Temporal aggregation of stationary and nonstationary discrete-time processes

HENGHSIU TSAI

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

htsai@stat.sinica.edu.tw

K. S. CHAN

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

kung-sik-chan@uiowa.edu

March 11, 2004

SUMMARY

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 AutoRegressive Fractionally Integrated Moving-Average (ARFIMA) process, after suitable number of differencing if necessary. We derive closed-form expressions for the limiting autocorrelation function and the normalized spectral density of the aggregates, as the extent of aggregation increases to infinity. These results are then used to assess the loss of forecasting efficiency due to aggregation.

Some key words: Asymptotic efficiency of prediction, Autocorrelation, ARFIMA models, Long memory, Spectral density.

1 Introduction

Time series data are often (temporally) aggregated before analysis. Thus, a fundamental problem is to characterize what features of the basic series are inherited by the aggregates. For a short-memory time series, the aggregate data approach white noises with increasing aggregation, owing to the Central Limit Theorem. However, Tiao (1972) 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 part uniquely determined by the differencing order. Tiao (1972) made use of the limiting structure of the aggregates to assess the loss of forecasting efficiency due to aggregation. Recently, there are some works examining this problem assuming that the basic series is fractionally integrated. Specifically, the basic series is assumed to be some autoregressive fractionally integrated moving average (ARFIMA) process. Man & Tiao (working paper, 2001a,b) found that temporal aggregation preserves the long-memory parameter of the basic series. They showed that the limiting structure of the temporal aggregates is an ARFIMA$(0, d, \infty)$ model that can be approximated (and is exact when d is an integer) by some ARFIMA$(0, d, \bar{d})$ model, where $\bar{d}$ is the greatest integer strictly less than $d + 1$. They gave for the d-differenced aggregates their limiting autocorrelation function with lag up to and including $\bar{d}$ in magnitude. They also investigated the approximate loss of 1-step forecasting efficiency due to aggregation. Ohanissian, Russell & Tsay (working paper, 2002, available at *www.crde.umontreal.ca/crde-cirano/russell.pdf*) made use of the fact that temporal aggregation of an ARFIMA model preserves the long-memory parameter to devise a test for spurious long-memory. Beran & Ocker (working paper, 2000, available at $http : //netec.mcc.aac.uk/WoPEc/data/knzcofedp.html$) derived explicit formulas for the limiting autocorrelation function of the aggregates of an ARFIMA$(p, d, 0)$ model for the case $-1/2 < d < 3/2$ and $d \notin \{0, 1/2, 1\}$.

Both Man & Tiao (working paper, 2001a,b) and Beran & Ocker (working paper, 2000) used time domain methods in their investigations. Now, a long-memory process can be usefully characterized in terms of the asymptotic behaviour of its spectral density function around the origin. Here, we consider the second moment structure of the temporally aggregated data, by using frequency-domain methods. The basic series is assumed to be an ARFIMA(p, d, q) process where d is any real number that is greater than $-1/2$ and is not a half-integer. In particular, we derive below the exact form of the spectral density function of the aggregate data and its limiting normalized form with increasing aggregation. Interestingly, the limiting normalized spectral density function are the same as that of temporal aggregates of a Continuous-time Autoregressive Fractionally Integrated Moving-average (CARFIMA$(p, r + d - 1/2, q)$) process; see Tsai & Chan (working paper, 2003) for details. Moreover, we derive the exact limiting autocorrelation function of the aggregate data, which is identical to that of Beran & Ocker (2000) for $-1/2 < d < 3/2$ and $d \notin \{0, 1/2, 1\}$.

This paper is organized as follows. In § 2, we briefly review the ARFIMA models. The main results on the limiting (normalized) spectral density and the autocorrelation function are given in § 3. Finally, in § 4, we consider the forecasting of aggregates. If the the basic series is available, the future aggregates can be forecasted more accurately than just using past aggregates. However, it is often expensive to measure the basic series while the aggregates may be more readily available. A natural question arises as to how much forecasting efficiency can be gained by using the basic series relative to that based on aggregated data. For one-step prediction, we show that, for the stationary case, the loss of forecasting efficiency due to aggregation is generally less than 10% and at most 20%. However, the loss of forecasting efficiency increases drastically with the (fractional) integration order, in the non-stationary case. All proofs are placed in an appendix.

