Statistica Sinica **16**(2006), 213-226

QUASI-MAXIMUM LIKELIHOOD ESTIMATION OF LONG-MEMORY LIMITING AGGREGATE PROCESSES

Henghsiu Tsai

Academia Sinica, Taipei

Abstract: We consider the application of the limiting aggregate model derived by Tsai and Chan (2005d) for modeling aggregated long-memory data. The model is characterized by the fractional integration order of the original process and may be useful for (i) modeling discrete-time data with sufficiently long sampling intervals, for example, annual data, and/or (ii) studying the fractional integration order of the original process. The fractional integration parameter is estimated by maximizing the Whittle likelihood. It is shown that the quasi-maximum likelihood estimator is asymptotically normal, and its finite-sample properties are studied through simulation. The efficacy of the proposed approach is demonstrated with three data analyses.

Key words and phrases: ARFIMA models, asymptotic normality, temporal aggregation, Whittle likelihood.

1. Introduction

Time series data is often temporally aggregated before analysis. Typical examples are the incidence rates of diseases, sales of products, industrial production, tree-ring widths, riverflows, and rainfall that can only be obtained through aggregation over a certain time interval. For a short-memory time series, the aggregate data approaches white noise with increasing aggregation, due to the Central Limit Theorem. The aggregates of a non-stationary process, however, do not have a white noise structure in the limit. It was first proved by Working (1960) that the aggregates of an ARIMA(0,1,0) process have a limiting ARIMA(0,1,1) structure with the MA(1) parameter equal to -0.268. This result was later generalized by Tiao (1972), who obtained the interesting result that if the basic series is nonstationary and follows an IMA(d, q) process with $d \geq 1$, then, with increasing aggregation, the model of the (appropriately re-scaled) aggregate data becomes an IMA(d, d) model with the MA parameters uniquely determined by the differencing order.

Tiao's (1972) aggregation result has practical relevance. Rossana and Seater (1995) investigated the effects of temporal aggregation on the estimated time series properties of a number of economic time series. They estimated ARIMA models for monthly, quarterly, and annual data and found that the annual time series models are usually ARIMA(0,1,1), or random walks. Rossana and Seater (1995, p.445) suggested that Tiao's (1972) asymptotic model of temporal aggregation is a good approximation for many economic time series.

Recently, Tsai and Chan (2005d) extended Tiao's (1972) limiting analysis by considering the problem of temporal aggregation of stationary and nonstationary processes that may have components of short-memory, long-memory, and anti-persistence. The long range dependence properties of time series data have diverse applications in many fields, including hydrology, finance, economics, and telecommunications; see Bloomfield (1992), Robinson (1993), Beran (1994), Baillie (1996) and Ray and Tsay (1997). Tsai and Chan (2005d) showed that temporal aggregation does not change the long-memory parameter of the underlying ARFIMA process (see Section 2). Furthermore, as the extent of aggregation increases to infinity, the limiting model still preserves the long-memory/anti-persistent parameter(s) of the original process, whereas the short-memory components vanish. The limiting normalized model, after an appropriate amount of differencing, essentially contains only one parameter, the fractional integration order of the original process. As suggested by Rossana and Seater (1995), the limiting aggregate model may be preferable for aggregate data with sufficiently long aggregation intervals, for example, annual data. If a limiting aggregate model fits an aggregated time series well, the estimate of the fractional integration order provides a good estimate of the fractional integration order of the original process.

The goal of the present paper is to investigate the use of the limiting aggregate model in data analysis, and to compare it with some existing models. In Section 2, the long-memory limiting aggregate model derived by Tsai and Chan (2005d) is briefly reviewed. Quasi-maximum likelihood estimation of the long-memory model, obtained by maximizing the Whittle likelihood, and the large sample properties of the estimator are discussed in Section 3. Finite sample properties of the estimator are studied via simulation in Section 4. The method is illustrated with three data applications in Section 5. Finally in Section 6, we present our conclusions.

2. Review of The Long-Memory Limiting Aggregate Model

Consider a fractionally integrated ARMA (ARFIMA$(p, r+d, q)$) process $\{\tilde{Y}_t\}$ that satisfies the difference equation

$$\phi(B)(1-B)^{r+d}\tilde{Y}_t = \theta(B)\epsilon_t, \tag{1}$$

where r is a non-negative integer and $0 < d < 1/2$. Alternatively, after rth differencing, the process $\{U_t\}$, where $U_t = (1-B)^r\tilde{Y}_t$, is a stationary ARFIMA(p, d, q) process that follows the equation

$$\phi(B)(1-B)^{d}U_t = \theta(B)\epsilon_t, \tag{2}$$

where $\{\epsilon_t\}$ is a sequence of white noises with mean 0 and variance σ^2, $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$, $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$. Here B is the backward shift operator, $(1-B)^d$ being defined by the binomial series

$$(1-B)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} B^k,$$

see Granger and Joyeux (1980), Hosking (1981) and Brockwell and Davis (1991, p.520). The roots of $\phi(B) = 0$ and those of $\theta(B) = 0$ are assumed to lie outside the unit circle. The spectral density of $\{U_t\}$ is given by

$$f_U(\omega) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} |1 - e^{-i\omega}|^{-2d},$$

see Brockwell and Davis (1991, p.525), so the process $\{U_t\}$ is long-memory. For a unified framework for studying non-stationary and/or long-memory discrete-time processes, see Beran (1995), Ling and Li (1997) and Beran, Bhansali and Ocker (1998).

