Empirical Economics (1993) 18:791–806

Small Sample Bias in Conditional Sum-of-Squares Estimators of Fractionally Integrated ARMA Models[1]

CHING-FAN CHUNG AND RICHARD T. BAILLIE

Department of Economics, Michigan State University, 101 Marshall Hall, East Lansing, Michigan 48824-1038, USA

Abstract: This paper considers estimation of the parameters for the fractionally integrated class of processes known as ARFIMA. We consider the small sample properties of a conditional sum-of-squares estimator that is asymptotically equivalent to MLE. This estimator has the advantage of being relatively simple and can estimate all the parameters, including the mean, simultaneously. The simulation evidence we present indicates that estimation of the mean can make a considerable difference to the small sample bias and MSE of the other parameter estimates.

JEL Classification System-Number: C22

1 Introduction

This paper is concerned with the estimation of fractionally integrated $I(d)$ time-series processes, which were originally introduced by Granger (1980), Granger and Joyeux (1980), and Hosking (1981). An $I(d)$ process exhibits long memory and generates very slow, but eventual, decay in its impulse-response weights, or moving-average representation. Hence, one attraction of $I(d)$ processes is that they avoid knife-edge choices between unit-root $I(1)$ processes, which generate complete persistence, and the alternative of stationary and invertible ARMA models, which imply relatively rapid exponential decay in their impulse-response weights. There has recently been considerable interest in the possibility that many macroeconomic and financial time series possess long-memory properties consistent with $I(d)$ behavior. Studies by Shea (1989, 1991), Diebold and Rudebusch (1989, 1991), Sowell (1992b), Baillie, Chung, and Tieslau (1992) and others have applied $I(d)$ processes to a variety of economic problems.

Section 2 of this paper provides a brief survey of some characteristics of the fractional white noise process and the ARFIMA(p, d, q) model where the $I(d)$ behavior is appended with ARMA behavior. Section 3 then discusses different estimation procedures that have been suggested for the ARFIMA process. The most attractive estimator currently available is a full maximum likelihood estimator (MLE) proposed by Sowell (1992a). A recent study by Cheung and Die-

[1] We gratefully acknowledge helpful comments from the editor, Baldev Raj, and three anonymous referees.

0377–7332/93/4/791–806 $2.50 © 1993 Physica-Verlag, Heidelberg

bold (1993) has provided further simulation evidence on the performance of Sowell's full MLE in estimating the fractional white noise, or ARFIMA(0, d, 0) model, compared to an approximate MLE of Fox and Taqqu (1986). When the mean of the process is known, Cheung and Diebold not surprisingly find Sowell's estimator to be quite satisfactory. However, when the mean of the process is unknown, the approximate MLE of Fox and Taqqu yields similar biases and mean squared errors of the parameter estimates.

The main contribution of this paper is to provide a further study about the small sample performance of the estimation of the ARFIMA model based on the conditional sum-of-squares (CSS) estimator. With the initial observations fixed, the CSS estimator is asymptotically equivalent to full MLE (a brief proof of this is provided in the appendix). The CSS estimator has been widely used for ARMA models, transfer function models, ARCH and GARCH models, and has been suggested for $I(d)$ processes by Hosking (1984) and Li and McLeod (1986). In section 3 of this paper we discuss some of the advantages of the CSS estimator. In particular, the CSS estimator can deal with quite complicated ARFIMA models with non-normal disturbances and innovations that exhibit time-dependent heteroskedasticity of the ARCH form.

The results of a detailed simulation study are reported in section 4. Unlike previous work in this area, this paper considers the estimation of ARFIMA(p, d, q) processes with p and q equal to 1 or 2, as opposed to the pure fractional white-noise case. Since the CSS estimator is asymptotically equivalent to MLE, our interest is in determining its performance in small samples. In particular, we find that when the unknown mean is estimated jointly with the fractional differencing parameter d and the ARMA parameters, substantial bias may result in small sample. Interestingly, when we evaluate three different estimators of the mean, we find that in some cases the sample median may perform better than the more usual sample mean or the MLE of the mean.

2 Fractional Integrated Processes

Following Granger (1980, 1981), Granger and Joyeux (1980), and Hosking (1981), a discretely observed time-series process y_t, with mean μ is said to be integrated of order d, i.e., $I(d)$, if

$$(1 - L)^d(y_t - \mu) = u_t \ , \tag{1}$$

where u_t is a weakly stationary $I(0)$ process and $-0.5 < d < 0.5$. For $0 < d < 0.5$, the process y_t exhibits long memory in the sense that its autocorrelations will have a hyperbolic rate of convergence. For $-0.5 < d < 0$, the process y_t is said to have intermediate memory. If u_t is a stationary and invertible ARMA process, then y_t is generated by the Autoregressive Fractionally Integrated