2 ARFIMA(p, d, q) processes

The time series $\{Y_t, t = 0, \pm 1, \pm 2, \cdots\}$ is said to be a stationary ARFIMA(p, d, q) process with $d \in (-1/2, 1/2)$ if $\{Y_t\}$ is stationary and satisfies the difference equations,

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

where $\{\epsilon_t\}$ is white noise with mean 0 and variance σ^2, $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$, $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ and B is the backward shift operator, $(1 - B)^d$ is 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 & 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 $\{Y_t\}$ is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2} |1 - e^{-i\lambda}|^{-2d}; \tag{2}$$

Brockwell & Davis (1991, p. 525).

For a unified framework for studying non-stationary and/or long-memory discrete-time process $\{Y_t\}$, see Beran (1995), Ling & Li (1997) and Beran *et al.* (1998). Throughout this paper, $\{Y_t\}$ is assumed to be an ARFIMA$(p, r+d, q)$ process, which means that the r-th difference of $\{Y_t\}$ is a stationary ARFIMA(p, d, q) process. The sum $r + d$ is referred to as its (fractional) integration order. The process is stationary if and only if the integration order is between $-1/2$ and $1/2$. In the stationary case, i.e., $r = 0$, the process is of long-memory for $0 < d < 1/2$, of short memory if $d = 0$, and anti-persistent for $-1/2 < d < 0$.

3 Aggregates of discrete-time processes

Let r be a non-negative integer, $-1/2 < d < 1/2$, and $\{Y_t, t = 0, \pm 1, \pm 2, \cdots\}$ an ARFIMA$(p, r + d, q)$ process:

$$\phi(B)(1-B)^{r+d} Y_t = \theta(B)\epsilon_t, \tag{3}$$

That is, after rth differencing, the process is a stationary ARFIMA(p, d, q) process defined by equation (1). 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 differencing operator. Below are the main results on the spectral density of the aggregates.

THEOREM 1 *Assume that $\{Y_t\}$ is an ARFIMA$(p, r+d, q)$ model defined by (3).*
(a) For $r \geq 0$ and $m = 2h+1$, the spectral density function of $\{\nabla^r X_T^m\}$ is given by

$$\begin{aligned} f_{r,m}(w) &= m^{-1}\{2(1-\cos w)\}^{r+1} \\ &\quad \times \sum_{k=-h}^{h} \left| 2\sin\left(\frac{w+2k\pi}{2m}\right)\right|^{-2r-2d-2} g\left(\frac{w+2k\pi}{2m}\right), \end{aligned} \tag{4}$$

where $-\pi < w < \pi$, and $g(\lambda) = \sigma^2(2\pi)^{-1}|\theta(e^{-i\lambda})|^2|\phi(e^{-i\lambda})|^{-2}$.
If $m = 2h$, the spectral density is given by equation (4) with the summation ranging from $-h+1$ to h for $-\pi < w < 0$ and from $-h$ to $h-1$ for $0 < w < \pi$.
(b) As $m \to 0$, the normalized spectral density function of $\{\nabla^r X_T^m\}$ converges to

$$f_r(w) = K\{2(1-\cos w)\}^{r+1} \sum_{k=-\infty}^{\infty} |w+2k\pi|^{-2r-2d-2}, \tag{5}$$

where K is the normalization constant ensuring that $\int_{-\pi}^{\pi} f_r(w)dw = 1$.

Note (i) both $f_{r,m}$ and f_r are of $O(w^{-2d})$ for $|w| \longrightarrow 0$, so that the aggregates and their limits preserve the long-memory parameter of the underlying ARFIMA process, and (ii) the limiting normalized spectral density function is independent of the short-memory parameter, confirming the central limit effect. It can be seen from the proof of Theorem 1 that part (b) of Theorem 1 actually holds for any discrete-time process whose spectral density function is of the form $w^{-2d}g(w)$, where g is a bounded, integrable function that is continuous at $w = 0$, with $g(0) > 0$. The convergence of the normalized spectral density functions of the m-aggregates ensures the convergence of the corresponding autocorrelation functions, and hence if the basic series is Gaussian then after suitable re-scaling the m-aggregates converge to some limiting Gaussian process, in finite-dimensional distributions; the re-scaling is a function of m and is proportional to $m^{-r-d-1/2}$, as can be seen from the proofs of the theorem.

