Journal of Econometrics 62 (1994) 301–316. North-Holland

# On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean*

Yin-Wong Cheung

*University of California, Santa Cruz, CA 95064, USA*

Francis X. Diebold

*University of Pennsylvania, Philadelphia, PA 19104, USA*

Received March 1991, final version received April 1993

There are two approaches to maximum likelihood (ML) estimation of the parameter of fractionally-integrated noise: approximate frequency-domain ML [Fox and Taqqu (1986)] and exact time-domain ML [Sowell (1992b)]. If the mean of the process is *known*, then a clear finite-sample mean-squared error ranking of the estimators emerges: the exact time-domain estimator is superior. We show in this paper, however, that the finite-sample efficiency of approximate frequency-domain ML relative to exact time-domain ML rises dramatically when the mean is unknown and so must be estimated. The intuition for our result is straightforward: the frequency-domain ML estimator is invariant to the true but unknown mean of the process, while the time-domain ML estimator is not. Feasible time-domain estimation must therefore be based upon de-meaned data, but the long memory associated with fractional integration makes precise estimation of the mean difficult. We conclude that the frequency-domain estimator is an attractive and efficient alternative for situations in which large sample sizes render time-domain estimation impractical.

*Key words*: Fractional differencing; Maximum likelihood estimation; Frequency domain estimation; Simulation

*JEL classification*: C15; C22

*Correspondence to*: Francis X. Diebold, Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297, USA.

* We are grateful for the constructive comments of two referees and a co-editor. Discussions with Richard Baillie, Larry Christiano, Christian Gourieroux, Roberto Mariano, Marc Nerlove, Peter Robinson, and Fallaw Sowell were extremely valuable. Useful comments were also received from participants at the 1990 NSF/NBER Time Series Conference at the University of California, San Diego, the 1990 North American Winter Econometric Society Meetings, and from seminar participants at New York University, the University of British Columbia, and the University of Southern California. All remaining errors and shortcomings are ours alone. Cheung gratefully acknowledges support from the Faculty Research Fund and the GICES, U.C. Santa Cruz. Diebold gratefully acknowledges support from the National Science Foundation, the Sloan Foundation, the University of Pennsylvania Research Foundation, and the Institute for Empirical Macroeconomics.

## 1. Introduction

The literature on long-memory time series processes and, in particular, autoregressive fractionally-integrated moving average (ARFIMA) processes has grown rapidly since the early contributions of Granger and Joyeux (1980), Hosking (1981), and Geweke and Porter-Hudak (1983). Recent theoretical work includes Li and McLeod (1986), Fox and Taqqu (1986), Robinson (1988, 1991), Sowell (1990, 1992b), Gourieroux, Maurel, and Monfort (1987), Yajima (1985, 1988), Dahlhaus (1989), and Cheng and Robinson (1992), among others.[1]

The theory has been used in applied econometric work, in which flexible characterization of long-run, or low-frequency, dynamics is often of crucial importance. Examples include real output dynamics and the unit-root hypothesis [Diebold and Rudebusch (1989), Sowell (1992a)], disposable income dynamics and the permanent-income hypothesis [Diebold and Rudebusch (1991)], predictability of stock returns and the efficient-markets hypothesis [Lo (1991)], variance bounds for the interest-rate term structure and Hicks' expectations hypothesis [Shea (1991)], real exchange rate dynamics and the purchasing-power-parity hypothesis [Diebold, Husted, and Rush (1991), Cheung and Lai (1993)], real wage dynamics and the intertemporal-substitution hypothesis [Hassett (1990)], and nominal exchange rate dynamics [Cheung (1993)].

Most such applied work, however, makes use of estimation procedures whose properties are incompletely understood and are likely to be suboptimal relative to maximum likelihood (ML) estimation under correct model specification.[2] Hence the interest in recent work on exact time-domain and approximate frequency-domain ML estimation of fractionally-integrated models. In particular, Fox and Taqqu (1986) construct an asymptotic approximation to the likelihood of an ARFIMA process in the frequency domain, and Sowell (1992b) constructs the exact likelihood function of an ARFIMA process in the time domain.

Although the two ML estimators are asymptotically equivalent, their finite-sample properties differ. Monte Carlo analyses, in particular Sowell (1992b), have shown that the time-domain ML estimator is substantially more efficient than the frequency-domain ML estimator, when the mean of the process is known. Thus, in spite of the fact that time-domain ML is more tedious than frequency-domain ML [due to the $(T\times T)$ covariance matrix that must be inverted at each evaluation of the likelihood function], it appears to be an attractive estimator.[3]

In practice, of course, the mean is not known, so that existing Monte Carlo results for time-domain ML correspond to an infeasible estimator. In applied

[1] Robinson (1990) provides an insightful survey.

[2] The leading such estimator was proposed in 1983 by Geweke and Porter-Hudak (GPH).

[3] Here and throughout $T$ denotes sample size.

work, a feasible estimator is used, obtained by replacing the unknown population mean by an estimate; that is, the time-domain ML procedure is applied to de-meaned data. What are the properties of this feasible time-domain ML estimator, and how does it compare to the frequency-domain ML estimator? This is the relevant question for applied work, and it is the subject of this paper. We shall motivate it in light of some important differences underlying the construction of the time- and frequency-domain ML estimators, and we shall provide answers.

The paper proceeds as follows. In section 2 we discuss the details of our Monte Carlo experiment, in which we contrast the efficiency of the frequency-domain ML estimator to that of the time-domain ML estimator (with the population mean assumed known and, alternatively, with the sample mean removed prior to analysis). In section 3 we report the results of the Monte Carlo analysis; the efficiency of frequency-domain ML relative to time-domain ML with estimated mean is strikingly different from that with known mean. In section 4 we offer additional discussion. Section 5 concludes.

## 2. The Monte Carlo experiment

We work with the stationary and invertible pure fractionally-integrated process

$$X_t = \mu + (1 - B)^{-d} e_t, \tag{1}$$

$$e_t \sim \text{iid}\,\mathrm{N}(0, 1), \tag{2}$$

for $t = 1, 2, \ldots, T$, where $B$ is the backshift operator, $-\frac{1}{2} < d < \frac{1}{2}$, and $\mu < \infty \in \mathbb{R}$.

