ	2	1.9964 (0.2787)	2.0036 (0.1919)	2.0107 (0.1585)
						{0.962} [0.3100]	{0.969} [0.2192]	{0.948} [0.1550]
7	irregularly	2	0	H	0.75	0.6952 (0.0890)	0.7178 (0.0633)	0.7342 (0.0425)
	spaced			α_1	-0.3	-0.2945 (0.0437)	-0.2970 (0.0305)	-0.2981 (0.0214)
				α_2	-0.2	-0.1982 (0.0695)	-0.1982 (0.0526)	-0.2002 (0.0366)
				σ	2	1.9350 (0.2723)	1.9546 (0.1795)	1.9735 (0.1246)

with increasingly large sample size. Similarly, the standard deviation of the estimators generally becomes smaller with larger sample size, "confirming" the consistency results in the previous section. Also note that the results of the CARFIMA(2,H,0) model are about the same for both regularly and irregularly spaced data. For regularly spaced data, we also compute the asymptotic standard errors using Theorem 2, and the empirical coverage rates of the 95% C.I.'s using the asymptotic standard errors. In general, the empirical coverage rates get closer to the nominal coverage rates with increasing sample sizes, except for the MA parameter for the case $N = 400$, which is somewhat lower than the nominal 95%. The latter problem is partly due to the relatively large bias of the MA(1) coefficient. Bias correction via bootstrap may be applied to mitigate the problem. In general, the empirical standard errors also get closer to the asymptotic standard errors with increasing sample sizes, except for σ of models 2 and 3. We have also conducted some other simulation studies with different sampling intervals, the results of which are similar to the reported results.

5. Application

Example: Field values of pH of wet deposition at the McNay Research Station in the Lucas County of Iowa, U.S., were collected on a more or less weekly basis since 1984. The McNay Research Station is one of the two monitoring sites of the Iowa Precipitation Monitoring Program for the National Trends Network. The monitoring program aims at providing an overview of chemical composition of atmospheric deposition in the U.S., see `http://ia.water.usgs.gov/projects/ia005.html`. Besides field values of pH, specific conductance and chemical analysis of the precipitation were recorded. Here, we focus on the pH measurements which measured the acidity of the wet deposition; there are 562 observations, collected from October 1, 1985 to September 18, 2001. For simplicity, we round the sampling times to days. The data are irregularly spaced with the sampling intervals between consecutive observations ranging from 1 to 98 days; the average and median sampling intervals are 10.394 and 7 days, respectively. The unit of time is taken as one day, so $h = 1$. The time series plot of the data is displayed in Fig. 1 which suggests an increasing trend in the data. We have fitted linear trend plus CARFIMA(p,H,0) models to the data, for $p = 0, ..., 4$, with a denoting the intercept and b the slope. For the case

Table 3. AIC and Maximum likelihood estimates of H for model defined by equations (15) and (16)

	p	0	1	2	3	4
original data	AIC	1031.73	927.60	927.91	929.91	931.89
(N=562)	H	0.5824	0.7285	0.7543	0.7543	0.7543
outlier-deleted	AIC	1005.31	880.15	881.47	883.47	885.46
data (N=559)	H	0.5869	0.7107	0.7281	0.7288	0.7288

of $p = 0$, the data are modeled as a linear trend with fractional Gaussian noises (FGN) defined by $Y_{t_i} = a + bt_i + \sigma(B^H_{t_i} - B^H_{t_{i-1}})$, $i = 1, ..., N$. Note that for regularly spaced data of unit sampling interval, the autocovariance function of FGN is given by $\gamma_Y(\tau) = 0.5\sigma^2(|\tau+1|^{2H} - 2|\tau|^{2H} + |\tau-1|^{2H})$, $\tau = ..., -1, 0, 1, ...$ For further discussion of the fractional Gaussian noise, see Beran (1994). For $p \leq 2$, the initial value of the parameter H is set to be 0.75 and those of the parameters α's are chosen to be -1. For $p \geq 3$, the initial value of H is set to be the estimate of H from $p = 2$, and then with H fixed at the initial value, those of the α's set to be the maximizer of the likelihood function with the optimization initiated with α's equal to -1. The initial values of a and b are estimated by regressing Y_{t_i} on t_i assuming the errors to be white noises.

In Table 3, we report for each order p, the maximum likelihood estimate of H and the corresponding Akaike information criterion $AIC = -2(l_Y(\hat{\theta}) - r)$, where r is the number of parameters in the model, and $\hat{\theta}$ is the maximum likelihood estimator of θ. AIC selects the autoregressive order $p = 1$. The parameter estimates are $(\hat{H}, \hat{\alpha}_1, \hat{\sigma}, \hat{a}, \hat{b}) = (0.7285, -0.4013, 0.3759, 4.8588, 9.0448 \times 10^{-5})$. An inspection of the standardized residuals reveals three outliers with values: 3.78, 3.66 and 3.64. These three outliers were then deleted and we re-fitted the models; see Table 3. While AIC still suggests $p = 1$, the difference between $p = 1$ and $p = 2$ in terms of AIC is relatively small; hence we report the model fits for both cases. Residual analysis suggests that both models provide adequate fit to the data. For $p = 1$, $(\hat{H}, \hat{\alpha}_1, \hat{\sigma}, \hat{a}, \hat{b}) = (0.7107, -0.3463, 0.3352, 4.8196, 9.875 \times 10^{-5})$, whereas for $p = 2$, $(\hat{H}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\sigma}, \hat{a}, \hat{b}) = (0.7281, -0.1576, -0.5246, 0.1468, 4.8178, 9.910 \times 10^{-5})$.

We then assess the uncertainty of the maximum likelihood estimator via two methods, namely, by computing the observed Fisher information based on (14) and, then checking the asymptotics by parametric bootstrap with bootstrap size 1,000; see Table 4. The 95%

Table 4. Maximum likelihood estimates of the parameters of the model defined by equations (15) and (16), with $p = 1$ and 2.