Tsai & Chan (working paper, 2003) considered the analogous problem for a continuous-time process $\{Y_t, t \in R\}$, where temporal aggregation is defined by integration, i.e., the aggregate series $\{Y_n^\Delta, n \in Z\}$ is defined by $Y_n^\Delta = \int_{(n-1)\Delta}^{n\Delta} Y_u du$. They showed that for a continuous-time autoregressive fractionally integrated moving-average (CARFIMA$(p, r+H, q)$) process, the normalized spectral density function of $\{\nabla^r Y_n^\Delta\}$ converges to the function $f_r(w)$ given by equation (5) with the fractional differencing parameter d replaced by $H-1/2$. Moreover, Tsai & Chan (working paper, 2003) derived a closed-form expression for the limiting autocorrelation function of $\{\nabla^r Y_n^\Delta\}$ as the extent of aggregation increases to infinity. Because the rth difference process of the aggregates of the CARFIMA$(p_1, r+d+1/2, q_1)$ and the ARFIMA$(p_2, r+d, q_2)$ model have the same expressions for their limiting normalized spectral densities, they must share the same limiting autocorrelation functions. Thus the following result follows directly from Theorem 2 of Tsai & Chan (working paper, 2003).

THEOREM 2 *Assume* $-1/2 < d < 1/2$.
(a) As $m \to \infty$, *the limiting autocorrelation function of* $\{\nabla^r X_T^m\}$ *is given by*

$$\begin{aligned} \rho_\infty(h) &= K \sum_{k=-r-1}^{r+1} (-1)^{r+k+1} \binom{2r+2}{k+r+1} |h+k|^{2d+2r+1} \\ &= K \sum_{k=0}^{2r+2} (-1)^k \binom{2r+2}{k} |r+1-h-k|^{2d+2r+1}, \end{aligned} \tag{6}$$

where K *is a normalization constant ensuring that* $\rho_\infty(0) = 1$.
(b) For $d = 0$, *the limiting model for* $\{\nabla^r X_T^m, T \in Z\}$ *is an ARIMA(0,r,r) model. The corresponding limiting autocorrelation function given by (6) can be simplified to*

$$\rho_\infty(h) = K \sum_{k=0}^{r-h} (-1)^k \binom{2r+2}{k} (r+1-h-k)^{2r+1}, \quad \textit{if } |h| \le r, \tag{7}$$

and 0 *otherwise, where* K *is a normalization constant ensuring that* $\rho_\infty(0) = 1$.

Equation (7) is a special case of equation (6), by equation (0.154.6) of Gradshteyn & Ryzhik (1994); namely, for any integers N, p, real number α and $N \ge p \ge 1$,

$$\sum_{k=0}^{N} (-1)^k \binom{N}{k} (\alpha+k)^{p-1} = 0. \tag{8}$$

Using the self-similarity property of the fractional Brownian motion, Tsai & Chan (working paper, 2003) showed that the limiting autocorrelation structure can be realized by some functional of the standard fractional Brownian motion with the fractional integration order equal to $H = d+r+1/2$. Note also that equation (7) is essentially the same as equation (2.8) of Tiao (1972). Below, we give some examples illustrating the preceding theorem.

Example: For $-1/2 < d < 1/2$, the limiting autocorrelation function of the aggregates takes the form $\rho_\infty(h) = K\{|h-1|^{2d+1} - 2|h|^{2d+1} + |h+1|^{2d+1}\}$, for $r = 0$. For $r = 1$, $\rho_\infty(h) = K\{|h-2|^{2d+3} - 4|h-1|^{2d+3} + 6|h|^{2d+3} - 4|h+1|^{2d+3} + |h+2|^{2d+3}\}$. For r=2, $\rho_\infty(h) = K\{|h-3|^{2d+5} - 6|h-2|^{2d+5} + 15|h-1|^{2d+5} - 20|h|^{2d+5} + 15|h+1|^{2d+5} - 6|h+2|^{2d+5} + |h+3|^{2d+5}\}$.

Note that for $r = 0$ and $r = 1$, we recover the results of Beran & Ocker (working paper, 2000).

4 Forecasting efficiency of aggregate series

Suppose we are interested in doing one step ahead prediction of the future aggregate observation X^m_{T+1}. This can be done by using the aggregate series $\{X^m_s, s \leq T\}$. In principle, more accurate prediction can be obtained from the basic discrete-time process $\{Y_k, k \leq Tm\}$ if it is available. This raises the issue of what is the possible gain one may obtain in the prediction using the basic series as compared to just using past and current aggregates. Let $\sigma^2_{1,r,d}(m)$ be the one-step prediction variance of X^m_{T+1} by using the aggregate series $\{X^m_s, s \leq T\}$ and $\sigma^2_{2,r,d}(m)$ be the corresponding prediction variance obtained by using the basic series. Following Tiao (1972), we define the limiting efficiency of one-step ahead prediction using the basic series as compared to that of using the temporally aggregated series to be the limiting variance ratio $\xi_1(r,d) = \lim_{m\to\infty}\{\sigma^2_{1,r,d}(m)/\sigma^2_{2,r,d}(m)\}$.