We first consider time-domain ML estimation. Assume first that the mean of the process is known and, without loss of generality, assume that it is zero. Under the normality assumption, construction of the likelihood simply amounts to expressing the autocovariances of the process in terms of the underlying parameters (in this case, $d$). Evaluation of the likelihood requires inversion of the $(T \times T)$ Toeplitz covariance matrix, $\Sigma(d)$, with $ij$th entry,

$$\gamma_x(|i-j|) = (-1)^{|i-j|} \frac{\Gamma(1-2d)}{\Gamma(1-d+|i-j|)\Gamma(1-d-|i-j|)},$$

where $1/\Gamma(d) = 0$ when $d$ is a nonpositive integer.[4] The first estimator we explore is precisely the one that maximizes this likelihood with the mean

[4] This expression for the covariance matrix may be traced at least to Adenstedt (1974) for the ARFIMA $(0, d, 0)$ case. It is extended by Sowell (1992b) to the ARFIMA $(p, d, q)$ case.

$\mu$ assumed known to be zero (and hence not estimated).[5] The estimator is denoted $ML1$. Formally,

$$ML1 = \underset{d}{\operatorname{argmax}}\, L(d; z),$$

where $L(\cdot)$ denotes the Gaussian likelihood function,

$$L(d; z) = (2\pi)^{-T/2} |\Sigma(d)|^{-1/2} \mathrm{e}^{(-1/2) z' \Sigma(d)^{-1} z},$$

and $z = (x_1 - \mu, x_2 - \mu, \ldots, x_T - \mu)'$. $ML1$ is, of course, not feasible in practice, because $\mu$ is never known; however, it will serve as a useful benchmark.

The obvious feasible counterpart to $ML1$ is obtained by first removing the sample mean from the data,

$$ML1a = \underset{d}{\operatorname{argmax}}\, L(d; \bar{z}),$$

where $L(\cdot)$ denotes the Gaussian likelihood function and $\bar{z} = (x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_T - \bar{x})'$. As long as $\bar{x}$ is consistent for $\mu$, the feasible estimator will perform satisfactorily in large samples.[6] The long memory associated with fractional integration with $d > 0$ may produce slow rates of convergence of $\bar{x}$ to $\mu$, however, leading to poor performance of $ML1a$ in samples of the size typically available in economics.

Now we consider frequency-domain ML estimation.[7] Following Fox and Taqqu (1986), we exploit the fact that maximization of the Gaussian likelihood is asymptotically equivalent to minimization of

$$\sum_{j=1}^{T-1} \frac{I_x(2\pi j/T)}{f_x(2\pi j/T; d)},$$

with respect to $d$, where $I_x(\lambda)$ is the periodogram of $X$ at frequency $\lambda$ given by

$$\frac{1}{T} \left| \sum_{t=1}^{T} x_t \mathrm{e}^{-i\lambda t} \right|^2,$$

[5] The FORTRAN code used to evaluate and maximize the likelihood was generously supplied by Fallaw Sowell.

[6] See Dahlhaus (1989) for a formal discussion.

[7] The frequency-domain procedure, which builds upon an important result of Whittle (1951), has received substantial attention in the estimation of short-memory ARMA and unobserved-components models. See, for example, Nerlove, Grether, and Carvalho (1979) and Harvey (1989).

and

$$f_x(\lambda, d) = |1 - e^{-i\lambda}|^{-2d}$$

is proportional to the spectral density of $X$ at frequency $\lambda$. We call the resulting estimator *ML2*; formally,

$$ML2 = \underset{d}{\operatorname{argmin}} \sum_{j=1}^{T-1} \frac{I_x(2\pi j/T)}{f_x(2\pi j/2; d)}.$$

Note that *ML2* is invariant to $\mu$, because the periodogram ordinate at frequency zero is not used.[8]

Some authors, such as Dahlhaus (1988), have argued that tapering may improve the finite-sample properties of the frequency-domain estimator; therefore, we also explore the properties of two frequency-domain ML estimators that make use of trapezoidally tapered data. The first is

$$ML2a = \underset{d}{\operatorname{argmin}} \sum_{j=1}^{T-1} \frac{I_{Xa}(2\pi j/T)}{f_x(2\pi j/T; d)}.$$

In this expression, $I_{Xa}(\lambda)$ is the periodogram of $Xa$ at frequency $\lambda$ and $xa_t = k_t x_t$, with

$$k_t = \begin{cases} t/a, & 1 \leqslant t \leqslant a, \\ 1, & a+1 \leqslant t \leqslant T-a, \\ (T+1-t)/a, & T+1-a \leqslant t \leqslant T, \end{cases}$$

and $a = [0.1T]$.

The second taper is identical, except that 25 percent of each end of the sample is tapered, rather than 10 percent. The estimator is

$$ML2b = \underset{d}{\operatorname{argmin}} \sum_{j=1}^{T-1} \frac{I_{Xb}(2\pi j/T)}{f_x(2\pi j/T; d)}.$$

$I_{Xb}(\lambda)$ is the periodogram of $Xb$ at frequency $\lambda$ and $xb_t = k_t x_t$, with

$$k_t = \begin{cases} t/b, & 1 \leqslant t \leqslant b, \\ 1, & b+1 \leqslant t \leqslant T-b, \\ (T+1-t)/b, & T+1-b \leqslant t \leqslant T, \end{cases}$$

and $b = [0.25T]$.

[8] It is well known [e.g., Priestley (1980, p. 417)] that only the zero-frequency periodogram ordinate depends on the mean.

Ten points in the parameter space are explored, corresponding to $d = \pm 0.050$, $\pm 0.150$, $\pm 0.250$, $\pm 0.350$, $\pm 0.450$. Sample sizes of $T = 50, 100, 300, 500$ are explored. Realizations of the white noise sequence (2) for each $T$ are generated by the IMSL subroutine DRNNOA, and then the corresponding realizations of the fractional noise (1) for each $\{d, T\}$ configuration are generated by multiplying the vectors of N(0, 1) deviates by the Choleski factor of the covariance matrix of $X$, as in Geweke and Porter-Hudak (1983). The sampling properties of the various time- and frequency-domain ML estimators are then explored. All likelihood maximizations are done with the Davidon–Fletcher–Powell algorithm as implemented in GQOPT, using the true value of $d$ to start the iterations. For each $\{d, T$, estimator$\}$ configuration, $N = 1000$ Monte Carlo replications are performed, and the mean-squared error (MSE) and bias across the replications are computed.

