Inference of Seasonal Long-memory Aggregate Time Series

BY Kung-Sik CHAN and Henghsiu TSAI

March 15, 2008

Abstract. Time-series data with regular and/or seasonal long-memory are often aggregated before analysis. Often, the aggregation scale is large enough to remove any short-memory components of the underlying process but too short to eliminate seasonal patterns of much longer periods. In this paper, we investigate the limiting correlation structure of aggregate time series within an intermediate asymptotic framework that attempts to capture the aforementioned sampling scheme. In particular, we study the autocorrelation structure and the spectral density function of aggregates from a discrete-time process. The underlying discrete-time process is assumed to be a stationary Seasonal AutoRegressive Fractionally Integrated Moving-Average (SARFIMA) process, after suitable number of differencing if necessary, and the seasonal periods of the underlying process are multiples of the aggregation size. We derive the limit of the normalized spectral density function of the aggregates, with increasing aggregation. The limiting aggregate (seasonal) long-memory model may then be useful for analyzing aggregate time-series data, which can be estimated by maximizing the Whittle likelihood. We prove that the maximum Whittle likelihood estimator is consistent and asymptotically normal, and study its finite-sample properties through simulation. The efficacy of the proposed approach is illustrated by a real-life internet traffic example.

KEY WORDS: Asymptotic normality; Consistency; Quasi-maximum likelihood estimation; Seasonal auto-regressive fractionally integrated moving-average models; Spectral density; Whittle likelihood.

1 Introduction

Data are often aggregated before analysis, for example, 1-minute data aggregated into half-hourly data or daily data aggregated into monthly data. On a fine sampling scale, many time series are of long memory in the sense that their spectral density function admit a pole at the zero frequency. A popular class of discrete time long memory processes are autoregressive fractionally integrated moving average (ARFIMA) models (see Granger and Joyeux 1980; Hosking 1981). Man and Tiao (2006) and Tsai and Chan (2005) showed that temporal aggregation preserves the long-memory parameter of the underlying ARFIMA process. Ohanissian, Russell, and Tsay (2008) made use of this property in developing a test for long-memory. Furthermore, as the extent of aggregation increases to infinity, the limiting model retains the long-memory parameter of the original process, whereas the short-memory components vanish.

In practice, the underlying process may admit seasonal long memory in that its

---

Kung-Sik Chan is Professor, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242 (E-mail: kung-sik-chan@uiowa.edu). Henghsiu Tsai is Associate Research Fellow, Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan (E-mail: htsai@stat.sinica.edu.tw). We thank Wilfredo Palma for sending us the aggregate http request data. KSC thanks the National Science Foundation (DMS-0405267) for partial support. HT thanks the National Science Council Taiwan (NSC 95-2118-M-001-027) for partial support.

spectral density function may have poles at certain non-zero frequencies. Such data may be modeled as some Seasonal Auto-Regressive Fractionally Integrated Moving-Average (SARFIMA) process, see Section 2. If the aggregation interval is much larger than the largest seasonal period, aggregation will intuitively merge the seasonal long-memory components with the regular long-memory component and eliminate the regular or seasonal short-memory components of the raw data. For example, within the framework of ARIMA models, Wei (1978) showed that aggregation removes seasonality if the frequency of aggregation is larger or the same as the seasonal frequency.

On the other hand, if the aggregation interval is large but is just some fraction of the seasonal periods of the original data, the aggregates may be expected to keep the seasonal short- and long-memory pattern, albeit with different periods. For many data, the latter scenario may be more relevant for analysis. For example, aggregating 1-minute data into half-hourly data may remove the short memory component on the minute scale but the daily or monthly correlation pattern of the raw data may persist in the aggregates.

Here, our purposes are twofold. First, we study the intermediate asymptotics of aggregating a SARFIMA process. In particular, we derive the limiting (normalized) spectral density function of an aggregated SARFIMA process via the asymptotic framework where the seasonal periods of the SARFIMA model are multiples of the aggregation interval and the aggregation interval is large. While the original time series is assumed to be a SARFIMA process, the limiting result is robust to the exact form of the short-memory and the regular long-memory components. The limiting spectral density functions then define a class of models suitable for analyzing aggregate time series that may have regular or seasonal long-memory and short-memory components. Second, we derive the large-sample properties of the Quasi-Maximum Likelihood Estimator (QMLE) of the limiting aggregate SARFIMA model, obtained by maximizing the Whittle likelihood.

The rest of the paper is organized as follows. The SARFIMA model is reviewed in Section 2. In Section 3, we derive the limiting spectral density function of an aggregate SARFIMA process, under the intermediate asymptotic framework. Qusai-maximum likelihood estimation of the limiting aggregate SARFIMA model and its large-sample properties are discussed in Section 4. We report some results on the empirical performance of the QMLE in Section 5, and illustrate the use of the limiting aggregate SARFIMA model with a real application in Section 6. We conclude in Section 7. All proofs are deferred to the appendix.

2 Seasonal autoregressive fractionally integrated moving average models

Over the last two decades, studies about long memory and cyclical behavior have gained popularity and importance in many fields including hydrology, telecommunication engineering and econometrics. See, for example, Porter-Hudak (1990), Ray (1993), Hassler and Wolters (1995), Montanari, Rosso, and Taqqu (2000), and Palma and Chan (2005). In particular, Porter-Hudak (1990) introduced the SARFIMA model and illustrated its use in studying the long-memory and cyclical behavior in the U.S. monetary aggregates. See also Palma and Bondon (2003), and Palma and Chan (2005). We now review the SARFIMA model, with particular attention to the spectrum and the long-memory properties of the SARFIMA models. Let $\{Y_t, t = 0, \pm 1, \pm 2, \cdots\}$ be a seasonal autoregressive fractionally integrated moving average (SARFIMA) model with multiple periods $s_1, ..., s_c$

$$\phi(B)(1-B)^d \prod_{i=1}^{c} \Phi_i(B^{s_i})(1-B^{s_i})^{D_i} Y_t = \theta(B) \prod_{i=1}^{c} \Theta_i(B^{s_i})\varepsilon_t, \tag{1}$$

where d and $D_i, i = 1, \ldots, c$, are real numbers, $s_c > s_{c-1} > \cdots > s_1 > 1$ are integers, $\{\varepsilon_t\}$ is an uncorrelated sequence of random variables with zero mean and common, finite variance $\sigma_\varepsilon^2 > 0$, $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$, $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$, and for $i = 1, ..., c$, $\Phi_i(z) = 1 - \Phi_{i,1} z - \cdots - \Phi_{i,P_i} z^{P_i}$, $\Theta_i(z) = 1 + \Theta_{i,1} z + \cdots + \Theta_{i,Q_i} z^{Q_i}$, B is the backward shift operator, and $(1-B)^d$ is defined by the binomial series expansion

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

where $\Gamma(\cdot)$ is the gamma function. Stationarity of $\{Y_t\}$ requires $D_i < 1/2$ for all i and $d + \sum_{i=1}^{c} D_i < 1/2$, see Palma and Bondon (2003). We assume that none of the roots of $\phi(\cdot)$ and $\Phi_i(\cdot)$, $i = 1, ..., c$, match any roots of $\theta(\cdot)$ and $\Theta_i(\cdot)$, $i = 1, ..., c$. Moreover, all roots of of the above polynomials are assumed to lie outside the unit circle. The conditions on the roots, the fractional orders d and D's ensure that $\{Y_t\}$ is stationary and the model is identifiable. It can be readily checked that the spectral density of $\{Y_t\}$ equals, for $-\pi < \omega \leq \pi$,

$$\begin{aligned} h(\omega) &= \frac{\sigma^2}{2\pi} \left| \frac{\theta(\exp(j\omega))}{\phi(\exp(j\omega))} \right|^2 |1 - \exp(i\omega)|^{-2d} \prod_{j=1}^{c} \left| \frac{\Theta_j(\exp(is_j\omega))}{\Phi_j(\exp(is_j\omega))} \right|^2 \prod_{j=1}^{c} |1 - \exp(is_j\omega)|^{-2D_j} \\ &= \frac{\sigma^2}{2\pi} \left| \frac{\theta(\exp(j\omega))}{\phi(\exp(j\omega))} \right|^2 \left| 2\sin\left(\frac{\omega}{2}\right) \right|^{-2d} \prod_{j=1}^{c} \left| \frac{\Theta_j(\exp(is_j\omega))}{\Phi_j(\exp(is_j\omega))} \right|^2 \prod_{j=1}^{c} \left| 2\sin\left(\frac{s_j\omega}{2}\right) \right|^{-2D_j} \\ &= \frac{\sigma^2}{2\pi} \left| \frac{\theta(\exp(j\omega))}{\phi(\exp(j\omega))} \right|^2 \left| 2\sin\left(\frac{\omega}{2}\right) \right|^{-2\delta_0} \prod_{j=1}^{c} \left| \frac{\Theta_j(\exp(is_j\omega))}{\Phi_j(\exp(is_j\omega))} \right|^2 \\ &\quad \times \prod_{j=1}^{c} \prod_{k=1}^{\tau_j} \left| \Big(\exp(i\nu_{jk}) - \exp(i\omega)\Big)\Big(\exp(-i\nu_{jk}) - \exp(i\omega)\Big) \right|^{-2\delta_{jk}}, \end{aligned} \tag{2}$$