p		estimated value	asymptotic 95% C.I.	asymptotic std. err.	bootstrap 95% C.I.	bootstrap std. err.
1	H	0.7107	(0.5751, 0.8463)	0.0678	(0.5000, 0.7903)	0.0760
	α_1	-0.3463	(-0.5791, -0.1135)	0.1164	(-0.6135, -0.1465)	0.1329
	a	4.820	(4.5606, 5.0794)	0.1297	(4.5700, 5.0875)	0.1314
	$\mathrm{b}\times 10^5$	9.875	(2.963 , 16.79)	3.456	(2.906 , 16.77)	3.543
	σ	0.3352			(0.2724, 0.4729)	0.0586
2	H	0.7281	(0.5949, 0.8613)	0.0666	(0.5429, 0.8117)	0.0675
	α_1	-0.1576	(-0.2966, -0.0186)	0.0695	(-0.6916, -0.0990)	0.1808
	α_2	-0.5246	(-1.1078, 0.0586)	0.2916	(-2.5816, -0.2995)	0.6547
	a	4.818	(4.5378, 5.0978)	0.1400	(4.543, 5.093)	0.1381
	$\mathrm{b}\times 10^5$	9.910	(2.575, 17.25)	3.668	(3.209 , 17.22)	3.598
	σ	0.1468			(0.0883, 0.8113)	0.2056

bootstrap confidence intervals are obtained by the percentile method (Chapter 13 of Efron and Tibshirani, 1993) with the 2.5 and 97.5 percentiles of the bootstrap estimates being the end points of the 95% confidence intervals. The parametric bootstrap estimates appear to have skewed distributions, and the bootstrap standard errors and confidence intervals are generally larger than their asymptotic counterparts, except for the regression parameters when $p = 2$. The 95% confidence intervals of H indicate substantial variability. While the bootstrap confidence interval of H includes 0.5 for $p = 1$, its counterpart for $p = 2$ excludes 0.5. The confidence intervals of b suggest that $b > 0$. Hence, we conclude that (i) there is some evidence that the data are of long memory and (ii) the wet deposition became less acidic over time.

The population spectral density function of the CARFIMA noises can be estimated by (8) with the unknown parameters there replaced by the maximum likelihood estimates. For $p = 1$, the spectral density must peak at 0 with a singularity there, but for $p = 2$, the spectral density may admit a secondary peak away from the origin. Indeed for $p = 2$, although the estimated spectral density peaks only at 0 with a singularity there (left diagram in Fig. 2, over the range $0.05 < \omega < 1$), the squared magnitude of the transfer function of the autoregressive filter, i.e. $1/|\alpha(i\omega)|^2$, peaks at $\omega = 0.1414$, with the corresponding period

Table 5. Sample autocorrelations of the parametric bootstrap estimates for $p = 2$, with their counterparts for $p = 1$ enclosed by parentheses.

	α_1	α_2	σ	a	b
H	0.119 (-0.793)	0.353	-0.076 (0.569)	0.040 (-0.003)	0.002 (0.023)
α_1		0.930	-0.105 (-0.939)	0.019 (0.027)	-0.040 (-0.039)
α_2			-0.149	0.040	-0.038
σ				-0.014 (-0.034)	-0.013 (0.037)
a					-0.755 (-0.785)

equal to $2\pi/0.1414 = 44.4$ days, essentially a one and half month pattern. Thus, besides the long-memory component, the pH series has a significant short-memory component.

The sample correlation matrices of the parametric bootstrap estimates are given in Table 5. It can be seen that $(\hat{a}, \hat{b})$ and the CARFIMA parameter estimators appear to be uncorrelated, which is consistent with the asymptotic independence result stated in Theorem 2.

6. Conclusions

We have introduced the CARFIMA models for continuous-time process that is analogous to the fractional ARMA model in the discrete time setting. The short-memory CARMA model relates to the CARFIMA model as its limiting case when $H = 1/2$. So far, H is restricted to lie between 1/2 and 1. An interesting problem is to extend the model to include $H \geq 1$. See Beran (1995) and Ling and Li (1997) for estimation of discrete-time non-stationary long memory processes. The real application illustrates that the CARFIMA model provides a useful framework for studying the temporal dependence structure and the spectrum of irregularly-spaced long-memory data. It is of interest to extend this approach to multiple time series.

Acknowledgements

We are grateful to Professor Rainer Dahlhaus for sending us a preprint of Dahlhaus (2004). We thank two referees, an Associate Editor and the Joint Editor for helpful comments. The authors gratefully acknowledge Academia Sinica, National Science Council (NSC 91-2118-M-001-011), R.O.C., and the National Science Foundation (DMS-0405267) for partial support.

Appendix 1.

The stationary CARFIMA process defined over all real t

The stationary CARFIMA process defined over non-negative t can be extended so that it is a stationary process over all real t. First note that for $H \in (0,1)$, we have the following integral representation of fractional Brownian motion for all real t:

$$B_t^H = D_H \int_{-\infty}^{\infty} \{(t-u)_+^{H-1/2} - (-u)_+^{H-1/2}\} dB_u,$$

where $D_H = [2H\Gamma(3/2-H)/\{\Gamma(H+1/2)\Gamma(2-2H)\}]^{1/2}$, and B_u is a standard Brownian motion, see Taqqu (2003). Then we can show that provided (i) $\int_{-\infty}^{t} e^{A(t-u)} dB_u^H$ exists and (ii) the eigenvalues of A all have negative real parts, the process $\{X_t\}$ defined by

$$X_t = \alpha_0 \int_0^t e^{A(t-u)} \delta_p du + \sigma \int_{-\infty}^{t} e^{A(t-u)} \delta_p dB_u^H \qquad \text{(A1)}$$

is a strictly stationary solution of (4) for $t \in (-\infty, \infty)$ with the corresponding CARFIMA process,

$$Y_t = \alpha_0 \int_0^t \beta' e^{A(t-u)} \delta_p du + \sigma \int_{-\infty}^{t} \beta' e^{A(t-u)} \delta_p dB_u^H.$$

Appendix 2.

Proof of Theorem 1

Proof of (a) We first prove equation (7). For $1/2 < H < 1$ and for $f, g \in L^2(R;R) \bigcap L^1(R;R)$, Gripenberg and Norros (1996) showed that

$$\text{cov}\left(\int_{-\infty}^{\infty} f(u) dB_u^H, \int_{-\infty}^{\infty} g(v) dB_v^H\right) = C_H \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u)g(v)|u-v|^{2H-2} du dv. \qquad \text{(A2)}$$

Note that equations (5) and (A1) are essentially equivalent. Without loss of generality, assume $\alpha_0 = 0$ and write $Y_0 = \sigma \int_{-\infty}^{0} \beta' e^{-Au} \delta_p dB_u^H$ and $B_t^H = \int_0^t dB_u^H$, then the result follows from equation (A2).

