2-5-2009

First-Order Bias Correction for Fractionally Integrated Time Series

Jaechoul Lee
Boise State University

Kyungduk Ko
Boise State University

First-order bias correction for fractionally integrated time series

Jaechoul Lee and Kyungduk Ko*

Department of Mathematics, Boise State University
Boise, ID 83725-1555. U.S.A.

September 11, 2009

SUMMARY

Most of the long memory estimators for stationary fractionally integrated time series models are known to experience non-negligible bias in small and finite samples. Simple moment estimators are also vulnerable to such bias, but can easily be corrected. In this paper, we propose bias reduction methods for a lag-one sample autocorrelation-based moment estimator. In order to reduce the bias of the moment estimator, we explicitly obtain the exact bias of lag-one sample autocorrelation up to the order n^{-1}. An example where the exact first-order bias can be noticeably more accurate than its asymptotic counterpart, even for large samples, is presented. We show via a simulation study that the proposed methods are promising and effective in reducing the bias of the moment estimator with minimal variance inflation.

Key words: Autoregressive fractionally integrated moving average; Bias correction; Long memory; Sample autocorrelations.

*Corresponding author: Kyungduk Ko, Department of Mathematics, Boise State University, 1910 University Dr., Boise, ID 83725-1555, U.S.A. E-mail: ko@math.boisestate.edu. Phone: 1-208-426-1123. Fax: 1-208-426-1356.

1 Introduction

Time series data realized from a long memory process have the distinctive feature that the autocorrelations between distant observations are not negligible in a sense that these autocorrelations are not summable. Due to the infinite sum of the autocorrelations, the spectral density of the long memory process approaches infinity as the frequency goes to zero. As a result, typical regularity conditions for time series limit theorems do not immediately apply. While these theoretical difficulties are inherent, the long memory models are popular in many disciplines such as finance, hydrology, and engineering.

The most versatile long memory time series model is the autoregressive fractionally integrated moving average process, ARFIMA(p, d, q), with $d \in (0, 0.5)$, which is first introduced by Granger & Joyeux (1980) and Hosking (1981). As the long memory features of the processes are subject to the fractional differencing parameter d, the accurate estimation procedure for d is crucial.

Parametric inferences for the ARFIMA(p, d, q) model involve calculation of the exact likelihood and its maximization with respect to the model parameters in time or frequency domains. In an early stage of the parametric estimation for Gaussian ARFIMA(p, d, q) models, approximate maximum likelihood (ML) methods based on Whittle's (1951) results were used by Fox & Taqqu (1986) and Li & McLeod (1986). However, these methods perform inaccurately in finite samples. Sowell (1992) calculated the exact form of the variance/covariance functions for Gaussian ARFIMA(p, d, q) processes and obtained the ML estimates of the model parameters. This exact likelihood method achieves relative accuracy, but is computationally exhaustive. As Bayesian approaches, Pai & Ravishanker (1996) adopted the Metropolis algorithm to estimate the model parameters, and Koop, Ley, Osiewalski & Steel (1997) used the importance sampling method with Sowell's (1992) exact form of variance/covariance functions. On the other hand, semiparametric estimation methods have been widely used in the literature due to their simplicity and computational speed. A typical one is the GPH estimator (Geweke & Porter-Hudak 1983) which uses the least squares estimation of the log-periodogram on low frequencies. Künsch (1987) and Robinson (1995) developed local Whittle estimators. However, an appropriate choice of 'm', which is the maximum number of the frequency index in Fourier transform, $\lambda_j = 2\pi j/n,\ j = 1, \ldots, m$, is an issue.

The performance of the existing estimators for the long memory parameter d has been demonstrated in a wealth of literature, including Hauser (1997, 1999) and Reisen, Abraham & Lopes (2001). Since most of these estimators have been developed based on asymptotic theory, and long memory estimators intrinsically need a 'large' sample to attain adequate accuracy, non-negligible biases are commonly present in the estimators, particularly for small and moderate samples (cf. Lieberman 2001; Nielsen & Frederiksen 2005).

Reducing the substantial bias of long memory estimators is therefore an important statistical task. The objective of this paper is to present bias reduction methods for ARFIMA(p, d, q) processes with minimum computational burden. As likelihood-based and numerically optimized long memory estimators are often not practical for bias evaluation, we use a simple moment estimator for easiness of bias reduction. This moment estimator is a function of lag-one sample autocorrelation, and the proposed bias reduction methods correct the first-order bias of the lag-one sample autocorrelation. The resulting estimators are less biased for small and moderate samples, and even for large samples, while their standard error inflation remains minimal. In addition, they are relatively simple in practical implementations compared to likelihood function-based or periodogram-based estimators.

The rest of this paper is organized as follows. In Section 2, the ARFIMA(p, d, q) model is briefly reviewed. In Section 3, we present the explicit form of the bias, up to the order n^{-1}, in the lag-one sample autocorrelation. In Section 4, the bias reduction methods for d in ARFIMA(p, d, q) models are proposed. In Section 5, we show through simulation studies that the proposed estimators work well for small and moderate samples. An application to the northern hemisphere data, which is a benchmark in long memory literature, is presented in Section 6. Some concluding remarks are given in Section 7.

2 ARFIMA(p, d, q) Models

A long memory process is characterized by a slow decay in its autocovariances: $\gamma(h) \sim Ch^{-\alpha}$ where C is a positive constant depending on the process, $0 < \alpha < 1$ and h is large. An ARFIMA(p, d, q) process $\{X_t\}_{t=1}^n$, first introduced by Granger & Joyeux (1980) and Hosking (1981), is defined as the stationary solution to the equation

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t, \tag{1}$$

where B is the backshift operator such that $BX_t = X_{t-1}$, $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, $\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$, and $\{\varepsilon_t\}_{t=1}^n$ is a white noise with zero mean and variance σ^2. Employing fractional d-differencing to $\{X_t\}_{t=1}^n$ results in an ARMA(p, q) model. The ARFIMA(p, d, q) process is stationary and invertible if $-0.5 < d < 0.5$ and all the roots of the polynomials $\Phi(\cdot)$ and $\Theta(\cdot)$ lie outside the unit circle. If $0 < d < 0.5$, the process has long range dependency between distant observations and the autocorrelations decay hyperbolically to zero as the lag increases. If $d = 0$, it becomes a Box-Jenkins ARMA(p, q) model. If $-0.5 < d < 0$, it has an intermediate memory and a summable autocorrelation function. Sowell (1992) explicitly derives the autocovariance functions of ARFIMA(p, d, q) models. In this paper, we concentrate on the region $d \in (0, 0.5)$.

A simple but important class of the ARFIMA(p, d, q) process is the fractionally integrated noise, or ARFIMA$(0, d, 0)$, model

$$(1-B)^d X_t = \varepsilon_t.$$

The autocorrelation function $\rho(h)$ of ARFIMA$(0, d, 0)$ model is, for $h = 1, \ldots, n-1$,

$$\rho(h) = \frac{\Gamma(1-d)\Gamma(d+h)}{\Gamma(d)\Gamma(1-d+h)} \tag{2}$$

with the process variance

$$\gamma(0) = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}.$$

When $0 < d < 0.5$, the autocorrelations are positive at all lags and decay monotonically and hyperbolically to zero as h increases.

3 Exact First-order Bias in Sample Autocorrelations

Here we study the first-order bias in the sample autocorrelations for a stationary time series. This exact form of bias holds for both stationary short memory and long memory time series models, and is employed into the bias-reduced estimators of d that will be presented in Section 4.

Consider a stationary time series $\{X_t\}_{t=1}^n$ with autocorrelations $\rho(h)$, for $h = 1, 2, \dots, n-1$. Among many variants of the sample autocorrelations with equivalent asymptotic properties, we consider the following sample autocorrelations defined by, for $k = 1, \dots, n-1$,

$$R_k = \frac{C_k}{C_0} = \frac{\frac{1}{n-k}\sum_{t=1}^{n-k}(X_t - \bar{X}_{[1:n-k]})(X_{t+k} - \bar{X}_{[k+1:n]})}{\frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})^2}, \tag{3}$$

where $\bar{X}_{[1:n-k]} = \sum_{t=1}^{n-k} X_t/(n-k)$, $\bar{X}_{[k+1:n]} = \sum_{t=k+1}^{n} X_t/(n-k)$, and $\bar{X} = \sum_{t=1}^{n} X_t/n$.