## 3. Results

### *3.1. Mean-squared error*

The finite-sample estimated MSEs of the various estimators are reported in table 1. As expected, for fixed $d$, the MSE of each estimator decreases as $T$ increases. Moreover, the relative efficiency MSE $(ML1a)$/MSE$(ML2)$ tends to be less than one, reflecting the somewhat better performance of *ML1a*. Due to the asymptotic equivalence of the two estimators, however, relative efficiency tends to increase as $T$ increases, approaching 1.0 in the limit. Importantly, the convergence of the relative efficiency to 1.0 is fast; the performance of the two estimators is nearly identical for $T > 150$.

A number of additional important results are apparent for fixed $T$ and $d$. First, time-domain ML distinctly dominates frequency-domain ML when the mean is known. Second, time-domain ML outperforms frequency-domain ML by a much smaller margin when the mean is estimated. Particularly for $d > 0$ and $T \geqslant 150$, there is little performance difference between the two estimators. Third, although neither taper consistently produces a reduction in the MSE of the frequency-domain estimator, both tapers reduce MSE for the more persistent parameterizations.[9] Of the two, *ML2a* tends to perform somewhat better.

Now consider the effects of varying $d$. First, for each sample size, the MSE of *ML1* decreases as $d$ increases. Evidently the reduced variance of that estimator, which is obtained through the greater unconditional variation induced by higher $d$ values, more than offsets the slightly increased squared bias associated

[9] This finding accords with Dahlhaus (1988).

Table 1

Mean-squared error.[a]

| $d$ | $ML1$ | $ML1a$ | $ML2$ | $ML2a$ | $ML2b$ |
|---|---|---|---|---|---|
| | | | $T = 50$ | | |
| − 0.45 | 0.0175 | 0.0191 | 0.0294 | 0.0320 | 0.0352 |
| − 0.35 | 0.0190 | 0.0235 | 0.0310 | 0.0329 | 0.0370 |
| − 0.25 | 0.0173 | 0.0246 | 0.0333 | 0.0337 | 0.0386 |
| − 0.15 | 0.0163 | 0.0241 | 0.0329 | 0.0354 | 0.0389 |
| − 0.05 | 0.0169 | 0.0278 | 0.0338 | 0.0340 | 0.0388 |
| 0.05 | 0.0156 | 0.0280 | 0.0335 | 0.0342 | 0.0380 |
| 0.15 | 0.0127 | 0.0274 | 0.0342 | 0.0353 | 0.0378 |
| 0.25 | 0.0128 | 0.0296 | 0.0343 | 0.0344 | 0.0362 |
| 0.35 | 0.0102 | 0.0297 | 0.0311 | 0.0295 | 0.0316 |
| 0.45 | 0.0079 | 0.0302 | 0.0307 | 0.0254 | 0.0287 |
| | | | $T = 100$ | | |
| − 0.45 | 0.0078 | 0.0086 | 0.0107 | 0.0114 | 0.0130 |
| − 0.35 | 0.0080 | 0.0094 | 0.0114 | 0.0126 | 0.0149 |
| − 0.25 | 0.0075 | 0.0092 | 0.0106 | 0.0117 | 0.0143 |
| − 0.15 | 0.0074 | 0.0104 | 0.0117 | 0.0135 | 0.0159 |
| − 0.05 | 0.0074 | 0.0102 | 0.0112 | 0.0117 | 0.0138 |
| 0.05 | 0.0076 | 0.0110 | 0.0121 | 0.0133 | 0.0152 |
| 0.15 | 0.0062 | 0.0103 | 0.0111 | 0.0118 | 0.0139 |
| 0.25 | 0.0060 | 0.0105 | 0.0112 | 0.0113 | 0.0125 |
| 0.35 | 0.0049 | 0.0103 | 0.0109 | 0.0104 | 0.0110 |
| 0.45 | 0.0035 | 0.0114 | 0.0113 | 0.0102 | 0.0120 |
| | | | $T = 300$ | | |
| − 0.45 | 0.0024 | 0.0023 | 0.0027 | 0.0030 | 0.0036 |
| − 0.35 | 0.0026 | 0.0024 | 0.0027 | 0.0032 | 0.0036 |
| − 0.25 | 0.0022 | 0.0025 | 0.0026 | 0.0032 | 0.0035 |
| − 0.15 | 0.0024 | 0.0024 | 0.0025 | 0.0029 | 0.0032 |
| − 0.05 | 0.0021 | 0.0028 | 0.0029 | 0.0030 | 0.0040 |
| 0.05 | 0.0022 | 0.0027 | 0.0027 | 0.0030 | 0.0036 |
| 0.15 | 0.0021 | 0.0026 | 0.0027 | 0.0029 | 0.0035 |
| 0.25 | 0.0020 | 0.0026 | 0.0027 | 0.0029 | 0.0034 |
| 0.35 | 0.0020 | 0.0027 | 0.0028 | 0.0029 | 0.0034 |
| 0.45 | 0.0014 | 0.0028 | 0.0026 | 0.0021 | 0.0023 |
| | | | $T = 500$ | | |
| − 0.45 | 0.0013 | 0.0014 | 0.0015 | 0.0014 | 0.0018 |
| − 0.35 | 0.0013 | 0.0015 | 0.0016 | 0.0017 | 0.0020 |
| − 0.25 | 0.0013 | 0.0015 | 0.0016 | 0.0015 | 0.0020 |
| − 0.15 | 0.0012 | 0.0015 | 0.0015 | 0.0017 | 0.0020 |
| − 0.05 | 0.0013 | 0.0016 | 0.0015 | 0.0016 | 0.0020 |
| 0.05 | 0.0014 | 0.0016 | 0.0016 | 0.0021 | 0.0023 |
| 0.15 | 0.0013 | 0.0014 | 0.0015 | 0.0017 | 0.0022 |
| 0.25 | 0.0012 | 0.0015 | 0.0015 | 0.0016 | 0.0018 |
| 0.35 | 0.0011 | 0.0014 | 0.0015 | 0.0016 | 0.0019 |
| 0.45 | 0.0008 | 0.0013 | 0.0012 | 0.0012 | 0.0014 |

[a] $d$ = fractional differencing parameter; $T$ = sample size; $ML1$ = time-domain maximum likelihood, true mean removed; $ML1a$ = time-domain maximum likelihood, arithmetic mean removed; $ML2$ = frequency-domain maximum likelihood; $ML2a$ = frequency-domain maximum likelihood, taper $a$; $ML2b$ = frequency-domain maximum likelihood, taper $b$.