Moving Average, or ARFIMA(p, d, q) process:

$$\phi(L)(1-L)^d(y_t-\mu)=\theta(L)\varepsilon_t\ , \tag{2}$$

where $\phi(L)=1-\sum_{j=1}^{p}\phi_jL^j$ and $\theta(L)=1+\sum_{j=1}^{q}\theta_jL^j$ are polynomials in the lag operator of order p and q respectively and have all their roots lying outside the unit circle, and ε_t is a white-noise sequence with a variance σ^2. Granger (1981) and Hosking (1981) show that the infinite autoregressive weights, the infinite moving-average representation weights and the autocorrelation functions all decline at a hyperbolic rate, as opposed to the conventional exponential rate associated with the stationary and invertible class of ARMA processes. For example, for the fractionally integrated white-noise process

$$(1-L)^d(y_t-\mu)=\varepsilon_t\ , \tag{3}$$

provided $d<0.5$, the process will be stationary and will possess the infinite moving-average representation,

$$y_t=\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}\ , \tag{4}$$

where $\psi_j=\Gamma(j+d)/[\Gamma(d)\Gamma(j+1)]$. Also, provided $d>-0.5$ the process will be invertible and can be expressed as an infinite autoregressive representation,

$$y_t=-\sum_{j=1}^{\infty}\pi_jy_{t-j}+\varepsilon_t\ , \tag{5}$$

where $\pi_j=\Gamma(j-d)/[\Gamma(-d)\Gamma(j+1)]$. The autocorrelation coefficients ρ_j for the fractional white-noise ARFIMA(0, d, 0) process are given by:

$$\rho_j=\frac{\Gamma(j+d)\Gamma(1-d)}{\Gamma(j-d+1)\Gamma(d)}\ .$$

Parametric expressions for the autocorrelations of the general ARFIMA (p, d, q) process are given by Sowell (1992a). They are complicated functions of the hypergeometric function. However, for large lags, hyperbolic decay takes place in the autoregressive representation weights, the moving-average representation weights and the autocorrelations of the ARFIMA(p, d, q) process. In particular, Granger (1980) and Hosking (1981) show that

$$\pi_j\sim j^{-(1+d)}\ ,\qquad \psi_j\sim j^{d-1}\qquad \text{and}\qquad \rho_j\sim j^{(2d-1)}\ .$$

3 Estimation of Fractionally Integrated Processes

We now consider estimation of the $p+q+3$ dimensional vector of parameters $\lambda=(d,\mu,\phi_1,\ldots\phi_p,\theta_1,\ldots\theta_q,\sigma^2)$. Several different approaches have been suggested for parameter estimation. First, Geweke and Porter-Hudak (1983) suggested a two-step estimator that allowed d to be estimated from a regression of ordinates of the periodogram of y_t on a trigonometric function. Although the estimator is straightforward to apply and is potentially robust to non-Gaussian disturbances, Agiakoglou, Newbold and Wohar (1992) have recently shown this estimator to be severely biased in the presence of strong autocorrelation in the u_t process. There is the additional disadvantage that it is a two-step estimator, and the ARMA parameters would have to be estimated from a series with the $I(d)$ component removed by filtering.

An approximate maximum likelihood procedure in the frequency domain has been suggested by Fox and Taqqu (1986). The estimator assumes unconditional normality and numerically minimizes the quantity

$$\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f(\omega)}\left|\sum_{t=0}^{T-1} y_t e^{it\omega}\right|^2 d\omega \ ,$$

where $f(\omega)=\dfrac{\sigma^2}{2\pi}\dfrac{\theta(e^{i\omega})\theta(e^{-i\omega})}{\phi(e^{i\omega})\phi(e^{-i\omega})}\dfrac{1}{|1-e^{-i\omega}|^{2d}}$ is the spectral density evaluated at frequency ω.

In a seminal paper, Sowell (1992a) was able to derive the full maximum likelihood estimator (MLE) for the ARFIMA(p, d, q) process with normally distributed innovations. Following Sowell (1992a), and under normality, the logarithm of the likelihood can be expressed in the time domain as

$$L(\lambda) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu) \ ,$$

where y is the T dimensional vector of y_t and Σ is the corresponding $T \times T$ autocovariance matrix, where each element is a non-linear function of hypergeometric functions. Sowell (1992b) has also provided a detailed example of the application of his estimator to US real GNP data.

In this paper we consider the properties of an alternative conditional sum-of-squares (CSS) estimator which minimizes

$$\begin{aligned} S(\lambda) &= \frac{1}{2}\log\sigma^2 + \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2 \\ &= \frac{1}{2}\log\sigma^2 + \frac{1}{2\sigma^2}\sum_{t=1}^{T}[\phi(L)\theta(L)^{-1}(1-L)^d(y_t-\mu)]^2 \ . \end{aligned}$$

If the initial observations $y_0, y_{-1}, y_{-2}, \ldots$ are assumed fixed, then minimizing the conditional sum-of-squares function will be asymptotically equivalent to MLE. The minimum CSS procedure in the context of ARFIMA processes was originally suggested by Hosking (1984). It is worth noting that similar estimation methods have been implemented in the stationary and invertible class of ARMA models. Box and Jenkins (1976) used the minimum CSS estimator, while Newbold (1974) considered the full MLE with the initial observations treated as stochastic. For an infinite number of observations the CSS estimator will be equivalent to MLE. In the appendix we show that the effect of initial observations is negligible asymptotically. Similar results for ARMA processes are provided by Pierce (1971).

There are several reasons for considering the CSS estimator of potential interest for practical applications. First, it is readily extendable to situations where the innovations possess time-dependent heteroskedasticity of the ARCH form. It is well known that even for relatively simple non-linear dynamic models, e.g., Engel's (1982) ARCH process, full MLE is intractable and a type of CSS estimator has to be used. Indeed, in a study on the variability of inflation, Baillie, Chung and Tieslau (1992) use the CSS approach to estimate the parameters of an ARFIMA(0, d, 1)-GARCH(1, 1) process with a conditional density of student t. While it is beyond the range of the present study to examine the properties of the CSS estimator in such a complex setting, it is desirable to look at the CSS estimator for at least some simple ARFIMA models.

