The Effects of Systematic Sampling and Temporal Aggregation on Discrete Time Long Memory Processes and Their Finite Sample Properties

March 1999

Soosung Hwang*

Department of Applied Economics
University of Cambridge
Austin Robin Building
Sidgwick Avenue
Cambridge CB3 9DE
UK
TEL) 44 1223 335282
FAX) 44 1223 335299
E-MAIL) sh211t@econ.cam.ac.uk

* I would like to thank Dr. D. Lee for making available his GAUSS programme, and Professor M. J. Chambers, Professor A. C. Harvey, Dr. S. E. Satchell, and an anonymous referee for helpful comments.

Abstract

This study investigates the effects of varying sampling intervals on the long memory characteristics of certain stochastic processes. We find that although different sampling intervals do not affect the decay rate of discrete time long memory autocorrelation functions in large lags, the autocorrelation functions in short lags are affected significantly. The level of the autocorrelation functions moves upward for temporally aggregated processes and downward for systematically sampled processes, and these effects result in a bias in the long memory parameter. For the ARFIMA(0,*d*,0) process, the absolute magnitude of the long memory parameter, |*d*|, of the temporally aggregated process is greater than the |*d*| of the true process, which is greater than the |*d*| of the systematically sampled process. We also find that the true long memory parameter can be obtained if we use a decay rate that is not affected by different sampling intervals.

1 Introduction

An important class of time series models are the so-called long memory processes which were introduced by Mandelbrot and Van Ness (1968), Granger and Joyeux (1980), and Hosking (1981). A simple property of long memory processes is that whilst the autocorrelations decrease, they decrease very slowly. Therefore, the past influences the future in a manner reminiscent of chaotic processes.

It is interesting to notice that many of the empirical studies use temporally aggregated data, such as monthly time series, for the test of long memory process. We now have available financial data that are sampled on many different frequencies. Consider foreign exchange data for example. This is now available on a quote-by-quote or trade-by-trade basis. It is also available on daily, weekly, and even much lower frequencies. This availability naturally raises questions of temporal aggregation in long memory processes.

Ding, Granger and Engle (1992) conjectured that temporal aggregation does not change the long memory property of the return series. Recently, Chambers (in press) shows that at low frequencies, the decay rate of the spectral density functions of long memory processes is not affected by sampling intervals. Therefore, the true long memory parameter can be estimated by considering low frequencies regardless of the sampling interval.

A sampling interval which is different from the dynamics of the true process affects the model specification of long memory processes. However, both the frequencies of the true process and the effects of the sampling interval on the model specification of long memory processes are not known. In this situation, an appropriate long memory model which takes into account the sampling interval is impossible. Most empirical studies which use long memory processes do not consider changes in model specification from varying sampling

intervals. For this reason, we assume that the sampling interval does not change the model specification for a long memory process and investigate changes in the long memory parameter. Thus, our focus is whether or not there is a change in the long memory parameter when different sampling intervals are used.

In what follows, we extend Chambers (in press) results by using both time and frequency domain analysis in discrete time long memory processes[1]. Firstly, the effects of systematic sampling and temporal aggregation on discrete time long memory processes are presented in an analytical way; autocovariance, autocorrelation, and spectral density functions are derived for the systematically sampled and temporally aggregated long memory processes. Then, Monte Carlo simulations using frequency domain maximum likelihood estimation methods are used for the finite sample properties of systematic sampling and temporal aggregation effects on discrete time long memory processes.

Our study confirms the results of Chambers (in press), in that the decay rate of discrete time long memory spectral density functions at low frequencies (or equivalently, the decay rate of discrete time long memory autocorrelation functions in large lags) is not affected by systematic sampling or temporal aggregation. However, significant effects of systematic sampling or temporal aggregation are found in the autocorrelation functions at short lags, which result in either upward or downward movements in the level of autocorrelation function. We find that the effects are different for systematically sampled and temporally aggregated long memory processes. The absolute value of the autocorrelation coefficient of systematically sampled long memory processes becomes smaller as the sampling interval increases, while the absolute value of the autocorrelation coefficient of temporally aggregated long memory processes becomes larger as the sampling interval increases. These effects result in a bias in the long memory parameter[2]. In systematically

sampled long memory processes, the long memory parameter is always biased towards zero as the sampling interval increases, while in temporally aggregated long memory processes the absolute value of the long memory parameter, $|d|$, is larger than its true value. Although it is not reported in this paper, we find that there is no temporal aggregation effect on continuous time long memory processes.

We also investigate whether or not the true long memory parameter can be obtained from the decay rate of long memory spectral density functions at low frequencies in finite samples. For this purpose, we use the semiparametric regression model suggested by Geweke and Porter-Hudak (1983). Our simulation results using the semiparametric regression for discrete time long memory processes show that the estimation biases are much less than the biases of ML estimates. Therefore, the true long memory parameter can be estimated in temporally aggregated long memory processes. However, the standard deviation of the estimates is very large and significant estimation bias still exists in systematically sampled long memory processes.

2 Discrete Time Long Memory Processes

A model appears to be more attractive when just one parameter is used for long range dependence. A long memory process is characterized as a process that should explain long range dependence between observations more effectively than conventional short memory processes. There are two major models for long memory process, continuous time models such as the fractional Gaussian noise (FGN) model introduced by Mandelbrot and Van Ness (1968), and discrete time models such as the autoregressive fractionally integrated moving average (ARFIMA) model introduced by Granger and Joyeux (1980) and Hosking (1981).

Although the FGN model has the benefit that it can be used together with ordinary Brownian motion, the ARFIMA model is generally preferred to the FGN model. The main reason is that the former can describe economic and financial time series better than the latter. Moreover, the ARFIMA model is a generalisation of the more familiar autoregressive integrated moving average (ARIMA) model; it captures both long and short memory, and it is easier to use than the FGN model.

Discrete time long memory processes have much more flexibility than ARIMA(p,d,q) models. The order of integration in the ARIMA process is confined only to integer values such as 0 or 1. In this process a shock has mean-reversion with an exponential decay rate when d=0, but has infinite persistence when d=1. This knife-edge distinction between ARIMA(p,0,q) and ARIMA(p,1,q) may be too narrow. Long memory processes add more flexibility by allowing for fractional orders of integration such as $0 \leq d \leq 1$.

A discrete time long memory process (fractionally integrated process), which was introduced by Granger and Joyeux (1980) and Hosking (1981), is defined to be a discrete time stochastic process which is represented as

$$\nabla^d x_t = (1-L)^d x_t = u_t, \qquad (1)$$

where L is the lag operator, u_t is an independent identically distributed random variable, and $-0.5<d<0.5$[3]. Using the binomial series expension, the fractional difference operator ∇^d can be represented as

$$\begin{aligned} \nabla^d &= (1-L)^d \\ &= \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} L^j, \end{aligned} \qquad (2)$$

where $\Gamma(\cdot)$ is the gamma function. The infinite moving average representation of x_t may be denoted by[4]

$$x_t = \Psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}\,, \tag{3}$$

where $\psi_j = \dfrac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}$ and $\psi_j \approx \dfrac{j^{d-1}}{\Gamma(d)}$ as $j \to \infty$ via Stirling's approximation.

The autocovariance, autocorrelation and spectral density functions of the fractionally integrated process for $s \geq 0$ and $0 < \omega \leq \pi$ are[5]

$$\begin{aligned} \gamma_{FI}(s) &= \frac{\Gamma(1-2d)\Gamma(s+d)}{\Gamma(d)\Gamma(1-d)\Gamma(s+1-d)}\sigma_u^2 \\ &\approx \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} s^{2d-1}\sigma_u^2, \qquad \text{as } s \to \infty, \end{aligned} \tag{4}$$

$$\begin{aligned} \rho_{FI}(s) &= \frac{\Gamma(1-d)\Gamma(s+d)}{\Gamma(d)\Gamma(s+1-d)} \\ &\approx \frac{\Gamma(1-d)}{\Gamma(d)} s^{2d-1}, \qquad \text{as } s \to \infty, \end{aligned} \tag{5}$$

$$\begin{aligned} S_{FI}(\omega) &= \frac{\sigma_u^2}{2\pi}\{2\sin(\frac{\omega}{2})\}^{-2d} \\ &\approx \frac{\sigma_u^2}{2\pi}\omega^{-2d}, \quad \text{as } \omega \to 0, \end{aligned} \tag{6}$$

where s is a lag between observations, ω is a frequency, and σ_u^2 is a variance of u_t. We can see that the autocorrelations of the fractionally integrated series decline at a slower rate than that of the autoregressive moving average (ARMA) model. The autocorrelation function (5) decays at a hyperbolic rate, while that of the ARMA model decays exponentially.

Fractionally integrated processes show different characteristics depending on the parameter d. A fractionally integrated process is covariance stationary and invertible when $-0.5 < d < 0.5$, and it is a long memory process when d lies between 0 and 0.5. The fractional differencing parameter d is defined by the behaviour of the series up to infinite cycles. As d goes to 0.5, the decay rate of the impact of a unit innovation becomes slower. Hence, the

fractional differencing parameter d decides the decay of the system's response to the innovation. Sowell (1990) shows that, while the variance of the partial sums of variables grows linearly with number of observations when $d = 0$, it grows faster than a linear rate when $0<d<0.5$. On the other hand, when $-0.5<d<0$, the process has short memory since each shock is negatively correlated with the others, thus making the variance of the partial sums of variables less than the variance of the individual shock.