Following Marriott & Pope (1954), the sample autocorrelation R_k has, up to the order n^{-1},

$$E(R_k) = \frac{E(C_k)}{E(C_0)}\left[1 - \frac{\mathrm{cov}(C_k, C_0)}{E(C_k)E(C_0)} + \frac{\mathrm{var}(C_0)}{E^2(C_0)}\right]. \tag{4}$$

The right-hand side of the equation (4) can be further expressed as $\rho(k) + Bias(R_k)$ for bias evaluation. With elaborate calculations repeated, including the result of Anderson (1971), we obtained the explicit forms of the means, variances, and covariances on the right-hand side of (4) in terms of autocorrelations only. In the following theorem, we present the exact first-order bias in R_1 for $\rho(1)$ in a closed form for the proposed bias correction in Section 4. Its proof is given in the Appendix.

THEOREM 1. For a zero forth-order cumulant stationary process with mean μ and autocorrelation $\boldsymbol{\rho} = (\rho(1), \ldots, \rho(n-1))'$, the bias of R_1 for $\rho(1)$ is, up to the order n^{-1},

$$Bias(R_1) = -\frac{G(n,\boldsymbol{\rho})\rho(1)}{1-g_0^{(1)}(n,\boldsymbol{\rho})} - \frac{[1-g_0^{(1)}(n,\boldsymbol{\rho}) - G(n,\boldsymbol{\rho})][g_0^{(1)}(n-1,\boldsymbol{\rho}) - g_0^{(1)}(n,\boldsymbol{\rho})\rho(1)]}{[1-g_0^{(1)}(n,\boldsymbol{\rho})]^2}, \tag{5}$$

where

$$\begin{aligned} G(n,\boldsymbol{\rho}) = & \frac{4g_1^{(2)}(n-1,\boldsymbol{\rho}) - v_1(n,\boldsymbol{\rho}) - v_2(n,\boldsymbol{\rho}) + 2\{f(n,\boldsymbol{\rho})\}^2}{\rho(1) - g_0^{(1)}(n-1,\boldsymbol{\rho})} \\ & - \frac{2[g_0^{(2)}(n,\boldsymbol{\rho}) - v_3(n,\boldsymbol{\rho}) + \{g_0^{(1)}(n,\boldsymbol{\rho})\}^2]}{1-g_0^{(1)}(n,\boldsymbol{\rho})}, \end{aligned}$$

$$g_m^{(j)}(n,\boldsymbol{\rho}) = \frac{1}{n}\left[\frac{n}{n+m}\rho^{j-1}(m) + (2-m)\sum_{h=1}^{n-1}\left\{1 - \frac{h+m}{n+m}\right\}\rho(h)\rho^{j-1}(h+m)\right],$$

$$f(n,\boldsymbol{\rho}) = \frac{1}{n}\left[1 + 2\sum_{h=1}^{n-1}\left\{1 - \frac{h-.5}{n-1}\right\}\rho(h)\right],$$

$$\begin{aligned} v_1(n,\boldsymbol{\rho}) = & \frac{2}{n(n-1)}\Bigg[1 - \frac{1}{n} + 4\sum_{h=1}^{n-1}(1 - \frac{h+.5}{n})\rho(h) \\ & + 2\sum_{h=1}^{n-1}\left\{1 - \frac{h+.5}{n} + \left(1 - \frac{2h+.5}{n}\right)_+ - \frac{1_{[n-2h<0]}}{2n}\right\}\rho^2(h) \\ & + 4\sum_{h=1}^{n-2}\sum_{h'=h+1}^{n-1}\left\{1 - \frac{h'}{n} + \left(1 - \frac{h+h'+.5}{n}\right)_+ - \frac{1_{[n-h-h'<0]}}{2n}\right\}\rho(h)\rho(h')\Bigg], \end{aligned}$$

$$\begin{aligned} v_2(n,\boldsymbol{\rho}) = & \frac{2}{n(n-1)}\Bigg[1 - \frac{1}{n-1} + 4\sum_{h=1}^{n-1}(1 - \frac{h}{n-1} + \frac{1_{[h=n-1]}}{2(n-1)})\rho(h) \\ & + 2\sum_{h=1}^{n-1}\left\{1 - \frac{h}{n-1} + \left(1 - \frac{2h}{n-1}\right)_+ + \frac{1_{[2h=n-1]}}{2(n-1)}\right\}\rho^2(h) \\ & + 4\sum_{h=1}^{n-2}\sum_{h'=h+1}^{n-1}\left\{1 - \frac{h'-.5}{n-1} + \left(1 - \frac{h+h'}{n-1}\right)_+ + \frac{1_{[h+h'=n-1]}}{2(n-1)}\right\}\rho(h)\rho(h')\Bigg], \end{aligned}$$

and

$$v_3(n,\boldsymbol{\rho}) = \frac{2}{n^2}\Bigg[1 + 4\sum_{h=1}^{n-1}(1-\frac{h}{n})\rho(h) + 2\sum_{h=1}^{n-1}\left\{1-\frac{h}{n}+\left(1-\frac{2h}{n}\right)_+\right\}\rho^2(h)$$
$$+ 4\sum_{h=1}^{n-2}\sum_{h'=h+1}^{n-1}\left\{1-\frac{h'}{n}+\left(1-\frac{h+h'}{n}\right)_+\right\}\rho(h)\rho(h')\Bigg].$$

Here $x_+ = \max(x, 0)$ and $1_{[A]} = 1$ if A is true, $1_{[A]} = 0$ otherwise.

Note that the bias in (5) is exact up to the order n^{-1} for any covariance stationary process if the forth-order cumulant of the process is zero. Only a slight modification in (5) is needed for non-zero cumulant stationary processes, but this is not pursued here. With the autocorrelation functions $\rho(h)$ provided, the exact bias of the sample autocorrelation can easily be evaluated for stationary processes including ARFIMA(p, d, q) models.

One advantage of the exact first-order bias (5) lies in its outstanding accuracy compared to asymptotic results, especially in long memory models. As a simple example of this, if $\{X_t\}$ follows an ARFIMA$(0, d, 0)$ model, Hosking (1996) provides the following asymptotic bias of R_1: for sufficiently large n,

$$Bias_\infty(R_1) \simeq -\frac{(1-2d)\Gamma(1-d)}{d(1-d)(1+2d)\Gamma(d)}n^{2d-1} \equiv \Delta_\infty(d). \tag{6}$$

We set $Bias_\infty(R_1) = \Delta_\infty(d)$ in (6) for emphasis that the asymptotic bias depends only on d. Similarly, we set $Bias(R_1) = \Delta_n(d)$ in (5). We see that $\Delta_\infty(d) < 0$ and $\Delta_n(d) < 0$ in the long-range dependence case where $d \in (0, 0.5)$, implying R_1 underestimates $\rho(1)$. Also observe that whereas $|\Delta_\infty(d)|$ is not monotone in d, $|\Delta_n(d)|$ monotonically increases with increasing d. These behaviors of the exact first-order bias $\Delta_n(d)$ explain the Newbold & Agiakloglou (1993) finding that R_1 tends to be much smaller than $\rho(1)$ for more strongly correlated processes even when n is large. Figure 1 shows the ratio of the asymptotic bias $\Delta_\infty(d)$ to the exact bias $\Delta_n(d)$ against d for the sample sizes $n = 25$ and $n = 500$. Other bias ratios for different sample sizes show a similar pattern. Note that $|\Delta_\infty(d)|$ is

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at Canadian Journal of Statistics, published by Wiley-Blackwell. Copyright restrictions may apply. doi: 10.1002/cjs.10022

considerably smaller than $|\Delta_n(d)|$ as d gets larger. It is surprising to us that this discrepancy between the exact and asymptotic biases is still noticeably large even for large n in strong correlation cases where d is close to 0.5. In short, the asymptotic bias does not satisfactorily reflect the severe bias in R_1 for $\rho(1)$ in this long memory setting. The other advantage is