Secondly, although Sowell's (1992a) work provides an elegant approach for computing the full MLEs of the "vanilla" ARFIMA model with unconditional normality and no ARCH effects, it is nevertheless computationally demanding, with every iteration of the likelihood requiring inversion of a T-dimensional covariance matrix and having each element a non-linear function of hypergeometric functions.

A third reason for being interested in the CSS estimator is that it does share many features with the the approximate MLE[2] of Fox and Taqqu (1986) when the basic vanilla ARFIMA model is being estimated; and in many cases the Fox-Taqqu estimator has at least as desirable characteristics as the full MLE of Sowell. Some simulation evidence on the relative performance of Sowell's full MLE and the approximate MLE of Fox and Taqqu (1986) has been provided by Cheung and Diebold (1993). In the case of the mean μ being known, Cheung and Diebold document the excellent performance of Sowell's (1992a) full MLE. However, when μ is unknown and has to be estimated, they found the Fox-Taqqu estimator to perform as well as the full MLE. The simulation study of Cheung and Diebold is confined to the ARFIMA(0, d, 0), case. It is of interest to see whether the CSS estimator shares similar properties with the Fox-Taqqu estimator. By extending the range of our investigation to include ARFIMA (p, d, q) models instead of just the fractional white-noise ARFIMA(0, d, 0) model, we hope to gain further insight into this problem. It is possible that the inclusion of ARMA parameters will exacerbate the bias problems caused by estimation of the unknown mean estimator. One of the purposes of our simulation study is to assess this issue on both the parameter estimates and their standard errors.

---

2 While it is not the purpose of this study to compare the CSS estimator with the full MLE of Sowell (1992a), it should be noted that we have some preliminary work on this issue and which is available from us on request. Initial work with a relatively small number of replications suggests the CSS estimator performs about the same as the full MLE in sample sizes of $T = 100$.

Li and McLeod (1986) have shown that with μ equal to zero, the estimates of the remaining parameters are $T^{1/2}$ consistent and have conjectured that they are also asymptotically normal. Taqqu (1975) has considered the sample mean of fractional white noise, while Yajima (1988) has considered the properties of OLS and GLS estimates in the regression model with $I(d)$ disturbances. A corollary of these results is that the sample mean is convergent at a rate of $T^{1/2-d}$ to its limiting distribution. Due to the unusually slow (when d is greater than 0) convergence rate of the estimator of μ, an important issue concerns the performance of the estimation with μ being jointly estimated. As previously described, the study by Cheung and Diebold (1993) noted the unexpected performance of Sowell's full MLE of the fractional-differencing parameter d when μ is unknown in an ARFIMA(0, d, 0) model. Our simulation study goes one step further to examine the effect of the estimation of μ on all parameters in an ARFIMA(p, d, q) model.

4 Simulation Results

In this section we report the results of the simulation of the small sample properties of the CSS estimator with all the parameters estimated simultaneously. We consider different values of p and q as 0, 1 or 2. Hence our study is the first to consider the estimation of the $I(d)$ process in the presence of ARMA parameters and an unknown mean.

Unless stated to the contrary all the simulations were based on 500 replications of a sample size of 100. For each replication 100 values were generated from the standard normal random variable and assigned into a column vector denoted by e. We then computed the analytic autocovariance matrix Σ together with its Cholesky decomposition matrix[3] denoted by C, i.e., $\Sigma = CC'$. As mentioned earlier, the autocovariance matrix Σ was first derived by Sowell (1992a) who presented complete formulas that involve combinations of hypergeometric functions. The vector y of 100 realized values is then constructed as:

$$y = \mu + Ce \ ,$$

where μ is the mean of the process.[4] The Monte Carlo results are independent of the mean and variance and for convenience both μ and σ^2 were set as one

[3] All our computations are conducted using the personal computer programming language GAUSS, including Cholesky decomposition, which is also available in many other software packages.

[4] It has been suggested in the literature that, in order for the long-memory ARFIMA series y not to be affected by "the initial values," it is advisable to simulate longer series and drop the first subset of values. We find that such a practice is unnecessary if y is simulated using our method.

throughout the study. For each configuration of parameter values the CSS estimator is computed for each realization. Apart from the CSS estimator of μ we also report the median, since this may be more robust to "outliers" caused by the persistence of the autocovariances of an ARFIMA process.

As usual, the estimator of the asymptotic covariance matrix of the parameter estimates is based on the numerically computed Hessian. Part of the interest in our simulation study is to assess the accuracy of the estimated standard errors with the theoretical standard errors.

The simulation results[5] are summarized in Tables 1A through 1C. In each block on the extreme left, the true parameter values used in the data-generating process are given. For the ARFIMA(1, d, 0) process, values of ϕ corresponding to 0.7, 0.2, -0.3 and -0.8 are used. For the ARFIMA(0, d, 1) process values of θ corresponding to 0.9, 0.5, -0.3 and -0.8 are used. For ARFIMA(1, d, 1) processes the assigned values for the ARMA parameters are the four different combinations of ± 0.2 and ± 0.5. For each data-generating process, the next three blocks in Tables 1A–1C, under the headings CSS, Mean, Median, contain the estimation biases, as well as the square roots of the mean square errors (SRMSE) in the parentheses, based on 500 replications. For each replication μ is estimated by the CSS, the sample mean, and the sample median, respectively. The CSS method estimates all the parameters simultaneously while the other methods estimate μ by the sample mean or median before estimating the other parameters. The simulation results for the estimates of σ^2 are omitted to save space. Since we are especially interested in the estimates of the fractional difference parameter d, the least bias in estimating d among the three estimation procedures is underlined in each row.