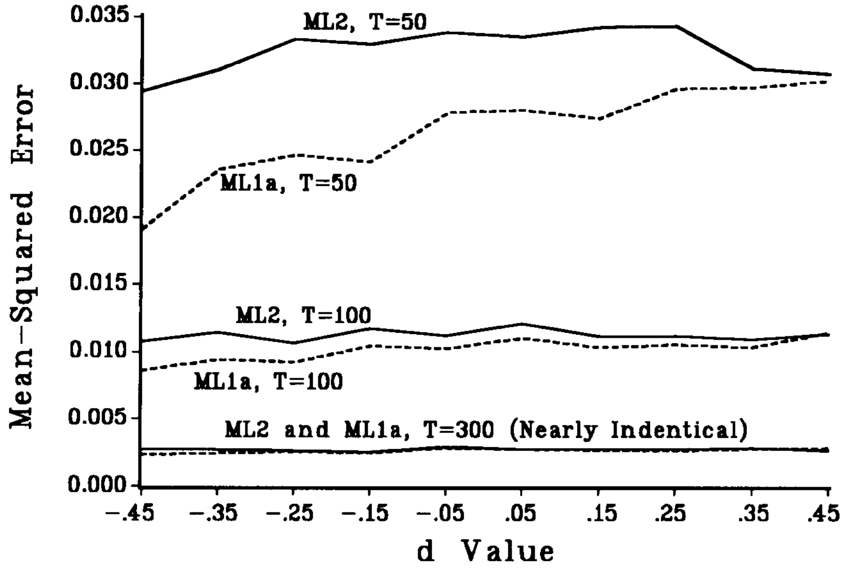


Fig. 1. Mean-squared error as a function of $d$ and $T$.
$d$ = fractional differencing parameter; $T$ = sample size; *ML1a* = time-domain maximum likelihood, arithmetic mean removed; *ML2* = frequency-domain maximum likelihood.

with higher $d$ values.[10] Second, for each sample size, the MSE of *ML1a* tends to *increase* as $d$ increases. Here, the reduced variance is not enough to offset the large bias increases associated with higher $d$ when the mean is estimated. Third, the MSEs of the frequency-domain estimators are not particularly sensitive to $d$.

A key insight is that the efficiency of *ML2* relative to *ML1a increases* as $d$ increases; the more persistent the process, the better the relative performance of frequency-domain ML. This result may be understood by recalling a result of Adenstedt (1974) and Sowell (1990), who show that for the pure fractionally-integrated process (1)–(2) $\mathrm{var}(\sum_{t=1}^{T} x_t) = O(T^{1+2d})$. Thus, $T^{-1/2-d}\sum_{t=1}^{T} x_t$ has a stable limiting distribution, so that $T^{1/2-d}\bar{x}$ has a stable limiting distribution; that is, the convergence rate of the sample mean depends inversely on $d$. When $d = 0$, the usual $\sqrt{T}$ consistency obtains; convergence is faster or slower than $\sqrt{T}$ as $d$ is less than or greater than zero. The larger $d$ is, the more slowly the sample mean converges and the poorer the performance of feasible time-domain ML.

The MSE comparison between *ML1a* and *ML2* is highlighted in fig. 1, which shows the MSEs of those estimators for a range of $d$ values and for $T$ = 50, 100, 300, and which makes the points raised in this section visually apparent.

[10] Bias will be discussed shortly in section 3.2.

### 3.2. Bias

The squared-error loss with which we evaluate the estimators may be decomposed into the sum of squared bias and variance, so that high bias is acceptable if it is sufficiently offset by low variance. Nevertheless, it may be of interest to examine bias in isolation. For this reason, we report the finite-sample biases of the various estimators in table 2. Three results emerge. First, bias is almost always negative; that is, $d$ tends to be underestimated. For fixed $d$, the bias of each estimator decreases as $T$ increases, as expected.

Second, for fixed $d$ and $T$, the absolute bias of *ML1* is generally smallest, followed by *ML1a*, which often has a smaller absolute bias than any of the frequency-domain estimators *ML2*, *ML2a*, and *ML2b* for small $T$ and small $d$. For larger $T$ and/or $d$, however, the frequency-domain estimators typically have smaller absolute bias than *ML1a*. In particular, the data tapers underlying *ML2a* and *ML2b* provide substantial bias reduction for the more persistent parameterizations (large positive $d$ values).[11]

Third, for fixed $T$, the absolute bias of *ML1* tends to increase slightly as $d$ increases, while the bias of *ML1a* tends to increase *sharply* as $d$ increases. This positive relationship between $d$ and bias is one manifestation of the negative relationship between $d$ and the convergence speed of $\bar{x}$. The biases of the other estimators show no clear relationship to $d$.

### 3.3. On the magnitude of Monte Carlo error

Approximate assessment of the Monte Carlo uncertainty associated with our estimates is straightforward. By virtue of the fact that

$$\sqrt{T}(\hat{d}_i - d) \xrightarrow{d} \mathrm{N}(0, 6/\pi^2), \qquad i = 1, 2, \ldots, N, \tag{3}$$

for all of the estimators considered in this paper, and by virtue of the independence of our $d$ estimates across Monte Carlo replications, it follows that, for large $T$ and $N$, $(1/N)\sum_{i=1}^{N}(\hat{d}_i - d)^2$ is approximately normal with variance $72/(NT^2\pi^4)$. Thus, as $N = 1000$ in all experiments, approximate standard errors for the MSEs are given by $\pm$ 0.00054 ($T = 50$), $\pm$ 0.00027 ($T = 100$), $\pm$ 0.00009 ($T = 300$), $\pm$ 0.00005 ($T = 500$), so that the Monte Carlo uncertainty appears quite small. Similar computations for the bias yield a variance expression of $6/(NT\pi^2)$, yielding approximate standard errors for the biases given by $\pm$ 0.00349 ($T = 50$), $\pm$ 0.00247 ($T = 100$), $\pm$ 0.00142 ($T = 300$), $\pm$ 0.00110 ($T = 500$), which is again small.

[11] Again, this finding accords with Dahlhaus (1988).