Let $\{\tilde{Y}_t\}$ be generated by (1), $s \geq 2$ be an integer, and $X_T^s = \sum_{k=s(T-1)+1}^{sT} \tilde{Y}_k$ be the non-overlapping s-temporal aggregates of $\{\tilde{Y}_t\}$. Tsai and Chan (2005d) showed that the rth differenced aggregate series $\{(1-B)^r X_T^s, T = 0, \pm 1, \pm 2, \cdots\}$ is a stationary process, the limiting normalized spectral density function of which, as $s \to \infty$, is given by

$$f_\infty(\omega; r, d) = K\{2(1-\cos\omega)\}^{r+1} \sum_{k=-\infty}^{\infty} |\omega + 2k\pi|^{-2r-2d-2}, \quad -\pi < \omega < \pi, \quad (3)$$

where K is the normalization constant ensuring that $\int_{-\pi}^{\pi} f_\infty(\omega; r, d) d\omega = 1$. Note that (i) $f_\infty(\omega; r, d)$ is of $O(|\omega|^{-2d})$ for $\omega \to 0$, so the limit of the aggregates preserves the long-memory parameter of the underlying ARFIMA process, (ii) the limiting normalized spectral density function is independent of the short-memory parameters $\{\phi_j\}$ and $\{\theta_j\}$, and (iii) when $r = 0$, $f_\infty(\omega; r, d)$ is the normalized spectral density of fractional Gaussian noise, so the limiting aggregate model can be regarded as an extension of fractional Gaussian noise.

3. Quasi-maximum Likelihood Estimator and Its Large Sample Properties

We are interested in the limiting aggregate model defined in (3). Note that the spectral density in (3) can be usefully reparameterized by letting $\eta = r + d$. The parameter η is called the fractional integration order of the original process.

We also introduce the parameter σ to account for the variance of the data. Let $\{Y_i\}_{i=1}^N$ be a stationary Gaussian process with the spectral density function

$$f(\omega;\eta,\sigma^2)=\sigma^2\{2(1-\cos\omega)\}^{[\eta]+1}\sum_{k=-\infty}^{\infty}|\omega+2k\pi|^{-2\eta-2},\quad -\pi<\omega<\pi, \tag{4}$$

where $[\eta]$ is the greatest integer $\leq \eta$. It can be verified that the auto-covariance function corresponding to the spectral density function defined in (4) is

$$\gamma(h)=\frac{2\pi\sigma^2\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\Big(\prod_{j=0}^{2r+1}\frac{1}{2d+j}\Big)\sum_{k=0}^{2r+2}(-1)^k\binom{2r+2}{k}|r+1-h-k|^{2d+2r+1}, \tag{5}$$

where $r=[\eta]$, $d=\eta-[\eta]$ and $\Gamma()$ is the Gamma function. The (negative) log-likelihood function of $\{Y_i\}$ can be approximated by the (negative) Whittle log-likelihood function (see Dahlhaus (1989))

$$-\tilde{l}(\eta,\sigma^2)=\sum_{j=1}^{T}\left\{\log f(\omega_j;\eta,\sigma^2)+\frac{I_N(\omega_j)}{f(\omega_j;\eta,\sigma^2)}\right\}, \tag{6}$$

where $\omega_j := 2\pi j/N \in (0,\pi)$ are the Fourier frequencies, T is the largest integer $\leq (N-1)/2$, $I_N(\omega)=\left|\sum_{j=1}^N (Y_j-\bar{Y})e^{ij\omega}\right|^2/(2\pi N)$, and $\bar{Y}=\sum_{j=1}^N Y_j/N$. In (6), the computation of $f(\omega_j;\eta,\sigma^2)$ requires evaluation of an infinite sum. Here, we adopt the method of Chambers (1996) to approximate $f(\omega;\eta,\sigma^2)$ by

$$\tilde{f}(\omega;\eta,\sigma^2)=\sigma^2\{2(1-\cos\omega)\}^{[\eta]+1}h(\omega;\eta),\qquad -\pi<\omega<\pi, \tag{7}$$

where $h(\omega;\eta)=(2\eta+1)^{-1}(2\pi)^{2\eta+1}\{(2\pi M-\omega)^{-2\eta-1}+(2\pi M+\omega)^{-2\eta-1}\}+\sum_{k=-M}^{M}|\omega+2k\pi|^{-2\eta-2}$ for some large integer M that is a function of N. Note that the approximation error is $O(M^{-2\eta-1})$, see Chambers (1996). It is shown below that the approximation has a negligible effect on the limiting properties of the quasi-maximum likelihood estimator if M is chosen such that $\sqrt{N}M^{-2\eta-1}\to 0$ for $N\to\infty$ (see Theorem 1). In practice, we set the value of M to be the value of N. Replacing $f(\omega_j;\eta,\sigma^2)$ with $\tilde{f}(\omega_j;\eta,\sigma^2)$ and letting $\tilde{g}(\omega_j;\eta)=\tilde{f}(\omega_j;\eta,\sigma^2)/\sigma^2$, the (negative) Whittle log-likelihood function (6) now becomes

$$-\tilde{l}_M(\eta,\sigma^2)=\sum_{j=1}^{T}\left\{\log\sigma^2+\log\tilde{g}(\omega_j;\eta)+\frac{I_N(\omega_j)}{\sigma^2\tilde{g}(\omega_j;\eta)}\right\}. \tag{8}$$