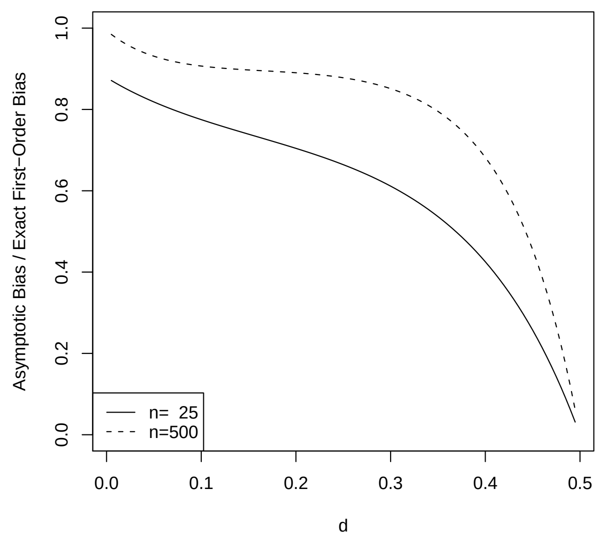

Figure 1: Ratio of Hosking's asymptotic bias $\Delta_\infty(d)$ to the exact first-order bias $\Delta_n(d)$ for $n = 25$ (solid line) and $n = 500$ (dashed line).

in its computational easiness. The exact first-order bias is presented in a closed form with the computational effort for each component of (5) minimized up to $(n-1)(n-2)/2$, not n^4 as in Anderson (1971, p. 452). In addition, the explicit expression does not require intensive computations such as Cholesky decomposition of large sized matrices, which is needed in Newbold & Agiakloglou (1993) for the evaluation of the biases in the sample autocorrelations.

4 Bias Reduction Methods

In this section we present bias reduction methods in long memory parameter estimation of ARFIMA (p, d, q) models. The bias correction methods are presented for the important special case of the ARFIMA(0,d,0) first and then are extended to the general ARFIMA(p, d, q) model.

4.1 Bias correction for ARFIMA$(0, d, 0)$ models

A simple estimation method in short memory time series models is to equate the sample autocorrelations to their corresponding true autocorrelations and solve for unknown parameters. This moment estimation method seems to be accurate because the bias in sample autocorrelations is negligible for large samples. In fact, this method of moments estimation is often asymptotically efficient in Box-Jenkins models, in that the limiting distribution of the moments estimators is the same as that of the ML estimators.

Suppose that $\{X_t\}$ is an ARFIMA(0,d,0) process with $d \in (0, 0.5)$. From (2),

$$\rho(1) = \frac{\Gamma(1-d)\Gamma(1+d)}{\Gamma(d)\Gamma(2-d)} = \frac{d}{1-d}.$$

Then, a simple moment estimator of d is

$$\hat{d}_0 = \frac{R_1}{1+R_1}, \tag{7}$$

where R_1 is the lag-one sample autocorrelation defined in (3). This simple moment estimator is used in Kettani & Gubner (2003) to evaluate a confidence interval of d. However, although the bias in the sample autocorrelations would be negligible for large n in short memory time series, naive use of the corresponding lag-one sample autocorrelation R_1 for $\rho(1)$ causes severe bias in long memory processes. Hosking (1996) showed that the sample autocorrelations of a long memory process are substantially negative-biased for the corresponding true autocorrelations, even for large samples. Newbold & Agiakloglou (1993) also empirically supported such non-negligible biases through an elaborate simulation study. On the other hand, the initial bias of $\hat{d}_0$ in Section 5 turns out to be up to n^{-1}. This kind of bias may be very crucial in estimating d over a relatively small range (0, 0.5) with small sample size. Thus bias correction for $\hat{d}_0$ is needed in small samples or even in large samples.

To reduce the bias of $\hat{d}_0$, we propose a bias correction of R_1 by using the exact first-order bias in (5). As the exact bias is a function of the unknown d only, an initial estimate of d is required. Although we do not favor any particular estimator in this initial stage, we use $\hat{d}_0$ in this paper. The resulting bias-corrected estimator is

$$\hat{d}_{\mathrm{BC}} = \frac{R_1 - \Delta_n(\hat{d}_0)}{1 + R_1 - \Delta_n(\hat{d}_0)}. \tag{8}$$

With the exact first-order bias correction applied to R_1, $\hat{d}_{\mathrm{BC}}$ would be expected to be closer to the true value of d than $\hat{d}_0$. This bias reduction can be explained by the relation between $\hat{d}_0$ and R_1 in (7). The relation is not linear but monotone because $\hat{d}_0 = R_1/(1 + R_1)$. Moreover, it is approximately linear. In the simple moment estimator $\hat{d}_0$, a linearization of $R_1/(1 + R_1)$ can be obtained using the first-degree Taylor polynomial at $\rho(1)$, which can be expressed as $T_1(R_1) = [\rho(1)^2 + R_1]/[1 + \rho(1)]^2$ such that $\hat{d}_0 = T_1(R_1) + E_1(R_1)$ where $E_1(R_1)$ is the remainder from the first-degree Taylor polynomial. Thus reducing the bias of R_1 is approximately equivalent to reducing the bias of $\hat{d}_0$.

While the bias of $\hat{d}_{\mathrm{BC}}$ becomes smaller than that of $\hat{d}_0$ through the proposed correction, $\hat{d}_{\mathrm{BC}}$ can still be biased because $\hat{d}_0$ often underestimates d and in turn, $\Delta_n(\hat{d}_0)$ is still smaller than its true counterpart $\Delta_n(d)$. To obtain a more accurate estimate of d, we propose a further refinement of $\hat{d}_{\mathrm{BC}}$ through a recursive iteration as follows.

1. Set $\hat{d}^{(0)}_{\mathrm{IBC}} = \hat{d}_0$, $\hat{d}^{(1)}_{\mathrm{IBC}} = \hat{d}_{\mathrm{BC}}$, and a desirable tolerance T.

2. In k-th iteration,
$$\hat{d}^{(k)}_{\mathrm{IBC}} = \frac{R_1 - \Delta_n(\hat{d}^{(k-1)}_{\mathrm{IBC}})}{1 + R_1 - \Delta_n(\hat{d}^{(k-1)}_{\mathrm{IBC}})}.$$

3. Repeat the step 2 until $|\hat{d}^{(k)}_{\mathrm{IBC}} - \hat{d}^{(k-1)}_{\mathrm{IBC}}| < T$.

The idea of the recursive iteration can be justified as follows: on average, since the initial estimate $\hat{d}_0$ tends to severely underestimate d, $\Delta_n(\hat{d}_0)$ still appears to be away from the true bias $\Delta_n(d)$ (due to the monotonic property of the exact bias in Section 3). Updating $\Delta_n(\hat{d}_0)$ by substituting $\hat{d}_0$ with $\hat{d}_{\text{BC}}$ can be considered a more accurate bias assessment. Using $\Delta_n(\hat{d}_{\text{BC}})$ for $\Delta_n(\hat{d}_0)$ in (8) results in an updated estimate of d. With its bias corrected more accurately, the updated estimate would be expected to be closer to the true value of d than $\hat{d}_{\text{BC}}$. This procedure can be iterated until no meaningful gain in accuracy is achieved.

Note that implementing the proposed estimators, $\hat{d}_{\text{BC}}$ and $\hat{d}_{\text{IBC}}$, in practice is much simpler than maximizing complicated likelihood functions on time or frequency domains, which has been typically adopted in the literature of long memory processes. In addition, it is worth pointing out that evaluation of the exact bias in the ML estimators is most often not feasible in long memory time series models.

On the other hand, for large n one might be attracted to the use of Hosking's (1996) asymptotic bias $\Delta_\infty(d)$ instead of $\Delta_n(d)$ in (8). The resulting asymptotic bias-adjusted estimator can be expressed as

$$\hat{d}_{\text{ASY}} = \frac{R_1 - \Delta_\infty(\hat{d}_0)}{1 + R_1 - \Delta_\infty(\hat{d}_0)}.$$

The results from Figure 1, however, imply that this asymptotic bias correction does not help satisfactorily reduce the bias of $\hat{d}_{\text{ASY}}$, especially when d is large. Such large bias inherent in $\hat{d}_{\text{ASY}}$ is also numerically confirmed through the simulation study in Section 5.1.

4.2 Bias correction for ARFIMA(p, d, q) models

In the presence of short memory components, one can equate the sample autocorrelations with their biases corrected to their corresponding population autocorrelations, which are the functions of the unknown autoregressive (AR), moving average (MA), and long memory parameters, and solve the resulting equations for the unknown parameters. However, it is very tedious and laborious to calculate the biases in the sample autocorrelations at various lags. Moreover Smith, Taylor & Yadav (1997) found that simultaneous estimators, including simultaneous ML estimators, for the parameters in ARFIMA(p, d, q) models can cause severe biases.