THEOREM 3 *For model (3) with a non-negative integer r and $-1/2 < d < 1/2$, the limiting prediction variance ratio is given by*

$$\begin{aligned}\xi_1(r,d) &= 2^{r+1}(2r+2d+1)\Gamma^2(r+d+1) \\ &\quad \times \exp\left\{\frac{r+1}{2\pi}\int_{-\pi}^{\pi}\log(1-\cos\lambda)d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\sum_{k=-\infty}^{\infty}|w+2k\pi|^{-2r-2d-2}dw\right\}.\end{aligned}$$

The computation of $\xi_1(r,d)$ requires evaluating the sum $\sum_{k=-\infty}^{\infty}|w+2k\pi|^{-2r-2d-2}$. If $d+r$ are integers, it is possible to evaluate this infinite sum using well-known methods, see Chambers (1996). For example, by equations (822) and (824) of Jolley (1961), we have the following two simplifications:

$$\xi_1(0,0) = 1; \quad \xi_1(1,0) = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left\{3 - 2\sin^2\left(\frac{w}{2}\right)\right\}dw\right\}.$$

But with non-integer values of the exponent, some approximation method is required to compute the series. Here we adopt the method of Chambers (1996) as follows. First note that

$$
\begin{aligned}
&\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\sum_{k=-\infty}^{\infty}|w+2k\pi|^{-2r-2d-2}dw \\
= \quad &(-2r-2H-1)\log(2\pi)+2\int_{0}^{1/2}\log\sum_{k=-\infty}^{\infty}|y+k|^{-2r-2d-2}dy.
\end{aligned}
$$

The series $\sum_{k=-\infty}^{\infty}|y+k|^{-2r-2d-2}$ can be approximated by $\sum_{j=1}^{M}(j-y)^{-2r-2d-2}+\sum_{j=0}^{M}(j+y)^{-2r-2d-2}+(2r+2d+1)^{-1}\{(M-y)^{-2r-2d-1}+(M+y)^{-2r-2d-1}\}$ for some large M. The results summarized in Table 1 are based on $M=10,000$. Indeed, the results based on $M=100$ are essentially the same, suggesting that setting $M=10,000$ provides an adequate approximation to the series.

Table 1 displays the variance ratio $\xi_1(r,d)$ for various r and d. For the purpose of comparison, some of the corresponding values computed by Man & Tiao (working paper, 2001) are included and enclosed in parentheses in the Table. It is interesting to note that for $r+d=0,1,2,3,4$, our results are identical to those of Man & Tiao (working paper, 2001) up to four significant digits. Theorem 3 for computing $\xi_1(r,d)$ is not applicable for $r+d=0.5,1.5,2.5,3.5$, so the values of $\xi_1(r,d)$ for $r+d=0.4999,1.4999,2.4999$ and 3.4999 are computed. Note that these values are close to but larger than those for $r+d=0.5,1.5,2.5,3.5$ of Man & Tiao (working paper, 2001), respectively. However, because the ARFIMA model is undefined when the fractional integration order $r+d$ is a half-integer, the meaning of $\xi_1(r,d)$ is unclear for such cases.

Table 1 indicates that for the stationary case, i.e., $-1/2<d+r<1/2$, the gain of forecasting efficiency by using the basic series as compared to just using the aggregates is at most 20% and generally less than 10% for $-0.5<d+r<=0.3$. However, the efficiency gain increases rapidly with $d+r$ for nonstationary processes.

Acknowledgement

We thank Academia Sinica, the University of Iowa and the National Science Council (NSC 92-2118-M-001-018), R.O.C. for partial support.

Table 1: *Values of the one-step prediction variance ratio $\xi_1(r, d)$ for various r and d (values of $\xi_1(r, d)$ produced by Man & Tiao (working paper, 2001b)).*

$d \backslash r$	0	1	2	3	4
-0.4999	1.001	1.181	3.721	25.508	307.970
-0.4000	1.087	1.267	4.374	31.974	405.347
-0.3000	1.058	1.375	5.182	40.306	535.958
-0.2000	1.027	1.507	6.182	51.080	711.646
-0.1000	1.007	1.669	7.427	65.070	948.832
0.0000	**1.000**	**1.866**	**8.982**	**83.311**	**1270.197**
0.0000	(1.000)	(1.866)	(8.982)	(83.312)	(1270.260)
0.1000	1.007	2.106	10.932	107.190	1707.161
0.2000	1.028	2.398	13.390	138.580	2303.387
0.3000	1.063	2.753	16.501	180.005	3119.724
0.4000	1.114	3.188	20.455	234.888	4241.231
0.4999	1.181	3.718	25.490	307.717	5783.730
0.5000	(1.144)	(3.457)	(23.355)	(279.086)	