Now, we show that if all the eigenvalues of A have negative real parts, then the solution to (4) with a deterministic initial condition is asymptotically stationary. It follows from (6) that $\mu_{X,t}$ tends to $-(\alpha_0/\alpha_1)\delta_1$ under the eigenvalue condition. By (5) and (A2), the variance of X_t equals

$$
\begin{aligned}
V_{X,t} &= \text{var}\left(e^{At}X_0 + \sigma\int_0^t e^{A(t-u)}\delta_p dB_u^H\right) \\
&= e^{At}V_0 e^{A^{'}t} + \text{cov}\left(e^{At}X_0, \sigma\int_0^t e^{A(t-u)}\delta_p dB_u^H\right) + \text{cov}\left(\sigma\int_0^t e^{A(t-u)}\delta_p dB_u^H, e^{At}X_0\right) \\
&\quad + C_H\sigma^2\int_0^t\int_0^t e^{A(t-u)}\delta_p\delta_p^{'}e^{A^{'}(t-v)}|u-v|^{2H-2}dudv. \qquad \text{(A3)}
\end{aligned}
$$

Assume that all eigenvalues of A have negative real parts. Then

$$
\begin{aligned}
\lim_{t\longrightarrow\infty} V_{X,t} &= C_H\sigma^2\lim_{t\longrightarrow\infty}\int_0^t\int_0^t e^{A(t-u)}\delta_p\delta_p^{'}e^{A^{'}(t-v)}|u-v|^{2H-2}dudv \\
&= C_H\sigma^2\int_0^\infty\int_0^\infty e^{Au}\delta_p\delta_p^{'}e^{A^{'}v}|u-v|^{2H-2}dudv, \qquad \text{(A4)}
\end{aligned}
$$

where the integral is finite and denoted by V_X. Similarly, the covariance kernel can be shown to be asymptotically a function of $|s-t|$. Therefore, the process $\{X_t, t\geq 0\}$ is asymptotically stationary if all the eigenvalues of A have negative real parts.

We now prove the necessity part. Let $t>0$ be fixed. Suppose that the process is asymptotically stationary and let X_0 be random. Then, by equation (7), it can be seen that

$$
\text{cov}(X_0, B_t^H) \approx H\sigma\int_0^\infty e^{Au}\delta_p\{(u+t)^{2H-1}-u^{2H-1}\}du, \qquad \text{(A5)}
$$

showing that (A5) is finite. This implies that the real part of all the eigenvalues of A must be negative. Indeed, it follows from the finiteness of (A5) that the series $\sum_{n=0}^\infty c_n$ is finite where $c_n = \int_n^{n+1}\delta_p' e^{Au}\delta_p\{(u+t)^{2H-1}-u^{2H-1}\}du$. By the mean value theorem, $\int_n^{n+1}\delta_p' e^{Au}\delta_p\{(u+t)^{2H-1}-u^{2H-1}\}du = \delta_p' e^{Au_n}\delta_p\int_n^{n+1}\{(u+t)^{2H-1}-u^{2H-1}\}du$, for some $n<u_n<n+1$. But $t(n+1+t)^{2H-2}/(2H-1)\leq\int_n^{n+1}\{(u+t)^{2H-1}-u^{2H-1}\}du\leq tn^{2H-2}/(2H-1)$. For example, the right side of the preceding inequality can be seen as follows: $\int_n^{n+1}\{(u+t)^{2H-1}-u^{2H-1}\}du = \int_n^{n+1}\int_0^t(u+v)^{2H-2}dvdu/(2H-1)\leq\int_0^t(n+v)^{2H-2}dv/(2H-1)\leq tn^{2H-2}/(2H-1)$. It can be readily checked that $\{A^i\delta_p, i=0,\cdots,p-1\}$ forms a basis of R^p. Hence, δ_p does not belong to any subspace invariant with respect to A that is a proper subset of R^p. This implies that all the eigenvalues of A must have negative real part in order for c_n to be summable. We show this by first assuming that A is diagonalizable so that $A=PDP^{-1}$ where D is a diagonal matrix with diagonal elements d_i and the ith column of P is the associated eigenvector. Then, $e^{Au}=Pe^{uD}P^{-1}$. Because δ_p does not belong to any invariant subspace except R^p, all entries of $P^{-1}\delta_p$ are non-zero.

Similarly, all entries of $\delta_p' P$ are non-zero as δ_p is not orthogonal to any eigenvector, otherwise we have a contradiction that it belongs to the invariant space that is orthogonal to some eigenvectors. For the case of non-diagonalizable A, we can argue similarly using the Jordan canonical form of A.

That the marginal stationary distribution is Gaussian follows from (5).

Proof of (b) Equations (5) and (A1) are essentially equivalent, so the result follows from (A1) and (A2).

Proof of (c) This follows from calculus and the definition of V^*.

Appendix 3.

A recursive procedure for computing the autocovariances

First write $\gamma_Y(h) = C_H \beta' \{L_1(h) + L_2(h) + e^{Ah} L_2(0)\} V^* \beta$, where $L_1(h) = \int_0^h e^{A(h-u)} u^{2H-2} du$, and $L_2(h) = \int_h^\infty e^{A(u-h)} u^{2H-2} du$. Assume that A is diagonalizable so that $A = PDP^{-1}$ and $e^{tA} = PE_t P^{-1}$, where P is the $p \times p$ matrix whose jth column is a right eigenvector corresponding to d_j, $D = \text{diag}(d_1, \cdots, d_p)$ and $E_t = \text{diag}(e^{d_1 t}, \cdots, e^{d_p t})$. Then we have $\gamma_Y(h) = C_H \beta' P \{M_1(h) + M_2(h) + E_h M_2(0)\} P^{-1} V^* \beta$, where $M_1(h) = \int_0^h E_{h-u} u^{2H-2} du$, and $M_2(h) = \int_h^\infty E_{u-h} u^{2H-2} du$. Note that the M_i's , $i = 1, 2$, are $p \times p$ diagonal matrices. Suppose we need to compute $\gamma_Y(h_j)$, for $j = 0, ..., r$, where $0 = h_0 < h_1 < \cdots < h_r$ is an increasing sequence. Then we need to compute $g_1(h_i, d_j) := \int_0^{h_i} e^{d_j(h_i - u)} u^{2H-2} du$ and $g_2(h_i, d_j) := \int_{h_i}^\infty e^{d_j(u-h_i)} u^{2H-2} du$, $i = 1, \cdots, r, j = 1, \cdots, p$, which we compute recursively as follows. Let $g_1(0, d_j) = 0$ for all $j = 1, \cdots, p$, then for fix j,

$$g_1(h_{i+1}, d_j) = e^{d_j(h_{i+1}-h_i)} g_1(h_i, d_j) + \int_{h_i}^{h_{i+1}} e^{d_j(h_{i+1}-u)} u^{2H-2} du, \qquad i = 0, \cdots, r-1,$$

and

$$g_2(h_i, d_j) = e^{d_j(h_{i+1}-h_i)} g_2(h_{i+1}, d_j) + \int_{h_i}^{h_{i+1}} e^{d_j(u-h_i)} u^{2H-2} du, \qquad i = 0, \cdots, r-1,$$

with $g_2(h_r, d_j) = \int_{h_r}^\infty e^{d_j(u-h_r)} u^{2H-2} du$, for all j. In practice, we set the integration limits of $L_2(h)$, to be from h_r to M, where M is some large number, say $100h_r$. We have tried different M in the program and the results are very robust to the choice of M.

Appendix 4.