Fractionally integrated processes can easily be generalised to the ARFIMA(p,d,q) process. More formally,

$$\Phi(L)(1-L)^d x_t = \Theta(L)u_t \qquad (7)$$

where $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$, $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_p L^q$, $-0.5<d<0.5$, and all the roots of $\Phi(L)$ and $\Theta(L)$ lie outside of the unit circle. In the ARFIMA model, d may be chosen to describe the autocorrelation structure of distant observations of a time series, while the θ and ϕ parameters can be chosen to describe the low lag autocorrelation structure. Therefore, the ARFIMA model not only overcomes the drawbacks that Mcleod and Hipel (1978) point out but also generalises the Box-Jenkins ARIMA model where only integral differencing is allowed.

3 The Effects of Systematic Sampling and Temporal Aggregation on Discrete Time Long Memory Processes

The sampling interval of observed economic and financial time series (e.g., daily, weekly, and monthly) does not necessarily correspond to the true unknown interval. We conjecture that the sampling interval is longer than the true interval, resulting in temporal aggregation in flow time series and systematic sampling in stock time series. We next investigate the effects of systematic sampling and temporal aggregation on discrete time long memory processes

under the assumption that the sampling interval is longer than the true interval.

This section shows the effects of systematic sampling and temporal aggregation on discrete time long memory processes: a systematically sampled ARFIMA(0,*d*,0) process and a temporally aggregated ARFIMA(0,*d*,0) process. For these two processes, properties such as autocovariance, autocorrelation, and spectral density functions are derived.

Assumption 1 *The dynamics of the true underlying discrete time process* x_t^D *take place at every unknown* $1/\eta$ *period where* η *is a positive integer, while the observations of the process take place at every unit time period.*

The underlying process is simply thought of as a series of logarithmic changes of the original positive time series at discrete time intervals. For example, it is a return in financial markets or a growth rate in economics.

In the assumption above, η is a sampling interval. More specifically, η is the number of times the dynamics of the true underlying process take place between observations. When η=1, the observed time series is equivalent to the true process. When η is a positive integer greater than 1, then the time series is observed less frequently than the frequency of the true process.

The true underlying process can be partially observed in two ways, systematic sampling and temporal aggregation. A systematically sampled process is a sequence of the true process at observation points and, thus, consists of every other η of the true process. A temporally aggregated process is a sequence of the aggregated true process between sampling intervals. Therefore, if the dynamics of the true process take place at every day, for example,

the systematically sampled process is like observing a daily process every Monday but not on other days and the temporally aggregated process is a weekly return process.

Definition 1 *A discrete systematically sampled process, whose true process has 1/η dynamic periods, consists of* $\{ x_t^{DS} = x_t^D ; t=1,2,3,... \}$. *A discrete aggregated process, whose true process has 1/η dynamic periods, consists of* $\{ x_t^{DA} = \sum_{k=0}^{\eta-1} x^D_{t-\frac{k}{\eta}} ; t=1,2,3,... \}$.

When the true process has *1/η* dynamic periods, the discrete true ARFIMA(0,*d*,0) process, x_t^{DF}, is defined as

$$(1-L^{\frac{1}{\eta}})^d x_t^{DF} = \varepsilon_t, \qquad t = 1,\ 1+1/\eta,\ 1+2/\eta,\ 1+3/\eta,\ \ldots, \qquad (8)$$

where ε_t is a white noise sequence with variance σ^2. Note that the variance σ^2 is measured with time interval *1/η*. When we assume that the frequency of the process defined in section 2 is *1/η* , the variance of the white noise, σ_u^2, in section 2 is identical to σ^2 in equation (8). Thus, without loss of generality, we assume $\sigma_u^2 = \sigma^2$ throughout this section.

We represent a theorem concerning autocovariance, autocorrelation, and spectral density functions of a systematically sampled ARFIMA(0,*d*,0) process. This generalises theorem 1 (a) of Chambers (in press), where he derives the spectral density functions of a systematically sampled ARFIMA(0,*d*,0) process.

Theorem 1 Systematically Sampled ARFIMA(0,*d*,0) Process

Under assumption 1 and definition 1, the discrete systematically sampled ARFIMA(0,d,0) process, x_t^{DSF}, *is represented as*

$$x_t^{DSF} = \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} \varepsilon_{t-\frac{k}{\eta}} \qquad t = 1, 2, 3, \ldots \tag{9}$$

The autocovariance, $\gamma_{DSF}(s,\eta)$, *autocorrelation,* $\rho_{DSF}(s,\eta)$, *and spectral density functions,* $S_{DSF}(\omega,\eta)$, *of* x_t^{DSF} *for* $s \geq 0$, $\eta \geq 1$, *and* $0 < \omega \leq \pi$ *are given by*

$$\gamma_{DSF}(s,\eta) = \frac{\Gamma(1-2d)\Gamma(\eta s+d)}{\Gamma(d)\Gamma(1-d)\Gamma(\eta s+1-d)}\sigma^2 \tag{10}$$

$$\approx \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\sigma^2\eta^{2d-1}s^{2d-1}, \quad as \quad \eta \text{ or } s \to \infty,$$

$$\rho_{DSF}(s,\eta) = \frac{\Gamma(1-d)\Gamma(\eta s+d)}{\Gamma(d)\Gamma(\eta s+1-d)} \tag{11}$$

$$\approx \frac{\Gamma(1-d)}{\Gamma(d)}\eta^{2d-1}s^{2d-1}, \quad as \quad \eta \text{ or } s \to \infty,$$

$$S_{DSF}(\omega,\eta) = \frac{\sigma^2}{2\pi}\left(2\sin\frac{\omega}{2\eta}\right)^{-2d} \tag{12}$$

$$\approx \frac{\sigma^2}{2\pi}\eta^{2d}\omega^{-2d}, \quad as \quad \omega \to 0,$$

where η *is a sampling interval,* s *is a lag between observations,* ω *is a frequency, and* σ^2 *is a variance of* ε_t.

Proof : See the Appendix.

The properties of the discrete systematically sampled ARFIMA(0,*d*,0) process have a sampling interval η that the true discrete ARFIMA(0,*d*,0) process does not have (see equations (4), (5), and (6)). The true process is observed differently by the sampling interval. As pointed out in Chambers (in press), the decay rate in the spectral density function at low frequencies, *-2d*, is the same as that of the true ARFIMA(0,*d*,0) process. The autocorrelation function for large *s* also confirms that the decay rate is not changed by systematic sampling

(compare equation (11) with equation (5)).

However, the absolute value of the autocorrelation function of the systematically sampled process is always less than that of the true ARFIMA(0,*d*,0) process, since for any positive integer $\eta>1$, η^{2d-1} is less than 1 when $-0.5<d<0.5$. This decreased autocorrelation function is expected to make the estimated *d* biased toward zero. Figures 1 to 4 show the changes of the autocorrelation functions for systematically sampled and aggregated ARFIMA(0,*d*,0) processes when $d=0.35(\eta=5)$ and $d=-0.35(\eta=5)$, respectively. In figures 1 and 3, we can see that the absolute values of the autocorrelation functions are decreased by systematic sampling. Figures 2 and 4 show the ratios of the autocorrelations of x_t^{DSF} to those of the true process in various lags, which are always less than 1. Therefore, although we may estimate the true *d* by considering low frequencies or equivalently remote autocorrelations, a systematically sampled ARFIMA(0,*d*,0) process becomes less persistent than the true ARFIMA(0,*d*,0) process.

A discrete aggregated ARFIMA(0,*d*,0) process, x_t^{DAF}, can be obtained by summing a discrete true ARFIMA(0,*d*,0) process, x_t^{DF}, up to $(\eta-1)/\eta$ lags.

$$\begin{aligned} x_t^{DAF} &= \Sigma_{j=0}^{\eta-1} x_{t-j/\eta}^{DF} \\ &= (1 + L^{\frac{1}{\eta}} + L^{\frac{2}{\eta}} + \ldots + L^{\frac{\eta-1}{\eta}})(1 - L^{\frac{1}{\eta}})^{-d} \varepsilon_t \\ &= (1 - L)(1 - L^{\frac{1}{\eta}})^{-d-1} \varepsilon_t \end{aligned} \qquad (13)$$

We present a theorem concerning autocovariance, autocorrelation, and spectral density functions of a temporally aggregated ARFIMA(0,*d*,0) process. This generalises theorem 1 (b) of Chambers (in press), where he only derives spectral density functions of temporally aggregated and systematically sampled ARFIMA(0,*d*,0) processes.