where $\delta_0 = d + D_1 + \cdots + D_c$; $\tau_j = [s_j/2]$, the greatest integer $\leq s_j/2$; $\nu_{jk} = 2\pi k/s_j$, for $j = 1, ..., c$, and $k = 1, ..., \tau_j$; $\delta_{jk} = D_j$, for $k = 1, ..., \tau_j - 1$, $\delta_{j\tau_j} = D_j$ if $s_j = 2\tau_j + 1$, and $\delta_{j\tau_j} = D_j/2$ if $s_j = 2\tau_j$. The last equality in (2) follows from the factorization of $x^{s_j} - 1$ in terms of its roots that consist of 1 and $\exp(\pm\nu_{jk}), k = 1, \ldots, \tau_j$. From (2), we see that, as $\omega \to 0$, the spectral density $f(\omega) = O(|\omega|^{-2d-2D_1-\cdots-2D_c})$, whereas for $j = 1, ..., c$, $k = 1, ..., \tau_j$, as $\omega \to \nu_{jk}$, $f(\omega) = O(|\omega - \nu_{jk}|^{-2D_j})$.

From the above discussion, it is clear that the SARFIMA process defined in (1) subsumes many important time-series models including pure long-memory processes, pure seasonal long-memory processes and seasonal ARMA processes. Given our interest in long-memory processes, throughout this paper, the parameters d and the D_j's are restricted by the inequality constraints: $0 \leq d+D_1+\cdots+D_c < 1/2$, and $0 \leq D_j < 1/2$, for $j = 1, ..., c$.

3 Aggregates of SARFIMA models

For non-stationary data, we assume that, after suitable regular and/or seasonal differencing, the data become stationary and follow some stationary SARFIMA model. Specifically, let r and R_j, $j = 1, ..., c$, be non-negative integers and $\{Y_t, t = 0, \pm 1, \pm 2, \cdots\}$ a time series such that $(1-B)^r(1-B^{s_1})^{R_1}\cdots(1-B^{s_c})^{R_c}Y_t$ is a stationary SARFIMA model defined by equation (1). Therefore, $\{Y_t\}$ satisfies the difference equation

$$\phi(B)(1-B)^{r+d}\prod_{j=1}^{c}\Phi_j(B^{s_j})(1-B^{s_j})^{R_j+D_j}Y_t = \theta(B)\prod_{j=1}^{c}\Theta_j(B^{s_j})\varepsilon_t, \tag{3}$$

which is referred to as the SARFIMA$(p, r+d, q) \times (P_1, R_1+D_1, Q_1)_{s_1} \times \cdots \times (P_c, R_c+D_c, Q_c)_{s_c}$ model.

Let $m \geq 2$ be an integer and

$$X_T^m = \sum_{k=m(T-1)+1}^{mT} Y_k$$

be the non-overlapping m-temporal aggregates of $\{Y_t\}$. Let $\nabla = 1 - B$ be the first difference operator, and $\nabla_s = 1 - B^s$ the lag-s difference operator. Let $D = (D_1, ..., D_c)$,

$R = (r, R_1, ..., R_c)$, $\xi = (d; D_j, j = 1, ..., c; \Phi_{i,j}, i = 1, ..., c, j = 1, ..., P_i; \Theta_{i,j}, i = 1, ..., c, j = 1, ..., Q_i)$, and assume $s_i = mz_i$, $i = 1, ..., c$. Below we derive the spectral density of the aggregates, and the limit of the normalized spectral densities with increasing aggregation. The normalization that makes the spectral densities integrate to 1 is necessary lest the variance of the aggregates may increase without bound with increasing aggregation.

THEOREM 1 *Assume that $\{Y_t\}$ satisfies the difference equation defined by (3).*
(a) For $r \geq 0$, $R_i \geq 0$, $i = 1, ..., c$, and $m = 2h + 1$, the spectral density function of $\{\nabla^r \nabla_{z_1}^{R_1} \cdots \nabla_{z_c}^{R_c} X_T^m\}$ is given by

$$\begin{aligned} f_{\xi,m}(\omega) &= \frac{1}{m} \left|2\sin\left(\frac{\omega}{2}\right)\right|^{2r+2} \prod_{j=1}^{c} \left|2\sin\left(\frac{z_j\omega}{2}\right)\right|^{-2D_j} \prod_{j=1}^{c} \left|\frac{\Theta_j(\exp(iz_j\omega))}{\Phi_j(\exp(iz_j\omega))}\right|^2 \\ &\quad \times \sum_{k=-h}^{h} \left|2\sin\left(\frac{\omega + 2k\pi}{2m}\right)\right|^{-2r-2d-2} g\left(\frac{\omega + 2k\pi}{m}\right), \end{aligned} \tag{4}$$

where $g(\omega) = \sigma^2(2\pi)^{-1}|\theta(\exp(i\omega))|^2|\phi(\exp(i\omega))|^{-2}$ and $-\pi < \omega \leq \pi$.
If $m = 2h$, the spectral density is given by equation (4) with the summation ranging from $-h+1$ to h for $-\pi < \omega \leq 0$ and from $-h$ to $h-1$ for $0 < \omega \leq \pi$.
(b) As $m \to \infty$, the normalized spectral density function of $\{\nabla^r \nabla_{z_1}^{R_1} \cdots \nabla_{z_c}^{R_c} X_T^m\}$ converges to

$$\begin{aligned} f_{\xi}(\omega) &= K_\xi \left|\sin\left(\frac{\omega}{2}\right)\right|^{2r+2} \prod_{j=1}^{c} \left|\sin\left(\frac{z_j\omega}{2}\right)\right|^{-2D_j} \prod_{j=1}^{c} \left|\frac{\Theta_j(\exp(iz_j\omega))}{\Phi_j(\exp(iz_j\omega))}\right|^2 \\ &\quad \times \sum_{k=-\infty}^{\infty} |\omega + 2k\pi|^{-2r-2d-2}, \end{aligned} \tag{5}$$

where K_ξ is the normalization constant ensuring that $\int_{-\pi}^{\pi} f_\xi(\omega)d\omega = 1$.

Remark 1: Note that $z_c > z_{c-1} > \cdots > z_1 \geq 1$. For $j = 1, ..., c$, $k = 0, 1, ..., [z_j/2]$, let $\omega_{jk} = \nu_{j(mk)} = 2\pi k/z_j$, then both $m^{-2r-2d-1} f_{\xi,m}$ and f_ξ are of order $O(|\omega|^{-2d-2D_1-\cdots-2D_c})$, for $\omega \to 0$, and of order $O(|\omega - \omega_{jk}|^{-2D_j})$, for $\omega \to \omega_{jk}$, $j = 1, ..., c$, $k = 1, ..., [z_j/2]$.

Remark 2: If $z_1 = 1$, the corresponding seasonal long-memory component is confounded with the regular long-memory component for the limiting aggregate process. Hence, without loss of generality, we shall set $D_1 = 0$ if $z_1 = 1$ in applications.

Remark 3: In order for the term $\sum_{k=-\infty}^{\infty} |\omega + 2k\pi|^{-2r-2d-2}$ to be well-defined, we assume $r + d > -1/2$.

Remark 4: As $m \to \infty$, the limiting autocorrelation function of $\{\nabla^r \nabla_{z_1}^{R_1} \cdots \nabla_{z_c}^{R_c} X_T^m\}$ is given by

$$\begin{aligned}\rho_\xi(h) \;&=\; C_\xi \int_0^\pi \cos(\omega h) \left|\sin\left(\frac{\omega}{2}\right)\right|^{2r+2} \prod_{j=1}^{c} \left|\sin\left(\frac{z_j \omega}{2}\right)\right|^{-2D_j} \prod_{j=1}^{c} \left|\frac{\Theta_j(\exp(iz_j\omega))}{\Phi_j(\exp(iz_j\omega))}\right|^2 \\ &\qquad \times \sum_{k=-\infty}^{\infty} |\omega + 2k\pi|^{-2r-2d-2} d\omega,\end{aligned}$$

where C_ξ is a normalization constant ensuring that $\rho_\xi(0) = 1$.

Remark 5: If $r = 0$, then the limiting model of the aggregates of $\{Y_t\}$ is simply a SARFIMA$(P_1, R_1 + D_1, Q_1)_{z_1} \times \cdots \times (P_c, R_c + D_c, Q_c)_{z_c}$ process with fractional Gaussian noise as the driving noise process, where the self-similarity parameter (Hurst parameter) of the underlying fractional Gaussian process equals $H = d + 1/2$. See Beran (1994) for definition of the fractional Gaussian noise.