Proof of Theorem 2

We first show that the maximum likelihood estimator of θ is asymptotically equal to that when γ is known. Write the true model in vector form as $Y = W\gamma^* + X$, where $Y = (Y_1, Y_2, \cdots, Y_N)'$, W the design matrix and $X = (X_1, X_2, \cdots, X_N)'$. Let $\mathcal{L}_N(\theta)$ be twice the negative profile log-likelihood of θ that is normalized by the sample size, and Σ_θ be the covariance matrix of Y induced by the stationary CARFIMA noises. For a fixed θ, the maximum likelihood estimator of γ can be readily seen to be $\hat{\gamma}(\theta) = (W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}Y$. Hence, up to an additive constant, we have

$$
\begin{aligned}
& \mathcal{L}_N(\theta) \\
= \; & \sup_\gamma \{\log(|\Sigma_\theta|)/N + (Y - W\gamma)'\Sigma_\theta^{-1}(Y - W\gamma)/N\} \\
= \; & \log(|\Sigma_\theta|)/N + Y'\{I - W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}\}'\Sigma_\theta^{-1}\{I - W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}\}Y/N \\
= \; & \log(|\Sigma_\theta|)/N + (W\gamma^* + X)'\{I - W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}\}' \\
& \qquad \Sigma_\theta^{-1}\{I - W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}\}(W\gamma^* + X)/N \\
= \; & \log(|\Sigma_\theta|)/N + X'\{I - W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}\}'\Sigma_\theta^{-1}\{I - W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}\}X/N \\
= \; & \{\log(|\Sigma_\theta|) + X'\Sigma_\theta^{-1}X\}/N \\
& \quad +\{-X'\Sigma_\theta^{-1}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}X - X'\Sigma_\theta^{-1}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}X \\
& \quad +X'\Sigma_\theta^{-1}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}X\}/N \\
= \; & \{\log(|\Sigma_\theta|) + X'\Sigma_\theta^{-1}X\}/N - X'\Sigma_\theta^{-1}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}X/N \\
= \; & \mathcal{L}_{1,N}(\theta) - \mathcal{L}_{2,N}(\theta). \qquad \text{(A6)}
\end{aligned}
$$

Note that $\mathcal{L}_{1,N}(\theta)$ equals twice the negative log-likelihood of θ when γ is known, and it is known that $\mathcal{L}_{1,N}(\theta) - \mathcal{L}_{1,N}(\theta^*) = O_p(1)$; see Dahlhaus (1989). We shall outline the proof that $\mathcal{L}_{2,N}(\theta)$ is uniformly $O_p(N^{-1+\delta})$ for some $0 < \delta < 1$, thus showing that twice the negative profile log-likelihood of θ asymptotically equals that when γ is known. To see this, let $|A| = \{\text{trace}(AA')\}^{1/2}$ and $\|A\|$ be the spectral norm of any square matrix A and noting the inequality $|AB| \leq |A| \, \|B\|$ (p. 1754 of Dahlhaus, 1989), we have

$$
\begin{aligned}
& E\{\mathcal{L}_{2,N}(\theta)\} \\
= \; & E[\text{trace}\{(W'\Sigma_\theta^{-1}W)^{-1/2}W'\Sigma_\theta^{-1}XX'\Sigma_\theta^{-1}W(W'\Sigma_\theta^{-1}W)^{-1/2}\}]/N \\
= \; & \text{trace}\{(W'\Sigma_\theta^{-1}W)^{-1/2}W'\Sigma_\theta^{-1}\Sigma_{\theta^*}\Sigma_\theta^{-1}W(W\Sigma_\theta^{-1}W)^{-1/2}\}/N
\end{aligned}
$$

$$
\begin{aligned}
&= \left\| (W'\Sigma_\theta^{-1}W)^{-1/2}W'\Sigma_\theta^{-1}\Sigma_{\theta^*}^{1/2} \right\|^2 /N \\
&\leq \left\| (W'\Sigma_\theta^{-1}W)^{-1/2}W'\Sigma_\theta^{-1/2} \right\|^2 \|\Sigma_\theta^{-1/2}\Sigma_{\theta^*}^{1/2}\|^2/N \\
&= d\|\Sigma_\theta^{-1/2}\Sigma_{\theta^*}^{1/2}\|^2/N,
\end{aligned}
$$

which can be shown to be $O(N^{-1+\delta})$ for some $0 < \delta < 1$; recall $d-1$ is the degree of the polynomial trend. This follows from the fact that as $N \to \infty$, the spectral norm of $\Sigma_\theta^{-1/2}\Sigma_{\theta^*}^{1/2}$ is uniformly bounded by $O(N^{\delta/2})$ for some $0 < \delta < 1$, by Lemma 5.3 of Dahlhaus (1989) and the compactness of Θ. Because $\mathcal{L}_{2,N}(\theta)$ is non-negative, it follows from Markov inequality that, for each θ,

$$
\mathcal{L}_{2,N}(\theta) = O_p(N^{-1+\delta}). \tag{A7}
$$

Moreover, this result can be strengthened to showing that $\mathcal{L}_{2,N}(\theta)$ is uniformly $O_p(N^{-1+\delta})$ for some $0 < \delta < 1$. This we do by adapting the argument of Lemma 6.2 of Dahlhaus (1989). Let $S = N^{-\beta}|\theta^{(1)} - \theta^{(2)}|^{-1}X'(B_{\theta^{(1)}} - B_{\theta^{(2)}})X$ where $\theta^{(i)} \in \Theta, i = 1,2$, $B_\theta = \Sigma_\theta^{-1}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}$ and $0 < \beta < 1$ is some constant to be determined below. Let l be a positive integer and $\text{cum}_\ell(S)$ denote the cumulant $\text{cum}(S, \cdots, S)$ where S appears ℓ times within the parentheses. By an argument of Dahlhaus (1989, p. 1765),

$$
\begin{aligned}
|\text{cum}_\ell(S)| &\leq (\ell-1)!2^{\ell-1}|\theta^{(1)} - \theta^{(2)}|^{-\ell}N^{-\ell\beta}|\text{trace}\{\Sigma_{\theta^*}^{1/2}(B_{\theta^{(1)}} - B_{\theta^{(2)}})\Sigma_{\theta^*}^{1/2}\}^\ell| \\
&\leq (\ell-1)!2^{\ell-1}|\theta^{(1)} - \theta^{(2)}|^{-\ell}N^{-\ell\beta} \left\| \Sigma_{\theta^*}^{1/2}(B_{\theta^{(1)}} - B_{\theta^{(2)}})\Sigma_{\theta^*}^{1/2} \right\|^\ell,
\end{aligned}
$$

where we have made use of the fact that $|\text{trace}(AB)| \leq \| A \| \| B \|$, and $\| AB \| \leq \| A \| \| B \|$. It follows from the mean value theorem that

$$
\left\| \Sigma_{\theta^*}^{1/2}(B_{\theta^{(1)}} - B_{\theta^{(2)}})\Sigma_{\theta^*}^{1/2} \right\| \leq \left\| \Sigma_{\theta^*}^{1/2}\nabla B_{\tilde\theta}\Sigma_{\theta^*}^{1/2} \right\| |\theta^{(1)} - \theta^{(2)}|,
$$