Theorem 2 Temporally Aggregated ARFIMA(0,*d*,0) Process

Under assumption 1 and definition 1, the discrete aggregated ARFIMA(0,d,0) process, x_t^{DAF}, *is defined as*

$$x_t^{DAF} = \sum_{j=0}^{\eta-1}\sum_{k=0}^{\infty}\frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}\varepsilon_{t-\frac{k}{\eta}-\frac{j}{\eta}} \qquad t = 1,\ 2,\ 3,\ \ldots \tag{14}$$

The autocovariance, $\gamma_{DAF}(s,\eta)$, *autocorrelation,* $\rho_{DAF}(s,\eta)$, *and spectral density functions,* $S_{DAF}(\omega,\eta)$, *of* x_t^{DAF} *for* $s\geq 0$, $\eta\geq 1$, *and* $0<\omega\leq\pi$ *are given by*

$$\begin{aligned}\gamma_{DAF}(s,\eta) = \frac{\sigma^2\Gamma(1-2d)}{2(1+2d)\Gamma(1+d)\Gamma(1-d)}\Big[&\frac{\Gamma(1+\eta s+d+\eta)}{\Gamma(\eta s-d+\eta)} \\ &+\frac{\Gamma(1+\eta s+d-\eta)}{\Gamma(\eta s-d-\eta)}-2\frac{\Gamma(1+\eta s+d)}{\Gamma(\eta s-d)}\Big],\end{aligned} \tag{15}$$

$$\rho_{DAF}(s,\eta) = \frac{\dfrac{\Gamma(1+\eta s+d+\eta)}{\Gamma(\eta s-d+\eta)}+\dfrac{\Gamma(1+\eta s+d-\eta)}{\Gamma(\eta s-d-\eta)}-2\dfrac{\Gamma(1+\eta s+d)}{\Gamma(\eta s-d)}}{2\dfrac{\Gamma(1+d+\eta)}{\Gamma(-d+\eta)}-2\dfrac{\Gamma(1+d)}{\Gamma(-d)}} \tag{16}$$

$$\approx \frac{\eta^{1+2d}[(s+1)^{1+2d}+(s-1)^{1+2d}-2s^{1+2d}]}{2\eta^{1+2d}-2\dfrac{\Gamma(1+d)}{\Gamma(-d)}} \qquad \textit{for large } \eta$$

$$\approx \frac{\eta^{1+2d}d(1+2d)}{\eta^{1+2d}-\dfrac{\Gamma(1+d)}{\Gamma(-d)}}s^{2d-1}, \qquad as \quad s\to\infty,$$

using Taylor series up to second order,

$$S_{DAF}(\omega,\eta) = \frac{\sigma^2}{2\pi}\left(2\sin\frac{\omega}{2\eta}\right)^{-2(d+1)}\left(2\sin\frac{\omega}{2}\right)^2 \tag{17}$$

$$\approx \frac{\sigma^2}{2\pi}\eta^{2d+2}\omega^{-2d}, \qquad as \quad \omega\to 0,$$

where η *is a sampling interval, s is a lag between observations,* ω *is a frequency, and* σ^2 *is a variance of* ε_t.

Proof : See the Appendix.

The limiting spectral density and autocorrelation functions show that the decay rate of the ARFIMA(0,*d*,0) process is not changed by temporal aggregation, identical to results in theorem 1 (b) of Chambers (in press). However, as we can see in figure 1, when *d*=0.35 and η=5, the autocorrelation function of the aggregated process is always larger than that of the true process. For positive *d*, $\frac{\eta^{1+2d} d(1+2d)}{\eta^{1+2d} - \frac{\Gamma(1+d)}{\Gamma(-d)}}$ in equation (16) is always larger than its true value $\frac{\Gamma(1-d)}{\Gamma(d)}$ in (5). This is shown in figure 2 by the ratio of the autocorrelation of the aggregated ARFIMA(0,*d*,0) process to that of the true ARFIMA(0,*d*,0) process when *d*=0.35 and η=5. Therefore, although the limiting autocorrelation and spectral density functions do not show any change in the decay rate, the level of the autocorrelation function moves upward, making the aggregated ARFIMA(0,*d*,0) process more persistent than the true ARFIMA(0,*d*,0) process. On the other hand, when -0.5<*d*<0, the ratio of the autocorrelation function of the aggregated ARFIMA(0,*d*,0) process to that of the true ARFIMA(0,*d*,0) process does not show a consistent pattern; see figure 4. However, the pattern of the autocorrelation function of the aggregated ARFIMA(0,*d*,0) process in figure 3 is typical for a larger negative *d* value. Therefore, for negative *d*, temporal aggregation makes the true ARFIMA(0,*d*,0) process have a higher negative *d*. As a result, we conclude that, the absolute value of the discrete time long memory parameter |*d*| of the aggregated ARFIMA(0,*d*,0) process is larger than that of its true process. We present our conclusions in Proposition 1.

Proposition 1

1. Systematic Sampling Effect

In the discrete systematically sampled ARFIMA(0,d,0) processes, the absolute values of d are biased toward zero, as the sampling interval η increases.

2. Temporal Aggregation Effect

In the discrete aggregated ARFIMA(0,d,0) process, the absolute value of d increases as the sampling interval η increases.

A referee suggests that the ARFIMA(p,d,q) model rather than the ARFIMA(0,d,0) model would be better to specify the autocorrelation functions of systematically sampled or temporally aggregated discrete time long memory processes. Figures 2 and 4 support this suggestion; the ratios of autocorrelation of systematically sampled and temporally aggregated ARFIMA(0,d,0) processes are not constant over lags. That is, the ratios vary for short lags, but they approach a constant for larger lags. In addition, the major changes in the levels of the autocorrelation of systematically sampled and temporally aggregated ARFIMA(0,d,0) processes occur in the first lag. Therefore, although the autocorrelation functions of systematically sampled and temporally aggregated ARFIMA(0,d,0) processes are $O(s^{2d-1})$ as $s\rightarrow\infty$, autocorrelation levels of systematically sampled and temporally aggregated ARFIMA(0,d,0) processes are changed in short lags, generating a short memory process. In these cases, ARFIMA(0,d,0) processes together with appropriate short memory processes may be better specified than ARFIMA(0,d,0) processes for systematically sampled and temporally aggregated ARFIMA(0,d,0) processes. However, we do not know what the exactly specified model is or the dynamic period of the true underlying process. In this study, we investigate the effects of sampling interval on the long memory parameter when sampling intervals are disregarded.

Whilst the above analysis reflects the impact of temporal aggregation on the true model, it does not deal with issues of estimation. We consider this problem in the next section.

4 Finite Sample Properties of the Effects of Systematic Sampling and Temporal Aggregation on Discrete Time ARFIMA(0,*d*,0) Processes

4.1 Maximum Likelihood Estimation

Many studies of long memory processes have used the frequency domain approach, since the spectral generating function is generally easy to evaluate. Time domain maximum likelihood is computationally difficult to use for large sample sizes because of the inversion of the $T \times T$ covariance matrix. The frequency domain approximate log-likelihood is represented as follows:

$$\ln L(\theta) = -\frac{1}{2}\sum_{k=1}^{T-1} \ln s(\omega_k) - \frac{1}{2}\sum_{k=1}^{T-1}\frac{I(\omega_k)}{s(\omega_k)}, \qquad (18)$$

where $I(\omega_k) = \dfrac{\left|\sum_{t=1}^{T} e^{i\omega_k t}(X_t - \overline{X_T})\right|^2}{2\pi T}$, $s(\omega_k)$ is a spectral density function, and $\omega_k = \dfrac{2\pi k}{T}$, and $k = 1, 2, \ldots, T-1$.

Fox and Taqqu (1986) prove that the estimates obtained using only the last term of equation (18) are consistent and asymptotically normal. These results are extended by Dahlhaus (1989) as in equation (18). Cheung and Diebold (1994) compare the exact time domain ML estimates of Sowell (1992a) with the approximate frequency domain ML estimates of Fox and Taqqu (1986). When the mean of a process is known, the time domain ML estimates are superior. However, time domain ML estimates with a known mean are not

feasible in practice. Time domain ML estimates with a sample mean do not seem to be preferred to frequency domain ML estimates. In addition, when the sample size is more than 150 and $d>0$, there is little difference between the two, and the frequency domain ML estimates are less biased than the time domain ML estimates. Therefore, Cheung and Diebold (1994) suggest that the frequency domain ML estimation method is an attractive and efficient alternative when large sample sizes are used. Hauser (1992) and Lee (1994) also find similar results.

The recursive Levinson (1947)-Durbin (1960) algorithm is used for the generation of the ARFIMA(0,d,0) process. The Cholesky decomposition is not a suitable method here, because, as noted in GPH, it needs memory proportional to T^2 and computation time proportional to T^3. Therefore, when as in this study T=1000, it becomes impractical to use the Cholesky decomposition.

The GAUSS computer package is used for all computations. We use the same GAUSS code as Lee (1994) for the generation of fractionally integrated processes. For the numerical optimization method, the Broyden, Fletcher, Goldfard and Shanno (BFGS) algorithm is used. However, when the BFGS algorithm cannot find the optimum, the Davidon, Fletcher and Powell (DFP) algorithm, or the Newton-Raphson algorithm, or the steepest descent method are used in that order. Convergence tolerance for the gradient of estimated coefficients is set to 10^{-5}.

Sampling intervals of η = 1, 5, 10, 15, 20 are explored. First, a sample size of T=1,000 is generated and then, the original time series are transformed into systematically sampled or temporally aggregated processes. Systematically sampled processes are obtained by using the following formulation:

$$x_{t,\eta}^{S} = x_{\eta t} \tag{19}$$

where $x_{t,\eta}^{s}$ is a systematically sampled process with a sampling interval η and x_t is the originally generated series. When η=1, the frequencies of the true process are the same as the observation frequencies. However, when η>1, the transformed process becomes a systematically sampled process with sampling interval η. To explore temporal aggregation effects, the original process is temporally aggregated as follows[6]:

$$x_{t,\eta}^{A} = \sum_{k=0}^{\eta-1} x_{\eta t-k} \tag{20}$$

Since the sample size decreases according to the increase in the sampling interval, the sample sizes for the sampling intervals η=1, 5, 10, 15, 20 are 1,000, 200, 100, 66, 50, respectively.