Appendix 1

Proof of Theorem 1

Proof of (a) Note that $\nabla^r Y_t$ admits the following spectral representation (equation (4.11.19) of Priestley, 1981): $\nabla^r Y_t = \int_{-\pi}^{\pi} e^{itw} dZ(w)$ with $E|dZ(w)|^2 = f(w)dw$ where $f()$ is given by equation (2). Routine calculus shows that, for $r \geq 1$ and $k \geq 1$, $Y_k = p(k) + \sum_{i=1}^{k} \prod_{j=1}^{r-1}(k-i+j)\nabla^r Y_i/(r-1)!$, where $p(k) = Y_0 + \sum_{s=1}^{r-1}\sum_{i=1}^{k}\prod_{j=1}^{s-1}(k-i+j)\nabla^s Y_0/(s-1)!$, $\sum_{j=1}^{0}$ is defined to be zero and $\prod_{j=1}^{0}$ is defined to be one. Because $p(k)$ vanishes upon r-th differencing, we can assume $\nabla^s Y_0 = 0$ for all $0 \leq s \leq r-1$ without loss of generality. Thus, for $T \geq 2$ and $r \geq 1$,

$$\begin{aligned} X_T^m &= \sum_{k=m(T-1)+1}^{mT} Y_k \\ &= \sum_{k=m(T-1)+1}^{mT} \frac{1}{(r-1)!} \sum_{t=1}^{k} \prod_{j=1}^{r-1} (k-t+j)\nabla^r Y_t \\ &= \frac{1}{(r-1)!} \sum_{t=1}^{m(T-1)+1} \sum_{k=m(T-1)+1}^{mT} \prod_{j=1}^{r-1} (k-t-j)\nabla^r Y_t \\ &\quad + \frac{1}{(r-1)!} \sum_{t=m(T-1)+2}^{mT} \sum_{k=t}^{mT} \prod_{j=1}^{r-1} (k-t+j)\nabla^r Y_t. \end{aligned}$$

Using the fact that $\sum_{k=t}^{mT}\prod_{j=1}^{r-1}(k-t+j) = \sum_{i=0}^{mT-t}\prod_{j=1}^{r-1}(i+j) = r^{-1}\sum_{i=0}^{mT-t}\{\prod_{j=1}^{r}(i+j) - \prod_{j=1}^{r}(i+j-1)\} = r^{-1}\sum_{i=1}^{mT-t+1}\prod_{j=1}^{r}(i+j-1) - r^{-1}\sum_{i=1}^{mT-t}\prod_{j=1}^{r}(i+j-1)$ $= r^{-1}\prod_{j=1}^{r}(mT-t+j)$, we have, for $T \geq 2$,

$$\begin{aligned} X_T^m &= \frac{1}{(r-1)!} \sum_{t=1}^{m(T-1)+1} \left\{ \frac{1}{r} \prod_{j=1}^{r} (mT-t+j) - \frac{1}{r} \prod_{j=1}^{r} (m(T-1)-t+j) \right\} \nabla^r Y_t \\ &\quad + \frac{1}{(r-1)!} \sum_{t=m(T-1)+2}^{mT} \frac{1}{r} \prod_{j=1}^{r} (mT-t+j)\nabla^r Y_t \\ &= Q_T - Q_{T-1}, \end{aligned}$$

where

$$\begin{aligned} Q_T &= \frac{1}{r!} \sum_{t=1}^{mT} \prod_{j=1}^{r} (mT-t+j)\nabla^r Y_t \\ &= \frac{1}{r!} \sum_{t=1}^{mT} \prod_{j=1}^{r} (mT-t+j) \int_{-\pi}^{\pi} e^{itw} dZ(w) \\ &= \int_{-\pi}^{\pi} \xi_{T,r} dZ(w), \end{aligned}$$

and

$$\begin{aligned}\xi_{T,r} &= \frac{1}{r!}\sum_{k=1}^{mT}\prod_{j=1}^{r}(mT-k+j)e^{iwk} \\ &= -\sum_{k=1}^{r}\frac{(1-e^{-iw})^{-k}}{(r-k+1)!}\prod_{s=0}^{r-k}(mT+s)-(1-e^{-iw})^{-r-1}(1-e^{imTw}),\end{aligned}$$