The main conclusion resulting from these tables is that, with a sample size of 100, the bias in estimating the ARFIMA model by CSS can be quite substantial when the mean μ is also estimated. A closer examination of the Monte Carlo results shows the following patterns: (a) when the true values of d, ϕ, and θ are all positive, the CSS estimator contains relatively small biases; (b) when the true values of d, ϕ, and θ are all negative, estimation based on the sample mean produces the least bias; (c) for all other cases, estimation based on sample medians generally produces the least bias, which may be due to the median being robust to outliers.

One interesting aspect of the study concerns the degree to which the parameter estimation bias is due to the intercept μ being estimated. Before examining this issue, we first note that Table 1C for the model ARFIMA(1, 0.3, 1) contains some of the most substantial biases. When the values of the ARMA parameters ϕ and θ are both positive, the best estimator (based on the MLE of μ) of d gives a downward bias that is greater than 0.2. Moderate bias is also observed when the ARMA parameters have negative values. These typical results based on the

[5] Simulations based on ARFIMA(2, d, 0) and ARFIMA(0, d, 2) processes have also been conducted. The results are similar to those reported in this paper.

Table 1A. Simulation of the ARFIMA (1, d, 0) model

True		CSS			Mean			Median		
d	ϕ	d	ϕ	μ	d	ϕ	μ	d	ϕ	μ
0.3	0.7	0.010	−0.042	0.121	−0.086	0.023	0.140	−0.099	0.034	0.020
		(0.233)	(0.201)	(2.142)	(0.185)	(0.149)	(1.469)	(0.169)	(0.136)	(1.398)
0.3	0.2	−0.240	0.192	0.031	−0.270	0.222	0.010	−0.221	0.182	−0.030
		(0.335)	(0.318)	(0.587)	(0.321)	(0.312)	(0.533)	(0.285)	(0.286)	(0.512)
0.3	−0.3	−0.106	0.081	0.004	−0.098	0.072	−0.001	−0.069	0.053	0.001
		(0.217)	(0.222)	(0.334)	(0.159)	(0.166)	(0.331)	(0.150)	(0.163)	(0.325)
0.3	−0.8	−0.054	0.031	−0.004	−0.057	0.032	−0.004	−0.035	0.023	−0.002
		(0.107)	(0.079)	(0.236)	(0.103)	(0.074)	(0.233)	(0.095)	(0.075)	(0.256)
0.	0.7	−0.063	0.007	0.012	−0.106	0.038	0.017	−0.090	0.033	0.023
		(0.217)	(0.173)	(0.499)	(0.170)	(0.126)	(0.308)	(0.166)	(0.130)	(0.339)
0.	0.2	−0.270	0.220	−0.001	−0.248	0.199	−0.005	−0.151	0.124	0.002
		(0.319)	(0.308)	(0.136)	(0.285)	(0.288)	(0.123)	(0.252)	(0.265)	(0.150)
0.	−0.3	−0.093	0.070	0.004	−0.092	0.068	0.004	−0.036	0.031	0.005
		(0.153)	(0.171)	(0.082)	(0.142)	(0.162)	(0.076)	(0.128)	(0.145)	(0.115)
0.	−0.8	−0.059	0.030	−0.001	−0.052	0.027	−0.000	0.023	0.010	−0.011
		(0.100)	(0.073)	(0.059)	(0.096)	(0.069)	(0.058)	(0.093)	(0.067)	(0.116)
−0.3	0.7	−0.100	0.037	0.007	−0.094	0.037	0.002	−0.011	−0.023	0.002
		(0.158)	(0.124)	(0.106)	(0.161)	(0.128)	(0.107)	(0.183)	(0.169)	(0.129)
−0.3	0.2	−0.179	0.149	0.000	−0.127	0.098	−0.002	0.050	−0.050	−0.001
		(0.270)	(0.281)	(0.040)	(0.237)	(0.256)	(0.041)	(0.196)	(0.211)	(0.086)
−0.3	−0.3	−0.085	0.062	−0.001	−0.049	0.033	0.003	0.118	−0.056	−0.000
		(0.149)	(0.157)	(0.025)	(0.118)	(0.138)	(0.027)	(0.131)	(0.117)	(0.087)
−0.3	−0.8	−0.028	0.024	−0.001	−0.013	0.022	−0.000	0.181	−0.012	0.001
		(0.105)	(0.073)	(0.019)	(0.091)	(0.070)	(0.022)	(0.138)	(0.065)	(0.114)

Among the three estimates of d in each row, the one with the least bias is underlined.
All simulations were based on 500 replications of a sample size of 100.