For the ARFIMA(0,*d*,0) process, equation (18) is used for the likelihood function. The spectral density functions for the ARFIMA(0,*d*,0) process are

$$S(\omega_k) = \frac{\sigma_\eta^2}{2^{2d+1}\pi(1-\cos(\omega_k))^{2d}} \tag{21}$$

where σ_η^2 is the variance of the white noise variables of systematically sampled or temporally aggregated ARFIMA(0,*d*,0) processes. Since the exact values of the variance are not known for systematically sampled and temporally aggregated processes[7], we estimate $\{d, \sigma_\eta^2\}$ for the ARFIMA(0,*d*,0) model for all η.

As shown in section 3, the autocorrelations of systematically sampled and aggregated processes are also a function of the parameters *d*. Therefore, ten points in both the parameter *d* of the ARFIMA process are explored: *d*=±0.05, ±0.15, ±0.25, ±0.35, ±0.45. For each $\{d, \eta\}$ set, 1000 Monte Carlo replications are performed. For the starting values of *d*, the true *d* values are used. Sample variances of systematically sampled or temporally aggregated

processes are used for the starting values of σ_η^2. The bias, standard deviation, and mean square error (MSE) of the estimated *d* are computed.

The results of 1000 replications of the ML estimation of the ARFIMA(0,*d*,0) processes are reported in tables 1 to 3. Table 1 shows the finite sample properties of the frequency domain approximate ML estimation for the original ARFIMA(0,*d*,0) process at sample sizes of *T*=1000, 200, 100, 66, 50, respectively. The results for the ARFIMA(0,*d*,0) process are consistent with those of Hauser (1992) and Lee (1994). As expected, the estimation bias and standard deviation of estimates tend to increase, as the sample size becomes smaller.

Table 2 shows the simulation results for the finite sample properties of the systematically sampled process. As explained in the previous section, for the systematically sampled process, long memory parameters are biased toward zero. The finite sample properties of the systematic sampling effects can be summarised as follows. Firstly, the systematic sampling effect appears in the ARFIMA(0,*d*,0) process. When the dynamics of the true process are more frequent than the observations of the process, the estimates of the observed systematically sampled process are always biased toward zero. Table 2 shows that the bias is larger in $-0.5<d<0$ than in $0<d<0.5$. This is because in $-0.5<d<0$, long memory processes have similar properties to those of short memory processes.

Our second investigation looks at the temporal aggregation effects, see table 3. As expected, in the aggregated ARFIMA(0,*d*,0) process, the aggregation effect has upward bias. The simulation results support what we found in the section 3, that is, the absolute magnitude of the long memory parameter, $|d|$, of the aggregated ARFIMA(0,*d*,0) process is greater than the $|d|$ of the true ARFIMA(0,*d*,0) process.

Table 3 show that the magnitude of the bias depends on the sampling interval; the bias of the aggregated ARFIMA(0,*d*,0) process increases as η increases. However, after some sampling intervals, it shows little difference. This can be explained by the autocorrelation function of (16). As η becomes infinitely large, the limiting autocorrelation function can be represented as

$$\rho^{DAF}(s,\eta,d) \approx \frac{\eta^{1+2d} d(1+2d)s^{2d-1}}{\eta^{1+2d} - \dfrac{\Gamma(1+d)}{\Gamma(-d)}} \qquad (22)$$

$$\approx d(1+2d)s^{2d-1}, \qquad \text{as} \quad \eta \to \infty.$$

Therefore, for an infinitely large η the autocorrelation is not a function of η. Note that the above limiting autocorrelation function is the same as that of the discrete time fractional Gaussian noise process, see Mandelbrot and Van Ness (1969). The dynamics of the true process which have an infinitestimal interval in continuous time long memory processes are equivalent to an infinitely large η in discrete time long memory processes.

4.2 Further Considerations

Chambers (in press) shows that the true long memory parameter *d* can be estimated regardless of sampling interval, since the decay rates of the spectral density function are not affected by systematic sampling or temporal aggregation. Our results agree with his for both spectral density and autocorrelation functions; see equations (11), (12), (16), and (17).

To further investigate whether or not estimates obtained using low frequencies reflect the true long memory parameter *d*, we use simulations with the semiparametric regression analysis of Geweke and Porter-Hudak (1983) (GPH). GPH concentrate on low frequencies because *d* is determined by the spectral density near zero frequencies. Their equation can be

represented as[8]

$$ln\ I(\omega_j) = c - d\ ln(2\sin(\frac{\omega_j}{2}))^2 + \varepsilon_j \qquad (23)$$

where c is a constant. The long memory parameter d can be estimated with j=1,2,...,$g(T)$, where $\lim_{T\to\infty} g(T) = \infty$ and $\lim_{T\to\infty} g(T)/T = 0$. That is, $g(T)$ should be sufficiently small compared with T. Here, $g(T)$[9] is set equal to $T^{0.5}$. As presented in equation (23), the semiparametric method of GPH uses only the decay rate of fractionally integrated processes. Since systematic sampling and temporal aggregation do not affect the decay rate of the true long memory process, the GPH method is a more appropriate method to estimate the true unobserved long memory parameter than parametric methods such as those used in the previous subsection.

Four points, d=±0.15, ±0.35 at sampling intervals η=1, 5, 10, 15, 20 are considered in both the systematically sampled and temporally aggregated ARFIMA(0,d,0) processes. Other simulation conditions are the same as those of the ML estimation of subsection 4.1.

Tables 4 to 6 show the results of 1,000 replications of GPH's semiparametric estimation. Table 4 represents finite sample properties of GPH's estimates. The estimation bias is small and seems robust to changes in sample size and d[10]. The standard deviation is also robust to the changes in the value of the true parameter. However, the standard deviations are larger than those of the ML estimates in table 1 and seem very large for small samples. This is because the number of periodograms used in GPH estimation is only $T^{0.5}$. Therefore, the MSEs of the GPH estimates are high compared with those of ML estimates and the GPH estimates are not efficient for small samples[11].

Let us first consider the finite sample properties of systematically sampled processes. Table 5 shows that the estimation bias is still present. Although the estimation biases are less than that of the ML estimates (see table 2), they are still very large. In addition, for negative d

values, there is little difference between GPH and ML estimates. This may be because for negative *d*, the discrete time long memory process has the properties of a short memory process rather than those of a long memory process. However, as expected, for a temporally aggregated long memory process, using low frequencies gives estimates close to the true long memory parameter. Table 6 shows that the bias of estimates obtained using low frequencies is small in the aggregated long memory process.

Therefore, for the finite samples considered in this study, the true *d* may not be obtained when observed time series are systematically sampled. On the other hand, for the temporally aggregated time series, we may obtain the true *d* by considering low frequencies. However, note that standard deviations are relatively very high in the GPH's semiparametric estimates.

Chambers (in press) concentrates only on the decay rate of long memory processes near the zero frequency and suggests that the true long memory parameter can be estimated using the decay rate of the spectral density function at low frequencies. Our finite sampling simulations show that GPH semiparametric estimates using only low frequencies may fail to obtain the true long memory parameter when the observed time series is systematically sampled or the number of observations is small. In addition, the true long memory parameter may not be obtained when $-0.5<d<0$.

5 Conclusion

This study investigated the effects of systematic sampling and temporal aggregation on long memory processes under the assumption that there was no change in model specification caused by systematic sampling or temporal aggregation. From the theoretical explanation in

section 3 and the simulations conducted in section 4, the following results are obtained: for the ARFIMA(0,*d*,0) process, the absolute value of the long memory parameter, |*d*|, of the temporally aggregated process is larger than the |*d*| of the true process which is larger than the |*d*| of the systematically sampled process.

Our results are consistent with those of Chambers (in press) and Ding, Granger and Engle (1992) who conjecture that temporal aggregation does not change the decay rate of autocorrelations. One of the interesting properties of discrete time long memory processes is that although the autocorrelation level is affected by systematic sampling or temporal aggregation, the decay rate of the autocorrelations in remote lags or equivalently, the decay rate of spectral densities at low frequencies is not affected by the sampling interval. For the finite samples used in this study, GPH's semiparametric regression gives a long memory parameter close to its true value for the temporally aggregated process, while the estimates are still biased toward zero for the systematically sampled long memory process.

However, it is worth pointing out that the autocorrelation function of discrete time long memory processes shifts downwards or upwards by systematic sampling or temporal aggregation. In this respect, although we may estimate the true long memory parameter using the decay rates, we conclude that the sampling interval does impact on the long memory parameter if we do not consider the changes in model specification. Empirical studies on the existence of long memory processes in economic and financial time series may be affected by the sampling interval. The probability that we find evidence of a long memory process increases as the sampling interval increases for a temporally aggregated time series, if there exists long memory. Our results in section 4 do suggest that for a large number of observations such as in financial time series, the true long memory parameter may be estimated using the semiparametric regression of Geweke and Porter-Hudak (1983).

NOTES

[1] The results for continuous time long memory processes can be obtained from the author on request.

[2] We use 'bias' for the difference between the true long memory parameter values and the long memory parameter values of the systematically sampled or temporally aggregated long memory processes. 'Bias' used in this study means a model misspecification bias. We use 'estimation bias' for the difference between the true values and the estimated parameter values to differentiate the 'bias' in this study.

[3] Although d may be any real number, d is assumed to lie in the interval (-0.5, 0.5) with a finite number of difference.