the last equality following from routine computation and the fact that $\xi_{T,r+1} = (1-e^{-iw})^{-1}\{\xi_{T,r}-\prod_{k=0}^{r}(mT+k)/(r+1)!\}$. Therefore, for $T\geq r+2$,

$$\begin{aligned}\nabla^r X_T^m &= \nabla^{r+1}Q_T \\ &= \sum_{i=0}^{r+1}\binom{r+1}{i}(-1)^i Q_{T-i} \\ &= \int_{-\pi}^{\pi}\sum_{j=0}^{r+1}\binom{r+1}{j}(-1)^j\xi_{T-j,r}dZ(w) \\ &= \int_{-\pi}^{\pi}\eta_T dZ(w),\end{aligned}$$

where

$$\begin{aligned}\eta_T &= -\sum_{j=0}^{r+1}\binom{r+1}{j}(-1)^j\sum_{k=1}^{r}\frac{(1-e^{-iw})^{-k}}{(r-k+1)!}\prod_{s=0}^{r-k}\{m(T-j)+s\} \\ &\quad -\sum_{j=0}^{r+1}\binom{r+1}{j}(-1)^j(1-e^{-iw})^{-r-1}\left(1-e^{im(T-j)w}\right) \\ &= I+II.\end{aligned}$$

Note that for each k, the coefficient of $(1-e^{-iw})^{-k}/(r-k+1)!$ in I is $\sum_{j=0}^{r+1}\binom{r+1}{j}(-1)^j$ $\prod_{s=0}^{r-k}\{m(T-j)+s\} = \sum_{j=0}^{r+1}\binom{r+1}{j}(-1)^j\sum_{i=1}^{r-k+1}a_i m^i(T-j)^i = 0$, where the a_i's represent some constants and the last equality follows from equation (8). Therefore, $I=0$, and the fact that $\sum_{j=0}^{r+1}\binom{r+1}{j}(-1)^j=0$ can be used to simplify η_T as follows,

$$\begin{aligned}\eta_T &= \sum_{j=0}^{r+1}\binom{r+1}{j}(-1)^j(1-e^{-iw})^{-r-1}e^{im(T-j)w} \\ &= (1-e^{-iw})^{-r-1}e^{im(T-r-1)w}(e^{imw}-1)^{r+1} \\ &= (1-e^{iw})^{-r-1}e^{iw\{mT-(m-1)(r+1)\}}(1-e^{imw})^{r+1},\end{aligned}$$

which implies that, for $r\geq 0, T\geq r+2$ and $m=2h+1$,

$$\begin{aligned}&\nabla^r X_T^m \\ &= \int_{-\pi}^{\pi}(1-e^{iw})^{-r-1}e^{iw\{mT-(m-1)(r+1)\}}(1-e^{imw})^{r+1}dZ(w)\end{aligned}$$

$$
\begin{aligned}
&= \int_{-(2h+1)\pi}^{(2h+1)\pi} (1 - e^{iy/m})^{-r-1} e^{iy\{mT-(m-1)(r+1)\}/m} (1 - e^{iy})^{r+1} dZ\left(\frac{y}{m}\right) \\
&= \sum_{s=-h}^{h} \int_{-\pi}^{\pi} \left(1 - e^{i(w+2s\pi)/m}\right)^{-r-1} e^{iwT} e^{i(1-m)(r+1)(w+2s\pi)/m} (1 - e^{iw})^{r+1} dZ\left(\frac{w+2s\pi}{m}\right) \\
&= \int_{-\pi}^{\pi} e^{iwT} dZ_m(w),
\end{aligned}
$$

where

$$
dZ_m(w) = \sum_{s=-h}^{h} \left(1 - e^{i(w+2s\pi)/m}\right)^{-r-1} e^{i(1-m)(r+1)(w+2s\pi)/m} (1 - e^{iw})^{r+1} dZ\left(\frac{w+2s\pi}{m}\right),
$$

$-\pi < w < \pi$. Validity of the above equalities for all integer T follows from stationarity. Therefore, by (4.11.19) of Priestley (1981),

$$
\begin{aligned}
f_{r,m}(w)dw &= E\left(|dZ_m(w)|^2\right) \\
&= \frac{1}{m} \sum_{k=-h}^{h} \left|1 - e^{i(w+2k\pi)/m}\right|^{-2r-2d-2} \left|1 - e^{iw}\right|^{2r+2} g\left(\frac{w+2k\pi}{m}\right) dw \\
&= \frac{1}{m} \{2(1 - cosw)\}^{r+1} \sum_{k=-h}^{h} \left|2\sin\left(\frac{w+2k\pi}{2m}\right)\right|^{-2r-2d-2} g\left(\frac{w+2k\pi}{m}\right) dw,
\end{aligned}
$$

$-\pi < w < \pi$. This proves the result for $m = 2q + 1$. The proof for the case of $m = 2q$ is similar and hence omitted.