Differentiating (8) with respect to σ^2 and equating to zero gives

$$\hat{\sigma}_M^2=\frac{1}{T}\sum_{j=1}^{T}\frac{I_N(\omega_j)}{\tilde{g}(\omega_j;\eta)}. \tag{9}$$

Substituting (9) into (8) yields the objective function

$$-\tilde{l}_M(\eta) = \sum_{j=1}^{T} \log \tilde{g}(\omega_j;\eta) + T \log\Big(\sum_{j=1}^{T} \frac{I_N(\omega_j)}{\tilde{g}(\omega_j;\eta)}\Big) + C, \tag{10}$$

where $C = T - T\log T$. The objective function is minimized with respect to η to get the quasi-maximum likelihood estimator (QMLE) $\hat{\eta}_M$; the estimator $\hat{\sigma}^2_M$ is then calculated by (9).

For simplicity, let $\theta = (\eta, \sigma^2)$, and $\hat{\theta}_M = (\hat{\eta}_M, \hat{\sigma}^2_M)$ be the quasi-maximum likelihood estimator that minimizes the (negative) Whittle log-likelihood function (8). The estimator $\hat{\theta}_M$ can be shown to be asymptotically normal by use of Theorem 2.1 of Dahlhaus (1989). We now state this result in the following theorem.

Theorem 1. *Let $Y = \{Y_i\}_{i=1}^N$ be sampled from a stationary Gaussian long-memory process with the spectral density* (4). *Let the quasi-maximum likelihood estimator $\hat{\theta}_M \in \Theta$, a compact parameter space, and the true parameter θ_0 be in the interior of the parameter space. Let M be such that $\sqrt{N}M^{-2\eta-1} \to 0$ for $N \to \infty$. Then $\sqrt{N}(\hat{\theta}_M - \theta_0)$ tends in distribution to a normal random vector with mean 0 and covariance matrix $\Gamma(\theta_0)^{-1}$ with*

$$\Gamma(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi} (\bigtriangledown \log f(\omega;\eta,\sigma^2))(\bigtriangledown \log f(\omega;\eta,\sigma^2))' d\omega, \tag{11}$$

where $\bigtriangledown$ denotes the derivative with respect to θ, and superscript $'$ denotes transpose.

See the Appendix for proof of the theorem. We note that $\Theta = \{(\bar{\eta}, \bar{\sigma}^2) \mid 0 \le \bar{\eta} - [\bar{\eta}] \le 1/2,\ \bar{\eta} \ge 0, \text{ and } \bar{\sigma}^2 \ge 0\}$. The compactness condition on the parameter space is taken from condition (A0) in Dahlhaus (1989) who pointed out that the quasi-maximum likelihood estimators may lie on the boundary of the compact parameter space.

4. Simulations

In this section, we report some finite sample performance of the quasi-maximum likelihood estimator for six models in which $\eta = 0.05$, 0.25, 0.45, 1.05, 1.25 and 1.45, and $\sigma = 2$. Because $\{Y_i\}$ is a stationary Gaussian process, we can use the method of Davies and Harte (1987) to simulate $\{Y_i\}_{i=1}^N$. The sample sizes considered are $N = 512, 1024$ and 2048. All the computations in this and the following section were performed using Fortran code with IMSL subroutines. The quasi-maximum likelihood estimator $\hat{\eta}_M$ is computed based on equation (10) using the following procedure. We first find the local maximum likelihood estimators of η between $[\eta]+0$ and $[\eta]+0.5$, for $[\eta]=0,\ldots,K$, for some integer K. In our experiments, we chose K to be 5. These local maximum likelihood estimators are then used to find the global maximum likelihood estimator of η. For each model, the averages and the standard deviations of 1,000 replicates of the estimators are summarized in Tables 1−3 for $N=512, 1024$ and 2048, respectively. The tables also show the asymptotic standard errors of the parameter estimators computed from $\Gamma(\theta)$ defined in equation (11). The value of M used in the computation of $h(\omega;\eta)$, defined below (7), is set to be the value of N. We have also tried $M=10N$ and $M=20N$ in the program and the results are essentially the same.

Table 1 shows that for sample size $N=512$, there is about 0.7 % to 12.1 % chance that $[\eta]$ will be incorrectly estimated, resulting in larger biases and standard errors of the estimates of η. The incorrect estimation rates of $[\eta]$ decrease to no more than 2.2 % for $N=1,024$. For $N=2,048$, all the incorrect rates become zero, and the empirical standard errors are very close to the asymptotic standard errors. Numerical computations show that the asymptotic standard error of $\hat{\eta}$ increases with a larger η value, although differences are small for $\eta>1$.