On the other hand, Hosking (1981) proposed a two-stage recursive algorithm, with which one can recursively estimate the AR, MA, and long memory parameters in the ARFIMA(p, d, q) models. In the first stage, the long memory parameter d is estimated and in the second stage, Box-Jenkins model procedures are used to estimate the short memory parameters in ARMA(p, q). The two-stage recursive estimation procedure provides an easy implementation in practice and gives reliable results once one has a plausible estimation algorithm for d in ARFIMA$(0, d, 0)$. Beveridge & Oickle (1993) and Reisen, Abraham & Lopes (2001) showed that the two-stage recursive algorithm performs well under various scenarios.

With the proposed bias-corrected estimators $\hat{d}_{\mathrm{BC}}$ and $\hat{d}_{\mathrm{IBC}}$, we here rewrite Hosking's (1981) two-stage recursive estimation procedure for the ARFIMA(p, d, q) model. Referring to the model (1):

1. Estimate d in the model $(1 - B)^d X_t = \varepsilon_t$ using $\hat{d}_{\mathrm{BC}}$ or $\hat{d}_{\mathrm{IBC}}$.

2. Obtain $Y_t = (1 - B)^{\hat{d}_{\mathrm{BC}}} X_t$ or $Y_t = (1 - B)^{\hat{d}_{\mathrm{IBC}}} X_t$.

3. Identify and estimate the ARMA parameters ϕ's and θ's in the model $\Phi(B)Y_t = \Theta(B)\varepsilon_t$.

4. Define $U_t = \{\hat{\Theta}(B)\}^{-1}\hat{\Phi}(B)X_t$.

5. Estimate d in the model $(1 - B)^d U_t = \varepsilon_t$ using $\hat{d}_{\mathrm{BC}}$ or $\hat{d}_{\mathrm{IBC}}$.

6. Repeat the steps 2 to 5 until $\hat{d}_{\mathrm{BC}}$ (or $\hat{d}_{\mathrm{IBC}}$), $\hat{\phi}$s, and $\hat{\theta}$s converge.

We conducted a simulation study in Section 5.2 to demonstrate that this two-stage procedure performs satisfactorily with the proposed estimators $\hat{d}_{\mathrm{BC}}$ and $\hat{d}_{\mathrm{IBC}}$ for ARFIMA(p, d, q) models. In fact, the procedure works very well.

5 Simulation Studies

The accuracy of the bias-corrected estimators $\hat{d}_{\mathrm{BC}}$ and $\hat{d}_{\mathrm{IBC}}$ is examined in several ARFIMA models. In Section 5.1, the simulation results for Gaussian ARFIMA$(0, d, 0)$ and non-Gaussian ARFIMA$(0, d, 0)$ models are summarized in terms of biases and root mean squared errors. The simulation studies for Gaussian ARFIMA$(1, d, 0)$ and ARFIMA$(1, d, 1)$ models are carried out in Section 5.2.

5.1 ARFIMA$(0, d, 0)$ models

For the simulation of Gaussian ARFIMA$(0, d, 0)$ models, the biases and root mean squared errors were calculated from ten thousand simulations of Gaussian ARFIMA $(0, d, 0)$ processes with the sample sizes $n = 50, 100, 200, 500, 1000, 5000$ and the long memory parameters $d = 0.05$, 0.15, 0.25, 0.35, 0.45. A unit innovation variance was chosen, i.e., $\sigma^2 = 1$.

For each combination of d and n, we computed the exact first-order bias-corrected estimators, $\hat{d}_{\mathrm{BC}}$ and $\hat{d}_{\mathrm{IBC}}$, and the existing estimators: the uncorrected estimator $\hat{d}_0$, the asymptotic bias-corrected estimator $\hat{d}_{\mathrm{ASY}}$, the GPH estimator $\hat{d}_{\mathrm{GPH}}$ with a trimming parameter $\sqrt{n}$, the fractionally integrated exponential model (FEXP) estimator $\hat{d}_{\mathrm{FEX}}$, and the Whittle estimator $\hat{d}_{\mathrm{WT}}$. The Whittle's estimator (Whittle 1951) of d uses a simple approximation of the variance-covariance matrix on frequency domain and is widely used in long memory literature. The GPH estimator (Geweke & Porter-Hudak 1983) is a least squares estimate of d at low frequency in the spectral density function of ARFIMA models. The FEXP estimator, introduced by Moulines & Soulier (1999), is a global semi-parametric estimator and has the advantage that it is adaptive (Bardet, Lang, Oppenheim, Philippe & Taqqu 2002) and does not require the presence of a trimming parameter in the GPH estimator.

Table 1 reports the biases and root mean squared errors (in parentheses) of the exact first-order bias-corrected estimators and the existing estimators. Most of these biases, except those of $\hat{d}_{\mathrm{GPH}}$, were negative and some were very large. The simulation results confirm that the proposed first-order bias-corrected estimators $\hat{d}_{\mathrm{BC}}$ and $\hat{d}_{\mathrm{IBC}}$ substantially reduce biases. Especially the iterated estimator $\hat{d}_{\mathrm{IBC}}$ produced almost negligible biases, even in the settings of small/moderate n and large d where $\hat{d}_0$, $\hat{d}_{\mathrm{ASY}}$, $\hat{d}_{\mathrm{FEX}}$ and $\hat{d}_{\mathrm{WT}}$ had relatively large biases. Moreover, the root mean squared errors of the proposed bias-corrected estimators were quite comparable to those of the Whittle and FEXP estimators. It was also observed that the iterated bias correction method did not seriously inflate the error margin of $\hat{d}_{\mathrm{IBC}}$. On the other hand, $\hat{d}_{\mathrm{GPH}}$ tended to be less biased especially for small d, but this semi-parametric estimator should not be trustworthy due to its very large variances. $\hat{d}_{\mathrm{ASY}}$ appeared to work fine only when d was small and n was large; however, it was not as satisfactory as $\hat{d}_{\mathrm{BC}}$, $\hat{d}_{\mathrm{IBC}}$ and $\hat{d}_{\mathrm{FEX}}$.

Overall, the proposed bias correction methods using the first-order bias of lag-one sample autocorrelation performed well and were very effective in reducing biases. In particular, the recursive bias-corrected estimator $\hat{d}_{\mathrm{IBC}}$ was least biased over all ranges of d and n (with a few exceptions in which $\hat{d}_{\mathrm{FEX}}$ had extremely small biases for large n, and the unreliable $\hat{d}_{\mathrm{GPH}}$ had negligibly smaller biases). This is attributable to the iterative bias update procedure in Section 4.1 through the exact first-order bias expression in (5). In addition, $\hat{d}_{\mathrm{BC}}$ and $\hat{d}_{\mathrm{IBC}}$ had their standard errors competitive to $\hat{d}_{\mathrm{WT}}$ and $\hat{d}_{\mathrm{FEX}}$.

For the simulation of non-Gaussian ARFIMA $(0, d, 0)$ processes, we generated ten thousand copies of ARFIMA$(0, d, 0)$ models with exponentially distributed white noises. Then biases and root mean squared errors were calculated with the sample sizes $n = 50$, 100, 200, 500 and the long memory parameters $d = 0.05$, 0.15, 0.25, 0.35, 0.45.

Table 2 summarizes the biases and mean squared errors of the proposed estimators in non-Gaussian ARFIMA$(0, d, 0)$ model setting. The patterns of the biases and mean squared errors of $\hat{d}_{\mathrm{BC}}$ and $\hat{d}_{\mathrm{IBC}}$ were similar to those in the Gaussian ARFIMA$(0, d, 0)$ simulation. The iterative $\hat{d}_{\mathrm{IBC}}$ were least biased, with some exceptions in which the GPH estimator had very small biases for d close to 0 and had variations close to $\hat{d}_{\mathrm{FEX}}$ and $\hat{d}_{\mathrm{WT}}$.