[4] Moving average (MA) representation will be used for our explanation. The autoregressive (AR) representation can be inferred from the MA representation. That is, the d of the MA representation is equal to $-d$ of the AR representation.

[5] See Granger and Joyeux (1980) and Hosking (1981) for proof.

[6] Sample size of both systematically sampled and aggregated processes when η=15 is 66 and only the first 990 observations from the original sample size of 1000 are used.

[7] Only when η=1, σ_1^2 is known to be 1.

[8] This equation is a logarithmic transformation of equation (6).

[9] The GPH method has some difficulties in selecting $g(T)$; see Sowell (1992b) and Hurvich and Beltrao (1993). GPH suggest that $g(T)=T^{0.5}$ is appropriate and Diebold and Rudebusch (1989) and Cheung (1993) also show that the exponent (0.5) is appropriate.

[10] The biases of the GPH estimates are found in the presence of AR or MA processes (Agiakloglou, Newbold and Wohar, 1993) and in either non-invertible (d=-0.5) or nonstationary (d=0.5) ARFIMA(0,d,0) processes (Hurvich and Ray, 1995).

[11] See Sowell (1992a) for the simulation results of the finite sample properties of GHP, time domain ML, and frequency domain ML estimates. Also see Robinson (1994, 1995) and Hidalgo and Yajima (1997) for further discussion of the GPH estimation and its modification.

REFERENCES

Agiakloglou, C., P. Newbold, and M. Wohar (1993) Bias in an Estimator of the Fractional Difference Parameter. *Journal of Time Series Analysis* 14, 235-246.

Chambers, M. (in press) Long Memory and Aggregation in Macroeconomic Time Series. *International Economic Review*.

Chambers, M. (1996) The Estimation of Continuous Parameter Long-memory Time Series Models. *Econometric Theory* 12, 374-390.

Cheung, Y. (1993) Long Memory in Foreign-Exchange Rates. *Journal of Business and Economic statistics* 11, 93-101.

Cheung, Y., and F. Diebold (1994) On Maximum Likelihood Estimation of the Differencing Parameter of Fractionally-Integrated Noise with Unknown Mean. *Journal of Econometrics* 62, 301-316.

Dahlhaus, R. (1989) Efficient Parameter Estimation for Self-Similar Processes. *The Annals of Statistics* 17, 1749-1766.

Diebold, F. X., and G. D. Rudebusch (1989) Long Memory and Persistence in Aggregate Output. *Journal of Monetary Economics* 24, 189-209.

Ding, Z., C. Granger, and R. F. Engle (1992) A Long memory Property of Stock Market Returns and a New Model. Discussion paper 92-21, Department of Economics, University of California, San Diego.

Fox, R., and M. S. Taqqu (1986) Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series. *The Annals of Statistics* 14,

517-532.

Geweke, J., and S. Porter-Hudak (1983) The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis* 4, 221-238

Gradshteyn, I. S., and I. M. Ryzhik (1994) Table of Integrals, Series, and Products. Fifth Edition, Academic Press.

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

Hauser, M. A. (1992) Long Range Dependence in International Output Series: A Reexamination. Working Paper, Department of Statistics, University of Economics and Business Administration.

Hidalgo, J., and Y. Yajima (1996) Semiparametric Estimation of the Long-range Parameter. Unpublished manuscript.

Hosking, J. (1981) Fractional Differencing. *Biometrika* 68, 165-76

Hurvich, C. M., and K. I. Beltrao (1993) Asymptotics for the Low-Frequency Ordinates of the Periodogram of Long-Memory Time Series. *Journal of Time Series Analysis* 14, 455-472.

Hurvich, C. M., and B. K. Ray (1994) Estimation of the Memory Parameter for Nonstationary or Noninvertible Fractionally Integrated Processes. *Journal of Time Series Analysis* 16, 17-41.

Lee, D. (1994) Asymptotic Theory for Long-Memory Time Series. Michigan State University PhD Dissertation.

Mandelbrot, B., and J. Van Ness (1968) Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review* 10, 422-437

Mcleod, A. I., and K. W. Hipel (1978) Preservation of the Rescaled Adjusted range, 1, A Reassessment of the Hurst Phenomenon. *Water Resource Research* 14, 491-508.

Robinson, P. M. (1994) Semiparametric Analysis of Long-memory Time Series. *Annals of Statistics* 22, 515-539.

Robinson, P. M. (1995) Log-periodogram Regression of Time Series with Long Range Dependence. *Annals of Statistics* 23, 1048-1072.

Sowell, F. (1990) The Fractional Unit Root Distribution. *Econometrica* 58, 495-505.

Sowell, F. (1992a) Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models. *Journal of Econometrics* 53, 165-188

Sowell, F. (1992b) Modeling Long-Run Behavior with the Fractional ARIMA Model. *Journal of Monetary Economics* 29, 277-302.

APPENDIX

Proof of Theorem 1 Systematically Sampled ARFIMA(0,*d*,0) Process.

The spectral density function of x^{DSF} in equation (12) is easily obtained from equation (8). The autocovariance function of x^{DSF}, $\gamma_{DSF}(s,\eta)$, has an ηs lag between observations in the true process. That is, a lag of magnitude *s* between observations in systematically sampled ARFIMA(0,*d*,0) is equivalent to a lag of magnitude ηs in the true ARFIMA(0,d,0) process. Therefore, we can use equation (4) to derive $\gamma_{DSF}(s,\eta)$ by simply replacing *s* with ηs. Mathematically,

$$
\begin{aligned}
\gamma_{DSF}(s,\eta) &= Cov(x_t^{DSF}, x_{t-s}^{DSF}) \\
&= E[(\varepsilon_t + \frac{\Gamma(1+d)}{\Gamma(d)\Gamma(2)}\varepsilon_{t-\frac{1}{\eta}} + .. + \frac{\Gamma(\eta s+d)}{\Gamma(d)\Gamma(\eta s+1)}\varepsilon_{t-s} + \frac{\Gamma(\eta s+d+1)}{\Gamma(d)\Gamma(\eta s+2)}\varepsilon_{t-s-\frac{1}{\eta}} + ..)\times \\
&\qquad (\varepsilon_{t-s} + \frac{\Gamma(1+d)}{\Gamma(d)\Gamma(2)}\varepsilon_{t-s-\frac{1}{\eta}} + \frac{\Gamma(2+d)}{\Gamma(d)\Gamma(3)}\varepsilon_{t-s-\frac{2}{\eta}} + ..)] \\
&= \frac{\sigma^2}{\Gamma(d)^2}[\frac{\Gamma(\eta s+d)\Gamma(d)}{\Gamma(\eta s+1)\Gamma(1)} + \frac{\Gamma(\eta s+1+d)\Gamma(1+d)}{\Gamma(\eta s+2)\Gamma(2)} \\
&\qquad + \frac{\Gamma(\eta s+2+d)\Gamma(2+d)}{\Gamma(\eta s+3)\Gamma(3)} + ...] \\
&= \frac{\sigma^2}{\Gamma(d)^2}\frac{\Gamma(\eta s+d)\Gamma(d)}{\Gamma(\eta s+1)}[1 + \frac{(\eta s+d)d}{(\eta s+1)1} + \frac{(\eta s+d)(\eta s+1+d)d(1+d)}{(\eta s+1)(\eta s+2)1\cdot 2} + ...] \\
&= \frac{\sigma^2}{\Gamma(d)^2}\frac{\Gamma(\eta s+d)\Gamma(d)}{\Gamma(\eta s+1)}F(d,\eta s+d;\eta s+1;1) \\
&= \frac{\sigma^2}{\Gamma(d)^2}\frac{\Gamma(\eta s+d)\Gamma(d)}{\Gamma(\eta s+1)}\frac{\Gamma(\eta s+1)\Gamma(1-2d)}{\Gamma(\eta s+1-d)\Gamma(1-d)} \\
&= \frac{\Gamma(\eta s+d)\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)\Gamma(\eta s+1-d)}\sigma^2
\end{aligned}
\tag{A1}
$$

where $F(\alpha,\beta;\gamma;1)$ is a *hypergeometric function* and

$$F(\alpha,\beta;\gamma;1) = 1+\frac{\alpha\cdot\beta}{\gamma\cdot 1}+\frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)\cdot 1\cdot 2}+\frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{\gamma(\gamma+1)(\gamma+2)\cdot 1\cdot 2\cdot 3}+\ldots$$
$$= \frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)} \qquad [\gamma > \alpha+\beta] \qquad \text{(A2)}$$

(9.122, p1068, Gradshteyn and Ryzhik, 1994).

The autocorrelation function in equation (11) follows straightforwardly.■

Proof of Theorem 2 Temporally Aggregated ARFIMA(0,*d*,0) Process.