4 Quasi-maximum likelihood estimator and its large sample properties

We are interested in applying the long-memory limiting aggregate process derived in Section 3 to data analysis. For this purpose, we assume (i) $0 \le d + D_1 + \cdots + D_c < 1/2$ and (ii) $0 \le D_j < 1/2$ for $j = 1, ..., c$. The limiting aggregate process is of long memory regularly or seasonally if either $0 < d + D_1 + \cdots + D_c < 1/2$ or $0 < D_j < 1/2$ for some $j \in \{1, ..., c\}$. Note that this implicitly implies $-c/2 < d < 1/2$. We also introduce the parameter σ to account for the variance of the data. Furthermore, we assume z_j, $j = 1, ..., c$, are known. Consider a time series $\{Y_i\}_{i=1-\delta}^N$, where δ is a positive integer to be defined below, such that, conditional on $\{Y_i\}_{i=1-\delta}^0$, $\{\nabla^r \nabla_{z_1}^{R_1} \cdots \nabla_{z_c}^{R_c} Y_i\}_{i=1}^N$ is a stationary process with its spectral density defined by

$$\begin{aligned} f(\omega; \xi, R, \sigma^2) &= \sigma^2 \left|\sin\left(\frac{\omega}{2}\right)\right|^{2r+2} \prod_{j=1}^{c} \left|\sin\left(\frac{z_j\omega}{2}\right)\right|^{-2D_j} \prod_{j=1}^{c} \left|\frac{\Theta_j(\exp(iz_j\omega))}{\Phi_j(\exp(iz_j\omega))}\right|^2 \\ &\quad \times \sum_{k=-\infty}^{\infty} |\omega + 2k\pi|^{-2r-2d-2}, \ \ -\pi < \omega \leq \pi, \end{aligned} \tag{6}$$

where $\delta = \max_r + \sum_{i=1}^c z_i \cdot \max_{R_i}$; $\max_r$ and $\max_{R_i}$, $i = 1, ..., c$, are the largest possible values of r and R_i, $i = 1, ..., c$, respectively, which we will consider in simulation studies and real data analysis in Sections 5 and 6. That is, the maximum likelihood estimators $\hat{r}$ and $\hat{R}_i$, $i = 1, ..., c$, satisfy the conditions that $\hat{r} \in \{0, ..., \max_r\}$ and $\hat{R}_i \in \{0, ..., \max_{R_i}\}$, for $i = 1, ..., c$. The auto-covariance function corresponding to the spectral density function defined by (6) equals, for any integer h,

$$\begin{aligned} \gamma_{\xi,R,\sigma^2}(h) &= 2\sigma^2 \int_0^{\pi} \cos(\omega h) \left|\sin\left(\frac{\omega}{2}\right)\right|^{2r+2} \prod_{j=1}^{c} \left|\sin\left(\frac{z_j\omega}{2}\right)\right|^{-2D_j} \prod_{j=1}^{c} \left|\frac{\Theta_j(\exp(iz_j\omega))}{\Phi_j(\exp(iz_j\omega))}\right|^2 \\ &\quad \times \sum_{k=-\infty}^{\infty} |\omega + 2k\pi|^{-2r-2d-2} d\omega. \end{aligned}$$

It can be easily checked that, conditional on $\{Y_i\}_{i=1-\delta}^0$, the joint distributions of $\{\nabla^r \nabla_{z_1}^{R_1} \cdots \nabla_{z_c}^{R_c} Y_i\}_{i=1}^N$ and $\{Y_i\}_{i=1}^N$ are the same. Therefore, conditional on $\{Y_i\}_{i=1-\delta}^0$, the (negative) log-likelihood function of $\{Y_i\}$ can be approximated by the (negative) Whittle log-likelihood function (see Hosoya, 1996)

$$-\tilde{l}(\xi, R, \sigma^2) = \sum_{j=1}^{T} \left\{ \log f(\omega_j; \xi, R, \sigma^2) + \frac{I_N(\omega_j; R)}{f(\omega_j; \xi, R, \sigma^2)} \right\}, \tag{7}$$

where $\omega_j := 2\pi j/N \in (0,\pi)$ are the Fourier frequencies, T is the largest integer $\leq (N-1)/2$, $I_N(\omega;R) = \left|\sum_{j=1}^N U_j(R)\exp(ij\omega)\right|^2/(2\pi N)$, and $U_i(R) = \nabla^r\nabla_{z_1}^{R_1}\cdots\nabla_{z_c}^{R_c}Y_i$, $i=1,...,N$. In (7), the computation of $f(\omega_j;\xi,R,\sigma^2)$ requires evaluation of an infinite sum. Here, we adopt the method of Chambers (1996) to approximate $f(\omega;\xi,R,\sigma^2)$ by

$$\begin{aligned}&\tilde{f}(\omega;\xi,R,\sigma^2)\\ =\;&\sigma^2\left|\sin\left(\frac{\omega}{2}\right)\right|^{2r+2}\prod_{j=1}^{c}\left|\sin\left(\frac{z_j\omega}{2}\right)\right|^{-2D_j}\prod_{j=1}^{c}\left|\frac{\Theta_j(\exp(iz_j\omega))}{\Phi_j(\exp(iz_j\omega))}\right|^2 h(\omega;\xi,R),\end{aligned} \tag{8}$$

where $h(\omega;\xi,R) = \{2\pi(2r+2d+1)\}^{-1}\{(2\pi M-\omega)^{-2r-2d-1}+(2\pi M+\omega)^{-2r-2d-1}\} + \sum_{k=-M}^{M}|\omega+2k\pi|^{-2r-2d-2}$ for some large integer M. By routine analysis, it can be shown that, under the conditions stated in Theorem 2, the approximation error of $h(\omega;R,\xi)$ to the infinite sum is of order $O(M^{-2r-2d-2})$. Also, the approximation error of the first partial derivative with respect to d is of order $O(M^{-2r-2d-1-\epsilon})$, for any positive ϵ less than 1. These error rates guarantee that if the truncation parameter M increases with the sample size at a suitable rate, then the truncation has negligible effects on the asymptotic distribution of the estimator, see Theorem 2 below. Replacing $f(\omega_j;\xi,R,\sigma^2)$ with $\tilde{f}(\omega_j;\xi,R,\sigma^2)$ and letting $\tilde{g}(\omega_j;\xi,R) = \tilde{f}(\omega_j;\xi,R,\sigma^2)/\sigma^2$, the (negative) Whittle log-likelihood function (7) now becomes

$$-\tilde{l}(\xi,R,\sigma^2) = \sum_{j=1}^{T}\left\{\log\sigma^2 + \log\tilde{g}(\omega_j;\xi,R) + \frac{I_N(\omega_j;R)}{\sigma^2\tilde{g}(\omega_j;\xi,R)}\right\}. \tag{9}$$

Differentiating (9) with respect to σ^2 and equating to zero gives

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{j=1}^{T}\frac{I_N(\omega_j;R)}{\tilde{g}(\omega_j;\xi,R)}. \tag{10}$$

Substituting (10) into (9) yields the objective function

$$-\tilde{l}(\xi,R) = \sum_{j=1}^{T}\log\tilde{g}(\omega_j;\xi,R) + T\log\left(\sum_{j=1}^{T}\frac{I_N(\omega_j;R)}{\tilde{g}(\omega_j;\xi,R)}\right) + C, \tag{11}$$

where $C = T - T\log T$. The objective function is minimized with respect to ξ and R to get the quasi-maximum likelihood estimators (QMLEs) $\hat{\xi}$ and $\hat{R}$; the estimator $\hat{\sigma}^2$ is then calculated by (10). Specifically, the quasi-maximum likelihood estimators $\hat{\xi}$ and $\hat{R}$ are computed based on equation (11) using the following procedure (Recall that $0 \le D_j < 1/2$, for $j = 1, ..., c$, and $0 \le d + D_1 + \cdots + D_c < 1/2$). For each $r \in \{0, ..., \max_r\}$ and $R_i \in \{0, ..., \max_{R_i}\}$, for $i = 1, ..., c$, we first find the local maximum likelihood estimator of ξ in the range that $0 \le D_j < 1/2$, $j = 1, ..., c$, and $r \le d + D_1 + \cdots + D_c < r + 1/2$. In our experiments, we let $\max_r = \max_{R_1} = \cdots = \max_{R_c} = 2$. These local maximum likelihood estimators are then used to find the global maximum likelihood estimator of ξ, as well as those of r and R's.

For simplicity, let $\theta = (\xi, \sigma^2)$, and $\hat{\theta} = (\hat{\xi}, \hat{\sigma}^2)$ be the quasi-maximum likelihood estimator that minimizes the (negative) Whittle log-likelihood function (9). Below, we derive the large-sample distribution of the QMLE.