Table 1: Biases and root mean squared errors (in parenthesis) in Gaussian ARFIMA$(0, d, 0)$ models.

d	n	$\hat{d}_{\mathrm{BC}}$	$\hat{d}_{\mathrm{IBC}}$	$\hat{d}_0$	$\hat{d}_{\mathrm{ASY}}$	$\hat{d}_{\mathrm{GPH}}$	$\hat{d}_{\mathrm{FEX}}$	$\hat{d}_{\mathrm{WT}}$
0.05	50	-0.027(0.165)	-0.018(0.172)	-0.050(0.156)	-0.028(0.159)	0.006(0.382)	-0.022(0.151)	-0.088(0.174)
	100	-0.009(0.107)	-0.007(0.109)	-0.023(0.101)	-0.010(0.105)	0.000(0.289)	-0.010(0.095)	-0.043(0.104)
	200	-0.006(0.070)	-0.005(0.071)	-0.013(0.068)	-0.006(0.070)	-0.001(0.232)	-0.007(0.062)	-0.024(0.066)
	500	-0.002(0.043)	-0.002(0.043)	-0.005(0.042)	-0.002(0.043)	-0.002(0.172)	-0.003(0.037)	-0.010(0.038)
	1000	-0.001(0.030)	-0.001(0.030)	-0.003(0.030)	-0.001(0.030)	0.001(0.136)	-0.001(0.026)	-0.005(0.026)
	5000	-0.001(0.013)	-0.001(0.013)	-0.001(0.013)	-0.001(0.013)	0.003(0.086)	-0.000(0.011)	-0.001(0.011)
0.15	50	-0.030(0.150)	-0.014(0.161)	-0.064(0.145)	-0.035(0.145)	0.006(0.383)	-0.022(0.154)	-0.089(0.176)
	100	-0.014(0.098)	-0.008(0.102)	-0.036(0.094)	-0.017(0.096)	-0.003(0.296)	-0.011(0.097)	-0.044(0.106)
	200	-0.006(0.064)	-0.004(0.065)	-0.020(0.062)	-0.008(0.063)	-0.003(0.233)	-0.006(0.063)	-0.023(0.067)
	500	-0.002(0.039)	-0.002(0.039)	-0.010(0.038)	-0.003(0.039)	0.002(0.171)	-0.002(0.038)	-0.010(0.039)
	1000	-0.001(0.027)	-0.001(0.027)	-0.006(0.026)	-0.001(0.027)	-0.001(0.138)	-0.001(0.026)	-0.005(0.026)
	5000	-0.001(0.012)	-0.001(0.012)	-0.002(0.012)	-0.001(0.012)	0.002(0.086)	-0.000(0.011)	-0.001(0.011)
0.25	50	-0.037(0.135)	-0.006(0.153)	-0.083(0.140)	-0.047(0.132)	0.006(0.388)	-0.017(0.154)	-0.084(0.174)
	100	-0.017(0.087)	-0.002(0.096)	-0.050(0.090)	-0.023(0.085)	0.007(0.295)	-0.007(0.097)	-0.040(0.105)
	200	-0.011(0.059)	-0.003(0.062)	-0.034(0.062)	-0.014(0.058)	0.007(0.230)	-0.005(0.063)	-0.022(0.066)
	500	-0.004(0.036)	-0.001(0.037)	-0.019(0.038)	-0.006(0.036)	0.003(0.171)	-0.001(0.037)	-0.009(0.038)
	1000	-0.002(0.026)	-0.000(0.026)	-0.013(0.026)	-0.003(0.025)	0.004(0.137)	-0.000(0.026)	-0.004(0.026)
	5000	-0.000(0.012)	-0.000(0.012)	-0.005(0.012)	-0.001(0.012)	0.009(0.086)	-0.000(0.011)	-0.001(0.011)
0.35	50	-0.055(0.127)	-0.002(0.148)	-0.113(0.150)	-0.071(0.128)	0.015(0.383)	-0.015(0.154)	-0.084(0.171)
	100	-0.033(0.083)	-0.001(0.095)	-0.078(0.101)	-0.042(0.083)	0.006(0.296)	-0.006(0.097)	-0.039(0.104)
	200	-0.020(0.055)	0.000(0.063)	-0.055(0.071)	-0.027(0.056)	0.010(0.233)	-0.000(0.063)	-0.018(0.065)
	500	-0.011(0.036)	0.001(0.039)	-0.037(0.047)	-0.015(0.036)	0.011(0.173)	0.001(0.037)	-0.007(0.038)
	1000	-0.008(0.026)	-0.001(0.028)	-0.029(0.036)	-0.011(0.027)	0.004(0.138)	0.000(0.026)	-0.004(0.026)
	5000	-0.004(0.014)	-0.001(0.015)	-0.017(0.021)	-0.005(0.015)	0.002(0.093)	0.000(0.011)	-0.001(0.011)
0.45	50	-0.084(0.173)	-0.002(0.140)	-0.152(0.173)	-0.106(0.138)	0.017(0.389)	-0.009(0.154)	-0.092(0.162)
	100	-0.058(0.087)	0.002(0.094)	-0.115(0.127)	-0.076(0.095)	0.016(0.295)	0.000(0.097)	-0.042(0.093)
	200	-0.044(0.063)	0.002(0.066)	-0.092(0.099)	-0.057(0.070)	0.016(0.234)	0.002(0.063)	-0.020(0.060)
	500	-0.031(0.044)	0.001(0.044)	-0.071(0.075)	-0.042(0.050)	0.015(0.171)	0.003(0.038)	-0.006(0.036)
	1000	-0.025(0.035)	0.001(0.033)	-0.060(0.063)	-0.034(0.040)	0.013(0.138)	0.002(0.026)	-0.002(0.026)
	5000	-0.016(0.021)	0.001(0.019)	-0.042(0.044)	-0.022(0.025)	0.010(0.086)	0.002(0.011)	0.001(0.011)

Table 2: Biases and root mean squared errors (in parenthesis) for non-Gaussian ARFIMA$(0, d, 0)$ models with exponential white noise.

d	n	$\hat{d}_{\mathrm{BC}}$	$\hat{d}_{\mathrm{IBC}}$	$\hat{d}_{\mathrm{GPH}}$	$\hat{d}_{\mathrm{FEX}}$	$\hat{d}_{\mathrm{WT}}$
0.05	50	-0.021(0.151)	-0.013(0.159)	-0.004(0.375)	-0.019(0.147)	-0.085(0.169)
	100	-0.011(0.103)	-0.008(0.105)	0.005(0.292)	-0.011(0.094)	-0.043(0.103)
	200	-0.005(0.069)	-0.005(0.070)	0.001(0.228)	-0.006(0.061)	-0.023(0.065)
	500	-0.001(0.043)	-0.001(0.043)	-0.001(0.171)	-0.002(0.037)	-0.009(0.038)
0.15	50	-0.026(0.137)	-0.010(0.150)	-0.000(0.382)	-0.018(0.148)	-0.085(0.170)
	100	-0.012(0.091)	-0.006(0.096)	0.002(0.289)	-0.010(0.094)	-0.043(0.103)
	200	-0.006(0.062)	-0.004(0.064)	0.007(0.227)	-0.005(0.061)	-0.022(0.065)
	500	-0.002(0.038)	-0.002(0.038)	0.002(0.169)	-0.002(0.037)	-0.010(0.038)
0.25	50	-0.032(0.123)	-0.001(0.142)	0.003(0.383)	-0.012(0.147)	-0.079(0.166)
	100	-0.018(0.082)	-0.003(0.091)	0.009(0.292)	-0.008(0.094)	-0.040(0.102)
	200	-0.008(0.057)	-0.001(0.061)	0.005(0.227)	-0.003(0.062)	-0.020(0.065)
	500	-0.005(0.036)	-0.001(0.037)	0.002(0.172)	-0.002(0.037)	-0.009(0.038)
0.35	50	-0.054(0.120)	-0.001(0.141)	0.003(0.384)	-0.014(0.150)	-0.084(0.166)
	100	-0.031(0.079)	0.001(0.092)	0.009(0.292)	-0.004(0.095)	-0.037(0.101)
	200	-0.020(0.055)	-0.000(0.062)	0.011(0.231)	-0.001(0.062)	-0.018(0.065)
	500	-0.011(0.036)	0.000(0.039)	0.010(0.170)	0.001(0.038)	-0.006(0.038)
0.45	50	-0.079(0.121)	0.004(0.134)	0.023(0.384)	-0.005(0.149)	-0.087(0.153)
	100	-0.058(0.085)	0.003(0.092)	0.016(0.295)	-0.000(0.096)	-0.043(0.092)
	200	-0.044(0.063)	0.001(0.065)	0.015(0.231)	0.001(0.063)	-0.021(0.059)
	500	-0.032(0.044)	0.001(0.044)	0.013(0.172)	0.002(0.038)	-0.006(0.036)

5.2 ARFIMA(p, d, q) models

We examined the performance of the exact bias correction in Gaussian ARFIMA $(1, d, 0)$ and ARFIMA $(1, d, 1)$ models. Five hundred simulations were independently generated from each combination of $d = 0.10,\ 0.25,\ 0.40$ and $n = 100,\ 200,\ 500$ for each model. For the ARFIMA$(1, d, 0)$ models, the autoregressive parameter $\phi = 0.4$ was used. The autoregressive and moving average parameters were set to $\phi = 0.4$ and $\theta = 0.4$ for the ARFIMA $(1, d, 1)$ models.