for some $\tilde\theta$ between $\theta^{(1)}$ and $\theta^{(2)}$, and where ∇B_θ is the partial derivative of B_θ w.r.t θ, i.e. it is an array whose jth component is the partial derivative of B_θ w.r.t. θ_j; for any matrices A and C of compatible dimensions, $A\nabla B_\theta C$ is an array whose jth component equals the matrix product of A times $\partial B_\theta/\partial\theta_j$ times C, and $\| A\nabla B_\theta C \|$ is the sum of the Euclidean norms of the component matrices of $A\nabla B_\theta C$. Now,

$$
\begin{aligned}
\nabla B_\theta &= -\Sigma_\theta^{-1}\nabla\Sigma_\theta\Sigma_\theta^{-1}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1} \\
&\quad +\Sigma_\theta^{-1}W(W'\Sigma_\theta^{-1}W)^{-1}(W'\Sigma_\theta^{-1}\nabla\Sigma_\theta\Sigma_\theta^{-1}W)(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}
\end{aligned}
$$

$$
\begin{aligned}
&\quad -\Sigma_\theta^{-1}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}\nabla\Sigma_\theta\Sigma_\theta^{-1} \\
&= -P_{1,\theta} + P_{2,\theta} - P_{3,\theta}.
\end{aligned}
$$

Consider

$$
\begin{aligned}
& \left| \Sigma_{\theta^*}^{1/2} P_{1,\theta} \Sigma_{\theta^*}^{1/2} \right| \\
= & \left| \Sigma_{\theta^*}^{1/2}\Sigma_\theta^{-1}\nabla\Sigma_\theta\Sigma_\theta^{-1}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1}\Sigma_{\theta^*}^{1/2} \right| \\
\le & \|\Sigma_{\theta^*}^{1/2}\Sigma_\theta^{-1/2}\| |||\Sigma_\theta^{-1/2}\nabla\Sigma_\theta\Sigma_\theta^{-1/2}||| \left| \Sigma_\theta^{-1/2}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1/2} \right| \|\Sigma_{\theta^*}^{1/2}\Sigma_\theta^{-1/2}\| \\
\le & \sqrt{d}KN^\beta,
\end{aligned}
$$

uniformly for $\theta \in \Theta$, for some constants K that may vary from occurrence to occurrence, $0 < \beta < 1$, by Lemma 5.3 and the proof of Lemma 5.4 (a) of Dahlhaus (1989); the notation $||| \cdot |||$ is defined as the sum of the spectral norms of the components of the enclosed array. Similarly, it can be shown that $\left| \Sigma_{\theta^*}^{1/2} P_{3,\theta} \Sigma_{\theta^*}^{1/2} \right| \le \sqrt{d}KN^\beta$, and also that

$$
\begin{aligned}
& \left| \Sigma_{\theta^*}^{1/2} P_{2,\theta} \Sigma_{\theta^*}^{1/2} \right| \\
\le & \|\Sigma_{\theta^*}^{1/2}\Sigma_\theta^{-1/2}\| \left| \Sigma_\theta^{-1/2}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1/2} \right| |||\Sigma_\theta^{-1/2}\nabla\Sigma_\theta\Sigma_\theta^{-1/2}||| \\
& \times \left| \Sigma_\theta^{-1/2}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1/2} \right| \|\Sigma_{\theta^*}^{1/2}\Sigma_\theta^{-1/2}\| \\
\le & dKN^\beta.
\end{aligned}
$$

Hence, $\left| \Sigma_{\theta^*}^{1/2} \nabla B_\theta \Sigma_{\theta^*}^{1/2} \right| \le KN^\beta$, uniformly for $\theta \in \Theta$. Hence, there exists a constant K such that for $\ell \ge 2$, $|\mathrm{cum}_\ell(S)| \le (\ell-1)!2^{\ell-1}K^\ell$. Moreover, it can be similarly verified that $|E(S)| \le K$, by enlarging K if necessary. Then the proof of Lemma 6.2 of Dahlhaus (1989) can be adapted to show that $\{N^{\beta-1}\mathcal{L}_{2,N}(\theta), \theta \in \Theta\}$ is equicontinuous. The compactness of Ω and (A7) then imply that $\mathcal{L}_{2,N}(\theta)$ is uniformly $O_p(N^{-1+\delta})$, with δ replaced by β if the latter is larger, where $1 > \delta > 0$. Similarly, it can be verified that $\{\nabla^2\mathcal{L}_{2,N}(\theta), \theta \in \Theta\}$ is equicontinuous. Moreover, it can be checked that for each $\delta > 0$,

$$
N^{-1}|X'\nabla B_{\theta^*}X| = O_p(N^{-1+\delta}) \tag{A8}
$$

$$
N^{-1}|X'\nabla^2 B_{\theta^*}X| = O_p(N^{-1+\delta}). \tag{A9}
$$

Here, $|X'\nabla B_{\theta^*}X|$ is the L^2-norm of the vector consisting of $X'(\partial B_{\theta^*}/\partial\theta_j)X$. To see (A8), consider $E|X'\nabla B_{\theta^*}X| \le \sum_{i=1}^3 E|X'P_{i,\theta^*}X|$ and hence it suffices to show that $E|X'P_{i,\theta^*}X| \le KN^\delta$ for some K and each $\delta > 0$. Below, we make use of the technical result that for any matrices C, H and D of compatible dimensions and H a non-negative definite matrix,

$$E|X'CHDX| \quad \leq \quad \|\Sigma_{\theta^*}^{1/2}C\| \, \Big| H^{1/2} \Big|^2 \|\Sigma_{\theta^*}^{1/2}D'\|, \tag{A10}$$

which follows from the Cauchy-Schwartz inequality. Specifically,

$$\begin{aligned} & E|X'CHDX| \\ \leq \quad & E\{(X'CHC'X)^{1/2}(X'D'HDX)^{1/2}\} \\ \leq \quad & E^{1/2}(X'CHC'X)E^{1/2}(X'D'HDX) \\ = \quad & \text{trace}^{1/2}(\Sigma_{\theta^*}^{1/2}CHC'\Sigma_{\theta^*}^{1/2})\text{trace}^{1/2}(\Sigma_{\theta^*}^{1/2}D'HD\Sigma_{\theta^*}^{1/2}) \\ = \quad & \Big|\Sigma_{\theta^*}^{1/2}CH^{1/2}\Big| \, \Big|\Sigma_{\theta^*}^{1/2}D'H^{1/2}\Big| \\ \leq \quad & \|\Sigma_{\theta^*}^{1/2}C\| \, \Big|H^{1/2}\Big|^2 \|\Sigma_{\theta^*}^{1/2}D'\|. \end{aligned}$$