THEOREM 2 *Let the data $Y = \{Y_i\}_{i=1}^N$ be such that $\{\nabla^r \nabla_{z_1}^{R_1} \cdots \nabla_{z_c}^{R_c} Y_i\}_{i=1}^N$ is sampled from a stationary Gaussian long-memory process with the spectral density given by (6). Let the quasi-maximum likelihood estimator $\hat{\theta} \in \Theta$, a compact parameter space, and the true parameter θ_0 be in the interior of the parameter space. Assume that each component of $R = (r, R_1, ..., R_c)$ is known to be between 0 and some integer K. Let r_0 and d_0 be the true values of r and d, and the truncation parameter M increase with the sample size so that $M \to \infty$. Then the QMLE $\hat{R}$ and $\hat{\theta}$ are consistent. Moreover, if $\sqrt{N} M^{-2r_0 - 2d_0 - 1} \to 0$ as $N \to \infty$, then $\sqrt{N}(\hat{\theta} - \theta_0)$ converges in distribution to a normal random vector with mean 0 and covariance matrix $\Gamma(\theta_0)^{-1}$ with*

$$\Gamma(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} (\bigtriangledown \log f(\omega; R, \theta))(\bigtriangledown \log f(\omega; R, \theta))' d\omega, \tag{12}$$

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

5 Simulation studies

In this section, we report some finite sample performance of the quasi-maximum likelihood estimator for models simulated from stationary and non-stationary Gaussian processes such that $\{\nabla^r \nabla_z^R Y_i\}_{i=-\delta}^N$ is a stationary process with its spectral density defined by

$$\begin{aligned} & f(\omega; r, d, D, \sigma^2) \\ = \; & \sigma^2 \left|\sin\left(\frac{\omega}{2}\right)\right|^{2r+2} \left|\sin\left(\frac{z\omega}{2}\right)\right|^{-2D} \sum_{k=-\infty}^{\infty} |\omega + 2k\pi|^{-2r-2d-2}, \end{aligned} \qquad (13)$$

and

$$\begin{aligned} & f(\omega; r, d, D, \Phi_{1,1}, \Theta_{1,1}, \sigma^2) \\ = \; & \sigma^2 \left|\sin\left(\frac{\omega}{2}\right)\right|^{2r+2} \left|\sin\left(\frac{z\omega}{2}\right)\right|^{-2D} \left|\frac{\Theta_1(\exp(iz\omega))}{\Phi_1(\exp(iz\omega))}\right|^2 \sum_{k=-\infty}^{\infty} |\omega + 2k\pi|^{-2r-2d-2}, \end{aligned} \qquad (14)$$

where $\Phi_1(z) = 1 - \Phi_{1,1}z$, $\Theta_1(z) = 1 + \Theta_{1,1}z$, $-\pi < \omega \leq \pi$, and $\delta = \max_r + \max_R z$. We consider $\sigma = 2$, and $z = 10$. The true values of the parameters are given in Tables 1 and 2. We used the method of Davies and Harte (1987) to simulate the model. The sample sizes considered are $N = 512$ and $N = 1{,}024$. All the computations in this and the following section were performed using Fortran code with IMSL subroutines. In our experiments, we chose both $\max_r$ and $\max_R$ to be 2. The averages and the standard errors of 1,000 replicates of the estimators for models (13) and (14) are summarized in Table 1 and 2, respectively. The asymptotic standard errors of the parameter estimators computed from $\Gamma(\theta)$ defined in equation (12) are also given in the tables. Again, $f(\omega; \xi, R, \sigma^2)$ is approximated by $\tilde{f}(\omega; \xi, R, \sigma^2)$ defined in (8). The value of M used in the computation of $h(\omega; \xi, R)$, defined below (8), is set to be 100. We have also tried $M = 1000$ in the program and the results are essentially the same.

Table 1: Averages (standard deviation) [asymptotic standard deviation] of 1,000 simulations of the quasi-maximum likelihood estimators of the parameters d, D, σ, r, and R for model (13).

parameter	true value	N=512 (z=10)	N=1,024 (z=10)
d	0.1	0.0988(0.0290)[0.0284]	0.0992(0.0213)[0.0201]
D	0.3	0.3214(0.0412)[0.0346]	0.3139(0.0276)[0.0244]
d+D	0.4	0.4203(0.0480)[0.0429]	0.4131(0.0341)[0.0303]
σ	2.0	1.9952(0.0838)[0.0820]	1.9944(0.0588)[0.0580]
r	0	0.0000(0.0000)[0.0000]	0.0000(0.0000)[0.0000]
R	0	0.0000(0.0000)[0.0000]	0.0000(0.0000)[0.0000]
d	0.2	0.1995(0.0324)[0.0308]	0.1998(0.0224)[0.0218]
D	0.2	0.2041(0.0382)[0.0346]	0.2041(0.0255)[0.0245]
d+D	0.4	0.4036(0.0473)[0.0443]	0.4039(0.0328)[0.0313]
σ	2.0	2.0044(0.0844)[0.0800]	2.0005(0.0594)[0.0566]
r	1	1.0000(0.0000)[0.0000]	1.0000(0.0000)[0.0000]
R	1	1.0000(0.0000)[0.0000]	1.0000(0.0000)[0.0000]

Table 1 shows that for models without a seasonal short memory component (defined by Equation (13)), the estimates are quite good in terms of biases and variances for both sample sizes 512 and 1,024. Table 2 shows that for models with a seasonal short memory component (defined by Equation (14)), the biases and the variances of the estimates tend to be larger compared with models without a seasonal short memory component, and the biases and the variances become smaller with increasing sample size, which is consistent with the theory developed in the previous section.

Table 2: Averages (standard deviation) [asymptotic standard deviation] of 1,000 simulations of the quasi-maximum likelihood estimators of the parameters d, D, σ, r, R, $\Phi_{1,1}$, and $\Theta_{1,1}$ for model (14).

parameter	true value	N=512 (z=10)	N=1,024 (z=10)
d	-0.1	-0.1021(0.0285)[0.0265]	-0.1013(0.0200)[0.0187]
D	0.3	0.3231(0.0652)[0.0507]	0.3184(0.0414)[0.0354]
$d+D$	0.2	0.2210(0.0699)[0.0554]	0.2171(0.0456)[0.0392]
$\Phi_{1,1}$	-0.3	-0.3058(0.1791)[0.1459]	-0.3134(0.1111)[0.1032]
$\Theta_{1,1}$	0.6	0.5874(0.1286)[0.1065]	0.5991(0.0798)[0.0753]
σ	2.0	2.0023(0.1125)[0.0988]	1.9951(0.0772)[0.0698]
r	0	0.0000(0.0000)[0.0000]	0.0000(0.0000)[0.0000]
R	0	0.0000(0.0000)[0.0000]	0.0000(0.0000)[0.0000]
d	0.1	0.0980(0.0349)[0.0308]	0.0990(0.0230)[0.0218]
D	0.25	0.2570(0.1057)[0.0798]	0.2578(0.0714)[0.0564]
$d+D$	0.35	0.3550(0.1081)[0.0844]	0.3568(0.0737)[0.0597]
$\Phi_{1,1}$	0.3	0.2805(0.1805)[0.1484]	0.2872(0.1185)[0.1049]
$\Theta_{1,1}$	0.2	0.1715(0.2278)[0.0992]	0.1959(0.1037)[0.0701]
σ	2.0	2.0410(0.1626)[0.1278]	2.0162(0.1117)[0.0904]
r	1	1.0000(0.0000)[0.0000]	1.0000(0.0000)[0.0000]
R	1	1.0320(0.1761)[0.0000]	1.0040(0.0063)[0.0000]

6 Applications

In this section, we report some analysis of a time series of counts of http requests to a World Wide Web server at the University of Saskatchewan, Canada, between 1 June and 31 December in year 1995, within the framework of the limiting aggregate seasonal long-memory model and Whittle maximum likelihood estimation. The original data set consists of time stamps of 1-second resolution, which can be downloaded from `http://ita.ee.lbl.gov/html/contrib/Sask-HTTP.html`. Palma and Chan (2005) analyzed the 30-minute (non-overlapping) aggregates, i.e., each data point represents the total number of requests sent to the Sakastchewan's server within a 30-minute interval. To make the data more Gaussian and to stabilize their variances, Palma and Chan (2005) applied a logarithmic transformation to the aggregate data. See Figures 1, 2, and 3 for the time series plot, the sample autocorrelation function, and the periodogram of the transformed aggregate data. Their fitted model is a SARFIMA $(1, d, 1) \times (0, D, 0)_s$ model with $(\hat{d}, \hat{D}, \hat{\phi}, \hat{\theta}) = (0.076, 0.148, 0.917, 0.583)$. Although this model explains roughly two thirds of the total variance of the data, the residuals display significant autocorrelations at several lags, in particular, at lags from 40 to 50 (Figure 6(a) of Palma and Chan 2005), suggesting a lack of fit. Hsu and Tsai (2007) also analyzed the same data set, pointing out the presence of both daily and weekly persistency in the data. Indeed, observe that there are two major peaks in the periodogram: one at the origin and another at frequency $\omega = 2\pi \times 189/9074 = 0.1309$. These features indicate a possible seasonal long-memory process with $z = 48$, i.e. a daily pattern. The third peak is at frequency $\omega = 2\pi \times 27/9074 = 0.0187$, indicating a possible weekly pattern.