Table 1. Averages (s.e.=standard errors) [asymp. s.e.=asymptotic standard errors] of 1,000 simulations of the quasi-maximum likelihood estimators of the parameters η for $\eta=0.05$, 0.25, 0.45, 1.05, 1.25 and 1.45, and $\sigma=2$. The sample size $N=512$.

True value of η	Estimated value of $[\eta]$	% of 1,000 replications	Quasi-MLE of η average (s.e.) (for each estimated $[\eta]$)	Quasi-MLE of σ average (s.e.) (for each estimated $[\eta]$)	Quasi-MLE of η average (s.e.) [asymp. s.e.] (overall)	Quasi-MLE of σ average (s.e.) [asymp. s.e.] (overall)
0.05	0	99.3	0.0572 (0.0261)	1.9776 (0.0665)	0.0638 (0.0828)	1.9807 (0.0765)
	1	0.7	1.0000 (0.0000)	2.4230 (0.0965)	[0.0279]	[0.0659]
0.25	0	94.5	0.2522 (0.0292)	1.9947 (0.0653)	0.2995 (0.1982)	2.0144 (0.1046)
	1	5.5	1.1111 (0.0378)	2.3519 (0.0702)	[0.0292]	[0.0644]
0.45	0	94.6	0.4529 (0.0292)	2.0048 (0.0662)	0.4999 (0.1989)	2.0223 (0.0999)
	1	5.4	1.3226 (0.3721)	2.3346 (0.0720)	[0.0298]	[0.0639]
	0	4.3	0.1922 (0.0299)	1.6552 (0.0575)	1.0312 (0.2222)	1.9878 (0.1025)
1.05	1	93.8	1.0500 (0.0291)	1.9971 (0.0632)	[0.0306]	[0.0634]
	2	1.9	2.0001 (0.0384)	2.2847 (0.0716)		
	0	5.3	0.3795 (0.0293)	1.6844 (0.0534)	1.2645 (0.3064)	2.0026 (0.1227)
1.25	1	87.9	1.2508 (0.0313)	1.9996 (0.0642)	[0.0307]	[0.0633]
	2	6.8	2.1317 (0.0371)	2.2889 (0.0690)		
	0	2.0	0.5000 (0.0000)	1.6893 (0.0499)	1.4862 (0.2549)	2.0190 (0.1063)
1.45	1	92.0	1.4521 (0.0301)	2.0081 (0.0664)	[0.0308]	[0.0633]
	2	6.0	2.3372 (0.0361)	2.2951 (0.0728)		

Table 2. Averages (s.e.=standard errors) [asymp. s.e.=asymptotic standard errors] of 1,000 simulations of the quasi-maximum likelihood estimators of the parameters η for $\eta = 0.05,\ 0.25,\ 0.45,\ 1.05,\ 1.25$ and 1.45, and $\sigma = 2$. The sample size $N = 1,024$.

True value of η	Estimated value of $[\eta]$	% of 1,000 replications	Quasi-MLE of η average (s.e.) (for each estimated $[\eta]$)	Quasi-MLE of σ average (s.e.) (for each estimated $[\eta]$)	Quasi-MLE of η average (s.e.) [asymp. s.e.] (overall)	Quasi-MLE of σ average (s.e.) [asymp. s.e.] (overall)
0.05	0	100	0.0565 (0.0190)	1.9792 (0.0478)	0.0565 (0.0190) [0.0197]	1.9792 (0.0478) [0.0466]
0.25	0	99.6	0.2516 (0.0207)	1.9954 (0.0465)	0.2551 (0.0586) [0.0206]	1.9968 (0.0516) [0.0456]
	1	0.4	1.1202 (0.0177)	2.3481 (0.0294)		
0.45	0	99.6	0.4524 (0.0214)	2.0031 (0.0471)	0.4559 (0.0596) [0.0211]	2.0043 (0.0513) [0.0452]
	1	0.4	1.3334 (0.0148)	2.3255 (0.0250)		
1.05	0	1.1	0.1835 (0.0147)	1.6537 (0.0332)	1.0426 (0.1067) [0.0217]	1.9951 (0.0605) [0.0448]
	1	98.6	1.0493 (0.0218)	1.9979 (0.0457)		
	2	0.3	2.0000 (0.0000)	2.3084 (0.0358)		
1.25	0	1.2	0.3713 (0.0153)	1.6836 (0.0348)	1.2482 (0.1339) [0.0217]	1.9988 (0.0653) [0.0448]
	1	97.8	1.2498 (0.0221)	1.9994 (0.0458)		
	2	1.0	2.1531 (0.0184)	2.3127 (0.0470)		
1.45	0	0.3	0.5000 (0.0000)	1.7019 (0.0364)	1.4585 (0.1104) [0.0218]	2.0073 (0.0592) [0.0448]
	1	98.6	1.4514 (0.0221)	2.0049 (0.0471)		
	2	1.1	2.3569 (0.0168)	2.3086 (0.0434)		

Table 3. Averages (s.e.=standard errors) [asymp. s.e.=asymptotic standard errors] of 1,000 simulations of the quasi-maximum likelihood estimators of the parameters η for $\eta = 0.05,\ 0.25,\ 0.45,\ 1.05,\ 1.25$ and 1.45, and $\sigma = 2$. The sample size $N = 2,048$.