Applying (A10), we have

$$\begin{aligned} & E|X'P_{1,\theta^*}X| \\ = \quad & E|X'\Sigma_{\theta^*}^{-1}\nabla\Sigma_{\theta^*}\Sigma_{\theta^*}^{-1}W(W'\Sigma_{\theta^*}^{-1}W)^{-1}W'\Sigma_{\theta^*}^{-1}X| \\ \leq \quad & |||\Sigma_{\theta^*}^{-1/2}\nabla\Sigma_{\theta^*}\Sigma_{\theta^*}^{-1/2}||| \, \Big|\{\Sigma_{\theta^*}^{-1/2}W(W'\Sigma_{\theta^*}^{-1}W)^{-1}W'\Sigma_{\theta^*}^{-1/2}\}^{1/2}\Big|^2 \|\Sigma_{\theta^*}^{1/2}\Sigma_{\theta^*}^{-1/2}\| \\ \leq \quad & d|||\Sigma_{\theta^*}^{-1/2}\nabla\Sigma_{\theta^*}\Sigma_{\theta^*}^{-1/2}||| \\ \leq \quad & KN^{\delta}, \end{aligned}$$

for some K and each $\delta > 0$, by the proof of Lemma 5.4 (a) in Dahlhaus (1989). The proofs for $E|X'P_{i,\theta^*}X| \leq KN^{\delta}, i = 2, 3$, are similar and hence omitted. Equation (A9) can be similarly established.

Because of the equicontinuity of $\{\nabla^2\mathcal{L}_{2,N}(\theta), \theta \in \Omega\}$, (A8) and (A9), the asymptotically normal limiting distribution of $\hat{\theta}$ then follows as in the proof of Theorem 3.2 of Dahlhaus (1989) and Dahlhaus (2004).

Let $\hat{\gamma}$ be the maximum likelihood estimator of γ given $\hat{\theta}$ was the true parameter value. We are going to show that the limiting distribution of $\hat{\gamma}$ is as given in Theorem 2. Note that given θ, the maximum likelihood estimator of γ equals $\hat{\gamma}(\theta) = (W'\Sigma_{\theta}^{-1}W)^{-1}W'\Sigma_{\theta}^{-1}Y$. Also, $\hat{\gamma}(\theta)$ is an unbiased estimator of γ because $E(\hat{\gamma}(\theta)) = (W'\Sigma_{\theta}^{-1}W)^{-1}W'\Sigma_{\theta}^{-1}W\gamma^* = \gamma^*$. For two fixed $\theta^{(1)}$ and $\theta^{(2)}$, consider the squared L^2-norm between $N^{1-H}P_N\hat{\gamma}(\theta^{(1)})$ and $N^{1-H}P_N\hat{\gamma}(\theta^{(2)})$ which equals

$$\rho^2(\theta^{(1)}, \theta^{(2)}) \quad = \quad \text{trace}(\text{cov}\{N^{1-H}P_N(\hat{\gamma}(\theta^{(1)}) - \hat{\gamma}(\theta^{(2)}))\}). \tag{A11}$$

Define $C(\theta) = N^{1-H}P_N[(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1} - (W'\Sigma_{\theta^{(2)}}^{-1}W)^{-1}W'\Sigma_{\theta^{(2)}}^{-1}]\Sigma^{1/2}$ where $\Sigma = \Sigma_{\theta^*}$ is the true covariance matrix of Y. Then $\rho^2(\theta^{(1)},\theta^{(2)}) = \sum_{i,j} C_{ij}^2(\theta^{(1)})$. Hence, by Theorem 5.19 of Rudin (1976), there exists an η between $\theta^{(1)}$ and $\theta^{(2)}$ such that $\rho^2(\theta^{(1)},\theta^{(2)}) \leq \sum_{i,j}(\theta^{(1)}-\theta^{(2)})_i \mathrm{trace}\{(\partial C(\eta)/\partial\theta_i)(\partial C(\eta)/\partial\theta_j)'\}(\theta^{(1)}-\theta^{(2)})_j$ where $\partial C/\partial\theta_i$ is the partial derivative of C with respect to the ith component of θ. Now, $\partial C/\partial\theta_i = N^{1-H}P_N(W'\Sigma_\theta^{-1}W)^{-1}\,[W'\Sigma_\theta^{-1}(\partial\Sigma_\theta/\partial\theta_i)\Sigma_\theta^{-1}W(W'\Sigma_\theta^{-1}W)^{-1}W'\Sigma_\theta^{-1} - W'\,\Sigma_\theta^{-1}(\partial\Sigma_\theta/\partial\theta_i)\Sigma_\theta^{-1}]\Sigma^{1/2}$. Letting $V_\theta = W'\Sigma_\theta^{-1}W$, we have

$$
\begin{aligned}
&\left|\mathrm{trace}\left(\frac{\partial C}{\partial\theta_i}\frac{\partial C'}{\partial\theta_j}\right)\right| \\
\leq\ &\left|\mathrm{trace}\left(N^{2-2H}P_N V_\theta^{-1}W'\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_i}\Sigma_\theta^{-1}WV_\theta^{-1}W'\Sigma_\theta^{-1}\Sigma\Sigma_\theta^{-1}WV_\theta^{-1}W'\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_j}\Sigma_\theta^{-1}WV_\theta^{-1}P_N\right)\right| \\
&+\left|\mathrm{trace}\left(N^{2-2H}P_N V_\theta^{-1}W'\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_i}\Sigma_\theta^{-1}WV_\theta^{-1}W'\Sigma_\theta^{-1}\Sigma\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_j}\Sigma_\theta^{-1}WV_\theta^{-1}P_N\right)\right| \\
&+\left|\mathrm{trace}\left(N^{2-2H}P_N V_\theta^{-1}W'\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_i}\Sigma_\theta^{-1}\Sigma\Sigma_\theta^{-1}WV_\theta^{-1}W'\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_j}\Sigma_\theta^{-1}WV_\theta^{-1}P_N\right)\right| \\
&+\left|\mathrm{trace}\left(N^{2-2H}P_N V_\theta^{-1}W'\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_i}\Sigma_\theta^{-1}\Sigma\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_j}\Sigma_\theta^{-1}WV_\theta^{-1}P_N\right)\right|.
\end{aligned}
$$

It can be shown that for any $\delta > 0$ and uniformly for θ in a sufficiently small neighbourhood of θ^*,

(a) $\|N^{2-2H}P_N V_\theta^{-1}P_N\| = O(N^\delta)$,

(b) $\|(N^{1-H}P_N)^{-1}W'\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_i}\Sigma_\theta^{-1}W(N^{1-H}P_N)^{-1}\| = O(N^\delta)$,

(c) $\|(N^{1-H}P_N)^{-1}W'\Sigma_\theta^{-1}\Sigma\Sigma_\theta^{-1}W(N^{1-H}P_N)^{-1}\| = O(N^\delta)$,