Table 3 summarizes the biases and root mean squared errors (in parentheses) of $\hat{d}_{\mathrm{BC}}$ and $\hat{d}_{\mathrm{IBC}}$ implemented by the two-stage algorithm for ARFIMA $(1, d, 0)$ models. To proceed with the two-stage method in Section 4.2, we need an estimate of ϕ. Although we have no specific preference in the estimator, we here use the bias-corrected moments estimator $\hat{\phi}_{\mathrm{BC}} = [(n-1)R_1 + 1]/(n-4)$ as in Patterson (2007). For a comparison, $\hat{d}_{\mathrm{GPH}}$, $\hat{d}_{\mathrm{FEX}}$

Table 3: Biases and root mean squared errors (in parenthesis) for Gaussian ARFIMA$(1, d, 0)$ models. The autoregressive parameter ϕ is set to 0.4.

d	n	$\hat{d}_{\mathrm{BC}}$	$\hat{d}_{\mathrm{IBC}}$	$\hat{d}_{\mathrm{GPH}}$	$\hat{d}_{\mathrm{FEX}}$	$\hat{d}_{\mathrm{WT}}$
0.10	100	-0.051(0.273)	0.034(0.254)	0.073(0.299)	0.205(0.283)	-0.333(0.407)
	200	-0.027(0.202)	0.005(0.197)	0.032(0.237)	0.145(0.196)	-0.223(0.314)
	500	-0.005(0.142)	0.005(0.135)	0.005(0.187)	0.077(0.110)	-0.118(0.209)
0.25	100	-0.120(0.247)	0.009(0.218)	0.072(0.290)	0.139(0.226)	-0.374(0.424)
	200	-0.092(0.220)	-0.032(0.206)	0.037(0.240)	0.080(0.144)	-0.283(0.360)
	500	-0.040(0.127)	-0.014(0.120)	0.030(0.170)	0.076(0.099)	-0.137(0.241)
0.40	100	-0.219(0.307)	-0.029(0.223)	0.072(0.286)	0.098(0.202)	-0.441(0.475)
	200	-0.156(0.230)	-0.040(0.190)	0.049(0.225)	0.079(0.151)	-0.402(0.450)
	500	-0.086(0.140)	-0.027(0.126)	0.019(0.171)	0.070(0.098)	-0.301(0.384)

and $\hat{d}_{\mathrm{WT}}$ were also obtained. Note that the semi-parametric estimator $\hat{d}_{\mathrm{GPH}}$, the global semiparametric estimator $\hat{d}_{\mathrm{FEX}}$, and the simultaneous approximate ML estimator $\hat{d}_{\mathrm{WT}}$ were computed only once for each simulation. The result showed that the proposed bias-corrected estimators overall achieved noticeable reduction in bias with $\hat{d}_{\mathrm{IBC}}$ resulting in the smallest biases and root mean square errors, even if the sample size was small. This desirable accuracy was again obtained by using the exact first-order bias (5).

Table 4 reports the biases and mean squared errors of the proposed estimators in an ARFIMA $(1, d, 1)$ setting. To implement the two-stage procedure with the first-order bias-corrected estimators, we estimated the ARMA parameters ϕ and θ using conditional sum of squares method. The accuracy of the proposed estimators appeared to be somewhat deteriorated but be still attained when compared to that of the other estimators. The iterative estimator $\hat{d}_{\mathrm{IBC}}$ performed satisfactorily for large value of d.

The performance of the Whittle estimator was not good for small/moderate n when the AR and MA components exist in ARFIMA model. This was also indicated in Reisen, Abraham and Lopes (2001) when the AR and MA components are involved in ARFIMA model.

Table 4: Biases and root mean squared errors (in parenthesis) for Gaussian ARFIMA$(1, d, 1)$ models. Both autoregressive parameter ϕ and moving-average parameter θ are set to 0.4.

d	n	$\hat{d}_{\text{BC}}$	$\hat{d}_{\text{IBC}}$	$\hat{d}_{\text{GPH}}$	$\hat{d}_{\text{FEX}}$	$\hat{d}_{\text{WT}}$
0.10	100	0.117(0.142)	0.171(0.197)	0.089(0.285)	0.278(0.475)	-0.422(0.518)
	200	0.101(0.119)	0.124(0.143)	0.045(0.231)	0.280(0.445)	-0.267(0.395)
	500	0.084(0.096)	0.091(0.103)	0.020(0.179)	0.190(0.359)	-0.124(0.243)
0.25	100	0.021(0.101)	0.115(0.169)	0.096(0.290)	0.124(0.302)	-0.417(0.512)
	200	0.017(0.082)	0.066(0.119)	0.047(0.227)	0.034(0.205)	-0.279(0.399)
	500	0.021(0.059)	0.041(0.078)	0.015(0.165)	-0.028(0.106)	-0.121(0.237)
0.40	100	-0.092(0.147)	0.044(0.149)	0.076(0.258)	0.011(0.187)	-0.502(0.562)
	200	-0.071(0.110)	0.021(0.111)	0.045(0.221)	-0.038(0.125)	-0.342(0.442)
	500	-0.056(0.085)	-0.007(0.082)	0.008(0.166)	-0.046(0.083)	-0.152(0.262)

6 An Application

The northern hemisphere monthly temperature data (the left plot in Figure 2) is a benchmark data in long memory literature and has been widely used for the study of global warming. We used the dataset collected by the Climate Research Unit of the University of East Anglia in England during 1854-1989. Beran (1994) fitted a linear trend model $y_t = \beta_0 + \beta_1 t + \epsilon_t$ to the data and applied the ARFIMA$(0, d, 0)$ model to the residuals from the ordinary least squares fit. The resulting Whittle estimate of d was 0.370. Also Beran & Feng (2002) obtained $\hat{d} = 0.33$ by SEMIFAR model, and Craigmile, Guttorp & Percival (2005) reported $\hat{d} = 0.361$ using an approximate maximum likelihood estimation method on wavelet domain.

On the other hand, one can see that the variability of the series at the beginning is larger than for the rest of the observations in the plot. In this paper, we applied our first-order corrected methods and the existing methods to the northern hemisphere data with the first 300 observations data excluded (the right plot in Figure 2). We detrended the resulting data ($n = 1332$) by the OLS estimates, $\hat{\beta}_0 = -0.4031$ and $\hat{\beta}_1 = 0.0004$. To identify the orders of ARMA polynomials in step 3 of the two stage procedure, Section 4.2, we first applied the Box-Ljung white noise test to the residuals by $\hat{d}_{\text{BC}}$ and $\hat{d}_{\text{IBC}}$. Table 5 summarizes the test results at the first three lags. The results confirm that the residuals are uncorrelated,

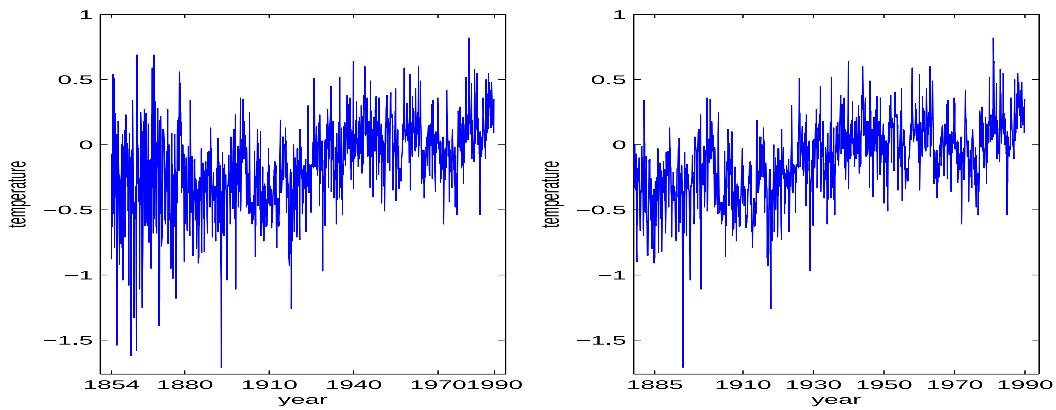

Figure 2: *Left*: The northern hemisphere data (1854-1989) with $n = 1632$. *Right*: The northern hemisphere data (1879-1989) with the first 300 data excluded ($n = 1332$).

which indicates an ARFIMA$(0, d, 0)$ model. Thus we fitted the data to an ARFIMA$(0, d, 0)$ model.