Proof of (b) Without loss of generality, consider $m = 2q + 1$, then

$$
m^{-2r-2d-1} f_{r,m}(w) = \{2(1 - cos(w)\}^{r+1} \sum_{k=-q}^{q} \left|2m\sin\left(\frac{w+2k\pi}{2m}\right)\right|^{-2r-2d-2} g\left(\frac{w+2k\pi}{2m}\right),
$$

which tends to $\{2(1 - \cos(w)\}^{r+1} \sum_{k=-q}^{q} |w + 2k\pi|^{-2r-2d-2} g(0)$ by the dominated convergence theorem, owing to (i) the inequality $|\sin(w)| \leq |w|$, (ii) the boundedness of g and its continuity at 0, and (iii) the fact that $\sum_{k=-\infty}^{\infty} |w + 2k\pi|^{-2r-2d-2} < \infty$ for $r \geq 0$ and $-1/2 < d < 1/2$. The convergence of the normalization constants of $f_{r,m}$ to K follows along similar arguments.

Proof of Theorem 3

The theorem follows readily from the expressions of $\sigma^2_{i,r,d}(m)$, $i = 1, 2$, to be derived below. Routine algebra shows that, for $r \geq 1$, the one-step prediction error of $\nabla^r X^m_{T+1}$ equals that of X^m_{T+1}. Hence, the one-step prediction variance of X^m_{T+1} by using the aggregate series $\{X^m_s, s \leq T\}$ can be computed by Kolmogorov's formula (Theorem 5.8.1. of Brockwell & Davis, 1991):

$$\begin{aligned}\sigma^2_{1,r,d}(m) &= 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_{r,m}(\lambda) d\lambda\right\} \\ &= m^{-1}2^{r+2}\pi \exp\left\{\frac{r+1}{2\pi}\int_{-\pi}^{\pi} \log(1-\cos\lambda) d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log \psi_m(\lambda) d\lambda\right\},\end{aligned}$$

where

$$\psi_m(w) = \sum_{k=-q}^{q} \left|2\sin\left(\frac{w+2k\pi}{2m}\right)\right|^{-2r-2d-2} g\left(\frac{w+2k\pi}{2m}\right).$$

Consequently,

$$\begin{aligned}\sigma^2_{1,r,d}(m) &= m^{2r+2d+1}2^{r+1}\sigma^2|\theta(1)|^2|\phi(1)|^{-2} \\ &\quad \times \exp\left\{\frac{r+1}{2\pi}\int_{-\pi}^{\pi} \log(1-\cos\lambda) d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log \sum_{k=-\infty}^{\infty} |w+2k\pi|^{-2r-2d-2} dw\right\} \\ &\quad + o(m^{2r+2d+1}).\end{aligned}$$

Now, we follow the technique of Man & Tiao (working paper, 2001) to compute $\sigma^2_{2,r,d}(m)$. For simplicity, write d_1 for $d+r$, and $\binom{d_1+k-1}{k}$ for $\Gamma(d_1+k)/\{\Gamma(k+1)\Gamma(d_1)\}$. First, consider the case that $r=0$ and $-1/2 < d < 1/2$. Let $\sum_{j=0}^{\infty}\chi_j z^j = \theta(z)/\phi(z)$ and $t \wedge s = \min(t,s)$. Clearly, $\chi_j \to 0$ exponentially fast. Then

$$\begin{aligned}X_T^m &= \sum_{i=0}^{m-1} B^i(1-B)^{-d_1}\theta(B)\phi^{-1}(B)\epsilon_{Tm} \\ &= \sum_{i=0}^{m-1} B^i \sum_{j=0}^{\infty}\binom{d_1+j-1}{j} B^j \sum_{s=0}^{\infty}\chi_s B^s \epsilon_{Tm} \\ &= \sum_{k=0}^{\infty}\sum_{s=0}^{k}\sum_{i=0}^{(k-s)\wedge(m-1)} \chi_s \binom{d_1+k-s-i-1}{k-s-i} B^k \epsilon_{Tm}.\end{aligned}$$