Let us define $\gamma_{DF}(s)$ as the autocovariance function of the true underlying fractionally integrated process x_t^{DF}, where s is a lag between observations. That is,

$$\gamma_{DF}(s) = Cov(x_t^{DF}, x_{t-s/\eta}^{DF})$$
$$= \frac{\Gamma(1-2d)\Gamma(s+d)}{\Gamma(d)\Gamma(1-d)\Gamma(s+1-d)}\sigma^2.$$

This is actually the same as equation (4) as explained after Definition 1 of page 10. Then the autocovariance function of x_t^{DAF} is

$$\gamma_{DAF}(s,\eta) = Cov(\sum_{k=0}^{\eta-1} x_{t-k/\eta}^{DF}, \sum_{k=0}^{\eta-1} x_{t-s-k/\eta}^{DF})$$
$$= -\eta\gamma_{DF}(\eta s)+\sum_{i=0}^{\eta-1}(\eta-i)\gamma_{DF}(\eta s-i)+\sum_{i=0}^{\eta-1}(\eta-i)\gamma_{DF}(\eta s+i)$$
$$= -\eta\gamma_{DF}(\eta s)+\sum_{i=0}^{\eta-1}\sum_{k=1}^{\eta-i}\gamma_{DF}(\eta s-i)+\sum_{i=0}^{\eta-1}\sum_{k=1}^{\eta-i}\gamma_{DF}(\eta s+i)$$
$$= \frac{\sigma^2\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\sum_{k=1}^{\eta}[-\frac{\Gamma(d+\eta s)}{\Gamma(1-d+\eta s)}+\sum_{i=0}^{\eta-k}\frac{\Gamma(d+\eta s-i)}{\Gamma(1-d+\eta s-i)} \qquad \text{(A3)}$$
$$+\sum_{i=0}^{-k}\frac{\Gamma(d+\eta s+i)}{\Gamma(1-d+\eta s+i)}]$$
$$= \frac{\sigma^2\Gamma(1-2d)}{2(1+2d)\Gamma(1+d)\Gamma(1-d)}[\frac{\Gamma(1+\eta s+d+\eta)}{\Gamma(\eta s-d+\eta)}$$
$$+\frac{\Gamma(1+\eta s+d-\eta)}{\Gamma(\eta s-d-\eta)}-2\frac{\Gamma(1+\eta s+d)}{\Gamma(\eta s-d)}],$$

using the following Lemma of Sowell (1990) twice;

$$\sum_{k=1}^{N}\frac{\Gamma(a+k)}{\Gamma(b+k)}=\frac{1}{1+a-b}[\frac{\Gamma(1+a+N)}{\Gamma(b+N)}-\frac{\Gamma(1+a)}{\Gamma(b)}]. \tag{A4}$$

The variance function of x_t^{DAF} is

$$\gamma_{DAF}(0,\eta)=\frac{\sigma^2\Gamma(1-2d)}{2(1+2d)\Gamma(1+d)\Gamma(1-d)}[\frac{\Gamma(1+d+\eta)}{\Gamma(-d+\eta)}+\frac{\Gamma(1+d-\eta)}{\Gamma(-d-\eta)}-2\frac{\Gamma(1+d)}{\Gamma(-d)}] \tag{A5}$$

Using the following equation (see 8.334, p946, Gradshteyn and Ryzhik, 1994)

$$\Gamma(1-x)\Gamma(x)=\frac{\pi}{sin(\pi x)} \tag{A6}$$

we have

$$\Gamma(1+d-\eta)\Gamma(-d+\eta)=\frac{\pi}{sin(\pi(-d+\eta))} \tag{A7}$$

$$\Gamma(1+d+\eta)\Gamma(-d-\eta)=\frac{\pi}{sin(\pi(-d-\eta))} \tag{A8}$$

Since η is a positive integer, equation (A7) is the same as equation (A8). Equating and rearranging the two equations, we have

$$\frac{\Gamma(1+d-\eta)}{\Gamma(-d-\eta)}=\frac{\Gamma(1+d+\eta)}{\Gamma(-d+\eta)} \tag{A9}$$

Therefore, we have

$$\gamma_{DAF}(0,\eta)=\frac{\sigma^2\Gamma(1-2d)}{2(1+2d)\Gamma(1+d)\Gamma(1-d)}[2\frac{\Gamma(1+d+\eta)}{\Gamma(-d+\eta)}-2\frac{\Gamma(1+d)}{\Gamma(-d)}] \tag{A10}$$

The autocorrelation function of x_t^{DAF} is obtained using autocovariance function (A3) and variance function (A9) of x_t^{DAF}, and the spectral density function of x_t^{DAF} follows from equation (13).■

Table 1 Finite Sample Properties of the Frequency Domain Approximate Maximum Likelihood Estimation of ARFIMA(0,*d*,0) Processes

1. Bias

	Sample Size				
True *d*	1000	200	100	66	50
0.45	0.0015	-0.0015	-0.0084	-0.0124	-0.0173
0.35	0.0000	-0.0041	-0.0103	-0.0141	-0.0188
0.25	-0.0010	-0.0068	-0.0132	-0.0165	-0.0216
0.15	-0.0017	-0.0083	-0.0154	-0.0187	-0.0249
0.05	-0.0020	-0.0089	-0.0163	-0.0199	-0.0261
-0.05	-0.0019	-0.0087	-0.0161	-0.0199	-0.0258
-0.15	-0.0015	-0.0077	-0.0146	-0.0186	-0.0239
-0.25	-0.0007	-0.0056	-0.0117	-0.0157	-0.0203
-0.35	0.0008	-0.0022	-0.0072	-0.0111	-0.0146
-0.45	0.0035	0.0032	-0.0006	-0.0045	-0.0066

2. Standard Deviation

	Sample Size				
True *d*	1000	200	100	66	50
0.45	0.0268	0.0671	0.1036	0.1267	0.1544
0.35	0.0264	0.0662	0.0995	0.1259	0.1546
0.25	0.0261	0.0656	0.0993	0.1264	0.1543
0.15	0.0260	0.0655	0.0989	0.1266	0.1529
0.05	0.0259	0.0655	0.0990	0.1267	0.1532
-0.05	0.0258	0.0655	0.0992	0.1268	0.1534
-0.15	0.0258	0.0656	0.0995	0.1270	0.1538
-0.25	0.0258	0.0657	0.1000	0.1274	0.1546
-0.35	0.0260	0.0661	0.1008	0.1282	0.1558
-0.45	0.0265	0.0671	0.1024	0.1296	0.1574

3. Mean-Squared Error

	Sample Size				
True *d*	1000	200	100	66	50
0.45	0.0007	0.0045	0.0108	0.0162	0.0241
0.35	0.0007	0.0044	0.0100	0.0161	0.0243
0.25	0.0007	0.0043	0.0100	0.0162	0.0243
0.15	0.0007	0.0044	0.0100	0.0164	0.0240
0.05	0.0007	0.0044	0.0101	0.0164	0.0242
-0.05	0.0007	0.0044	0.0101	0.0165	0.0242
-0.15	0.0007	0.0044	0.0101	0.0165	0.0242
-0.25	0.0007	0.0044	0.0101	0.0165	0.0243
-0.35	0.0007	0.0044	0.0102	0.0166	0.0245
-0.45	0.0007	0.0045	0.0105	0.0168	0.0248

Notes: The results are based on 1000 replications of frequency domain ML estimates.

Table 2 Finite Sample Properties of the Frequency Domain Approximate Maximum Likelihood Estimation of Systematically Sampled ARFIMA(0,*d*,0) Processes

1. Bias

	Sampling Interval (τ)				
True *d*	1	5	10	15	20
0.45	0.0015	-0.1188	-0.1638	-0.2001	-0.2200
0.35	0.0000	-0.1320	-0.1798	-0.2159	-0.2335
0.25	-0.0010	-0.1289	-0.1690	-0.1996	-0.2114
0.15	-0.0017	-0.1007	-0.1244	-0.1472	-0.1518
0.05	-0.0020	-0.0441	-0.0512	-0.0679	-0.0684
-0.05	-0.0019	0.0355	0.0390	0.0251	0.0261
-0.15	-0.0015	0.1289	0.1370	0.1238	0.1248
-0.25	-0.0007	0.2289	0.2379	0.2243	0.2249
-0.35	0.0008	0.3315	0.3396	0.3253	0.3255
-0.45	0.0035	0.4348	0.4411	0.4262	0.4262

2. Standard Deviation

	Sampling Interval (τ)				
True *d*	1	5	10	15	20
0.45	0.0268	0.0623	0.0981	0.1263	0.1523
0.35	0.0264	0.0609	0.0959	0.1258	0.1509
0.25	0.0261	0.0601	0.0951	0.1243	0.1499
0.15	0.0260	0.0606	0.0955	0.1238	0.1490
0.05	0.0259	0.0622	0.0961	0.1226	0.1495
-0.05	0.0258	0.0631	0.0965	0.1217	0.1510
-0.15	0.0258	0.0632	0.0967	0.1212	0.1529
-0.25	0.0258	0.0628	0.0965	0.1209	0.1548
-0.35	0.0260	0.0623	0.0963	0.1208	0.1562
-0.45	0.0265	0.0618	0.0962	0.1209	0.1571

3. Mean-Squared Error

	Sampling Interval (τ)				
True *d*	1	5	10	15	20
0.45	0.0007	0.0180	0.0364	0.0560	0.0716
0.35	0.0007	0.0211	0.0415	0.0624	0.0773
0.25	0.0007	0.0202	0.0376	0.0553	0.0672
0.15	0.0007	0.0138	0.0246	0.0370	0.0452
0.05	0.0007	0.0058	0.0119	0.0196	0.0270
-0.05	0.0007	0.0052	0.0108	0.0154	0.0235
-0.15	0.0007	0.0206	0.0281	0.0300	0.0390
-0.25	0.0007	0.0564	0.0659	0.0649	0.0746
-0.35	0.0007	0.1138	0.1246	0.1204	0.1304
-0.45	0.0007	0.1929	0.2038	0.1963	0.2063

Notes: The results are based on 1000 replications of frequency domain ML estimates. Sample sizes for sampling intervals 1, 5, 10, 15, and 20 are 1000, 200, 100, 66, and 50, respectively.