Table 5: The Box-Ljung white noise test with $\hat{d}_{\text{BC}}$ and $\hat{d}_{\text{IBC}}$

Residuals by	Lag	Box-Ljung statistic	d.f.	p-value
$\hat{d}_{\text{BC}}$	1	3.161	1	0.075
	2	3.950	2	0.139
	3	4.083	3	0.253
$\hat{d}_{\text{IBC}}$	1	1.995	1	0.158
	2	2.493	2	0.288
	3	2.725	3	0.436

Table 6 shows the estimation results of the long memory parameter in the northern hemisphere data. The bias corrected estimates $\hat{d}_{\text{BC}}$ and $\hat{d}_{\text{IBC}}$ were 0.3769 and 0.3869, respectively.

Table 6: Estimates of the long memory parameter in the northern hemisphere data (1879-1989).

$\hat{d}_{\text{BC}}$	$\hat{d}_{\text{IBC}}$	$\hat{d}_{\text{ASY}}$	$\hat{d}_{\text{GPH}}$	$\hat{d}_{\text{WT}}$	$\hat{d}_{\text{FEX}}$
0.377	0.387	0.373	0.282	0.402	0.405

7 Concluding Remarks

We proposed several bias correction methods of the moment long memory estimator in ARFIMA (p, d, q) models. For an accurate bias correction, we presented the explicit form of the bias, up to the order n^{-1}, in the lag-one sample autocorrelation for the corresponding true autocorrelation. Although this first-order bias is exact for any zero forth-order cumulant stationary process, the greatest advantage is gained in long memory time series settings. We showed how the proposed estimators work well in terms of bias and root mean squared error, via simulation studies, compared to the existing estimators. We confirm from the empirical results that overall the iterated bias-reduced estimator has the smallest biases with its root mean squared errors comparable to those of the Whittle and FEXP estimators. In addition, the proposed bias-reduced estimators are of simple forms and so, take an advantage in their computational easiness compared to the existing ML estimators. Even though our empirical results are promising, we note that there is a possibility that the tolerance threshold T might be reached without getting very close to the true value of d over a relatively small range $(0, 0.5)$.

The asymptotic properties of the proposed bias-reduced estimators can be further studied. In particular, for $0 < d \leq 0.25$ where the sample autocorrelations are asymptotically normally distributed (Hosking 1996), the limiting distribution of $\hat{d}_{\text{BC}}$ is normal, which can be derived by the delta method. For $0.25 < d < 0.50$ where the sample autocorrelations have a Rosenblatt-type distribution, the limiting distribution is uncertain, and further theoretical investigation is needed for the distribution of a function of a Rosenblatt quantity.

8 ACKNOWLEDGEMENTS

The authors would like to thank an associate editor and two referees for very constructive comments that greatly improved the presentation of the paper.

Appendix

Proof of Theorem 1. Let $\gamma(h)$ be the autocovariance function of $\{X_t\}$. We algebraically evaluate the expected value of R_k in (4) when $k = 1$. The first moments of C_1 and C_0 are

$$E(C_0) = \gamma(0) - \frac{1}{n}\left[\gamma(0) + 2\sum_{h=1}^{n-1}\left(1 - \frac{h}{n}\right)\gamma(h)\right] = \gamma(0)\{1 - g_0^{(1)}(n, \boldsymbol{\rho})\} \tag{9}$$

and

$$\begin{aligned} E(C_1) &= \gamma(1) - \frac{1}{n-1}\left[\left(1 - \frac{1}{n-1}\right)\gamma(0) + 2\sum_{h=1}^{n-2}\left(1 - \frac{h}{n-1}\right)\gamma(h) + \frac{\gamma(n-1)}{n-1}\right] \\ &= \gamma(0)\left[\rho(1) - g_0^{(1)}(n-1, \boldsymbol{\rho})\right] + O(n^{-(2+\alpha)}), \end{aligned} \tag{10}$$

respectively. For a simple expression of $\mathrm{var}(C_0)$, we further reduce the expressions in Anderson (1971, p.452–453) to

$$\sum_{h=1}^{n}\sum_{h'=1}^{n}\gamma^2(h - h') = n\gamma^2(0) + 2\sum_{h=1}^{n-1}(n-h)\gamma^2(h),$$

$$\begin{aligned}
&\sum_{h=1}^{n} \sum_{h',k'=1}^{n} \gamma(h-h')\gamma(h-k') = \sum_{h,k=1}^{n} \sum_{h'=1}^{n} \gamma(h-h')\gamma(k-h') \\
&= n\gamma^2(0) + 4\sum_{h=1}^{n-1}(n-h)\gamma(0)\gamma(h) + 2\sum_{h=1}^{n-1}[n-h+(n-2h)_+]\gamma^2(h) \\
&\quad + 2\sum_{h=1}^{n-2} \sum_{h'=h+1}^{n-1} [2(n-h') + 2(n-h-h')_+]\gamma(h)\gamma(h')
\end{aligned}$$

and

$$\sum_{h,k=1}^{n} \sum_{h',k'=1}^{n} [\gamma(h-h')\gamma(k-k')+\gamma(h-k')\gamma(k-h')] = 2\left[n\gamma(0) + 2\sum_{h=1}^{n-1}(n-h)\gamma(h)\right]^2 .$$

Using these explicit expressions, we have

$$\begin{aligned}
\operatorname{var}(C_0) &= \frac{2}{n}\left[\gamma^2(0) + 2\sum_{h=1}^{n-1}\left(1-\frac{h}{n}\right)\gamma^2(h)\right] \\
&\quad - \frac{4}{n^2}\Bigg\{\gamma^2(0) + 4\sum_{h=1}^{n-1}\left(1-\frac{h}{n}\right)\gamma(0)\gamma(h) + 2\sum_{h=1}^{n-1}\left[1-\frac{h}{n}+\left(1-\frac{2h}{n}\right)_+\right]\gamma^2(h) \\
&\quad + 4\sum_{h=1}^{n-2}\sum_{h'=h+1}^{n-1}\left[1-\frac{h'}{n}+\left(1-\frac{h+h'}{n}\right)_+\right]\gamma(h)\gamma(h')\Bigg\} \\
&\quad + \frac{2}{n^2}\left[\gamma(0) + 2\sum_{h=1}^{n-1}\left(1-\frac{h}{n}\right)\gamma(h)\right]^2 \\
&= 2\gamma^2(0)\left[g_0^{(2)}(n,\boldsymbol{\rho}) - v_3(n,\boldsymbol{\rho}) + \{g_0^{(1)}(n,\boldsymbol{\rho})\}^2\right]. \qquad (11)
\end{aligned}$$