True value of η	Estimated value of $[\eta]$	% of 1,000 replications	Quasi-MLE of η average (s.e.) (for each estimated $[\eta]$)	Quasi-MLE of σ average (s.e.) (for each estimated $[\eta]$)	Quasi-MLE of η average (s.e.) [asymp. s.e.] (overall)	Quasi-MLE of σ average (s.e.) [asymp. s.e.] (overall)
0.05	0	100	0.0563 (0.0132)	1.9801 (0.0324)	0.0563 (0.0132) [0.0139]	1.9801 (0.0324) [0.0330]
0.25	0	100	0.2513 (0.0144)	1.9955 (0.0312)	0.2513 (0.0144) [0.0146]	1.9955 (0.0312) [0.0322]
0.45	0	100	0.4515 (0.0152)	2.0008 (0.0313)	0.4515 (0.0152) [0.0149]	2.0008 (0.0313) [0.0319]
1.05	1	100	1.0494 (0.0152)	1.9985 (0.0306)	1.0494 (0.0152) [0.0153]	1.9985 (0.0306) [0.0317]
1.25	1	100	1.2498 (0.0153)	1.9993 (0.0306)	1.2498 (0.0153) [0.0154]	1.9993 (0.0306) [0.0317]
1.45	1	100	1.4510 (0.0157)	2.0021 (0.0311)	1.4510 (0.0157) [0.0154]	2.0021 (0.0311) [0.0316]

5. Applications

We now illustrate the limiting aggregate model with three data sets posted on the website http://www-personal.buseco.monash.edu.au/~hyndman/TSDL.

Example 1. Annual tree ring measurements from California, USA, from 1027 A.D. through 1987 A.D., a total of 961 years. 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 CA531.DAT under the tree-rings category. The tree ring series is of considerable climatological interest. Also, a lot of tree ring data exhibits long range dependence properties, see Baillie (1996). The annual tree ring data can be considered as aggregates of the underlying continuous-time growth rate process. In Section 2, we motivate the limiting model by temporal aggregation of a discrete-time ARFIMA process; however, as proved by Tsai and Chan (2005c), the limiting aggregate model of a continuous-time ARFIMA process is the same as that of a discrete-time ARFIMA process. Therefore, the limiting model might be applied to the tree ring series.

Figure 1(a) shows the time series plot of the tree ring measurements, while Figure 1(b) shows the sample auto-correlation of the data. We have fitted the limiting aggregate model and the continuous-time ARFIMA$(p, H, 0)$ models of Tsai and Chan (2005a) with the autoregressive order $1 \leq p \leq 4$ for the tree ring data. For each model, we computed the corresponding Akaike Information Criterion $AIC = -2(l_Y(\hat{\theta}) - r)$, where r is the number of parameters in the model, and $-l_Y$ is the log-likelihood function. The AIC of the limiting aggregate model is $-3,807.94$, and those of the continuous-time ARFIMA$(p, H, 0)$ models for $p = 1, 2, 3$, and 4 are $-3,807.88$, $-3,806.65$, $-3,807.61$ and $-3,802.17$ respectively. These are all larger than the limiting aggregate model; therefore, based on Akaike's Information Criterion, the limiting aggregate model is the preferred one. However, it should be noted that AIC is a criterion for selection among nested. As the limiting aggregate model and the continuous-time ARFIMA models are not nested, the AIC can only be used as a reference in this example.

The parameter estimates of the limiting aggregate model are $\hat{\eta} = 0.2863$ and $\hat{\sigma} = 0.0915$. To verify that the model is an adequate fit for the data, we compute the Ljung-Box statistic, $\tilde{Q}_m = N(N+2)\sum_{k=1}^{m} \hat{r}_k^2/(N-k)$, where the $\hat{r}_k$s are the auto-correlation functions of the residuals, and the residuals are computed by the innovations algorithm (Brockwell and Davis (1991)). The Ljung-Box statistic has an asymptotic chi-squared distribution with $m - r$ degrees of freedom, where r is the number of parameters, if the fitted model is adequate (Ljung and Box (1978) and Li (2004, p.145)). The Ljung-Box statistic using $m = 20$ is $\tilde{Q}_m = 19.48$, suggesting that this model provides an adequate fit. The asymptotic standard errors of the estimates are 0.0293 and 0.0029 for η and σ, respectively, suggesting that the data is indeed long-memory.