Table 3 Finite Sample Properties of the Frequency Domain Approximate Maximum Likelihood Estimation of Temporally Aggregated ARFIMA(0,*d*,0) Processes

1. Bias

True *d*	Sampling Interval (τ)				
	1	5	10	15	20
0.45	0.0015	0.0916	0.1041	0.0977	0.0909
0.35	0.0000	0.0694	0.0802	0.0776	0.0687
0.25	-0.0010	0.0480	0.0562	0.0540	0.0463
0.15	-0.0017	0.0267	0.0307	0.0262	0.0177
0.05	-0.0020	0.0052	0.0046	-0.0030	-0.0119
-0.05	-0.0019	-0.0165	-0.0224	-0.0335	-0.0427
-0.15	-0.0015	-0.0380	-0.0500	-0.0651	-0.0745
-0.25	-0.0007	-0.0587	-0.0774	-0.0968	-0.1068
-0.35	0.0008	-0.0773	-0.1028	-0.1259	-0.1370
-0.45	0.0035	-0.0910	-0.1223	-0.1473	-0.1598

2. Standard Deviation

True *d*	Sampling Interval (τ)				
	1	5	10	15	20
0.45	0.0268	0.0708	0.1038	0.1348	0.1603
0.35	0.0264	0.0673	0.0995	0.1288	0.1530
0.25	0.0261	0.0655	0.0969	0.1280	0.1530
0.15	0.0260	0.0652	0.0962	0.1272	0.1520
0.05	0.0259	0.0652	0.0963	0.1265	0.1532
-0.05	0.0258	0.0651	0.0962	0.1267	0.1542
-0.15	0.0258	0.0651	0.0961	0.1271	0.1554
-0.25	0.0258	0.0652	0.0965	0.1278	0.1568
-0.35	0.0260	0.0657	0.0976	0.1290	0.1584
-0.45	0.0265	0.0675	0.1004	0.1314	0.1599

3. Mean-Squared Error

True *d*	Sampling Interval (τ)				
	1	5	10	15	20
0.45	0.0007	0.0134	0.0216	0.0277	0.0340
0.35	0.0007	0.0093	0.0163	0.0226	0.0281
0.25	0.0007	0.0066	0.0125	0.0193	0.0255
0.15	0.0007	0.0050	0.0102	0.0169	0.0234
0.05	0.0007	0.0043	0.0093	0.0160	0.0236
-0.05	0.0007	0.0045	0.0097	0.0172	0.0256
-0.15	0.0007	0.0057	0.0117	0.0204	0.0297
-0.25	0.0007	0.0077	0.0153	0.0257	0.0360
-0.35	0.0007	0.0103	0.0201	0.0325	0.0439
-0.45	0.0007	0.0129	0.0250	0.0390	0.0511

Notes: The results are based on 1000 replications of frequency domain ML estimates. Sample sizes for sampling intervals 1, 5, 10, 15, and 20 are 1000, 200, 100, 66, and 50, respectively.

Table 4 Finite Sample Properties of the Semiparametric Regression Analysis of GPH for ARFIMA(0,*d*,0) Processes

1. Bias

	Sample Size				
True *d*	1000	200	100	66	50
0.35	0.0023	0.0087	0.0134	0.0168	0.0065
0.15	-0.0038	-0.0021	0.0026	0.0061	-0.0044
-0.15	-0.0051	0.0045	0.0055	0.0003	-0.0065
-0.35	0.0084	0.0268	0.0252	0.0186	0.0132

2. Standard Deviation

	Sample Size				
True *d*	1000	200	100	66	50
0.35	0.1378	0.2360	0.2921	0.3546	0.3908
0.15	0.1372	0.2426	0.2946	0.3495	0.3981
-0.15	0.1404	0.2384	0.2943	0.3577	0.4045
-0.35	0.1415	0.2365	0.2909	0.3484	0.3976

3. Mean Squared Error

	Sample Size				
True *d*	1000	200	100	66	50
0.35	0.0190	0.0558	0.0855	0.1260	0.1528
0.15	0.0189	0.0589	0.0868	0.1222	0.1585
-0.15	0.0197	0.0569	0.0866	0.1279	0.1637
-0.35	0.0201	0.0567	0.0853	0.1217	0.1582

Notes: The results are based on 1000 replications of GPH semiparametric regression estimates.

Table 5 Finite Sample Properties of Semiparametric Regression Analysis of GPH for Systematically Sampled ARFIMA(0,*d* ,0) Processes

1. Bias

	Sampling Interval (τ)				
True *d*	1	5	10	15	20
0.35	0.0023	-0.0567	-0.1212	-0.1533	-0.2013
0.15	-0.0038	-0.0796	-0.1101	-0.1229	-0.1355
-0.15	-0.0051	0.1374	0.1413	0.1539	0.1477
-0.35	0.0084	0.3413	0.3422	0.3489	0.3527

2. Standard Deviation

	Sampling Interval (τ)				
True *d*	1	5	10	15	20
0.35	0.1378	0.2311	0.3067	0.3508	0.3976
0.15	0.1372	0.2332	0.2842	0.3438	0.3924
-0.15	0.1404	0.2362	0.2871	0.3399	0.3935
-0.35	0.1415	0.2472	0.2852	0.3454	0.3773

3. Mean Squared Error

	Sampling Interval (τ)				
True *d*	1	5	10	15	20
0.35	0.0190	0.0567	0.1088	0.1466	0.1986
0.15	0.0189	0.0607	0.0929	0.1333	0.1723
-0.15	0.0197	0.0747	0.1024	0.1392	0.1767
-0.35	0.0201	0.1776	0.1984	0.2410	0.2668

Notes: The results are based on 1000 replications of semiparametric regression estimates. Sample sizes for sampling intervals 1, 5, 10, 15, and 20 are 1000, 200, 100, 66, and 50, respectively.

Table 6 Finite Sample Properties of Semiparametric Regression Analysis of GPH for Temporally Aggregated ARFIMA(0,d,0) Processes

1. Bias

	Sampling Interval (τ)				
True d	1	5	10	15	20
0.35	0.0023	0.0033	0.0019	0.0125	0.0105
0.15	-0.0038	-0.0061	-0.0073	-0.0043	-0.0071
-0.15	-0.0051	-0.0145	-0.0238	-0.0276	-0.0377
-0.35	0.0084	-0.0036	-0.0229	-0.0382	-0.0493

2. Standard Deviation

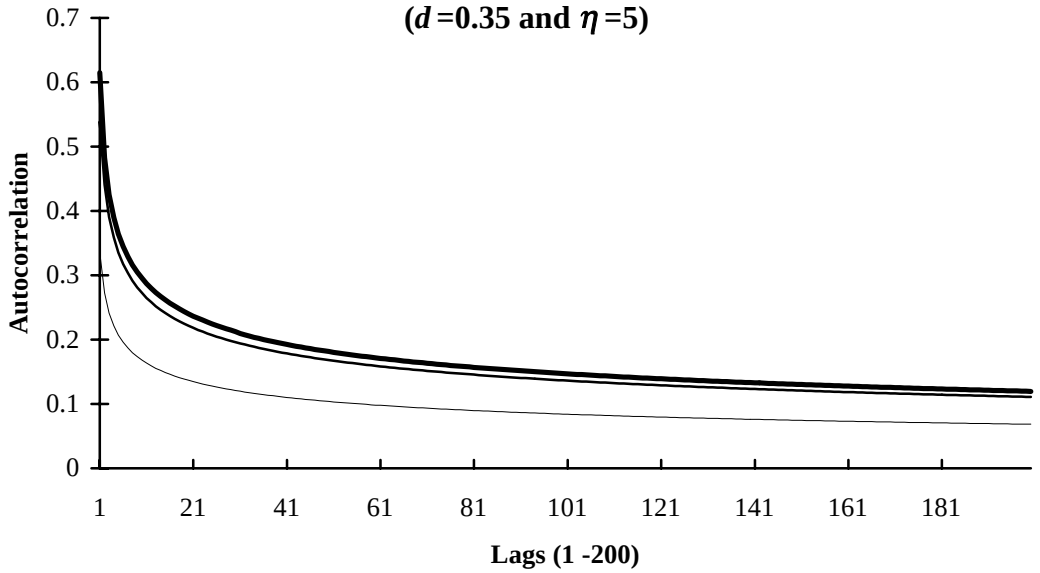
	Sampling Interval (τ)				
True d	1	5	10	15	20
0.35	0.1378	0.2351	0.2918	0.3562	0.4002
0.15	0.1372	0.2342	0.2904	0.3650	0.3973
-0.15	0.1404	0.2377	0.3041	0.3627	0.4132
-0.35	0.1415	0.2423	0.2954	0.3545	0.3970

3. Mean Squared Error

	Sampling Interval (τ)				
True d	1	5	10	15	20
0.35	0.0190	0.0553	0.0851	0.1270	0.1603
0.15	0.0189	0.0549	0.0844	0.1332	0.1579
-0.15	0.0197	0.0567	0.0930	0.1323	0.1722
-0.35	0.0201	0.0587	0.0878	0.1271	0.1601