Here, we re-analyze this dataset with the limiting aggregate seasonal long-memory model defined by (6) with $c = 3$, $r = R_1 = R_2 = R_3 = 0$, $z_1 = 1$, $z_2 = 48$ (corresponding to daily effects), and $z_3 = 48 \times 7 = 336$ (corresponding to weekly effects). Our new approach may be justified as the 30-minute aggregation may well fall within the intermediate asymptotic framework studied in Section 3. As discussed in Remark 2 of Section 3, we assume $D_1 = 0$. Specifically, if $\{Y_t\}$ is the observed time series, the spectral density function of $\{Y_t\}$ can be written as

$$
\begin{aligned}
f_\xi(\omega) &= \sigma^2 \left|\sin\left(\frac{z_1\omega}{2}\right)\right|^2 \left|\sin\left(\frac{z_2\omega}{2}\right)\right|^{-2D_2} \left|\sin\left(\frac{z_3\omega}{2}\right)\right|^{-2D_3} \left|\frac{\Theta_1(\exp(i\omega))}{\Phi_1(\exp(i\omega))}\right|^2 \\
&\quad \times \left|\frac{\Theta_2(\exp(iz_2\omega))}{\Phi_2(\exp(iz_2\omega))}\right|^2 \left|\frac{\Theta_3(\exp(iz_3\omega))}{\Phi_3(\exp(iz_3\omega))}\right|^2 \sum_{k=-\infty}^{\infty} |\omega + 2k\pi|^{-2d-2}. \qquad (15)
\end{aligned}
$$

We have considered models of orders $(P_1, Q_1, P_2, Q_2, P_3, Q_3) = (P_1, Q_1, 0, 0, 0, 0)$ with $0 \le P_1 \le 2$, and $0 \le Q_1 \le 2$. The model with the smallest AIC (Akaike information criterion) is $(P_1, Q_1, P_2, Q_2, P_3, Q_3) = (2, 2, 0, 0, 0, 0)$. Before using this model for drawing inference, it is pertinent to examine its goodness of fit. This can be more conveniently carried out in frequency domain, by checking whether or not $\{I_N(\omega_j)/\tilde{g}(\omega_j; \hat{\xi}_M)\}$ are roughly independent and identically distributed. Consider the test statistic $W_{obs} = \max_{1\le p\le \tilde{N}} |T_p - p/\tilde{N}|$ (Priestley, 1981), where $\tilde{N} = [N/2]$ is the largest integer $\le N/2$, and $T_p = \sum_{j=1}^{p}\{I_N(\omega_j)/\tilde{g}(\omega_j; \hat{\xi}_M)\}/\sum_{j=1}^{\tilde{N}}\{I_N(\omega_j)/\tilde{g}(\omega_j; \hat{\xi}_M)\}$, for $p = 1, \ldots, \tilde{N}$. A large value of W_p signifies possible lack of fit. A bootstrap procedure (Hidalgo and Kreiss, 2006) in the frequency domain is used to compute the bootstrap p-value of W_{obs}.

Let $\tilde{Y}_t = (Y_t - \bar{Y})/\hat{\sigma}_Y$, where $\bar{Y} = \sum_{t=1}^{N} Y_t/N$ and $\hat{\sigma}_Y^2 = \sum_{t=1}^{N}(Y_t - \bar{Y})^2/(N-1)$, be the standardized data of $\{Y_t\}_{t=1}^{N}$. The bootstrap procedure runs as follows.

Step 1: Draw a random sample of size N with replacement from the empirical distribution of $\tilde{Y}_t$. Denote that sample as $Y^* = (Y_1^*, ..., Y_N^*)'$.

Step 2: For j=1,…,$\tilde{N}$, compute the bootstrap periodogram

$$I_N^*(\omega_j) = \tilde{f}(\omega_j; \hat{\xi}_M, \hat{\sigma}_M^2) I_{N,Y}^*(\omega_j),$$

where $I^*_{N,Y}(\omega_j) = |\sum_{t=1}^{N} Y_t^* \exp(it\omega_j)|^2/N$.

Step 3: Compute the bootstrap objective function

$$-\tilde{l}_M(\xi) = \sum_{j=1}^{T} \log \tilde{g}(\omega_j;\xi) + T \log \left(\sum_{j=1}^{T} \frac{I_N^*(\omega_j)}{\tilde{g}(\omega_j;\xi)} \right) + C,$$

where $C = T - T\log T$. The objective function is minimized with respect to ξ to get the bootstrap quasi-maximum likelihood estimator $\hat{\xi}^*_M$.

Step 4: Compute $T_p^* = \sum_{j=1}^{p}\{I_N^*(\omega_j)/\tilde{g}(\omega_j;\hat{\xi}^*_M)\}/\sum_{j=1}^{\tilde{N}}\{I_N^*(\omega_j)/\tilde{g}(\omega_j;\hat{\xi}^*_M)\}$, for $p = 1,\ldots,\tilde{N}$, and $W^* = \max_{1\le p\le \tilde{N}} |T_p^* - p/\tilde{N}|$.

The above four steps are run B times (here B = 1,000) to get W_1^*,...,W_B^* and $\hat{\xi}^*_{M,1}$,..., $\hat{\xi}^*_{M,B}$. The W_j^*'s are used to compute the bootstrap p-value of W_{obs} and the $\hat{\xi}^*_{M,j}$'s are used to compute the bootstrap confidence intervals of the parameters. Specifically, the bootstrap p-value of W_{obs} equals $\sharp\{j|W_{obs} < W_j^*, j = 1,...,B\}/B$, where $\sharp\{A\}$ denotes the cardinal number of the set A.

Alternatively, we could replace steps 1 and 2 by

Step $1'$: Draw independent exponential variables χ_1,...,$\chi_{\tilde{N}}$ with unit mean.

Step $2'$: For $j = 1,\ldots,\tilde{N}$, compute the bootstrap periodogram $I_N^*(\omega_j) = \tilde{f}(\omega_j;\hat{\xi}_M,\hat{\sigma}^2_M)\chi_j$.

The bootstrap p-values of W_{obs} based on steps 1-4 for the $(P_1,Q_1,P_2,Q_2,P_3,Q_3) = (2,2,0,0,0,0)$ model is 0.098, suggesting a good fit to the data.

The maximum likelihood estimates of the parameters and the 95% bootstrap confidence intervals based on steps 1-4 are summarized in Table 3. The asymptotic standard deviations and the asymptotic 95% confidence intervals are also included in Table 3. It is clear that the bootstrap confidence intervals of the parameters are comparable to their asymptotic counterparts. The confidence intervals of the parameters $d+D_2+D_3$, D_2 and D_3 indicate that the long-memory pattern, the daily seasonal long-memory pattern and the weekly seasonal long-memory pattern are all significant.

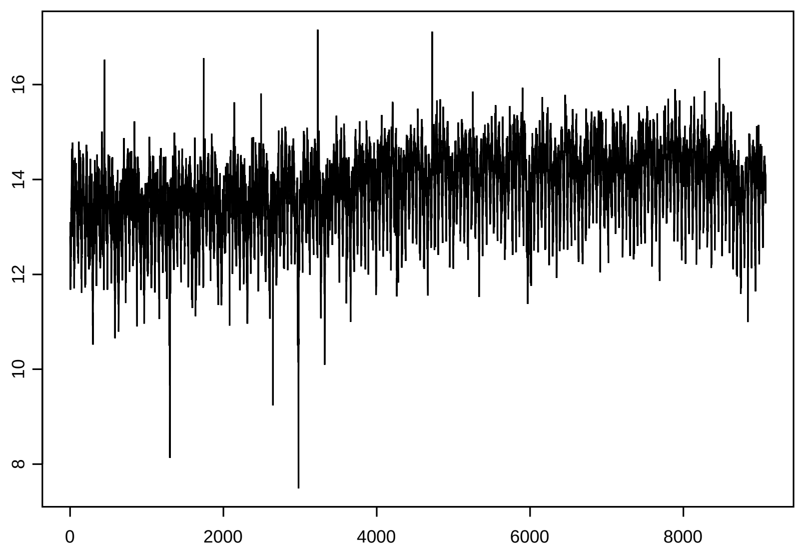

Figure 1: Log transformed 30-minute (non-overlapping) aggregates.

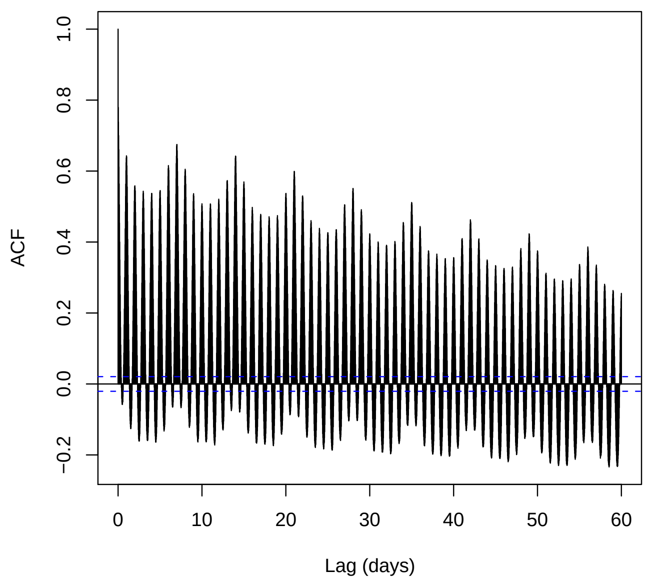

Figure 2: The autocorrelation function of the log transformed 30-minute (non-overlapping) aggregates.