Table 1B. Simulation of the ARFIMA (0, d, 1) model

True		CSS			Mean			Median		
d	θ	d	θ	μ	d	θ	μ	d	θ	μ
0.3	0.9	<u>0.005</u> (0.123)	−0.018 (0.055)	−0.029 (1.564)	−0.048 (0.101)	−0.037 (0.070)	−0.005 (0.801)	−0.046 (0.100)	−0.034 (0.068)	−0.029 (0.819)
0.3	0.5	<u>−0.045</u> (0.123)	0.021 (0.119)	−0.055 (0.736)	−0.047 (0.117)	0.020 (0.108)	0.014 (0.613)	−0.050 (0.111)	0.023 (0.108)	−0.003 (0.623)
0.3	−0.3	−0.138 (0.212)	0.123 (0.221)	−0.018 (0.311)	−0.140 (0.199)	0.127 (0.213)	−0.011 (0.290)	<u>−0.108</u> (0.190)	0.097 (0.209)	0.014 (0.318)
0.3	−0.8	−0.398 (0.357)	0.359 (0.347)	0.006 (0.087)	−0.307 (0.334)	0.279 (0.338)	−0.008 (0.090)	<u>−0.075</u> (0.272)	0.086 (0.249)	0.002 (0.123)
0.	0.9	−0.038 (0.104)	−0.018 (0.058)	−0.002 (0.259)	−0.047 (0.099)	−0.021 (0.060)	−0.005 (0.187)	<u>−0.033</u> (0.091)	−0.021 (0.060)	0.002 (0.208)
0.	0.5	−0.062 (0.116)	0.037 (0.112)	0.011 (0.150)	−0.055 (0.109)	0.023 (0.112)	0.005 (0.148)	<u>−0.036</u> (0.094)	0.023 (0.110)	−0.008 (0.179)
0.	−0.3	−0.139 (0.202)	0.123 (0.213)	0.006 (0.071)	−0.145 (0.186)	0.132 (0.190)	−0.004 (0.068)	<u>−0.021</u> (0.204)	0.020 (0.229)	−0.003 (0.099)
0.	−0.8	−0.317 (0.342)	0.304 (0.337)	0.001 (0.021)	−0.165 (0.296)	0.157 (0.287)	0.002 (0.024)	<u>0.105</u> (0.191)	0.023 (0.134)	0.002 (0.096)
−0.3	0.9	−0.039 (0.100)	−0.012 (0.056)	−0.002 (0.061)	−0.033 (0.093)	−0.017 (0.056)	−0.006 (0.066)	<u>0.026</u> (0.088)	−0.017 (0.059)	−0.002 (0.115)
−0.3	0.5	−0.056 (0.119)	0.022 (0.112)	−0.003 (0.046)	−0.043 (0.110)	0.023 (0.110)	0.002 (0.048)	<u>0.032</u> (0.104)	−0.009 (0.114)	−0.002 (0.088)
−0.3	−0.3	−0.127 (0.206)	0.122 (0.217)	0.002 (0.023)	<u>−0.080</u> (0.180)	0.080 (0.204)	−0.001 (0.025)	0.305 (0.280)	−0.246 (0.266)	0.004 (0.089)
−0.3	−0.8	−0.112 (0.243)	0.150 (0.214)	−0.000 (0.011)	<u>0.015</u> (0.183)	0.061 (0.142)	−0.001 (0.015)	0.309 (0.182)	0.016 (0.068)	−0.005 (0.101)

Among the three estimates of d in each row, the one with the least bias is underlined.
All simulations were based on 500 replications of a sample size of 100.

Table 1C. Simulation of the ARFIMA (1, d, 1) model

True			CSS				Mean				Median			
d	ϕ	θ	d	ϕ	θ	μ	d	ϕ	θ	μ	d	ϕ	θ	μ
0.3	0.5	0.2	−0.205	0.056	0.109	0.050	−0.331	0.157	0.112	−0.048	−0.294	0.135	0.116	0.045
			(0.350)	(0.342)	(0.161)	(1.398)	(0.287)	(0.274)	(0.137)	(1.018)	(0.285)	(0.291)	(0.148)	(1.039)
0.3	0.2	0.5	−0.258	0.177	0.043	−0.017	−0.302	0.216	0.039	0.015	−0.294	0.217	0.033	−0.045
			(0.379)	(0.387)	(0.127)	(1.051)	(0.355)	(0.383)	(0.125)	(0.809)	(0.321)	(0.350)	(0.126)	(0.806)
0.3	−0.2	−0.5	−0.215	−0.074	0.269	0.005	−0.223	−0.066	0.266	−0.002	−0.108	−0.049	0.152	0.001
			(0.258)	(0.252)	(0.345)	(0.190)	(0.245)	(0.246)	(0.334)	(0.175)	(0.227)	(0.203)	(0.327)	(0.213)
0.3	−0.5	−0.2	−0.094	−0.039	0.125	−0.003	−0.109	−0.036	0.128	0.001	−0.077	−0.023	0.084	0.006
			(0.228)	(0.148)	(0.316)	(0.280)	(0.190)	(0.146)	(0.269)	(0.233)	(0.184)	(0.155)	(0.277)	(0.257)
0.	0.5	0.2	−0.362	0.185	0.129	−0.009	−0.328	0.163	0.124	0.005	−0.238	0.107	0.095	−0.010
			(0.289)	(0.285)	(0.143)	(0.256)	(0.270)	(0.276)	(0.145)	(0.233)	(0.261)	(0.282)	(0.144)	(0.259)
0.	0.2	0.5	−0.322	0.241	0.042	−0.004	−0.291	0.222	0.033	−0.000	−0.197	0.132	0.038	−0.009
			(0.352)	(0.370)	(0.125)	(0.183)	(0.303)	(0.340)	(0.123)	(0.185)	(0.280)	(0.330)	(0.133)	(0.205)
0.	−0.2	−0.5	−0.223	−0.084	0.287	−0.002	−0.193	−0.057	0.234	−0.002	0.121	−0.073	−0.035	−0.002
			(0.227)	(0.218)	(0.335)	(0.044)	(0.210)	(0.230)	(0.325)	(0.043)	(0.289)	(0.160)	(0.306)	(0.095)
0.	−0.5	−0.2	−0.122	−0.039	0.134	−0.004	−0.108	−0.029	0.128	−0.003	0.075	−0.016	−0.045	0.004
			(0.192)	(0.142)	(0.280)	(0.053)	(0.172)	(0.151)	(0.264)	(0.055)	(0.225)	(0.142)	(0.319)	(0.104)
−0.3	0.5	0.2	−0.284	0.135	0.110	−0.001	−0.208	0.078	0.104	0.002	−0.057	−0.035	−0.081	−0.006
			(0.277)	(0.291)	(0.154)	(0.080)	(0.270)	(0.296)	(0.159)	(0.078)	(0.254)	(0.311)	(0.179)	(0.103)
−0.3	0.2	0.5	−0.186	0.121	0.043	0.001	−0.154	0.095	0.038	0.001	−0.002	−0.032	0.039	0.006
			(0.275)	(0.351)	(0.146)	(0.063)	(0.241)	(0.319)	(0.141)	(0.061)	(0.217)	(0.297)	(0.152)	(0.106)
−0.3	−0.2	−0.5	−0.128	−0.067	0.192	−0.000	−0.060	−0.050	0.128	−0.001	0.496	−0.146	−0.252	−0.005
			(0.235)	(0.205)	(0.312)	(0.014)	(0.221)	(0.204)	(0.308)	(0.018)	(0.264)	(0.127)	(0.201)	(0.100)
−0.3	−0.5	−0.2	−0.086	−0.031	0.110	0.001	−0.028	−0.015	0.041	0.000	0.452	−0.023	−0.366	0.003
			(0.203)	(0.151)	(0.284)	(0.019)	(0.177)	(0.140)	(0.264)	(0.022)	(0.335)	(0.123)	(0.333)	(0.097)