Similarly, for the faster evaluation of $\operatorname{cov}(C_1, C_0)$, the results by Anderson (1971, p.452–453) are reexpressed as follows:

$$\sum_{h=1}^{n-1}\sum_{h'=1}^{n}\gamma(h-h')\gamma(h+1-h') = 2\left[(n-1)\gamma(0)\gamma(1) + \sum_{h=1}^{n-2}(n-1-h)\gamma(h)\gamma(h+1)\right],$$

$$\begin{aligned}
&\sum_{h=1}^{n-1}\sum_{h',k'=1}^{n}[\gamma(h-h')\gamma(h+1-k') + \gamma(h-k')\gamma(h+1-h')]\\
&= 2\Bigg\{(n-1)\gamma^2(0) + 2\sum_{h=1}^{n-1}[2(n-h)-1]\gamma(0)\gamma(h)\\
&\quad + \sum_{h=1}^{n-1}[2(n-h)-1+(2(n-2h)-1)_+ - 1_{[2(n-2h)<0]}]\gamma^2(h)\\
&\quad + 2\sum_{h=1}^{n-2}\sum_{h'=h+1}^{n-1}[2(n-h')+(2(n-h-h')-1)_+ - 1_{[2(n-h-h')<0]}]\gamma(h)\gamma(h')\Bigg\},
\end{aligned}$$

$$\begin{aligned}
&\sum_{h,k=1}^{n-1}\sum_{h'=1}^{n}[\gamma(h-h')\gamma(k+1-h') + \gamma(h-h')\gamma(k+1-h')]\\
&= 2\Bigg\{(n-2)\gamma^2(0) + 2\sum_{h=1}^{n-1}[2(n-1-h)+1_{[h=n-1]}]\gamma(0)\gamma(h)\\
&\quad + \sum_{h=1}^{n-1}[2(n-1-h)+2(n-1-2h)_+ + 1_{[2h=n-1]}]\gamma^2(h)\\
&\quad + 2\sum_{h=1}^{n-2}\sum_{h'=h+1}^{n-1}[2(n-1-h')+1+2(n-1-h-h')_+ + 1_{[h+h'=n-1]}]\gamma(h)\gamma(h')\Bigg\}
\end{aligned}$$

and

$$\sum_{h,k=1}^{n-1} \sum_{h',k'=1}^{n} [\gamma(h-h')\gamma(k+1-k') + \gamma(h-k')\gamma(k+1-h')]$$
$$= 2\left[(n-1)\gamma(0) + \sum_{h=1}^{n-1}(2n-2h-1)\gamma(h)\right]^2.$$

With these simpler summations, we obtain

$$
\begin{aligned}
\operatorname{cov}(C_1, C_0) = {} & \frac{4}{n-1}\left[\left(1-\frac{1}{n}\right)\gamma(0)\gamma(1) + \sum_{h=1}^{n-2}\left(1-\frac{h+1}{n}\right)\gamma(h)\gamma(h+1)\right] \\
& - \frac{2}{n(n-1)}\Bigg\{\left(1-\frac{1}{n}\right)\gamma^2(0) + 4\sum_{h=1}^{n-1}\left(1-\frac{h+.5}{n}\right)\gamma(0)\gamma(h) \\
& + 2\sum_{h=1}^{n-1}\left[1-\frac{h+.5}{n}+\left(1-\frac{2h+.5}{n}\right)_+ - \frac{1_{[n-2h<0]}}{2n}\right]\gamma^2(h) \\
& + 4\sum_{h=1}^{n-2}\sum_{h'=h+1}^{n-1}\left[1-\frac{h'}{n}+\left(1-\frac{h+h'+.5}{n}\right)_+ - \frac{1_{[n-h-h'<0]}}{2n}\right]\gamma(h)\gamma(h')\Bigg\} \\
& - \frac{2}{n(n-1)}\Bigg\{\left(1-\frac{1}{n-1}\right)\gamma^2(0) + 4\sum_{h=1}^{n-1}\left[1-\frac{h}{n-1}+\frac{1_{[h=n-1]}}{2(n-1)}\right]\gamma(0)\gamma(h) \\
& + 2\sum_{h=1}^{n-1}\left[1-\frac{h}{n-1}+\left(1-\frac{2h}{n-1}\right)_+ + \frac{1_{[2h=n-1]}}{2(n-1)}\right]\gamma^2(h) \\
& + 4\sum_{h=1}^{n-2}\sum_{h'=h+1}^{n-1}\left[1-\frac{h'-.5}{n-1}+\left(1-\frac{h+h'}{n-1}\right)_+ + \frac{1_{[h+h'=n-1]}}{2(n-1)}\right]\gamma(h)\gamma(h')\Bigg\} \\
& + \frac{2}{n^2}\left[\gamma(0)+2\sum_{h=1}^{n-1}\left(1-\frac{h-.5}{n-1}\right)\gamma(h)\right]^2 \\
= {} & \gamma^2(0)\left\{4g_1^{(2)}(n-1,\boldsymbol{\rho}) - v_1(n,\boldsymbol{\rho}) - v_2(n,\boldsymbol{\rho}) + 2[f(n,\boldsymbol{\rho})]^2\right\}.
\end{aligned}
\tag{12}
$$

Substituting (9), (10), (11), and (12) into the equation (4) leads to the bias of R_1 in (5). This completes the proof.

REFERENCES

T. W. Anderson (1971). *The statistical analysis of time series*. Wiley, New York.

J. M. Bardet, G. Lang, G. H. Oppenheim, A. Philippe & M. Taqqu (2002). Semi-parametric estimation of the long-range dependent processes: A Survey. In *Theory and Applications of Long-Range Dependence* (P. Doukhan, G. Oppenheim & M. Taqqu, eds.), Birkhauser, Boston, pp. 557–577.

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

J. Beran & Y. Feng (2002). SEMIFAR models - A semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Computational Statistics and Data Analysis*, 40, 393–419.

S. Beveridge & C. Oickle (1993). Estimating fractionally integrated time series models. *Economics letters*, 43, 137–142.

P. F. Craigmile, P. Guttorp & D. B. Percival (2005). Wavelet-based parameter estimation for polynomial contaminated fractionally differenced processes. *IEEE Transactions on Signal Processing*, 53, 3151–61.

R. Fox & M. S. Taqqu (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics*, 14, 517–532.

J. F. Geweke & S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journals of Time Series Analysis*, 4, 221–238.

C. W. Granger & R. Joyeux (1980). An introduction to long memory time series models and fractional differencing. *Journals of Time Series Analysis*, 1, 15–29.

M. A. Hauser (1997). Semiparametric and nonparametric testing for long memory: A Monte Carlo study. *Empirical Economics*, 22, 247–271.

M. A. Hauser (1999). Maximum likelihood estimators for ARMA and ARFIMA models: A Monte Carlo study. *Journal of Statistical Planning and Inference*, 80, 229–255.

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

J. R. M. Hosking (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *Journal of Econometrics*, 73, 261–284.

H. Kettani & J. A. Gubner (2003). Estimation of the long-range dependence parameter of fractional ARIMA precesses. *Proceedings of the 28th Annual IEEE International Conference on Local Computer Networks*, 28, 307–308.

G. Koop, E. Ley, J. Osiewalski, & M. F. J. Steel (1997). Bayesian analysis of long memory and persistence using ARFIMA models. *Journal of Econometrics*, 76, 149–169.

H. R. Künsch (1987). Statistical aspects of self-similar processes. *Proceedings of the First World Congress of the Bernoulli Society*, 1, 67–74.

W. K. Li & A. I. McLeod (1986). Fractional time series modelling. *Biometrika*, 73, 217–221.

O. Lieberman (2001). Penalized maximum likelihood estimation for fractional Gaussian processes. *Biome-trika*, 88, 888–894.

F. H. C. Marriott & J. A. Pope (1954). Bias in the estimation of autocorrelations. *Biometrika*, 41, 390–402.

E. Moulines & P. Soulier (1999). Broadband log-periodogram regression of time series with long-range dependence. *Annals of Statistics*, 27, 1415–1439.

P. Newbold & C. Agiakloglou (1993). Bias in the sample autocorrelations of fractional noise. *Biometrika*, 80, 698–702.

M. Ø. Nielsen & P. H. Frederiksen (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews*, 24, 405–443.

J. S. Pai & N. Ravishanker (1998). Bayesian analysis of autoregressive fractionally integrated moving-average processes. *Journals of Time Series Analysis*, 19, 99–112.

K. D. Patterson (2007). Bias reduction through first-order mean correction, bootstrapping and recursive mean adjustment. *Journal of Applied Statistics*, 34, 23–45.

V. Reisen, B. Abraham & S. Lopes (2001). Estimation of parameters in ARFIMA processes: A simulation study. *Communications in Statistics-Simulation and Computation*, 30, 787–803.

P. M. Robinson (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 23, 1630–1661.

J. Smith, N. Taylor & S. Yadav (1997). Comparing the bias and misspecification in ARFIMA models. *Journals of Time Series Analysis*, 18, 507–527.

F. Sowell (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53, 165–188.

P. Whittle (1951). *Hypothesis testing in time series analysis*. Hafner, New York.