Let

$$\psi_k = \sum_{s=0}^{k}\sum_{i=0}^{(k-s)\wedge(m-1)} \chi_s \binom{d_1+k-s-i-1}{k-s-i}.$$

Then

$$X_{T+1}^m = \left(\sum_{k=0}^{m-1} + \sum_{k=m}^{\infty}\right)\psi_k B^k \epsilon_{(T+1)m},$$

and $\sigma^2_{2,r,d}(m) = \sigma^2\sum_{k=0}^{m-1}\psi_k^2$. Note that the preceding formula continues to hold even for the nonstationary case when $d_1 \geq 1/2$ (Brockwell & Davis, 1991, p. 318).

Using the equality that $\sum_{j=0}^{k} \binom{g+j-1}{j} = \binom{k+g}{k}$, we have, for $k \le m-1$,

$$\begin{aligned}
\psi_k &= \sum_{s=0}^{k} \sum_{i=0}^{k-s} \chi_s \binom{d_1 + k - i - s - 1}{k - i - s} \\
&= \sum_{s=0}^{k} \sum_{j=0}^{k-s} \chi_s \binom{d_1 + j - 1}{j} \\
&= \sum_{s=0}^{k} \chi_s \binom{d_1 + k - s}{k - s}.
\end{aligned}$$

For large k and fixed g, Stirling's formula (Feller 1968, p.52) implies that $\binom{g+k}{k} \approx k^g/\Gamma(g+1)$, i.e. the ratio of the two expressions is bounded and equals $1 + o(1)$. Hence,

$$\psi_k = k^{d_1} \sum_{s=0}^{k} \chi_s \frac{(1 - s/k)^{d_1}}{\Gamma(d_1 + 1)} (1 + o(1)) = (1 + o(1)) \frac{k^{d_1}}{\Gamma(d_1 + 1)} \sum_{s=0}^{k} \chi_s.$$

Therefore, for large m,

$$\begin{aligned}
\sigma^2_{2,r,d}(m) &= \sigma^2 \sum_{k=0}^{m-1} \psi_k^2 \\
&= \frac{\sigma^2}{\Gamma^2(d_1 + 1)} \sum_{k=0}^{m-1} \left\{ \left(\sum_{s=0}^{k} \chi_s \right)^2 k^{2d_1} (1 + o(1)) \right\} \\
&= \frac{\sigma^2 m^{2d_1+1}}{\Gamma^2(d_1 + 1)} \left(\sum_{s=0}^{m-1} \chi_s \right)^2 \sum_{k=0}^{m-1} \left\{ \left(\frac{\sum_{s=0}^{k} \chi_s}{\sum_{s=0}^{m-1} \chi_s} \right)^2 \left(\frac{k}{m} \right)^{2d_1} \frac{1}{m} (1 + o(1)) \right\} \\
&= \frac{\sigma^2 \left(\sum_{s=0}^{m-1} \chi_s \right)^2 m^{2d_1+1}}{\Gamma^2(d_1 + 1)} \left(\int_0^1 x^{2d_1} dx \right) (1 + o(1)) \\
&= \frac{\sigma^2 \left(\sum_{s=0}^{\infty} \chi_s \right)^2 m^{2d_1+1}}{(2d_1 + 1)\Gamma^2(d_1 + 1)} (1 + o(1)).
\end{aligned}$$

But $\sum_{s=0}^{\infty} \chi_s = |\theta(1)|/|\phi(1)|$. This proves the desired result.

References

[1] 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**, 4, 659-72.

[2] Beran, J., Bhansali, R.J. & Ocker, D. (1998). On unified model selection for stationary and nonstationary short-and long-memory autoregressive processes. *Biometrika* **85**, 4, 921-34.

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

[4] Chambers, M. J., (1996). The estimation of continuous parameter long-memory time series models. *Econometric Theory* **12**, 374-90.

[5] Feller, W. (1968). An introduction to probability theory and its applicaitons, Volume 1. New York: Wiley.

[6] Gradshteyn, I. S. & Ryzhik, I.M. (1994). *Table of integrals, series, and products.* Boston: Academic Press.

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

[8] Jolley, L.B.W. (1961). *Summation of Series.* New York: Dover.

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

[10] Ling, S. & Li, W. K., (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Amer. Statist. Assoc.* **92**, no. 439, 1184-1194.

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

[12] Tiao, G. C. (1972). Asymptotic behavior of temporal aggregates of time series. *Biometrika* **59**, No. 3, 525-531.