Among the three estimates of d in each row, the one with the least bias is underlined.
All simulations were based on 500 replications of a sample size of 100.

Table 2. Simulation of the ARFIMA (1, 0.3, 1) model with known mean

True			CSS		
d	ϕ	θ	d	ϕ	θ
0.3	0.5	0.2	−0.063 (0.268)	−0.035 (0.332)	0.081 (0.189)
0.3	0.2	0.5	−0.085 (0.255)	0.047 (0.310)	0.021 (0.142)
0.3	−0.2	−0.5	−0.003 (0.255)	−0.061 (0.188)	0.066 (0.330)
0.3	−0.5	−0.2	0.029 (0.198)	−0.009 (0.143)	−0.016 (0.294)

All simulations were based on 500 replications.

ARFIMA(1, 0.3, 1) process are a useful starting point for the investigation of the impact of estimating μ.

Table 2 presents results for the same model, conditional on μ being known and with a sample size of 100. The magnitude of the bias drops to an order that is quite similar to those reported in Sowell's (1992a) Monte Carlo study. The contrast between Tables 2 and 1C offers the strongest evidence yet for the argument that estimating μ substantially increases the degree of bias in the other parameter estimates.

Table 3 gives Monte Carlo results based on a larger sample size of 300; μ is again assumed unknown and is estimated. Comparing Tables 3 and 1C indicates a considerable reduction in the bias. Clearly, the small sample bias is substantially reduced as the sample size increases. In particular, the estimate of the fractional differencing parameter d experiences more than a 40 percent reduction in the bias. In three out of the four cases the reduction is more than 50 percent.

In contrast to the large bias in estimating d and the ARMA parameters, the bias of the estimates of μ is reasonably small. When $d > 0$, the SRMSE of the μ estimates are usually quite large due to the slower convergence rate of $T^{1/2-d}$ as mentioned earlier. It is surprising that large SRMSE do not necessarily imply a large bias in the estimates of μ. These results confirm the assertion of Yajima (1988) and Samarov and Taqqu (1988) that the efficiency of the sample mean and the MLE of μ is about the same when $d > -0.3$.

5 Conclusion

The estimation of ARFIMA processes is likely to be an important activity for economists and area of research for econometricians for some time to come. As

Table 3. Simulation of the ARFIMA (1, 0.3, 1) model based on the sample size 300

True			CSS				Mean				Median			
d	ϕ	θ	d	ϕ	θ	μ	d	ϕ	θ	μ	d	ϕ	θ	μ
0.3	0.5	0.2	<u>−0.123</u>	0.060	0.057	0.022	−0.188	0.114	0.057	−0.025	−0.176	0.103	0.059	−0.088
			(0.256)	(0.236)	(0.100)	(1.149)	(0.236)	(0.215)	(0.089)	(0.753)	(0.242)	(0.226)	(0.094)	(0.773)
0.3	0.2	0.5	<u>−0.107</u>	0.088	0.010	−0.037	−0.109	0.089	0.009	−0.057	−0.124	0.106	0.005	0.010
			(0.240)	(0.261)	(0.080)	(0.773)	(0.215)	(0.237)	(0.084)	(0.620)	(0.214)	(0.245)	(0.086)	(0.631)
0.3	−0.2	−0.5	−0.066	−0.032	0.092	0.003	−0.079	−0.019	0.094	−0.005	<u>−0.050</u>	−0.020	0.065	−0.002
			(0.169)	(0.107)	(0.212)	(0.155)	(0.150)	(0.108)	(0.191)	(0.145)	(0.154)	(0.102)	(0.195)	(0.151)
0.3	−0.5	−0.2	−0.031	−0.013	0.039	0.009	−0.044	−0.014	0.059	0.000	<u>−0.031</u>	−0.007	0.036	0.001
			(0.125)	(0.083)	(0.186)	(0.183)	(0.107)	(0.087)	(0.165)	(0.186)	(0.104)	(0.079)	(0.152)	(0.198)