Table 2
Bias.[a]

| *d* | *ML1* | *ML1a* | *ML2* | *ML2a* | *ML2b* |
|---|---|---|---|---|---|
| | | | $T = 50$ | | |
| − 0.45 | − 0.0193 | − 0.0389 | − 0.0781 | − 0.0933 | − 0.0997 |
| − 0.35 | − 0.0291 | − 0.0630 | − 0.0846 | − 0.0932 | − 0.0983 |
| − 0.25 | − 0.0206 | − 0.0695 | − 0.0991 | − 0.1016 | − 0.1081 |
| − 0.15 | − 0.0237 | − 0.0740 | − 0.0979 | − 0.1002 | − 0.1044 |
| − 0.05 | − 0.0329 | − 0.0893 | − 0.1022 | − 0.1014 | − 0.1035 |
| 0.05 | − 0.0283 | − 0.0950 | − 0.1030 | − 0.1007 | 0.0989 |
| 0.15 | − 0.0213 | − 0.0916 | − 0.0972 | 0.0979 | 0.0971 |
| 0.25 | − 0.0318 | − 0.1068 | − 0.1031 | − 0.0966 | − 0.0906 |
| 0.35 | − 0.0365 | − 0.1182 | − 0.0929 | − 0.0796 | 0.0664 |
| 0.45 | − 0.0433 | − 0.1339 | − 0.0886 | − 0.0473 | 0.0121 |
| | | | $T = 100$ | | |
| − 0.45 | − 0.0134 | − 0.0230 | − 0.0349 | − 0.0474 | − 0.0499 |
| − 0.35 | − 0.0156 | − 0.0331 | − 0.0437 | − 0.0529 | − 0.0570 |
| − 0.25 | − 0.0072 | − 0.0308 | − 0.0392 | − 0.0436 | − 0.0468 |
| − 0.15 | − 0.0165 | − 0.0451 | − 0.0524 | − 0.0555 | − 0.0583 |
| − 0.05 | − 0.0091 | − 0.0410 | − 0.0472 | − 0.0480 | − 0.0494 |
| 0.05 | − 0.0150 | − 0.0494 | − 0.0530 | − 0.0543 | − 0.0540 |
| 0.15 | − 0.0143 | − 0.0503 | − 0.0502 | − 0.0490 | − 0.0504 |
| 0.25 | − 0.0137 | − 0.0528 | − 0.0471 | − 0.0437 | − 0.0418 |
| 0.35 | − 0.0188 | − 0.0594 | − 0.0435 | − 0.0353 | − 0.0299 |
| 0.45 | − 0.0271 | − 0.0789 | − 0.0412 | − 0.0119 | 0.0049 |
| | | | $T = 300$ | | |
| − 0.45 | − 0.0028 | − 0.0053 | − 0.0057 | − 0.0178 | − 0.0185 |
| − 0.35 | − 0.0044 | − 0.0110 | − 0.0119 | − 0.0183 | − 0.0184 |
| − 0.25 | − 0.0036 | − 0.0136 | − 0.0148 | − 0.0172 | − 0.0205 |
| − 0.15 | − 0.0029 | − 0.0143 | − 0.0160 | − 0.0182 | − 0.0141 |
| − 0.05 | − 0.0017 | − 0.0167 | − 0.0177 | − 0.0168 | − 0.0207 |
| 0.05 | − 0.0048 | − 0.0173 | − 0.0184 | − 0.0195 | − 0.0204 |
| 0.15 | − 0.0041 | − 0.0174 | − 0.0170 | − 0.0166 | − 0.0182 |
| 0.25 | − 0.0041 | − 0.0186 | − 0.0156 | − 0.0141 | − 0.0136 |
| 0.35 | − 0.0064 | − 0.0224 | − 0.0147 | − 0.0121 | − 0.0101 |
| 0.45 | − 0.0133 | − 0.0349 | − 0.0253 | − 0.0088 | 0.0082 |
| | | | $T = 500$ | | |
| − 0.45 | − 0.0036 | − 0.0048 | − 0.0041 | − 0.0104 | − 0.0104 |
| − 0.35 | − 0.0016 | − 0.0075 | − 0.0070 | − 0.0100 | − 0.0126 |
| − 0.25 | − 0.0033 | − 0.0091 | − 0.0092 | − 0.0111 | − 0.0120 |
| − 0.15 | − 0.0022 | − 0.0100 | − 0.0107 | − 0.0139 | − 0.0103 |
| − 0.05 | − 0.0030 | − 0.0122 | − 0.0121 | − 0.0120 | − 0.0117 |
| 0.05 | − 0.0019 | − 0.0095 | − 0.0135 | − 0.0186 | − 0.0193 |
| 0.15 | − 0.0013 | − 0.0094 | − 0.0116 | − 0.0090 | − 0.0108 |
| 0.25 | − 0.0020 | − 0.0119 | − 0.0097 | − 0.0090 | − 0.0086 |
| 0.35 | − 0.0018 | − 0.0113 | − 0.0057 | − 0.0054 | − 0.0057 |
| 0.45 | − 0.0056 | − 0.0195 | − 0.0105 | − 0.0018 | − 0.0010 |

[a] $d$ = fractional differencing parameter; $T$ = sample size; *ML1* = time-domain maximum likelihood, true mean removed; *ML1a* = time-domain maximum likelihood, arithmetic mean removed; *ML2* = frequency-domain maximum likelihood; *ML2a* = frequency-domain maximum likelihood, taper *a*; *ML2b* = frequency-domain maximum likelihood, taper *b*. The bias is the mean estimate over the Monte Carlo replications, less the true value.

Alternatively, we can estimate the standard errors of the MSEs in a way that avoids the use of asymptotics, as follows. Let $m_i = (\hat{d}_i - d)^2$, $i = 1, 2, \ldots, 1000$, so that our reported MSEs are just the sample means, $\bar{m} = (1/1000)\sum_{i=1}^{1000} m_i$. Now, $\bar{m} \overset{a}{\sim} \mathrm{N}(m, \sigma^2/1000)$, where $m = \mathrm{E}(m_i)$ and $\sigma^2 = \mathrm{E}[(m_i - m)^2]$. The standard deviation of interest, $\sigma/\sqrt{1000}$, may therefore be estimated by

$$\frac{\hat{\sigma}}{\sqrt{1000}} = \frac{1}{1000}\sqrt{\sum_{i=1}^{1000} (m_i - \bar{m})^2}\,. \tag{4}$$

These standard errors are presented in table 3. In the case of *ML1*, the standard errors computed via (3) and (4) are in close agreement, as expected. For the more important cases of *ML1a* and *ML2*, there is greater disagreement, with the standard errors based on (4) being generally larger. Either way, however, it appears that the Monte Carlo variability associated with our MSE estimates is very small.