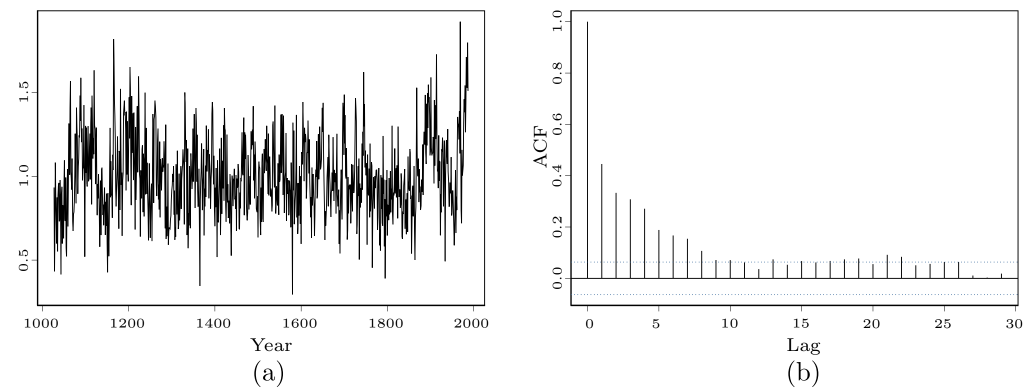


Figure 1. The annual tree ring measurements from California, USA: (a) time series plot; (b) sample auto-correlation function.

Example 2. Annual Swedish Fertility Rates (per thousand), 1750-1849. (McCleary and Hay (1980)). The fertility rate in the ith year is defined as f_i=births per 1,000 female population. See McCleary and Hay (1980) for further details. The data, which is posted in the file MCCLEARY15.DAT under the health category, contains 100 observations. Figures 2(a) and (b) show the time series plot and the sample auto-correlation of the data, respectively.

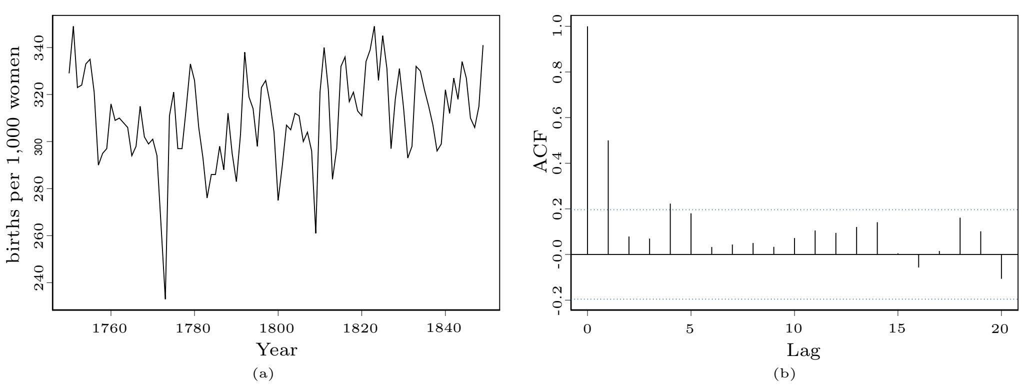
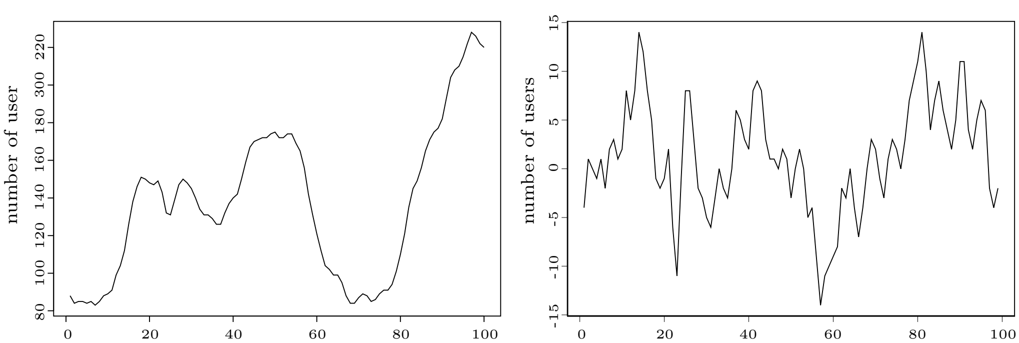


Figure 2. Annual Swedish Fertility Rates, 1750-1849: (a) time series plot; (b) sample auto-correlation function.

McCleary and Hay (1980) fitted the data with the AR(2) model, $f_i = 0.62f_{i-1} - 0.23f_{i-2} + a_i$, where a_i is a white noise process with a variance of 267.84. Note that 267.84 is also the estimated one-step mean square prediction error of the $AR(2)$ process. Meanwhile, the parameter estimates of the limiting aggregagte model are $\hat{\eta} = 2.1593$ and $\hat{\sigma} = 9.8812$. The one-step mean square prediction error of the limiting process can be computed by Kolmogorov's formula (Theorem 5.8.1 of Brockwell and Davis (1991)): $2\pi \exp\{(2\pi)^{-1} \int_{-\pi}^{\pi} \log f(\omega) d\omega\}$, where $f(\omega)$ is the spectral density of the process defined in (4). The error is estimated to be 262.19, which is smaller than that of the AR model.