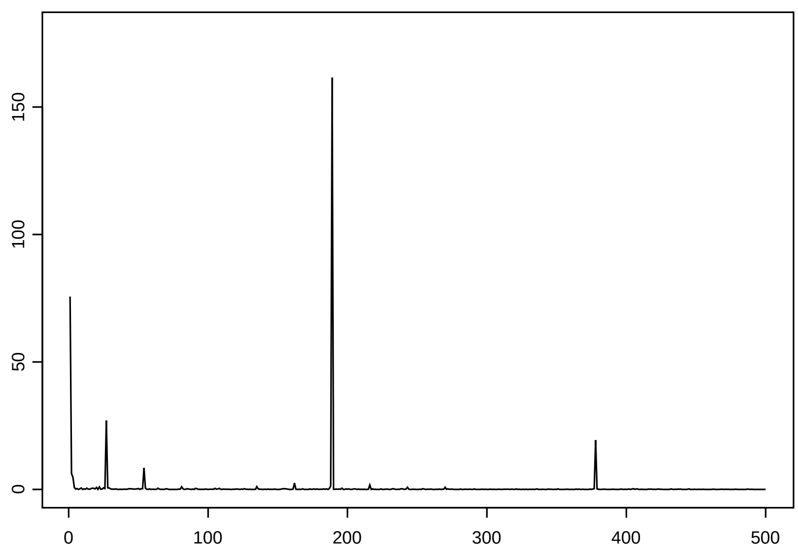

Figure 3: The periodgram of the log transformed 30-minute (non-overlapping) aggregates.

Table 3: Maximum likelihood estimates of the parameters of the model defined by equation (15), with $(P_1, Q_1, P_2, Q_2, P_3, Q_3) = (2, 2, 0, 0, 0, 0)$

Parameter	*Estimated value*	*Bootstrap* 95% *confidence interval*	*Asymptotic standard error*	*Asymptotic* 95% *confidence interval*
d	0.2326	(0.1268, 0.2608)	0.0436	(0.1471, 0.3181)
D_2	0.1274	(0.1085, 0.1429)	0.0083	(0.1111, 0.1437)
D_3	0.1271	(0.1083, 0.1430)	0.0083	(0.1108, 0.1434)
$d + D_2 + D_3$	0.4871	(0.3821, 0.5000)	0.0441	(0.4007, 0.5735)
$\phi_{1,1}$	1.1277	(0.8916, 1.5089)	0.1256	(0.8815, 1.3739)
$\phi_{1,2}$	-0.2610	(-0.5773, -0.0508)	0.1009	(-0.4588, -0.0632)
$\theta_{1,1}$	-1.1788	(-1.4586, -0.9315)	0.0936	(-1.3623, -0.9953)
$\theta_{1,2}$	0.3593	(0.1237, 0.5755)	0.0831	(0.1964, 0.5222)
σ	0.3117	(0.3017, 0.3194)	0.0051	(0.3017, 0.3217)

7 Concluding remarks

We have derived the limiting structure of the temporal aggregates of a (possibly non-stationary) SARFIMA model, with increasing aggregation, under the condition that the seasonal periods of the underlying process are multiples of the aggregation size. The limiting model can be estimated by maximizing the Whittle likelihood. We have also derived some large sample properties of the quasi-maximum likelihood estimators including consistency and asymptotic normality. Monte Carlo experiments showed that the QMLE performed well with sample size $N = 512$ or above. The efficacy of our proposed methodology is illustrated with an analysis of an internet traffic data. Model diagnostic using a bootstrap procedure in the frequency domain suggests a good fit. Future research problems include extending the model to include covariates and developing other tools for model diagnostics.

APPENDIX

Proof of Theorem 1. (a) First note that $\nabla^r \nabla_{s_1}^{R_1} \cdots \nabla_{s_c}^{R_c} Y_t$ admits the following spectral representation (Priestley 1981, equation 4.11.19): $\nabla^r \nabla_{s_1}^{R_1} \cdots \nabla_{s_c}^{R_c} Y_t = \int_{-\pi}^{\pi} \exp(it\omega) dZ(\omega)$ with $E|dZ(\omega)|^2 = h(\omega)d\omega$, where $h(\cdot)$ is given by equation (2). Also note that

$$
\begin{aligned}
& (1 - B^{z_1})^{R_1} \cdots (1 - B^{z_c})^{R_c} X_T^m \\
= & \sum_{k_1=0}^{R_1} \binom{R_1}{k_1} (-1)^{k_1} B^{k_1 z_1} \cdots \sum_{k_c=0}^{R_c} \binom{R_c}{k_c} (-1)^{k_c} B^{k_c z_c} X_T^m \\
= & \sum_{k_1=0}^{R_1} \cdots \sum_{k_c=0}^{R_c} \binom{R_1}{k_1} \cdots \binom{R_c}{k_c} (-1)^{k_1+\cdots+k_c} X^m_{T-k_1 z_1 - \cdots - k_c z_c} \\
= & \sum_{k_1=0}^{R_1} \cdots \sum_{k_c=0}^{R_c} \binom{R_1}{k_1} \cdots \binom{R_c}{k_c} (-1)^{k_1+\cdots+k_c} \sum_{t=m(T-1)+1}^{mT} Y_{t-k_1 s_1 - \cdots - k_c s_c} \\
= & \sum_{t=m(T-1)+1}^{mT} (1 - B^{s_1})^{R_1} \cdots (1 - B^{s_c})^{R_c} Y_t. \qquad (16)
\end{aligned}
$$

Thus, by (16) and the technique used in the proof of Theorem 1 of Tsai and Chan (2005), we have, for $m = 2h + 1$,

$$
\begin{aligned}
& (1-B)^r (1-B^{z_1})^{R_1} \cdots (1-B^{z_c})^{R_c} X_T^m \\
= \; & (1-B)^r \sum_{t=m(T-1)+1}^{mT} (1-B^{s_1})^{R_1} \cdots (1-B^{s_c})^{R_c} Y_t \\
= \; & \int_{-\pi}^{\pi} \exp(i\omega T) dZ_m(\omega),
\end{aligned}
$$

where

$$
\begin{aligned}
dZ_m(\omega) \; = \; & \sum_{k=-h}^{h} \left\{ \left(1 - \exp\Big(\frac{i(\omega + 2k\pi)}{m}\Big)\right)^{-r-1} \exp\left(\frac{i(1-m)(r+1)(\omega + 2k\pi)}{m}\right) \right. \\
& \left. \times (1 - \exp(i\omega))^{r+1} dZ\left(\frac{\omega + 2k\pi}{m}\right) \right\},
\end{aligned}
$$

$-\pi < \omega \leq \pi$. Therefore, for $m = 2h + 1$ and $-\pi < \omega \leq \pi$,

$$
\begin{aligned}
f_{\xi,m}(\omega) d\omega \; = \; & \mathrm{E}(|dZ_m(\omega)|^2) \\
= \; & \frac{1}{m} |1 - \exp(i\omega)|^{2r+2} \prod_{j=1}^{c} |1 - \exp(iz_j\omega)|^{-2D_j} \prod_{j=1}^{c} \left| \frac{\Theta_j(\exp(iz_j\omega))}{\Phi_j(\exp(iz_j\omega))} \right|^2 \\
& \times \sum_{k=-h}^{h} \left| 1 - \exp\left(\frac{i(\omega + 2k\pi)}{m}\right) \right|^{-2r-2d-2} g\left(\frac{\omega + 2k\pi}{m}\right) d\omega \\
= \; & \frac{1}{m} \left| 2\sin\left(\frac{\omega}{2}\right) \right|^{2r+2} \prod_{j=1}^{c} \left| 2\sin\left(\frac{z_j\omega}{2}\right) \right|^{-2D_j} \prod_{j=1}^{c} \left| \frac{\Theta_j(\exp(iz_j\omega))}{\Phi_j(\exp(iz_j\omega))} \right|^2 \\
& \times \sum_{k=-h}^{h} \left| 2\sin\left(\frac{\omega + 2k\pi}{2m}\right) \right|^{-2r-2d-2} g\left(\frac{\omega + 2k\pi}{m}\right) d\omega.
\end{aligned}
$$

This proves part (a) for $m = 2h + 1$. The proof for the case of $m = 2h$ is similar and hence omitted.