Identical arguments may be used to obtain a nonasymptotic standard error estimator for the bias. Let $b_i = (\hat{d}_i - d)$, $i = 1, 2, \ldots, 1000$, so that our reported biases are again the sample means, $\bar{b} = (1/1000)\sum_{i=1}^{1000} b_i$. Now, $\bar{b} \overset{a}{\sim} \mathrm{N}(b, \sigma^2/1000)$, where $b = \mathrm{E}(b_i)$ and $\sigma^2 = \mathrm{E}[(b_i - b)^2]$, so we take

$$\frac{\hat{\sigma}}{\sqrt{1000}} = \frac{1}{1000}\sqrt{\sum_{i=1}^{1000} (b_i - \bar{b})^2}\,. \tag{5}$$

These standard errors are presented in table 4. Parallel to the MSE case, the standard errors computed via (3) and (5) are in close agreement for *ML1*, with greater disagreement for the other estimators. In all cases, however, the Monte Carlo variability associated with our estimates is very small.

## 4. Additional discussion

Here we focus on certain aspects of the analysis that merit additional discussion, with some attention paid to directions for future research.

First, we intentionally neglect to include the GPH estimator in our Monte Carlo comparison. Such a comparison would be unfairly biased against the semi-parametric GPH estimator, because we work only under correct model specification. It might be desirable to study in future work the comparative properties of GPH, time-domain ML, and frequency-domain ML under model misspecification of various types. Such a study, however, would be very challenging in terms of experimental design.

Second, we intentionally focus on the case of pure fractional noise, that is, the ARFIMA$(0, d, 0)$ case. The pure fractional noise is of substantial interest in its

Table 3

Estimated standard error of mean-squared error.[a]

| $d$ | $ML1$ | $ML1a$ | $ML2$ | $ML2a$ | $ML2b$ |
|---|---|---|---|---|---|
| | | | $T = 50$ | | |
| − 0.45 | 0.00074 | 0.00079 | 0.00114 | 0.00125 | 0.00140 |
| − 0.35 | 0.00080 | 0.00096 | 0.00126 | 0.00130 | 0.00148 |
| − 0.25 | 0.00073 | 0.00099 | 0.00131 | 0.00131 | 0.00154 |
| − 0.15 | 0.00070 | 0.00098 | 0.00130 | 0.00138 | 0.00153 |
| − 0.05 | 0.00074 | 0.00107 | 0.00125 | 0.00129 | 0.00157 |
| 0.05 | 0.00070 | 0.00103 | 0.00132 | 0.00136 | 0.00159 |
| 0.15 | 0.00057 | 0.00108 | 0.00136 | 0.00136 | 0.00148 |
| 0.25 | 0.00058 | 0.00077 | 0.00136 | 0.00138 | 0.00150 |
| 0.35 | 0.00045 | 0.00056 | 0.00123 | 0.00128 | 0.00146 |
| 0.45 | 0.00032 | 0.00133 | 0.00125 | 0.00118 | 0.00134 |
| | | | $T = 100$ | | |
| − 0.45 | 0.00035 | 0.00037 | 0.00045 | 0.00045 | 0.00052 |
| − 0.35 | 0.00036 | 0.00040 | 0.00046 | 0.00051 | 0.00059 |
| − 0.25 | 0.00037 | 0.00040 | 0.00046 | 0.00050 | 0.00060 |
| − 0.15 | 0.00035 | 0.00043 | 0.00047 | 0.00055 | 0.00067 |
| − 0.05 | 0.00036 | 0.00043 | 0.00047 | 0.00049 | 0.00057 |
| 0.05 | 0.00033 | 0.00044 | 0.00048 | 0.00054 | 0.00066 |
| 0.15 | 0.00032 | 0.00043 | 0.00047 | 0.00049 | 0.00058 |
| 0.25 | 0.00026 | 0.00042 | 0.00048 | 0.00049 | 0.00057 |
| 0.35 | 0.00020 | 0.00036 | 0.00049 | 0.00046 | 0.00048 |
| 0.45 | 0.00014 | 0.00030 | 0.00046 | 0.00046 | 0.00050 |
| | | | $T = 300$ | | |
| − 0.45 | 0.00011 | 0.00010 | 0.00011 | 0.00013 | 0.00016 |
| − 0.35 | 0.00012 | 0.00011 | 0.00011 | 0.00014 | 0.00015 |
| − 0.25 | 0.00009 | 0.00011 | 0.00011 | 0.00014 | 0.00016 |
| − 0.15 | 0.00010 | 0.00010 | 0.00011 | 0.00012 | 0.00014 |
| − 0.05 | 0.00010 | 0.00013 | 0.00013 | 0.00013 | 0.00017 |
| 0.05 | 0.00010 | 0.00011 | 0.00012 | 0.00013 | 0.00016 |
| 0.15 | 0.00009 | 0.00012 | 0.00012 | 0.00013 | 0.00015 |
| 0.25 | 0.00009 | 0.00011 | 0.00012 | 0.00013 | 0.00015 |
| 0.35 | 0.00011 | 0.00012 | 0.00013 | 0.00013 | 0.00015 |
| 0.45 | 0.00005 | 0.00009 | 0.00008 | 0.00008 | 0.00010 |
| | | | $T = 500$ | | |
| − 0.45 | 0.00006 | 0.00006 | 0.00007 | 0.00006 | 0.00008 |
| − 0.35 | 0.00006 | 0.00007 | 0.00007 | 0.00007 | 0.00009 |
| − 0.25 | 0.00006 | 0.00007 | 0.00007 | 0.00006 | 0.00010 |
| − 0.15 | 0.00005 | 0.00006 | 0.00007 | 0.00007 | 0.00009 |
| − 0.05 | 0.00006 | 0.00006 | 0.00006 | 0.00007 | 0.00009 |
| 0.05 | 0.00006 | 0.00007 | 0.00007 | 0.00010 | 0.00011 |
| 0.15 | 0.00006 | 0.00006 | 0.00007 | 0.00009 | 0.00011 |
| 0.25 | 0.00006 | 0.00007 | 0.00007 | 0.00008 | 0.00009 |
| 0.35 | 0.00005 | 0.00006 | 0.00007 | 0.00007 | 0.00008 |
| 0.45 | 0.00004 | 0.00005 | 0.00004 | 0.00005 | 0.00007 |

[a] $d$ = fractional differencing parameter; $T$ = sample size; $ML1$ = time-domain maximum likelihood, true mean removed; $ML1a$ = time-domain maximum likelihood, arithmetic mean removed; $ML2$ = frequency-domain maximum likelihood; $ML2a$ = frequency-domain maximum likelihood, taper $a$; $ML2b$ = frequency-domain maximum likelihood, taper $b$.