Using $m = 10$, the Ljung-Box statistics are 7.00 and 9.67 for the limiting aggregate model and the AR(2) model, respectively, indicating adequate fits for both models. In fact, the residuals produced by these two models show similar patterns. However, as the AIC of the AR model is larger than that of the limiting aggregate model (846.22 versus 842.81), the limiting aggregate model may provide a better fit for the data based on the AIC. Still, the AR model and the limiting aggregate model are not nested models. It is interesting to note that the integer part of the fractional integration order of the limiting aggregate model is the same as the AR order of the model of McCleary and Hay (1980).

Example 3. The number of users logged onto an Internet server each minute over a 100-minute period. (Makridakis, Wheelwright and Hyndman (1998)). The data is posted in the file COMPUTER.DAT under the miscellaneous category. See Figure 3(a) for the time series plot of the original data, Figure 3(b) for the data after first differencing, and Figure 4(a) for the sample auto-correlation function of the first-differenced data.



(a) (b)

Figure 3. Number of internet users: (a) time series plot of the number of internet users; (b) time series plot of the change in the number of internet users.

Makridakis, Wheelwright and Hyndman (1998) fitted the data $\{Y_t\}$ with the ARIMA(3,1,0) model $U_t = 1.151U_{t-1} - 0.661U_{t-2} + 0.341U_{t-3} + a_t$, where $\{U_t\}$ is the first-differenced time series and a_t is a white noise process with a variance of 9.66. Meanwhile, the parameter estimates of the limiting aggregate model are

$\hat{\eta} = 3.4222$ and $\hat{\sigma} = 2.3463$. The estimated one-step mean square prediction error of the limiting process is 10.07, which is larger than that of the ARIMA model. Using $m = 10$, the Ljung-Box statistics are 6.99 and 4.34 for the limiting aggregate model and the ARIMA(3,1,0) model, respectively, indicating adequate fits for both models. The AIC of the limiting aggregate model is 511.43, which is larger than the criterion 509.99 of the ARIMA(3,1,0) model. (These are not nested models). It is also noteworthy that the integer part of the fractional order of the limiting aggregate model is the same as the AR order of the ARIMA model.

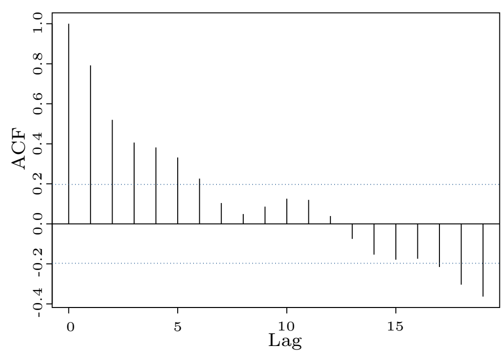


Figure 4. Sample auto-correlation function of the change in the number of internet users.

6. Conclusions

We have successfully applied the long-memory model derived by Tsai and Chan (2005d) to data analysis by quasi-maximum likelihood estimation. The residuals computed by the innovations algorithm provide us with a tool to perform diagnostic checks. Application of the model to more aggregated time series, especially annual data exhibiting long range dependence properties, is an interesting issue for future research.

The limiting aggregate model considered is a Gaussian process. It is known that Gaussianity is not needed for a central limit theorem for Whittle estimators of the discrete-time fractional ARIMA process, see Giraitis, L. and Surgailis, D. (1990). It would be interesting to extend the model by relaxing the Gaussian assumption. We will consider this in our future work.

Acknowledgements

Support for this research was provided by Academia Sinica and the National Science Council (Grant No: NSC 93-2118-M-001-013), R.O.C., and the National Science Foundation (DMS-0405267). The author is grateful to the two anonymous referees for their insightful and helpful comments, to Kung-Sik Chan, Paul

Dunne and John Aston for careful reading and valuable suggestions leading to a streamlined version of the paper. Special thanks go to the Co-Editor for encouraging me to work on a revision and to Han-Wei Ho for many helpful discussions.

Appendix

Proof of Theorem 1 Let $\hat{\theta} = (\hat{\eta}, \hat{\sigma}^2)$ be the quasi-maximum likelihood estimator that minimizes the (negative) Whittle log-likelihood function (6). By equations (i)−(iii) of Dahlhaus (1989, p.1752) and the discussion in the third paragraph of Dahlhaus (1989, p.1753), the asymptotic result stated in the theorem holds if we can show (a) $\sqrt{N}(\hat{\theta} - \theta_0)$ tends in distribution to a normal random vector with mean 0 and covariance matrix $\Gamma(\theta_0)^{-1}$, (b) $\sqrt{N}\{\nabla \tilde{l}_M(\theta_0) - \nabla \tilde{l}(\theta_0)\} = O_p(\sqrt{N} M^{-2\eta-1})$, (c) uniformly for $\theta \in \Theta$, $\tilde{l}_M(\theta) = \tilde{l}(\theta) + O_p(M^{-2\eta-1})$, and (d) $\nabla^2 \tilde{l}_M(\theta) = \nabla^2 \tilde{l}(\theta) + O_p(M^{-2\eta-1})$.

Note that (a) follows from Theorem 2.1 of Dahlhaus (1989) if we can verify conditions (A0)−(A6) listed therein. Verifications of (A1−6) are similiar to those of (A1-6) in Tsai and Chan (2005b) and are, therefore, omitted.