(d) $\|(N^{1-H}P_N)^{-1}W'\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_i}\Sigma_\theta^{-1}\Sigma\Sigma_\theta^{-1}W(N^{1-H}P_N)^{-1}\| = O(N^\delta)$,

(e) $\|(N^{1-H}P_N)^{-1}W'\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_i}\Sigma_\theta^{-1}\Sigma\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_j}\Sigma_\theta^{-1}W(N^{1-H}P_N)^{-1}\| = O(N^\delta)$,

where $\|\cdot\|$ denotes the spectral norm of the enclosed matrix. Consequently, there exists some constant K such that, uniformly for $\theta^{(1)}$ and $\theta^{(2)}$ in a sufficiently small neighbourhood of θ^*, $\rho^2(\theta^{(1)},\theta^{(2)}) \leq KN^\delta|\theta^{(1)}-\theta^{(2)}|^2$, which can be checked by making use of (a)-(e) and by observing that for any square $r \times r$ matrix A, $|\!|\!| A |\!|\!| \leq \sqrt{r}\|A\|$. Then, we can adapt the chaining argument in the proof of Theorem 3.1 in Dahlhaus (1995) to show that $N^{1-H}P_N(\hat\gamma(\hat\theta) - \hat\gamma(\theta^*)) = o_p(1)$. Because $\hat\gamma(\theta^*)$ has the limiting distribution stated in Theorem 2, by Theorems 2.1 and 2.3 of Dahlhaus (1995), so has $\hat\gamma$.

We now verify (a) which requires us to show that $\|N^{1-H}P_N(W'\Sigma_\theta^{-1}W)^{-1}N^{1-H}P_N\| \leq KN^\delta$ for some constant K and any $\delta > 0$. This inequality holds iff for all vector x of appropriate dimension, (below each occurrence of K may denote a different constant)

$$x'(N^{1-H}P_N)^{-1}W'\Sigma_\theta^{-1}W(N^{1-H}P_N)^{-1}x \geq \frac{K}{N^\delta}|x|^2.$$

But the left side of the preceding inequality can be written as
$x'(N^{1-H}P_N)^{-1}W'\Sigma^{-1/2}\Sigma^{1/2}\Sigma_\theta^{-1}\Sigma^{1/2}\Sigma^{-1/2}W(N^{1-H}P_N)^{-1}x$ which is not less than $KN^{-\delta}|\Sigma^{-1/2}W(N^{1-H}P_N)^{-1}x|^2$ iff the spectral norm of $\Sigma^{-1/2}\Sigma_\theta\Sigma^{-1/2}$ is less than KN^δ, uniformly for θ in a sufficiently small neighbourhood of θ^*, which follows from applying Lemma 5.3 of Dahlhaus (1989). Finally, note that
$|\Sigma^{-1/2}W(N^{1-H}P_N)^{-1}x|^2 = x'(N^{1-H}P_N)^{-1}W'\Sigma^{-1}W(N^{1-H}P_N)^{-1}x$
and the matrix $(N^{1-H}P_N)^{-1}W'\Sigma^{-1}W(N^{1-H}P_N)^{-1}$ converges to a positive definite matrix so that the quadratic form is larger than a certain multiple of $|x|^2$. This completes the proof of (a); those of (b)-(e) are similar and hence omitted.

We now outline the proof that $\hat{\gamma}$ is asymptotically independent of $\hat{\theta}$. First, $N^{1-H}P_N(\hat{\gamma} - \gamma^*)$ is asymptotically equal to $N^{1-H}P_N(W'\Sigma_{\theta^*}^{-1}W)^{-1}W'\Sigma_{\theta^*}^{-1/2}\epsilon$ where $\epsilon = \Sigma_{\theta^*}^{-1/2}X$ consists of N independent standard normal variables. Also, it follows from the proof of Theorem 3.2 of Dahlhaus (1989) that $\Gamma_h(\theta^*)(\hat{\theta} - \theta^*)$ is asymptotically equal to $-\sqrt{N}\nabla\mathcal{L}_{1,N}(\theta^*)/2$, where $\Gamma_h(\theta^*)$ is defined in Theorem 2. Up to an additive deterministic function of θ^*, $N\nabla\mathcal{L}_{1,N}(\theta^*)$ is a random vector consisting of the quadratic forms $-X'\Sigma_{\theta^*}^{-1}(\partial\Sigma_{\theta^*}/\partial\theta_k)\Sigma_{\theta^*}^{-1}X$ $= -\epsilon'\Sigma_{\theta^*}^{-1/2}(\partial\Sigma_{\theta^*}/\partial\theta_k)\Sigma_{\theta^*}^{-1/2}\epsilon$, which has zero covariance with $N^{1-H}P_N(W'\Sigma_{\theta^*}^{-1}W)^{-1}W'\Sigma_{\theta^*}^{-1/2}\epsilon$, owing to the fact that the components of ϵ are independent standard normal variables. This is because any linear form of $\epsilon = (\epsilon_1, \cdots, \epsilon_N)'$, say $\ell = \sum_{i=1}^N \ell_i\epsilon_i$, must be uncorrelated with any quadratic form, say $Q = \sum_{1\leq j,k\leq N} q_{jk}\epsilon_i\epsilon_j$, from the fact that $\text{cov}(\ell, Q) = \sum_{i,j,k} \ell_i q_{jk} E(\epsilon_i\epsilon_j\epsilon_k) = 0$. The joint asymptotic normality of $\hat{\theta}$ and $\hat{\gamma}$ then implies that they are asymptotically independent.

References

Baillie, R. (1996) Long memory processes and fractional integration in econometrics. *J. Econometrics* **73**, 5-59.

Beran, J. (1994) *Statistics for Long-Memory Presses.* New York: Chapman and Hall.

Beran, J. (1995) Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. R. Statist. Soc.* **B 57**, 659-672.

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

Brockwell, P. J. (1993) Threshold ARMA processes in continuous time. In *Dimension Estimation and Models*, Ed. H. Tong, pp. 170-90. River Edge: World Scientific.

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

Brockwell, P. J. and Marquardt, T. (2005) Lévy-driven and fractionally integrated ARMA processes with continuous time parameter. *Statistica Sinica* **15**, 477-494.

Chambers, M. J. (1996) The estimation of continuous parameter long-memory time series models. *Econometric Theory* **12**, 374-390.

Comte, F. (1996) Simulation and estimation of long memory continuous time models. *J. Time Ser. Anal.* **17**, 19-36.

Comte, F. and Renault, E. (1996) Long memory continuous time models. *J. Econometrics* **73**, 101-149.

Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes. *Ann. Stat.* **17**, 1749-1766.

Dahlhaus, R. (1995) Efficient location and regression estimation for long range dependent regression models. *Ann. Statist.* **23**, 1029-1047.