Among the three estimates of d in each row, the one with the least bias is underlined.
All simulations were based on 500 replications.

indicated in our results, the assumption of μ being known is far from innocuous. The estimation of μ in small sample sizes corrupts the CSS estimates of the other parameters. One corollary of our results is that estimation of ARFIMA models for small samples, e.g., T less than 150, should only be attempted with extreme caution. The resulting bias will be sufficiently large to make inference extremely unreliable. A useful check of the specification of the ARFIMA model in small sample cases would be to use all three estimates of μ (the CSS, the sample mean, as well as the sample median). Radically different estimates of the other parameters are likely to suggest a specification problem.

The CSS estimator appears to be a useful technique for quite general ARFIMA models in moderate to large samples but can possess substantial small-sample bias. The CSS estimator is of interest since it can be extended to more complicated models with non-normal conditional densities and time-dependent heteroskedasticity of the ARCH type. Some applications in this context are reported in Baillie, Chung and Tieslau (1992).

Appendix

In this appendix we show that setting the initial values $y_0, y_{-1}, y_{-2}, \ldots$ to zero is immaterial in examining the asymptotic distribution of the CSS estimator in the simple fractionally integrated white-noise ARFIMA(0, d, 0) model. Similar results for ARMA processes and univariate random walks are given by Pierce (1971) and Phillips (1987), respectively.

As indicated in the text, the ARFIMA(0, d, 0) model

$$(1 - L)^d y_t = \varepsilon_t \ ,$$

has the infinite moving average and autoregressive representations, respectively,

$$y_t = \psi(L)\varepsilon_t \qquad \text{and} \qquad \pi(L)y_t = \varepsilon_t \ ,$$

where

$$\psi(L) \equiv \sum_{j=0}^{\infty} \psi_j L^j \qquad \text{and} \qquad \pi(L) \equiv \sum_{j=0}^{\infty} \pi_j L^j \ .$$

If we assume $y_0 = y_{-1} = y_{-2} = \cdots = 0$, then these infinite-series representations will be truncated as follows:

$$y_t = \psi^*(L)\varepsilon_t \qquad \text{and} \qquad \pi(L)^* y_t = \varepsilon_t \ ,$$

where

$$\psi^*(L) \equiv \sum_{j=0}^{t} \psi_j L^j \qquad \text{and} \qquad \pi^*(L) \equiv \sum_{j=0}^{t} \pi_j L^j \ ,$$

since $\pi^*(L)\psi^*(L) = 1$, which is the truncated version of the fact $\pi(L)\psi(L) = 1$. Note that $\pi^*(L)\psi^*(L) = 1$ is due to the equalities $\sum_{j=0}^{t} \pi_j \psi_{t-j} = 0$, for all $t > 0$, which are in turn due to the equality $\pi(L)\psi(L) = 1$. Furthermore, if we compare the derivatives of ε_t with respect to d that are based on the three alternative forms $(1-B)^d y_t$, $\pi(B)y_t$ and $\psi^{-1}(B)y_t$ of ε_t, we find

$$\sum_{k=0}^{j} \frac{\partial \psi_k}{\partial d} \pi_{j-k} = \frac{1}{j} = -\sum_{k=0}^{j} \frac{\partial \pi_k}{\partial d} \psi_{j-k}\,, \qquad \text{for} \qquad j = 1, 2, \ldots \tag{A1}$$

The computation of the CSS estimator of d is based on minimizing the CSS function $S(d)$ or, equivalently, by solving the first-order condition

$$\frac{\partial S(d)}{\partial d} = \frac{1}{\sigma^2} \sum_{t=1}^{T} \varepsilon_t \frac{\partial \varepsilon_t}{\partial d} = 0\ . \tag{A2}$$

With the initial conditions $y_0 = y_{-1} = y_{-2} = \cdots = 0$, the derivative of ε_t with respect to d in (A2) is

$$\frac{\partial \varepsilon_t}{\partial d} = \frac{\partial \pi^*(L)}{\partial d} y_t = \frac{\partial \pi^*(L)}{\partial d} \psi^*(L)\varepsilon_t \equiv \tau(L)\varepsilon_t\ ,$$

and, by (A1), we have

$$\tau(L) \equiv \frac{\partial \pi^*(L)}{\partial d} \psi^*(L) = \sum_{j=1}^{t-1} \sum_{k=1}^{j} \frac{\partial \pi_k}{\partial d} \psi_{j-k} L^j + \sum_{j=t}^{2t-2} \delta_{jt} L^j$$

$$= -\sum_{j=1}^{t-1} \frac{1}{j} L^j + \sum_{j=t}^{2t-2} \delta_{jt} L^j\ ,$$

where

$$\delta_{jt} \equiv \sum_{k=j-(t-1)}^{t-1} \frac{\partial \pi_k}{\partial d} \psi_{j-k}\,, \qquad \text{for} \qquad j = t, \ldots, 2(t-1)\ .$$