(b) Without loss of generality, consider $m = 2h + 1$.

$$
m^{-2r-2d-1} f_{\xi,m}(\omega) \; = \; \{2(1 - \cos\omega)\}^{r+1} \prod_{j=1}^{c} \left| 2\sin\left(\frac{z_j\omega}{2}\right) \right|^{-2D_j} \prod_{j=1}^{c} \left| \frac{\Theta_j(\exp(iz_j\omega))}{\Phi_j(\exp(iz_j\omega))} \right|^2
$$

$$\times \sum_{k=-h}^{h} \left| 2m \sin\left(\frac{\omega + 2k\pi}{2m}\right)\right|^{-2r-2d-2} g\left(\frac{\omega + 2k\pi}{m}\right),$$

which tends to $\{2(1-\cos\omega)\}^{r+1} \prod_{j=1}^{c} |2\sin(z_j\omega/2)|^{-2D_j} \prod_{j=1}^{c} |\Theta_j(\exp(iz_j\omega))|^2 |\Phi_j(\exp(iz_j\omega))|^{-2}$ $\sum_{k=-\infty}^{\infty} |\omega + 2k\pi|^{-2r-2d-2} g(0)$ by the dominated convergence theorem, owing to (i) the inequality $|\sin\omega| \leq |\omega|$, (ii) the boundedness of g and its continuity at 0, and (iii) the fact that $\sum_{k=-\infty}^{\infty} |\omega + 2k\pi|^{-2r-2d-2} < \infty$ based on the assumption given in Remark 3 in Section 3. The convergence of the normalization constants of $f_{\xi,m}$ to K_ξ follows along similar arguments.

Proof of Theorem 2. Let the true differencing orders be r_0 and $R_{j,0}, j = 1, \ldots, c$. There are two cases dependent on whether or not all of the differencing orders r and $R_j, j = 1, \ldots, c$, are greater than or equal to their true counterparts, in which case it shall be designated as Case 1 and otherwise Case 2. In other words, Case 1 concerns the case that the differenced series is stationary whereas Case 2 implies that the differenced series is non-stationary.

For Case 2, we claim that the approximate negative Whittle likelihood defined by (7), when normalized by the sample size, is eventually uniformly unbounded, with probability 1. Hence, the QMLE must have the differencing orders greater than or equal to their true counterparts. To prove the claim, note that the root condition preceding Equation (2) ensures that $\tilde{f}(\omega_j; R, \theta)$ is bounded away from 0 for all sufficiently large M and uniformly for all θ (over the compact parameter space). Moreover, it can be readily checked that $F(\theta) = \int_0^\pi \log \tilde{f}(\omega; R, \theta) d\omega$ is uniformly bounded for θ. Consequently, $F_N(\theta) = \sum_{j=1}^{T} 2\pi \log \tilde{f}(\omega_j; R, \theta)/N$ must be uniformly bounded for all θ, which can be verified by deriving a contradiction if it is not true, as follows. Suppose that the normalized sums are unbounded over the compact parameter space and for all positive integer N. Then, without loss of generality, there exists a sequence θ_N in the compact parameter space such that $F_N(\theta_N) \to \infty$ as $N \to \infty$ and $\theta_N \to \theta_0$, the true parameter vector. However, this implies that $F(\theta_0) = \infty$, by Lebesgue dominated convergence theorem, leading to a contradiction. The proof of the claim is completed by observing that there exists a positive constant C_1 such that

$$\frac{1}{N}\sum_{j=1}^{T}\frac{I_N(\omega_j;R)}{\tilde{f}(\omega_j;R,\theta)} \quad \geq \quad \frac{C_1}{N}\sum_{j=1}^{T} I_N(\omega_j;R),$$

and $\sum_{j=1}^{T} I_N(\omega_j;R)/(T-1)$ equals the sample variance of the under-differenced series which tends to infinity almost surely, due to the non-stationarity of the differenced series.

Henceforth, we focus on Case 1 so that the differenced series must be stationary. The parameters r and R's are discrete and admit only finitely many values. However, for fixed r and R's, the other parameters are continuous-valued, and Lemma 1 of Hosoya (1996) then implies the consistency of the QMLE. Hence, for ease of exposition, we shall only give the proof for the case when r and R's are set to their true values. We first note that the techniques used in proving Theorem 1 of Tsai (2006) can be adapted to show that the approximation of the Whittle likelihood (7) by (9) is asymptotically negligible under the stated growth rate of M. Below, we implicitly assume that the negative Whittle likelihood is normalized by the sample size. Because the parameter space is compact, it can be shown by routine analysis that the first partial derivatives of the approximate Whittle likelihood differ from those of the true Whittle likelihood by an error of order $o_p(1)$, uniformly over the parameter space. That the approximation of the Whittle likelihood has negligible effects on the large-sample asymptotics of the QMLE follows from the proof of Theorem 1 and Lemma 1 of Hosoya (1996), and the fact that, over a sufficiently small neighborhood of the true parameter vector, the approximation errors of the first partial derivatives are of order $O_p(M^{-2r_0-2d_0-1})$, see the discussion below (8).

Thus, without loss of generality, we can assume that the estimator is the exact QMLE, in which case Theorem 2 follows from Theorem 2 of Hosoya (1996) if we can verify Conditions A, C, and D listed there. (For ease of exposition, we confine the proof for the case that the true differencing orders $R = 0$ and $r = 0$, as the proof for the general case is similar.) We now verify these conditions. Let $\delta(x, y) = 1$ if $x = y$, and $\delta(x, y) = 0$ otherwise. First, write $f(\omega; \xi, R, \sigma^2) = \sigma^2 2^{-2r-2+2D_1+\cdots+2D_c}|k(\omega)|^2$, where

$$
\begin{aligned}
k(\omega) &= (1-\exp(i\omega))^{r+1}\prod_{j=1}^{c}(1-\exp(iz_j\omega))^{-2D_j}\prod_{j=1}^{c}\frac{\Theta_j(\exp(iz_j\omega))}{\Phi_j(\exp(iz_j\omega))} \\
&\quad \times\left\{\sum_{k=-\infty}^{\infty}|\omega+2k\pi|^{-2r-2d-2}\right\}^{1/2}.
\end{aligned}
$$

For $t = 1, ..., N$, write $Y_t = \sum_{j=0}^{\infty} G_j e_{t-j}$, where $\{e_j\}$ is a sequence of is a sequence of independent and identically distributed normal random variables with zero mean and unit variance, and $\{G_j\}$ is such that $k(\omega) = \sum_{j=0}^{\infty} G_j \exp(ij\omega)$. Condition A can be checked easily, and therefore omitted. In what follows, we write $f(\omega)$ for $f(\omega; R, \theta)$ for simplicity. We now verify condition C.

(i) (a) We will show that $\int_{-\pi}^{\pi} f^u(\omega)d\omega < \infty$ for some u such that $1 < u \leq 2$. By Remark 1 in Section 3, it can be easily checked that there exists non-negative constants b_0, b_1, and b_{jk}, ε_{jk}, $j = 1, \cdots, c, k = 1, \cdots, [z_j/2]$, such that

$$
\begin{aligned}
&\int_{-\pi}^{\pi} f^u(\omega)d\omega \\
\leq\ & b_0\int_0^{\varepsilon}\omega^{-2u(d+D_1+\cdots+D_c)}d\omega+\sum_{j=1}^{c}\sum_{k=1}^{[zj/2]}b_{jk}\int_{\omega_{jk}-\varepsilon_{jk}}^{\omega_{jk}}(\omega_{jk}-\omega)^{-2uD_j}d\omega+b_1 \\
<\ & \infty.
\end{aligned}
$$

(b) We need to show that there exists $\gamma > 0$ such that

$$
\sup_{|\lambda|<\varepsilon}\|f^{-1}(\cdot)\{f(\cdot)-f(\cdot-\lambda)\}\|_u = O(\varepsilon^{\gamma}), \tag{17}
$$

where $\|g\|_p = \{\int_{-\pi}^{\pi} |g(\omega)|^p d\omega\}^{1/p}$. This can be shown by making use of two inequalities, the first one being that for any functions a and b, there exists a constant $K > 0$ such that for all $|\lambda|$ small,

$$\begin{aligned} & \|a^{-1}(\cdot)b^{-1}(\cdot)\{a(\cdot)b(\cdot) - a(\cdot - \lambda)b(\cdot - \lambda)\}\|_u \\ \leq \; & K(\|a^{-1}(\cdot)\{a(\cdot) - a(\cdot - \lambda)\}\|_u + \|b^{-1}(\cdot)\{b(\cdot) - b(\cdot - \lambda)\}\|_u), \end{aligned}$$