Dahlhaus, R. (2004) Correction note for efficient parameter estimation for self-similar processes, working paper.

Davies, R.B. and Harte, D.S. (1987) Tests for Hurst effect. *Biometrika* **74**, 95-101.

Duncan, Tyrone E., Hu, Yaozhong and Pasik-Duncan Bozenna (2000) Stochastic calculus for fractional Brownian motion. I. Theory. *SIAM J. Control Optim.* **38**, 582–612.

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

Elliott, R. and van der Hoek, J. (2003) A general fractional white noise theory and applications to finance. *Math. Finance* **13**, 301-330.

Granger, C. W. J. and Joyeux, R. (1980) An introduction to long memory time series models and fractional differencing. *J. Time Ser. Anal.* **1**, 15-29.

Gripenberg, G. and Norros, I. (1996) On the prediction of fractional Brownian motion. *J. Appl. Probab.* **33**, 400-410.

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika* **68**, 165-76.

Hu, Y. and Øksendal, B. (2003) Fractional white noise calculus and application to finance. *Inf. Dim. Anal. Wuantum Probab. Rel. Top.* **6**, 1-32.

Ling, S. and Li, W. K. (1997) On fractionally integrated autoregressive moving-average time series models with conditional heteroscedascity. *J. Amer. Statist. Assoc.* **92**, 1184-1194.

Mandelbrot, B.B. and Van Ness, J. W. (1968) Fractional Brownian motions, fractional noises and applications. *SIAM Review* **10**, 422-437.

Øksendal, B. (2003) Fractional Brownian motion in finance. University of Oslo. Pure Mathematics Preprint Series 2003, No. 28. Can be downloaded from `http://www.math.uio.no/eprint/pure_math/2003/28-03.pdf`

Palma, W. and Chan, N.H. (1997) Estimation and forecasting of long-memory processes. *J. Forecasting* **16**, 395-410.

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

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

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

Rudin, W. (1976) *Principles of mathematical analysis*. New York: McGraw-Hill.

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

Shapiro, H. S. and Silverman, R. A. (1960) Alias-free sampling of random noise. *Journal of the Society for Industrial and Applied Mathematics* **8**, 225-248.

So, M. K. P. (1999) Time series with additive noise. *Biometrika* **86**, 474-482.

Sowell, F. (1992) Modeling long-run behaviour with the fractional ARMA Model. *J. Monetary Econ.* **29**, 277-302.

Stramer,O., Tweedie, R.L. and Brockwell, P.J. (1996) Existence and stability of continuous time threshold ARMA processes. *Statistica Sinica* **6**, 715-732.

Taqqu, M. S. (2003) Fractional brownian motion and long-range dependence. In *Theory and Applications of Long-Range Dependence* (eds. Doukhan, Oppenheim, and Taqqu), pp. 3-39.

Tsai, H. and Chan, K. S. (2005) Quasi-maximum likelihood estimation for a class of continuous-time long-memory processes. *J. Time Ser. Anal.* In press.

Valkeila, E. (1999) On some properties of geometric fractional Brownian motion. Preprint, Univ. of Helsinki, May 1999.

Viano, M. C., Deniau, C. and Oppenheim, G. (1994) Continuous-time fractional ARMA processes. *Statist. Probab. Lett.* **21**, 323-336.

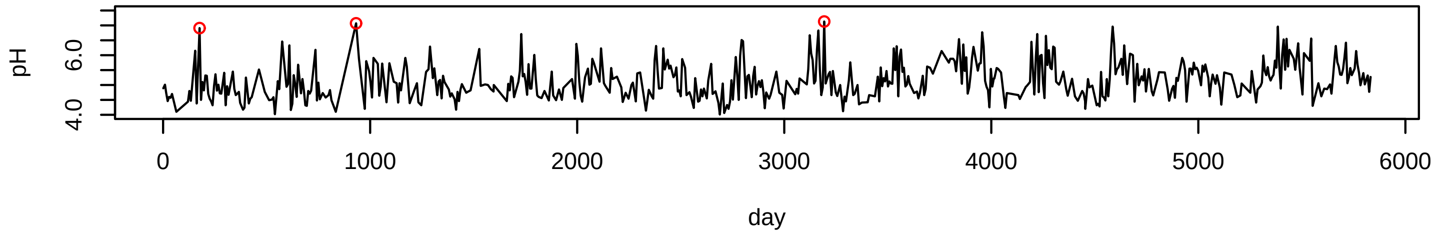

Fig. 1. Time series plot of the field values of PH of wet deposition at the McNay Research Station; open circles mark the 3 outliers.

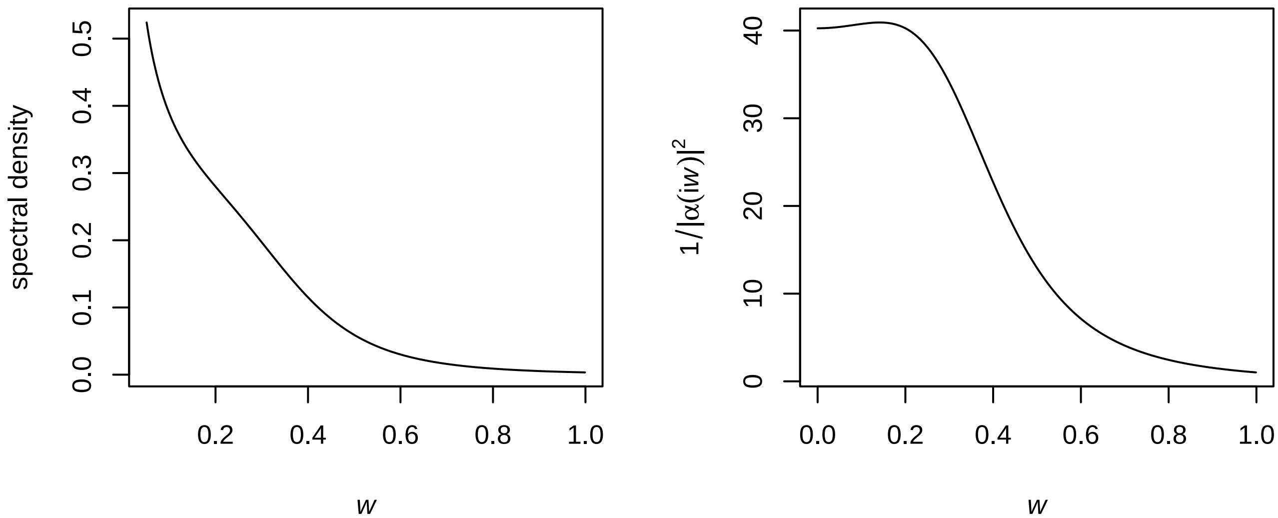

Fig. 2. Left figure is the spectral density estimate and right figure is the plot of squared magnitude of the transfer function of the autoregressive part of the model.