Now, given that the expectation of the first-order condition (A1) is zero:

$$E\left[\frac{\partial S(d)}{\partial d}\right] = \frac{1}{\sigma^2} E\left[\sum_{t=1}^{T} \varepsilon_t \frac{\partial \varepsilon_t}{\partial d}\right] = \frac{1}{\sigma^2} \sum_{t=1}^{T} E[\varepsilon_t \tau(B)\varepsilon_t] = 0\ .$$

then the CSS estimator $\hat{d}$ converges almost surely to the true value of d under standard regularity conditions. We now show the asymptotic variance of $\hat{d}$ is the same as that of the MLE, which is $6/\pi^2$. Consider the expectation of the second-order derivative of $S(d)$:

$$E\left[\frac{\partial^2 S(d)}{\partial d^2}\right] = \frac{1}{\sigma^2} E\left[\sum_{t=1}^{T} \left(\frac{\partial \varepsilon_t}{\partial d}\right)^2 + \sum_{t=1}^{T} \varepsilon_t \frac{\partial^2 \varepsilon_t}{\partial d^2}\right] = \frac{1}{\sigma^2} \sum_{t=1}^{T} E\left(\frac{\partial \varepsilon_t}{\partial d}\right)^2$$

$$= \frac{1}{\sigma^2} \sum_{t=1}^{T} E[\tau(B)\varepsilon_t]^2 = \frac{1}{\sigma^2} \sum_{t=1}^{T} E\left[-\sum_{j=1}^{t-1} \frac{1}{j} B^j \varepsilon_t + \sum_{j=t}^{2t-2} \delta_{jt} B^j \varepsilon_t\right]^2$$

$$= \frac{1}{\sigma^2} \sum_{t=1}^{T} \left[\sum_{j=1}^{t-1} \frac{1}{j^2} E(\varepsilon_{t-j}^2) + \sum_{j=t}^{2t-2} \delta_{jt}^2 E(\varepsilon_{t-j}^2) \right]$$

$$= \sum_{t=1}^{T} \left[\sum_{j=1}^{t-1} \frac{1}{j^2} + \sum_{j=t}^{2t-2} \delta_{jt}^2 \right] = \sum_{t=1}^{T} (T - t)\frac{1}{t^2} + o(T) \ .$$

So $\lim_{T\to\infty} \frac{1}{T} E\left[\frac{\partial^2 S(d)}{\partial d^2}\right] = \sum_{t=1}^{\infty} \frac{1}{t^2} = \frac{\pi^2}{6}$, which is the inverse of the asymptotic variance of $\hat{d}$.

References

Agiakloglou, Newbold P, Wohar M (1992) Bias in an estimator of the fractional differencing parameter. Journal of Time Series Analysis, forthcoming

Baillie RT, Chung CF, Tieslau MA (1992) The long memory and variability of inflation: A reappraisal of the friedman hypothesis. Michigan State University Working Paper 9102

Box GEP, Jenkins GM (1976) Time series analysis: forecasting and control, second edition, Holden-Day San Francisco

Cheung YW, Diebold FX (1993) On maximum-likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. Journal of Econometrics, forthcoming

Diebold FX, Rudebusch GD (1989) Long memory and persistence in aggregate output. Journal of Monetary Economics 24:189–209

Diebold FX, Rudebusch GD (1991) Is consumption too smooth? Long memory and the deaton paradox. The Review of Economics and Statistics 73:1–9

Engel RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. Econometrica 50:987–1008

Fox R, Taqqu MS (1986) Large sample properties of parameter estimates for strongly dependent stationary gaussian time series. Annals of Statistics 14:517–532

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

Granger CWJ (1980) Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14:227–238

Granger CWJ (1981) Some properties of time series data and their use in econometric model specification. Journal of Econometrics 16:121–130

Granger CWJ, Joyeux R (1980) An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis 1:15–39

Hosking JRM (1981) Fractional differencing. Biometrika 68:165–176

Hosking JRM (1984). Modeling persistence in hydrological time series using fractional differencing. Water Resources Research 20:1898–1908

Li WK, McLeod AI (1986) Fractional time series modeling. Biometrika 73:217–221

Newbold P (1974) The exact likelihood function for a mixed autoregressive-moving average process. Biometrika 61:423–426

Phillips PCB (1987) Time series regression with a unit root. Econometrica 55:277–301

Pierce DA (1971) Least squares estimation in the regression model with autoregressive-moving average errors. Biometrika 58:299–312

Samarov A, Taqqu MS (1988) On the efficiency of the sample mean in long-memory noise. Journal of Time Series Analysis 9:191–200

Shea GS (1989) Ex-post rational price approximations and the empirical reliability of the present-value relation. Journal of Applied Econometrics 4:139–159

Shea GS (1991) Uncertainty and implied variance bounds in long-memory models of the interest rate term structure. Empirical Economics 16:287–312

Sowell FB (1992a) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics 53:165–188

Sowell FB (1992b) Modelling long-run behavior with the fractional ARIMA model. Journal of Monetary Economics 29:277–302

Taqqu MS (1975) Weak convergence to fractional brownian motion and to the Rosenblatt process. Zeitschrift Wahrscheinlichkeitstheorie verw. Gebiete 31:287–302

Yajima Y (1988) On estimation of A regression model with long-memory stationary errors. Annals of Statistics 16:791–807