Table 4

Estimated standard error of bias.[a]

| $d$ | $ML1$ | $ML1a$ | $ML2$ | $ML2a$ | $ML2b$ |
|---|---|---|---|---|---|
| | | | $T = 50$ | | |
| − 0.45 | 0.00414 | 0.00419 | 0.00483 | 0.00483 | 0.00503 |
| − 0.35 | 0.00426 | 0.00442 | 0.00488 | 0.00492 | 0.00523 |
| − 0.25 | 0.00411 | 0.00445 | 0.00485 | 0.00484 | 0.00519 |
| − 0.15 | 0.00397 | 0.00432 | 0.00483 | 0.00500 | 0.00529 |
| − 0.05 | 0.00398 | 0.00445 | 0.00483 | 0.00487 | 0.00530 |
| 0.05 | 0.00385 | 0.00436 | 0.00478 | 0.00491 | 0.00531 |
| 0.15 | 0.00350 | 0.00436 | 0.00498 | 0.00507 | 0.00533 |
| 0.25 | 0.00343 | 0.00427 | 0.00487 | 0.00501 | 0.00529 |
| 0.35 | 0.00298 | 0.00397 | 0.00474 | 0.00481 | 0.00521 |
| 0.45 | 0.00245 | 0.00350 | 0.00478 | 0.00481 | 0.00534 |
| | | | $T = 100$ | | |
| − 0.45 | 0.00276 | 0.00284 | 0.00308 | 0.00303 | 0.00324 |
| − 0.35 | 0.00279 | 0.00288 | 0.00308 | 0.00313 | 0.00341 |
| − 0.25 | 0.00273 | 0.00287 | 0.00301 | 0.00313 | 0.00348 |
| − 0.15 | 0.00267 | 0.00289 | 0.00299 | 0.00323 | 0.00354 |
| − 0.05 | 0.00271 | 0.00292 | 0.00300 | 0.00307 | 0.00337 |
| 0.05 | 0.00272 | 0.00293 | 0.00305 | 0.00322 | 0.00350 |
| 0.15 | 0.00245 | 0.00279 | 0.00293 | 0.00307 | 0.00337 |
| 0.25 | 0.00241 | 0.00278 | 0.00300 | 0.00306 | 0.00328 |
| 0.35 | 0.00213 | 0.00260 | 0.00300 | 0.00303 | 0.00318 |
| 0.45 | 0.00166 | 0.00227 | 0.00310 | 0.00317 | 0.00346 |
| | | | $T = 300$ | | |
| − 0.45 | 0.00155 | 0.00151 | 0.00163 | 0.00164 | 0.00180 |
| − 0.35 | 0.00161 | 0.00151 | 0.00160 | 0.00169 | 0.00181 |
| − 0.25 | 0.00148 | 0.00152 | 0.00154 | 0.00170 | 0.00175 |
| − 0.15 | 0.00155 | 0.00148 | 0.00150 | 0.00160 | 0.00173 |
| − 0.05 | 0.00145 | 0.00159 | 0.00161 | 0.00165 | 0.00189 |
| 0.05 | 0.00148 | 0.00155 | 0.00154 | 0.00162 | 0.00178 |
| 0.15 | 0.00144 | 0.00152 | 0.00155 | 0.00162 | 0.00178 |
| 0.25 | 0.00141 | 0.00150 | 0.00157 | 0.00164 | 0.00179 |
| 0.35 | 0.00140 | 0.00148 | 0.00161 | 0.00166 | 0.00182 |
| 0.45 | 0.00111 | 0.00126 | 0.00140 | 0.00142 | 0.00149 |
| | | | $T = 500$ | | |
| − 0.45 | 0.00113 | 0.00117 | 0.00122 | 0.00114 | 0.00130 |
| − 0.35 | 0.00114 | 0.00120 | 0.00125 | 0.00126 | 0.00136 |
| − 0.25 | 0.00114 | 0.00119 | 0.00123 | 0.00117 | 0.00136 |
| − 0.15 | 0.00109 | 0.00118 | 0.00118 | 0.00123 | 0.00138 |
| − 0.05 | 0.00114 | 0.00120 | 0.00116 | 0.00121 | 0.00136 |
| 0.05 | 0.00118 | 0.00123 | 0.00119 | 0.00132 | 0.00139 |
| 0.15 | 0.00114 | 0.00115 | 0.00117 | 0.00127 | 0.00144 |
| 0.25 | 0.00109 | 0.00117 | 0.00119 | 0.00123 | 0.00131 |
| 0.35 | 0.00105 | 0.00113 | 0.00121 | 0.00125 | 0.00137 |
| 0.45 | 0.00088 | 0.00096 | 0.00104 | 0.00109 | 0.00118 |

[a] $d$ = fractional differencing parameter; $T$ = sample size; $ML1$ = time-domain maximum likelihood, true mean removed; $ML1a$ = time-domain maximum likelihood, arithmetic mean removed; $ML2$ = frequency-domain maximum likelihood; $ML2a$ = frequency-domain maximum likelihood, taper $a$; $ML2b$ = frequency-domain maximum likelihood, taper $b$.

own right, and moreover, it is best to attempt a thorough understanding of the pure fractional noise before proceeding to more complex processes. The insights gained from its study are likely to provide useful guidance to behavior in more complex environments.[12]

Third, in accord with common practice, we use the arithmetic sample mean to estimate the population mean. The prospects for improving the performance of the feasible time-domain estimator by using alternative estimators of the mean are very limited. Samarov and Taqqu (1988) have shown analytically that, for a variety of sample sizes, the efficiency of the arithmetic mean estimator relative to Adenstedt's BLUE estimator is close to 100 percent over most of the parameter space and for a wide range of sample sizes. Similar results emerge from the Monte Carlo analyses of Mohr (1981) and Graf (1983).

Fourth, as correctly pointed out by a referee, it should be noted that the *tapered* frequency-domain estimators are not invariant to the mean of the process. In our Monte Carlo experiment, none of the data used in our frequency-domain procedures was de-meaned. This is irrelevant for the nontapered frequency-domain procedure, as we emphasized. However, it implicitly corresponds to centering the data used for the tapered frequency-domain procedures by the true mean, which is not known in practice. Had an estimated mean been removed, the tapered frequency-domain procedures probably would not have performed as well; the MSEs reported for the tapered frequency-domain estimators should therefore be viewed as lower bounds on the MSEs obtainable in practice. Because of the prohibitive cost involved and the fact that our attention here centers largely on *ML1a* vs. *ML2*, neither of which involves tapering, we have left detailed exploration of this issue to future research.