if the ratio $a^{-1}(\cdot)\{a(\cdot) - a(\cdot - \lambda)\}$ is uniformly bounded for all λ that are sufficiently small. The second inequality is the trivial observation that for any function a and any finite partition $[-\pi, \pi] = \cup A_k$, $\|a\|_u \leq \sum \|aI_{A_k}\|_u$ where I_A is the indicator function of the set A. Note that the spectral density function has poles possibly at 0, and integer multiples of $2\pi/z_j, j = 1, \ldots, c$, that are not greater than π in magnitude, and that the spectral density function is uniformly continuous over any compact set that does not include any pole of the function. The second inequality implies that (17) holds if it holds when the spectral density function is restricted to any small interval containing only one pole of the spectral density function. Thus, it suffices to show that (17) holds for $f(\omega) = |\omega|^{-\kappa}$ with some $0 < \gamma < 1$, where $1 > \kappa > 0$, which we now verify. Let $u > 1$ be a constant such that $0 < u\kappa < 1$. Because $0 \leq (\omega + |\lambda|)^\kappa - \omega^\kappa \leq |\lambda|^\kappa$, for $\omega \geq 0$, we have

$$\begin{aligned} & \|f^{-1}(\cdot)\{f(\cdot) - f(\cdot - \lambda)\}\|_u^u \\ \leq \; & 2\int_0^\pi \frac{\{(\omega + |\lambda|)^\kappa - \omega^\kappa\}^u}{(\omega + |\lambda|)^{\kappa u}} d\omega + \int_0^{|\lambda|} \left| \frac{\omega^\kappa - (|\lambda| - \omega)^\kappa}{(|\lambda| - \omega)^\kappa} \right|^u d\omega \\ \leq \; & 2|\lambda|^{\kappa u} \int_0^\pi (\omega + |\lambda|)^{-\kappa u} d\omega + \int_0^{|\lambda|} |\omega^\kappa/(|\lambda| - \omega)^\kappa - 1|^u d\omega \\ \leq \; & 2|\lambda|^{\kappa u}\{(\pi + |\lambda|)^{1-\kappa u} - |\lambda|^{1-\kappa u}\}/(1 - \kappa u) + 2^u|\lambda|(1 + |\lambda|^{\kappa u}/(\kappa u + 1)). \end{aligned}$$

(ii) Let $h_j(\omega) = \partial f^{-1}(\omega; \theta)/\partial\theta_j$, where θ_j is the jth component of θ. For any $\varepsilon > 0$, there exists $a > 0$ and Hermitian-valued bounded functions $\tilde{h}_j$ and $\bar{h}_j$ such that, if $|\theta_1 - \theta| < a$, $\tilde{h}_j(\omega) \leq h_j(\omega, \theta_1) \leq \bar{h}_j(\omega)$ and $\left\| \{\bar{h}_j(\cdot) - \tilde{h}_j(\cdot)\} f(\cdot) \right\|_v < \varepsilon$, where $v = (u-1)/u$ for u given in (17) above.

The Hermitian-value requirement is automatically satisfied since the time series is univariate. Note that $\partial f^{-1}/\partial\theta_j = -f^{-1}\partial \log f/\partial\theta_j$ and that f^{-1} is a uniformly continuous function in ω and θ in a sufficiently small neighborhood of θ_0, and so are the first partial derivatives of $\log f$. Thus, the requirement in (ii) can be readily verified.

(iii) Let $V_j(\theta) = H_j(\theta) + \int_{-\pi}^{\pi} \mathrm{tr}\{h_j(\omega, \theta) f(\omega)\} d\omega$, where $f(\omega)$ is the true spectral density function, $h_j(\omega, \theta) = \partial f^{-1}(\omega, \theta)/\partial\theta_j$, and $H_j(\theta) = \partial \int_{-\pi}^{\pi} \log|f(\omega; \theta)| d\omega/\partial\theta_j$. Then, $V_j(\theta)$ has a unique zero for all j at $\theta = \theta_0$, where θ_0 is an interior point of θ.

This condition can be proved as follows. Note that the true spectral density function is, by assumption, equal to $f(\omega, \theta_0)$. Consider the function $Q(\theta) = \int_{-\pi}^{\pi} \log f(\omega, \theta) d\omega - \int_{-\pi}^{\pi} \log f(\omega, \theta_0) d\omega + \int_{-\pi}^{\pi} f(\omega, \theta_0)/f(\omega, \theta) d\omega$. Note that the partial derivative of Q with respect to the jth component of θ equals $V_j(\theta)$, for all j. Condition (iii) holds if Q attains its unique minimum at $\theta = \theta_0$, which is shown below. Define $T(x) = \exp(x) - x$ which is a convex function that is always ≥ 1. Jensen's inequality implies that

$$\begin{aligned} Q(\theta)/(2\pi) &= \int_{-\pi}^{\pi} T(\log f(\omega, \theta_0) - \log f(\omega, \theta)) d\omega/(2\pi) \\ &\geq T\left(\int_{-\pi}^{\pi} \{\log f(\omega, \theta_0) - \log f(\omega, \theta)\} d\omega/(2\pi) \right) \\ &\geq 1, \end{aligned}$$

with both equalities obtained if and only if $\theta = \theta_0$.

(iv) $H_j(\theta)$ is continuous on θ.

This condition holds trivially.

Parts (i)–(iii) of condition D can be proved by arguments similar to those used in proving conditions (i) and (ii) of condition C. Condition (v) of condition D can be easily verified if the spectral density function admits no poles but otherwise it can be proved by adapting the arguments presented in Example 3.1 of Hosoya (1996). It remains to consider

(iv) $\left|V(\theta)\right| \geq \alpha_1 \left|\theta_1 - \theta_0\right|$ for some $\alpha_1 > 0$ in a neighborhood of θ_0, where V is the vector consisting of all the first partial derivatives V_j.

This condition holds because it is straightforward to show that

$$\frac{\partial^2 Q}{\partial \theta_i \partial \theta_j} = \int_{-\pi}^{\pi} \frac{\partial \log f(\omega, \theta_0)}{\partial \theta_i} \frac{\partial \log f(\omega, \theta_0)}{\partial \theta_j d\omega},$$

which is positive definite since the partial derivatives are linearly independent.

References

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

[2] Chambers, M. J. (1996), "The Estimation of Continuous Parameter Long-memory Time Series Models," *Econometric Theory*, 12, 374-90.

[3] Davies, R.B. and Harte, D.S. (1987), "Tests for Hurst Effect," *Biometrika*, 74, 95-101.

[4] Granger, C. W. J. and Joyeux, R. (1980), "An Introduction to Long Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15-29.

[5] Hassler, U. and Wolters, J. (1995), "Long Memory in Inflation Rates: International Evidence," *Journal of Business and Economic Statistics*, 13, 37-45.

[6] Hidalgo, J. and Kreiss, J.-P. (2006), "Bootstrap Specification Tests for Linear Covariance Stationary Processes," *Journal of Econometrics*, 133, 807-839.

[7] Hosking, J.R.M. (1981), "Fractional Differencing," *Biometrika*, 68, 165-76.

[8] Hosoya, Y. (1996), "The Quasi-likelihood Approach to Statistical Inference on Multiple Time-series with Long-range Dependence," *Journal of Econometrics*, 73, 217-236.

[9] Hsu, N.-J. and Tsai, H. (2007), "Semiparametric Inference for Long-Memory Seasonality Using Generalized Exponential Models," unpublished manuscript.

[10] Man, K.S. and Tiao, G.C. (2006), "Aggregation Effect and Forecasting Temporal Aggregates of Long Memory Processes," *International Journal of Forecasting*, 22, 267-281.

[11] Montanari, A., Rosso, R. and Taqqu, M. (2000), "A Seasonal Fractional ARIMA Model Applied to Nile River Monthly Flows at Aswan," *Water Resources Research*, 36, 1249-1259.

[12] Ohanissian, A., Russell, J, and Tsay, R. (2008), "True or Spurious Long Memory? A New Test," *Journal of Business and Economic Statistics*, in press.

[13] Palma, W. and Bondon, P. (2003), "On the Eigenstructure of Generalized Fractional Processes," *Statistics & Probability Letters*, 65, 93-101.

[14] Palma, W. and Chan, N. H. (2005), "Efficient Estimation of Seasonal Long-range-dependent Processes," *Journal of Time Series Analysis*, 26, 863-892.

[15] Porter-Hudak, S. (1990), "An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates," *Journal of the American Statistical Association*, 85, 338-344.

[16] Priestley, M. B. (1981), *Spectral Analysis and Time Series*. New York: Academic Press.

[17] Ray, B. K. (1993), "Long-range Forecasting of IBM Product Revenues Using a Seasonal Fractionally Differenced ARMA Model," *International Journal of Forecasting*, 9, 255-269.

[18] Tsai, H. (2006), "Quasi-maximum likelihood estimation of long-memory limiting aggregate processes," *Statistica Sinica*, 16, 213-226.

[19] Tsai, H. and Chan, K. S. (2005), "Temporal aggregation of stationary and non-stationary discrete-time processes," *Journal of Time Series Analysis*, 26, 613-624.

[20] Wei, W.W.S. (1978), "Some Consequences of Temporal Aggregation in Seasonal Time Series Models," In *Seasonal Analysis of Economic Time Series*, Zellner A. (eds.) US Department of Commerce, Bureau of Census: Washington, DC.