Condition (A0) is about the identifiability of the model. We now verify (A0). Write the spectral density of $\{Y_i\}$ as $f(\omega) = \sigma^2\{2(1-\cos\omega)\}^{[\eta]+1}|\omega|^{-2\eta-2}(R(\omega)+1)$, where $R(x) = |x|^{2\eta+2}\sum_{k\neq 0}|x+2k\pi|^{-2\eta-2}$. Then use the fact that $\lim_{\omega\to 0}\{\log(1-\cos\omega)/\log|\omega|\} = 2$ to get $\lim_{\omega\to 0}\{\log f(\omega)/\log|\omega|\} = -2(\eta - [\eta])$, implying the identifiability of $\eta - [\eta]$. Let $R^{(k)}(\omega)$ be the kth derivative of R with respect to ω. Then

$$\begin{aligned}
&\lim_{\omega\to 0}\frac{\partial^2}{\partial\omega^2}\{\log f(\omega) + 2(\eta - [\eta])\log|\omega|\} \\
&= ([\eta]+1)\lim_{\omega\to 0}\frac{\partial}{\partial\omega}\Big\{\frac{\sin\omega}{1-\cos\omega} - \frac{2}{\omega}\Big\} + \lim_{\omega\to 0}\frac{\partial}{\partial\omega}\frac{R^{(1)}(\omega)}{R(\omega)+1} \\
&= ([\eta]+1)\lim_{\omega\to 0}\Big(\frac{1}{\cos\omega - 1} + \frac{2}{\omega^2}\Big) + \lim_{\omega\to 0}\frac{R^{(2)}(\omega)(R(\omega)+1) - (R^{(1)(\omega)})^2}{(R(\omega)+1)^2} \\
&= -\frac{1}{6}([\eta]+1).
\end{aligned}$$

Hence $[\eta]$, and therefore η, are identifiable. The identifiability of σ^2 follows from the identifiability of η.

We now verify (d). By Chambers (1996), $\tilde{f}(\omega;\theta) = f(\omega;\theta) + \mathrm{O}(M^{-2\eta-1})$, for $M \to \infty$, uniformly for $\theta \in \Theta$. Let $\theta_1 = \eta$ and $\theta_2 = \sigma^2$. It can then be verified that for $i, j \in \{1, 2\}$, $\partial\tilde{f}(\omega;\theta)/\partial\theta_i = \partial f(\omega;\theta)/\partial\theta_i + \mathrm{O}(M^{-2\eta-1})$, and $\partial^2\tilde{f}(\omega;\theta)/\partial\theta_i\partial\theta_j = \partial^2 f(\omega;\theta)/\partial\theta_i\partial\theta_j + \mathrm{O}(M^{-2\eta-1})$. Therefore,

$$\frac{\partial^2\tilde{l}_M(\theta)}{\partial\theta_i\partial\theta_j} = \frac{\partial^2\tilde{l}(\theta)}{\partial\theta_i\partial\theta_j} + O_p(M^{-2\eta-1}).$$

This proves (d). Parts (b) and (c) can be similarly proved. This completes the proof of the theorem.

References

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

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

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

Beran, J., Bhansali, R. J. and Ocker, D. (1998). On unified model selection for stationary and nonstationary short-and long-memory autoregressive processes. *Biometrika* **85**, 921-934.

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

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

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

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

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

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

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.

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

Li , W. K. (2004). *Diagnostic Checks in Time Series.* Chapman & Hall, New York.

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

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65**, 209-255.

Makridakis, S, Wheelwright, S. and Hyndman, R. (1998). *Forecasting: Methods and Applications.* Wiley, New York.

McCleary, R. and Hay, A. (1980). *Applied Time Series Analysis for the Social Sciences.* Sage Publications, Beverly Hills, CA.

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

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

Rossana, R. J. and Seater, J. J. (1995). Temporal aggregation and economic time series. *J. Bus. Econom. Statist.* **13**, 441-451.

Tiao, G. C. (1972). Asymptotic behavior of temporal aggregates of time series. *Biometrika* **59**, 525-531.

Tsai, H. and Chan, K. S. (2005a). Maximum likelihood estimation of linear continuous time long memory processes. with discrete time data. *J. Roy. Statist. Soc. Ser. B* **67**, 703-716.

Tsai, H. and Chan, K. S. (2005b). Quasi-maximum likelihood estimation for a class of continuous-time long-memory processes. *J. Time Ser. Anal.* **26**, 691-713.

Tsai, H. and Chan, K. S. (2005c). Temporal aggregation of stationary and nonstationary continuous-time processes. *Scand. J. Statist.* **32**, 583-597.

Tsai, H. and Chan, K. S. (2005d). Temporal aggregation of stationary and nonstationary discrete-time processes. *J. Time Ser. Anal.* **26**, 613-624.

Working, H. (1960). Note on the correlation of first differences of averages in a random chain. *Econometrica* **28**, 916-918.

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

E-mail: htsai@stat.sinica.edu.tw

(Received March 2004; accepted November 2004)