Finally, we note that inclusion of the determinant term in the frequency-domain Gaussian likelihood is likely to improve the finite-sample performance of the frequency-domain estimator *ML2*, as indicated by the results of Nerlove et al. (1979) for short-memory processes and Boes, Davis, and Gupta (1989) for long-memory processes. If this conjecture is true, then the prospects for improving the performance of *ML2* are *not* so limited (in contrast to those for *ML1a*), and our main result would be even stronger: The efficiency of a simple variation of *ML2* would be *even better* than that reported here for *ML2*.

## 5. Summary and conclusions

We have examined the finite-sample performance of ML estimators of the parameter of a pure fractionally-integrated process. We first showed that the efficiency of frequency-domain ML relative to time-domain ML is poor when the mean of the process is known; our results were in complete accord with those

[12] Cheung (1990) explores the cases of ARFIMA(1, $d$, 0), ARFIMA(0, $d$, 1), and ARFIMA(1, $d$, 1).

of Sowell (1992b). Time-domain ML with known mean is not a feasible estimator, however, whereas frequency-domain ML is. We therefore compared the finite-sample efficiency of the feasible time-domain ML estimator to the frequency-domain estimator. The comparison is of key importance, because it is the one relevant for applied work. The results were striking: the relative efficiency of approximate frequency-domain ML was drastically improved.

To determine the ultimate implications of our results for applied work, one must weigh costs and benefits. The feasible time-domain ML estimator, while much less efficient than its infeasible counterpart, nevertheless usually has somewhat lower MSE than the frequency-domain ML estimator for the sample sizes and parameter values examined here. (The biggest differences, of course, arise in the smallest samples.) Time-domain ML, however, requires tedious $(T\times T)$ covariance matrix inversion at each evaluation of the likelihood. Conversely, the frequency-domain ML estimator has the virtue of a very light computational burden.

One might be tempted to conclude that the lighter computational burden associated with frequency-domain ML more than offsets its slightly higher MSE. We do not necessarily agree. Today's powerful computing environment makes Sowell's exact time-domain estimator viable for the small/medium sample sizes in which it can really make a difference. The good news provided by this paper is that, for the medium/large sample sizes in which time-domain ML is likely to be prohibitively tedious (or impossible), frequency-domain ML is likely to perform very well.

## References


Adenstedt, R.K., 1974, On large-sample estimation for the mean of a stationary random sequence, Annals of Statistics 2, 1095–1107.

Boes, D.C., R.A. Davis, and S.N. Gupta, 1989, Parameter estimation in low order fractionally differenced ARMA processes, Stochastic Hydrology and Hydraulics 3, 97–110.

Cheung, Y.W., 1990, Long memory in foreign exchange rates and sampling properties of some statistical procedures related to long memory series, Ph.D. dissertation (University of Pennsylvania, Philadelphia, PA).

Cheng, B. and P.M. Robinson, 1992, Semiparametric estimation from time series with long range dependence, Manuscript (London School of Economics, London).

Cheung, Y.W., 1993, Long memory in foreign exchange rates, Journal of Business and Economic Statistics 11, 93–101.

Cheung, Y.W. and K.S. Lai, 1993, A fractional cointegration analysis of purchasing power parity, Journal of Business and Economic Statistics 11, 103–112.

Dahlhaus, R., 1988, Small sample effects in time series analysis: A new asymptotic theory and a new estimator, Annals of Statistics 16, 808–841.

Dahlhaus, R., 1989, Efficient parameter estimation for self-similar processes, 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.

Diebold, F.X. and G.D. Rudebusch, 1991, Is consumption too smooth? Long memory and the Deaton paradox, Review of Economics and Statistics 71, 1–9.

Diebold, F.X., S. Husted, and M. Rush, 1991, Real exchange rates under the gold standard, Journal of Political Economy 99, 1252–1271.
Fox, R. and M.S. Taqqu, 1986, Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, 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.
Gourieroux, C., F. Maurel, and A. Monfort, 1987, Regression and nonstationarity, Working paper no. 8708 (INSEE, Paris).
Graf, H.P., 1983, Long-range correlations and estimation of the self-similarity parameter, Ph.D. dissertation (Swiss Federal Institute of Technology, Zurich).
Granger, C.W.J. and R. Joyeux, 1980, An introduction to long-memory time series models and fractional differencing, Journal of Time Series Analysis 1, 15–39.
Harvey, A.C., 1989, Forecasting, structural time series models and the Kalman filter (Cambridge University Press, Cambridge).
Hassett, K., 1990, Is the aggregate labor market exploding?, Manuscript (Graduate School of Business, Columbia University, New York, NY).
Hosking, J.R.M., 1981, Fractional differencing, Biometrika 68, 165–176.
Li, W.K. and A.E. McLeod, 1986, Fractional time series modelling, Biometrika 73, 217–221.
Lo, A.W., 1991, Long-term memory in stock market prices, Econometrica 59, 1279–1314.
Mohr, D.L., 1981, Modeling data as a fractional gaussian noise, Ph.D. dissertation (Princeton University, Princeton, NJ).
Nerlove, M., D.M. Grether, and J.L. Carvalho, 1979, Analysis of economic time series: A synthesis (Academic Press, New York, NY).
Priestley, M.B., 1980, Spectral analysis and time series (Academic Press, New York, NY).
Robinson, P.M., 1988, Nonparametric function estimation for long-memory time series, Manuscript (London School of Economics, London).
Robinson, P.M., 1990, Time series with strong dependence, Manuscript (London School of Economics, London).
Robinson, P.M., 1991, Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression, Journal of Econometrics 47, 67–84.
Samarov, A. and M.S. Taqqu, 1988, On the efficiency of the sample mean in long-memory noise, Journal of Time Series Analysis 9, 191–200.
Shea, G.S., 1991, Uncertainty and implied variance bounds for the expectations hypothesis of the interest rate term structure, Empirical Economics 16, 287–312.
Sowell, F.B., 1990, Fractional unit-root distribution, Econometrica 58, 495–506.
Sowell, F.B., 1992a, Characterizing long-run behavior, Journal of Monetary Economics 29, 277–302.
Sowell, F.B., 1992b, Maximum likelihood estimation of stationary univariate fractionally-integrated time-series models, Journal of Econometrics 53, 165–188.
Whittle, P., 1951, Hypothesis testing in time series (Almqvist and Wiksell, Uppsala).
Yajima, Y., 1985, On estimation of long-memory time series models, Australian Journal of Statistics 27, 303–320.
Yajima, Y., 1988, On estimation of a regression model with long-memory stationary errors, Annals of Statistics 16, 791–807.