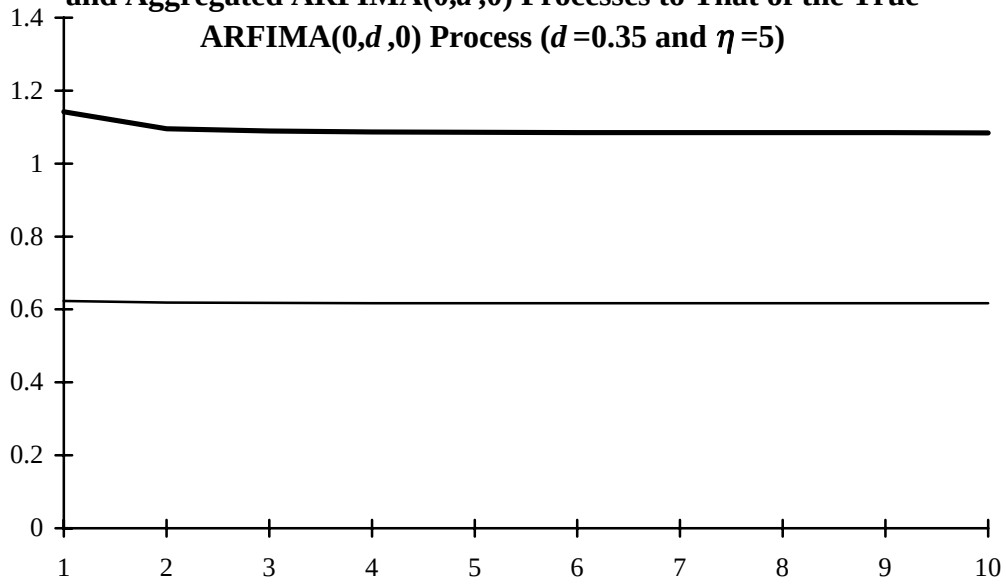
Notes: The results are based on 1000 replications of semiparametric regression estimates. Sample sizes for sampling intervals 1, 5, 10, 15, and 20 are 1000, 200, 100, 66, and 50, respectively.

Figure 1 Autocorrelation Functions of True, Systematically Sampled and Aggregated ARFIMA(0,d,0) Processes (d=0.35 and η=5)

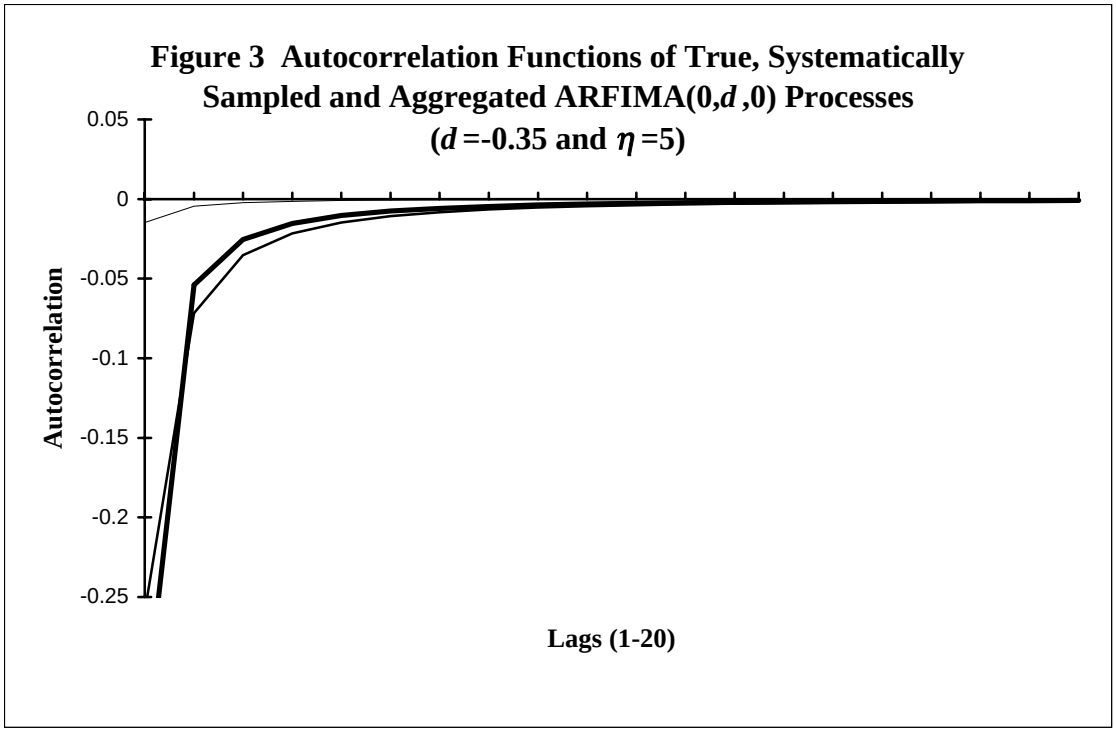


Notes : The upper line represents the autocorrelation function of the aggregated ARFIMA (0,d,0) process, the middle line represents the autocorrelation function of the true ARFIMA(0,d,0) process, and the lower line represents the autocorrelation function of the systematically sampled ARFIMA(0,d,0) process. The autocorrelations up to 30 lags are calculated using exact autocorrelation functions and autocorrelations from 31 to 200 lags are calculated using limiting autocorrelation functions (see equations (11) and (16)).

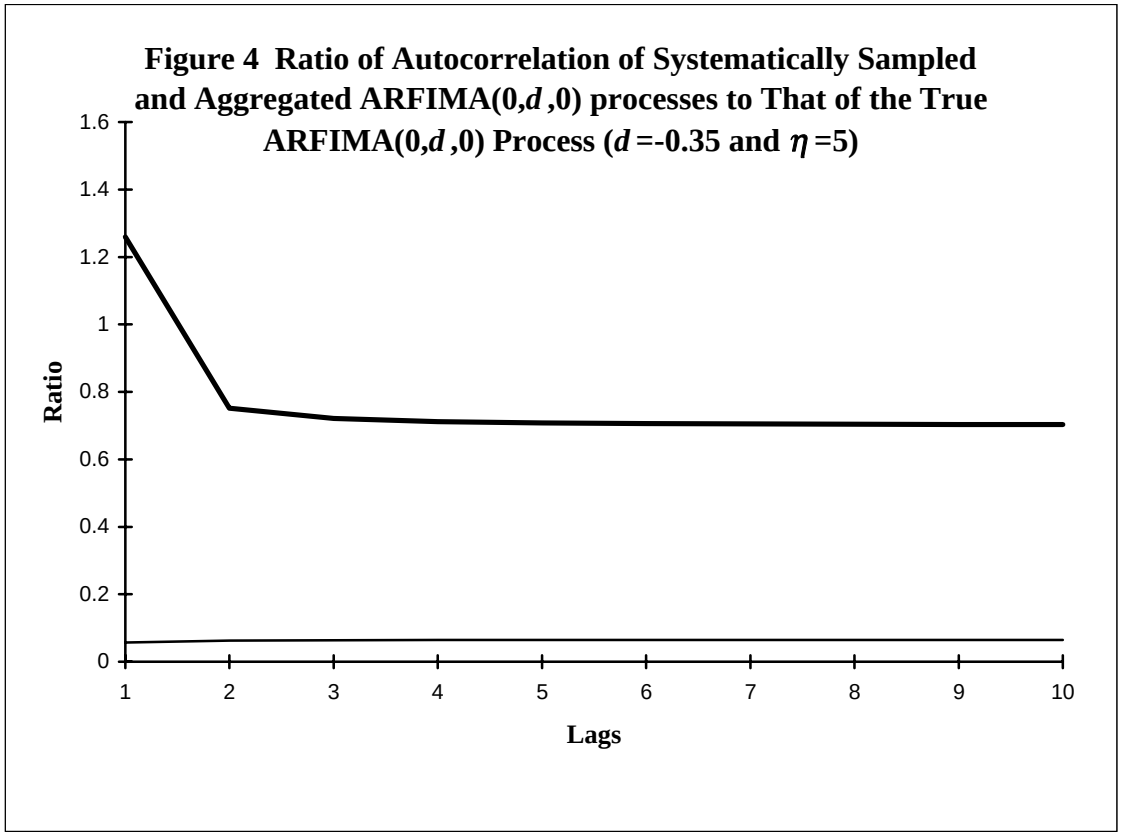
Figure 2 Ratio of Autocorrelation of Systematically Sampled and Aggregated ARFIMA(0,d,0) Processes to That of the True ARFIMA(0,d,0) Process (d=0.35 and η=5)



Notes : The upper line represents the ratio of autocorrelations of the aggregated ARFIMA (0,d,0) process to those of its true ARFIMA (0,d,0) process. The lower line represents the ratio of autocorrelations of the systematically sampled ARFIMA(0,d,0) process to those of the true ARFIMA(0,d,0) process.



Notes : The thick line represents the autocorrelation function of the aggregated ARFIMA (0,d,0) process, the middle line represents the autocorrelation function of the true ARFIMA(0,d,0) process, and the thin (upper) line represents the autocorrelation function of the systematically sampled ARFIMA(0,d,0) process. The autocorrelations are calculated using exact autocorrelation functions.



Notes : The upper line represents the ratio of autocorrelations of the aggregated ARFIMA (0,d,0) process to those of its true ARFIMA (0,d,0) process. The lower line represents the ratio of autocorrelations of the systematically sampled ARFIMA(0,d,0) process to those of the true ARFIMA(0,d,0) process.